Methods in
Molecular Biology 2263

Springer Protocols

Tina Daviter
Christopher M. Johnson
Stephen H. McLaughlin
Mark A. Williams *Editors*

Protein-Ligand Interactions

Methods and Applications

Third Edition

Humana Press

METHODS IN MOLECULAR BIOLOGY

For further volumes:
http://www.springer.com/series/7651

For over 35 years, biological scientists have come to rely on the research protocols and methodologies in the critically acclaimed *Methods in Molecular Biology* series. The series was the first to introduce the step-by-step protocols approach that has become the standard in all biomedical protocol publishing. Each protocol is provided in readily-reproducible step-by-step fashion, opening with an introductory overview, a list of the materials and reagents needed to complete the experiment, and followed by a detailed procedure that is supported with a helpful notes section offering tips and tricks of the trade as well as troubleshooting advice. These hallmark features were introduced by series editor Dr. John Walker and constitute the key ingredient in each and every volume of the *Methods in Molecular Biology* series. Tested and trusted, comprehensive and reliable, all protocols from the series are indexed in PubMed.

Protein-Ligand Interactions

Methods and Applications

Third Edition

Edited by

Tina Daviter

The Institute of Cancer Research, London, UK

Christopher M. Johnson

MRC Laboratory of Molecular Biology, Cambridge, UK

Stephen H. McLaughlin

MRC Laboratory of Molecular Biology, Cambridge, UK

Mark A. Williams

ISMB BiophysX Centre, Birkbeck, University of London, London, UK

Humana Press

Editors
Tina Daviter
The Institute of Cancer Research
London, UK

Christopher M. Johnson
MRC Laboratory of Molecular Biology
Cambridge, UK

Stephen H. McLaughlin
MRC Laboratory of Molecular Biology
Cambridge, UK

Mark A. Williams
ISMB BiophysX Centre, Birkbeck
University of London
London, UK

ISSN 1064-3745 ISSN 1940-6029 (electronic)
Methods in Molecular Biology
ISBN 978-1-0716-1196-8 ISBN 978-1-0716-1197-5 (eBook)
https://doi.org/10.1007/978-1-0716-1197-5

Front cover: Interaction of initiator tRNA (teal) and initiation factors IF1 (purple) and IF3 (orange) with the small ribosomal subunit (30S) from structural work on the pre-initiation complexes determined in the lab of Venki Ramakrishnan. Copyright Visual Aids, MRC Laboratory of Molecular Biology, Cambridge.

This Humana imprint is published by the registered company Springer Science+Business Media, LLC, part of Springer Nature.
The registered company address is: 1 New York Plaza, New York, NY 10004, U.S.A.

Preface

Proteins are central in life at the molecular level. They are catalysts, transporters, and messengers. They provide or create much of the cellular architecture and go beyond the cell to seek nutrients and neutralize antigens. They each function in large interaction networks with other proteins, small molecules, and the genome: dynamically binding, unbinding, competing for and changing partners, or forming functional multimolecular assemblies. All these interactions are intricately regulated, and their dysregulation often results in disease. One way of understanding the detail of the function and regulation of proteins is to dissect and measure their interactions.

As in previous editions, this volume aims to provide an overview of common methods for characterizing the interactions of proteins. This edition does not replace previous editions, but seeks to supplement them with new methods and perspectives. Unlike the previous editions, which were structured by types of method, this edition is organized by the types of ligand with which proteins interact. We hope that this will make the book even more useful to researchers new to these techniques of biophysical investigation by suggesting approaches suited to particular biological problems from the great, and perhaps confusing, range of possibilities.

Many of our authors lead research facilities in academic or pharmaceutical settings and have considerable experience of making their instrumentation and expertise available to the wider research community. Chapters provide clear instructions for the nonspecialist, for graduate students, or researchers new to a given approach and cover all stages of the research process, from the screening and discovery of interactions to the detailed quantitative understanding of the mechanisms involved.

London, UK — *Tina Daviter*
Cambridge, UK — *Christopher M. Johnson*
Cambridge, UK — *Stephen H. McLaughlin*
London, UK — *Mark A. Williams*

Contents

Part IV Nucleotide Binding and Hydrolysis

Part V Binding Nucleic Acids

Part VI Membrane Binding

Contributors

Guillaume Bec • *Institut de Biologie Moléculaire et Cellulaire, CNRS, Université de Strasbourg, Strasbourg, France*

Holly L. Birchenough • *Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK; Wellcome Trust Centre for Cell Matrix Research, University of Manchester, Manchester, UK*

Andrea Bogutzki • *Hannover Medical School, Institute for Biophysical Chemistry, Hannover, Germany*

Harland E. Brandon • *Department of Chemistry and Biochemistry, Alberta RNA Research and Training Institute (ARRTI), University of Lethbridge, Lethbridge, AB, Canada*

Chad A. Brautigam • *Departments of Biophysics and Microbiology, UT Southwestern Medical Center, Dallas, TX, USA*

Sébastien Brûlé • *Plate-forme de Biophysique Moléculaire, Institut Pasteur, UMR 3528 CNRS, Paris, France*

John E. Burke • *Department of Biochemistry and Microbiology, University of Victoria, Victoria, BC, Canada; Department of Biochemistry and Molecular Biology, The University of British Columbia, Vancouver, BC, Canada*

Ute Curth • *Hannover Medical School, Institute for Biophysical Chemistry, Hannover, Germany*

Gareth Davies • *Hit Discovery, Discovery Sciences, R&D BioPharmaceuticals, AstraZeneca, Alderley Park, UK*

Ben J. Davis • *Vernalis Research, Cambridge, UK*

Sarah R. Dennison • *School of Pharmacy and Biomedical Sciences, University of Central Lancashire, Preston, UK*

Eric Ennifar • *Institut de Biologie Moléculaire et Cellulaire, CNRS, Université de Strasbourg, Strasbourg, France*

Ewa Folta-Stogniew • *W.M. Keck Biotechnology Resource Laboratory, Department of Molecular Biophysics and Biochemistry, Yale University School of Medicine, New Haven, CT, USA*

David J. Hamelin • *Department of Biochemistry and Microbiology, University of Victoria, Victoria, BC, Canada*

Paul E. Hemsley • *AstraZeneca R&D Discovery Sciences, Cambridge, UK*

Matthew R. Hicks • *Linear Diagnostics Limited, The BioHub, Birmingham, UK*

Reece M. Hoffmann • *Department of Biochemistry and Microbiology, University of Victoria, Victoria, BC, Canada*

Geoffrey A. Holdgate • *Hit Discovery, Discovery Sciences, R&D BioPharmaceuticals, AstraZeneca, Alderley Park, UK*

Per Jemth • *Department of Medical Biochemistry and Microbiology, BMC, Uppsala University, Uppsala, Sweden*

Christopher M. Johnson • *MRC Laboratory of Molecular Biology, Cambridge, UK*

Thomas A. Jowitt • *Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK; Wellcome Trust Centre for Cell Matrix Research, University of Manchester, Manchester, UK*

Elin Karlsson • *Department of Medical Biochemistry and Microbiology, BMC, Uppsala University, Uppsala, Sweden*
Stefan H. Knauer • *Biopolymers, University of Bayreuth, Bayreuth, Germany*
Simone Kunzelmann • *Structural Biology Science Technology Platform, The Francis Crick Institute, London, UK*
David M. Lawson • *Department of Biological Chemistry, John Innes Centre, Norwich Science Park, Norwich, UK*
Xiaochun Li-Blatter • *Biophysics Facility, Biozentrum, University of Basel, Basel, Switzerland*
Stephen R. Martin • *Structural Biology Science Technology Platform, The Francis Crick Institute, London, UK*
Laura Masino • *Structural Biology Science Technology Platform, The Francis Crick Institute, London, UK*
Benoit Meyer • *Institut de Biologie Moléculaire et Cellulaire, CNRS, Université de Strasbourg, Strasbourg, France*
David Murray • *Hit Discovery, Discovery Sciences, R&D BioPharmaceuticals, AstraZeneca, Alderley Park, UK*
Adewale Olamoyesan • *Department of Molecular Sciences, Macquarie University, Sydney, NSW, Australia*
Lori A. Passmore • *MRC Laboratory of Molecular Biology, Cambridge, UK*
Helen Plant • *Hit Discovery, Discovery Sciences, R&D BioPharmaceuticals, AstraZeneca, Alderley Park, UK*
Andres Ramos • *Department of Structural & Molecular Biology, University College London, London, UK*
Bertrand Raynal • *Plate-forme de Biophysique Moléculaire, Institut Pasteur, UMR 3528 CNRS, Paris, France*
Alison Rodger • *Department of Molecular Sciences, Macquarie University, Sydney, NSW, Australia*
Taiichi Sakamoto • *Institut de Biologie Moléculaire et Cellulaire, CNRS, Université de Strasbourg, Strasbourg, France; Department of Life Science, Faculty of Advanced Engineering, Chiba Institute of Technology, Chiba, Japan*
Emma Schenckbecher • *Institut de Biologie Moléculaire et Cellulaire, CNRS, Université de Strasbourg, Strasbourg, France*
Maria J. Schilstra • *Structural Biology Technology Platform, The Francis Crick Institute, London, UK*
Maximilian Seidel • *MRC Laboratory of Molecular Biology, Cambridge, UK*
Hannah Semple • *Hit Discovery, Discovery Sciences, R&D BioPharmaceuticals, AstraZeneca, Alderley Park, UK*
Timothy Sharpe • *Biophysics Facility, Biozentrum, University of Basel, Basel, Switzerland*
Joseph Shaw • *Discovery Biology, Discovery Sciences, IMED Biotech Unit, AstraZeneca, Cambridge, UK*
Ian Sinclair • *Hit Discovery, Discovery Sciences, R&D BioPharmaceuticals, AstraZeneca, Alderley Park, UK*
Jordan T. B. Stariha • *Department of Biochemistry and Microbiology, University of Victoria, Victoria, BC, Canada*
Clare E. M. Stevenson • *Department of Biological Chemistry, John Innes Centre, Norwich Research Park, Norwich, UK*
James A. W. Stowell • *MRC Laboratory of Molecular Biology, Cambridge, UK*

Christopher Stubbs • *Discovery Biology, Discovery Sciences, IMED Biotech Unit, AstraZeneca, Cambridge, UK*
Terence T. L. Tang • *MRC Laboratory of Molecular Biology, Cambridge, UK*
Shih-Chia Tso • *Departments of Biophysics and Microbiology, UT Southwestern Medical Center, Dallas, TX, USA*
Andrew P. Turnbull • *Cancer Research UK Therapeutic Discovery Laboratories, London Bioscience Innovation Centre, London, UK*
Stephan Uebel • *Max-Planck-Institut für Biochemie, Martinsried, Germany*
Hans-Joachim Wieden • *Department of Chemistry and Biochemistry, Alberta RNA Research and Training Institute (ARRTI), University of Lethbridge, Lethbridge, AB, Canada*
Xiaoqiu Wu • *Advanced Drug Delivery, Pharmaceutical Sciences, BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden*
Ludovit Zweifel • *Biophysics Facility, Biozentrum, University of Basel, Basel, Switzerland*

Part I

Overview

Chapter 1

Assessing and Improving Protein Sample Quality

Bertrand Raynal, Sébastien Brûlé, Stephan Uebel, and Stefan H. Knauer

Abstract

One essential prerequisite of any experiment involving a purified protein, such as interaction studies or structural and biophysical characterization, is to work with a "good-quality" sample in order to ensure reproducibility and reliability of the data. Here, we define a "good-quality" sample as a protein preparation that fulfills three criteria: (1) the preparation contains a protein that is pure and soluble and exhibits structural and functional integrity, (2) the protein must be structurally homogeneous, and (3) the preparation must be reproducible. To ensure effective quality control (QC) of all these parameters, we suggest to follow a simple workflow involving the use of gel electrophoresis, light scattering, and spectroscopic experiments. We describe the techniques used in every step of this workflow and provide easy-to-use standard protocols for each step.

Key words Purity, Homogeneity, Identity, Oligomeric state, Structural integrity, Protein stability, Optimization of storage conditions, Batch-to-batch consistency

1 Introduction

The overall quality control (QC) workflow consists of a series of experiments that are intended to assess the purity, structural homogeneity, and stability of a protein sample (Fig. 1). The data obtained can demonstrate, inasmuch as is reasonably practicable, that the sample is of good quality, can be used to verify batch-to-batch consistency of protein preparations, or, where relevant criteria are not met, can be used to guide iterative improvement of the quality of protein samples.

Prior to any experimental work, it is essential to gather all information about the sample that is necessary for the setup of experiments and data analysis. Researchers should know the full amino acid sequence of their target protein (including tags, additional residues from cleavage sites, etc.) and the identity of all other known chemical components in the sample (i.e., those derived from its preparative process). Full information on the safety of the protein and other components of the sample should be obtained as this can have an impact on the practical performance of experiments.

Tina Daviter et al. (eds.), *Protein-Ligand Interactions: Methods and Applications*, Methods in Molecular Biology, vol. 2263, https://doi.org/10.1007/978-1-0716-1197-5_1,

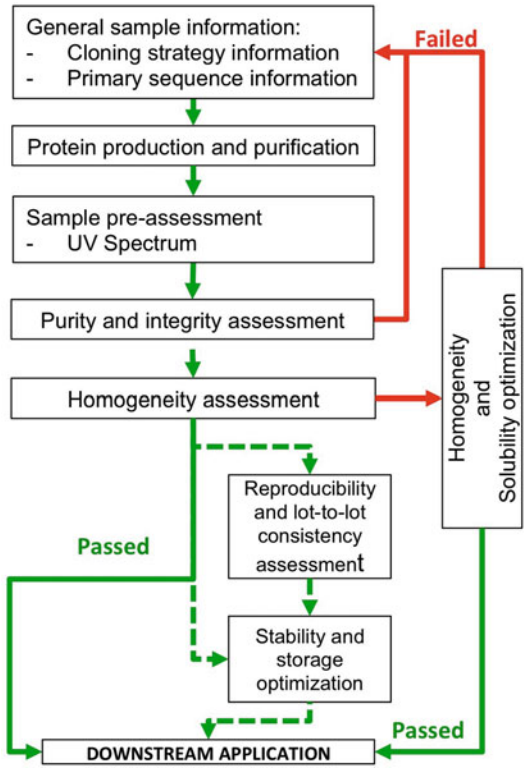


Fig. 1 Protein quality control workflow. The workflow has to be followed step-by-step starting with the essential preassessment. For each step, there is an associated section in this chapter. Achievement of relevant QC criteria is indicated by green arrows (passed), whereas failure is indicated by red arrows (failed) with the following step in the workflow to be applied to improve the situation. The green dotted arrows indicate optional measurements

Knowledge of the expression and purification protocols and prior storage conditions of the sample are very often also important in interpreting experimental results and improving protein quality.

1.1 Preassessment of Purity and Concentration Determination by Ultraviolet Spectroscopy

Proteins exhibit a characteristic ultraviolet (UV) absorption spectrum (Fig. 2) between 200 and 350 nm with most proteins having a single maximum around 280 nm caused mainly by the aromatic amino acids tyrosine (Tyr) and tryptophan (Trp). UV absorbance at 280 nm is routinely used to determine the protein concentration via the Beer–Lambert law. However, this method will give inaccurate or erroneous results if the protein has no or only a few aromatic amino acids or if it contains UV-absorbing nonprotein components such as bound nucleic acids, nucleotide cofactors, porphyrins, or iron–sulfur centers (*see* **Note 1**). In these cases, the absorbance at 205 or 214 nm, which is largely due to peptide bonds present in all proteins, can be used instead to determine the protein

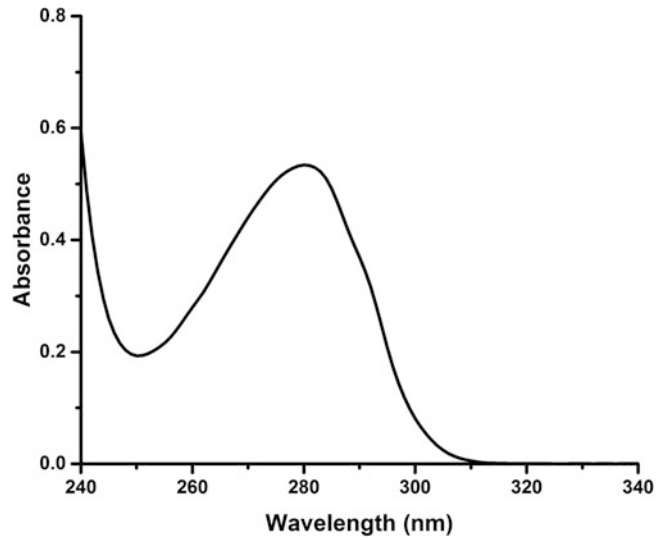


Fig. 2 UV spectrum of a protein. The absence of an absorbance signal >320 nm and the A_{260}/A_{280} ratio of 0.6 show the good quality of the sample (i.e., this sample shows no sign of aggregation or contamination). Measurements were performed in a 1-cm quartz cuvette

concentration [1, 2], as long as the buffer does not absorb light significantly in this wavelength regime.

The absorbance measured at 280, 214, or 205 nm can be used to calculate the sample concentration employing the protein molar absorption coefficient at the working wavelength and the Beer–Lambert law.

$$A_i = \varepsilon_i \cdot c \cdot d \tag{1}$$

with A_i being the absorption at wavelength i, ε_i the molar absorption coefficient at that wavelength, c the concentration (M), and d the cuvette path length (cm).

Independent of the wavelength used for concentration determination, the sample must be pure, i.e., not contain other protein(s) as contaminants, as such impurities will falsify the concentration. The advantages of UV absorbance-based protein quantification are that (1) the sample can be recovered, (2) it is accurate as long as the extinction coefficient is known, and (3) there is less variability as compared to colorimetric reactions as the latter are strongly dependent on the protein composition, which will be different from the one used as a standard [3].

In addition to the determination of the concentration, UV spectroscopy is a very convenient tool for detecting nonprotein contaminants. Firstly, nucleic acids have an absorption maximum at 260 nm and reducing agents (especially dithiothreitol (DTT)) [2, 4] at 250 nm. Thus, both kinds of contaminants will alter the shape of the 280 nm absorbance peak of the sample, resulting in a

high A_{260}/A_{280} ratio (>0.6) [5]. Secondly, UV spectroscopy can detect the presence of large particles. Aggregate-free protein samples do not absorb light at wavelengths >320 nm. Thus, an absorbance signal above 320 nm can be attributed exclusively to the scattering of light by large aggregates (hydrodynamic radius larger than 200 nm) present in the sample.

In order to assess aggregation, the aggregation index (AI) can be calculated as

$$\mathrm{AI} = \frac{100 \cdot A_{340}}{A_{280} - A_{340}} \tag{2}$$

with A_{280} and A_{340} being the absorbance signals at 280 nm and 340 nm, respectively. As a rule of thumb, the AI should have a value lower than 2 for a homogeneous sample without aggregation. This simple measurement can quickly provide qualitative information about aggregation in the sample, which is followed up using more quantitative techniques (Subheading 1.2.2).

1.2 Assessing Protein Purity, Homogeneity, and Oligomeric State

Assessing the purity of a protein sample means detecting and visualizing impurities that it contains. It is a necessary control during the sample production process and is one of the most essential QC checks as most biophysical and structural biology experiments require high sample purity. It is necessary to choose a visualization method with sufficient sensitivity to detect contamination at the lowest level that may be detrimental to the interpretation of subsequent functional studies of the sample.

1.2.1 Assessing Purity

Electrophoretic methods are the most common tests to check the purity of a protein sample [6, 7]. In gel electrophoresis, proteins are run through agarose or polyacrylamide (PA) gels in an electric field. The electrophoretic mobility of proteins, i.e., how they migrate through the gel, is, depending on the method, determined by their charge, size, and shape. Agarose gels have relatively large pores and are thus ideal to separate proteins >500 kDa or DNA. Here we present the most common PA-based gel electrophoresis methods that are better suited to most protein applications, but also note that alternatives such as capillary electrophoresis may be used for protein QC. Subsequent staining of the proteins reveals their position within the gel, with the staining method being chosen based on the required detection limit. Adaption of protocols for membrane proteins is possible in most cases.

Discontinuous, denaturing *sodium dodecyl sulfate polyacrylamide gel electrophoresis* (SDS-PAGE) is the primary method of choice and was first described by Laemmli [8]. It is simple, fast, and economic and thus often used to follow the purification progress during sample production. A reducing and denaturing sample buffer containing SDS, an anionic detergent, is added to the sample, followed by an incubation at >95 °C for a few minutes.

Proteins are denatured and disulfide bridges are broken so that the unfolded monomers bind the negatively charged SDS homogeneously along the amino acid chain, abolishing the intrinsic charge of the proteins and resulting in ellipsoid micelles of a constant negative charge proportional to the protein length (1.4 g SDS are bound per 1 g protein). Thus, migration during electrophoresis mostly depends on the proteins' size in their unfolded state and one protein will give just one band. In discontinuous electrophoresis systems, the gel matrix contains two layers: a stacking gel with neutral pH and large pores to enable the concentration of the proteins, and a resolving/separating gel with basic pH and small pores to separate the proteins. Subsequent staining visualizes the proteins that have migrated as separate bands. Many staining methods are available, which differ in their sensitivity and thus their detection limit (Subheading 2.2.1). Appropriate staining is the most crucial aspect in detecting contaminants by gel electrophoresis. A (commercially available) size marker is usually loaded in one lane, allowing the estimation of the molecular weight (MW) of the proteins in the sample by comparison of positions. This molecular weight estimate can be made more accurate by careful preparation of the gel and subsequent quantitative analysis of the protein positions [*see* **Note 2**].

In *native PAGE*, the sample buffer does not contain SDS, proteins remain in their native state, and complexes (e.g., protein: protein or protein:nucleic acid) are not disrupted if their affinity is high enough. The electrophoretic mobility of the proteins/protein complexes depends on shape, charge, and molecular weight. Hence, in addition to purity, structural homogeneity of the sample can be assessed and the molecular weight of complexes may be estimated via comparison with proteins/complexes of similar shape and known molecular weight. Here, however, molecular weight standards allow only a rough approximation due to inevitable differences in shape. As no SDS is present, buffers must be chosen carefully to ensure that proteins/protein complexes are charged (i.e., that the pH is away from the protein's isoelectric point) and will migrate during electrophoresis. Usually, in native PAGE, both the pH and the acrylamide concentration have to be adapted and optimized for each sample. The variant method of *blue native PAGE* is often used for the analysis of membrane proteins. Here negatively charged Coomassie Brilliant Blue dye is present in the running buffer and binds to hydrophobic patches of the proteins, replacing the detergent without denaturing them (although this detergent effect can itself cause the dissociation of complexes). For reviews on the significance, methods, and protocols for native gel analysis of protein complexes as well as membrane protein complexes, *see* Refs. 9 and 10.

1.2.2 Assessing Homogeneity

Depending on the context, the term "homogeneity" describes different intrinsic properties of the sample. For example, researchers can question, if the protein is monomeric or oligomeric, if there is only one conformational state, if there are one or many states of assemblies, if all proteins carry the same posttranslational modifications (PTMs), or if the sample contains soluble aggregates. No technique will answer definitively all these questions in a single experiment. In protein QC, "homogeneity" normally means that the protein or protein complex is monodisperse, i.e., in one stable assembly of fixed stoichiometry, and that no aggregates are present.

Due to its speed and low sample consumption, *dynamic light scattering (DLS)* is a very convenient method to assess simultaneously the apparent monodispersity of the species of interest and the presence of soluble high-order assemblies and aggregates [11]. The phenomenon of DLS arises because of Brownian motion, the extent of which is related to the size of the particles. When light from a laser hits small particles, the light is scattered in all directions, and the scattering intensity at a particular location fluctuates over time due to the Brownian motion of the particles. The autocorrelation curve of the scattering intensity provides information on the particles' motion—the scattering intensity will fluctuate, and the autocorrelation will decay, more slowly for larger particles. Using the translational diffusion deduced from this motion, a hydrodynamic radius R_h (typically a Stokes radius, i.e., the radius of a sphere that would diffuse with the same rate as that observed) can be calculated for each species present [3] (Fig. 3). Light scattering intensity is proportional to the mass squared. Consequently, DLS is the method of choice to detect small quantities of large aggregates in a sample. However, a key weakness of DLS is that it does not have high mass resolution and will typically not distinguish between monomeric and dimeric forms. In order to distinguish separate species correctly, their radii must differ approximately twofold (which corresponds to an approximately tenfold difference in molecular weight for globular proteins). DLS only allows the determination of the hypothetical radii and not the actual mass of the protein. Thus, one should be aware that any indication of molecular weight given by instrument software is based on theoretical calculations assuming a certain shape and protein density.

Because of this lack in resolution, DLS measurements can be complemented by *analytical size-exclusion chromatography* (SEC), which is currently the standard column-based chromatography technique to quantify protein oligomers and, like DLS, a hydrodynamic technique [12]. The chromatography column contains a matrix with many fine pores of various sizes and SEC separates molecules according to their size because molecular species with a greater radius can enter fewer pores and thus take a shorter path and elute before smaller ones from the column. Aggregates, contaminants, and potentially different oligomeric states of the protein

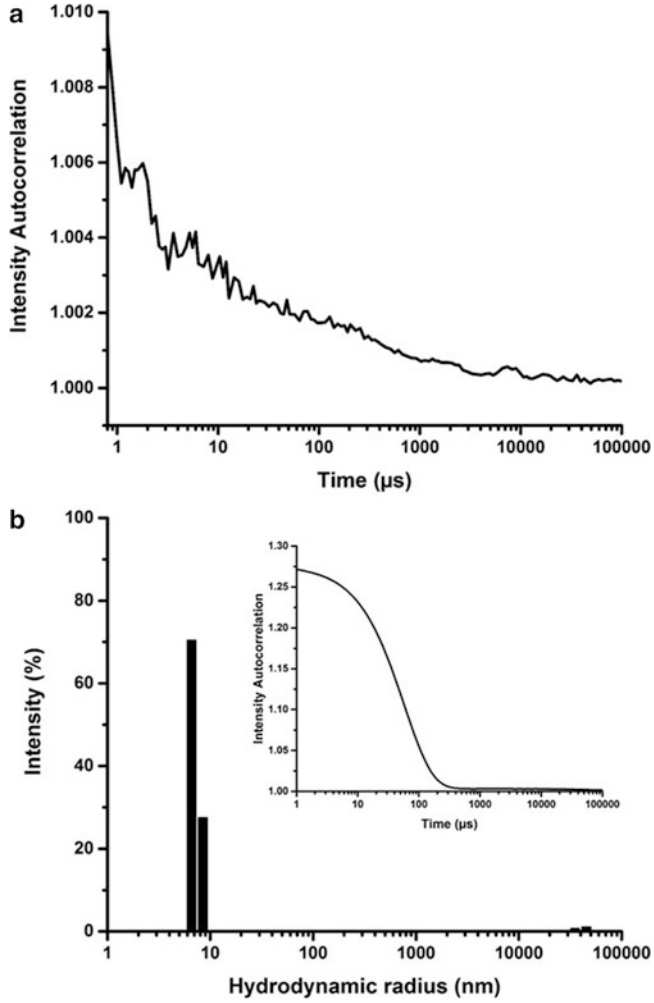


Fig. 3 Typical DLS autocorrelation curve and analysis. (**a**) Autocorrelation curve of a pure buffer. Notice the small and noisy signal due to little scattering and the sharp decrease at early correlation time indicative of very small particles. (**b**) Analysis of the distribution of a protein sample by DLS. The single peak is indicative of a homogeneous sample. (Inset) The autocorrelation curve correspondingly shows a low-noise smooth decay

of interest can be readily separated and quantified via inline concentration detection. However, although the gel filtration resins are intended to be "inert," some proteins interact with them, falsifying results (or even making SEC impossible). One should also keep in mind that the protein sample will be significantly diluted during the analytical SEC (at least by a factor of 2), which might alter the equilibria between any oligomeric species present.

In order to determine the absolute molecular weight of each eluted species, an inline *static light scattering* (SLS) detector [13], in combination with a UV or a refractive index (RI) detector, must

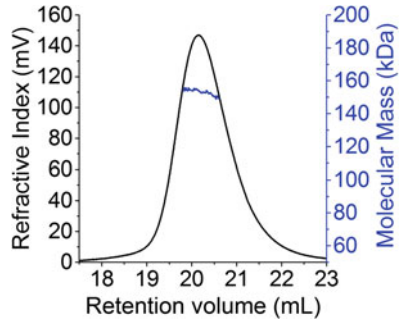


Fig. 4 Typical SEC-SLS elution profile and analysis. The black line represents the elution profile detected by refractometry. The blue square represents the calculated molecular mass along the peak

be coupled to the SEC system (SEC-SLS). A typical elution profile is presented in Fig. 4. Currently, different types of SLS systems are available on the market, e.g., *multiangle light scattering* (MALS) detectors and *low-angle light scattering* (LALS) detectors. MALS detectors measure the intensity of the scattered light at different angles θ followed by extrapolation of the intensity to $\theta = 0°$, which allows calculation of the radius of gyration (R_g) of the protein. The main reason for extrapolation is to compensate for the angular dependence of light scattering intensity for molecules larger than 1/20th of the laser wavelength (i.e., R_g = 10–15 nm). LALS detectors measure the scattered light at a small angle (e.g., 7° for some systems), assuming that the difference in intensity between 7° and 0° is negligible, even for megadalton molecules. Thus, the main difference between MALS and LALS detectors is the way they obtain light scattering intensity at $\theta = 0°$.

1.2.3 Assessing Identity and Chemical Integrity

Mass spectrometry (MS) has become an indispensable technique for protein QC (*see* **Note 3**) and is also widely used as a tool complementing biophysical techniques for the study of protein complexes and protein:ligand interactions [14]. Many quality-relevant aspects of a protein affect molecular mass, e.g., primary sequence, proteolytic degradation, and PTMs. The high mass accuracy that can be achieved by MS techniques (in the range of 1 Da) allows sensitive detection of PTMs, such as phosphorylation (≈+80 Da), or proteolytic processing that would evade detection in gel electrophoresis.

Only with the establishment of *electrospray ionization* (ESI) [15] and *matrix-assisted laser desorption ionization* (MALDI) [16] in the late 1980s it was possible to ionize and desolvate large biomolecules reliably. These soft ionization processes transfer only a moderate excess of energy to the analyte, which is crucial for fragile and nonvolatile compounds such as proteins. Ionization of proteins and peptides by ESI results usually in ions of a number of charge states (Fig. 5a) due to different protonation states of the

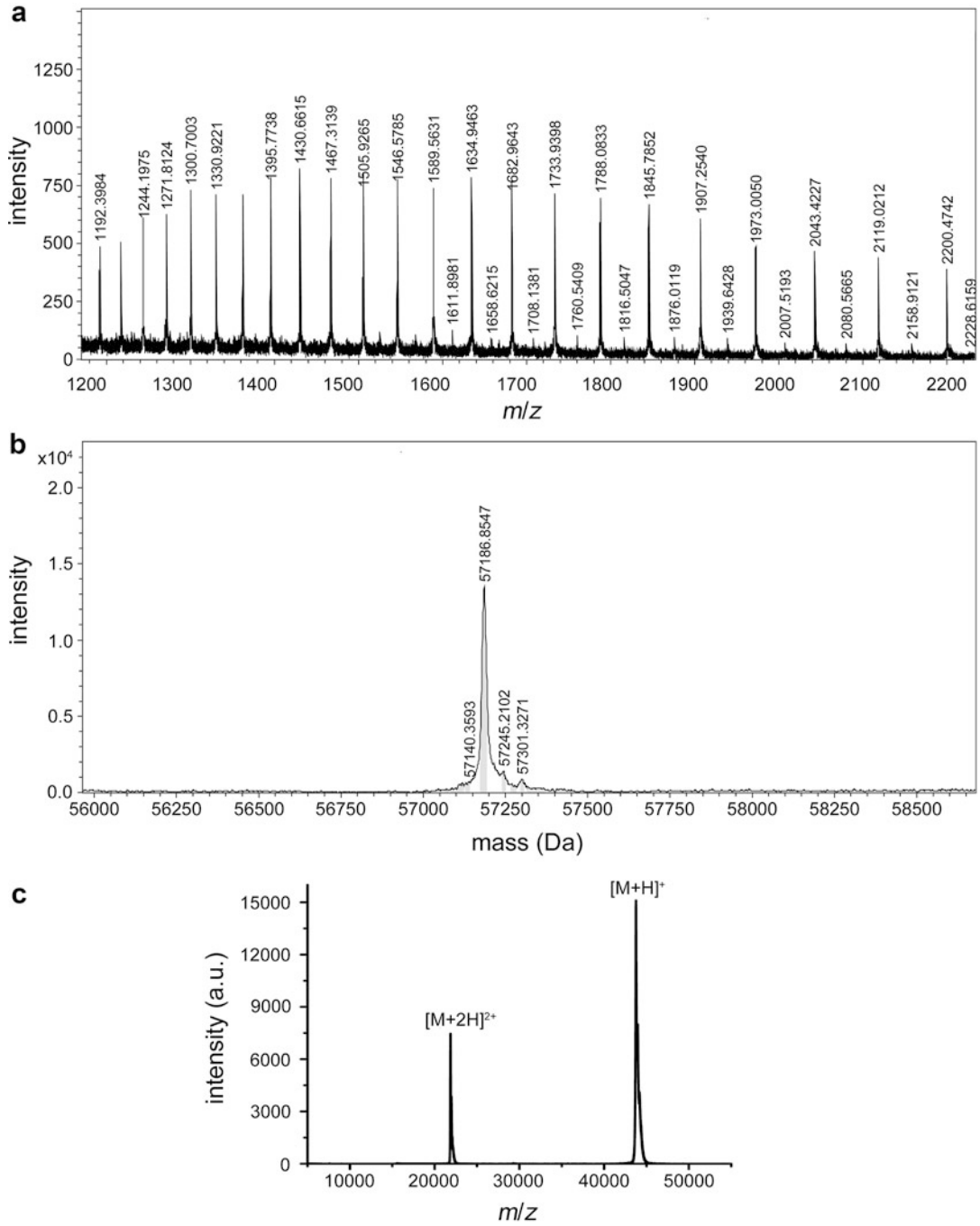


Fig. 5 MS spectra. (**a**) MS spectrum of proteins and peptides generated using an ESI ionization source and analysis in positive mode resulting in ions representing a number of charge states due to different protonation states of the molecule. (**b**) Deconvolution of the distribution of charge states of (**a**) to give a molecular mass. (**c**) MS spectrum generated using a MALDI ionization source, resulting in doubly and singly charged ion species

molecule. This distribution of charge states, each corresponding to a particular mass/charge ratio, is then deconvoluted to give a molecular mass (Fig. 5b). In contrast, MALDI is an ionization process from a solid phase. The protein of interest is dispersed in a saturated solution of a small aromatic molecule called the matrix and the whole is co-crystallized by evaporation of the solvent. The solid phase obtained is then irradiated by a laser beam. The role of the matrix is to absorb the energy of the laser beam, which leads to transfer of the sample into the gas phase and subsequent desolvation. In MALDI, peptides, and even proteins, often result in only a singly charged ion species, or at least a considerably lower number of ion species with fewer charges as compared to ESI (Fig. 5c). One of the main advantages of MALDI is that the sample droplets can be prepared rapidly, allowing the analysis of a large number of soluble or membrane proteins in a very short time, making it particularly suited for QC screening (*see* Subheading 2.2.4).

A mass spectrometer consists of three main parts: (1) the ion source, where analytes are ionized and transferred into the gas phase and subsequently into the vacuum inside the instrument, (2) a mass analyzer that separates the analytes according to their m/z ratio (m being the mass of the analyte; z being the charge the ionized analyte carries; separation in space, frequency space, or speed), and (3) the detector registering the ions. It is worth noting that mass spectrometers do not measure mass per se but the m/z ratio.

For protein QC, an *intact mass MS* analysis (top-down strategy) is highly recommended to check the identity and integrity of the protein of interest. This QC parameter is crucial to ensure that the target protein is actually the desired protein (i.e., that there has not been some mistake during cloning) and is highly pure and without any degradation at the N- or C-terminus (which may result from contaminating peptidase enzymes not being removed by the purification process). The precision of intact mass MS of big proteins (e.g., bovine serum albumin, 66.4 kDa) is better with an ESI than with a MALDI source. However, buffer components such as detergents or glycerol can easily interfere with the ESI measurement and make the analysis of membrane proteins or proteins in glycerol more challenging than with a MALDI source. Other potential chemical modifications of a protein that may degrade its functionality, such as deamination and oxidation, aspartic acid isomerization, and nonreducible crosslinking, can also be identified using MS [17]. Ideally, to rule out degradation later in the process, the intact mass should be checked at different timepoints, such as immediately postpurification and at the conclusion of downstream experiments.

Alternatively, a peptide mass fingerprint (bottom-up strategy) can identify and characterize the amino acid sequence of the target

protein and detect posttranslational modifications. This approach is possible with both a MALDI and an ESI ionization source and is based on the analysis of peptide fragments generated by proteolytic digestion of the target protein. Trypsin is often used for this purpose as it has good specificity, produces homogeneous peptide fragments, and can be applied to proteins that have been analyzed by gel electrophoresis and stained [17] (see also Subheading 2.2.1). Before cleavage of the protein with trypsin, disulfide bonds are reduced and free cysteines are alkylated. After digestion, the peptides are extracted and desalted. Overall, a peptide mass fingerprint takes approximately 24 h from sample preparation to data acquisition. Note, however, that in most cases terminal truncations cannot be detected.

1.3 Assessing Structural Integrity

Typically, proteins adopt a defined 3D structure in their physiological/native state that is essential for their function. It is critical to assess the structural integrity, i.e., that each batch of a protein adopts the same (average) structure before using it in downstream applications. In general, the 3D structure of a protein may depend on solution conditions, such as buffer, pH, salt concentration, reducing agents, or detergents, and it is important to examine this possibility. The most common methods to assess the structural state of a protein are *circular dichroism (CD) spectroscopy* and *nuclear magnetic resonance (NMR) spectroscopy*. Both methods can clearly distinguish folded compact structures from unfolded structures and provide a characteristic spectrum of a protein, which can be used to detect structural variation by comparison. The characteristic spectrum of a protein allows the assessment of the folding state of a protein, comparison of the structure of a protein from different sources, production processes or batches (*see* Subheading 1.5), and potential differences between protein variants, as well as studying protein stability by thermal or chemical denaturation, characterizing conformational changes, and analyzing protein:ligand interactions [18–21]. Introduction to protein NMR spectroscopy is beyond the scope of this chapter and can be found elsewhere [22, 23]. However, we note that, if NMR spectroscopy is available, it is highly recommended to include it in the QC workflow for small-to-medium-sized proteins (<30 kDa). The simplest NMR experiment, a one-dimensional proton NMR spectrum, may be recorded in a few minutes and not only allows assessment of folding but also detection of the presence of low-molecular-weight impurities or aggregation. Two-dimensional spectra, which are readily obtainable for proteins <30 kDa by ^{15}N isotope labeling, can provide information on the local chemical environment of all individual residues. The residue chemical shifts (peak positions) are very sensitive to the local chemical environment so that even slight structural changes, e.g., during thermal or chemical denaturation, or due to changes in buffer composition or

degradation or PTM inhomogeneity, can be detected. Certainly, if a protein is to be produced and investigated repeatedly, NMR is the benchmark standard for checking batch-to-batch consistency, the impact of freeze-thaw cycles, effects of changes in buffer composition, or the production process. NMR spectroscopy may also be the only viable structural assessment method for proteins that are fully or largely intrinsically disordered.

Here, however, we will concern ourselves only with the application of CD spectroscopy given its widespread availability and relative simplicity for novice users.

Protein CD spectroscopy is based on the fact that amino acids and secondary and tertiary structures of proteins are chiral and as such absorb left- and right-handed polarized light differently. CD spectroscopy measures this difference (ΔA):

$$\Delta A = A_{\text{left}} - A_{\text{right}} = \Delta\varepsilon \cdot c \cdot d \tag{3}$$

where A_{left} is the absorption of left-handed polarized light; A_{right} the absorption of right-handed polarized light. This measured quantity, ΔA, depends on the intrinsic differential molar extinction coefficient $\Delta\varepsilon$ in $M^{-1}\ cm^{-1}$, the concentration c in M, and the path length of the sample through which the light passes d in cm. Due to the historical use of polarimetry in measuring CD, it is also commonly reported as an ellipticity θ in millidegrees according to

$$\theta = \Delta A \cdot 32.982 \tag{4}$$

As the differences in absorbance are small (typically in the range of 1 part in 10^3 to 10^4), the measurements have to be very accurately subtracted from the background. Consequently, the instruments have to be much more sensitive than classical UV/Visible (UV/Vis) spectrophotometers.

Generally, one distinguishes far-UV CD spectroscopy (190–250 nm) and near-UV CD spectroscopy (250–350 nm). In the far-UV region, the peptide bond is the chromophore. Peptide bonds in defined secondary structure elements such as α-helices and β-sheets, as well as random-coil structures, yield specific CD spectra with characteristic features (Fig. 6). Thus, the far-UV CD spectrum of a protein can distinguish a folded protein that forms secondary structures from an unfolded protein. The differing proportions of the individual secondary structure elements in a particular protein will give a unique "fingerprint" spectrum characteristic of its particular structure, which can be used to detect any difference between samples (Subheading 1.5). Deconvolution of the CD spectrum using a set of reference spectra for each of the secondary structure types can produce an estimate of the percentage of each type present in the protein (and thus quantitate any secondary structural changes).

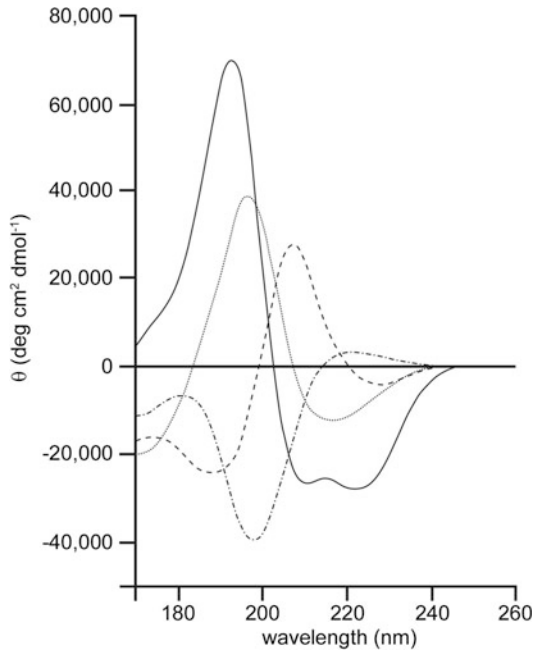


Fig. 6 Typical far-UV CD profiles of various types of protein secondary structure. α-helix (———), antiparallel β-sheet (·····), β-turn (------), and random coil (-·-·-·-·)

Aromatic side chains and disulfide bonds are the chromophores in the near-UV regime. As the signal arising from these groups is very sensitive to their environment, information on the overall tertiary structure of the protein can be retrieved from these spectra and even subtle changes in the structure can be detected, e.g., that may occur upon binding to a ligand or changes in buffer composition, which do not alter secondary structure elements.

Overall, CD spectroscopy can be used to determine the folding state of a protein to compare the structure (1) of proteins from different sources/batches, (2) of different protein variants, and (3) of proteins before and after changes in the production process, as well as to study protein stability by both chemical and thermal denaturation allowing the determination of thermodynamic parameters, to characterize conformational changes, and to analyze protein:ligand interactions [18–20].

1.4 Assessing Protein Stability and Solubility

Generally, different types of "stability" can be distinguished. In the QC context, it may be divided into "chemical" and "physical" stability. Chemical stability describes how prone a protein is to chemical modifications such as deamination and oxidation as well as aspartic acid isomerization and nonreducible crosslinking; such modifications can be checked by MS (*see* Subheading 1.2.3). In order to facilitate reproducible downstream experimentation, it is

necessary that the protein is physically stable, i.e., that its structural integrity and homogeneity, and thus its functionality, are maintained over a reasonably long period of time. Physical stability relates to several distinct physical properties: structural/conformational stability (is there a single conformation of the protein or is the protein flexible), thermodynamic stability (the ratio of folded to unfolded protein molecules), thermostability (the resistance of the structure to the effects of temperature change), chemostability (the resistance of the structure to the effects of denaturing agents), and colloidal stability (the resistance to self-aggregation = solubility). These physical stabilities are often related to each other as aggregation frequently occurs via unfolded or partly unfolded states. In protein QC, thermostability and colloidal stability are the two parameters that are most commonly tested.

1.4.1 Thermal Unfolding Assays

Thermostability is an important feature that should be optimized to improve the protein behavior in order to facilitate crystallization or other structural, biophysical, or functional studies [24–26]. The melting temperature T_m (the temperature at which half of the protein is folded and half is unfolded) is a useful measure of the thermostability and a characteristic feature of each protein. T_m may be used to guide optimization of buffer/storage conditions, to improve stability, or to compare the stability of different protein variants or proteins from different sources. The thermal stability of a structure and T_m can be determined via denaturation experiments, in which the temperature is scanned across a wide range and an experimental property related to the extent of folding or unfolding is recorded at the different temperatures. There are a number of techniques that can be used to determine the T_m of the sample such as near- or far-UV CD spectroscopy that monitors the loss of tertiary or secondary structure as the signature of unfolding (Subheading 1.3), *differential scanning fluorimetry* (DSF) that monitors changes in the environment of intrinsic or extrinsic fluorophores, and *differential scanning calorimetry* (DSC) that measures the change in heat capacity. DSF based on both intrinsic and extrinsic fluorophores requires relatively small amounts of sample due to the high intrinsic sensitivity of fluorescence measurements and are inexpensive, whereas DSC is the method of choice in order to characterize a sample thermodynamically. However, one should keep in mind that thermodynamic parameters can only be determined if the unfolding reaction is fully reversible. In order to test the reversibility of denaturation, the sample should be cooled down following thermal denaturation and heated up again, checking if the measurement can be duplicated. Reversibility is also desirable for accuracy of measurements, but is not strictly required, for QC purposes.

DSF of intrinsic fluorophores monitors the changing behavior of the protein's fluorescent residues, usually Trp whose fluorescence

responds very sensitively to properties of its local microenvironment. Thermal unfolding is measured by following the intrinsic Trp steady-state fluorescence intensity and the position of the emission maximum [27, 28]. The ratio of the fluorescence intensities at two different wavelengths, 330 nm (folded) and 350 nm (unfolded), is used to follow the change in the environment of Trp residues during temperature-dependent unfolding of the protein. This ratio shows a sharp increase of the signal near the midpoint of thermal unfolding. T_m can be calculated from the maximum of the first derivative. The applicability of DSF of intrinsic fluorophores necessarily requires the presence of Trp (or high levels of Tyr) in the folded core of the protein that will be exposed upon unfolding. Moreover, it is necessary to exclude (e.g., with complementary DLS experiments) the possibility that observed signal changes are caused by any aggregation as this will also lead to changes in the fluorophore environment. A comprehensive understanding of the measurement, advantages, and drawbacks can be found in Ref. 29, and an example is presented in Fig. 7.

An alternative strategy is based on *DSF using extrinsic fluorophores*. Several nonspecific protein-binding dyes increase their fluorescence upon binding to the hydrophobic parts of a protein, which are exposed upon partial or full unfolding of proteins during heat

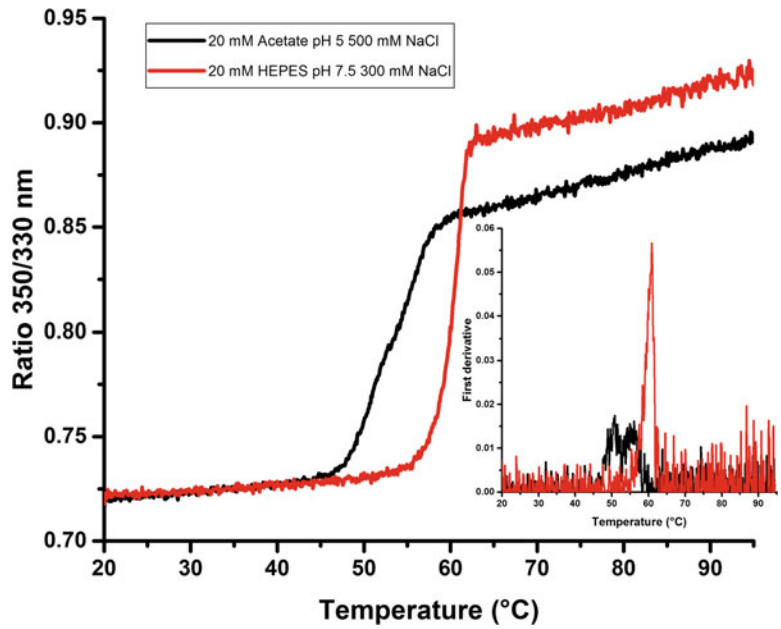


Fig. 7 Following thermal unfolding by DSF of intrinsic fluorophores. Thermal denaturation of the same protein in two different buffers. One of the buffers (red) clearly improves the thermal stability of the protein. (Inset) First derivative of the fluorescence intensity ratio showing a single transition with a T_m of 61 °C in the HEPES buffer

denaturation. This approach, developed by Pantoliano *et al.* in 2001 [30], has been originally registered under the trademark "*Thermofluor*," a name now used as a synonym for all such extrinsic dye-based experiments. Ideally, the rapid increase in the proportion of unfolded protein will lead to a sharp increase of the fluorescent signal over a short temperature range, generating a sharp sigmoidal curve of fluorescence and again allowing the calculation of T_m from the maximum of the first derivative. These measurements can be carried out in multiwell plates in a quantitative polymerase chain reaction (qPCR) machine. This approach is very popular and accessible because of the wide availability and relatively low cost of such instruments. However, there are some drawbacks vs. the intrinsic approach: (1) the addition of the extrinsic dye modifies the buffer composition, (2) some dyes may be incompatible with some buffers, and (3) binding of the dye may itself destabilize the protein. An exhaustive review on Thermofluor has been written by Boivin and coworkers [31].

An alternative is DSC, the only direct technique to study the thermodynamics of protein thermal stability. Its advantages are that no modification of the protein or any additional component is required. On the negative side, DSC is typically both more time- and sample-consuming than DSF of intrinsic fluorophores or Thermofluor [32]. However, automated DSC machines are becoming more widely available, reducing the time required for the experiments.

DSC measures the molar heat capacity C_p of the sample as a function of temperature T and allows the determination of the change in enthalpy and entropy upon denaturation. Folded and unfolded states have different heat capacities (largely due to their different interactions with water molecules), and changes in populations of the two states as the temperature is increased mean that a peak appears in the C_p vs. T plot with its maximum at T_m and its integral corresponding to the change in enthalpy upon denaturation. As DSC directly measures the heat capacity, it is very well suited to measure the protein activation energy that is directly correlated to thermal flexibility [33]. The protein activation energy corresponds to the energy barrier that needs to be passed to allow unfolding.

1.4.2 Assessing Colloidal Stability/Aggregation

DLS is the method of choice to detect the formation of protein aggregates and to follow the effects of changes in the experimental protocols to improve solubility. A variety of changes can be made either upstream to/during the production/purification process (for details *see* Refs. 26 and 34 and **Note 4**), or to buffer conditions. Many groups use DLS as a technique to improve the solution conditions of their proteins, in particular before crystallization studies [35, 36]. Recent DLS instrumental developments

permit the processing of a large number of samples in plate format, making buffer condition screening easier. An automated setup can simplify the measurement of the homogeneity of the sample and the detection of aggregates (and high-order, physiologically irrelevant oligomers) as described above (Subheading 1.2.2). Moreover, measurements can be performed over time to predict the stability under the condition of a downstream experiment, as a "good" buffer at the start of an experiment may appear "not so good" after a few hours at 20 °C.

1.4.3 Buffer Optimization

The physical stability of a protein strongly depends not only on the protein itself but also on the buffer composition, with pH, salinity, the presence of additives, cofactors, or ligands all having an impact. There is presently no effective way to accurately predict the effect of solution conditions on a protein's thermostability or solubility from its intrinsic properties (amino acid composition, pI, secondary structure elements, etc.). Consequently, researchers must systematically screen a range of different buffer compositions. It is highly recommended to combine methods such as DLS and DSF that will separately give information on solubility/aggregation and thermal stability. This combined approach allows the user to define the best buffer by seeking conditions that improve both properties. The buffer matrices for multiparametric screening of pH, salinity, buffer nature, additives, and cofactors can be generated by hand or using simple robotics that dilutes the protein in all the different conditions prior to measurement [28].

1.4.4 Storage Issues

It is recommended to work with freshly prepared proteins. However, this is not always practicable, and proteins may have to be stored for weeks or months before being used. One method that is commonly used to store proteins is flash-freezing of small aliquots (1 mL) and subsequent storage of the samples at −80 °C. Thin-walled plastic tubes should be used to hold the sample and plunged into liquid nitrogen so that freezing is as fast as possible to avoid the formation of ice crystals that may damage the sample. If retaining catalytic activity is paramount, proteins are usually stored in a buffer containing 50% (v/v) glycerol at −20 °C. An alternative is lyophilization, which has the advantage that long-term storage can be at 4 °C or even room temperature. Ideally, the protein is in pure water before lyophilization. In this case, when the protein is needed, an aliquot of lyophilisate is weighed and resuspended in the buffer of choice. Often proteins are not stable in pure water, so more typically a defined sample volume of the protein in its target buffer is used for lyophilization. In this case, the lyophilisate must be resuspended in the same volume of water. Not all proteins are suitable for lyophilization as they may undergo irreversible aggregation when concentrated and not be resuspended properly.

Independent of the method to store a protein, it is necessary to assess the quality of the sample after storage, i.e., before it is used in downstream applications. This can be done using a batch-to-batch consistency protocol combining the basic QC tests described above (Subheadings 1.1–1.3). It is essential that the protein's physicochemical properties are not altered during storage. Processes such as oxidation, precipitation, or (partial) degradation may impede the study of protein structure and function or render it completely impossible. This problem has been extensively studied for proteins that are used as therapeutics, and it has been found that the most crucial aspect in stabilizing the protein, and allowing storage for 18–24 months without loss of integrity and functionality, is the final buffer composition; thus, it is worth investing time and consumables in the optimization of the buffer used for storage, which may be distinct from that used downstream.

1.4.5 Membrane Protein Buffer Optimization

Membrane proteins have large surface patches of hydrophobic residues required for their insertion into or interaction with membranes. In aqueous solution, these patches promote protein aggregation through strong hydrophobic interactions. Thus, the conditions for solubilizing membrane proteins are distinct from those of cytoplasmic proteins. The most common method for solubilizing (integral) membrane proteins is the addition of detergents or amphipols—polymers with a hydrophilic backbone decorated with hydrophobic sidechains, making them amphiphilic and thus enabling them to stabilize membrane proteins in aqueous solution by binding with one side to the hydrophobic part of the protein and the other facing the surrounding water [37]. However, many detergents do not maintain the structure and function of membrane proteins, possibly due to the lack of lateral pressure or due to the loss of "core lipids," i.e., the tightly bound, quasi-integral lipids that are necessary for activity. Proteoliposomes are model systems where membrane proteins are reconstructed into a lipid bilayer after their extraction from the membrane with detergents, and this can already lead to an improved stability of membrane proteins in solution. More recently, membrane scaffold proteins (MSPs) [38] and new types of polymers made of styrene–maleic acid (SMA) [39] or diisobutylene–maleic acid (DIBMA) [40] have been used as they can keep the proteins in their original lipid environment [41]. Since all these materials have different characteristics, the choice of the optimal method for a particular membrane protein cannot be predicted. Conditions for efficient and functional solubilization need to be tested systematically using screening approaches (typically DSF or DLS) with different buffers and solubilizing molecules (Subheading 2.4).

1.5 Batch-to-Batch Consistency

If different batches of a protein are used or if production, storage, or assay conditions are altered, it is necessary to demonstrate batch-to-batch consistency; i.e., the different lots must exhibit the same purity, homogeneity, oligomeric state, and conformational/structural integrity. While the first three criteria can be assessed via the methods described (Subheading 1.2), consistency in conformation can be checked by recording a "spectral signature" or "spectral fingerprint," with a CD or NMR spectrum being the most common ones (Subheading 1.3). The spectral signature spectrum of a target protein is recorded when it is successfully produced and purified for the first time, when it fulfills all QC criteria, and when it is fully characterized ("batch 0"). This spectrum is then used as a reference for all future batches. If a newer batch shows differences to the reference signature, further quality tests are required or conditions must be altered before proceeding to downstream applications.

2 Materials and Methods

2.1 UV Spectrum and Concentration Determination

The most common causes of inaccuracy in concentration determination are miscalculation of dilutions or pipetting errors during sample preparation. Special care must therefore be taken during the preparation of the sample for concentration measurement and in subsequent downstream uses.

Prior to the sample measurement, the upper absorbance limit of the UV/Vis spectrometer in the wavelength range used must be determined. It is critical that the total absorbance of cuvette, buffer, and protein must not exceed this limit, as above this value the instrument will not produce a measurement in proportion to the actual sample absorbance and consequently an incorrect concentration would be obtained [42]. Consult the instrument manual or manufacturer to find this limit—if a specific value for the linear range of the instrument cannot be found, then ensure that absorbance measurements fall between 0.1 and 1 AU. It is highly recommended to record the spectrum using a 1-cm quartz cuvette to minimize errors in the measurement (*see* **Note 5**).

Although most spectrometer software will have options for determining concentration from absorbance based on average proteins or other standards such as BSA, these are meant for working with undefined protein mixtures, not pure recombinant proteins. Each pure protein will have a distinct molar absorption coefficient that depends on its amino acid composition. In order to determine concentration from the UV spectrum using Eq. 1, an estimate of the protein's molar absorption coefficient must be made based on its primary sequence, e.g., using the ProtParam program for ε at 280 nm (available on the web at https://web.expasy.org/protparam/) or the Protein Calculator for ε at 205 nm (https://spin.

niddk.nih.gov/clore/). For measurements at 214 nm, standard amino acid extinction coefficients and a formula for computing ε can be found in Ref. 43.

A simple protocol for a UV spectrum measurement is given below:

1. Switch on the instrument and allow it to warm up for 20 min.
2. Record a spectrum with an empty sample compartment from 340 nm down to at least 240 nm and preferably 190 nm.
3. Insert a 1-cm quartz cuvette (*see* **Note 5**) and record a spectrum of the empty cuvette (in order to identify any possible absorption of the cuvette itself or from contaminants at the required wavelengths).
4. Add buffer and measure the absorbance of cuvette plus buffer. This should be exactly the same buffer that used to prepare the protein solution (*see* **Note 6**). The absorbance will go up at low wavelengths as carbonyl and amide bonds found in many buffers as well as salts absorb here; DTT and β-mercaptoethanol absorb at higher wavelengths (*see* **Note 7**). It is best that the buffer is transparent at the required wavelengths or, if there is some absorbance, make sure that the absorbance is much smaller than the limit of the instrument. If not, use a lower concentration or less absorbent buffer.
5. Re-measure the buffer to make sure that it gives consistent results. This test checks whether the lamp is warm and excludes lamp intensity fluctuations.
6. Run a baseline/blank with buffer (this will be subtracted from future measurements to produce a spectrum for the sample alone).
7. Run a spectrum with the protein solution and inspect the spectrum shape: check that the maximum absorption is below the high absorption limit of the instrument. If this is not the case, dilute the protein solution with buffer and repeat the measurement.
8. Calculate AI and the A_{260}/A_{280} ratio (Subheading 1.1). The AI should be below 2 and the A_{260}/A_{280} ratio below 0.6 to be sure that the sample is sufficiently free of aggregation and nucleic acid contamination. If these conditions are met, use the A_{280}, A_{214}, or A_{205} value and the corresponding extinction coefficient to calculate the concentration of the protein solution in the cuvette (*see* **Note 6**).

Bear in mind that both the measurement and the extinction coefficient calculation will always have some error and the true concentration is frequently 5% different from the value estimated by UV measurement. However, no other method is more accurate

(except weighing a large quantity of pure protein, which is only practicable in rare cases) if a pure protein sample is used.

2.2 Methods to Determine Purity, Homogeneity, and Oligomeric State

2.2.1 SDS-PAGE and Staining

Gel electrophoresis equipment (apparatus, running buffers, sample buffers, precast gels, etc.) can be obtained from many manufacturers. Protein gel electrophoresis is usually performed in vertical systems where the gel is positioned between two glass plates with the top part being immersed in anode buffer and the bottom in contact with cathode buffer.

Polyacrylamide Gels

PA gels can be easily prepared in the lab, but commercially made precast gels are recommended for QC purposes because they are much more consistent in their behavior. Follow the manufacturer's instructions for their use. PA gels are produced by copolymerization of acrylamide monomers and a crosslinker, usually *N*,*N*′-methylenebisacrylamide, with the pore size being defined by the total concentration of acrylamide and the degree of consequent crosslinking. The acrylamide concentration (pore size) of the resolving gel is chosen depending on the sizes of proteins to be resolved (usually 10–19% total acrylamide concentration). With appropriate formulation, they can be used to separate proteins <800 kDa; larger proteins will not enter the gel. Use of gradient gels (increasing acrylamide concentration from top to bottom) allows the separation and visualization of large (200 kDa) and small (10 kDa) proteins on the same gel as the smaller pores slow down and retain small proteins in the bottom of the gel while large proteins are still separated at the top.

Discontinuous Denaturing SDS-PAGE

For sample preparation, a sample buffer containing SDS is added to the protein solution and the sample is incubated at >95 °C for some minutes in order to enhance denaturation. Typical sample buffer (4× concentrated): 260 mM Tris-HCl (pH 6.8), 400 mM DTT, 8% (w/v) SDS, 0.02% (w/v) bromophenol blue, 40% (v/v) glycerol. N.B. some proteins may precipitate or undergo degradation when heated (especially highly concentrated proteins or membrane protein) and sample preparation should be tested with and without boiling. The reducing agent (DTT) should ensure that each natively disulfide-bonded protein should run at the expected molecular weight of its monomer. In order to test the presence of disulfide bridges, samples can be prepared with and without reducing agent and should then be analyzed on the same gel. If intramolecular disulfide bridges are intact, the protein will be more compact and the band should be shifted toward lower molecular weight as compared to the reduced sample. If intermolecular disulfide bridges are present, the protein will move toward higher molecular weight.

For running the SDS-PAGE, a tris(hydroxymethyl)-aminomethane (Tris)-glycine-HCl running buffer of 25 mM Tris-base (pH 8.3), 192 mM glycin, and 0.01% (w/v) SDS is usually used; the same buffer is used as anode and cathode buffer. Many alternatives to the Tris-glycine system are available, which may be better suited for a particular target protein. Of particular note, the Tris-glycine-HCl buffer system has relatively poor separation for proteins and peptides <15 kDa and above 200 kDa. Good separation from 5 to 30 kDa can be achieved by the method developed by Schägger and von Jagow [44], which uses a small pore gel formulation and a Tris-tricine buffer system. A Tris-acetate system can be used to improve separation of higher molecular weight species. Glycoproteins do not bind as many SDS molecules as nonglycosylated proteins and, consequently, migrate more slowly than non-glycosylated proteins of the same size. A Tris-borate-EDTA buffer can be used for preparing and running such samples (replacing the Tris-HCl) as borate will bind the sugars, giving them a negative charge so that the migration speed is increased.

Even slight modifications of a gel electrophoresis protocol may significantly affect the results for a certain sample. Thus, in order to obtain high reproducibility, a standard operating procedure should be established that is strictly followed, especially for assessing batch-to-batch consistency.

Staining

The choice of the staining method depends on the required sensitivity, reproducibility, and downstream applications. For example, if 10μg of protein is loaded into a lane of the gel and contaminants of <1% should be detected, a staining method with a sensitivity <100 ng per band is required.

The most common method for in-gel protein staining is with Coomassie Brilliant Blue G-250 or R-250 [45, 46] as it involves a ready-to-use reagent that can be easily prepared in the lab, is quick, and does not permanently chemically modify the target proteins. Under acidic conditions, the Coomassie dye binds to basic and hydrophobic residues while acetic acid fixes the proteins. As the proteins are not chemically modified, excised protein bands can be destained completely and the target proteins can be recovered for downstream applications, such as intact mass or sequence analysis by MS. The detection limit of Coomassie Brilliant Blue is approximately 25 ng, although in some cases amounts down to 10 ng per band may be detected. Coomassie staining is routinely used in many labs.

A simple protocol is presented below:

1. Wash the gel with water to remove residual SDS, which interferes with dye binding.
2. Incubate the gel in a staining solution at approximately 50 °C for 10–15 min (there are different compositions for staining

solutions, a typical one is 0.025% (w/v) Coomassie Brilliant Blue G-250 in 10% (v/v) acetic acid; ethanol or methanol may be added in some formulations).

3. Incubate the gel in a destaining solution at approximately 50 °C for 10–15 min to wash out non-bound dye (again, there are different compositions for destaining solutions according to the manufacturer, e.g., 10% (v/v) acetic acid; ethanol/methanol may be added).
4. Incubate the gel in water overnight, where, optionally, a piece of paper tissue may be added to fully wash out non-bound dye.

Usually, larger proteins tend to bind more dye than small ones. Consequently, the intensity of bands can only be compared in terms of weight concentrations (mg/mL) and not in terms of molar concentrations. If quantitation by densitometric analysis is required, a comparison with samples of the same protein at known concentration is necessary; as well as even staining throughout the gel.

Staining with the Coomassie Brilliant Blue dye is the most common method and is usually sufficient when working with purified recombinant proteins. If there is sufficient protein available, but protein concentration is simply low, the sample can be concentrated by trichloroacetic acid (TCA), TCA/deoxycholate, or TCA/acetone precipitation in order to allow more total protein to be loaded. However, if available protein is low, it should be noted that Coomassie staining is at least 10 times less sensitive than silver staining, reverse zinc staining, visualization by fluorescent dyes, or radioactive staining (rarely used in QC). The higher sensitivity stains should also be considered whenever the main objective of the gel electrophoretic run is the detection of impurities in the sample. There are also stain-free techniques available, where proteins can be visualized without the need of a stain. *See* **Note 8** for further details on other visualization approaches.

2.2.2 Size-Exclusion Chromatography: Static Light Scattering (SEC-LALS/SEC-MALS)

An appropriate chromatography column is required. Depending on the manufacturer and the chemistry of the matrix, each column will be specified in terms of a particular range of molecular sizes that it can effectively separate. This is normally given in terms of the molecular mass of globular proteins. One should keep in mind that the molecular mass is not identical to the hydrodynamic size, and for nonglobular proteins, substantial adjustment may be required (i.e., an elongated 60-kDa protein may behave like a 160-kDa globular protein). In practice, the column should be chosen for its ability to elute the protein of interest at a volume toward the center of the upper and lower limit, and in particular that the target protein elution is well-separated from the void volume so that aggregates are removed. For quantitative, analytical

measurements, columns with small diameter are preferred as they have greater resolution. In order to select the best column and ensure correct handling, one should read the manufacturer's instruction, which gives the molecular weight limit, pressure limit, the volume of the column, the void volume, and the recommended flow rate. Test runs with small injection volumes of the protein of interest may be performed to choose a good column setup.

The running buffer for the sample should be selected such that it is stable over time, as any change, such as oxidation of a reducing agent, will lead to a drift in the baseline of the refractometer. The buffer should contain at least 50 mM of salt to avoid interactions with the column matrix and 0.1% (w/v) NaN_3 to prevent microbial growth. All solutions must be made with high purity water (preferably >18 MΩ), freshly filtered (0.1-μm filter), and degassed to avoid background scattering and an unstable baseline.

A simple protocol for a SEC-SLS experiment is given below:

1. Switch on the instrument.
2. Purge the system with high purity water.
3. Connect the column to the SEC-LALS/SEC-MALS system so that eluting proteins are analyzed immediately inline.
4. Choose the flow speed and pressure limit according to the column manufacturer's instruction.
5. Equilibrate the system with at least 2 column volumes of water. Columns are usually stored in 20% (v/v) ethanol. Therefore, the column must first be washed extensively with water (at least 2 column volumes) as ethanol may otherwise precipitate salt in the buffer when it is introduced.
6. Purge the system with a buffer of interest.
7. Equilibrate the system with at least 2 column volumes of the buffer. If necessary, equilibration should be continued until the baseline is stable.
8. Purge the refractometer extensively until the baseline remains stable for one column run.
9. Prepare your sample in the same buffer at a concentration of at least 1 mg/mL.
10. Centrifuge the sample for 5 min in a benchtop centrifuge at full speed.
11. Load the sample into the injection loop (avoiding creating any bubbles). The volume that will be injected onto the column should be lower than 1% of the total column volume to ensure good separation (usually 100μL for a column volume of 24 mL). For accurate quantification, it is necessary to flush

the injection loop with twofold the injection volume of buffer to ensure complete loading onto the column.

12. At the end of the run (1 column volume), analyze all peaks of the chromatogram to assess sample recovery and for molecular mass determination according to the manufacturer's manual. The baseline for the chromatograms should be set up in a region where no proteins elute, such as the region prior to the void volume of the column (first third of the column volume). The baseline should remain constant if the sample is in the same buffer as the elution buffer and if the temperature is controlled. If the column type and the buffer conditions are chosen correctly to avoid nonspecific binding of the protein, you should recover nearly 100% of your sample.
13. If you intend to run another sample in the same buffer, rinse the injection syringe and injection loop with buffer and you may then start another run at (7).
14. When not in use, the instrument should not be left with buffer for more than an hour. Rinse the injection syringe and injection loop and purge the pump refractometer and LALS/MALS detector with water and then fill with 20% (v/v) ethanol. Columns may be left in buffer for short periods but should be rinsed with least 2 column volumes of water and then filled with 20% (v/v) ethanol for longer-term storage.

If the sample is pure and homogeneous, SEC will give one sharp and symmetric peak in the elution profile/chromatogram associated with the molecular mass of the target protein. Polydispersity due to oligomers or complex formation will be immediately obvious if there is more than one peak. The amounts of each species may be estimated by integrating the area under each peak. Protein aggregates will usually elute in the void volume. However, one should keep in mind that SEC is not an equilibrium method; not only will it apply shear forces on the sample that may disturb protein oligomers/complexes, but the sample will also be diluted during the SEC run and will consequently change species populations as the run proceeds.

2.2.3 DLS Measurement of Homogeneity

DLS is a quick method to test the homogeneity of a sample and the presence of aggregates. The detected signal is proportional to the sixth power of the particle radius, meaning that even small numbers of large aggregates can be easily detected. However, this strong size dependence also means that the even small fractions of a large species (such as an oligomer) can lead to inaccurate results or entirely hide the presence of a smaller one. For example, in a solution containing 99.9% of homogeneous protein with a 3 nm R_h and 0.1% by mass of an aggregate with a 30 nm R_h, the peaks corresponding to the two populations will have approximately the

same intensity. Furthermore, the resolution of DLS permits one to distinguish between monomeric and oligomeric states of a protein only if the molecular mass of the two states differs significantly (*see* Subheading 1.2.2). Thus, samples that contain a mixture of monomers and dimers of a protein will usually give rise to only one peak.

Modern backscattering DLS instruments can measure scattering even for protein concentrations >50 mg/mL but will tend to provide progressively more inaccurate results at high concentrations. Each instrument will have an ideal operating concentration range, which depends on the particle sizes present. Larger particles scatter more and lower concentrations are better. The manufacturer's information should be followed, but a typical rule of thumb would be to aim for a concentration = 10/(particle molecular weight in kDa) in mg/mL, i.e., about 1 mg/mL for a 10-kDa particle and 0.1 mg/mL for a 100-kDa species. Experiments should be repeated at concentrations differing by a factor of 2 to ensure that there are no substantial concentration-dependent effects.

DLS experiments can be performed using nearly any buffer. However, the buffer must be filtered as it is very important to remove dust or salt crystals that would cause scattering and interfere with measurements. When working with membrane proteins containing detergent micelles, or other buffers containing colloids, or crowding agents, it is important to measure all conditions with and without protein to be able to distinguish the behavior of empty micelles/buffer from the solubilized protein. In such cases, the objective is to identify conditions where a monodisperse protein: detergent or protein:amphipol complex is formed.

A simple protocol for a DLS measurement is given below:

1. Switch on the instrument and select the experimental temperature.
2. Wait for temperature stabilization.
3. If using reusable cuvettes or plates, clean them with detergent (e.g., Hellmanex 2% (v/v)) and water, and dry it with ethanol.
4. Filter and degas buffers using a 0.1-μm or, preferably, 0.02-μm filter.
5. Prepare the protein sample at an appropriate concentration, diluting with buffer as necessary. If the purpose of the DLS measurement is to detect aggregation, then the sample should not be filtered or centrifuged. Otherwise, it is advisable to filter the sample and centrifuge it for 5 min in a benchtop centrifuge at full speed prior to loading.
6. Load buffer and samples into the wells or into the cuvettes. Avoid the formation of bubbles. However, if bubbles are present, centrifuge the plate or cuvette (e.g., for plates: 1 min at 500–1000 × *g*).

7. Set up a program containing 10–30 acquisitions with an acquisition time of 2–10 s per sample. The appropriate combination of number and time depends on whether or not dust or other larger particles disturb the measurement. Specifically, use a large number of short acquisitions if individual acquisitions give erratic results. Measurements should be done in technical triplicates using automatic laser power and detector attenuation.
8. If using a multiwell plate DLS instrument with a camera, an image of each well should be captured, as this may help in interpretation of data analysis by visualizing artifacts (e.g., phase separation, bubbles).
9. During data analysis, first the data from buffer alone should be checked to confirm that the buffer itself does not contain large particles. Figure 3a shows the typical autocorrelation curve for a clean buffer with a weak intensity of the autocorrelation curve.
10. Check the shape of the autocorrelation curve for the protein (Fig. 3b, inset). Check the fit and the error of the measurement. The presence of aggregates is indicated if particles with R_h of more than 10 times that of the expected size of the monomer can be observed (Fig. 3b).
11. If the protein is not detectable due to the presence of too much aggregate or other artifacts, the sample should be centrifuged using a benchtop centrifuge (15 min at full speed). Remeasure the concentration after centrifugation as protein content will be reduced.

2.2.4 Protein Intact Mass by MALDI-TOF Mass Spectrometry

Protein intact mass can be assessed by either ESI or MALDI MS approaches. Here, we will focus on MALDI-TOF MS. During sample preparation for a MALDI experiment, a droplet of the target protein:matrix solution is deposited onto a MALDI plate and the solvent is evaporated to allow co-crystallization. There are several ways to prepare deposits, varying in matrix concentration, solvent, method of crystallization, etc. The "dried droplet" method is the most commonly used technique because it is extremely fast and simple to implement.

A simple protocol for intact mass by MALDI is given below:

1. Prepare the protein sample. For a typical 15-kDa protein, 1μL at 0.1 mg/mL is sufficient, and for a 150 kDa protein, around 1 mg/mL is recommended. These concentrations are acceptable for protein samples in phosphate-buffered saline (PBS)-type buffers. If your protein sample buffer is supplemented with more than 1% (v/v) glycerol or detergent or in high salt, the protein should be prepared at a 10× higher concentration and then diluted with water to reach the required

concentration. Alternatively, pipette tips containing reverse-phase chromatography media (such as ZipTip C18 for proteins <20 kDa and C4 for proteins >20 kDa) may be used to concentrate and desalt the protein of interest.

2. Freshly prepare the matrix solution. The chosen matrix will depend on protein size and glycosylation state. HCCA (α-cyano-4-hydroxycinnamic acid, a.k.a. α-CHCA) or DHB (2,5-dihydroxybenzoic acid) are typically used for peptides and proteins below 20 kDa. SuperDHB (DHB with 10% (w/w) 2-hydroxy-5-methoxybenzoic acid) or sinapinic acid for larger proteins. Follow the instrument manufactures' instructions for matrix preparation. However, 25 mg/mL in 50% (v/v) acetonitrile and 0.1% (v/v) TFA in HPLC water is typically effective.
3. The protein sample can be mixed with the matrix directly on the MALDI plate, or the mixture can be prepared beforehand in a tube and a droplet is then deposited on the target plate. Typically, 1μL of protein is applied to the target plate and 1μL of matrix is immediately added. Alternatively, the sample and the matrix can be mixed by repeated pipetting before the deposit, which is still in the form of a droplet, is allowed to dry.
4. Acquire data on the MALDI time-of-flight (TOF) mass spectrometer. It is recommended to use the minimum laser power required to get sufficient signal to noise in the spectrum to detect the protein as a high laser power leads to a loss of resolution and precision on the m/z.
5. Calibrate the experiment with a standard protein mixture as recommended by the manufacturer.

2.3 Protein Structural Integrity and Stability

2.3.1 CD Spectroscopy

The signal/noise (S/N) ratio in a CD spectrum is strongly dependent on the UV absorbance of the sample—a low absorbance gives little measurable differential absorption (i.e., a low CD signal), and high absorbance means that little light reaches the detector. Optimal signal to noise is obtained a little below 1 AU, and an absorbance of the sample between 0.5 and 1.5 AU will give good results (above 2 AU insufficient light will reach the detector to obtain a reliable measurement). Protein sequence (for near-UV) and buffer composition, their concentrations, and cuvette path length will alter the total absorbance. It is necessary to select or adjust those quantities that can be changed to obtain a satisfactory S/N for each sample.

Choice of Cuvette, Buffer, and Sample Preparation

There are a variety of cuvettes available with path lengths ranging from 0.01 to 10 mm. The required concentration of the protein sample depends on the spectral range and type of cuvette. For far-UV CD, typically, a 1-mm cuvette is used, and for this, a good

starting point is to employ 100–150μg/mL of protein in a low-absorbance buffer. With this path length, lower concentrations can be used successfully by collecting data for longer, but very high concentrations (>300μg/mL) will absorb too much light. For near-UV experiments where buffer and protein absorbance (dependent on aromatic content) is smaller and the CD signal is also smaller, higher concentrations and/or longer path lengths (e.g., 1 cm) will typically be used.

A low concentration phosphate buffer with little salt (e.g., 5 mM phosphate, 5 mM NaF) is recommended as a good low--absorbance spectroscopic buffer. However, in protein QC, it is most important to work with solution conditions in which the protein is well behaved, and which will be used for downstream applications. These will not necessarily be optimal from a spectroscopy perspective. To carry out CD experiments, the buffer alone must not have UV absorbance >1 (and preferably much lower) anywhere in the range of the spectrum, as it will, otherwise, absorb most of the light. Consequently, buffers with compounds containing aromatic groups cannot be employed. High concentrations of chloride ions absorb at wavelengths <215 nm and should thus be avoided. If high salt is required, chloride may be replaced by fluoride. Many components of the sample buffer may prevent reliable recording near or below 200 nm (*see* **Note 7**), compromising the reliability of secondary structure calculation; the validity of spectral comparison, however, is not affected. It is possible to reduce the detrimental effects of absorbing buffers on measurements by working with higher protein concentration in shorter path-length cuvettes.

Scattering from contaminants/aggregates is indicated by a nonflat baseline at wavelengths longer than those where the sample absorbs light (for far-UV >250 nm, for near-UV >320 nm). The buffer should be filtered (0.1 or 0.02μm), and the sample should be centrifuged (10 min at full speed in a benchtop centrifuge); both the buffer and the sample should be degassed. Further details on the choice of cuvettes and buffers as well as sample preparation can be found in Ref. 18.

Spectrometer Setup

CD spectrometers must be purged continuously with nitrogen (1) to exclude O_2 from the sample compartment as O_2 absorbs incident radiation, which, in turn, limits the lowest wavelength that can be measured and (2) to prevent the production of ozone by the lamp, which would damage the optics. To ensure robust quantitative analysis or comparison of samples over time, the spectrometer should be calibrated on a weekly basis to validate wavelength accuracy, wavelength repeatability, intensity accuracy, intensity repeatability, baseline flatness, baseline stability, and the noise level (for details and protocols, see instrument manuals and Ref. 18).

Several instrument parameters can be adjusted to improve signal to noise by increasing the amount of light measured at a given wavelength (data pitch, scan mode, scan speed, and response time) (*see* **Note 9**). A quick, preliminary scan is helpful in guiding parameter adjustment for new proteins. Usually, the CD signal, i.e., the ellipticity θ, the high tension (HT) voltage (or gain), and the absorbance are recorded. The detector will have a limited range of amplification in which it is linear; thus, the HT voltage, which controls the gain of the detector on most laboratory instruments, should remain below 600 V (or manufacturer-specified limit). If it is too high, you will need to reduce absorbance by decreasing buffer and/or protein concentration or path length.

Measurement of a CD Spectrum

A CD spectrum recorded in the far-UV region will produce a spectrum that has a characteristic shape and magnitude determined by the secondary structure elements of the protein. Even a low-quality spectrum can readily distinguish folded from unfolded protein. However, any spectral comparisons used to monitor structural changes (e.g., due to solution conditions or batch-to-batch variation or calculation of secondary structure content) will require care to optimize spectral quality. A baseline spectrum must be collected of the buffer (the actual buffer used to prepare the protein) and the cuvette inserted into the instrument in the same orientation as for the sample measurement. The baseline spectrum must be recorded with the same instrument parameters as the subsequent protein sample. Usually, a 1-mm cuvette is used, requiring ~200 μL of buffer and protein solution (mininum 160 μl, maximum 350 μl). As experience is built, it is a good idea to create a list of usable path lengths and instrument settings for particular buffers and proteins. A comparative analysis of spectra is facilitated by reusing the same experimental parameters. All comparison should be carried out in terms of molar ellipticity and thus will require correction for any differences in protein concentration (*see* **Note 10**).

A simple protocol to record a CD spectrum is given below:

1. Purge the instrument with N_2 according to the manufacturer's instructions (usually 10–15 min).
2. Switch on the instrument (and the temperature control/water bath, if necessary).
3. Switch on the lamp and let it warm up for 10–20 min.
4. Set appropriate instrument parameters following the manufacturer's instruction (*see* **Note 9**).
5. Record a spectrum of the empty 1-mm quartz cuvette (a single accumulation is sufficient). Even slight characteristics of a protein spectrum for the empty cuvette indicate contamination from previous experiments, and thorough cleaning of the

cuvette is required, either by rinsing with water and drying with methanol or, if that is ineffective cleaning, with an appropriately diluted Hellmanex solution according to the manufacturer's protocol (e.g., 0.5–2% (v/v) Hellmanex for 20–180 min at 25–30 °C) or HCl. Remeasure the cuvette spectrum after cleaning.

6. Filter and degas the buffer. Fill the cuvette with 200 μL of buffer and record a reference spectrum.
7. Remove liquid and dry the cuvette.
8. Prepare the sample solution, centrifuge the sample in a benchtop centrifuge for 5–10 min at full speed, and degas it.
9. Fill the cuvette with 200 μL of sample solution and record a spectrum (same parameters as in **step 5**). Note that the sample spectrum has to be recorded with the same cuvette as the reference/buffer spectrum as the cuvette influences the background more strongly than the buffer.
10. Repeat **steps 4–10** as required for additional samples or to optimize instrument parameters, concentrations, etc. (or for additional samples).
11. Switch off the lamp and the instrument, and purge with N_2 according to the manufacturer's instruction (usually 15–20 min).
12. Switch off N_2 flux.

Estimation of Secondary Structure Composition

The CD spectrum in the far-UV region (throughout 180–250 nm) is largely determined by the secondary structure elements of the protein. The fraction of each secondary structure element can be approximated using different methods, which all make a linear combination of a set of reference spectra, in the simplest case spectra of purely α-helical, β-sheet, and random-coil structures (Fig. 6), plus a noise term containing the contribution of aromatic chromophores and prosthetic groups. Analysis may be carried out by either instrument-specific software or online (e.g., using the Beta Structure Selection (BeStSel) server (http://bestsel.elte.hu)) [47, 48]. An overview of different methods can be found in Ref. 18. Bear in mind that absolute accuracy of methods is very dependent on data quality, particularly at low wavelengths (where buffers often interfere). However, even without absolute accuracy, values are useful in comparative analysis, e.g., of samples, mutants, and denaturation.

Determination of Protein Thermostability

Prior to a thermal denaturation experiment, far-UV CD spectra of the sample should be recorded at 20 °C and any parameters optimized as necessary (as above). Ideally, a thermal denaturation experiment should be carried out in a magnetically stirred 10-mm quartz cuvette in order to enable the use of a low protein

concentration (and thus reduce aggregation during denaturation) and to ensure even heating. However, it is also possible to record unfolding experiments in open (unstoppered) 1-mm cuvettes, e.g., when buffer absorption necessitates their use.

A temperature-controlled instrument will typically have a range of 5–10 °C through to 90–95 °C. A denaturation experiment should run from near the low- to near the high-temperature limit of the instrument and subsequently cool back down to the starting temperature, recording a full spectrum at the start and end. If the CD spectrum following cooling is identical with the initial spectrum, this will indicate reversibility. If the unfolding reaction is reversible, thermodynamic parameters may be determined [19]. The wavelength is chosen where the difference between the folded and unfolded protein (random coil) is large and of low noise (222 nm for α-helical proteins and ~216 nm for beta-sheet-dominated structures).

For reasons of improving signal-to-noise (and thus enabling low concentrations to be used), the CD signal is typically recorded at a single wavelength during the unfolding process (although if the instrument is sufficiently fast, recording full spectra has advantages in analysis) [19].

A simple protocol for a thermal denaturation experiment is given below:

Steps 1–6 as in section "Measurement of a CD Spectrum" (above) to ensure that the cuvette is free from contamination.

7. Set the low- and high-temperature limits (e.g., 10 and 90 °C) and cool instrument to low-temperature limit.
8. Prepare the sample solution, centrifuge the sample in a bench-top centrifuge for 5–10 min at full speed, and degas it.
9. Fill a 1-cm cuvette with the sample solution. A buffer baseline is not necessary as only differences between spectra will be analyzed.
10. Place cuvette in holder, switch on stirrer (if applicable), and allow sample time to cool to the low-temperature limit.
11. Record a full spectrum of the sample at the low-temperature limit.
12. Select the wavelength for data recording where there is a large signal and where the signal for a random coil (unfolded protein) will be small.
13. Heat the sample at 1 °C/min. Record the CD signal at the selected wavelength at closely spaced intervals. The temperature should be monitored via a temperature probe in the cuvette.
14. After heating to 90 °C, record a full spectrum of the sample.

15. Cool the sample to the starting temperature at 1 °C/min, recording the CD signal at the selected wavelength at closely spaced intervals.
16. Record a full spectrum of the sample.
17. Wait for 30 min and record a full spectrum of the sample again.
18. Switch off the lamp and the instrument, and purge with N_2 according to the manufacturer's instruction (usually 15–20 min).
19. Switch off N_2 flux.

Compare the high-temperature spectrum to that expected for random coil and to the low-temperature spectrum to confirm that the protein did fully unfold. Compare the initial and final two low-temperature spectra. Does the sample fully refold? If so, this will allow a thermodynamic analysis of the unfolding data [19]. The refolding rate may differ substantially for each protein. If the two low-temperature spectra differ, and the latter is closer to the initial spectrum, then this implies that refolding is slow and future experiments on the protein should use slower cooling.

Calculate the difference spectrum of the spectra corresponding to the low- and high-temperature limits. What was the wavelength with the greatest difference in the signal? If this maximum difference is much larger than at the selected wavelength for recording the CD signal, data may be improved by repeating the experiment and selecting the maximum difference wavelength.

Inspect the plot of the CD signal at the selected wavelength against temperature. Typically, there will be a single sharp transition, the midpoint of which corresponds to (or at least approximates) the T_m. Where there is a more complex structure to the plot (e.g., multiple transitions that may correspond to different structured regions within the protein), this implies that less structurally resolved methods (e.g., DSF) will give results that cannot be automatically analyzed for T_m. In general, CD can be used to provide a check on the T_m derived by other methods and to compare T_m values of different batches or different protein variants.

Over what temperature range does the unfolding transition occur? Knowledge of this range can be used to optimize future thermostability experiments on this protein (e.g., is it possible to monitor unfolding over smaller range and thus carry out the experiment more quickly and reduce potential for aggregation?).

Information About the Tertiary Structure

Near-UV CD spectra show signals arising from aromatic residues, and disulfide bonds are extremely sensitive to their environment. The presence of defined signals in the near-UV region indicates that the protein is folded, whereas their absence suggests a nondefined tertiary structure (e.g., the protein is unfolded, misfolded, or in a molten globule state). The protocols for far-UV spectra can be

used, but the wavelength range must be adapted (250–340 nm) and greater protein concentration and/or path lengths are usually used to compensate for the smaller CD signal (and taking advantage of the lower molar absorbance) in this region.

2.3.2 DSF of Intrinsic Fluorophores

DSF of intrinsic tryptophan fluorescence is typically carried out on a dedicated instrument (such as the Prometheus—Nanotemper Technologies where samples are held and heated in capillaries). A backscattering measurement is also performed simultaneously to detect aggregation that may occur (before or after denaturation). DSF with intrinsic fluorophores can be effective with membrane proteins; however, in interpreting results, bear in mind that the detergent or amphipols used to solubilize them may disturb the 330-nm to 350-nm frequency shift upon which the method relies.

A simple protocol for buffer screening via DSF is presented below:

1. Switch on the instrument and let the system warm up for at least 15 min.
2. Prepare a stock solution of the target protein with an absorbance at 280 nm of ~1 AU.
3. Mix 1μL of protein with 10μL of each buffer to be screened. If available, preparation of buffers in multiwell plates with a robot or the use of commercial buffer-screening kits is recommended to avoid mistakes and to increase reproducibility.
4. Repeat the procedure for all buffer conditions to be tested.
5. Fill the capillaries with the sample solutions and put them into the DSF instrument.
6. Adjust the excitation power so that the fluorescence signal is higher than 2000 counts to get good signal-to-noise. If the signal remains lower than 2000 counts, increase the initial concentration of the protein.
7. Set the temperature range from 20 to 95 °C and select a temperature gradient. The choice of the temperature gradient is important as it is linked to the unfolding activation energy via the Arrhenius equation. Typically, a gradient of 0.5 °C/min is used.
8. Start the measurement.

Automated analysis provides T_m as the temperature of the maximum rate of change in the emission ratio. Visual inspection of the ratio vs. temperature plots should always be used to provided assurance that a simple two-state model of the unfolding process is appropriate. Changes in buffer can also lead to more complex changes in the plots (i.e., not only simple shifts of T_m) such as a transition from a complex multitransition trace (perhaps explicable

by partial unfolding and oligomerization) to a more desirable single transition.

2.3.3 Thermofluor/DSF of Extrinsic Fluorophores

Various fluorescent dyes are available to probe changes in protein conformation [38], and any of these may be used as long as the optical system of the instrument is compatible with their absorption and fluorescent emission properties. Common dyes are Nile red with a maximum excitation at 550–580 nm and emission in the range 590–690 nm (environment dependent), 1-anilinonaphthalene-8-sulfonic acid (1,8-ANS) with 370 nm excitation and 450–550 nm emission, and 4,4′-dianilino-1,1′-binaphthyl-5,5-′-disulfonic acid (bis-ANS) with 385 nm excitation/450–550 nm emission. SYPRO Orange is the most widely used dye for working with soluble proteins [45] because it has excitation and emission (460–520/550–650 nm) wavelengths that are compatible with almost all qPCR machines.

Thermofluor can also be applied to membrane proteins. However, interactions of the dye with the apolar external surface of the membrane protein or with detergents/amphipols may lead to a high initial signal background making changes in signal difficult to measure. Alternative dyes with lower background signal have been developed such as the thiol-specific fluorochrome *N*-[4 (7-diethylamino)-4-methyl-3-coumarinyl] maleimide (CPM) (385/470 nm) and Proteostat® dye (530/560–650 nm) [28]. Only a subset of qPCR instruments have optics suitable for use with these dyes.

In a thermofluor assay, the total volume per well is 25μL. Each condition tested requires 2μL of purified protein at an initial concentration of 20μM. SYPRO Orange (Invitrogen) is delivered as a 5000× concentrated stock solution in DMSO and needs to be diluted to a 62.5× solution in ultrapure water (3μL of SYPRO 5000× and 237μL of water), and 2μL of 62.5× SYPRO Orange is added per well. Do not premix the protein and the dye (the high concentration of DMSO may damage the protein) and follow the order of addition of components below for best results.

A simple protocol for buffer optimization via thermofluor is given below:

1. Put the plate on ice to equilibrate the sample at the starting temperature of the experiment.
2. Put sufficient buffer into each well to have finally a total volume of 25μL.
3. Add 2μL of 10× additive if needed.
4. Add 2μL of 20μM protein.
5. Add 2μL of 62.5× SYPRO Orange solution.
6. Seal the plate with highly transparent optical-clear quality sealing tape.

7. Centrifuge the plate at 4 °C at 2500 × *g* for 30 s to remove possible air bubbles.
8. Put the plate into the qPCR machine.
9. Program the qPCR machine with a 5-min equilibration time at 5 °C.
10. Set a heating ramp from 5 to 95 °C (20–95 °C if no cooling system is available) and a heating rate of 0.5 °C/min.
11. Record the fluorescence emission of SYPRO Orange at ~580 nm (excitation at ~485 nm).

Plot the change and rate of change of fluorescence against temperature to identify the number and position of transitions.

2.3.4 DSC

DSC is more sample- and time-consuming than Thermofluor but offers the possibility of analysis of thermostability of all proteins without the possible confounding effects of extrinsic dyes. Also, new calorimeters are able to use very dilute protein solutions (μM level) and work with much smaller volumes than has been historically the case, widening the scope for their use. Automated systems are also available. This technique should be used to complement intrinsic DSF or Thermofluor data to better characterize the condition in which the protein seems to be more stable.

A simple protocol for a DSC experiment is given below:

1. Preferably, switch on the DSC instrument the day before.
2. Set the temperature of the sample holding compartment to 5 °C.
3. Dialyze the sample against the buffer that will be used as a reference for the experiment (or collect both protein and the elution buffer used in the SEC column purification step).
4. Prepare a sample at a final concentration of 0.5–1 mg/mL.
5. Degas the sample and the buffer in vacuum (3–5 min with stirring).
6. Centrifuge or filter sample and buffer to remove any dust or aggregates if visible.
7. Load the reference buffer in both the reference cell and sample cell.
8. Set start and final temperature. Both depend on the protein. When prior knowledge is available, start >10 °C below and end >10 °C above the T_m to enable baseline matching in data analysis; otherwise, use the full range of the instrument.
9. Set scanning rate (e.g., 1 °C/min lower scan rate if multiple transitions are expected).
10. Set a wait time of 10 min before and after each temperature scan.

11. Set the instrument to repeatedly cycle.
12. Start the buffer:buffer run. Preferably, cycle multiple buffer: buffer runs overnight.
13. Just as a buffer:buffer run is complete, carefully remove buffer from the sample cell (do not dry the cell and do not stop the program) and load the protein sample.
14. Start the sample:buffer run.
15. When carrying out a series of experiments, rinse the sample cell with buffer twice between runs.

After completion of the experiment, data can be analyzed to obtain the transition temperature (T_m), the calorimetric enthalpy (ΔH), and the van't Hoff enthalpy (ΔH_{VH}) for the protein under study (more detail can be found in Ref. 32).

When comparisons are to be made between runs, use the same instrument settings (temperature range, scan rate, etc.) in each case. In a buffer-screening context, because the denaturation of most proteins is not reversible or not reversible in all conditions, data from the increasing temperature ramp only may be used for comparison (but not thermodynamic analysis).

2.4 Buffer Optimization

Optimizing the solution conditions of a protein sample is important but can be a tedious task. It is much simplified where there is access to instruments that can measure multiwell plates. Additionally, most (crystallography) core facilities have liquid handling and pipetting robots that can prepare small amounts of buffers with high precision directly into such plates. Alternatively, commercial buffer screens are available in preprepared plates. In order to exchange the buffer of the protein sample by the screening buffers, the protein sample is usually simply diluted ~10-fold with the screening buffer.

Systematic screening to maximize protein solubility would involve varying the pH from about 4 to about 9 using different buffers and testing each pH with concentrations of different salts (NaCl, KCl) in the range from 0 to 1000 mM. Other additives may be added, such as reducing agents (for proteins containing free cysteine residues), divalent ions or cofactors for those proteins dependent on them, and solubilizing agents such as detergents, sugars, and amino acids. In selecting buffers and salt concentrations, the downstream application must also be borne in mind; e.g., NMR studies benefit greatly from lowering conductivity of the buffer by lowering salt concentration and possibly compensating with increased buffer or additives such as amino acids, and many enzymes only function as catalysts in particular pH ranges. Typically, two rounds of screening will be required: the first to identify a broad range of pH, salt, and additives that are beneficial and the second with finer increments of the concentration of components

and finer increments of pH (and alternative buffers for the beneficial pH range) to enable optimization of solubility and stability.

A typical strategy would be to combine screening by DLS for homogeneity and solubility and a DSF method for thermostability. Having identified solution conditions giving good results using these methods, further investigation by CD (and DCS) should be used to confirm structural integrity in the conditions.

3 Notes

1. Whereas the presence of cofactors can confound concentration determination, a UV/Vis spectrum can also enable a fast and efficient batch-to-batch comparison relating to cofactor content and thus evaluate functional equivalence of different samples. Flavoproteins, for example, contain a flavin as a cofactor, and a UV/Vis spectrum of such a sample provides information on which flavin is bound (flavin mononucleotide (FMD) or flavin adenine dinucleotide (FAD)), how much flavin is present, and if the flavin is in the reduced or oxidized state [46].
2. In a PA concentration range of 10–19% containing 0.1% (w/v) SDS, there is a logarithmic relationship between MW and the migration distance of the SDS-polypeptide micelles, which may be exploited to determine more accurate estimates of the masses of proteins. The relative migration distance is determined as the ratio of the distance of the band of the protein to the distance of the buffer (which corresponds approximately to the distance covered by bromophenol blue in the sample buffer) from the beginning of the resolving gel. Then, a standard curve is generated by plotting the relative migration distances of the proteins of the size marker semi-logarithmically against their known molecular weights, followed by a least-squares fit of these data points. This standard curve can be used to estimate the molecular weights of the unknown proteins.
3. The European consensus on protein QC can be found on the websites of the European networks P4EU (https://p4eu.org/protein-quality-standard-pqs) and ARBRE-MOBIEU (https://arbre-mobieu.eu/guidelines-on-protein-quality-control).
4. In order to "polish" the purified protein and, in particular, to remove protein aggregates, the last step of the purification process should always be a SEC run. It is often necessary to further concentrate protein samples for their downstream applications. Unfortunately, this process, which is often done by ultrafiltration using spin concentrators or precipitation/resolublization protocols, very frequently induces aggregation.

Consequently, samples should not be concentrated beyond the absolutely required concentrations, avoiding overly high concentrations. Moreover, concentration of the samples should be followed by analytical SEC or DLS using an aliquot of the concentrated sample to ensure that it does not contain aggregates.

5. When measurements are performed with short path lengths, small manufacturing errors in path length will represent a larger percentage of the measurement; consequently, for quantitative analysis, short path-length cuvettes should be calibrated. Equipment that is cuvette-free (such as a Nanodrop (Thermo Scientific)) should be checked and maintained regularly for path length accuracy. Furthermore, be aware that the path length of such instruments may be disturbed by highly concentrated protein or glycerol-containing samples because of their high viscosity.
6. Buffer mismatch can be detected when the absorbance signal remains constant between 320 and 340 nm but not equal to zero. This is due to the buffer difference between the blank and the sample and should be a difference of not more than a few mAU. In that case, the baseline absorbance can be adjusted to zero, using the absorbance at 340 nm. See the manufacturer's instructions for the procedure. However, be aware that some software may use this option as default setting, which should be avoided! Similarly, if scattering affects the overall absorbance, the contribution of scattering can be removed by tracing a log–log plot of absorbance versus wavelength in the 320–340 nm region and then extrapolating the curve to the rest of the spectrum [1, 42]. The resulting value at 280 nm can then be removed from the A_{280} signal in order to calculate the concentration. However, concentration measurements will be less accurate if this correction is applied, and it should be avoided for measurements at 205 or 214 nm because the greater extrapolation will be more inaccurate.
7. Of commonly used buffers, phosphate, borate, MES, MOPS, and PIPES have relatively low absorbance in the far-UV, TRIS and HEPES have moderate absorbance and can only be used in low concentration, and DTT, 2-mercaptoethanol, and DMSO have very high absorbance in the main region of the far-UV CD signal (200–230 nm) and cannot be used in CD samples. If a reducing agent is required, TCEP has moderate absorbance. See http://www.uslims.aucsolutions.com/labresources.php for absorption spectra.
8. For other staining approaches to detect smaller quantities of the target protein or contaminants, we recommend the use of commercial staining kits as their protocols and formulations are

optimized and consistently manufactured to maximize reproducibility. Always bear in mind that the sensitivity of stains may be dependent on each protein's chemical composition, and practical detection limits for some proteins may be 10× higher than the optimal case.

The most sensitive colorimetric method is *silver staining* (optimally ~0.5 ng/band) [49, 50]. Proteins are usually fixed in the gel using trichloroacetic acid (TCA) and then incubated in a staining solution of silver nitrate. Silver ions bind to carboxylic acid, imidazole, and sulfhydryl sidechain groups and to amines. The subsequent development process reduces the protein-bound silver ions quickly to metallic silver, resulting in brown-black colored bands (although formulations exist to vary the color depending on protein charge and other features). A pH change is used to stop the reaction before all silver ions in the gel are reduced. The specificity and efficiency of silver ions binding to proteins as well as effective development require various sensitizers and enhancers, some of which (e.g., glutaraldehyde, formaldehyde) chemically crosslink the proteins in the gel matrix, limiting destaining and elution for downstream MS.

Reverse zinc staining is also highly sensitive (~1 ng/band) and works by staining areas not containing SDS-bound protein [51]; thus, bands are clear in an opaque background. Zinc ions bind to imidazole buffer, and the resulting complex precipitates in the gel matrix except where SDS-saturated proteins are located. Staining is fast (15 min), no fixation steps are needed, and the stain can be easily removed.

Fluorescent dye stains can be used if a fluorescence imaging system is available. The stains have high sensitivity (in the range 3–10 ng/band dependent on their chemistry) and high reproducibility. Most fluorescent stains are based on a noncovalent dye-binding mechanism, allowing complete destaining and recovering of the protein. SYPRO family dyes, which exhibit fluorogenic enhancement when interacting with hydrophobic groups of protein or SDS bound to protein, are commonly used, but alternative families of molecules such as LUCY, Krypton, and Flamingo may offer advantages of photostability and sensitivity.

Glycoproteins or phosphoproteins carry certain chemical moieties to which *functional group-specific stains* can be coupled. Other stains are able to selectively stain His-tags in PA gels without the need for His-tag-specific antibodies. However, such stains are not typically used in a QC context as they will only reveal the targeted subset of proteins and not contaminants.

In commercial, *stain-free methods*, the intrinsic fluorescence of Trp residues is enhanced by UV-light-controlled

formation of a covalent bond with a trihalo compound contained in the PA gel [52]. The sensitivity is comparable to that of Coomassie staining (20–50 ng protein). The modifications of the protein are minimal and do not mimic post-translational modifications so that the protein can be directly used in downstream applications such as MS and western blotting. The modified proteins can be visualized repeatedly on the gel or on blotting membranes. Provided the protein has Trp residues, stain-free methods are simple and highly reliable.

The most sensitive method to detect proteins is *autoradiography* or *fluorography*. Radioactive labeled proteins are visualized after gel electrophoresis by detecting radioactive emission of the protein bands on X-ray films. Autoradiography, however, is usually not used on an every-day-basis in protein QC and requires specialized equipment; *see* Ref. 53 for protocols.

9. Detailed settings will be instrument dependent. A typical starting point is 1 nm bandwidth, 0.2 –0.5 nm data pitch, 0.5 –1 s response time per point or scanning speed 50 nm/min, and 5–10 accumulations. The far-UV wavelength range is 260–185 nm and the near-UV wavelength range is 340–250 nm.

10. Quantitative comparison of spectra may require, or can be enhanced by, additional analyses. Since the measured CD signal will be proportional to the number of amino acids in the protein in addition to the concentration, in order to enable comparison between spectra of different proteins, the ellipticity θ is often converted to the molar ellipticity per residue θ_{MR}:

$$\theta_{MR} = 100 \cdot \frac{\theta}{N \cdot c \cdot d} \left[\frac{\text{deg} \cdot \text{cm}^2}{\text{dmol}}\right]$$

with θ being the measured ellipticity in mdeg, N the number of amino acids, c the protein concentration in mM, and d the path length in cm. Some instruments allow the simultaneous collection of an absorption spectrum; consequently, the absorption at 205 nm may be measured during data acquisition and can be used to determine the actual concentration in the cuvette (Subheadings 1.1 and 2.1). For accurate absorbance measurement, an additional background measurement of the instrument response without a cuvette will be required. For cuvettes of <1 mm, the path length should be determined experimentally via the Lambert–Beer law using a potassium chromate solution of known concentration.

Acknowledgments

The authors would like to thank Margarida Bastos for helpful comments on the manuscript.

References

1. Noble JE (2014) Quantification of protein concentration using UV absorbance and Coomassie dyes. Methods Enzymol 536:17–26. https://doi.org/10.1016/B978-0-12-420070-8.00002-7
2. Noble JE, Bailey MJA (2009) Chapter 8 Quantitation of protein. Methods Enzymol 463:73–95. https://doi.org/10.1016/S0076-6879(09)63008-1
3. Raynal B, Lenormand P, Baron B, Hoos S, England P (2014) Quality assessment and optimization of purified protein samples: why and how? Microb Cell Factories 13:180. https://doi.org/10.1186/s12934-014-0180-6
4. Pace CN, Vajdos F, Fee L, Grimsley G, Gray T (1995) How to measure and predict the molar absorption coefficient of a protein. Protein Sci 4:2411–2423. https://doi.org/10.1002/pro.5560041120
5. Glasel J (1995) Validity of nucleic acid purities monitored by 260nm/280nm absorbance ratios. BioTechniques 18:62–63
6. Andrews AT (1986) Electrophoresis: theory, techniques and biochemical and clinical applications. Clarendon Press, Oxford
7. Rothe GM (1994) Electrophoresis of enzymes. In: Laboratory methods. Springer, Berlin
8. Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680–685. https://doi.org/10.1038/227680a0
9. Krause F (2006) Detection and analysis of protein-protein interactions in organellar and prokaryotic proteomes by native gel electrophoresis: (Membrane) protein complexes and supercomplexes. Electrophoresis 27:2759–2781. https://doi.org/10.1002/elps.200600049
10. Schägger H, von Jagow G (1991) Blue native electrophoresis for isolation of membrane protein complexes in enzymatically active form. Anal Biochem 199:223–231. https://doi.org/10.1016/0003-2697(91)90094-a
11. Nobbmann U, Connah M, Fish B, Varley P, Gee C, Mulot S, Chen J, Zhou L, Lu Y, Shen F, Yi J, Harding SE (2007) Dynamic light scattering as a relative tool for assessing the molecular integrity and stability of monoclonal antibodies. Biotechnol Genet Eng Rev 24:117–128. https://doi.org/10.1080/02648725.2007.10648095
12. Fekete S, Beck A, Veuthey J-L, Guillarme D (2014) Theory and practice of size exclusion chromatography for the analysis of protein aggregates. J Pharm Biomed Anal 101:161–173. https://doi.org/10.1016/j.jpba.2014.04.011
13. Sahin E, Roberts CJ (2012) Size-exclusion chromatography with multi-angle light scattering for elucidating protein aggregation mechanisms. In: Voynov V, Caravella JA (eds) Therapeutic proteins, vol 899. Springer Nature, Cham, pp 403–423
14. Heck AJR, Van Den Heuvel RHH (2004) Investigation of intact protein complexes by mass spectrometry. Mass Spectrom Rev 23:368–389. https://doi.org/10.1002/mas.10081
15. Fenn JB, Mann M, Meng CK, Wong SF, Whitehouse CM (1989) Electrospray ionization for mass spectrometry of large biomolecules. Science 246:64–71. https://doi.org/10.1126/science.2675315
16. Tanaka K, Waki H, Ido Y, Akita S, Yoshida Y (1988) Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry. Rapid Commun Mass Spectrom 2:151–153. https://doi.org/10.1002/rcm.1290020802
17. Eakin CM, Miller A, Kerr J, Kung J, Wallace A (2014) Assessing analytical methods to monitor isoAsp formation in monoclonal antibodies. Front Pharmacol 5:87. https://doi.org/10.3389/fphar.2014.00087
18. Daviter T, Chmel N, Rodger A (2013) Circular and linear dichroism spectroscopy for the study of protein-ligand interactions. In: Williams M, Daviter T (eds) Protein-ligand interactions: methods and applications. Humana Press, Totowa, NJ, pp 211–241
19. Greenfield NJ (2006) Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. Nat Protoc 1:2527–2535. https://doi.org/10.1038/nprot.2006.204

20. Greenfield NJ (2006) Using circular dichroism spectra to estimate protein secondary structure. Nat Protoc 1:2876–2890. https://doi.org/10.1038/nprot.2006.202
21. Greenfield NJ (2006) Determination of the folding of proteins as a function of denaturants, osmolytes or ligands using circular dichroism. Nat Protoc 1:2733–2741. https://doi.org/10.1038/nprot.2006.229
22. Rule GS, Hitchens TK (2006) Fundamentals of protein NMR spectroscopy. Springer, Dordrecht
23. Zerbe O, Jurt S (2014) Applied NMR spectroscopy for chemists and life scientists. Wiley-VCH, Weinheim
24. Gräslund S et al (2008) Protein production and purification. Nat Methods 5:135–146. https://doi.org/10.1038/nmeth.f.202
25. Medrano G, Dolan MC, Condori J, Radin DN, Cramer CL (2012) Quality assessment of recombinant proteins produced in plants. In: Lorence A (ed) Recombinant gene expression. Humana Press, Totowa, NJ, pp 535–564
26. Dupeux F, Röwer M, Seroul G, Blot D, Márquez JA (2011) A thermal stability assay can help to estimate the crystallization likelihood of biological samples. Acta Crystallogr D Biol Crystallogr 67:915–919. https://doi.org/10.1107/S0907444911036225
27. Monsellier E, Bedouelle H (2005) Quantitative measurement of protein stability from unfolding equilibria monitored with the fluorescence maximum wavelength. Protein Eng Des Sel 18:445–456. https://doi.org/10.1093/protein/gzi046
28. Moon CP, Fleming KG (2011) Using tryptophan fluorescence to measure the stability of membrane proteins folded in liposomes. Methods Enzymol 492:189–211. https://doi.org/10.1016/B978-0-12-381268-1.00018-5
29. Žoldák G, Jancura D, Sedlák E (2017) The fluorescence intensities ratio is not a reliable parameter for evaluation of protein unfolding transitions. Protein Sci 26:1236–1239. https://doi.org/10.1002/pro.3170
30. Pantoliano MW, Petrella EC, Kwasnoski JD, Lobanov VS, Myslik J, Graf E, Carver T, Asel E, Springer BA, Lane P, Salemme FR (2001) High-density miniaturized thermal shift assays as a general strategy for drug discovery. J Biomol Screen 6:429–440. https://doi.org/10.1177/108705710100600609
31. Boivin S, Kozak S, Meijers R (2013) Optimization of protein purification and characterization using Thermofluor screens. Protein Expr Purif 91:192–206. https://doi.org/10.1016/j.pep.2013.08.002
32. Durowoju IB, Bhandal KS, Hu J, Carpick B, Kirkitadze M (2017) Differential scanning calorimetry - a method for assessing the thermal stability and conformation of protein antigen. J Vis Exp (121):e55262. https://doi.org/10.3791/55262
33. Quezada AG, Díaz-Salazar AJ, Cabrera N, Pérez-Montfort R, Piñeiro A, Costas M (2017) Interplay between protein thermal flexibility and kinetic stability. Structure 25:167–179. https://doi.org/10.1016/j.str.2016.11.018
34. Lebendiker M, Danieli T (2014) Production of prone-to-aggregate proteins. FEBS Lett 588:236–246. https://doi.org/10.1016/j.febslet.2013.10.044
35. Jancarik J, Pufan R, Hong C, Kim SH, Kim R (2004) Optimum solubility (OS) screening: an efficient method to optimize buffer conditions for homogeneity and crystallization of proteins. Acta Crystallogr D Biol Crystallogr 60:1670–1673. https://doi.org/10.1107/S0907444904010972
36. Wang J, Matayoshi E (2012) Solubility at the molecular level: development of a critical aggregation concentration (CAC) assay for estimating compound monomer solubility. Pharm Res 29:1745–1754. https://doi.org/10.1007/s11095-012-0730-8
37. Le Maire M, Champeil P, Møller JV (2000) Interaction of membrane proteins and lipids with solubilizing detergents. Biochim Biophys Acta 1508:86–111. https://doi.org/10.1016/S0304-4157(00)00010-1
38. Schuler MA, Denisov IG, Sligar SG (2013) Nanodiscs as a new tool to examine lipid–protein interactions. In: Kleinschmidt JH (ed) Lipid-protein interactions: methods and protocols. Humana Press, Totowa, NJ, pp 415–433
39. Stroud Z, Hall SCL, Dafforn TR (2018) Purification of membrane proteins free from conventional detergents: SMA, new polymers, new opportunities and new insights. Methods 147:106–117. https://doi.org/10.1016/j.ymeth.2018.03.011
40. Oluwole AO, Danielczak B, Meister A, Babalola JO, Vargas C, Keller S (2017) Solubilization of membrane proteins into functional lipid-bilayer nanodiscs using a diisobutylene/maleic acid copolymer. Angew Chem Int Ed 56:1919–1924. https://doi.org/10.1002/anie.201610778
41. Hardy D, Bill RM, Jawhari A, Rothnie AJ (2016) Overcoming bottlenecks in the membrane protein structural biology pipeline. Biochem Soc Trans 44:838–844. https://doi.org/10.1042/bst20160049

42. Aitken A, Learmonth MP (2002) Protein determination by UV absorption. In: Walker JM (ed) The protein protocols handbook. Humana Press, Totowa, NJ, pp 3–6
43. Kuipers BJH, Gruppen H (2007) Prediction of molar extinction coefficients of proteins and peptides using UV absorption of the constituent amino acids at 214 nm to enable quantitative reverse phase high-performance liquid chromatography-mass spectrometry analysis. J Agric Food Chem 55:5445–5451. https://doi.org/10.1021/jf0703371
44. Schägger H, von Jagow G (1987) Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. Anal Biochem 166:368–379. https://doi.org/10.1016/0003-2697(87)90587-2
45. Lo M-C, Aulabaugh A, Jin G, Cowling R, Bard J, Malamas M, Ellestad G (2004) Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. Anal Biochem 332:153–159. https://doi.org/10.1016/j.ab.2004.04.031
46. Macheroux P (1999) UV-visible spectroscopy as a tool to study flavoproteins. In: Chapman SK, Reid GA (eds) Flavoprotein protocols. Humana Press, Totowa, NJ, pp 1–7
47. Micsonai A, Wien F, Kernya L, Lee Y-H, Goto Y, Réfrégiers M, Kardos J (2015) Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. Proc Natl Acad Sci U S A 112: E3095–E3103. https://doi.org/10.1073/pnas.1500851112
48. Micsonai A, Wien F, Bulyáki E, Kun J, Moussong E, Lee Y-H, Goto Y, Réfrégiers M, Kardos J (2018) BeStSel: a web server for accurate protein secondary structure prediction and fold recognition from the circular dichroism spectra. *Nucleic Acids Res 46:W315–W322*. https://doi.org/10.1093/nar/gky497
49. Rabilloud T (1990) Mechanisms of protein silver staining in polyacrylamide gels: a 10-year synthesis. Electrophoresis 11:785–794. https://doi.org/10.1002/elps.1150111003
50. Rabilloud T (1992) A comparison between low background silver diamine and silver nitrate protein stains. Electrophoresis 13:429–439. https://doi.org/10.1002/elps.1150130190
51. Fernandez-Patron C (2005) Zn^{2+} reverse staining technique. In: Walker JM (ed) The proteomics protocol handbook. Humana Press, Totowa, NJ, pp 215–222
52. Edwards RA, Jickling G, Turner RJ (2002) The light-induced reactions of tryptophan with halocompounds. Photochem Photobiol 75:362–368. https://doi.org/10.1562/0031-8655(2002)0750362tlirot2.0.co2
53. Voytas D, Ke N (2001) Detection and quantitation of radiolabeled proteins in gels and blots. Curr Protoc Cell Biol Chapter 6:Unit 6.3. https://doi.org/10.1002/0471143030.cb0603s10

Chapter 2

A Familiar Protein–Ligand Interaction Revisited with Multiple Methods

Xiaochun Li-Blatter, Ludovit Zweifel, and Timothy Sharpe

Abstract

The interaction of hen egg white lysozyme with the trisaccharide tri-*N*-acetyl glucosamine has been well-characterized by biophysical methods and structural biology. In this chapter, we present a series of experiments designed to detect and quantify that interaction using several commonly available biophysical methods: thermal shift assay, fluorescence intensity, microscale thermophoresis, isothermal titration calorimetry, and surface plasmon resonance.

These experiments have been used for teaching and troubleshooting in a core facility. By taking a set of representative data from several years of practical courses, we are able to demonstrate the robustness of the protocols, calculate confidence intervals for the dissociation constant from each method, and illustrate the degree of consistency between those methods when applied to a simple system in a single location by different experimenters.

Key words Dissociation constant, Teaching, Orthogonal assay, Thermal shift, Fluorescence, Microscale thermophoresis, Isothermal titration calorimetry, Surface plasmon resonance, Global fitting

1 Introduction

1.1 Several Techniques, One Experimental System

This chapter presents a series of experiments using different techniques to detect binding and to measure the dissociation constant (K_d) for the interaction of the trisaccharide tri-*N*-acetyl glucosamine (NAG3) with the protein hen egg white lysozyme (HEWL). The idea to develop set experiments to measure the same interaction using different methods arose in the context of a biophysics core facility, where there is a regular need to train people to use instruments, and to troubleshoot and benchmark instruments after malfunction and repair.

This demands a simple and robust experimental system, with reagents that are readily available, easy to handle, and relatively inexpensive. The binding of NAG3 to HEWL seemed a good candidate system, and a series of experiments were developed over a few years, starting with ITC (*see* **Note 1**) and expanding to

Tina Daviter et al. (eds.), *Protein-Ligand Interactions: Methods and Applications*, Methods in Molecular Biology, vol. 2263, https://doi.org/10.1007/978-1-0716-1197-5_2,

include thermal shift by differential scanning fluorimetry, fluorescence intensity, microscale thermophoresis, and surface plasmon resonance.

It is undoubtedly the case that certain techniques are better suited to some experimental systems than others for a variety of physical reasons, and that some experimental systems afford a better chance to showcase the full information content of particular techniques. However, using a single experimental system for as many techniques as possible illustrates and encourages confirmation of results by orthogonal assays, and promotes discussion of the strengths and weaknesses of each type of measurement.

The body of data arising from many repeats of these experiments also provides a basis for discussion of techniques for analysis of replicate datasets, and for calculation of meaningful confidence intervals for K_d. Not only does this better inform the comparison of K_d between different techniques, but it also illustrates the level of agreement that can be expected for multi-technique analysis of a well-behaved and well-characterized system, when measured in a single location with a single set of instruments. The observed degree of agreement or variation gives a more realistic framework within which to evaluate the significance of variation observed for other projects, either between techniques, or by comparison with literature.

1.2 A Brief Introduction to HEWL and NAG3

Lysozymes are enzymes that cleave the glycosidic bond between the C-1 of *N*-acetyl muramic acid and the C-4 of *N*-acetyl glucosamine in bacterial cell wall peptidoglycan [1]. The apo structure of HEWL was the first enzyme structure to be solved by X-ray crystallography [2], and the structure of the NAG3-bound form was later solved by the same group [3]. NAG3 binds in the first three (A–C) of six subsites in the substrate-binding cleft in wild-type (WT) HEWL. NAG3 makes many interactions with the protein including an extensive hydrogen bond network, and hydrophobic and hydrogen-bonding interactions with two tryptophans (W62, W63). NAG3 binding causes small changes in the conformation and dynamics of the protein, particularly in the binding site [4, 5], and changes in water structure in and around the binding site [3]. Binding of NAG3 does not result in catalysis because the active site lies between subsites C and D, so NAG3 acts as a competitive inhibitor. The association of HEWL and NAG3 has been characterized by calorimetry and absorbance spectroscopy [6–8], giving estimates for K_d in the range 5–10 μM, with experimental conditions varying in the range pH 4.7–5.3 and ionic strength 68–100 mM.

1.3 A Brief Introduction to the Techniques Used

The aim of this chapter is not to give an extensive theoretical introduction to each technique, since a wealth of such information is available elsewhere in this volume. However, for the sake of those readers who may be unfamiliar with a particular technique and have not yet consulted other chapters, a short introduction is given below.

Thermal shift assay: This assay exploits the stabilization of proteins against thermal denaturation by mass action (Le Chatelier's principle) when small-molecule ligands bind selectively to the folded state and not to the unfolded state [9]. This causes a shift in the midpoint of the typically sigmoidal unfolding transition, which can be monitored by a variety of biophysical methods including differential scanning fluorimetry (DSF) using dyes that change their signal upon interacting with unfolded protein or using changes in intrinsic protein fluorescence upon unfolding. It is a high-throughput, low-consumption assay that can be used to screen for the presence of binding, to identify promising candidates for analysis by other methods, and to rank molecules that bind to a particular site in order of affinity [10, 11]. For further details about DSF and thermal shift assays, readers should consult Chapter 8.

Fluorescence intensity: This technique measures changes in intrinsic protein fluorescence intensity upon ligand binding, resulting from changes in the electronic and solvent environment of tryptophan and tyrosine residues in the bound and unbound states of the protein. This assay uses unadulterated native protein and is simple to perform for small-molecule ligands that do not absorb strongly or fluoresce in the same wavelength range as proteins.

Microscale thermophoresis (MST): This technique measures changes in partitioning of macromolecules across a local temperature gradient (thermophoresis) upon ligand binding, resulting from changes in surface area, ionic-shielding entropy, and solvation entropy [12]. Changes in distribution of one binding partner (which must be fluorescent, *see* **Note 2**) due to thermophoresis at different concentrations of ligand are quantified via changes in fluorescence intensity in a region that is transiently heated by an infrared laser. This is a low-consumption assay and applicable to many experimental contexts [13]. For further details about MST, readers should consult Chapter 6.

Isothermal titration calorimetry (ITC): This technique measures the heat absorbed or evolved upon titration of ligand solution into protein solution, using power compensation calorimetry. The characteristic pattern of heat transfer during the titration depends upon the enthalpy change for binding and the dissociation constant for the interaction [14, 15]. Although requiring relatively large amounts of material, this technique is close to universal in applicability to biomolecular binding events and can afford a more complete picture of the thermodynamics of binding than other

techniques [16, 17]. However, the thermodynamics of binding are very hard to interpret unambiguously from a single experiment, so in this set of experiments, the primary use of ITC is to measure the dissociation constant for the interaction. For further details about ITC, readers should consult Chapter 5.

Surface plasmon resonance (SPR): This technique measures the properties of light reflected from a metal–aqueous interface, which are sensitive to the refractive index in the aqueous layer at the metal surface [18, 19]. If one binding partner (the "ligand" in SPR terminology) is immobilized at the surface and another binding partner (the "analyte") is introduced in flow across the surface, the refractive index in the aqueous layer changes as the analyte is accumulated due to interaction with the ligand, giving rise to a measurable change in the properties of the reflected light. By alternating between flow containing analyte, and flow without analyte, an association phase, an equilibrium response, and a dissociation phase can be observed. For interactions with association and dissociation rates in the right range, information on both binding kinetics and dissociation constants can be obtained. However, for the interaction presented here, the kinetics are too fast to be measured by the instrument at concentrations that give measurable signal changes, and so the rate constants cannot be obtained. Therefore, the focus of the experiment is to obtain the dissociation constant for the interaction from the steady-state value of the instrumental response. For further details about SPR, readers may consult Chapter 17.

For the sake of brevity, only those technical details necessary to ensure successful replication of the experiments will be included, together with typical results taken in the most part from student practicals, perhaps the ultimate testing ground for experimental robustness and replicability. Some of the experimental details are necessarily instrument-specific, and it is hoped that readers can use these as a starting point for adaptation to their own instruments.

The experimental protocols should not be regarded as optimized, nor should the results be regarded as definitive, though most of the protocols have been successfully reproduced in different labs. Rather, it is hoped that this information will prove helpful to others engaged in practical teaching of biophysics or embarking on a project to quantify a protein–ligand interaction, and that this chapter might serve as a starting point for further optimization and expansion.

1.4 Nature and Analysis of Example Data

To illustrate the experimental protocols and to give an indication of their robustness, this chapter will present and analyze five or more datasets for each technique, in most cases collected during student practical classes in the last 5 years. As might be expected for student practicals, the standards of pipetting and sample-handling were somewhat variable. Datasets were selected essentially at random,

with the only proviso being that they exhibited typical experimental scatter (without obvious errors in the sample preparation) and were from a range of different years.

It is important to consider that these data were not collected with systematic analysis or multimethod benchmarking in mind. Firstly, for systematic analysis, the stock solutions would be ideally identical (to control for variability in sample preparation). However, the stocks for these experiments were often prepared fresh for individual experiments (HEWL solutions, NAG3 dilutions), or stocks were refreshed at least once a year (NAG3 stocks). Only the stock of fluorescently labeled HEWL for MST measurements was constant for all experiments. Secondly, for systematic analysis, a larger sample of different instruments, operated by experts, would be required. Therefore, the analyses in this chapter are emphatically not intended as a statistically reliable exploration of the real error and consistency in determining the K_d for this interaction by different methods. Rather, the data are an example of the typical variation that one might expect when using these protocols in the context of teaching, at a single location (*see* **Note 3**).

There are a number of different ways in which the data from independent replicate experiments could be analyzed, and there is not sufficient space here to explore them all. Nor are the data necessarily of sufficient quality and consistency to make such a comparison worthwhile. Therefore, to assess the robustness of the assay over the replicates presented for each technique, we have chosen a single approach: to globally fit the five datasets to a shared value of K_d. If all of the replicates for each technique are measuring the same interaction under the same conditions, with sufficient data quality, it should be possible to fit the datasets to a single shared value of K_d.

In performing a global fit with more than one fitting parameter (in addition to experimental dependent and independent variables), one must make a decision on whether to fit other parameters in the fitting equation as a global parameter (a shared value across all datasets) or a local parameter (an individual value for each dataset). This decision depends on whether that parameter is measured by an instrument with an absolute calibration, and upon the apparent consistency of that absolute calibration.

For example, in fitting fluorescence intensity data from different plate readers, with different (fixed) bandwidths of excitation and emission, it is not expected that the fluorescence intensities of free and bound HEWL would be consistent between datasets, even after normalization, since the amplitude of signal change will be bandwidth-dependent to some degree. Therefore, it seems sensible to allow the fluorescence intensities of free and bound HEWL to be fitted as a local parameter. A similar treatment has been applied here to MST data (where small variations between datasets can be seen in the signals for free and bound HEWL, even for the same

instrument over time), and SPR data (where the saturating biosensor response is proportional to the amount of HEWL immobilized, which varies between experiments). However, for ITC instruments from a single manufacturer, it might be expected that the calibration and measurement of differential power are sufficiently accurate and precise that the enthalpies of binding should be consistent between instruments. Therefore, in globally fitting the ITC data presented here, both the K_d and ΔH are fitted as global parameters. Further details of the fitting are provided in later sections.

On advantage of fitting K_d as a global parameter for all five datasets for each technique is that it allows the method of error surface projection (ESP) to be used to calculate a single confidence interval for that global K_d, taking into account all of the variability in the replicates. There is insufficient space to cover this method in detail here, but it has been widely employed for determination of confidence limits on fitted parameters and is described in more detail elsewhere [20]. Essentially it calculates the range of one fitted parameter that can fit the data at a given confidence limit, if other floating parameters are allowed to vary. The confidence limits derived in this way can be asymmetric about the best-fit value, in contrast to other methods that assume a Gaussian error distribution. This approach of global fitting and error analysis using ESP seems the most straightforward way to include many sources of experimental error into a realistic estimate of the confidence in a single parameter of interest measured by a single technique.

2 Materials

2.1 Buffers

- Standard assay buffer: 0.1 M sodium acetate/acetic acid, pH 5.0 (*see* **Note 4**).
- MST Buffer: 0.1 M sodium acetate/acetic acid, pH 5.0, 0.1% v/v Tween 20 (*see* **Note 5**).
- MST labeling buffer: 50 mM sodium phosphate, pH 7.5, 150 mM NaCl.
- SPR immobilization buffer: 50 mM sodium phosphate, pH 8.5.

2.2 Protein and Ligand

- Hen egg white lysozyme (HEWL): Molar mass: 14,313 g/mol, ε_{280} 37,970 M^{-1} cm^{-1}.
- Tri-*N*-acetyl glucosamine (NAG3): Molar mass: 627.6 g/mol (*see* **Note 6**).
- Various stock concentrations of HEWL and NAG3 are used in this chapter to suit different types of experiment. For detailed preparation instructions, please *see* Subheading 3.1.

2.3 Other Reagents

- DyLight488 NHS-ester dye: 1 × 50 μg aliquot.
- SYPRO Orange: 1 vial of 5000× stock solution (*see* **Note 7**).
- Biacore Amine Coupling Kit.

2.4 Instruments

- **Thermal shift assay:** Differential scanning fluorimetry (DSF) measurements on HEWL in the presence and absence of NAG3 should be performed using a thermocycler designed for quantitative polymerase chain reactions (qPCR). The instrument should be capable of measuring fluorescence emission for the dye SYPRO Orange (excitation maximum 492 nm, emission maximum 610 nm) during a continuous thermal ramp. The data presented here were measured using a Qiagen Rotor-Gene Q, employing a combination of the green excitation and orange emission channels for fluorescence detection. Examples of other suitable instruments are given in Ref. 10. Samples should be prepared and measured in PCR tubes or plates suited to the instrument used.
- **Fluorescence intensity titration:** The intensity of intrinsic fluorescence for titration of NAG3 vs. HEWL should be measured with a plate reader with monochromators or filters allowing measurement in the UV. The data presented here were measured using a BioTek Synergy H1 plate reader, with an excitation wavelength of 290 nm, an emission wavelength of 350 nm, and a fixed bandwidth of 18 nm in excitation and emission (manufacturer's specification). Essentially identical results were achieved with other plate readers (Tecan, Molecular Devices) with similar settings. The samples for fluorescence intensity measurements presented here were prepared and measured in Corning nonbinding surface flat-bottom 384-well black microplates (e.g., Corning 3575). It is anticipated that other brands of microplates with nonbinding surfaces ought to give equivalent results.
- **MST:** Measurements for titration of NAG3 vs. HEWL should be performed using an instrument with fluorescence excitation and emission optics that are suitable for the DyLight488 dye (excitation maximum 493 nm, emission maximum 518 nm). The data presented here were measured with a Nanotemper Monolith NT.115 instrument using the blue channel excitation and emission optics and standard capillaries. The samples for MST measurements presented here were prepared in 200-μL microtubes provided by Nanotemper, though it is anticipated that other microtubes or microplates with a nonbinding surface could be used.
- **ITC:** Measurements should be made using a titrating power compensation calorimeter with a suitable sensitivity and working volume. The data presented here were measured using two

Microcal VP-ITC instruments and a Microcal iTC200. These ITC instruments have similar cell geometries, but the working volume of the VP-ITC is sevenfold larger.

- **SPR:** Measurements should be made with an instrument with appropriate sensitivity for small molecule binding to proteins (the analyte:ligand mass ratio is approximately 0.04). The instrument should be capable of accepting sensor chips amenable to immobilization of ligand via free amines using NHS-ester chemistry. The data presented here were measured using a Biacore T200 with Biacore S-Series CM5 (carboxymethyl dextran coated) sensor chips. Samples should be prepared and measured in plasticware recommended by the instrument manufacturer.

3 Methods

3.1 Preparation of Protein and Ligand Solutions

3.1.1 Preparation of HEWL Solution

1. The HEWL solution should be prepared by raising the lyophilized powder in the standard assay buffer at approximately the correct concentration (a typical stock concentration would be 0.5 mg/mL, corresponding to 35 μM).
2. This solution should be filtered through a 0.2-μm syringe filter (with a polyvinylidene fluoride membrane) to remove any larger particulate material.
3. The accurate concentration of the HEWL solution should then be measured by absorbance spectrophotometry, using the standard assay buffer as a reference (*see* **Note 8**). The measured absorbance should be corrected for any contributions from scattering using the formula $A_{280}^{\text{corrected}} = A_{280}^{\text{measured}} - 2*A_{333}$ (*see* **Note 9**). The concentration should be calculated from the corrected 280 nm absorbance using the Beer–Lambert law and a molar extinction coefficient (ε_{280}) of 37,970 M^{-1} cm^{-1} calculated using ProtParam from the primary sequence (*see* **Note 10**).
4. The HEWL solution should be prepared fresh before experiments and can be kept at room temperature for a few hours before use. We do not use the solution for more than 1 day, preferring always to make a new stock.

Obtaining accurate concentration for the HEWL solution is crucial for consistency and accuracy of the experiments, especially in those experiments where the working concentration is on the same order as the K_d (fluorescence, ITC), such that concentration errors have a significant effect on fitted parameters. Where possible, it is best to measure the concentration of the experimental solution directly to avoid concentration errors from dilution. For example, a working concentration of 35 μM for an ITC experiment gives an absorbance of 1.33 in a 1 cm pathlength cuvette, which can be measured reliably.

3.1.2 Preparation of NAG3 Solution

1. Weigh out an appropriate mass of the lyophilized powder of NAG3 into a microtube using an analytical balance. Make sure the mass is sufficient for accurate weighing in the balance you are using (usually at least 1 mg).
2. Raise the powder directly in the required volume of the standard assay buffer to give the required concentration. NAG3 is soluble up to at least 2 mM and is apparently stable for >6 months when stored at −80 °C at that concentration.

It is important that the solution of NAG3 is prepared accurately to ensure consistency and accuracy of experimental parameters. In all experiments, concentration errors for the NAG3 solution have a significant effect on fitted parameters. These methods for determining protein and ligand concentration have in our hands provided apparently accurate concentrations, as judged by routine observation of interaction stoichiometries between 0.95 and 1.05 in ITC experiments (*see* **Note 11**).

3.2 Thermal Shift Assay

3.2.1 Thermal Shift Assay Measurement

The aim is to measure the midpoint of thermal denaturation for HEWL in the presence and absence of a saturating concentration of NAG3.

1. Make 500 μL of 20 μM HEWL in standard assay buffer.
2. Make 500 μL of 2 mM NAG3 solution in standard assay buffer.
3. Add SYPRO Orange dye directly to the protein solution at a 250-fold dilution of the dye stock concentration, i.e., 2 μL dye solution in 500 μL protein solution. Mix thoroughly with a pipette. The stock solution is supplied by the manufacturer at a fixed, but unstated concentration labeled as "5000×."
4. Mix samples for differential scanning fluorimetry according to Table 1. After mixing, the final concentration of HEWL will be 10 μM with 10× SYPRO Orange in the presence and absence of 1 mM NAG3 (*see* **Note 12**).
5. Place an appropriate volume of the sample into micro PCR tubes or plates, preferably in triplicate for each condition.
6. Perform a thermal ramp from 25 to 95 °C at a ramp rate in the region of 5 °C/min, monitoring the fluorescence intensity of

Table 1
Volumes for preparation of samples for the thermal shift assay

Sample	HEWL (μL) (in buffer with dye)	Buffer (μL) (no dye)	2 mM NAG3 (μL) (in buffer, no dye)
1	200	200	0
2	200	0	200

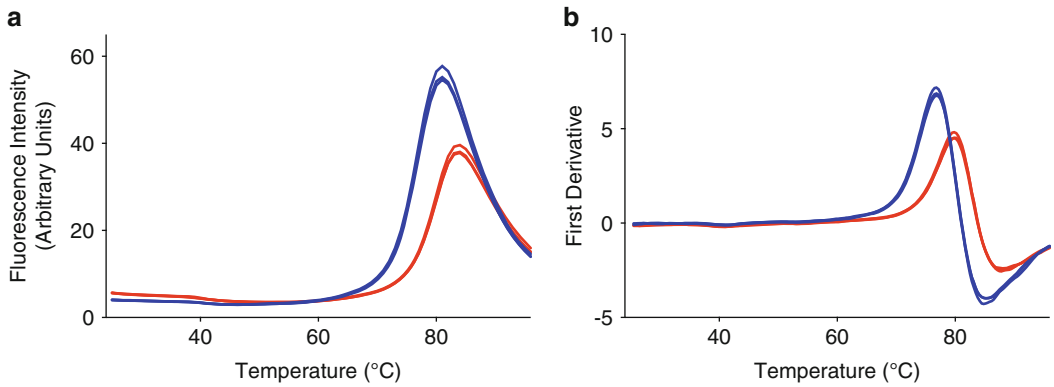


Fig. 1 DSF raw fluorescence intensity data (**a**) and first derivative (**b**) for HEWL in the presence (red) and absence (blue) of 1 mM NAG3

SYPRO Orange using appropriate optics and a detector gain (*see* **Note 13**) that ensures that the dynamic range of the instrument is not exceeded at higher temperatures (i.e., do not allow the fluorescence intensity to exceed the saturating value for the detector).

7. Fluorescence intensity data should be acquired at 1 °C intervals, or more frequently if greater reproducibility and precision of thermal denaturation midpoint is desired (*see* **Note 14**).

3.2.2 Thermal Shift Assay Data Analysis and Typical Results

1. The thermal denaturation experiment should yield an approximately sigmoid denaturation curve with a steeply sloping post-transition baseline, similar to that shown in Fig. 1a.
2. Take the first derivative of the fluorescence intensity data. Most thermocycler software will perform this automatically as part of the standard analysis (*see* **Note 15**). The first derivative of the raw data in Fig. 1a is shown in Fig. 1b.
3. Find the maximum of the first derivative, which corresponds (approximately) to the apparent midpoint of thermal denaturation (T_m). Most thermocycler software will do this automatically, with an adjustable threshold so that noisy baseline regions can be excluded from the search for maxima.
4. Average the T_m values from replicate measurements and calculate the ΔT_m according to $\Delta T_m = T_m^{\text{HEWL+NAG3}} - T_m^{\text{HEWL}}$.

In five replicates of this experiment, each consisting of triplicate samples for each condition, the mean T_m for HEWL was 76.5 ± 0.2 °C and the mean T_m for HEWL + 1 mM NAG3 was 79.6 ± 0.3 °C, and the mean ΔT_m was 3.2 ± 0.4 °C (*see* **Note 16**). Thus, there is clear evidence of binding of NAG3 to HEWL under these conditions, indicated by a significant stabilization of HEWL in the presence of 1 mM NAG3.

3.3 Fluorescence Intensity

3.3.1 Preparation of Titration for Fluorescence Intensity

The aim is to make a 24-point titration of NAG3 against a constant concentration of HWEL, with an appropriate range and point-density for accurate fitting of the dissociation constant (K_d).

1. Make a 4-μM solution of HEWL and a 450-μM solution of NAG3, both in the standard assay buffer, according to the method outlined in Subheading 3.1.
2. Choose one row of a 384-well microplate and dispense 80 μL of the 450 μM NAG3 solution into the first well of that row.
3. Dispense 20 μL of the standard assay buffer into the remaining 23 wells of that row.
4. Take 60 μL of solution from the first well in the row (containing the NAG3 solution), touching the tip on the inside of the well as you remove it, to remove residual solution on the outside of the tip.
5. Pipette the solution into the second well in the row (which should already contain 20 μL of buffer). Mix at least 4 times, aspirating at least two-thirds of the total solution in the well. Keep the pipette tip under the meniscus at all times to prevent air bubbles. Touch the tip on the inside of the well as you withdraw it.
6. Change the pipette tip and transfer 60 μL of the mixed solution in the second well into the third well. Mix, then repeat this process for the remaining wells.
7. Remove 60 μL of solution from the final well in the row so that all wells contain 20 μL of solution.
8. This will produce a 24-point serial dilution of NAG3 with the dilution ratio of 3:1 (a dilution factor at each step of 1.33).
9. Now add 50 μL of the 4 μM HEWL solution to each well of the row containing the NAG3 solutions. Mix the solution well (*see* **step 5** above).
10. This will give a 24-point titration with a varying concentration of NAG3 and a constant concentration of 2.85 μM HEWL in a working volume of 70 μL (*see* **Note 17**).

To obtain a more reliable estimate of the K_d, with more realistic errors, you could include one or more replicate titrations in other rows of the same plate. It is important to repeat the serial dilution for the replicate titration so that the serial dilution errors are accounted for in the error estimates for the fitted parameters.

3.3.2 Fluorescence Intensity Measurement

1. Incubate the titration at the measurement temperature for 10 min prior to measurement. Place a lid on the microplate to avoid sample evaporation.

2. Place the microplate into the plate reader and measure fluorescence intensity at 25 °C, or as close to that temperature as the ambient conditions in the plate reader will allow.
3. Fluorescence should be excited at 290 nm and emission should be measured at 350 nm, with the narrowest bandwidth available in the plate reader used (typically somewhere between 2 and 20 nm) (*see* **Note 18**).
4. The gain for the fluorescence detector should be set by scanning all wells in the titration and setting the gain such that the well with the highest fluorescence intensity has a measured value below the saturating value for the detector. Most plate readers can perform this adjustment automatically. The well with the lowest concentration of NAG3 should have the highest fluorescence intensity.
5. Some plate readers can also perform an automatic *z*-height adjustment for optimal signal intensity. If possible, this should be performed on the same well.

3.3.3 Fluorescence Intensity Data Analysis and Typical Results

The aim is to analyze the fluorescence intensity titration to obtain a value of K_d for the interaction between HEWL and NAG3.

1. Plot the fluorescence intensity against the molar concentration of NAG3. It is easier to visualize the scatter of the data and the quality of fits to the data if they are plotted on a logarithmic *x*-axis.
2. Fit the data to Eq. 1 using software capable of performing nonlinear regression.

$$S_{obs} = S_f + (S_b - S_f).\frac{([P]_t + [L]_t + K_d) - \sqrt{([P]_t + [L]_t + K_d)^2 - 4[P]_t[L]_t}}{2[P]_t} \quad (1)$$

where S_{obs} is the observed experimental signal (the *y*-axis variable); S_f is the experimental signal for free HEWL (this parameter should be fitted, an initial estimate can be obtained from the signal at the lowest concentration of NAG3) (*see* **Note 19**); S_b is the experimental signal for HEWL bound to NAG3 (this parameter should be fitted, an initial estimate can be obtained from the signal at the highest concentration of NAG3); $[P]_t$ is the total concentration of HEWL (this parameter should fixed to the known value); $[L]_t$ is the total concentration of NAG3 at a given titration point (the *x*-axis variable); and K_d is the dissociation constant for the interaction (this parameter should be fitted, an initial estimate can be obtained by taking the concentration of NAG3 at the midpoint of the sigmoid curve when the data are plotted on a logarithmic *x*-axis).

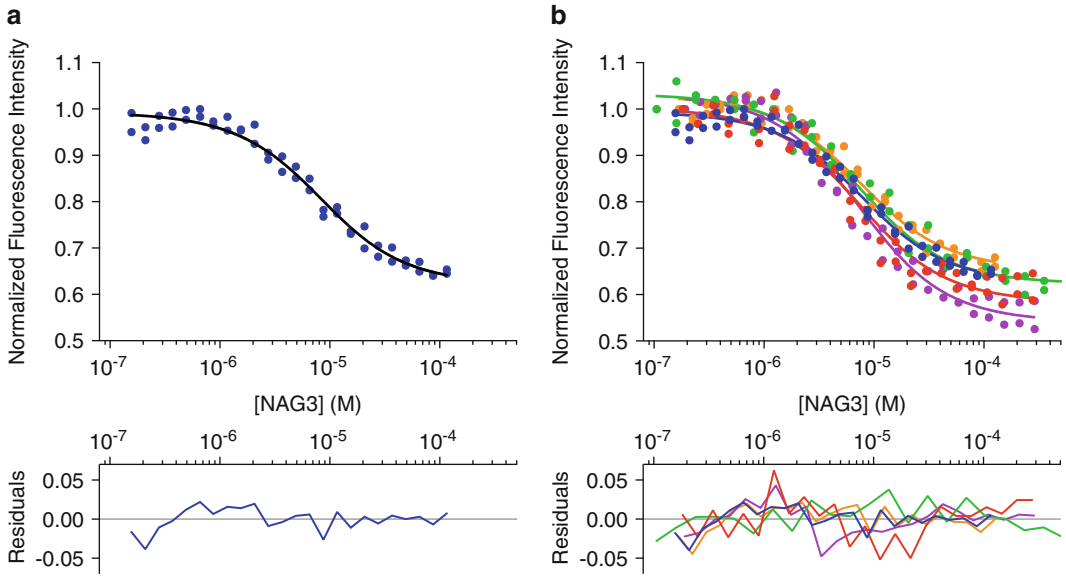


Fig. 2 (**a**) Fluorescence intensity at 350 nm vs. NAG3 concentration, for a single dataset containing two replicate titrations of HEWL with increasing concentrations of NAG3 in standard assay buffer at approximately 25 °C. (**b**) Five datasets, each consisting of two replicate titrations. Solid lines indicate the best local (**a**) or global (**b**) fit to Eq. 1, and fitted parameters are given in the main text

An example dataset containing two replicate titrations from the same stocks, together with a fit to Eq. 1, is shown in Fig. 2a. To facilitate comparison of datasets collected using different instruments, the raw fluorescence intensity was normalized by dividing each intensity value by the maximum intensity in the whole titration. The solid black line in Fig. 2a is the best fit to that single dataset using Eq. 1, treating each point from the two replicate titrations as a separate data point. The best-fit value of K_d is 6.5 μM, with a 95% confidence interval of 5.2–8.2 μM.

To assess the robustness of this experiment, five repeats, each consisting of two replicate titrations, were analyzed in a global fit. The five datasets were fitted to individual values of free and bound fluorescence intensities and a shared value of K_d. The normalized data and the lines of best fit for the global fit to Eq. 1 are shown in Fig. 2b. The globally fitted K_d was 6.1 μM with 95% confidence limits of 5.5–6.8 μM (*see* **Note 20**).

3.4 Microscale Thermophoresis (MST)

3.4.1 Preparation of Fluorescently Labeled HEWL for MST

The aim is to produce HEWL labeled with a fluorescent dye suitable for detection by the optics in the MST instrument at a stoichiometry of approximately 1 dye molecule per HEWL molecule.

1. Make 0.5 mL of a 1-mg/mL HEWL solution in MST labeling buffer.
2. Mix that HEWL solution with one 50-μg aliquot of DyLight 488 (*see* **Note 21**).

3. Incubate the labeling reaction for 1 h at room temperature in a foil-covered tube.
4. After incubation, dilute the labeling reaction solution to a total of 2 mL volume using the labeling buffer.
5. Dialyze the labeling reaction solution against three changes of 300 mL of standard assay buffer using a dialysis cassette or membrane with a nominal molecular weight cutoff of 10 kDa.
6. Measure the absorbance of the dialyzed labeling solution at 280 and 493 nm using fresh standard assay buffer as a reference. It should be possible to measure the absorbance directly using a small-volume 1-cm pathlength cuvette.
7. Calculate the concentration of DyLight488 from the absorbance at 493 nm using the Beer–Lambert law and a molar extinction coefficient of 70,000 M^{-1} cm^{-1}.
8. Calculate the contribution of DyLight488 to the measured absorbance at 280 nm by multiplying the measured absorbance at 493 nm by 0.147 (this factor is specific to the dye and represents its relative absorbance at 280 nm compared to 493 nm).
9. Subtract this contribution from DyLight488 from the total measured absorbance at 280 nm.
10. Use the remainder, corresponding to the absorbance from HEWL at 280 nm, to calculate the HEWL concentration using the Beer–Lambert law and a molar extinction coefficient of 37,970 M^{-1} cm^{-1}.
11. Calculate the ratio of DyLight488:HEWL. This should be approximately 1:1 (*see* **Note 22**).
12. The Dylight488 labeled HEWL (HEWL-D488) should be dispensed into PCR tubes in 20-μL aliquots, cryo-cooled using liquid nitrogen, and stored at −80 °C. It has proved stable under these conditions for more than 5 years.

3.4.2 Preparation of Titration for MST

The aim is to make a 16-point titration of NAG against a constant concentration of HEWL-D488, with an appropriate range and point-density for accurate fitting of the dissociation constant (K_d).

1. Make 90 μL of 300 μM NAG3 in MST buffer in a PCR tube.
2. Using that 90 μL as the first point, make a 16-point, 2:1 serial dilution from that stock. Follow a similar procedure to that outlined in detail in Subheading 3.3.1, but in this case, mix 60 μL of the preceding higher concentration in the serial dilution with 30 μL of buffer at each step. Remember to remove 60 μL of solution from the final tube in the titration series so that all tubes contain 30 μL of solution.

3. Thaw out one or more aliquots of the HEWL-D488 and make 600 μL of a 300 nM solution in MST buffer.
4. Add 30 μL of this HEWL-D488 solution to the 30 μL of NAG3 solution in each tube of the serial dilution and mix well by pipetting (*see* **Note 23**).
5. Incubate the samples for 15 min at 25 °C and then load them into standard MST capillaries (*see* **Note 24**).

3.4.3 MST Measurement

1. Set the instrument temperature to 25 °C and load the capillaries into the holder.
2. Use the blue channel excitation and detection with an LED intensity of 40% (*see* **Note 25**).
3. Run the "find capillaries" routine and confirm that all capillaries exhibit fluorescence intensity between 200 and 1500 units, with minimal adsorption to the capillary surface evinced by an approximately Gaussian distribution of intensity across the capillary profile. If the fluorescence intensity lies outside the desired range, adjust the LED intensity accordingly.
4. Perform an MST experiment, with a heating laser intensity of 40%, a laser on time of 30 s, and a laser off time of 5 s. Following normalization of the fluorescence intensity (by dividing the value at all time points by the value in the initial pre-heating period), the raw thermophoresis curves for the titration should resemble those in Fig. 3.

3.4.4 MST Data Analysis and Typical Results

The aim is to analyze the MST titration data to obtain a value of K_d for the interaction between HEWL-D488 and NAG3.

1. Calculate F_{norm} (‰) for the "thermophoresis plus jump" phase of the fluorescence vs. time curve in the software (*see* **Note 26**). For the data shown in this chapter, the *per mille* ratio of the mean normalized fluorescence values was calculated for

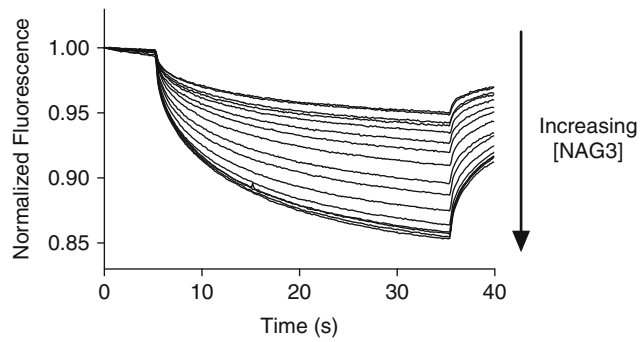


Fig. 3 Normalized fluorescence thermophoresis curves for titration of HEWL-D488 with increasing concentration of NAG3 in MST buffer at 25 °C

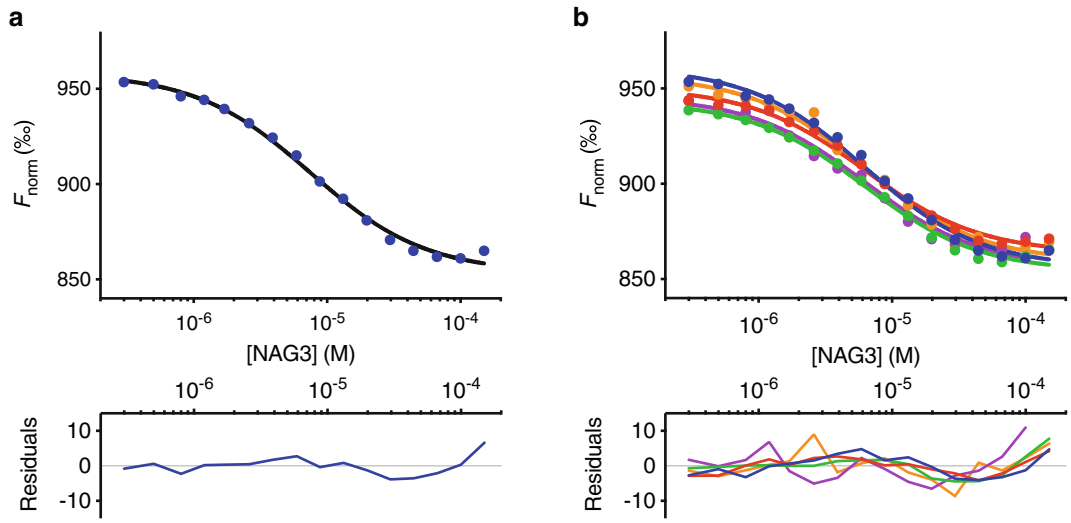


Fig. 4 (**a**) F_{norm} (‰) vs. NAG3 concentration, from a single dataset containing microscale thermophoresis measurements of a titration of HEWL-D488 with increasing concentration of NAG3 in MST buffer at 25 °C. (**b**) Five datasets, each consisting of a single titration, including that shown in (**a**). Solid lines indicate the best local (**a**) or global (**b**) fit to Eq. 1, and fitted parameters are given in the main text

windows of 3 s, centered at time points of 30 s (hot) and 2.5 s (cold). Since the signal plateau amplitude decreases in the presence of increasing NAG3 concentration, the *per mille* ratio also decreases.

2. Plot F_{norm} (‰) against the molar concentration of NAG3 and fit the data to Eq. 1 (Subheading 3.3.3). This can be performed either in the Nanotemper Analysis software (the equation used is essentially identical to Eq. 1) or by exporting the F_{norm} (‰) data to any software capable of nonlinear regression.

An example of a single dataset together with a fit to Eq. 1 is shown in Fig. 4a. The solid black line is the best fit to that single dataset using Eq. 1. The best-fit value of K_d is 7.5 μM, with a 95% confidence interval of 6.2–9.0 μM.

To assess the robustness of this experiment, five repeats, each consisting of a single titration, were analyzed in a global fit. The five datasets were fitted to individual values of free and bound F_{norm} and a shared value of K_d. The data and the lines of best fit for the global fit to Eq. 1 are shown in Fig. 4b. The globally fitted K_d was 6.2 μM with a 95% confidence interval of 5.5–7.0 μM.

This fitted value gives, on first inspection, apparently good agreement with the value determined for unlabeled HEWL using fluorescence intensity measurements. However, the situation is more complicated. Inspection of the fluorescence intensity for HEWL-D488 in the titration for MST measurements shows that there is a significant and systematic increase in fluorescence

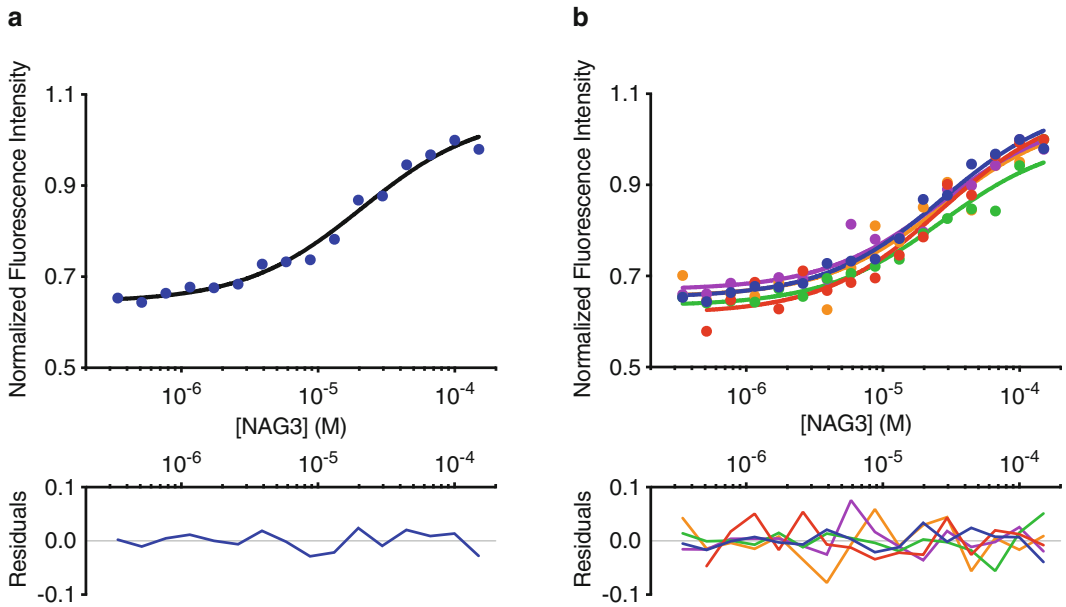


Fig. 5 (**a**) Normalized fluorescence intensity vs. NAG3 concentration, from a single titration of HEWL-D488 with increasing concentration of NAG3 in MST buffer at 25 °C. Fluorescence intensity was normalized by dividing each intensity value by the maximum intensity in the whole titration. (**b**) Five datasets, each consisting of a single titration, including that shown in (**a**). Solid lines indicate the best local (**a**) or global (**b**) fit to Eq. 1, and fitted parameters are given in the main text

intensity with increasing NAG3 concentration (this can already be seen when performing the "find capillaries" routine). Since fluorescence intensity is primarily sensitive to the local electronic environment of the dye, this indicates that NAG3 binds in the vicinity of the dye on HEWL in some or all of the HEWL-D488 population. This is illustrated in Fig. 5, using fluorescence intensity from the same titration that provided the MST data shown in Fig. 4.

In the instrument manual and other literature from the manufacturer, caution is advised when attempting to interpret MST data in the presence of a significant systematic change in fluorescence intensity. The recommendation is to fit the starting (pre-heating) fluorescence intensity data instead. Although the asymptotic fluorescence intensity at high NAG3 concentration is not well-defined by the data, it was possible to fit the data to Eq. 1 giving a best-fit value for K_d of 21 μM with a 95% confidence interval of 15–30 μM. This is not a peculiarity of one dataset but was present in all five datasets used in the global fit of MST data detailed above. The fluorescence intensity data from all five titrations could be globally fitted to a shared K_d value of 27 μM with a 95% confidence interval of 21–36 μM. This is clearly significantly higher than the value of K_d determined from the fit to the MST data shown above.

It is possible using a more complicated set of equations to account for the effect of the difference in fluorescence intensity of free and bound HEWL-D488 on the signal averaging in the thermophoresis experiments (*see* **Note 27**). These equations were used to globally fit the five MST datasets, accounting for the mean fluorescence intensity increase of 1.6-fold observed upon binding of NAG3 to HEWL-D488. This gave a shared K_d value of 10.0 μM with a 95% confidence interval of 8.9–11.2 μM.

It was not possible to use this system of equations to fit both the fluorescence intensity data and the MST data for NAG3 binding to HEWL-D488 to a common model with a shared value of K_d. The K_d measured by fluorescence intensity is apparently significantly higher than that measured by MST. It is not clear at present why this is the case. It is probable either (1) that there are additional binding events, or changes in solution properties, that take place at high NAG3 concentrations and are not related to the binding event that is monitored by MST; or (2) that there is some heterogeneity in the location of the dye within the population of HEWL-D488, giving rise to different fluorescence intensity changes and different K_d values for different subpopulations (*see* **Note 28**).

3.5 Isothermal Titration Calorimetry (ITC)

3.5.1 ITC Measurement

The aim is to perform an ITC titration in the standard configuration for small-molecule binding, titrating NAG3 (in the injector syringe) against HEWL (in the calorimeter cell). Essentially the same experiment is discussed in detail, together with much useful advice and many recommendations for best practice, in Chapter 5.

1. Consider performing a thorough cleaning of the instrument with suitable detergent solution at elevated temperature, followed by a water vs. water titration to confirm that the instrument is functional and that the level of noise arising from contamination or mechanical imperfection is low.
2. Prepare an appropriate volume of solution of 35 μM HEWL and a solution of 400 μM NAG3 (*see* **Note 29**) in standard assay buffer, in a volume suitable for the instrument that you are using (*see* **Note 30**).
3. Set the temperature of the calorimeter to 25 °C.
4. For the larger cell-volume instruments, degas both the HEWL and NAG3 solutions for 5 min under a weak vacuum. For the smaller cell-volume instrument, it is usually sufficient only to degas the cell solution. If possible, degassing should be performed slightly below the desired experimental temperature.
5. Wash the cell-loading syringe, the calorimeter cell, and the injector syringe extensively with filtered distilled water, and check that the reference cell is filled with degassed distilled water.

6. Wash the cell-loading syringe, the calorimeter cell, and the injector syringe extensively with standard assay buffer.
7. Empty the cell and the injector syringe completely by repeated aspiration to leave them as dry as practicable.
8. Fill the calorimeter cell with the HEWL solution, then the injector syringe with the NAG3 solution, according to the procedures recommended by the manufacturer (*see* **Note 31**).
9. Insert the injector into the calorimeter cell.
10. Set up the instrumental parameters (*see* **Note 32**), enter the correct concentrations for both species, and start the experiment, making sure to check that the differential power baseline settles close to the expected value, indicating that the cell is correctly filled (*see* **Note 33**).

3.5.2 ITC Data Analysis and Typical Results

The aim is to analyze the ITC titration to obtain fitted values for the dissociation constant K_d and the enthalpy of association ΔH.

1. A differential power baseline should be fitted to the raw differential power vs. time data. The curve should be integrated, and the heats should be converted into molar heats by dividing by the amount of NAG3 in each injection (*see* **Note 34**).
2. The resulting binding isotherm should be fitted by least-squares nonlinear regression to a 1:1 binding model (*see* **Note 35**).
3. Fit the K_a (transform the fitted value to K_d according to $K_a = 1/K_d$), the molar enthalpy of association, and either a stoichiometric ratio ("*n* value"), an incompetent fraction of NAG3, or a concentration correction for NAG3 (depending on the software used for the analysis).
4. The heat of dilution should be accounted for either by measuring it in an independent experiment or estimating the value from the heats in the final injections, then subtracting it from the molar heats of injection, or by allowing the asymptotic value of the molar heat at an infinite concentration of titrant to be determined in the fit.

Two example datasets are shown in Fig. 6; the data were acquired using (a) a large cell-volume calorimeter (Malvern Panalytical VP-ITC) and (b) a small cell-volume (Malvern Panalytical iTC200) calorimeter. The data were integrated and baseline subtracted using the software NITPIC [21, 22] and then individually fitted to 1:1 binding models using the software Sedphat [23].

The data in Fig. 6a gave a best-fit value for K_d of 6.6 μM with a 95% confidence interval of 6.4–6.8 μM and a best-fit value for ΔH of −13.5 kcal/mol with a 95% confidence interval of −13.6 to −13.4 kcal/mol. A concentration correction factor of 0.956 for NAG3 and the heat of dilution were both fitted to the data. The data in (b) gave a best-fit value for K_d of 7.3 μM with a 95%

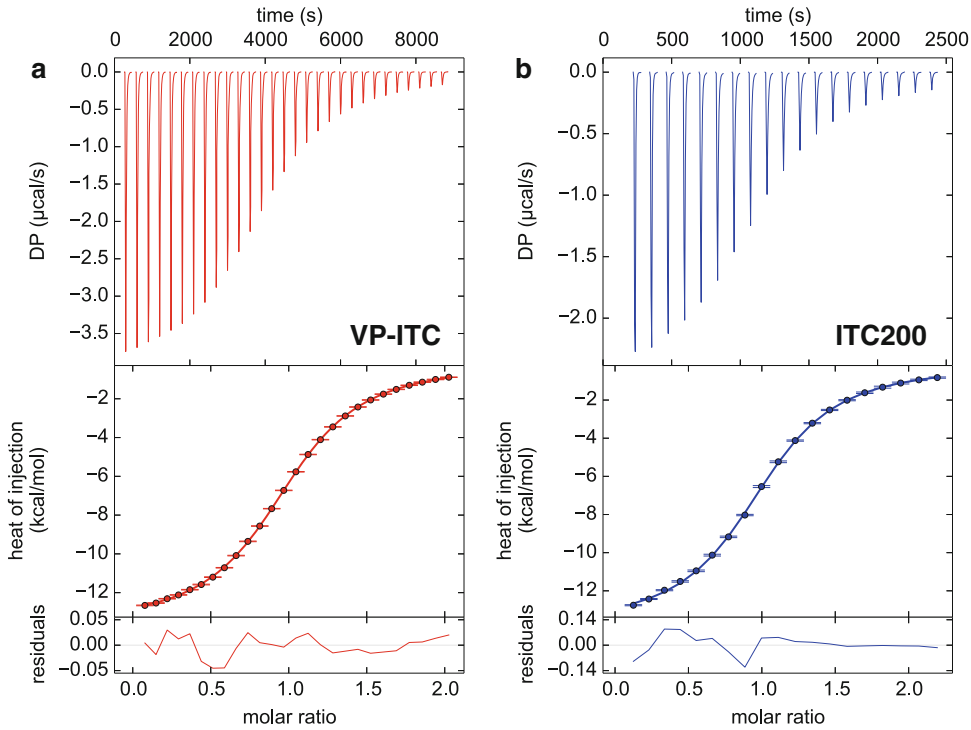


Fig. 6 ITC titrations of 1 mM NAG3 (syringe) against 96 μM HEWL in standard assay buffer at 25 °C, using a Malvern Panalytical (Microcal) VP-ITC (**a**) or iTC200 (**b**) calorimeter. The differential power data were plotted using GUSSI, after baseline subtraction in NITPIC, only showing the regions of the data that contribute to the final integrated heats of injection. Individual best fits to the integrated and normalized data were performed using a 1:1 binding model in Sedphat. Solid lines in the binding isotherms indicate the best fit to that model, and fitted parameters are given in the main text. Note that the isotherm data are plotted with the fitted concentration correction applied

confidence interval of 6.5–8.2 μM, and a best-fit value for ΔH of −13.6 kcal/mol with a 95% confidence interval of −14.1 to −13.1 kcal/mol. A concentration correction factor of 1.012 for NAG3 and the heat of dilution were both fitted to the data.

To assess the robustness of this experiment, five repeats from VP-ITC instruments and five repeats from an iTC200 instrument, each consisting of a single titration, were analyzed in a global fit. The data were integrated and baseline subtracted using NITPIC and then globally fitted to 1:1 binding models using Sedphat. The data from VP-ITCs are shown in Fig. 7a, and those from the iTC200 in Fig. 7b. The global fit of all ten datasets to shared values of K_d and ΔH, shown as solid lines in Fig. 7, gave a best-fit value for K_d of 6.5 μM with a 95% confidence interval of 6.1–7.2 μM and a best-fit value for ΔH of −13.3 kcal/mol with a 95% confidence interval of −13.7 to −13.0 kcal/mol. These values are in agreement with those previously reported in the literature for this interaction [6, 7]. Concentration correction factors in the range

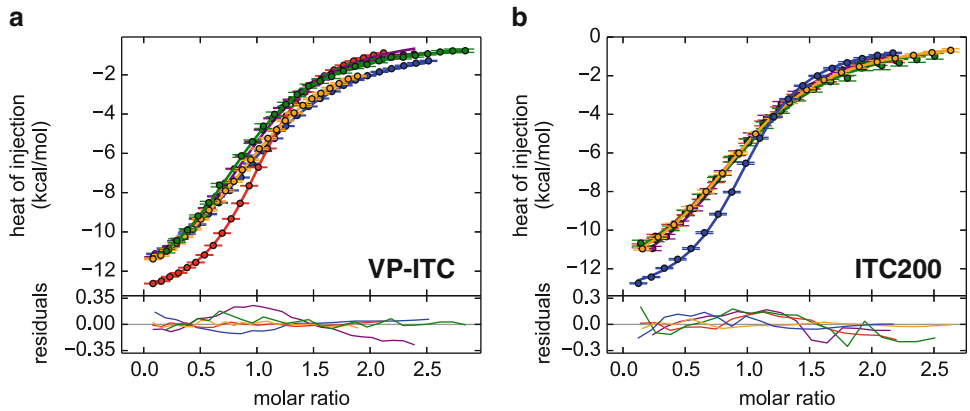


Fig. 7 (**a**) Five ITC titrations acquired with a Malvern Panalytical (Microcal) VP-ITC and (**b**) five ITC titrations acquired with an iTC200 for various concentrations of NAG3 (330–1000 μM) against various concentrations of HEWL (31–96 μM) in standard assay buffer at 25 °C. In many cases, the exact HEWL concentration was determined after the titration was set up, leading to differences in the observed molar ratios at completion. In general, for lower HEWL concentrations (lower Wiseman *c*-values), the titration was continued to higher molar ratios, better to define the heat of dilution. Solid lines in the binding isotherms indicate the best global fit of all datasets to a 1:1 binding model, and fitted parameters are given in the main text. Note that the isotherm data are plotted with the fitted concentration correction applied

0.94–1.02 for the concentration of NAG3 were fitted to each dataset individually, and the heats of dilution were fitted to each dataset individually.

3.6 SPR

3.6.1 Immobilization of HEWL

The aim is to immobilize HEWL on a carboxymethyl dextran-derivatized SPR sensor chip (Biacore CM5). Standard amine coupling chemistry is used to immobilize the protein via lysine residues. The solutions, concentrations, contact times, and flow rates specified below were used to perform the experiment using a CM5 sensor chip in a Biacore T200 and may need to be adjusted to suit other instruments. For the T200, follow the recommendations for reagent volumes and tubes suggested by the software based on the contact times and flow rates.

1. Make 500 μL of a 4-mg/mL solution of HEWL in SPR immobilization buffer. The solution should be filtered and degassed under a weak vacuum for 3–5 min before use.
2. Run the Biacore desorption protocol to ensure that the instrument is clean at the start of the measurements. Mount a new CM5 sensor chip and prime the instrument with immobilization buffer.
3. Expose two or more flow channels on the CM5 sensor chip surface to a mixture of 0.4 M *N*-ethyl-**N**′-dimethylaminopropyl carbodiimide (EDC) and 0.1 M *N*-hydroxysuccinimide (NHS) for 7 min at a flow rate of 2 μL/min. This creates an activated

ester on the chip surface, which is reactive toward lysine side-chains (and the protein N-terminal amine group).

4. One channel on the sensor chip should be inactivated for use as a reference surface to evaluate nonspecific binding. To achieve this, it can be exposed to an excess of a soluble reactive amine. Inject 1 M ethanolamine at pH 8.5 for 7 min at a flow rate of 2 μL/min.
5. Expose the other channels on the sensor chip that you intend to use for binding experiments to the 4-mg/mL HEWL solution in SPR immobilization buffer for 20 min at a flow rate of 2 μL/min.
6. Wash with immobilization buffer until the signal reaches a stable baseline.
7. After washing the sensor chip, you should determine the amount of immobilized HEWL by comparison with the activated surface before application of the HEWL solution. In five replicates, using a CM5 sensor chip in a Biacore T200, we typically obtained 4000–7000 RU of immobilized HEWL (*see* **Note 36**).
8. Inactivate any remaining unreacted activated ester groups by injecting 1 M ethanolamine at pH 8.5 for 7 min at a flow rate of 2 μL/min to the channels on the sensor chip where HEWL was immobilized. A sensorgram for the complete immobilization procedure is shown in Fig. 8.

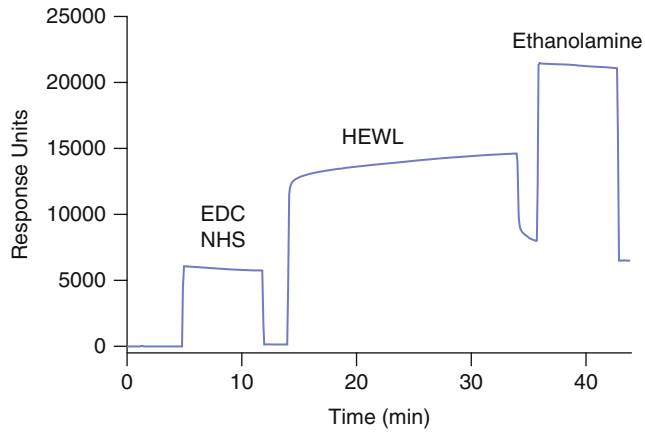


Fig. 8 SPR sensorgram for the immobilization of HEWL using EDC/NHS chemistry. The data are shown after subtraction of the initial baseline response observed when the sensor chip was exposed to immobilization buffer. The reagents applied during the three peaks in the sensorgram are indicated above the data

3.6.2 SPR Measurement

1. Wash the reference channel (FC1) and all channels containing immobilized HEWL in standard assay buffer to establish a stable baseline; a washing time of 30 min at a flow rate of 10 μL/min should be sufficient.
2. Make 400 μL of a 400-μM solution of NAG3 in standard assay buffer, and then make a 12-point 1:1 serial dilution in the range 0.2–400 μM with a final volume of 200 μL for each point. Degas the solutions for 3–5 min under a weak vacuum before use.
3. Perform a series of injections of the serial dilution of NAG3, from the lowest to the highest concentration. Injections should be made to the reference channel (FC1) and to all channels with immobilized HEWL.
4. The contact time for each injection should be 10 min at a flow rate of 10 μL/min (total injection volume of 100 μL). This should be sufficient to establish a stable equilibrium (steady-state) response at that concentration. After each injection, flow standard assay buffer across all the relevant flow channels for 20 min at 10 μL/min so that any bound NAG3 is completely dissociated and the response returns to the pre-injection baseline.

3.6.3 SPR Data Analysis and Typical Results

1. Subtract the signal from the reference channel of the sensor chip from the signal in the channels that have immobilized HEWL to correct for any response arising from changes in solution bulk refractive index or from nonspecific binding of NAG3 to the inactivated sensor chip surface. Doing this for the sensorgram from each injection will yield the sensor response for binding at that concentration of NAG3, as shown in Fig. 9, from which a response isotherm can be derived.

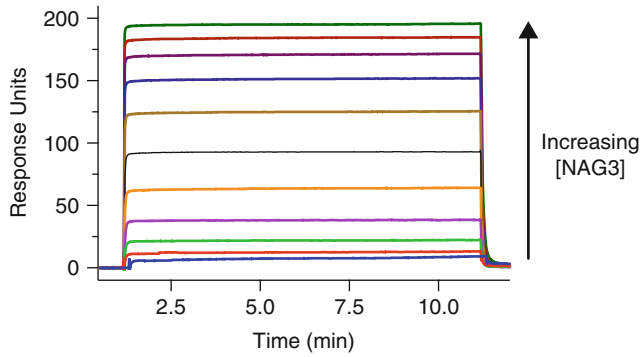


Fig. 9 SPR sensorgrams for sequential injections of increasing concentrations of NAG3. These data correspond to the individual dataset analyzed in Fig. 10a. The data are shown after subtraction of the response of the reference channel to the same injection and subtraction of the initial baseline response at the start of each injection

2. Fit the response isotherm to a 1:1 binding model in the BiaEvaluation software or to Eq. 2 in other software capable of nonlinear regression.

$$R_{eq} = \frac{R_{max}[L]}{K_d + [L]} + R_0 \quad (2)$$

where R_{eq} is the response from the sensor chip (the y-axis variable), $[L]$ is the concentration of NAG3 in flow at a particular injection, R_0 is the response baseline before an injection, R_{max} is the asymptotic response at infinite NAG3 concentration, and K_d is the dissociation constant for the interaction.

An example of a single dataset together with a fit to Eq. 2 is shown in Fig. 10a. The solid black line is the best fit to that single dataset using Eq. 2. The best-fit value of K_d is 12.2 μM, with a 95% confidence interval of 11.2–13.2 μM.

To assess the robustness of this experiment, five repeats, each consisting of a single titration, were analyzed in a global fit. The datasets were fitted to individual values of R_0 and R_{max} (the baseline response, and the response at saturation of the surface) and a shared value of K_d. The datasets and the lines of best fit for the global fit to Eq. 2 are shown in Fig. 10b. The globally fitted K_d was 11.6 μM with a 95% confidence interval of 10.9–12.3 μM.

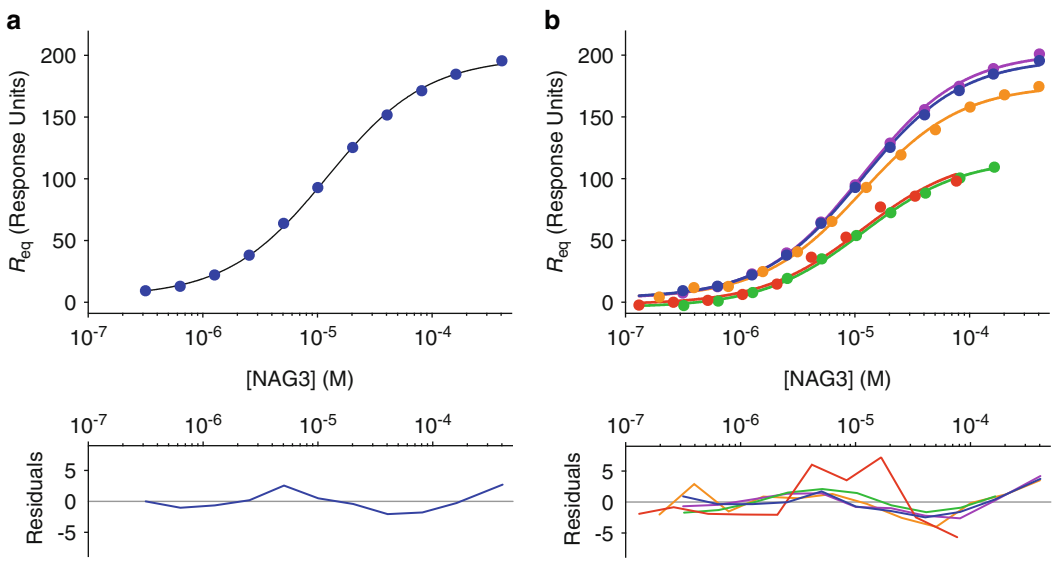


Fig. 10 (**a**) SPR response vs. NAG3 concentration, from a single titration of immobilized HEWL with increasing concentrations of NAG3 in standard assay buffer at 25 °C. (**b**) Five datasets, each consisting of a single titration. Solid lines indicate the best local (**a**) or global (**b**) fit to Eq. 2, and fitted parameters are given in the main text

4 Notes

1. The conditions for the ITC experiment were initially found in lecture notes from Professor Alan Cooper and were used as a starting point for all other experiments. Calorimetric studies of the interaction have also been reported in the literature under these conditions [7] and under slightly different conditions [6].
2. It is important that only the binding partner present at a constant concentration in a titration exhibits fluorescence in the wavelength range used. Depending on the type of instrument used and the experimental context, either intrinsic protein fluorescence or fluorescence from an extrinsic dye can be used. The choice to use an extrinsic dye in the case presented in this chapter was determined solely by the capabilites of the available instrument.
3. The data were collected under supervision from expert users, and the graduate students who participated in the course were also engaged in collection of data for publication. Therfore, it might be argued that the data offer a realistic representation of nonexpert use after training.
4. pH 5 acetate/acetic acid is used for three reasons: firstly, it is the pH at which the affinity of the interaction is reported to be highest; secondly, it is below the pH at which significant populations of dimer form for lysozyme [24]; and thirdly, the enthalpy of ionization of acetate/acetic acid buffers is essentially zero, meaning that they do not change pH significantly with temperature (useful for the thermal shift assay) and do not contribute to observed enthalpies of binding if proton-transfer occurs. All the buffers should ideally be prepared by weighing appropriate amounts of acidic and basic buffer species and dissolving them in an appropriate volume of pure distilled water, calculated taking into account the intended working temperature. An excellent resource for such calculations is www.liverpool.ac.uk/pfg/Research/Tools/BuffferCalc/Buffer.html. For example, to make 1 L of the standard assay buffer for use at 25 °C, 0.0314 mol acetic acid and 0.0685 mol sodium acetate should be dissolved in a total 1 L solution in a volumetric flask. It is also possible to prepare the buffer by carefully titrating a solution of 0.1 M sodium acetate to pH 5.0 using concentrated HCl, monitored by a calibrated pH meter. The buffer should be filtered through a 0.22-μm filter to sterilize it and remove particulate matter before use in experiments.
5. The MST experiment uses a low nanomolar concentration of fluorescently labeled lysozyme solution. To suppress binding of protein to plastic and glass surfaces during sample preparation

and measurement, 0.1% (v/v) Tween 20 is added to the buffer used for both serial dilution of NAG3 and dilution of HEWL from the stock solution.

6. In case it proves important for reproducibility, it should be noted that all of the data presented here were acquired using reagents from Sigma Aldrich (HEWL: L4919, NAG3: T2144). The IUPAC name for NAG3 is *N′N″N‴*-triacetyl chitotriose, but it is also known by several other names listed at Pubchem under the compound ID 123774.
7. When ordering multiple vials of SYPRO Orange, consistency of fluorescence intensity over time can be improved by pooling and mixing the aliquots and storing the result in a black Eppendorf tube at −20 °C.
8. We use a single-beam instrument and a 1-cm pathlength cuvette with a fill volume of 200 μL, which gives acceptable reproducibility with relatively little sample consumption.
9. This simple and readily employed correction [25] is a linearization of the inverse fourth power dependence of scattering upon wavelength, appropriate for any protein without cofactor absorbance at 333 nm. If the solutions are prepared correctly, the scattering correction should be very small (<5% of the total absorbance at 280 nm).
10. This is essentially identical to the experimentally determined value (37,789 M^{-1} cm^{-1}) [25].
11. If weighing proves unreliable for the determination of the NAG3 concentration, another (albeit less satisfactory approach) would be to use several replicate ITC experiments on a NAG3 stock to benchmark the concentration on the assumption that the lysozyme concentration can be accurately and reproducibly determined by spectrophotometry. It would also be possible to determine the concentration from peak integrals in a 1D NMR experiment, though we have not attempted to do so.
12. This experiment is designed to show how the thermal shift assay can be used as an initial high-throughput screen for binding. In those circumstances, it is usual to use ligand concentrations that are sufficiently high to ensure saturation of binding for a wide range of dissociation constants, but not so high as to cause shifts in pH of the experimental buffer.
13. Since the fluorescence intensity from SYPRO Orange usually increases in the presence of unfolded protein, it is not possible to optimize the detector gain setting from the initial fluorescence. Instead, an appropriate value must be determined empirically for each instrument (and sometimes for a new batch of dye) in a test experiment.

14. In normal use, the ramp rate for thermal denaturation should be chosen carefully as a trade-off between equilibration time at a given temperature and minimization of perturbation by aggregation. In this case, a relatively rapid ramp rate ensures a short experiment time, showing the utility of the method as a rapid initial screen or check for binding. Ramp rates of this order have been used successfully for high-throughput ligand screening [26].
15. The software sometimes automatically inverts the sign of the first derivative because it is primarily designed for experiments with decreasing signal as a function of temperature. Usually, there is a selectable option to negate this. If the software does not calculate the first derivative for you, most common graphing software can calculate it automatically. Depending on the degree of noise in the data, it is sometimes necessary to smooth the data before taking the first derivative.
16. More recently, we have obtained very similar results for thermal shift experiments monitored by DSF of the instrinsic fluorescence of HEWL, using a Nanotemper Prometheus instrument. Using the same concentrations of HEWL and NAG3 but omitting the SYPRO Orange dye, and running a continuous thermal ramp from 20 to 95 °C at a rate of 1 °C/min in standard capillaries, we obtained a T_m of 76.2 °C for HEWL, with a ΔT_m of 3 °C upon addition of 1 mM NAG3.
17. You should check that this working volume is suitable for the type of microplate that you are using in the assay.
18. Some experimentation may be required on the first attempt with a new plate reader to identify optimal settings for the best signal-to-noise. At the concentrations employed, the signal for all wells containing HEWL and NAG3 should be greatly in excess of that of the buffer with the same measurement settings. Excitation at longer wavelength (e.g., 295 nm) gives in principle selectivity for tryptophan over tyrosine residues due to the broader range of excitation bands for tryptophan (advantageous in this case, since the binding site of HEWL contains two tryptophans that contact NAG3). In practice, when measured with a typical plate reader, this selectivity is lost due to the wide bandwidth of excitation. Similarly, wide bandwidth in the emission optics means that the fluorescence intensity is effectively integrated over a broad range of wavelengths, and the exact magnitude of the signal change will depend upon the precise properties of the plate reader optics. One of the reasons for using a plate reader in this experiment is to demonstrate that modern plate readers can give data of sufficient quality for measurements of intrinsic protein fluorescence, with an acceptable trade-off of speed, sample economy, and ease of use vs. selectivity and sensitivity.

19. This fitting equation calculates the observed signal at a particular point in the titration from an observed experimental signal for each protein species (free HEWL and bound HEWL) and the mole fraction for that species. The observed experimental signal for both species will be dependent on the total protein concentration and the type and configuration of the plate reader used. In the following figures, it was necessary to perform a simple normalization on fluorescence intensity data to allow datasets from different instruments to be plotted together. However, it is usually preferable to analyze directly the non-normalized data in the original measurement units; for example, this allows comparison of consistency of experimental parameters between repeated runs on the same instrument with the same settings.
20. All confidence intervals on K_d presented in this chapter are calculated by the method of error surface projection. For all techniques except ITC, this was performed using Graphpad Prism v7 ("asymmetric likelihood CI") (https://www.graphpad.com/guides/prism/7/curve-fitting/index.htm?reg_confidence_tab.htm). For ITC, this was performed in Sedphat.
21. The aim is to obtain a labeling stoichiometry of approximately 1 dye per HEWL molecule. To achieve this, a slight molar excess (1.4-fold) of the NHS-activated dye over HEWL is used in the labeling reaction. The pH and composition of the labeling buffer were among those recommended by the dye manufacturer. At pH 7.5, it is expected that both lysine side-chain amino groups and the N-terminus could be labeled, raising the pH to around 8.5 would increase reactivity toward the former, and lowering the pH to around 6.5 would selectively target the latter.
22. If it is much higher than 1:1, it is possible that the dialysis has not proceeded to completion and there is considerable free dye left in solution. This will significantly lower the signal-to-noise in the final measurement.
23. This gives a final concentration of 150 nM HEWL-D488, sufficient to give excellent signal-to-noise at moderate laser intensities on the older instrument used for these measurements. Since the K_d is in the low μM range, it is possible to use a relatively high concentration or protein without limiting the ability of the fit to discriminate K_d. On newer instruments, it should be possible to reduce this concentration significantly to save material, if desired. For other interactions with K_d in the nM range, it is wise to use the lowest possible concentration of the constant species that affords acceptable signal-to-noise.

24. The cheaper "standard" MST capillaries give data with little scatter and minimal adhesion in the presence of Tween 20; in the absence of Tween 20, other capillaries should be used.
25. This experiment was developed on a rather old Nanotemper Monolith instrument, lacking the latest software due to incompatibility. Therefore, it might be necessary to make some small alterations to the protocol to suit newer instruments and software.
26. Here F_{norm} was calculated from the mean fluorescence signals at 2.5 and 30 s to give the best signal-to-noise. As noted in the excellent paper by Brautigam and colleagues [27], the observed K_d may vary as a function of heating time due to kinetic relaxation of the system at the higher temperature achieved through heating. We observed no significant difference in fitted K_d for any of the datasets when F_{norm} was calculated for "thermophoresis plus jump" phases with the endpoint at 30, 20, or 10 s (a global fit of data from the three heating times to a shared K_d was statistically indistinguishable from fits to local K_d values at the 95% confidence level according to an *F*-test performed in Graphpad Prism 7).
27. Thermophoresis is monitored via changes in fluorescence intensity in a detection volume due to depletion or enrichment of macromolecules during heating. For a mixture of free and bound macromolecule with equal fluorescence intensity, the thermophoresis signal is an average of the F_{norm} signals of the free and bound species, weighted by their mole fraction (Eq. 1). When the fluorescence intensity of the two species are different, the thermophoresis signal is an average of the F_{norm} signals of the free and bound species, weighted by their mole fraction of the total fluorescence intensity; i.e., even for equal populations of free and NAG3-bound HEWL-D488 (equal mole fractions), the NAG3-bound form would contribute significantly more to the observed thermophoresis signal since it has a significantly higher fluorescence intensity (a factor of approximately 1.6). This can be accounted for using the following system of equations:

$$[\mathrm{PL}] = \frac{([P]_t + [L]_t + K_d) - \sqrt{([P]_t + [L]_t + K_d)^2 - 4[P]_t[L]_t}}{2}$$

$$[P] = [P]_t - [\mathrm{PL}]$$

$$S_{obs} = \frac{[P] \cdot \varepsilon_P \cdot \alpha_P + [\mathrm{PL}] \cdot \varepsilon_{PL} \cdot \alpha_{PL}}{[P] \cdot \varepsilon_P + [\mathrm{PL}] \cdot \varepsilon_{PL}}$$

where $[P]_t$ is the total concentration of HEWL (fixed in the fit), $[L]_t$ is the total concentration of NAG3 at a given titration point, K_d is the dissociation constant for the interaction, α_P

and α_{PL} are the F_{norm} signals for free and bound HEWL-D488, ε_P and ε_{PL} are the relative or absolute fluorescence intensities of free and bound HEWL-D488 (fixed in the fit to values obtained from the fitting of the fluorescence intensity data), and S_{obs} is the observed F_{norm} signal. It is important to note that this is not a situation unique to MST. Essentially the same set of equations can be used to analyze any experimental variable that is measured indirectly using another spectroscopic means of detection and so does not necessarily report directly on the population of states (e.g., fluorescence anisotropy, fluorescence emission maximum, sedimentation coefficient). The equations presented here are similar to those presented elsewhere [28, 29].

28. Prompted by comments from Dr. Christopher Johnson during the editing of this chapter, based on his own experimental observations while reproducing these experiments, we performed a titration to higher NAG3 concentration (700 μM). In this single experiment, we did observe additional MST and fluorescence signal changes above 150 μM NAG3. The MST signal change is anticorrelated with that observed at lower concentration, whereas the fluorescence signal change continues in the same direction. These additional signal changes are only clearly apparent above the highest concentration used in the MST titrations presented here (chosen based on the dissociation constant). The cause is not clear: data from the other techniques do not show deviation from single-site binding, and a second binding site is not predicted from the structure of HEWL bound to NAG3. Additional titration phases of uncertain origin have been observed in other MST datasets and can complicate analysis [27]. It might be that a different labeling strategy or dye mitigated these effects.
29. These are minimal concentrations designed for economy, while giving acceptable curvature in the binding isotherm for robust data analysis. We have also used 100 μM HEWL and 1000 μM NAG3 to achieve a steeper curvature and better constrained estimates for the enthalpy of binding.
30. For an instrument with a working cell volume in the region of 1.4 mL, make 2 mL of the HEWL solution and 500 μL of the NAG3 solution. For an instrument with a working cell volume in the region of 200 μL, make 300 μL of the HEWL solution and 100 μL of the NAG3 solution. These volumes are sufficient to fill the cell and syringe of Malvern Panalytical (Microcal) calorimeters, with some margin of safety. As you become more proficient in the loading process, you may be able to reduce these volumes. If suitably low-volume cuvettes are available, it is also possible to use the excess HEWL solution recovered at

the end of the filling process to check the concentration once again, to account for any dilution during the degassing and cell-loading processes.

31. In instruments where the injection is made by a screw-driven Hamilton syringe, the volume of the first one or two injections can be reduced due to slack in the connection between the driving screw and the Teflon plunger [30]. This potential artifact can be avoided by advancing the plunger of the injector by a small fraction (a few percent) of its total travel before inserting the injector into the calorimeter cell.
32. For an instrument with a 1.4-mL cell volume and a 300-μL injector syringe volume, a typical injection protocol would be 1×1 μL injection followed by 29×10 μL injections. For an instrument with a 200-μL cell volume and a 40-μL injector syringe volume, a typical injection protocol would be 1×0.4 μL injection followed by 19×2 μL injections. The reference power value should be set in the middle of the dynamic range of the instrument. The time between injections should be such that the differential power returns to and remains in the baseline for some time after the injection. For an instrument with a 1.4-mL cell volume, a typical delay between injections would be 300 s; for an instrument with a 200-μL cell volume, typical delay would be 120 s. The averaging time for a single point should be set short enough to ensure that there is sufficient data density properly to sample the shape of the injection peaks, but not so short as to give excessive noise. The reference power value should be set in the middle of the dynamic range of the instrument, and the feedback in the power compensation circuit should be set to give the fastest response possible.
33. The differential power signal will deviate significantly from the set value if the heat capacity of either cell is very different from that of aqueous solution due to the presence of trapped air. A misfilled sample cell containing air bubbles will cause the differential power to settle at a value lower than the set value. This is explained in more detail in Chapter 5.
34. The signal-to-noise of the titration under these conditions is such that the baseline determination and integration should be robust and reliable with little manual intervention. We have obtained essentially identical results using Microcal Origin for ITC with small manual adjustments of the baseline, or using the automated procedures in Affinimeter and NITPIC. In general, we would prefer the automated procedures because they are less prone to user bias and provide estimates of the error on integrations that are important for determining the true error on fitted parameters.

35. We have obtained very similar results using Microcal Origin for ITC, Affinimeter, and Sedphat. Affinimeter or Sedphat offers a variety of additional possibilities for global data analysis and modeling. For this chapter, we have used Sedphat in order to perform error surface projection calculations for direct comparison with the fits for other methods.
36. The SPR experiments were more recently developed and therefore have been subject to relatively little repetition compared to the other experiments. The protocol has been reproduced in another lab with essentially identical results, but we have observed considerable variability in the number of response units of HEWL immobilized (and hence in signal change in the final titration). We suspect that this is correlated with the age and storage of the coupling reagents and would therefore recommend starting with fresh reagents or reagents that have been stored in aliquots at −80 °C.

References

1. Jollès P, Jollès J (1984) What's new in lysozyme research. Mol Cell Biochem 63:165–189
2. Blake CCF, Koenig DF, Mair GA, North ACT, Phillips DC, Sarma VR (1965) Structure of hen egg-white lysozyme: a three-dimensional Fourier synthesis at 2 Å resolution. Nature 206:757
3. Cheetham JC, Artymiuk PJ, Phillips DC (1992) Refinement of an enzyme complex with inhibitor bound at partial occupancy: hen egg-white lysozyme and tri-N-acetylchitotriose at 1.75 Å resolution. J Mol Biol 224:613–628
4. Moorman VR, Valentine KG, Wand AJ (2012) The dynamical response of hen egg white lysozyme to the binding of a carbohydrate ligand. Protein Sci 21:1066–1073
5. Lumb KJ, Cheetham JC, Dobson CM (1994) 1H nuclear magnetic resonance studies of hen lysozyme-n-acetylglucosamine oligosaccharide complexes in solution: application of chemical shifts for the comparison of conformational changes in solution and in the crystal. J Mol Biol 235:1072–1087
6. García-Hernández E, Zubillaga RA, Chavelas-Adame EA, Vázquez-Contreras E, Rojo-Domínguez A, Costas M (2003) Structural energetics of protein-carbohydrate interactions: insights derived from the study of lysozyme binding to its natural saccharide inhibitors. Protein Sci 12:135–142
7. Bjurulf C, Wadsö I (1972) Thermochemistry of lysozyme-inhibitor binding. Eur J Biochem 31:95–102
8. Banerjee SK, Rupley JA (1973) Temperature and pH dependence of the binding of oligosaccharides to lysozyme. J Biol Chem 248:2117–2124
9. Brandts JF, Lin LN (1990) Study of strong to ultratight protein interactions using differential scanning calorimetry. Biochemistry 29:6927–6940
10. Niesen FH, Berglund H, Vedadi M (2007) The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. Nat Protoc 2:2212–2221
11. Holdgate GA, Ward WHJ (2005) Measurements of binding thermodynamics in drug discovery. Drug Discov Today 10:1543–1550
12. Duhr S, Braun D (2006) Why molecules move along a temperature gradient. Proc Natl Acad Sci 103:19678–19682
13. Seidel SAI, Dijkman PM, Lea WA, van den Bogaart G, Jerabek-Willemsen M, Lazic A, Joseph JS, Srinivasan P, Baaske P, Simeonov A, Katritch I, Melo FA, Ladbury JE, Schreiber G, Watts A, Braun D, Duhr S (2013) Microscale thermophoresis quantifies biomolecular interactions under previously challenging conditions. Methods 59:301–315
14. Wiseman T, Williston S, Brandts JF, Lin LN (1989) Rapid measurement of binding constants and heats of binding using a new titration calorimeter. Anal Biochem 179:131–137
15. Freire E, Mayorga OL, Straume M (1990) Isothermal titration calorimetry. Anal Chem 62:950A–959A

16. Ladbury JE, Klebe G, Freire E (2010) Adding calorimetric data to decision making in lead discovery: a hot tip. Nat Rev Drug Discov 9:23–27
17. Olsson TSG, Williams MA, Pitt WR, Ladbury JE (2008) The thermodynamics of protein–ligand interaction and solvation: insights for ligand design. J Mol Biol 384:1002–1017
18. Karlsson R, Michaelsson A, Mattsson L (1991) Kinetic analysis of monoclonal antibody-antigen interactions with a new biosensor based analytical system. J Immunol Methods 145:229–240
19. Schasfoort RBM (2017) Handbook of surface plasmon resonance. The Royal Society of Chemistry, London
20. Kemmer G, Keller S (2010) Nonlinear least-squares data fitting in Excel spreadsheets. Nat Protoc 5:267–281
21. Keller S, Vargas C, Zhao H, Piszczek G, Brautigam CA, Schuck P (2012) High-precision isothermal titration calorimetry with automated peak-shape analysis. Anal Chem 84:5066–5073
22. Scheuermann TH, Brautigam CA (2015) High-precision, automated integration of multiple isothermal titration calorimetric thermograms: new features of NITPIC. Methods 76:87–98
23. Zhao H, Piszczek G, Schuck P (2015) SEDPHAT--a platform for global ITC analysis and global multi-method analysis of molecular interactions. Methods 76:137–148
24. Sophianopoulos AJ, Van Holde KE (1964) Physical studies of muramidase (lysozyme) II. pH-dependent dimerization. J Biol Chem 239:2516–2524
25. Pace CN, Vajdos F, Fee L, Grimsley G, Gray T (1995) How to measure and predict the molar absorption coefficient of a protein. Protein Sci 4:2411–2423
26. Fedorov O, Niesen FH, Knapp S (2012) Kinase inhibitor selectivity profiling using differential scanning fluorimetry. In: Kuster B (ed) Kinase inhibitors: methods and protocols. Humana Press, Totowa, NJ
27. Scheuermann TH, Padrick SB, Gardner KH, Brautigam CA (2016) On the acquisition and analysis of microscale thermophoresis data. Anal Biochem 496:79–93
28. Brown PH, Balbo A, Schuck P (2008) Characterizing protein-protein interactions by sedimentation velocity analytical ultracentrifugation. Curr Protoc Immunol 81:18.15.1–18.15.39
29. Pope AJ, Haupts UM, Moore KJ (1999) Homogeneous fluorescence readouts for miniaturized high-throughput screening: theory and practice. Drug Discov Today 4:350–362
30. Mizoue LS, Tellinghuisen J (2004) The role of backlash in the "first injection anomaly" in isothermal titration calorimetry. Anal Biochem 326:125–127

Part II

Universal Methods for Protein Interactions

Chapter 3

Interactions of a Signal Transduction Protein Investigated by Fluorescence Stopped-Flow Kinetics

Stephen R. Martin and Maria J. Schilstra

Abstract

To understand cellular processes such as biochemical pathways and signaling networks, we need to understand binding and reaction rates of often competing reactions, their dependence on cellular concentrations of participating molecules, and the regulation of these rates through allostery, posttranslational modifications, or other mechanisms. To do so, we break these systems down into their elementary steps, which are almost invariably either unimolecular or bimolecular reactions that frequently occur on sub-second, often sub-millisecond, time scales. Rapid mixing techniques, which generally achieve mixing in less than 2 ms, are generally suitable for the study of such reactions. The application of these techniques to the study of enzyme mechanisms is described in several excellent texts (Cornish-Bowden, Fundamentals of enzyme kinetics, 1995; Gutfreund, Kinetics for the life sciences. Receptors, transmitters and catalysis, 1995); flow techniques are used to study individual steps by monitoring the approach to equilibrium (the pre-steady state) under single turnover conditions.

The individual steps in complex biochemical reaction schemes determine how fast systems can respond to incoming signals and adapt to changed conditions [1, 2]. This chapter is concerned with in vitro techniques that have been developed to study fast reactions in solution, and we present the study of various interactions of calmodulin as an example. The kinetic information obtained with these techniques is indispensable for understanding the dynamics of biochemical processes and complements the static structural and thermodynamic information available from X-ray crystallography, NMR, and equilibrium binding studies.

Key words Calmodulin, Kinetics, Rate constants, Fluorescence, Stopped-flow, Data analysis and modeling

1 Introduction

1.1 The Biological System Under Study

Changes in calcium concentration act as a ubiquitous intracellular signal responsible for controlling many biological processes such as contraction, secretion, fertilization, and cell proliferation [3]. In all eukaryotic cells, one of the important proteins that mediates Ca^{2+} signaling is calmodulin (CaM). Upon Ca^{2+} stimulation, CaM binds to and modulates the activity of a diverse number of enzymes, including a family of CaM-dependent serine/threonine protein kinases [4]. CaM is a 148-amino acid protein whose crystal

Tina Daviter et al. (eds.), *Protein-Ligand Interactions: Methods and Applications*, Methods in Molecular Biology, vol. 2263, https://doi.org/10.1007/978-1-0716-1197-5_3,

structure shows an extended dumbbell shape, with the N- and C-terminal domains connected by a solvent-exposed α-helix [5] that is flexible in solution [6]. This flexibility is crucial for the binding and activation of target enzymes.

The binding of Ca^{2+} to CaM induces conformational changes in its N- and C-terminal domains, resulting in the exposure of hydrophobic pockets that are important for target protein binding [7]. The structural basis of the interaction of CaM with target proteins has largely been investigated using short synthetic peptides corresponding to the binding domains of the intact target proteins. This chapter describes methods for determining kinetic constants for the binding of these synthetic peptides and of the intact proteins.

1.2 The Choice of Techniques

There are various ways to monitor changes in concentration of reactants, intermediates, and products after mixing, but the most common way is to use changes in optical signals (generally either absorbance or fluorescence) which often accompany reactions. Although absorbance can sometimes be used, fluorescence is often preferred because of its greater sensitivity, particularly in monitoring conformational changes. Such methods are continuous with good time resolution but they seldom permit the direct determination of the concentrations of individual species. Alternatively, samples may be taken from the reaction volume, mixed with a chemical quenching agent to stop the reaction, and their contents assessed by techniques such as HPLC. These methods can directly determine the concentrations of different species, but are discontinuous and have a limited time resolution.

In all rapid mixing, or "flow" techniques, the reactant solutions are driven at high velocity into a special mixing chamber. The mixing and subsequent passage to the point of observation take a finite amount of time, so that the mixed solution already has a certain "age" before it can be observed. The interval between the start of the mixing and the earliest possible observation time is called the instrument's *dead time*.

In **stopped-flow**, the commonest flow technique, the mixed solution rapidly flows into an observation chamber, where it is stopped and monitored by recording the change in some suitable optical signal as a function of time (*see* **Notes 1** and **2**). The dead time of a stopped-flow instrument is typically 1–2 ms and reactions occurring on a faster time scale cannot be studied.

In **quenched-flow**, the reactants are mixed and flow down an "aging tube" at constant velocity before mixing with a "quenching agent," generally acid, that stops the reaction. The quenched reaction mixture is then analyzed using an appropriate method, such as HPLC. Because the age of the quenched sample is determined by the flow rate and the flow tube volume, a series of time points is built up by doing experiments with different flow rates and/or tube

volumes. Time points between ~5 ms and ~150 ms can usually be obtained in this way. Quenched-flow methods have the advantage that they can be used when no optical signal is available [8], but are much more labor intensive than stopped-flow methods.

In **continuous flow**, the reactants are mixed and an optical signal is monitored at different positions downstream from the mixer and converted into a time-dependent signal change on the basis of the known flow rate. Continuous flow has the potential to measure reactions on a much faster time scale than stopped-flow [9], but stopped-flow is generally preferred because of its better sample economy and its ability to measure the kinetics out to longer times.

In this study, we chose fluorescence stopped-flow to investigate the interactions in question, as it was the most convenient, covered the relevant time scales and suitable fluorescent probes were available.

1.3 Reaction Kinetics and Thermodynamics

Throughout this chapter, we use the symbols P and L to indicate "protein" and "ligand"—as many intracellular interactions are between proteins and small molecules—but these may be any two reactants, proteins, nucleic acids, lipids, biomolecular assemblies, etc. PL is used to indicate a complex between P and L, and P*, L*, and PL* indicate different conformational states of these species.

The simplest reversible reaction is one where both the forward and reverse steps are unimolecular processes with first-order rate constants k_1 and k_{-1} (units: s^{-1}).

$$P \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} P^* \qquad \text{Scheme A}$$

The dimensionless equilibrium constant, K, for this reaction is defined as $K = k_1/k_{-1}$. If the system is subjected to a change which alters the equilibrium constant, the concentrations of P and P* will change until the new equilibrium position is established. If the reaction is accompanied by a change in an optical signal, S, this will change with time following a single exponential according to:

$$S(t) = S_{eq} - \left(S_{eq} - S_0\right) \exp\left(-k_{obs} t\right) \tag{1}$$

where S_{eq} and S_0 are the signals at equilibrium and time zero, $(S_{eq} - S_0)$ is the total signal change (amplitude) of the reaction, and k_{obs} is the observed rate for the reaction (*see* **Note 3** and Fig. 1). In the case of fluorescence measurements, Eq. 1 can be written as

$$F(t) = \Delta F \exp\left(-k_{obs} t\right) + F_\infty \tag{2}$$

where ΔF is the amplitude and F_∞ is the final florescence value.

k_{obs} for **Scheme A** is equal to $(k_1 + k_{-1})$ and as such is independent of concentration. Such reactions can be studied with rapid

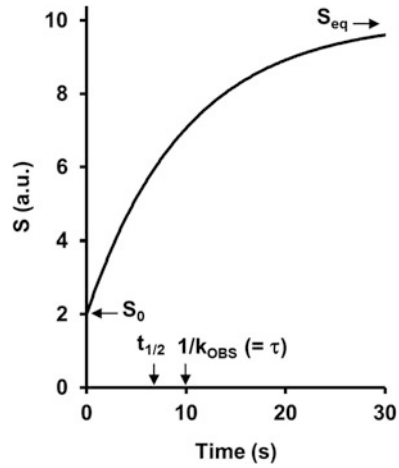


Fig. 1 A single exponential time course. A single exponential generated with Eq. 1 using $k_{obs} = 0.1\ s^{-1}$, $S_0 = 2$ and $S_{eq} = 10$ (see text and **Note 3**)

mixing devices using the solvent jump method in which the composition of the solvent is abruptly changed by mixing two solutions at equilibrium in different solvents, or a solution at equilibrium with another solvent. Analysis of kinetic traces using Eq. 1 does not give the individual rate constants but if the equilibrium constant K is known then they can be calculated using:

$$k_{-1} = \frac{k_{obs}}{1 + K} \qquad k_1 = \frac{K \cdot k_{obs}}{1 + K} \tag{3}$$

Reversible binding reactions, such as those in which a ligand associates with a protein, have a second-order association process, and a first-order dissociation process and are described by:

$$P + L \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} PL \qquad \text{Scheme B}$$

Here k_1 is the second-order association rate constant (units: $M^{-1}\ s^{-1}$), and k_{-1} is the first-order dissociation rate constant (units: s^{-1}). The equilibrium *dissociation* constant for this reaction, K_d, is equal to k_{-1}/k_1 (units: M), whereas the equilibrium *association* constant, K_a, is its reciprocal k_1/k_{-1} (units: M^{-1}). There is no simple general analytical solution for the differential rate equation (*see* Subheading 3.5) that describes the change in [PL] with time. However, if one of the reactants is in large excess over the other ($[L_{tot}] \gg [P_{tot}]$ or $[P_{tot}] \gg [L_{tot}]$), the concentration of the component in large excess remains effectively constant during the reaction because $[X_{tot}] - [PL] \approx [X_{tot}]$, where $[X_{tot}]$ is the total concentration of the component (P or L) present in excess. The formation of PL is then said to follow pseudo-first-order kinetics,

and an optical signal will change with time according to Eq. 1 but with the observed rate k_{obs} now given by:

$$k_{obs} = k_1[X_{tot}] + k_{-1} \quad (4)$$

Individual rate constants can then be extracted from the dependence of k_{obs} on $[X_{tot}]$ (*see* Subheading 3.1). More complex schemes will often show more than a single kinetic phase. Nevertheless, under the appropriate conditions, the observed time course will be the sum of two, or more, exponentials and analysis requires an appropriately extended version of Eq. 1 with two, or more, k_{obs} values (*see* Subheading 3.4). The approach, however, remains the same; the experimental transients are analyzed to give k_{obs} values and the rate constants are determined from the dependence of these k_{obs} values on concentration(s).

2 Materials

2.1 Instrumentation

Instrumentation for performing rapid kinetic measurements is available from several suppliers: TgK Scientific Ltd. (Supplier of HiTech Instruments: http://www.hi-techsci.com/); The KinTek Corporation (http://www.kintekcorp.com/); OLIS, Inc. (http://olisweb.com/); Applied Photophysics (http://www.photophysics.com/); and Biologic Science Instruments (http://www.bio-logic.net/). The principal detection methods employed are fluorescence and absorbance. Fluorescence detection is widely employed because it is more sensitive than absorption and therefore allows measurements to be made at lower concentrations. Circular dichroism (CD) detection is widely employed in studies of protein unfolding, but the inherently poor signal-to-noise ratios of CD signals limit its use in the study of protein–ligand interactions. Small hand-driven devices that can be used in conjunction with regular spectrophotometers are relatively inexpensive and permit the study of reactions with half times as short as 10 ms (depending on the response time of the spectrometer). Most stopped-flow instruments are designed to mix equal volumes of the two reactants, but some will allow different volumes to be used. This technique is most widely used in studies of protein folding using chemical denaturants, where large and rapid changes in denaturant concentration are required.

2.2 Instrument Settings

As with any scientific instrument, the user must understand the characteristics and limitations of the equipment being used [10]. In the case of stopped flow, it is useful, and instructive, to demonstrate that mixing is efficient and to determine the dead time of the instrument. Detailed methods for doing this and for performing temporal calibration of a quenched-flow instrument have been given elsewhere [11]. Selection of the appropriate settings is always facilitated by a steady-state investigation of the fluorescence or absorbance changes using a conventional spectrophotometer.

Detection of Emission: Emitted fluorescence is generally detected by a photomultiplier after passage through cutoff and/or band-pass filters selected to pass fluorescence while excluding any scattered exciting light (*see* **Note 4**). It is important that the filters selected maximize the signal change relative to the total signal.

Lamp Selection: Xenon arc lamps have a relatively smooth emission spectrum while mercury or xenon/mercury lamps have several intense emission bands which can be used when doing time-based acquisition at a single wavelength. Emission from deuterium or quartz halide lamps is less intense but is also less noisy, and these lamps can be used in absorbance measurements and for fluorescence excitation in the visible region.

Slit Widths: A large slit width can be used to increase the light intensity for fluorophores with a large Stokes shift (the wavelength difference between the excitation and emission maxima). If the Stokes shift is small then the excitation slit width may need to be reduced to exclude scattered light from the photomultiplier. Alternatively, the wavelength of the exciting light may be set to a shorter wavelength than the excitation maximum.

Time Constant: The signal-to-noise ratio (S/N) in rapid kinetic measurements is proportional to the square root of the instrumental time constant and should be selected to be <10% of the half time of the fastest process being observed (*see* **Note 3**). The S/N ratio can also be improved by averaging several individual records or, with some instruments, by collecting data at very high sampling rates and averaging appropriate blocks of data to give the individual time points.

In most cases, the data may be collected using linear time scales (*see* **Note 5**). Analysis of kinetic transients by fitting one or more exponential terms to the curves obtained is often straightforward, and the software supplied with commercially available equipment is generally adequate (*see* Subheading 3.5).

2.3 Samples Used in This Study

Wild-type and cysteine-containing mutants of *Drosophila* calmodulin were prepared and purified as described elsewhere [12]. The Asn111Cys calmodulin mutant was labeled with dansyl maleimide using standard methods. CaM-dependent protein kinase I (CamKI) was prepared and purified as described [13].

Peptides corresponding to the target sequences from CaMKI (CaMKIp: IKKNFAKSKWKQAFNATAVVRHMRK) and neuromodulin (NMp: ATKWQASFRGHITRKKLKG) were synthesized in-house with and without an N-terminal dansyl label. The chromophoric calcium chelator Quin 2 was obtained from local suppliers.

3 Methods

3.1 Simple Bimolecular Binding Reactions

Many protein–ligand interactions are simple, reversible second-order reactions (**Scheme B**), and we focus first on the determination of the rate constants for such processes. Stopped-flow kinetic studies of reactions of this type are generally performed under pseudo-first-order conditions with the concentration of one of the reagents (either P or L) at least tenfold higher than the other (*see* Subheading 1.3). The choice of the component to be used in excess may be dictated by availability of material; if this is not the case, it should be selected so that the ratio of signal change to total background signal is maximized (*see* **Note 6**). In this section, we describe an investigation of the interaction of Ca_4–CaM with a fluorescently labeled peptide NMp. A typical experiment will involve the following steps:

1. Perform a few "blank" shots with the fluorescent component alone to establish the starting fluorescence level.
2. Perform initial measurements with the component in excess, $[X_{tot}]$, at a tenfold excess. The actual concentration selected will depend on the instrument being used and on the intensity of the fluorophore but in the studies reported here the protein is generally the component in excess and initial measurements typically start with the protein concentration in the range 1–10 μM.
3. Optimize the concentration of $[X_{tot}]$ if necessary; If the reaction is too fast (more or less complete within the dead time), the concentrations of both components will need to be reduced. The lowest usable value of $[X_{tot}]$ will be that which still maintains pseudo-first-order conditions but is high enough to give good S/N. If the reaction is slow then the concentrations can be increased if this improves S/N.
4. When a suitable signal change has been obtained examine it over a wide range of time scales in order to confirm that there is only a single exponential process (*see* Fig. 2). Slow changes sometimes observed in fluorescence measurements can be caused by photobleaching of the fluorophore. This is not usually a problem on time scales of <1 s and can be quantified by mixing the fluorophore with buffer and recording any decrease in fluorescence intensity. Reducing the excitation slit width will reduce photobleaching.
5. Perform measurements at several different values of $[X_{tot}]$. Inspect all transients and average at least five at each concentration for subsequent analysis. If inconsistent results

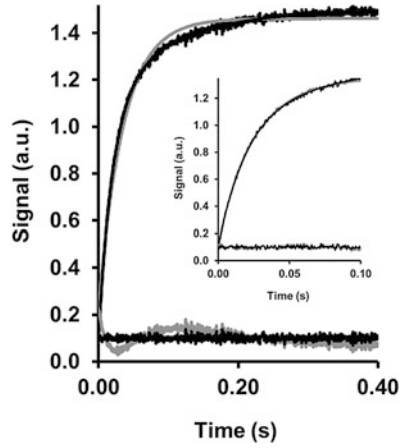


Fig. 2 Simulated double exponential time courses. The ability to detect deviations from single exponential behavior depends, of course, on the relative rates and relative amplitudes of the two phases and on the time range over which the transient is recorded. These transients were simulated over different time ranges with rates (and amplitudes) of 50 s^{-1} (amplitude = 1) and 10 s^{-1} (amplitude = 0.4). When analyzed over 10 ms (Inset), the one exponential (grey line) and two exponential (black line) fits are almost indistinguishable. When analyzed over a 40 ms time scale (main panel), it is clearly evident that the single exponential fit is inadequate. In both panels, the residuals (observed signal minus fitted signal) are shown as grey and black lines for the one and two exponential fits, respectively

are obtained from push to push, it is advisable to check carefully for artifacts (*see* **Note 8**).

6. Use Eq. 2 for the first attempt at curve fitting. Studies of simple bimolecular reactions performed under pseudo-first-order conditions should yield single exponential transients for which Eq. 2 gives the observed rate (k_{obs}). Eq. 2 can also be used to analyze fluorescence anisotropy data if there is no change in fluorescence intensity accompanying the reaction (*see* **Note 7**).
7. Extend the study to higher values of $[X_{tot}]$ while keeping the concentration of the other component unchanged. Although the largest possible concentration range should be studied this is not always possible. The largest usable value of $[X_{tot}]$ will depend upon the values of the individual rate constants. Thus, for example, with an association rate constant (k_1) of the order of 10^7 M^{-1} s^{-1} and an instrument dead time of 2 ms, the largest value would be of the order of 50 μM ($(k_{obs} = k_{+1}[X_{tot}] + k_{-1}) > 500$ s^{-1}) but would be significantly lower if k_{-1} is large.
8. Inspect average and fit traces for each concentration measured.

9. Plot k_{obs} vs. $[X_{tot}]$. If the reaction conforms to **Scheme B**, the plot will be linear with slope k_1 and intercept k_{-1} (*see* **Note 9**). It is important that the largest possible concentration range be covered because the demonstration that k_{obs} varies linearly over an extended concentration range is necessary to confirm that **Scheme B** is an adequate description of the process.

3.1.1 Results and Common Problems

Although the use of a large range of $[X_{tot}]$ should permit accurate determination of both kinetic constants this will not always be the case. For low-affinity interactions, where the dissociation rate constant (k_{-1}) is likely to be large, it may be possible to cover only a limited range of $[X_{tot}]$ before k_{obs} becomes too fast to measure and the association rate constant (k_1) will not be accurately determined. A very common problem with high-affinity interactions is that the value of k_{-1} is too small to be accurately determined. Figure 3 shows data for the interaction of Ca_4–CaM with the fluorescently labeled peptide NMp. The association rate constant (k_1) is well determined ($1.14 \pm 0.07 \times 10^7$ M^{-1} s^{-1}) but the dissociation rate constant (k_{-1}) is clearly not (0.48 ± 2.1 s^{-1}). If both rate constants are determined, they should be used to calculate a value for the equilibrium dissociation constant (K_d) for the interaction using the relationship $K_d = k_{-1}/k_1$ (*see* **Note 10**). In this particular case, the ratio of the kinetic constants (k_{-1}/k_1) gives an essentially meaningless K_d of 42 ± 185 nM.

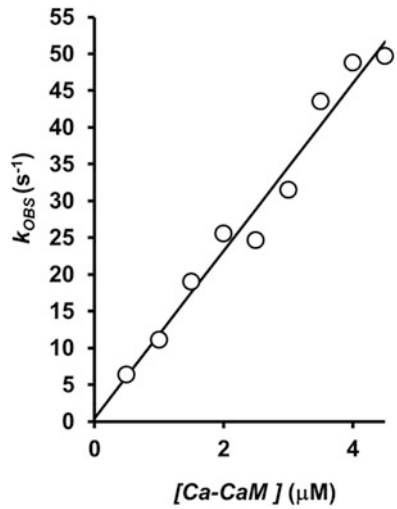


Fig. 3 Determination of kinetic constants for the interaction of Ca_4–Calmodulin with a fluorescently labelled peptide (NMp). The association rate constant (k_1, the slope, *see* Eq. 3) is well determined ($1.14 \pm 0.07 \times 10^7$ M^{-1} s^{-1}) but the dissociation rate constant (k_{-1}, the *y*-axis intercept) is clearly not (0.48 ± 2.1 s^{-1}). See text for full details

3.1.2 Alternative Approach

An accurate value for k_{-1} ($0.8 \pm 0.15\ s^{-1}$) was determined using the methods described in Subheading 3.2 in which the fluorescently labeled peptide was displaced by the addition of a large concentration of the corresponding unlabeled peptide. The ratio of the kinetic constants then gave a K_d of 70 ± 14 nM, in good agreement with a value of 60 ± 9 nM determined by direct fluorometric titration.

3.2 Competitive Binding

As noted in Subheading 3.1, it is not always possible to determine accurate **dissociation rate constants** from the intercept of a plot of k_{obs} vs. $[X_{tot}]$ if the dissociation rate is very slow. Likewise, it might not be possible to measure very fast **association rate constants** if the concentrations required for a good S/N ratio are such that the reaction is complete within the instrument dead time. Flow methods can be used to measure a more accurate value of the dissociation rate constant, if a displacement experiment can be performed. For example, a fluorescent ligand (L) can be displaced from its complex with the protein (PL) by adding an excess of a nonfluorescent ligand (N) which must be known to bind to the same site. The relevant reactions are shown in **Scheme C**:

$$P + L \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} PL$$

$$P + N \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} PN$$

Scheme C

3.2.1 Measuring k_{-1} Using an Excess of N to Compete Labeled L from PL

1. Select concentrations of P and L that give a reasonable saturation of P. This requires that the K_d for formation of PL is known. If not known from other experimental approaches, it should be determined as outlined in Subheading 3.1.
2. Stopped-flow mix such a solution of P *and* L with a solution of N and observe the dissociation of L from P. This should be a single exponential process and analysis using Eq. 2 will give an observed rate constant k_{obs}.
3. When $[N_{tot}]$ is high enough (so that when L dissociates from P it cannot reassociate before N binds), the observed dissociation rate of PL (k_{obs}) will be equal to k_{-1}. To confirm that this condition is met, it is important to measure k_{obs} using several different values of $[N_{tot}]$. If k_{obs} increases with $[N_{tot}]$ then the condition is not met and the true value of k_{-1} will only be determined if k_{obs} reaches a limiting (or plateau) value at high $[N_{tot}]$.

Dissociation rate constants for nonfluorescent ligands N can be determined using similar approaches involving competition with a fluorescent ligand L [14] (*see* **Note 11**).

3.2.2 Measuring the Association Rate Constant for N Binding to P

The **association rate constant** for binding of N to P (k_{+2} in **Scheme C**) can be measured using competition with L.

In this case, the experiment involves the following steps:

1. Stopped-flow mix P with a premixed solution of N *and* L. The concentration of L ($[L_{tot}]$) should be selected as that concentration that gives reasonable saturation of P in the absence of N. This requires that the K_d for formation of PL is known. If not known from other experimental approaches, it should be determined as outlined in Subheading 3.1.
2. Repeat the measurement using different $[N_{tot}]$ at a fixed value of $[L_{tot}]$.
3. If the dissociation rate constants (k_{-1} and k_{-2}) are small (Note: more complex behavior will be observed if the dissociation rate constants are not small) and both L and N are in large excess over P, then the observed first-order rate constant will be given by:

$$k_{obs} = k_1[L_{tot}] + k_2[N_{tot}] \quad (5)$$

A plot of k_{obs} vs. $[N_{tot}]$ should give a straight line with slope k_2 and intercept $k_1[L_{tot}]$ (Eq. 5). Any value for k_1 determined in this way should be compared with that determined using approach described in Subheading 3.1.

Figure 4 shows data for the system described in Subheading 3.1 in which Ca_4–calmodulin (0.1 μM) was reacted with a solution containing a fixed concentration of fluorescently labelled peptide NMp ($[L_{tot}] = 2$ μM) and varying concentrations of the unlabeled peptide ($[N_{tot}]$). The intercept ($=k_1[L_{tot}] = 23.95 \pm 2.3$) gives a k_1 value of ~1.2×10^7 M^{-1} s^{-1} in excellent agreement with the value determined in Subheading 3.1. The slope corresponds to a k_2 value of 6.5×10^6 M^{-1} s^{-1}, indicating that the kinetic properties of the unlabeled peptide differ somewhat from those of the labeled one. This was confirmed when direct fluorescence competition titrations showed that the unlabeled peptide has a K_d of 18 ± 5 nM (cf. 70 ± 14 nM for the labeled peptide). The k_2 and K_d values correspond to a dissociation rate constant (k_{-2}) of ~0.12 s^{-1} (cf. ~0.8 s^{-1} for the labelled peptide).

3.3 Ternary Complex Formation

There are many cases where a protein binds two different ligands to form a ternary complex. **Scheme D** is for a protein P interacting with two ligands (X and Y) to form the ternary complex PXY.

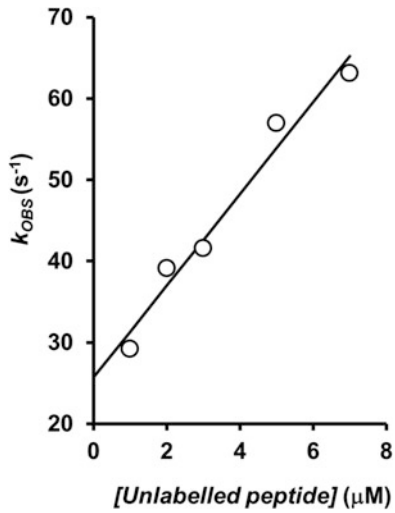
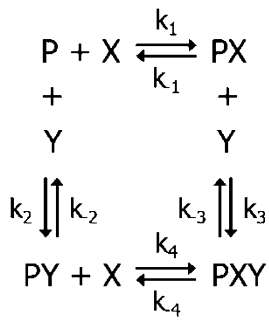


Fig. 4 Determination of the association rate constant for the interaction of Ca_4–Calmodulin with an unlabeled peptide. Ca_4–calmodulin (0.1 μM) was reacted with a solution containing a fixed concentration of fluorescently labelled peptide (NMp: 2 μM) and varying concentrations of the corresponding unlabeled peptide. The slope corresponds to an association rate constant for the unlabeled peptide of 6.5×10^6 M^{-1} s^{-1}. See text for full details



Scheme D

Given suitable optical signals, it would be possible, at least in principle, to study all four steps of the cycle individually. The interaction of P with X and P with Y could be studied using the methods described in Subheading 3.1 for **Scheme B**. The interaction of Y with PX would be studied in a similar way by mixing a solution of P *and* X (at concentrations of X that saturates P) with a solution of Y under pseudo-first-order conditions ($[\Upsilon_{tot}] \gg [P_{tot}]$), and so on. Only in ideal cases will it be possible to study all four steps individually; nevertheless, it is sometimes still possible to derive useful information from the study of a subset of the steps, particularly regarding the possibility of cooperativity between binding sites (*see* **Note 12**). For example, if the equilibrium dissociation constants determined for the interaction of Y with P and for Y with

PX are not the same then the presence of X alters the affinity for Y and the system is cooperative.

3.4 Multistep Reactions

Although many protein–ligand interactions conform to **Scheme B**, this is not always the case. The most obvious indication of additional complexity is the observation of more than a single kinetic phase, i.e., a time course of fluorescence change does not fit to a single exponential. Or, when only a single kinetic phase is observed, complexity is most often indicated by the observation that the variation in k_{obs} with concentration is not linear. One of the most commonly encountered complexities is the presence of an additional step involving an isomerization. This can be a first-order isomerization of PL following an initial second-order binding event (**Scheme E**) or a first-order isomerization of P (or L) followed by a second-order binding event (**Scheme F**).

$$\mathrm{P + L} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{PL} \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} \mathrm{PL^*} \qquad \text{Scheme E}$$

$$\mathrm{P} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{P^* + L} \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} \mathrm{PL^*} \qquad \text{Scheme F}$$

Simple equilibrium binding measurements cannot show that different conformational states exist for P or PL as analysis of binding curves for both of these schemes will always appear to be consistent with **Scheme B** with the single measured dissociation constants given by:

$$K_d = \frac{K_{d1}K_{d2}}{1 + K_{d2}} \quad \text{(Scheme E)} \tag{6}$$

$$K_d = K_{d2}(1 + K_{d1}) \quad \text{(Scheme F)} \tag{7}$$

The individual equilibrium dissociation constants are defined as $K_{d1} = k_{-1}/k_{+1}$ and $K_{d2} = k_{-2}/k_{+2}$ for both schemes.

The experimental approach is essentially the same as for studies of **Scheme B** and many of the same considerations apply. For example, if L is chosen as the component in excess, then pseudo-first-order conditions should be maintained ($[L_{tot}] \gg [P_{tot}]$) and the widest possible range of $[L_{tot}]$ should be covered. If the second-order binding step in **Scheme E** is very much faster than the isomerization step, and L is in large excess over P, then a stopped-flow record will, in ideal cases, have two kinetic phases. In the case of fluorescence measurements, the appropriate equation for a two exponential function would be:

$$F(t) = \Delta F_F \exp(-k_{obs}(F)t) + \Delta F_S \exp(-k_{obs}(S)t) + F_\infty \tag{8}$$

where $k_{obs}(F)$ and $k_{obs}(S)$ are the observed rate constants of the fast and slow components, ΔF_F and ΔF_S are the associated amplitudes, and F_∞ is the final florescence value.

The observed rate for the fast process ($k_{obs}(F)$) should vary linearly with $[L_{tot}]$, and the observed rate for the slow process ($k_{obs}(S)$) should vary hyperbolically with $[L_{tot}]$ [15] (*see* **Note 13**):

$$k_{obs}(F) = k_1[L_{tot}] + k_{-1} \tag{9}$$

$$k_{obs}(S) = \frac{k_2[L_{tot}]}{K_{d1} + [L_{tot}]} + k_{-2} \tag{10}$$

What will actually be observed experimentally will depend on the relative contributions of the different species to the optical signal being monitored, as well as on the values of the individual rate constants [10, 11]. Only in the most favorable cases where two easily resolvable kinetic events are observed over a wide range of $[L_{tot}]$ values (with the bimolecular step *always* remaining very much faster than the isomerization for all $[L_{tot}]$) will it be possible to extract all four rate constants by analysis of Eqs. 9 and 10 (*see* **Note 14**). Figure 5 shows data obtained for the interaction of a fluorescently labeled Ca_4–calmodulin (an Asn111Cys mutant labeled with dansyl maleimide) with CaM-dependent protein kinase I (CamKI) where all four rate constants could be determined. Analysis of the fast phase data using Eq. 9 gave $k_1 = 5.17 \pm 0.33 \times 10^6\ M^{-1}\ s^{-1}$ and $k_{-1} = 3.33 \pm 0.78\ s^{-1}$, so that $K_{d1} = 0.644 \pm 0.157\ \mu M$. Analysis of the slow phase data using Eq. 10 with K_{d1} fixed at 0.644 μM gave $k_2 = 1.04 \pm 0.05\ s^{-1}$ and $k_{-2} = 0.093 \pm 0.034\ s^{-1}$, so that $K_{d2} = 0.089 \pm 0.033\ \mu M$. Eq. 6 gives $K_d = 53 \pm 14$ nM, in reasonable agreement with a value of 33 ± 6 nM determined by direct fluorometric titration [13].

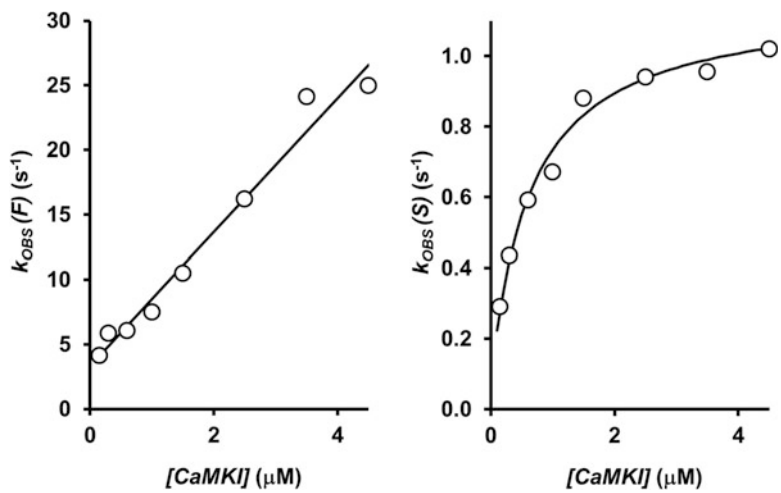


Fig. 5 Variation of $k_{obs}(S)$ and $k_{obs}(F)$ for the interaction of CaMKI with a fluorescently labeled calmodulin. $k_{obs}(F)$ (Left) varies linearly with [CaMKI] with slope and intercept k_{+1} and k_{-1}, respectively. $k_{obs}(S)$ (Right) varies hyperbolically from k_{-2} at low [CaMKI] to (k_{-2} + k_{-2}) at high [CaMKI]. See text for full details

In many cases, it will not be possible to extract all four rate constants. Thus, for example, although inspection of Eq. 10 shows that $k_{obs}(S)$ should *increase* from k_{-2} when $[L_{tot}] \ll K_{d1}$ to $(k_{-2} + k_2)$ when $[L_{tot}] \gg K_{d1}$, this will not always be observable. If K_{d1} is low then given the nature of the typical stopped-flow experiment, it is unlikely that it will be possible to work under conditions where $[L_{tot}] \ll K_{d1}$, and $k_{obs}(S)$ will only vary significantly with $[L_{tot}]$ when $k_{-2} \ll k_2$, so that k_{-2} will generally be difficult to determine. In extreme cases, K_{d1} may be so low that $k_{obs}(S)$ may well be completely independent of $[L_{tot}]$ under all attainable experimental conditions and only the sum of the rate constants for the isomerization step will be measurable.

If, on the other hand, the bimolecular step is the fast diffusion controlled formation of an encounter complex with very low overall affinity (high $K_{d1} \sim 1$ mM) then the value of K_{d2} would need to be 10^{-3} to give a typical overall equilibrium dissociation constant of 1 μM (*see* Eq. 6). A typical stopped-flow experiment could then show only a single transient process because occupancy of the intermediate PL would always be very low. The single observed rate might then vary linearly with $[L_{tot}]$ (since deviations from linearity would only be observed when $[L_{tot}]$ approached K_{d1}, *see* Eq. 10) with an apparent second-order rate constant of k_{+2}/K_{d1} and dissociation rate constant k_{-2}. When $[L_{tot}]$ does approach K_{d1}, some curvature may, of course, be observed and the *initial* slope can then be taken as equal to k_{+2}/K_{d1} [16]. Association rate constants measured for protein–ligand interactions are, in fact, often significantly lower than the values predicted using theoretical calculations based on diffusion coefficients, shape, and viscosity [2]. **Scheme E** with a high K_{d1} and a low k_{+2} is frequently invoked as an explanation for the observation of these unexpectedly low values [11, 17].

In the case of **Scheme F**, with the bimolecular step very much faster than the isomerization step, the expressions for experiments performed under the condition that $L_{tot} > (P_{tot} + P^*_{tot})$ are [15]:

$$k_{obs}(F) = k_2[L_{tot}] + k_{-2} \tag{11}$$

$$k_{obs}(S) = \frac{k_{-1}K_{d2}}{K_{d2} + [L_{tot}]} + k_1 \tag{12}$$

Scheme F can, at least in principle, be distinguished from **Scheme E** by the fact that the observed rate for the slow process should *decrease* from $(k_{-1} + k_{+1})$ when $[L_{tot}] \ll K_{d2}$ to k_{+1} when $[L_{tot}] \gg K_{d2}$. However, as for **Scheme E**, only in the most favorable cases will it be possible to extract all four rate constants for the reaction.

Many multistep mechanisms will consist of a series of three or more first- and second-order reactions and it is seldom, if ever, possible to derive analytical solutions for a kinetic analysis using

flow methods. In this situation, the most commonly used approach is to try to study the individual steps in isolation [16, 18, 19].

3.5 Data Analysis and Simulation

In the preceding discussion, we have assumed that fitting the time dependence of the observed signal to one or more exponential terms will always be possible. However, the requirement for explicit analytical solutions to the rate equations places severe constraints on the experimental conditions that can be employed, and it will not always be possible to work within these constraints. For example, if it is not possible to work under pseudo-first-order conditions, it will be necessary to analyze progress curves using an iterative method based on numerical integration of the appropriate differential rate equations [20].

Global analysis methods allow one to fit multiple kinetic data sets obtained under different concentration conditions [17]. The simultaneous analysis of the different data sets has the potential to achieve better definition of the rate constants common to all the sets. In favorable cases, it may allow the determination of kinetic constants not obtainable by traditional methods and can be used to distinguish between different kinetic models. Another strong point of global analysis is that the different data sets can be obtained using different methods, e.g., fluorescence intensity and anisotropy data, in which the kinetic constants are nevertheless the same. In such cases, it is important to weight the different data sets correctly. This can be done by determining the standard deviation in the signal of a reaction that has reached equilibrium. For example, using the last 5 ms of the transient shown in Fig. 2 would give a good estimate of the standard deviation.

Having extracted rate constants by any of the methods described here, it is almost always instructive to simulate the results in order to see how well the data actually fits the assumed mechanism. This is most often done at the level of simulating how k_{obs} values depend upon the concentrations of the reagents. It can also be very instructive to simulate individual reaction traces. This can be done using any one of several freely available packages (http://sbml.org/SBML_Software_Guide/SBML_Software_Summary) that will simulate changes in concentrations with time. Although many of these methods are very sophisticated, the principles are relatively easy to understand and the simplest methods can be implemented in a conventional spreadsheet. For example, **Scheme E** is described by the following set of coupled ordinary differential equations (ODEs):

$$d[P]/dt = d[L]/dt = -k_1[P][L] + k_{-1}[\mathrm{PL}]$$

$$d[\mathrm{PL}]/dt = k_1[P][L] - k_{-1}[\mathrm{PL}] + k_{-2}[\mathrm{PL}^*] - k_2[\mathrm{PL}]$$

$$d[PL^*]/dt = k_2[PL] - k_{-2}[PL^*]$$

There is no analytical solution to this set of ODEs. However, if an initial set of concentrations is provided, it is possible to create a *numerical* solution. In the simplest implementation [21], a time step Δt is chosen over which none of the concentrations are expected to change by more than a very small amount. The concentration changes after dt are calculated by multiplying the expressions for the rate by the time interval. For example:

$$d[PL^*] = [PL^*]_{t+dt} - [PL^*]_t \approx dt(k_2[PL] - k_{-2}[PL^*])$$

where the subscripts t and $t + dt$ indicate current concentration and predicted concentration after the time step dt, respectively. The new concentrations are then calculated by adding these changes to their (known) current values. This is done for all equations in the set, and the process is repeated until a preset end time is reached (*see* **Note 15**).

Multiplying the calculated concentrations by the appropriate optical constants (such as extinction coefficients) will then generate the theoretical (noise free) transient and what might actually be observed experimentally can be created by the addition of normally distributed random noise to this theoretical curve. In Microsoft Excel, for example, one may do this using the function NORMINV by writing = NORMINV(RAND(),T,SD), where T is the theoretical value and SD is the required standard deviation on this value. Whichever fitting package or program one is using can then be tested to see how well it actually performs under a variety of different conditions.

Finally, computer simulation is also invaluable as a teaching tool and a useful aid in the design of experiments. In our experience, intuitive arguments can frequently be wrong, even in apparently simple situations.

4 Notes

1. If no intrinsic optical signal is available, it is often possible to introduce a fluorescent label and comprehensive guides to probe selection and labeling procedures are readily available (https://www.thermofisher.com/uk/en/home/brands/molecular-probes.html). Labeling can be difficult if the protein contains more than a single site for the label because it may be difficult to obtain a reproducible product. Even when only a single site is available for labeling, this may be far from the binding site for the reaction partner, and not therefore report on the interaction. An alternative approach is to use genetic engineering to create a protein with a single cysteine residue that can then be specifically labeled. For all of these approaches,

it is essential that the modified protein is fully characterized and that the ratio of probe to protein should be determined. It should also be demonstrated that the modification does not affect any biological activity of the protein. Finally, equilibrium binding measurements should be performed to determine the affinity of the modified protein for the ligand and this should be compared with that of the native protein. This can be usually done using suitable fluorescence-based competition or displacement experiments or by using some of the other biophysical techniques described in this book.

2. If the fluorescence of the labeled protein does not change upon binding, it may be possible to study the interaction using anisotropy measurements. Such measurements require an instrument equipped with a polarizer filter in the excitation path which can be rotated to give either vertically or horizontally polarized light. Measurements are best made in what is known as the "T" format, with two detection photomultipliers equipped with polarizers positioned at right angles to the incident light direction for measurement of the intensity of the emitted light polarized parallel ($I_{\|}$) and perpendicular ($I_{\perp}$) to the plane of polarization of the exciting light. The two photomultipliers will respond differently to the parallel and perpendicular light and must first be normalized. This is done by exciting the fluorophore with horizontally polarized light and adjusting the high voltage on each photomultiplier so that they give the same output signal.

   The fluorophore is excited with vertically polarized light, and the intensity of the emitted light polarized parallel ($I_{\|}$) and perpendicular ($I_{\perp}$) to the plane of polarization of the exciting light is recorded. The total fluorescence intensity is given by ($I_{\|} + 2I_{\perp}$), and the anisotropy is calculated as $r = (I_{\|} - I_{\perp})/(I_{\|} + 2I_{\perp})$. The anisotropy is related to the fluorophore's rotational correlation time (τ_c) by the equation $r = r_o/(1 + \tau/\tau_c)$, where r_o is the limiting anisotropy of the fluorophore and τ is its excited state lifetime. Anisotropy measurements are particularly appropriate in the study of the binding of small fluorescent ligands to large macromolecules because τ_c is related to size and such reactions will therefore generally be accompanied by large increases in anisotropy. However, because anisotropy can be measured with high precision, it is also possible to use this approach using proteins labeled with a fluorophore.

3. The reciprocal of k_{obs} is called the *relaxation time*, or *time constant*, τ, of the system and is the time taken for the signal to change from S_0 to $(S_{eq} - (S_{eq} - S_0)/e)$. Although k_{obs} and τ^{-1} are identical, the former is generally used to describe

transients observed in flow experiments, whereas the latter is generally used to describe relaxation (or small perturbation) experiments. The *half-life*, $t_{1/2}$, of the reaction is defined as the time taken for the signal to change from S_0 to $(S_{eq} - (S_{eq} - S_0)/2)$ and is related to the relaxation time and observed rate through $t_{1/2} = 0.693\tau = 0.693/k_{obs}$ (Note: Ln $(0.5) = -0.693$).

4. Scattered light arises from three sources: Rayleigh scattering of the exciting light (observed at the excitation wavelength λ_{Ex}), Rayleigh scattering of the first harmonic of the exciting light (observed at $2 \times \lambda_{Ex}$), and Raman scattering from the water. The wavelength (in nanometers) for the Raman scattering peak (λ_R) for water depends on the excitation wavelength according to $\lambda_R = \lambda_{Ex}/(1 - 0.00034\ \lambda_{Ex})$.
5. In more complex systems, the observable processes may occur on very different time scales and it is then generally more appropriate to collect data with a logarithmic time base which allows data to be collected at longer time intervals as the reaction proceeds. Although the time constant will need to be set as less than the fastest process, the data can sometimes be collected in the oversampling mode (collecting and averaging blocks of data) to improve the *S/N* for long time points.
6. Thus, for example, if the ligand (L) and the complex (PL) are fluorescent, but the protein is not, then the protein should be the component used in excess. This may not be possible in all cases, and the ligand will then have to be the component in excess. In this case, it may be advantageous to use resonance energy transfer if a suitable donor/acceptor pair is available with a combination of intrinsic and/or extrinsic fluorophores. For example, the emission spectrum of tryptophan overlaps the excitation spectrum of 2′(3′)-*O*-(*N*-methylanthraniloyl)-adenine nucleotides and this has been taken advantage of in stopped-flow studies of the myosin subfragment 1 ATPase mechanism [22]. By exciting the tryptophan at 280 nm and observing the methylanthraniloyl emission, the bound fluorophore is preferentially excited over free fluorophore. This allows much higher concentrations of the excess fluorophore to be used compared to the situation where the methylanthraniloyl is excited directly.
7. If there is a significant change in fluorescence intensity accompanying the reaction, then the time-dependent change in anisotropy, $r(t)$, must be analyzed using [11]:

$$r(t) = r_{PL} + \frac{(r_L - r_{PL})}{1 - D + De^{k_{obs}t}}$$

where r_L and r_{PL} are the anisotropies of L and PL, and D is the fluorescence intensity of PL divided by that of L.

8. Many artifacts can occur in stopped-flow experiments and some of them can give rise to apparently perfect exponentials. One of the most common problems is the presence of air bubbles in the observation cell, and it is therefore advisable to use degassed solutions for all stopped-flow measurements. Inefficient mixing, poor thermal equilibration, and small leaks in the system may all give rise to apparently real transients. Mixing solutions with very different densities may also be problematic. A suitable control experiment will usually identify problems. For example, if the reaction being studied is that of a fluorescently labeled protein with a ligand, the control would be to mix the protein solution with the ligand solution, but with the ligand omitted.
9. Whenever possible, it is best to determine the variance in k_{obs} values for each value of the independent concentration variable, $[X_{tot}]$. The resulting sample variances may then be used to weight each k_{obs} value by the inverse of its estimated variance. In some cases, it may not be possible to obtain variances for individual samples, and it is then reasonable to assume that the *relative* error in k_{obs} is constant. The fitting should then be done to the logarithm of k_{obs}, since the error in $\log(k_{obs})$ will be constant. This is particularly important in cases where k_{obs} values vary by more than an order of magnitude.
10. A significant difference between the values may indicate that **Scheme B** is not an adequate description of the process. The observed variation in reaction amplitude should also be shown to be consistent with an independently measured K_d. The concentration of the protein–ligand complex formed following stopped-flow mixing is, of course, readily calculated from the total concentrations of protein and ligand present *after* mixing and the known K_d using:

$$[\mathrm{PL}] = \frac{([P_{tot}] + [L_{tot}] + K_d) - \sqrt{([P_{tot}] + [L_{tot}] + K_d)^2 - 4[P_{tot}][L_{tot}]}}{2}$$

11. The **dissociation rate constant** for N (k_{-2} in **Scheme C**) can be measured using fluorescently labelled L to induce the displacement of N from PN, although this may be technically difficult if high values of $[L_{tot}]$ are required, when the strong fluorescence from L will probably result in poor S/N.

    Dissociation of PN can also be induced by mixing with an excess of a compound that reacts with N rather than P. For example, the dissociation of Ca^{2+} (N) from calcium-binding

proteins such as calmodulin (P) can be studied by mixing with an excess of a fluorescent Ca^{2+} chelator (L) such as Quin 2, which forms a strongly fluorescing, high affinity 1:1 complex with Ca^{2+} [14].

12. Reactions such as those shown in **Scheme C** often exhibit cooperativity in ligand binding. That is, for example, the affinity of X for P may be increased (positive cooperativity) or decreased (negative cooperativity) when Y is also bound. Changes in affinity may be caused by changes in either or both of the rate constants defining the interaction with X. Conservation of free energy for this scheme dictates that $(k_{-1}k_{-3})/(k_1k_3)$ must be equal to $(k_{-2}k_{-4})/(k_2k_4)$.
13. If the second-order binding step in **Scheme E** is very much **slower** than the isomerization step, and L is in large excess over P, then a stopped-flow record will show single exponential behavior with k_{obs} given by the following equation:

$$k_{obs} = k_1[L_{tot}] + \frac{k_{-1}k_{-2}}{k_2 + k_{-2}}$$

14. When analyzing rate expressions such as that given in Eq. 8 it is, in general, not good practice to transform them into linear functions because the associated errors transform accordingly [23]. There are now numerous mathematical procedures available for χ^2 minimization of nonlinear functions such as these; for example, the Levenberg-Marquardt procedure is both efficient and relatively robust [24]. Fitting using the logarithms of rate and equilibrium constants is advisable because it forces them to be physically meaningful (positive) values.
15. The accumulation process described here is called numerical integration. Selecting smaller time steps will result in smaller relative changes, and in more accurate solutions, but also in an increased total simulation time. If the time steps taken are too large, the solution will not only lose accuracy but also may become unstable. In an unstable solution, the calculated values typically oscillate wildly, with amplitudes that increase with every new time step.

References

1. Cornish-Bowden A (1995) Fundamentals of enzyme kinetics. Portland Press, London
2. Gutfreund H (1995) Kinetics for the life sciences. Receptors, transmitters and catalysis. Cambridge University Press, Cambridge
3. Berridge MJ, Bootman MD, Lipp P (1998) Calcium-a life and death signal. Nature 395:645–648
4. Soderling TR, Stull JT (2001) Structure and regulation of calcium/calmodulin-dependent protein kinases. Chem Rev 101:2341–2352
5. Chattopadhyaya R, Meador WE, Means AR et al (1992) Calmodulin structure refined at 1.7 Å resolution. J Mol Biol 228:1177–1192
6. Heidorn DB, Trewhella J (1988) Comparison of the crystal and solution structures of

calmodulin and troponin C. Biochemistry 27:909–915

7. Finn BE, Evenas J, Drakenberg T et al (1995) Calcium-induced structural changes and domain autonomy in calmodulin. Nat Struct Biol 2:777–783
8. Barman TE, Bellamy SR, Gutfreund H et al (2006) The identification of chemical intermediates in enzyme catalysis by the rapid quench-flow technique. Cell Mol Life Sci 63:2571–2583
9. Shastry MCR, Luck SD, Roder H (1998) A continuous-flow capillary mixer to monitor reactions on the microsecond time scale. Biophys J 74:2714–2721
10. Martin SR, Schilstra MJ (2013) Rapid mixing kinetic techniques. In: Williams M, Daviter T (eds) Protein-ligand interactions. Methods in molecular biology (methods and protocols), vol 1008. Humana Press, Totowa, NJ, pp p119–p138
11. Eccleston JF, Martin SR, Schilstra MJ (2008) Rapid kinetic techniques. Methods Cell Biol 84:445–477
12. Browne JP, Strom M, Martin SR et al (1997) The role of β-sheet interactions in domain stability, folding, and target recognition reactions of calmodulin. Biochemistry 36:9550–9561
13. Clapperton JA, Martin SR, Smerdon SJ et al (2002) Structure of the complex of calmodulin with the target sequence of calmodulin-dependent protein kinase I: studies of the kinase activation mechanism. Biochemistry 41:14669–14479
14. Martin SR, Andersson-Teleman A, Bayley PM et al (1985) Kinetics of calcium dissociation from calmodulin and its tryptic fragments. A Quin 2 stopped-flow fluorescence study reveals a two-domain structure. Eur J Biochem 151:543–550
15. Halford SE (1971) Escherichia coli alkaline phosphatase. An analysis of transient kinetics. Biochem J 125:319–327
16. De La Cruz EM, Ostap EM, Sweeney HL (2001) Kinetic mechanism and regulation of myosin VI. J Biol Chem 276:32373–32381
17. Eccleston JF, Hutchinson JP, White HD (2001) Stopped-flow techniques. In: Harding SE, Chowdhry BZ (eds) Protein-ligand interactions: structure and spectroscopy. Oxford University Press, Oxford
18. De La Cruz EM, Wells AL, Rosenfeld SS et al (1999) The kinetic mechanism of myosin V. PNAS 96:13726–13731
19. Eccleston JF, Petrovic A, Davis CT et al (2006) The kinetic mechanism of the SufC ATPase: the cleavage step is accelerated by SufB. J Biol Chem 281:8371–8378
20. Kuzmic P (1996) Program DYNAFIT for the analysis of enzyme kinetic data: application to HIV proteinase. Anal Biochem 237:260–273
21. Schilstra MJ, Martin SR, Keating SM (2008) Methods for simulating the dynamics of complex biological processes. Methods Cell Biol 84:807–842
22. Woodward SKA, Eccleston JF, Geeves MA (1991) Kinetics of the interaction of 2′(3′)-O-(N-methylanthraniloyl)-ATP with myosin subfragment 1 and actomyosin subfragment 1: characterization of two acto.S1.ADP complexes. Biochemistry 30:422–430
23. Johnson ML (2008) Nonlinear least-squares fitting methods. Methods Cell Biol 84:781–805
24. Press WH, Teukolsky BP, Vetterling WT et al (1990) Numerical recipes. The art of scientific computing. Cambridge University Press, Cambridge

Chapter 4

Kinetic Methods of Deducing Binding Mechanisms Involving Intrinsically Disordered Proteins

Elin Karlsson and Per Jemth

Abstract

There are multiple examples of protein–protein interactions involving one intrinsically disordered protein region binding to an ordered protein domain in a coupled binding and folding reaction. Similarly to protein folding studies, much effort has been devoted to understanding the mechanisms of such coupled binding and folding reactions. In this chapter, we describe how kinetics can be used to assess binding mechanisms with focus on fluorescence-monitored stopped-flow experiments. The approach can be applied more generally to any protein interaction with or without a coupled conformational change and to other kinetic techniques. Determining binding mechanisms is a great challenge and while "proving" a mechanism may be futile, it is possible to deduce the simplest scenarios, which are consistent with experimental data.

Key words Binding mechanism, Intrinsically disordered proteins, Kinetics, Stopped flow

1 Introduction

1.1 Intrinsically Disordered Proteins

A large fraction of the proteome contains regions that are disordered, i.e., amino acid sequences that do not fold into well-defined compact structures but remain highly dynamic under physiological conditions [1]. These intrinsically disordered proteins (IDPs) could contain transient structure, and they may have short unfolded regions or disorder which may be a property the whole protein. Disordered regions can contain interaction motifs that bind to folded protein domains. Upon binding, the disordered region usually adopts a well-defined conformation, but there are examples where disorder prevails in the protein–protein complex [2]. In other cases, the disordered regions flanking the binding motifs modulate binding through short-lived attractive or unfavorable repulsive interactions within the complex [3]. Interactions involving IDPs are being increasingly identified using bioinformatics as well as in structural studies and so this mode of binding must be advantageous. For example, signal transduction pathways and the transcriptional machinery provide multiple examples of such

Tina Daviter et al. (eds.), *Protein-Ligand Interactions: Methods and Applications*, Methods in Molecular Biology, vol. 2263, https://doi.org/10.1007/978-1-0716-1197-5_4,

interactions [1]. One reason appears to be flexibility in binding, both regarding structure (plasticity with regard to bound conformation) and number of binding partners (promiscuity). Because of this, several biophysical studies of IDP interactions have tried to address the mechanism(s) of binding and find quantitative evidence for these predicted advantages of disorder [4].

IDPs and their interactions have been studied using the whole battery of available biophysical techniques such as circular dichroism, calorimetry, fluorescence spectroscopy, single molecule techniques, and NMR. The latter is particularly powerful, since it provides both structural and dynamic information [5], but a combination of multiple experimental approaches is always preferable when trying to deduce mechanisms. Importantly, kinetics is the only method that can unambiguously determine mechanisms in protein–protein interactions, protein folding, and all other biological reactions [6]. This chapter considers how to study interactions involving IDPs with the goal of determining mechanism, in particular the collection and analysis of kinetic data obtained using stopped-flow spectroscopy. We discuss basic principles of the methodology, experimental design, caveats and limitations of the approach, and the analysis and interpretation of the data. While we describe stopped flow, the approach is valid for other kinetic methods.

1.2 A Kinetic Approach to Assess Mechanism

The association constant (K_a) between biomolecules tells us how much they will bind to each other at equilibrium under a given set of conditions and concentrations. Kinetic experiments tell us what happens before this equilibrium is reached. If performed properly, kinetic experiments can therefore provide clues to the reaction mechanism; in the case of IDPs, what happens along the binding trajectory as the free proteins interact, fold, and form a complex. During protein folding and in binding reactions, it is surprisingly common that intermediate states do not accumulate and therefore cannot be detected easily in experiments. The process then consists of one visible step, it is apparently "two state" (only unfolded and folded, or free and bound, states are populated). Such a one-step (two state) binding mechanism (Scheme 1) will give rise to one "kinetic phase," i.e., an experimental binding trace, which follows a single exponential function with the observed rate constant k_{obs}. A two-step binding mechanism (three state; free, intermediate, and bound, Schemes 2 and 3) gives two kinetic phases (with two k_{obs} values) and a three-step (four state) three phases, and so on.

$$A + B \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} AB \qquad \text{Scheme 1}$$

$$A + B \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} AB \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} AB^* \qquad \text{Scheme 2}$$

$$A + B \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} A + B^* \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} AB^* \qquad \text{Scheme 3}$$

While this relationship between number of states and theoretical number of kinetic phases is clear, there are complicating factors facing the experimentalist in the real laboratory world when they attempt to reverse this argument by inferring mechanism from the number of kinetic phases. Except for the simplest one-step binding scenario this is not trivial. For example, consider a two-step binding, which results in two kinetic phases (Scheme 2). Kinetic phase one is usually not solely dependent on reaction step 1 and phase two on reaction step 2. Instead both steps contribute to both observed rate constants. How much each step contributes to the respective kinetic phase depends on the microscopic rate constants (k_1, k_{-1}, k_2, and k_{-2}) and the concentrations of proteins. Moreover, late binding steps can be kinetically silent such that a multistep mechanism appears as one step. Equally, there are quite likely several high-energy intermediates on any coupled binding and folding pathway; a two-state process is then observed, but it represents an average of several steps. Finally, the kinetic phases may be related to off-pathway events, i.e., nonproductive interactions not leading to the complex visualized in the crystal or NMR structure. In fact, proteins and in particular IDPs are very dynamic molecules. Therefore, a large number of states are most likely populated during the binding and folding reaction. Such rapidly interconverting ensembles may be structurally characterized by NMR [7], but exactly how they interconvert is usually experimentally inaccessible. Thus, the important take-home message is that we can only determine the simplest mechanism that is consistent with our data. We can often determine that a certain binding and folding reaction involves accumulating intermediates, as evinced by more than one kinetic phase. For some multistep binding mechanisms, it is even possible to deduce the order of events, for example, if an IDP (or part of it) folds before or after binding to its interaction partner. However, it is imperative to exercise the upmost care in interpretation of data and be humble to its limitations. In this chapter, we outline how to approach this task systematically.

Classically, techniques such as stopped-flow, continuous-flow, quenched-flow, pressure jump, and temperature jump were used to measure pre-equilibrium kinetics of chemical reactions. More recently, sensor-based technologies such as surface plasmon resonance, NMR, and the emerging field of single-molecule experiments provide an expanded toolbox for the kinetic experimentalist. Out of these methods, stopped flow is the simplest and arguably the most robust method to measure kinetics for biomolecules. Much of the basics of protein–ligand interactions were covered by Williams [8] and rapid mixing techniques by Martin and Schilstra [9]. We will cover some basic kinetic concepts here to facilitate discussion but for a more comprehensive description of biomolecular kinetics, we recommend the excellent book from Bagshaw [10].

In the simplest case, a one-step binding between proteins A and B, we can determine the association rate constant k_{on} and the dissociation rate constant k_{off} (Scheme 1). The ratio of the rate constants is the equilibrium dissociation constant K_d ($K_d = 1/K_a$).

The rate, v for formation of the bimolecular complex AB is

$$v = \frac{d[AB]}{dt} = k_{on}[A][B] - k_{off}[AB] \quad (1)$$

To simplify analysis, kinetic experiments are usually performed under "pseudo-first-order conditions" where, for example, $[B] \gg [A]$ such that the concentration of $[B]$ does not change significantly during the course of the binding reaction. The experimentally measured rate constant k_{obs} under such pseudo-first-order conditions is given by the solution of the differential to Eq. 1;

$$k_{obs} = k_{on}[B] + k_{off} \quad (2)$$

It is easy to confuse rate constants with reaction rate and so care should be given to the use of these terms as recently pointed out [11]. What are the meanings of the rate constants determined in Eq. 2? k_{off} is best understood as its reciprocal, $1/k_{off} = \tau_{off}$, which is the average lifetime or time constant of the AB complex. Thus, a low k_{off} means a high τ_{off} and a long-lived complex. k_{on} is a more deceptive parameter since it often lures even experienced kineticists into the conclusion of "fast binding." The overall formation of AB from A and B depends on the rate constants *and their concentrations*. For example, in a binding experiment under conditions of 50% complex formation at equilibrium (i.e., when $[B] = K_d$), the contribution to the time constant τ_{obs} ($=1/k_{obs}$) is equal for k_{on} and k_{off}. Thus, how much k_{on} contributes to k_{obs} depends on the concentration(s) of the interacting molecules. Indeed, a larger k_{on} will result in higher initial rate v_0, which is the rate v at time = 0, when $[AB] = 0$ and only $k_{on}[A][B]$ (and not $k_{off}[AB]$) contributes to k_{obs} (Eq. 1). But v_0 is dependent on the concentrations of A and B, so we could have a large k_{on} but a low v_0.

2 Materials

2.1 The Stopped-Flow Instrument

In the current chapter, we will mainly use stopped-flow experiments as examples. However, the approach is valid for any kinetic method. Whichever is the method of choice, the important thing is to have control over potential artifacts and other caveats and to know the limitations of the technique. Stopped flow and the other classical kinetic methods perturb the equilibrium of the system and measure the "relaxation" to the new equilibrium. In this respect, stopped flow is very efficient since it mixes two solutions, which enables large perturbations. The mixing time ("dead time") of the very best stopped-flow devices with T-type mixers is claimed to be around 1 ms; however, in our experience 2 ms is more common. The mixing time is also dependent on the properties of the solvent and the ratio of mixing (usually 1:1 in binding experiments but for folding experiments often an asymmetric 10:1 mixing). Each experimenter should determine the mixing time for their particular instrument and setup using established model reactions (*see* **Note 1**). The mixing time and the kinetic amplitude determine how large k_{obs} values can be determined reliably in a particular experiment. The observed kinetic amplitudes will usually start to decrease when the reaction is too fast, i.e., when the time constant approaches the mixing time. The instrument then captures only part of the kinetic transient. For example, if the instrumental mixing time is 2 ms, an average lifetime τ of the reaction in the same order would correspond to a k_{obs} of 500 s^{-1}. Then, after 2 ms 63% of the reaction is complete, and the final 37% of the transient can be used for curve fitting. This might be feasible if the signal-to-noise of the kinetic amplitude is good enough. However, in practice, 200–400 s^{-1} can be the upper limit for measured k_{obs} values. At the other end, slow transients and long measuring times suffer from backflow of solutions in the tubings, photobleaching of the samples, or other instrumental problems, which limit the measuring time to 10–100 s (k_{obs} of 0.01–0.1 s^{-1}) depending on experimental setup and protein samples. The temporal limitation of the instrument and experiment must be judged on a case-by-case basis.

2.2 Protein Samples

In protein binding as well as folding studies using stopped flow, fluorescence is the most common way to monitor the binding and/or folding reaction. The great advantage with fluorescence is that it is very sensitive to conformational changes. Any interactions involving change of local environment of a Trp (or to some extent Tyr) side chain, or other extrinsic fluorescent probes if such are used, will translate into a change in fluorescence that can be monitored over time. However, one problem can be that there is no native Trp or Tyr in the interacting proteins, or they are located in positions not affected by the interaction. In such cases, a Trp could

be engineered at an appropriate position using site-directed mutagenesis. However, this might affect the kinetics and affinity of the interaction and proper controls must be performed. For example, k_{off} for the wild-type protein variant should be determined in a displacement reaction described below and the overall K_d with isothermal titration calorimetry [12] or other equilibrium method. In this way, the K_d, k_{off}, and k_{on} ($=k_{off}/K_d$) can be compared for wild-type and Trp variant before extensive kinetic experiments are initiated. Another control is to put the Trp at a different position or engineer a fluorescence signal into the other binding protein and check for consistency in the equilibrium and kinetic parameters. In fact, a screen of different Trp variants is usually necessary to find the most appropriate one for detailed kinetic studies.

In experiments involving coupled binding and folding of IDPs, it is sometimes appropriate to use synthesized peptides corresponding to only the disordered region involved in binding. Peptide synthesis provides an easy route to introduce Trp into the peptide at a suitable position. In addition, peptides can be chemically ligated to larger and often brighter fluorophores during synthesis. This has advantages in giving a discrete and selectable wavelength for this component of the interaction. Chemical labeling of expressed proteins with these fluorophores is also done routinely and has been used to study IDPs [13] but requires additional steps.

In order to obtain an accurate k_{on} value, the concentration of the varied protein or peptide must be known accurately. Absorbance is by far the best method for this [12, 14] particularly if a Tyr, Trp, or other extrinsic fluorophore with a strong absorbance in the UV or visible spectrum is present. If this is not possible, as in Subheading 3.6.3 below, then absorbance at 205 nm can be used but the concentration should be confirmed using quantitative amino acid analysis. The concentration is usually the main source of error in k_{on} and K_d determinations and this should be kept in mind when interpreting data (*see* **Note 2**).

2.3 Buffers

Buffers should be selected in which the proteins of interest remain stable for the duration of stopped-flow measurement that may be considerably longer than the time of individual kinetic runs. Systematic changes in buffer composition can also be used to investigate binding mechanism. For example, properties such as ionic strength, pH, chaotropic, or antichaotropic activity have all been classically used in protein folding studies and can likewise be applied to IDP interactions [15–17]. Such variation in buffer conditions can stabilize or destabilize intermediates, or affect the transition state(s) of the binding reaction. If an intermediate is stabilized, it may be detected in the kinetic experiment and in some cases be shown to be present on the productive binding pathway. In other words, an apparent two-state binding (Scheme 1) could be "tuned"

into a three-state binding mechanism (Schemes 2 or 3). Similarly, high salt concentrations may be used to show that electrostatic interactions are important for the association of the proteins (Subheading 3.6.5).

3 Methods

3.1 Designing and Performing Binding Experiments

Experiments must be optimized with regard to sample availability, protein concentrations, solubility, affinity, presumed mechanism, etc. Other factors may also be important but these may only become evident following initial pilot experiments. The result of a kinetic experiment is a trace (or transient) in which fluorescence (or absorbance, fluorescence polarization, etc.) changes over time. In a typical binding experiment, if the fluorescent probe such as a Trp is in protein A, then it is mixed rapidly with varying concentrations of an excess of protein B. However, it should not matter whether protein A or B is in excess, the result should be the same in either case (*see* **Note 3**).

1. Stopped-flow instruments use conventional arc lamp light sources. Switch on the lamp 15–30 min before use such that the light output is stable when starting the experiments.
2. It is important to perform kinetic experiments at a well-defined temperature. Switch on the thermostat of your system (usually a water bath), set the experimental temperature, and let the system equilibrate before making any measurements. Use the internal temperature probe as the experimental temperature rather than the temperature of the water bath as the instrumental probe is closer to the mixing cell. Even if this instrumental probe is not accurately reporting an exact temperature at the point of measurement, it still allows experimental temperature to be precisely reproduced. Depending on the magnitude of the observed rate constants of the reaction under investigation, the temperature should be set such that reliable data can be recorded (i.e., rate constants within the stopped-flow range). As a rule of thumb, rate constants increase by a factor of 2 for every 10 °C increase and measuring k_{obs} at different temperatures is therefore a good control experiment to rule out artifacts (*see* **Note 4**). On the other hand, decreasing temperature reduces collisional quenching from solvent and therefore enhances fluorescence levels and the size of associated changes in fluorescence during binding. Stopped-flow instruments can even work at temperatures close to 0 °C providing care is taken that the system does not leak.
3. For fluorescence measurements, select an excitation wavelength and choose an appropriate fluorescence emission filter. For example, Trp is typically excited around 280 nm and

fluorescence emission monitored using a 320 nm cutoff (long-pass) filter, which ensures that scattered excitation light does not reach the detector. However, the exact choice of excitation wavelength and emission filter will be dictated by the fluorescent probe in the experimental system, for example, in the case of extrinsic probes that are excited and emit at longer wavelengths. It can also be very useful to check the effects of using different emission filters. For example, if there are two or more Trps involved in the binding, the changes in fluorescence signal upon binding might counteract each other (one increasing and one decreasing in fluorescence upon binding). Therefore it is sometimes better to use a bandpass (interference) filter, which transmits light centered around 330 nm with defined upper and lower limits (±25 nm), or use a long-pass filter transmitting above 360 nm. These options can be determined empirically or with reference to equilibrium measurements (*see* **Note 5**). If an intermediate accumulates during the binding, its fluorescence might also be better captured with these alternative emission filters, resulting in larger amplitudes for one or both kinetic phases.

4. Set the acquisition ("push") volume using the threaded backstop adjuster that is located under the stop syringe on Applied Photophysics SX instruments. Binding experiments are usually performed with a 1:1 mixing using 2500 μL sample syringes, allowing rather small volumes (e.g., 100 + 100 μL). In folding studies, it is common to use asymmetric mixing and larger drive volumes should then be used to ensure proper mixing (e.g., 30 + 300 μL). The volume must not affect the observed kinetics and this should be checked systematically during initial experiments. To do this, test the reaction at different push volumes. As long as a decrease in volume does not affect the rate and amplitude of the kinetic transient the lower volume can be safely used.
5. Check the integrity of the fluidic system carefully before making stopped-flow measurements as any leaks will generate a level of continued flow in the observation cell that may appear as an extra kinetic phase. Begin every experimental session by a 1–2 s acquisition using "pressure hold" on the sample syringes. (Typically, a pressure of 3 bar is used in symmetrical mixing experiments.) Monitor the sample syringes carefully; they should stand completely still while the pressure is held on and slightly bounce back when the pressure is released. If there is a leakage, every connection from the stop syringe all the way back to the sample syringes must be checked to locate the leak. If the leak is in one of the valves then it is likely that it must be replaced with a new one. Leaking tubing connections can usually be retightened. After prolonged use or work at low

temperatures, the sample syringes can become worn and prone to leaking and so may need to be replaced. However, before this step it is worthwhile to remove the syringe plungers and boil them for a couple of minutes. The cleaning and expansion of the plunger tip can make them work without leaking and extend their useable lifespan.

6. Introduce protein A into one of the sample syringes at 1 μM final concentration after mixing if using Trp fluorescence (*see* **Note 6**) and buffer into the other syringe. Let the solutions equilibrate to the set temperature (*see* **Note 7**). Make 2–3 drive shots until the fluorescence signal has stabilized, i.e., the tubing is filled with the respective solutions. The number of drive shots necessary to fill the tubing will depend on the push volume and the length of tubing connecting the syringe to the cell. Record a measurement of protein A mixed with buffer from the other drive syringe and check if this gives a constant signal and flat transient. When measuring over longer acquisitions this may reveal a slow linear decrease in signal that is due to photobleaching. Adjust the detection gain on the instrument photo multiplier tube (PMT) such that this signal level is in the middle of the instrument's dynamic range (e.g., 5 V for a 0–10 V scale).
7. Replace the buffer solution with protein B at 10 μM final concentration after mixing, and make 2–3 drives such that the tubing is filled with protein B all the way to the observation cell. At this point, you may observe whether the fluorescence signal decreases or increases upon addition of B. If the signal increases such that it is close to the PMT upper limit adjust the gain such that the end point signal is not higher than 80% of detection maximum. If B contains a fluorescent probe, this must also be taken into account before conducting an experiment with a series of concentrations of B; the highest concentration of B, and thus highest fluorescence, will then dictate what PMT voltage can be used. In this respect, remember that the fluorescence may decrease upon binding but the initial fluorescence of unbound A and B could potentially be even higher and saturate the detector. In such a case, set the PMT voltage so that the extrapolated starting fluorescence is around 80% of detection maximum.
8. Make the first acquisitions with an intermediate time base of 1 s and look for a transient (*see* **Note 8**). Then adjust the time base according to what you observe in the experiment in terms of time for the transient (*see* **Note 9**) and make further adjustments of the fluorescence signal if necessary. As a rule of thumb, a kinetic transient should be monitored over 5τ for optimal curve fitting. Even if you observe a transient that is completed within 1 s, make a new experiment where you

monitor the reaction on shorter and longer timescales to look for additional kinetic phases (*see* **Note 10**). It is advisable to collect at least four individual recordings and average them before proceeding to analysis. If the signal-to-noise is poor or the model ambiguous then more acquisitions can be made and averaged. Sometimes one or more trace(s) in a series of acquisitions displays obvious artifacts. Such traces should be discarded and not included in the average. However, to reduce any risk of bias by "cherry picking," any trace that is removed from the data set should be replaced with a couple of new acquisitions.

9. If the experiment gives a clear kinetic trace, it should be fitted to a single exponential equation with a fixed end point, preferably using the software provided with the instrument since it provides instant feedback for how to proceed. Initial data points should be removed according to the mixing time analysis (*see* **Note 1**). The residuals of the curve fitting should be inspected. Are these residuals randomly distributed over the entire recording or are there systematic deviations? In the event of nonrandom residuals, a second exponential term should be included and the data refitted. In rare cases, triple exponential transients are observed. Judging if a trace is appropriately fit or whether additional exponentials are needed is a critical part of kinetic studies. Figure 1 provides examples of fitting to three sets of experimental data.
10. After observing initial kinetic traces it may be useful to reconsider the concentration of protein A. The recommendation of 1 μM of A is a good starting point but if there is sufficient signal amplitude then this concentration can be reduced. The greater the ratio of $[B]/[A]$ the more optimal are pseudo-first-order conditions. On the other hand, sometimes a higher concentration of A may be required to obtain a clear change in fluorescent signal upon binding, in particular, if intrinsic Tyr fluorescence is monitored.
11. The next step, whether a single, double, or triple exponential is observed, is to measure binding at different concentrations of B while keeping the concentration of A constant and check how k_{obs} changes. Observe if any systematic deviations in the fitting residuals are reproducible and present at all concentrations of B.

Having collected data at different concentrations of B and fit these to an appropriate model allows for some inferences about binding mechanism as is explained and exemplified in more detail using a number of case studies below.

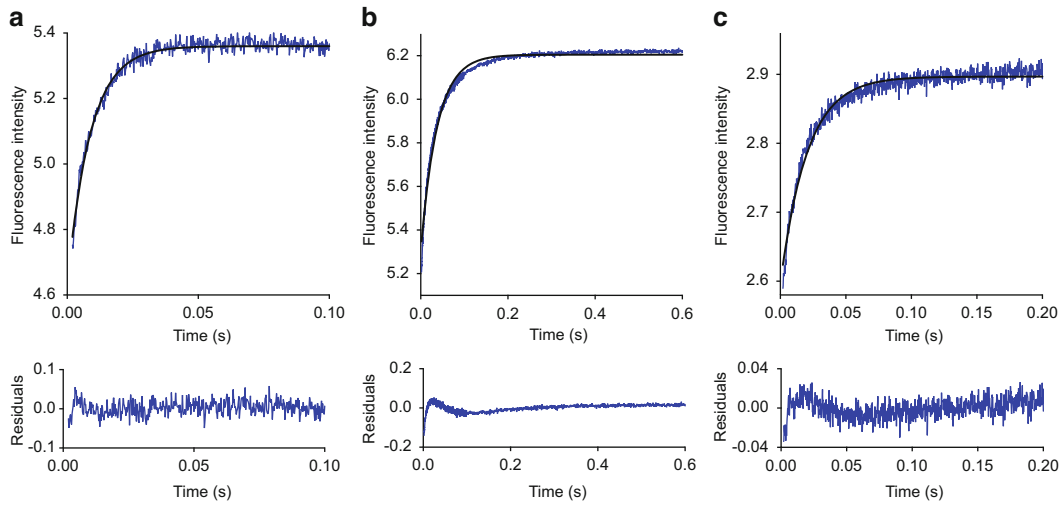


Fig. 1 Examples of curve fitting to experimental kinetic traces. Each trace shown is an average of 4–5 individual experiments run back-to-back. The data were fitted to a single exponential function, and residuals are reported below each trace. (**a**) The trace fits well to a single exponential. There is a tiny tendency of a trend in the residuals but not clear enough to warrant a double exponential. (**b**) A clear example where a double exponential equation is valid. The trend in the residuals is clear. (**c**) An ambiguous case. There is a trend in the residuals but the noise is almost of the same magnitude as the "amplitude" of the trend. Sampling of more kinetic traces might improve signal-to-noise further. In this case, data need to be acquired over a range of concentrations and analyzed with both single and double exponential functions. The concentration dependence of k_{obs} value(s), and kinetic amplitudes might indicate whether this is a true multistep binding (*see* Subheadings 3.6.1–3.6.4)

3.2 Dealing with Second-Order Conditions

Sometimes experimental conditions may be limited by protein concentrations, signal amplitudes, and rate constants and must therefore be performed under second order rather than pseudo-first-order conditions. The analysis is a little more complicated but there are two approaches for dealing with second-order conditions. The first is to analyze all binding traces simultaneously using numerical fitting in free software such as Dynafit (http://www.biokin.com) [18] or other commercial software such as KinTek [19] or MATLAB. In such global fitting of experimental traces, k_{obs} values will not be obtained but the output will be the k_{on} and k_{off} values, and fits to each experimental transient where residuals can be analyzed. The second option is to fit individual traces to a single exponential despite being in the second-order region. If they fit well to a single exponential, k_{obs} may be plotted versus $[B]$ and fitted to

$$k_{obs} = \sqrt{k_{on}^2\left([A]_0 - [B]_0\right)^2 + k_{off}^2 + 2k_{on}k_{off}\left([A]_0 + [B]_0\right)} \quad (3)$$

where $[A]_0$ and $[B]_0$ are the initial concentrations of the respective proteins. This equation accounts for second-order conditions at low $[B]_0$ and approaches a linear function at high $[B]_0$ as in Eq. 2 under pseudo-first-order conditions [20].

3.3 Three State Binding Reactions

Any binding mechanism beyond one-step becomes very complex very rapidly, and data from multistep mechanisms are challenging to analyze. Microscopic rate constants from two different steps (Schemes 2 and 3) will couple when they are of similar magnitude, which happens at lower concentrations of B. Moreover, the magnitude of the kinetic amplitudes may be very different such that one is small in relation to the other. Both of these scenarios can produce ambiguous looking exponentials (Fig. 1c). For example, we consider a two-step binding (Schemes 2 and 3), which theoretically will result in two kinetic phases where one increases linearly and the other hyperbolically with $[B]$ according to Eq. 4 (Scheme 2) or Eq. 5 (Scheme 3), under pseudo-first-order conditions.

$$k_{\mathrm{obs1,2}} = \frac{k_1[B] + k_{-1} + k_2 + k_{-2} \pm \sqrt{(k_1[B] + k_{-1} + k_2 + k_{-2})^2 - 4(k_1[B]k_2 + k_{-1}k_{-2} + k_1[B]k_{-2})}}{2} \tag{4}$$

$$k_{\mathrm{obs1,2}} = \frac{k_1 + k_{-1} + k_2[B] + k_{-2} \pm \sqrt{(k_1 + k_{-1} + k_2[B] + k_{-2})^2 - 4(k_1k_2[B] + k_{-1}k_{-2} + k_1k_{-2})}}{2} \tag{5}$$

What often happens in a real experiment is that in the time window where we can measure kinetics, the kinetic phases are on similar scales and one of the associated amplitudes might dominate over the other. The resulting experimental trace will not be perfectly represented by a single exponential. Neither is the data good enough to justify the use of a double exponential. A fit to a single exponential will then yield k_{obs} values intermediate between the two real ones given by the respective theoretical kinetic phase. Generally, k_{obs} values that differ by a factor of 3–4 or less result in exponential transients, which may be very difficult to resolve independently unless there are favorable kinetic amplitudes. Therefore, giving general advice on "borderline" curve fitting cases is particularly difficult. The best thing to do in such cases is to perform experiments over a range of protein concentrations, fitting the data for both single and double exponentials and plot k_{obs} and amplitudes versus $[B]$ and, if possible, versus $[A]$ as well (*see* Subheading 3.6.2 below). If the dependences of k_{obs} and amplitudes from the double exponential fit are in accordance with what can be expected from a two-step binding, the fitted parameters are likely sound. It may also be useful to change experimental conditions such that a suspected intermediate is stabilized, for example, by increasing the ionic strength or using sodium sulfate, 2,2,2-trifluoroethanol or trimethylamine *N*-oxide (*see* **Note 11**).

3.4 Displacement Experiments to Determine k_{off}

It is common that k_{off} is not well determined in binding experiments by linear extrapolation to the *y*-axis (zero concentration of B). This occurs when k_{off} is small in relation to the lowest measured k_{obs}. The k_{off} should then be determined separately in a

displacement experiment. In such experiment, the dissociation of the complex is turned into an essentially irreversible reaction by trapping one of the interacting proteins with an alternative protein partner C (Scheme 4). Protein C could be anything that binds either A or B as long as there is a change in fluorescence upon displacement, i.e., the fluorescence of the AC complex must be distinct from that of the AB complex.

$$AB + C \xrightarrow{k_{off}^{AB}} A + B + C \underset{k_{off}^{AC}}{\overset{k_{on}^{AC}[\mathrm{C}]}{\rightleftharpoons}} AC + B \qquad \text{Scheme 4}$$

In practice, C is chosen as a variant of A or B with different fluorescence properties. For example, if A contains a Trp, which is used to monitor the binding, a variant without the Trp can be used. If a synthesized peptide is used as the disordered binding motif, it can also easily be modified such that its fluorescent properties upon binding are different, for example, by using another fluorescent probe than that used for monitoring binding. k_{obs} is then determined by rapidly mixing the AB complex (*see* **Note 12**) with a large excess of C (preferably 50–100-fold) and monitoring the formation of AC complex. Since the contribution of rebinding of B gets lower at higher concentration of C, k_{obs} will accordingly approach $k_{\mathrm{off}}^{\mathrm{AB}}$ at high C (*see* **Note 13**) as the dissociation of AB becomes irreversible, as depicted in Scheme 4. Thus, k_{obs} should be determined at a range of $[C]$ (e.g., 20-, 50-, and 100-fold excess) and in reaching a common value confirming that the excess is large enough to justify the approximation $k_{\mathrm{obs}} \approx k_{\mathrm{off}}^{\mathrm{AB}}$.

3.5 Further Control Experiments and Common Artifacts

Instrumental and biological–based artifacts can often appear as exponential kinetic phases or as linear drift. The uttermost care must be taken to corroborate that what is analyzed is something real in the binding reaction and to rule out anything related to instrumental or other errors. Artifacts often occur at the end of longer recordings due to photobleaching or diffusion of solutions in the instrument. Therefore, an important control experiment is to mix protein A with buffer to confirm that this is a flat line over the longest time window used in the experiment. If protein B also contains a fluorescent side chain, then it should also be mixed with buffer in a separate experiment. Alternatively, mix protein–protein complex with an identical protein–protein complex solution to obtain a fluorescence signal identical to the one obtained in the binding experiment. Again, this should be a flat line. Photobleaching may appear as a decay of the signal during long acquisition times (10 s to minutes) and this could interfere with interpretation. If the photobleaching (or other linear drift) is

established in the control experiments described above, a linear term may be added to the exponential equation used to analyze the transient, to account for this effect.

It is common to observe artifacts resulting from (bad) mixing in the early parts of the trace. This is another reason to be careful when interpreting rate constants in the range 300–400 s^{-1} or higher, in particular if the rate constants do not increase linearly with increasing protein concentration under pseudo-first-order conditions. Thus, a hyperbolic dependence of k_{obs} as expected from a three-state binding model (Schemes 2 and 3) may be due to an instrumental mixing limitation rather than any conformational change in the protein(s). While such artifact-based changes do not appear perfectly hyperbolic, they can be hard to distinguish from a genuine three-state binding. Decreasing the temperature or adding a cosolvent might reduce k_{obs} values into a range where data is reliable. Alternatively, the hyperbolic kinetics can be confirmed by another kinetic method such as temperature jump that does not use mixing as the source of equilibrium perturbation.

We have also noted an unusual artifact on a long timescale with a "rate constant" approximately 0.2 s^{-1}, i.e., in a range that could interfere with slow conformational changes in kinetic experiments. This artifact is not related to the more common photobleaching effect. Interestingly, the amplitude changes with temperature and reverses sign around room temperature while k_{obs} for the phase remains unaltered. The basis for the artifact is possibly related to small differences in temperature between flow lines and the observation cell.

3.6 Example Studies and What They Tell Us About IDP Binding

3.6.1 Case Study 1: p53TAD and MDM2, an Apparent One-Step Binding

p53 is a central transcription factor in cell cycle regulation. The transactivation domain of p53 (p53TAD) contains a conserved Trp residue, which is directly involved in binding to the TAD-binding domain of MDM2, a negative regulator of p53. p53TAD is intrinsically disordered in the free state but forms an α-helix upon binding to MDM2. Stopped-flow kinetic traces were obtained using a fixed concentration of MDM2 (1 μM) and a range of p53TAD concentrations (2–10 μM) and were well described by a single exponential fit (Fig. 2).

Observed rate constants were fitted using Eq. 2 (pseudo-first order) and Eq. 3 (second order) to obtain a slope of around 8 μM^{-1} s^{-1}, which represents k_{on} for the binding reaction (Fig. 2d). However, the extrapolated k_{off} values from these fits were much lower than any measured k_{obs} value, were close to zero and thus associated with a moderate error. Therefore, k_{off} was also measured independently using a displacement experiment. A complex formed by 1 μM MDM2 and 1 μM p53TAD was mixed with an excess (20 μM) of dansylated p53TAD peptide. Fitting this kinetic trace to a single exponential gave a k_{obs} of 0.63 s^{-1}, which

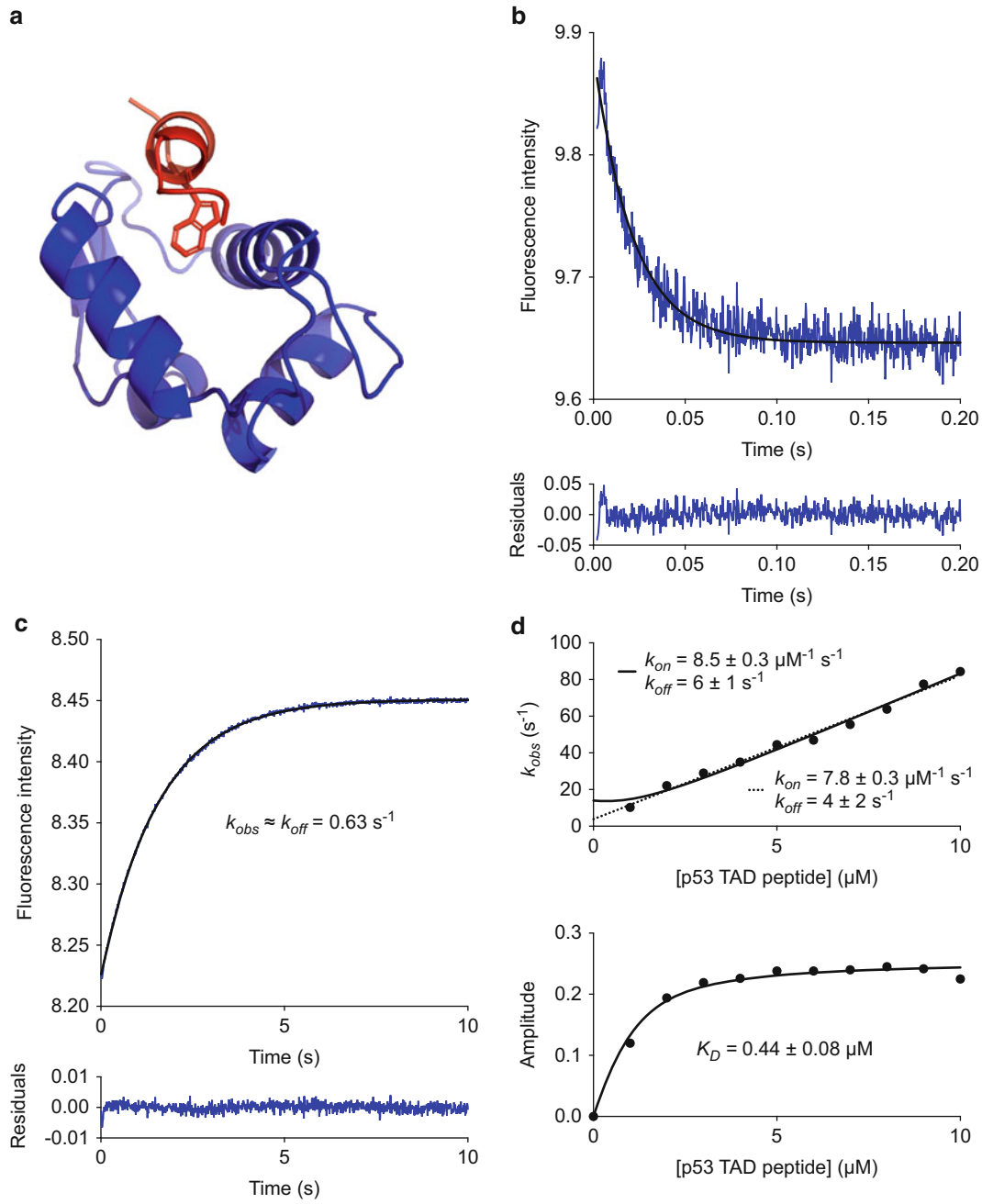


Fig. 2 p53TAD and MDM2 interact according to an apparent one-step (two state) mechanism. (**a**) Crystal structure of the complex between the p53TAD-binding domain of MDM2 and a peptide corresponding to the binding motif in p53TAD (PDB code: 1YCR). Note the Trp residue in p53TAD, which is crucial for the binding interface and thus appears to be a perfect probe for fluorescence-monitored binding. (**b**) Example of an experimental trace for binding that fits perfectly to a single exponential. The kinetic amplitude is surprisingly low and fluorescence intensity decreases upon binding, rather than increases, which could be anticipated

is a good approximation of k_{off}. There is a tenfold difference between extrapolated value of k_{off} (from Eq. 2 or 3) and that determined in the displacement experiment. The displacement measurements always report on the overall k_{off}, i.e., including all potential steps in the (un)binding pathway as well as concentration-independent forward rate constants. For example, Scheme 2 would result in an apparent $k_{off} = k_{-1}k_{-2}/(k_{-1} + k_{-2} + k_2)$. In the case of p53TAD and MDM2, there is no kinetic evidence for a slow conformational change. The crystal structure of MDM2 suggests that a helix must be displaced to allow binding of p53TAD. However, the observed kinetics fit perfectly to a two-state scenario [21]. Thus, any conformational changes of the helix must be on a faster timescale than is accessible by stopped-flow methodology.

Conclusions: *Single exponential binding kinetics where* k_{obs} *increases linearly with protein concentration is consistent with a one-step binding mechanism (two state). However, it does not rule out more complex mechanisms involving, e.g., fast conformational changes.* k_{off} *should be determined in a separate displacement experiment if it is low in comparison with the smallest* k_{obs} *directly measured.*

3.6.2 Case Study 2: N_{TAIL} and XD Domain, a Two-Step Binding

Measles virus expresses the nucleoprotein N_{TAIL}, which interacts with a domain from a viral phosphoprotein called XD. Upon binding, the disordered N_{TAIL} folds into an α-helix, a mechanism observed for several IDP interactions. The binding kinetics were found to occur on a faster timescale where stopped flow could not be used, and the authors instead employed temperature jump to perturb the binding equilibrium [15]. Using temperature jump is not nearly as efficient as mixing two solutions. The K_d for the binding equilibrium may not be particularly sensitive to temperature and in addition the protein complex continues to get more and

Fig. 2 (continued) from the crystal structure (burial of a Trp). (**c**) A displacement experiment using a dansylated p53TAD peptide was used to determine $k_{off} = 0.63\ s^{-1}$. (**d**) Upper panel, k_{obs} values increases linearly as a function of [p53TAD]. Since the concentration of MDM2 is 1 μM, the data points at low [p53TAD] are not determined under pseudo-first-order conditions. In such cases, Eq. 3 may be fitted to data (solid line). In the present case, a fit to a linear function (Eq. 2, pseudo-first-order conditions, dashed line) gives a similar result, but in either case k_{off} is not accurately determined. Lower panel, kinetic amplitudes associated with the k_{obs} values showed saturation as expected. The amplitudes were fitted to a binding isotherm (solid line), which yielded a $K_d = 0.44 \pm 0.08$ μM, which is different from the much more accurate K_d calculated from k_{off}/k_{on} (0.63/8.5 = 0.074 μM). The large discrepancy is due to the fact that the total [MDM2] in the experiment is 1 μM. Due to the high affinity, the titration becomes practically stoichiometric and the free [p53TAD peptide] is not equal to the plotted total concentration. Note the slight decrease in the amplitude at the highest concentrations resulting from an underestimation of the amplitude in the curve fitting when data is not corrected for the dead time of the instrument

more stable the higher the concentration of the protein added in excess. Despite this, it was possible to measure the coupled folding and binding reaction between N_{TAIL} and XD using an engineered Trp residue as fluorescence probe. The experimental traces obtained from the temperature jumps were well described by a single exponential as in the first case study. However, in contrast to p53TAD/MDM2, k_{obs} values were not increasing linearly with N_{TAIL} concentration as expected from a one-step binding mechanism, but instead displayed a hyperbolic dependence with a limiting value around 1000 s^{-1}. Such behavior is consistent with a two-step mechanism (three state). While a two-step mechanism should theoretically yield double exponential traces, it is quite common that one of the kinetic phases is not detected because it is too fast for the instrument or the amplitude is too small, or both. There has been considerable interest in whether three-state mechanisms involving IDPs follow an induced fit or conformational selection model. Induced fit requires the two proteins associate first and then fold into the equilibrium conformation of the complex (Scheme 2). Conformational selection requires that folding of a fraction of the IDP population precedes the binding step (Scheme 3). With regard to IDP interactions, the question often oversimplifies the problem since it is very hard to imagine a disordered polypeptide folding into the precise geometry of the complex before associating with its folded binding partner. However, in the case of the N_{TAIL} binding motif that is helical this might be possible since isolated helices may form transiently in absence of tertiary structure. The authors used a kinetic trick first devised by Olson et al. [22] and later rediscovered [23] to demonstrate that the N_{TAIL}/XD interaction indeed follows induced fit. This requires changing the concentration of N_{TAIL} and XD in separate experiments. If k_{obs} shows a hyperbolic dependence in both cases, the binding occurs before the conformational change (folding) with induced fit. If the dependence is hyperbolic for one of the proteins but linear for the other, folding happens before binding in conformational selection.

The authors could also demonstrate that the conformational change associated with the hyperbolic behavior of k_{obs} was associated with the folding of N_{TAIL} using the helix-stabilizing agent, 2,2,2-trifluoroethanol (TFE). Addition of TFE produced increases of k_{obs} at high [N_{TAIL}] and a less pronounced hyperbolic behavior. This is consistent with a lower barrier and larger rate constant for folding of the N_{TAIL} helix (an increase of k_2 in Scheme 2). Thus, N_{TAIL} binds to XD in a disordered state and folds into an α-helix in a second step.

Conclusions: *Single exponential binding kinetics where* k_{obs} *increases hyperbolically with ligand concentration is consistent with a two-step (three state) binding mechanism. This could be an induced fit*

*mechanism (Scheme 2) or a fast conformational change followed by binding (Scheme 3). "Fast" in this respect refers to a conformational change occurring on a timescale significantly shorter than that for the association reaction under the chosen ligand concentration (*see **Note 14***). The data do not rule out more complex mechanisms, for example, there could be several conformational changes (induced fit or conformational selection) occurring on faster timescales.*

3.6.3 Case Study 3: ACTR and NCBD, Multistep Binding with Several Kinetic Phases

ACTR and NCBD are interaction domains from two different transcriptional coactivators. Both are regarded as disordered although NCBD has a hydrophobic core and folds into a dynamic structure. ACTR and NCBD interact in a complicated coupled binding and folding reaction with several kinetic phases. Here, we will consider two kinetic phases detected in experiments in the presence of high ionic strength. The concentration of ACTR was varied (1–20 μM) at constant NCBD (1 μM). The kinetic traces were double exponential as assessed from the residuals (Fig. 3).

One of the kinetic phases increased linearly with [ACTR] while the other one appeared rather constant. The data were fitted to models describing a three-state binding: induced fit and conformational selection, respectively [16]. It is clear that both models describe the data equally well. The kinetic traces can also be fitted directly to each model using a numerical global fitting approach. The advantage of this is that the kinetic amplitudes can be taken into consideration. However, in this case it did not resolve the mechanism since we cannot assign a fluorescence signal to the intermediate state. Thus, it could not be concluded from the data whether the conformational change that produces the observed slow kinetic phase occurs before or after binding. What about the kinetic trick of varying NCBD at constant ACTR as used in the study of N_{TAIL} and XD binding? Unfortunately, this experiment was also inconclusive because the kinetic traces were not clearly double exponential but something in between, as illustrated in Fig. 1c. Comparison of k_{obs} values determined at excess NCBD and ACTR, respectively, suggested that curve fitting to a single exponential resulted in a weighted average k_{obs} value from the two kinetic phases. Thus, the combined observations are at best "more consistent" with an induced fit model [23].

Conclusions: *Double exponential binding kinetics proves that the interaction involves (at least) two distinct events (three state). If one kinetic phase increases linearly with the concentration of the varied protein and the other one appears hyperbolic, the experiment is*

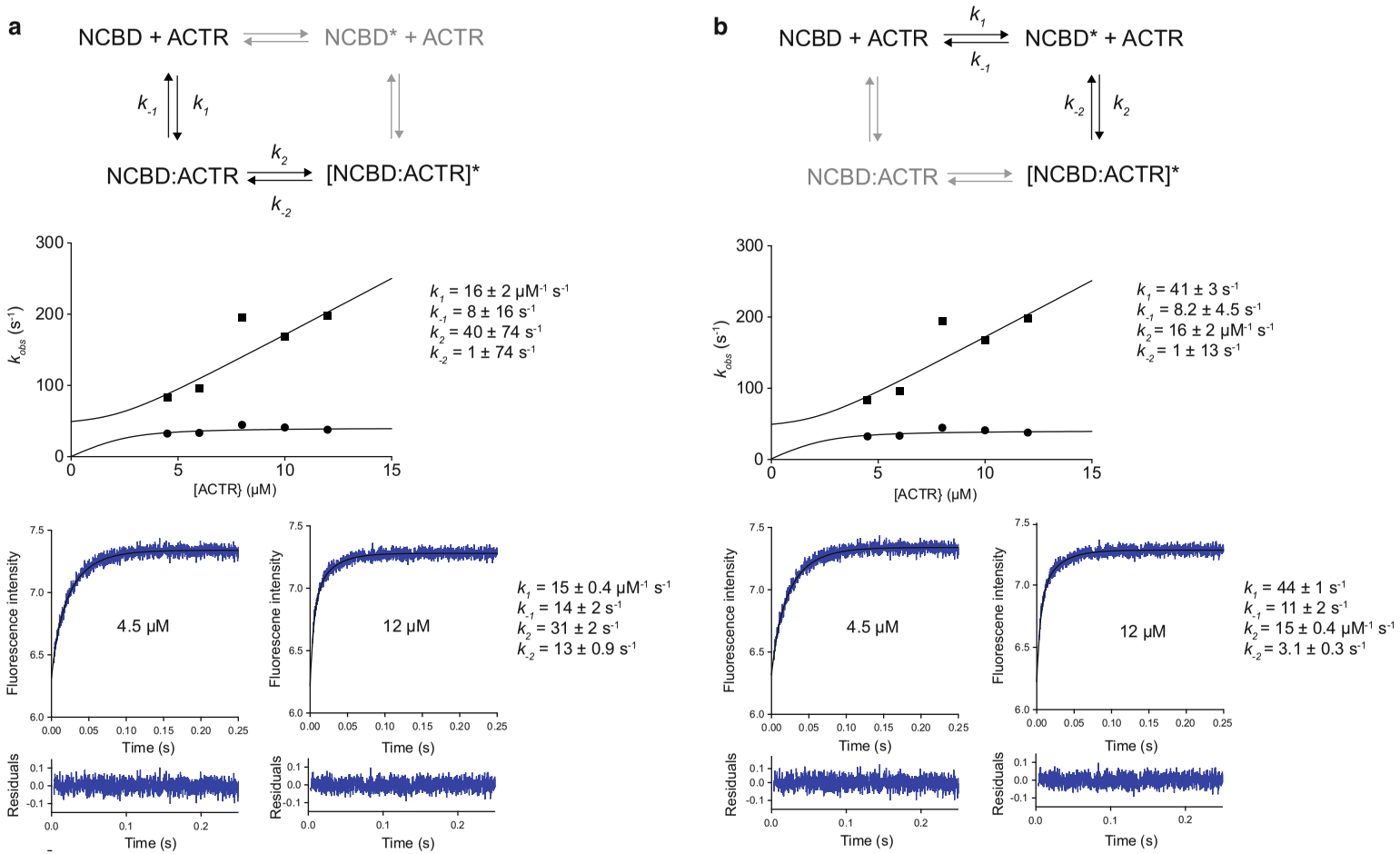


Fig. 3 (**a**, **b**) Double exponential binding kinetics for ACTR and NCBD. A comparison of fitting an induced fit or conformational selection model, respectively (upper panels), to kinetic data for ACTR and NCBD from Dogan et al. [16]. Middle panels; k_{obs} values were obtained from double exponential curve fits to each individual trace, plotted versus [ACTR] and fitted to an equation for induced fit (Eq. 4) or conformational selection (Eq. 5). Lower panels; Alternatively, the experimental traces were globally fitted numerically to each model using KinTek [19], taking into account the amplitudes of the traces (the resulting fit to two of the five traces are shown). In either case, it is impossible to distinguish the induced fit and conformational selection models. In the original reference, two more k_{obs} values were included, but these were determined at a lower voltage and the traces could not be used in the present global fit of the traces. Nevertheless, the global fit returns rate constants in good agreement with the original analysis and with lower fitting errors. The overall k_{off} was determined in a displacement experiment as 1.4 s^{-1}, while a value of 3.1 s^{-1} was obtained from both models in the global fit

*consistent with a two-step binding mechanism (three state) (*see **Note 15***). Curve fitting using Eqs. 4 and 5 can usually not distinguish induced fit from conformational selection when the conformational change is fast. However, if fluorescence yields can be assigned to free proteins, intermediate, and bound complex, the kinetic amplitudes can be used to rule out mechanisms. If experimentally possible, the concentrations of both proteins should be varied to assess the hyperbolic dependence of the slow kinetic phase.*

3.6.4 Case Study 4: HPV E7 and Rb, a Multistep Binding with Several Kinetic Phases

An example of a multistep binding mechanism deduced by stopped-flow kinetics is that for the intrinsically disordered N-terminus of the human papillomavirus E7 protein (HPV E7) and the folded human protein, Retinoblastoma tumor suppressor (Rb) [13]. The interaction acts to inhibit the Rb protein, thus promoting cell proliferation and HPV virus replication. In this work, a FITC label on HPV E7 was used as fluorescence probe giving the advantage that low concentrations of labeled HPV E7 protein could be used (5–50 nM). The authors compared the binding kinetics of HPV E7 as a short peptide containing only the binding motif with the kinetics of the entire disordered N-terminus. The shorter peptide displayed two-state kinetics while the longer disordered region followed a four-state mechanism. Interestingly, the longer protein bound 10 times tighter to Rb, illustrating the important role of disordered regions outside the designated interacting binding motif. Using experiments at a range of NaCl concentrations, the authors also highlighted the role of electrostatics in this IDP interaction and how they can be characterized. Following extensive kinetic experiments at different [NaCl] and with structural considerations, the authors concluded that the four-state mechanism most likely involve initial conformational selection in the disordered HPV E7 protein (about 50% is in a binding competent conformation in the free state) which is followed by association to Rb and a rearrangement of the complex. While a detailed description of the work is beyond the scope of this protocol, it is recommended as further reading illustrating an approach to very complex binding kinetics.

Conclusions: *Triple exponential binding kinetics proves that the interaction involves (at least) three distinct events (four state). If one kinetic phase increases linearly with [*B*] and the other two appears constant or hyperbolic, the experiment is consistent with one association step and two conformational changes. If one kinetic phase decreases with [*B*], it suggests that the initial step is a slow conformational change in A or B.*

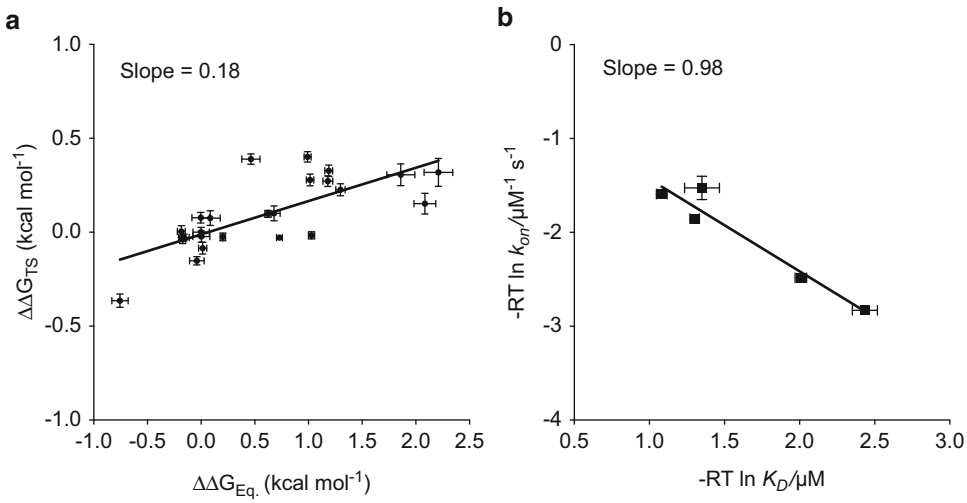


Fig. 4 Linear free energy diagrams for coupled binding and folding. (**a**) Site-directed mutagenesis to probe the interaction between ACTR and NCBD. (**b**) Ionic strength dependence of the interaction between ACTR and NCBD shows that electrostatic interactions play a role in the transition state. The effect ionic strength is not in k_{off} but only in k_{on}, as usually observed for protein–protein interactions. Data from Dogan et al. [27]

3.6.5 Case Study 5: Using Linear Free Energy Relationships to Access the Overall Properties of the Transition State for Binding and Folding

So far examples have considered how populated intermediates along the coupled binding and folding pathway result in detectable kinetic phases. But the transition states of the binding reactions are also accessible via kinetic measurements [24–26]. It has been seen in Subheadings 3.6.2 and 3.6.4 how systematic changes in buffer conditions can be used to probe the binding reaction. Another method of systematically perturbing protein–protein interactions is by using site-directed mutagenesis. Mutations will potentially affect low-energy populated ground states (free proteins, intermediates, bound complex) as well as high-energy transition states along the binding pathway. One strategy is to measure kinetics and plot logarithms of rate constants (which reflect the free energy barrier of the transition state) versus an equilibrium parameter that reflects changes in overall free energy of the system, for example, logarithms of equilibrium constants or concentration of a perturbing agent such as denaturant concentration. As an example, the coupled binding and folding of ACTR and NCBD described in Subheading 3.6.3 can be studied in this way. A large number of site-directed mutants were generated in this system [24] and the change in free energy upon mutation for the kinetic barrier ($RT \ln (k_{on}^{\text{wild-type}}/k_{on}^{\text{mutant}})$) was plotted versus the change in free energy at equilibrium ($RT \ln (K_d^{\text{mutant}}/\ln K_d^{\text{wild-type}})$) for each mutation (Fig. 4a).

The slope of this plot reports on the extent of formation of the mutated interaction(s) in the transition state. In the case of ACTR/NCBD, the mutations were all deletion of hydrophobic moieties. The slope of the plot was 0.18 suggesting that most native hydrophobic interactions have not formed in the conformation present at the top of the transition state barrier. However, the scatter in the data points suggests substantial local variation. These kinetic data can be combined with molecular dynamics simulations to obtain a clearer picture of the transition state. This combined approach suggested that the transition state does indeed retain high levels of disorder [28].

One recurring question in the IDP field is whether intrinsic disorder per se promotes protein association as predicted by a "fly-casting hypothesis" [29]. This is however not easy to test experimentally. Electrostatic interactions usually promote IDP interactions by increasing the association rate constant (Fig. 4b) masking any effects of the intrinsic disorder. Therefore, the salt dependence of several IDP interactions has been investigated to determine the basal association rate constant in absence of electrostatic interactions (i.e., at extrapolated infinite ionic strength) [17, 21, 30, 31]. This parameter has been compared between different protein–protein interactions to assess the role of intrinsic disorder on protein association [30]. While this might be the best experimental way to address the fly-casting hypothesis, the conclusions so far are not clear because of the limited number of experimental examples in the literature.

Conclusions: *Variation in buffer conditions and mutagenesis in conjunction with detailed kinetic experiments provide clues to binding mechanisms and the transition state of the interaction for IDPs.*

4 Notes

1. There is some confusion about what stopped-flow "dead time" represents and we prefer the use of the term "mixing time" which is the time from the true "time zero" of mixing (which may be in negative time relative to the triggering of data acquisition) and the first reliable data points in the kinetic traces recorded. Time zero and mixing time can be determined in the same simple experiment using a chemical reaction that is two state and gives clean single exponential kinetics with good signal to noise, for example, quenching of fluorescence of *N*-acetyl tryptophanamide (NATA) using *N*-bromosuccinimide

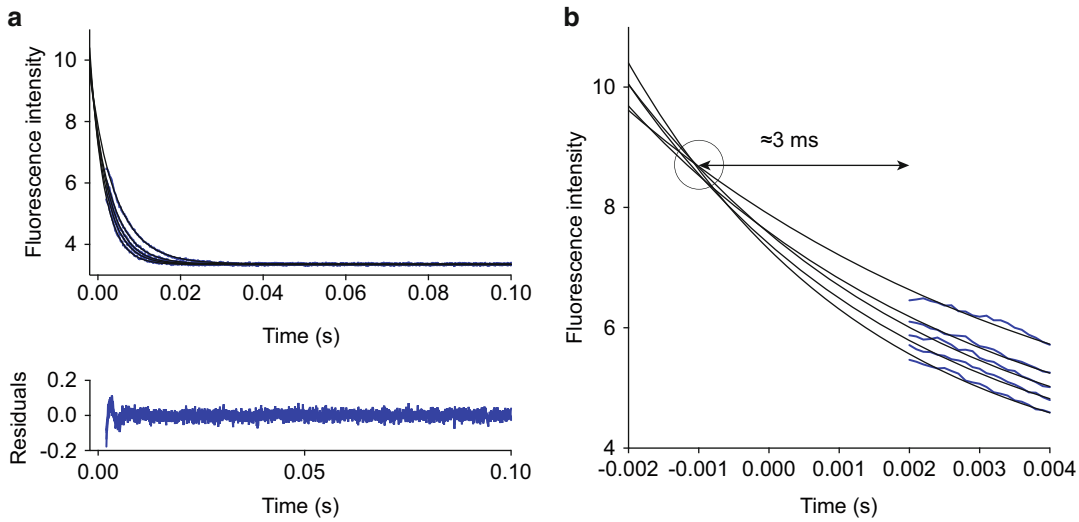


Fig. 5 Determination of mixing (dead) time in a stopped flow. (**a**) The reaction between NATA (1 μM) and NBS (500–900 μM) was measured by excitation at 280 nm and using a 320 nm long-pass emission filter. Each trace was fitted to a single exponential. (**b**) When experimental data points up to 2 ms were removed, the fitted curves intersected around −1 ms, which is the true time zero for the measured reaction. The first data points, which fall on the fitted curve are around 2 ms. Thus, the mixing time, or dead time, is 3 ms, which is the time between time zero and complete mixing

(NBS). The mixing time measurement should be performed under the same conditions of buffer and mixing ratio (1:1, 1:10, etc.) as used in the binding experiment. The reaction kinetics between NATA (0.5–1 μM) and an excess NBS are measured over a range of 5–6 different NBS concentrations that depends on the temperature of measurement but usually in the range of 100–2000 μM. The k_{obs} values measured should be between 200 and 1000 s^{-1}. The entire set of kinetic traces are then fit simultaneously to single exponential functions. Then, some early data points are removed systematically across the data set and the fit repeated iteratively until all the exponentials intersect at a common "time zero." It is very difficult to obtain data with sufficient quality that all traces indeed will intersect in a single point. But it is usually easy to determine a time region where all traces are close to each other. This is the true time zero with regard to the mixing of NATA and NBS under the experimental conditions used. The mixing time is the time from time zero to the time of the first good data point, i.e., where data were excluded/included to obtain the curve fit giving a good enough intersection of all single exponentials (Fig. 5).

2. Site-directed mutagenesis is a powerful method to help establishing mechanisms since it can pinpoint the significance of specific amino acid residues. In such cases, if the mutation is made in the non-varied protein, it is good practice to use the same stock solution for the varied species. Then, even if the absolute concentration (and thus k_{on}) is not correct, any error will cancel out when the effect of the mutation on free energy is assessed when comparing mutant and wild type: $\Delta\Delta G = RT \ln (k_{on}^{\text{wild-type}} / k_{on}^{\text{mutant}})$.
3. The fluorescence signal is adjusted with a voltage for the PMT. If the varied protein is fluorescent at the wavelength of detection then the overall signal will increase linearly as protein concentration is increased. There is then a risk that the detector output becomes saturated requiring the PMT voltage to be lowered. While the PMT voltage used during the experiment does not affect the k_{obs} values, it will change the fluorescence end points and amplitudes of the kinetic traces. Thus, whenever an amplitude analysis is included in a study, it is important to set the voltage such that the whole range of protein concentrations can be measured at a single PMT voltage. However, a PMT voltage of >300 V is desirable as below this the detector output can become nonlinear.
4. The rate constant for folding, k_F, is often not particularly sensitive to changes in temperature. However, unless k_F is measured directly, for example, using ligand trapping of the folded state [32], k_{obs} for folding ($=k_F + k_U$) will be temperature dependent due to the contribution of k_U even when measured in an experiment where the experiment (and signal change) monitors the folding reaction.
5. If nothing is known regarding the kinetics of a binding reaction, it is advisable to first perform equilibrium binding experiments with an appropriate technique to obtain the K_d and also to record fluorescence emission spectra of the respective protein(s) alone and in complex to help selection of an excitation wavelength and detection emission filter. If the K_d is in the single μM range or tighter, then excite 1 μM protein A at 280 nm and record fluorescence emission between 310 and 450 nm on a fluorimeter. Then record emission for 5 μM of protein B, if it contains any aromatic residues. Finally, take a spectrum of the complex 1 μM A and 5 μM B, and calculate the difference spectrum (AB − (A + B)). The difference spectrum will show at which wavelengths the fluorescence will change maximally during binding and assist selection of emission filter as well as indicating the relative increase or decrease in signal compared to the initial levels. Other affinities or fluorescent

probes will require appropriate concentrations of B to form the complex and the different excitation wavelength and emission range.

6. In the SX series stopped-flow instrument from Applied Photophysics, the tubing going from the left sample syringe to the observation cell is longer than the tubing from the right syringe. It is therefore better to have the non-varied protein in the left syringe and the protein whose concentration will be changed in the right syringe, since this will save on required volumes.
7. In some cases, it is advisable to degas solutions before the experiment; for example, if experiments are performed at higher than room temperature, there is an increased risk of formation of air bubbles during mixing. If solutions are prepared at room temperature and the experiments are performed at lower temperature then there is little risk for bubble formation. If pressure hold is used during kinetic acquisitions, this will also suppress outgassing (but *see* **Note 8**). In temperature jump experiments, solutions should always be degassed as the temperature perturbation is always an increase.
8. There is the option of "pressure hold" in stopped-flow instruments. If pressure hold is chosen, the pressure on the drive rams (pushing the sample syringes) is maintained during the whole acquisition. If pressure is not held on, then typically it will be released after 20–30 ms from triggering acquisition. Close inspection of this time region may reveal artifacts in the data due to the pressure release in the system. We therefore routinely use pressure hold for acquisition times up to 500 ms since the initial 50–100 ms of data is where the majority of signal change is occurring. However, pressure hold is not advisable for longer acquisitions (>1 s) as the repeated pressurization over longer time courses may eventually generate a leak in the fluidic system and premature wear in the valves.
9. Having a measurement of the equilibrium, K_d can help anticipate the range of k_{obs} values that may be expected. As a rule of thumb, if Trp fluorescence is used, preliminary experiments should be performed with 1 μM A and 10 μM B (final concentrations after mixing). Generally, k_{on} values are in the range 10^6–10^8 M^{-1} s^{-1}. Thus, if the measured K_d is 1 μM, k_{off} will be 1–100 s^{-1} since $K_d = k_{on}/k_{off}$. In this case, k_{obs} will then be 10 μM × k_{on} + k_{off} (Eq. 2), i.e., between 11 and 1100 s^{-1}. If no transient is observed and the fluorescence level is that measured for the AB complex, then the next step would be to reduce $[B]$ to 5 μM and/or reduce the

temperature of measurement to try and obtain a k_{obs} value that is within the accessible range (<400 s^{-1}). If brighter external fluorescent probes are employed then 50 nM of A and 500 nM of B may be a good starting point with the reduction in $[B]$ leading to a slower binding reaction and easier to measure values.

10. We recommend separate preliminary experiments to confirming the presence of different kinetic phases that may be over different time regimes. However, when the experimental protocol is optimized and the kinetics of a concentration series is recorded, it may be useful to acquire data with a logarithmic or split time base to distribute enough data points for each kinetic phase within the one experiment. Curve fitting of the kinetic transient may be biased depending on the nonuniform distribution of data points (linear, split, logarithmic) and residuals must therefore be carefully analyzed for systematic deviations with this in mind.
11. In general, certain salts promote "structuring" in proteins according to the Hofmeister series [33]. Sodium sulfate is particularly useful for stabilizing weakly structured intermediates that often contain only a collapsed hydrophobic core (for example, in protein folding). Likewise, trimethylamine *N*-oxide (TMAO) promotes tertiary structure in proteins whereas 2,2,2-trifluoroethanol (TFE) is more specific for stabilizing hydrogen bonds and thereby stabilizes helical elements. On the other hand, denaturants such as urea or guanidinium chloride can be used to destabilize protein structure and favor unfolded states. The additives glycerol and sucrose are sometimes used to increase solution viscosity. This is a common way to investigate association reactions and determine if they are diffusion limited since molecular encounters will slow with increased viscosity. However, their use is complicated because the increased viscosity will affect the instrument mixing time (important for fast reactions) and they may also affect the stability of folded conformations.
12. The concentration of AB is dependent on K_d for the interaction and the concentrations of A and B. In general, low concentrations of A and B and a high relative concentration of AB complex are preferable in a displacement experiment. The limiting factors are the K_d value and the difference in fluorescence between AB and AC. As a useful starting point, if $K_d = 100$ nM, the concentrations of A and B should be 0.1–1 μM (resulting in 38–73% complex if $[A]$ is equal to $[B]$). If $K_d = 1$ μM, then the concentrations of A and B should be between 1 and 10 μM, etc. With lower concentrations of B, less C is needed to compete out B, but there will be more free A, which will give an additional kinetic binding phase not

related to the displacement. The latter phase is avoided if $[B] \gg K_d$, but a higher concentration of C is then necessary to make the AB dissociation irreversible. To reduce the concentration of free A, it could be useful to have a small excess of B over A, e.g., 1 μM A and 2 μM B (resulting in 59% AB complex if $K_d = 1$ μM).

13. An attractive feature of displacement experiment is that any errors in the concentrations of A, B, or C do not matter, since there is no concentration-dependence of k_{obs} at high enough concentration of C. Measurement of the overall dissociation rate constant determined in a displacement experiment at this limit is therefore one of the more accurate kinetic parameters that can be obtained.
14. Fast conformational changes occurring before binding are difficult to distinguish from an induced fit model. However, when a slow conformational change precedes binding, a very characteristic decrease of k_{obs} occurs upon increasing protein/ligand concentration. In such cases, it is straightforward to rule out induced fit and favor conformational selection. Experimentally, this will appear exactly as a displacement experiment where the dissociation of the protein–protein complex corresponds to the intramolecular conformational change. Considering Scheme 3, in the rare perfectly intermediate case between "fast" and "slow" pre-equilibrium where $k_1 = k_{-1} = k_2$, the slow phase will have a constant value (the same as the microscopic rate constants) at all concentrations of B.
15. A triangular mechanism, in which binding can occur via either an intermediate or directly to the bound complex, will also give double exponential kinetics since it is still a three-state mechanism where the third step is not independent of the other two. Likewise, a square mechanism, which is obtained if Schemes 2 and 3 are combined, will theoretically give triple exponential kinetics.

References

1. Wright PE, Dyson HJ (2015) Intrinsically disordered proteins in cellular signalling and regulation. Nat Rev Mol Cell Biol 16:18–29. https://doi.org/10.1038/nrm3920
2. Fuxreiter M (2019) Fold or not to fold upon binding - does it really matter? Curr Opin Struct Biol 54:19
3. Ivarsson Y, Jemth P (2018) Affinity and specificity of motif-based protein-protein interactions. Curr Opin Struct Biol 54:26–33. https://doi.org/10.1016/j.sbi.2018.09.009
4. Dogan J, Gianni S, Jemth P (2014) The binding mechanisms of intrinsically disordered proteins. Phys Chem Chem Phys 16:6323. https://doi.org/10.1039/c3cp54226b
5. Sugase K, Dyson HJ, Wright PE (2007) Mechanism of coupled folding and binding of an intrinsically disordered protein. Nature 447:1021–1025. https://doi.org/10.1038/nature05858
6. Dogan J, Jemth P (2014) Only kinetics can prove conformational selection. Biophys J 107:1997–1998. https://doi.org/10.1016/j.bpj.2014.08.037
7. Schneider R, Blackledge M, Jensen MR (2018) Elucidating binding mechanisms and dynamics

of intrinsically disordered protein complexes using NMR spectroscopy. Curr Opin Struct Biol 54:10–18. https://doi.org/10.1016/j.sbi.2018.09.007

8. Williams MA (2013) Protein–ligand interactions: fundamentals. In: Williams MA, Daviter T (eds) Protein-ligand interactions. Humana Press, Totowa, NJ, pp 3–34. https://doi.org/10.1007/978-1-62703-398-5_1.
9. Martin SR, Schilstra MJ (2013) Rapid mixing kinetic techniques. Methods Mol Biol 1008:119–138. https://doi.org/10.1007/978-1-62703-398-5_5
10. Bagshaw CR (2017) Biomolecular kinetics: a step-by-step guide. CRC Press, Taylor & Francis Group, Boca Raton, FL
11. Gianni S, Jemth P (2017) How fast is protein-ligand association? Trends Biochem Sci 42:847–849. https://doi.org/10.1016/j.tibs.2017.08.007
12. Johnson CM (n.d.) Isothermal titration calorimetry
13. Chemes LB, Sánchez IE, de Prat-Gay G (2011) Kinetic recognition of the retinoblastoma tumor suppressor by a specific protein target. J Mol Biol 412:267–284. https://doi.org/10.1016/j.jmb.2011.07.015
14. Li-Blatter X, Zweifel L, Sharpe T (2020) A familiar protein-ligand interaction revisited with multiple methods. In: Protein-ligand interactions. Wiley-VCH, Weinheim
15. Dosnon M, Bonetti D, Morrone A, Erales J, di Silvio E, Longhi S, Gianni S (2015) Demonstration of a folding after binding mechanism in the recognition between the measles virus NTAIL and X domains. ACS Chem Biol 10:795–802. https://doi.org/10.1021/cb5008579
16. Dogan J, Schmidt T, Mu X, Engström Å, Jemth P (2012) Fast association and slow transitions in the interaction between two intrinsically disordered protein domains. J Biol Chem 287:34316–34324. https://doi.org/10.1074/jbc.M112.399436
17. Ou L, Matthews M, Pang X, Zhou H-X (2017) The dock-and-coalesce mechanism for the association of a WASP disordered region with the Cdc42 GTPase. FEBS J 284:3381–3391. https://doi.org/10.1111/febs.14197
18. Kuzmic P (2009) DynaFit--a software package for enzymology. Methods Enzymol 467:247–280. https://doi.org/10.1016/S0076-6879(09)67010-5
19. Johnson KA, Simpson ZB, Blom T (2009) Global kinetic explorer: a new computer program for dynamic simulation and fitting of kinetic data. Anal Biochem 387:20–29. https://doi.org/10.1016/j.ab.2008.12.024
20. Malatesta F (2005) The study of bimolecular reactions under non-pseudo-first order conditions. Biophys Chem 116:251–256. https://doi.org/10.1016/j.bpc.2005.04.006
21. Åberg E, Karlsson OA, Andersson E, Jemth P (2018) Binding kinetics of the intrinsically disordered p53 family transactivation domains and MDM2. J Phys Chem B 122:6899–6905. https://doi.org/10.1021/acs.jpcb.8b03876
22. Olson ST, Srinivasan KR, Björk I, Shore JD (1981) Binding of high affinity heparin to antithrombin III. Stopped flow kinetic studies of the binding interaction. J Biol Chem 256:11073–11079
23. Gianni S, Dogan J, Jemth P (2014) Distinguishing induced fit from conformational selection. Biophys Chem 189C:33–39. https://doi.org/10.1016/j.bpc.2014.03.003
24. Dogan J, Mu X, Engström Å, Jemth P (2013) The transition state structure for coupled binding and folding of disordered protein domains. Sci Rep 3:2076. https://doi.org/10.1038/srep02076
25. Giri R, Morrone A, Toto A, Brunori M, Gianni S (2013) Structure of the transition state for the binding of c-Myb and KIX highlights an unexpected order for a disordered system. Proc Natl Acad Sci U S A 110:14942–14947. https://doi.org/10.1073/pnas.1307337110
26. Rogers JM, Oleinikovas V, Shammas SL, Wong CT, De Sancho D, Baker CM, Clarke J (2014) Interplay between partner and ligand facilitates the folding and binding of an intrinsically disordered protein. Proc Natl Acad Sci 111:15420–15425. https://doi.org/10.1073/pnas.1409122111
27. Dogan J, Jonasson J, Andersson E, Jemth P (2015) Binding rate constants reveal distinct features of disordered protein domains. Biochemistry 54:4741–4750. https://doi.org/10.1021/acs.biochem.5b00520
28. Karlsson E, Andersson E, Dogan J, Gianni S, Jemth P, Camilloni C (2019) A structurally heterogeneous transition state underlies coupled binding and folding of disordered proteins. J Biol Chem 294:1230. https://doi.org/10.1074/jbc.RA118.005854
29. Shoemaker BA, Portman JJ, Wolynes PG (2000) Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. Proc Natl Acad Sci U S A

97:8868–8873. https://doi.org/10.1073/pnas.160259697

30. Shammas SL, Travis AJ, Clarke J (2013) Remarkably fast coupled folding and binding of the intrinsically disordered transactivation domain of cMyb to CBP KIX. J Phys Chem B 117:13346–13356. https://doi.org/10.1021/jp404267e
31. Lindström I, Dogan J (2017) Native hydrophobic binding interactions at the transition state for association between the TAZ1 domain of CBP and the disordered TAD-STAT2 are not a requirement. Biochemistry 56:4145–4153. https://doi.org/10.1021/acs.biochem.7b00428
32. Ivarsson Y, Travaglini-Allocatelli C, Jemth P, Malatesta F, Brunori M, Gianni S (2007) An on-pathway intermediate in the folding of a PDZ domain. J Biol Chem 282:8568–8572. https://doi.org/10.1074/jbc.M611026200
33. Okur HI, Hladílková J, Rembert KB, Cho Y, Heyda J, Dzubiella J, Cremer PS, Jungwirth P (2017) Beyond the hofmeister series: ion-specific effects on proteins and their biological functions. J Phys Chem B 121:1997–2014. https://doi.org/10.1021/acs.jpcb.6b10797

Chapter 5

Isothermal Titration Calorimetry

Christopher M. Johnson

Abstract

Calorimetry is a classical biophysical method that by definition measures heat. In isothermal titration calorimetry (ITC), the heat is the result of titrating interacting components together and allows direct determination of the thermodynamics for this process. The measured heat reflects the enthalpy change (ΔH), and the prospect of determining this in biological systems where high-resolution structural information is available has led to the possibility of rational thermodynamics-guided design of ligands. Although there are limitations to this approach due to the participation of solvent in the thermodynamics, ITC has become an established technique in many labs providing a valuable tool with which to quantify protein–-protein interactions. With careful use, ITC can also provide additional insights into the binding process or be used in increasingly complex systems and where interaction is coupled to other molecular events.

Key words Isothermal titration calorimetry, ITC, Thermodynamics, Enthalpy, Entropy, Free energy, Dissociation constant, Heat capacity, Stoichiometry, Binding affinity, Binding kinetics

1 Introduction

1.1 ITC: A Measurement Nirvana?

Many techniques for studying protein–ligand interactions discussed elsewhere in this volume depend on reporter signals, such as fluorescence or absorbance, which change as a result of complex formation. Sometimes these changes can be a rather indirect consequence of the binding event. Other techniques, such as fluorescence polarization, light scattering, or surface-based sensors using SPR or BLI, report on a change in physical properties such as mass or size during complex formation. Changes in these properties that produce adequate signal-to-noise measurements can often be large when compared to the changes expected for protein–ligand interactions.

In contrast, calorimetry simply measures directly the heat associated with making and breaking interactions that are intrinsic to complex formation irrespective of any change in size or mass and in the absence of any additional reporter labels required to give suitable spectroscopic properties. It is this "directness" of working with unmodified materials and the ubiquitous nature of the predominantly non-covalent forces, with their associated heats, that drive

Tina Daviter et al. (eds.), *Protein-Ligand Interactions: Methods and Applications*, Methods in Molecular Biology, vol. 2263, https://doi.org/10.1007/978-1-0716-1197-5_5,

protein–ligand interactions which has led to biocalorimetry, and ITC in particular, being described as the "gold standard" label-free technique for their characterization. Furthermore, under constant pressure (atmospheric), the measured heat of a protein–ligand interaction equates to the enthalpy change (ΔH) of the process that gives ITC an additional unique and direct insight into the fundamental underlying thermodynamics.

In the simplest scenario where protein (A) and ligand (B) interact forming a 1:1 complex then:

$$\mathrm{A} + \mathrm{B} \underset{k_{\mathrm{off}}}{\overset{k_{\mathrm{on}}}{\rightleftharpoons}} \mathrm{AB} \tag{1}$$

$$k_{\mathrm{obs}} = [\mathrm{A}][\mathrm{B}]k_{\mathrm{on}} + [\mathrm{AB}]k_{\mathrm{off}} \tag{2}$$

$$K_{\mathrm{a}} = \frac{[\mathrm{AB}]}{[\mathrm{A}] \cdot [\mathrm{B}]} = \frac{1}{K_{\mathrm{d}}} = \frac{k_{\mathrm{on}}}{k_{\mathrm{off}}} \tag{3}$$

$$\Delta G = -RT \ln K_{\mathrm{a}} = \Delta H - T\Delta S \tag{4}$$

The binding equilibrium changes when perturbed with observed kinetics (k_{obs}) that are the sum of the on and off rate constants for binding. K_{a} and K_{d} are the equilibrium association and dissociation constants reflecting the "strength" of the interaction (the equilibrium bias toward complex formation rather than being unbound).

ITC can in principal yield ΔH, the total heat in going from A + B to AB, as well as K_{a} ($=1/K_{\mathrm{d}}$) and the stoichiometry "n" of the interaction ($n = 1$ for 1:1 binding) by measuring the extent of complex formation when varying the ligand concentration during a titration. The free energy of binding (ΔG) is obtained from K_{a} and the entropy of the process (ΔS) is obtained from ΔG and ΔH by subtraction, giving a complete thermodynamic description of the event.

Thus, ITC has the potential to give additional information about the magnitude and the balance of thermodynamic forces that are driving complex formation as well as quantifying the overall "strength" of the interaction reflected in its K_{a}. The tantalizing prospect of this unique window into underlying thermodynamics further enhances the status of ITC as the biophysical method of choice for protein–ligand interactions.

1.2 Why Read This Chapter?

ITC instruments with sufficient sensitivity to measure protein–ligand interactions are available commercially from different manufacturers. These are supplied with comprehensive documentation as hard copy or online. The manufacturer's installation process should also include appropriate levels of hands-on training covering the basic operation and experiments using robust test chemistry (e.g., the chelation of metal ions by EDTA). If ITC is already an

established technique in the lab but there is no recognized expert available to give instruction, then the instrument manual and the instructions below are a good starting point.

This chapter attempts to establish good basic experimental practice and then to build on this initial platform to hopefully confer additional experience gained over the many years of ITC use in many different biological settings. It considers the real practical aspects of ITC that are encountered when working with proteins and details test chemistry that uses a protein–ligand interaction that has been verified using other techniques that might be available for comparison (*see* Chapter 2). The numerous advantages and unique insights given by ITC are considered but not forgetting also to discuss problem areas and practical limitations of the technique. Data processing, fitting issues, and data accuracy are also covered so that the results obtained from ITC can be set in a reliable framework and used in a realistic way.

2 Materials

2.1 ITC Instrumental Basics

The MicroCal iTC200 instrument has two metal cells, sample and reference, that are located in the core of the instrument in an insulated adiabatic environment (*see* **Note 1**). The cells cannot be seen or removed and are accessed through filling tubes that connect to the surface of the instrument as illustrated in Fig. 1.

The sample cell has a syringe titration system that can be introduced via the filling tube. The syringe has a paddle at its end (either twisted, as shown in Fig. 1, or flat) and is rotated at 750 or 1000 rpm to mix the cell contents very efficiently. It can make small μL volume injections of ligand using a stepper motor that drives the plunger (*see* **Note 2**). The reference cell is typically filled with water (*see* **Note 3**).

The instrument maintains a small difference in temperature between these two cells using a number of electrical heaters attached to their surface. The heaters are driven in a feedback loop from the voltage output of a very precise thermopile that is measuring the temperature difference as it is arranged between the two cells. It is then levels of excess differential power applied to the sample cell that become the instrument output varying during a titration as ligand is injected, binding occurs, and heat is produced or absorbed. By maintaining a positive level of differential power to the sample cell during measurement, the instrument can increase further this energy input, in the case of heat-absorbing endothermic (ΔH +ve) events on addition of ligand by the syringe, or it can reduce the energy input in the event of heat-producing exothermic (ΔH −ve) events (*see* **Note 4**). The instrument thus has an upper limit of measurement, where the differential power circuit is

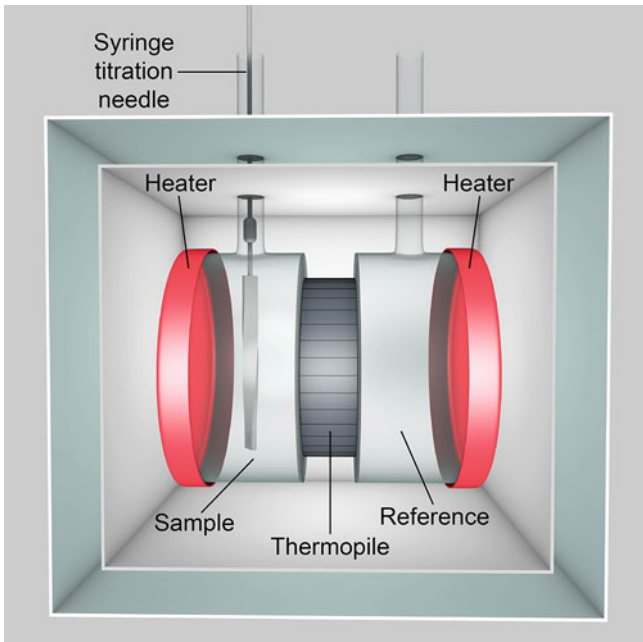


Fig. 1 Schematic of iTC200 instrument. The solution that is displaced from the cell by the serial injections of the experiment is thought to be prevented from mixing back with the active measurement volume in the cell by a small Teflon plug on the syringe needle that sits snugly in the neck of the cell (*see* **Note 6**)

applying its maximum electrical heating and a lower limit where differential power reaches zero; beyond which the instrument would require "cooling" at the sample cell.

Sequential injections of ligand from the syringe into the sample cell produce endothermic or exothermic heat effects in the cell and the instrument "compensates" by adjusting the differential power via the feedback circuit for the duration of the effect. This maintains the difference in temperature between sample and reference cell at its fixed value. The differential power vs. time trace is characterized by these peaks where the signal deflects from, and then returns to, a baseline value as the reaction equilibrium is perturbed and then reestablished. These "peaks" in differential power can be integrated over time using baselines that are interpolated from levels before and after each injection (*see* **Note 5**). It is this integral over time (units of heat) normalized by the amount of injected ligand (units of molar heat) that reflects heat of binding and the extent of the protein–ligand complex formation once changes in the protein and ligand concentration are accounted for and any corrections for nonbinding contributions are made (*see* **Note 6**).

Typical experimental parameters involve using 15–20 equal injections from the syringe. These are preceded by a smaller "pre-injection" that eliminates material at the end of the needle that may

have mixed with the cell component by diffusion during the instrument's equilibration phases after loading, but before measurement. The data from this injection are not normally included during analysis (*see* **Note 7**).

The injections are spaced at intervals of 2–4 min, depending on the instrument, its settings, and the duration of the heat effect that is determined by the kinetics of binding (*see* Eq. 2). Initially the interval is better left longer so all the heat can be captured by the instrument and a stable baseline reestablished before any subsequent injections. Once initial binding is characterized, this interval and the number of injections may be reduced if appropriate.

2.2 A Realistic Test Reaction: Lysozyme Binding a Simple Trisaccharide Ligand

1. Binding of tri-acetyl glucosamine (*N*,*N*′,*N*″-Triacetylchitotriose) to hen egg-white lysozyme is a useful test reaction to gain experience in performing ITC measurements. Both components are available as lyophilized materials of high purity and at modest cost. The binding K_d is single μM level and the interaction has been quantified by ITC and by a variety of orthogonal binding assays with comparable results (*see* Chapter 2).
2. Both should be carefully weighed using an analytical balance and dissolved in an experimental buffer at pH 5 or lower since this avoids the possibility of lysozyme forming higher order species such as dimers [1]. 0.1 M acetate buffer pH 5, is a suitable choice as it has been widely used in the literature for this interaction. It also has an advantage of a near zero enthalpy of ionization and will thus maintain its pH at different temperatures if required (*see* **Note 8**). The buffer can be prepared using an online recipe calculator (https://www.liverpool.ac.uk/pfg/Research/Tools/BuffferCalc/Buffer.html) and the appropriate molarities of basic and acidic components can simply be mixed in the calculated quantities (using weight or volume). The pH should be 5 if checked with a calibrated pH meter. The mixing approach is highly reproducible and avoids the use of a pH meter that is normally a communal resource in a lab and requires careful maintenance and calibration for reliable performance.
3. Measuring the concentration of the materials used is key to obtaining reliable and quantitative ITC data as discussed in more detail below. Most proteins contain aromatic amino acids that contribute absorbance at 280 nm giving a quick and accurate method of determining concentration. Molar extinction coefficients can be obtained from the literature or using the protein sequence and the online program ProtParam (https://web.expasy.org/protparam/) (*see* **Note 9**). For hen egg-white lysozyme this value is 37,970 M^{-1} cm^{-1}. Unfortunately, there is not an easy way of checking the concentration of tri-acetyl glucosamine stock solutions and so these must be

based on weight and the volume of buffer used to dissolve (Mr 628 Da).

4. Measuring protein concentration using absorbance requires a careful and methodical approach. Instruments must be blanked with appropriate solvent, and the measured absorbance should be within a reliable absorbance range and be corrected for any scattering contributions determined by measuring the full spectrum including longer wavelength data where the protein should not absorb (*see* **Note 10**). A suitable lysozyme solution for this test reaction is ~40 μM (having absorbance of ~1.5 cm^{-1} that can be measured in a 1 cm cuvette or at 1 mm which is common for micro volume measurement such as on a Nanodrop). For the iTC200 instrument and using the supplied loading syringe, a volume of ~350 μL is required.
5. The lysozyme should be loaded into the ITC cell using the syringe supplied with the instrument and following protocols recommended by the manufacturer to produce a total fill without any trapped air. It is also common that users develop their own variants of standard loading protocols and these "local practices" are also useful if they have a proven track record. The loading process is the tricky part of performing an ITC experiment and so it is worth practicing this with test reactions (*see* Subheading 3.1, **step 3** to evaluate success). There is an established history that degassing the solutions should help with loading but there is the risk that using low pressure can lead to evaporation and changes in concentration. Worse still is the possibility of denaturation of protein samples at the solution surface if stirring is used. So degassing should be done with care and the concentrations and integrity of material checked afterwards. What seems more useful in loading is to ensure that the ITC cell is spotlessly clean which is best achieved by storing the instrument with dilute laboratory detergent in the cells when not in use. This approach can be combined with more vigorous treatments such as incubating the cells at 50 °C overnight in detergent or the use of strong acid or alkali solutions (taking great care to comply with the detailed chemical compatibility of the cell construction that will be supplied by the manufacturer).
6. The tri-acetyl glucosamine should be prepared and loaded into the syringe as instructed in the instrument manual at ~20-fold higher concentration than the lysozyme (~800 μM). For the iTC200, this requires ~70 μL.
7. Both of the required volumes are larger than the active cell volume and syringe capacity (~200 μL and 40 μL, respectively for the iTC200) and some solution is left after loading. These small aliquots can be used to check concentration and sample integrity (at least for the lysozyme) since they most closely resemble the true experimental solutions being used.

3 Methods

3.1 Running the Test Reaction

1. The instrument should be programmed to execute 15–20 injections, including a small pre-injection (*see* **Note 7**), in volumes that use fully the 40 μL capacity of the syringe. In this way, the final concentration of tri-acetyl glucosamine in the iTC200 will approach ~130 μM representing a considerable molar excess over the lysozyme. Injections are spaced by 180 s. The melting temperature of lysozyme at pH 5 in this buffer is ~75 °C and so the test can be performed over a wide range of temperatures.
2. The differential power level to the sample cell should be set to the middle of the instrument's measurement range, which for the iTC200 is 6 μcal/s (*see* **Note 11**). The ability to control this level is important for experiments generating large heats, but for biological systems at typical concentrations the signals are typically more modest and so this setting will reliably capture exothermic or endothermic interactions. There is also an option to control the data density (time interval for averaging differential power into a point) and the instrumental "feedback gain" that sets levels of electrical filtering on the measured signal. It is best practice to set intervals of 1–2 s and feedback gain at its fastest level, both producing more "noisy" data. This allows use of the data in kinetic analysis, either for simple visual inspection or the more detailed analysis discussed below. Indeed it is true of any time-based measurement that inappropriate choice of data interval and excessive electrical damping (slow instrument response) during collection will mask kinetic information that cannot then be recovered. In contrast, the cosmetic effects of reducing noise can easily be obtained post data collection with the use of software.
3. Once the experiment is running, the differential power level is a good indication of how successful the loading has been. The various power levels of the instrument, including that of the final feedback heaters, are configured during instrument construction with both cells containing aqueous solutions and are therefore a measure of the heat capacity of the sample cell. If the background baseline differential power levels are lower than configured in the experiment (>1 μcal/s off) then this suggests that the sample cell may contain one or more micro bubbles; the power to this cell is reduced because of the very low heat capacity of air compared to aqueous solution. Equally, if the power levels are higher than expected then this can be indicative of poor loading of the reference cell.

3.2 Data Inspection and Fitting

Typical ITC raw data at 25 °C from the lysozyme–tri acetyl glucosamine test and the integrated excess power from each injection (excluding the pre-injection) plotted against the molar ratio of ligand:protein during the titration are shown in Fig. 2 in the "final figure" format that is commonly used in the ITC literature.

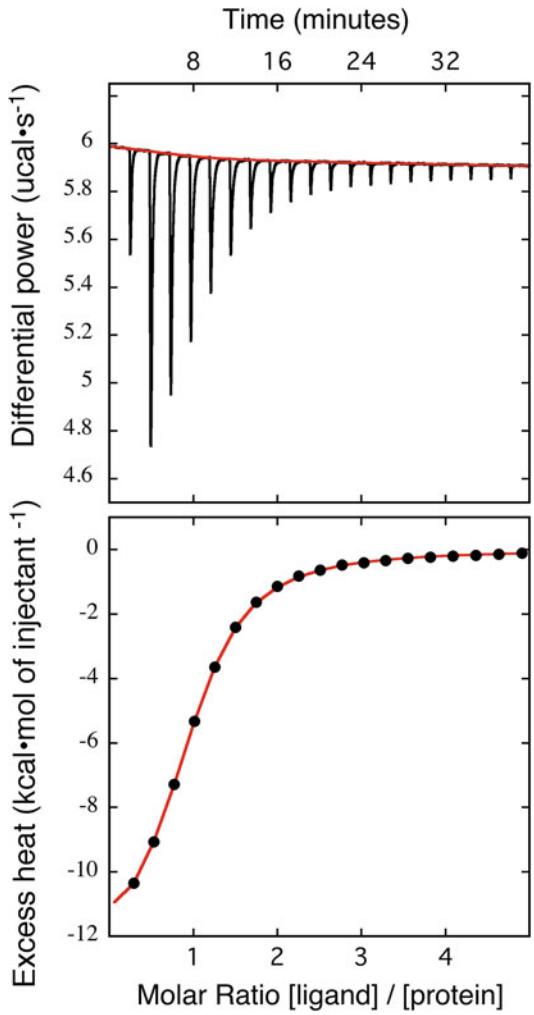


Fig. 2 Raw ITC data for lysozyme–tri-acetyl glucosamine test is shown with baseline generated for integrating the injection peaks shown in red (upper panel). The integrated heats of each injection have been corrected with the observed heats from an identical control experiment where the same stock tri-acetyl glucosamine solution was injected into buffer. The corrected integrals are shown with the fit to a simple binding model in red (lower panel). Note the first pre-injection of 0.5 μL produces a smaller heat. The concentration change of this injection is included but the integral is not plotted or included in the fit (*see* **Note 7**)

The experiment was run at 6 μcal/s and the raw data show a very small decrease in this baseline level. It is recommended that the raw data is presented in this way so that any drift or acute changes (steps) in the course of the experiment can be seen. In some software, the generation of the baselines required for the integration of the peaks gives the opportunity to set all the baseline at a "zero level" making the whole data collection look perfectly flat at zero differential power. This type of representation should be avoided.

These data have been fit using Malvern's PEAQ software. This program automatically integrates the excess heat effect from each injection and normalizes this to the number of moles of ligand added during each injection giving molar heats in kcal/mol. The integration was checked manually for each peak and baselines adjusted if there was an obvious inconsistency. The integrals are plotted against the molar ratio of [ligand]/[protein] at the end of each injection and the resulting binding isotherm can be fit to a K_d value of 6 μM, an enthalpy of −13 kcal/mol with stoichiometry of 0.92. The small heats at the end of the titration (high molar ratio) represent the background control heat arising at each injection when the protein is fully saturated with ligand and can be included as a fitted parameter. Preferably, the end point can be confirmed by measuring separately in a control experiment where the identical trisaccharide solution used is injected into buffer alone (*see* **Note 12**). The heats determined in this control experiment can be subtracted from the original binding integrals, as has been done here, and the corrected data fit with a zero end point (*see* Fig. 2 lower). Where the experimental data do not reach saturation of all binding sites on the protein, it becomes more important to perform such a control since the fitted end point is being extrapolated to values that are not measured directly in the binding experiment.

3.3 Are the Concentrations of Protein and Ligand Optimal?

The sigmoidal shape of the ITC binding titration can be defined by a unit-less parameter, the "*c*-value," which expresses how far the protein concentration in the cell is above or below the K_d value ($c = [\text{protein}]/K_d$, in Fig. 2 where [lysozyme] was 37 μM $c \sim 6$). It is possible to simulate data with the same K_d over a variety of experimental *c*-values as shown in Fig. 3.

Experiments performed at high *c*-value >250 (cell concentration far above K_d) essentially generate a step function plot. All the added ligand becomes bound until all sites on the protein are filled at a molar ratio of 1, and then no binding occurs and the control end point heat is observed. In this case, the enthalpy and stoichiometry of the interaction are very well defined but the K_d cannot be quantified other than concluding it is significantly below the cell concentration. Working in this *c*-value regime is known as a stoichiometric titration. At intermediate *c*-values between 5 and 250, the plots are more sigmoidal and K_d can be determined with

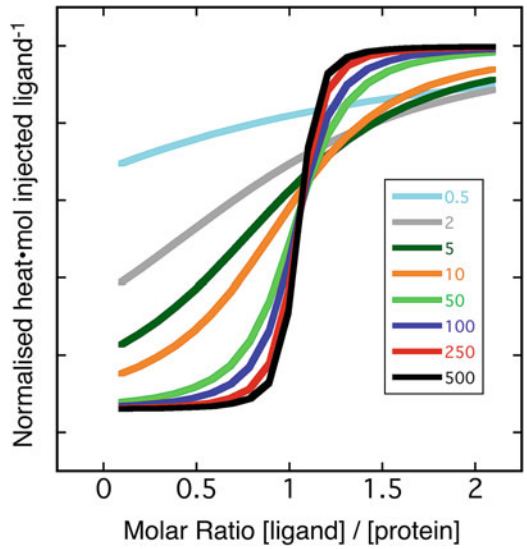


Fig. 3 Effect of different *c*-values on the shape of the integrated ITC binding curves

some confidence since only a fraction of added ligand binds with each injection and this fraction changes as the total concentration of added ligand increases. In this intermediate *c*-value regime, it is possible to determine ΔH, n, and K_d with some confidence as in Fig. 2.

At low *c*-values <5, the plots become rather featureless and at the limit become a straight line. Just as ΔH and n are well defined at high *c*-value, they now become poorly constrained at low *c*-value. These plots have many combinations of ΔH and n that can describe the data equally well (1 site with ΔH, 2 sites with $\Delta H/2$, etc.). In this case, the data can be better fit if either parameter is fixed during fitting, either from the results of a separate stoichiometric titration or by assuming binding of $n = 1$ [2].

Thus ITC experiments are optimally performed in an experimental *c*-value window of roughly 5–250 in order to obtain a full description of the interaction. These experiments have three regions; early injections have the most heat and help define the enthalpy, intermediate injections constrain the stoichiometry and K_d value while the last injections have least heat and indicate the background control heats or end point. In weak binding interactions, the *c*-value will tend to be low. However, it may not be possible to optimize this because of limitations on availability of materials, their solubility, and magnitude of the heat signal which may be larger than can be measured (e.g., K_d 1 mM would require 10 mM in the cell for *c*-value of 10 and thus ~200 mM of ligand in the syringe). There are different problems for tight binding interactions where the *c*-value tends to be high. Obtaining an optimal

c-value could be achieved by lowering the cell protein concentration. However, the total amount of ligand that can be bound and therefore the total amount of heat that can be released or absorbed during the titration is determined by this cell concentration. Therefore, there is a lower limit to the protein concentration defined by both the smallest amount of heat detectable in a single injection and the smallest number of injections required to define a binding curve. As a consequence, ITC is currently able to provide optimal characterization of binding interactions using a single standard experiment where the K_d is between 10's of nM and 100's of μM.

3.4 How Do the Concentrations Influence the Results?

Determining the concentration of protein and ligand in an ITC experiment is essential for accurate quantitation of the binding. UV–VIS absorbance spectroscopy is the easiest option (*see* **Notes 9** and **10**) while amino acid analysis can be used for proteins or peptides lacking aromatic residues. These methods should be able to reproducibly quantify concentration within a few percent. Weighing lyophilized solid and/or colorimetric methods are also options but are potentially much less accurate.

The effects of concentration errors on ITC data have been examined in detail elsewhere [3, 4] but as an exercise in underlining its importance it is informative to refit the test reaction data using deliberately inaccurate values mimicking a ±20% error in concentration as seen in Fig. 4.

It is evident that errors in the syringe concentration translate into equivalent percentage errors in all of the directly fitted parameters (although not in ΔG because of the logarithmic relation to K_d). Interestingly, the cell concentration only affects the value of stoichiometry n, while K_d and ΔH overlay and are unaltered. Since the concentration of material used in the syringe for ITC titrations is often quite high it may not be possible to measure directly its UV absorbance. In this case a dilution should be prepared so that absorbance is in a good range for measurement and this dilution should be repeated a few times for independent measurement and averaging.

Since both the cell and the syringe concentrations affect the determination of stoichiometry, this value will potentially have an error of at least 5% or more from the combined measurements. These errors can compensate if one is an overestimate while the other underestimates concentration, but stoichiometries within the generous range 0.9–1.1 should probably be considered as consistent with standard 1:1 binding with $n = 1$. Reflecting this it is also possible to fit ITC data with a fixed stoichiometry of binding of 1 but with the concentration of either cell or syringe component as a variable.

There can also be errors in concentration that are not the result of inaccuracy of measurement. For example, the "active" or "binding competent" concentration of a sample can be lower than the

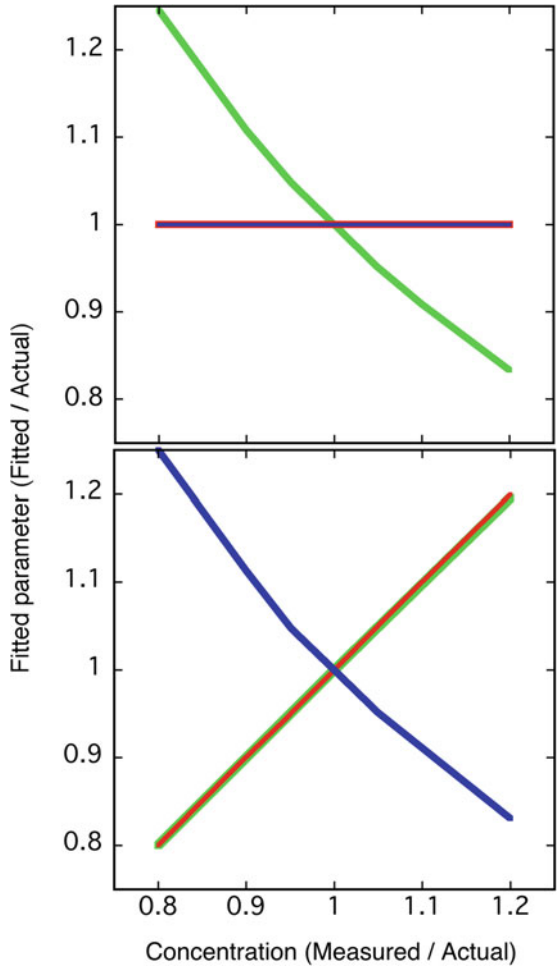


Fig. 4 Effect of errors in the measured concentration of lysozyme in the cell (upper panel) or tri-acetyl glucosamine in the syringe (lower panel) on the fitted K_d (red), ΔH (blue), and stoichiometry n (green). Fitted values are plotted as a fraction of the values determined in Fig. 2

measured total protein concentration as a result of inactive/modified/aggregated fractions, which do not bind the ligand or bind with much weaker affinity. Similarly, the presence of contaminant proteins that have been copurified along with the system of interest will reduce the active concentration. These could be unrelated contaminants or could be the binding partner of interest or a related ligand.

3.5 Occam's Razor and the Use of Advanced Data Fitting

Fitting ITC data using the software provided with instruments is very quick and simple compared to the measurement itself. These fitting programs also allow more complex models of binding to be considered (multiple sites, cooperativity, etc.) but these should only be employed if there is good evidence from other techniques that

the binding may be more complex or if multiple ITC experiments reproduce systematic deviations from a simpler model. Some programs also report fitted values of K_d, ΔH, and n with unrealistic levels of precision that are based purely on the NLLS curve fitting exercise. There are many other sources of error that contribute to uncertainty in fitted values from individual experiments (discussed above) and so fitted values and errors should be quoted with a more generous margin than indicated by the fit or preferably determined rigorously by repeats of the experiment using freshly prepared materials and based on new concentration determinations.

ITC data can be exported from the fitting programs (*see* **Note 6**) and used to generate a cumulative sum of total heat produced during the titration for plotting in alternate graphing packages. This can be fit with binding models such as one would use for changes in a spectroscopic property and graphed to generate more familiar looking binding hyperbola. Processing data in this way was a requirement of employing more complex binding models in the early era of ITC experimentation, although the software supplied with instruments has improved greatly in recent years. Thus in parallel with the development of ITC as a technique, a number of alternate fitting packages have been developed as academic and commercial exercises. These provide increasingly complex binding models, global fitting multiple experiments to common binding parameters and even fitting the shape of individual ITC injection profiles to gain kinetic information [5–10]. Reviewing these options is beyond the scope of this practical introduction. However, following Occam's principle, it is not advisable to introduce additional complexity when fitting a single data set unless supported by other experimental evidence. If reliable and robust ITC data that does not fit a simple model are obtained as a matter of routine, then global fitting can help to constrain the additional binding parameters that are required (*see* Chapter 2).

3.6 Titrations Either Way Round: Varying Ligand or Protein

The choice in an ITC binding experiment of which component to load into the cell and which to titrate from the syringe may be dictated by considerations such as the availability of materials or their solubility. A typical starting concentration in a small cell volume instrument such as the iTC200 is 20–40 μM. This can be adjusted up or down depending on the K_d and/or the signal amplitude (magnitude of ΔH). It is typical that the ligand in the syringe is added to a higher final molar concentration in order to fully saturate the binding partner in the cell, thus requiring more material. Also the syringe volume is typically a factor of ~5 or so smaller than the cell volume so to achieve the final excess concentration, it requires that the ligand be 10–20 times the concentration of the cell material. It is therefore important that the ligand can be concentrated to these levels without aggregating or undergoing other types of specific self-association.

In the absence of any such constraints, it is possible to perform ITC experiments in either configuration, say with the protein in the cell and the reverse titration with the protein in the syringe since ITC is directly measuring the heat of the non-covalent interactions that are involved. Indeed, it can be highly informative to perform experiments in both orientations since it can help to dissect the mechanism of more complex interactions as well as confirm reactant concentrations. For example, if a protein has two independent identical binding sites for a ligand with $\Delta H = x$, then a stoichiometry of $n = 2$ will be recorded, while on reversing the titration $\Delta H = 2x$ and $n = 0.5$. Similarly, if the binding mode of an interaction is established as 1:1 but the titration produces $n = 0.8$, then this could indicate a 20% error in the protein concentration either through inaccurate measurement or as a result of 20% of the protein being "inactive" and unable to bind the ligand. The reverse titration should then yield $\Delta H = 0.8x$ and $n = 1.2$ (since the enthalpy of interaction is determined by the concentration of the syringe component which is 20% lower). More complex effects are produced in the case of multiple sites that are nonequivalent or exhibit any cooperativity, and the binding curves can look really different when reversing the titration. However, the additional information content in these experiments will always help discriminate competing models and give increased confidence in fitted parameters.

Other scenarios may occur where the data from reverse titrations are more radically different and this indicates the need for further biophysical characterization of the system being studied. Processes of self-association in one component are an example; protein coiled-coil dimers may be stable and fully formed at required concentrations when in the ITC cell, but in a reverse titration may dissociate when injected from the syringe since despite the higher concentration required in the syringe, there is a dilution factor of 100 for a 2 μL injection into the cell volume of 200 μL. Thus, in the reverse titration, the initial injections of coiled-coil will include additional heat effects as the dimer dissociates upon dilution followed by some amount of reassociation and binding in the presence of its stabilizing binding partner in the cell.

3.7 All Heat Looks the Same: There Are No Different "Colors"

The heat produced during an ITC experiment can have many sources but these are measured collectively during formation of a protein–ligand complex. Some of the heat is background signal that must be subtracted from the data before fitting, heat inherent to the mechanics of injection and mixing (injection heats are still seen for example when injecting water into water; *see* **Note 13**) and some originates from the dilution of the ligand when injecting small volumes of concentrated stock in the syringe into the larger volume of the cell (injection heat seen when injecting ligand in buffer into identical buffer). These background heats are typically determined in separate control measurements of ligand titrated into

buffer but can also be determined during a protein–ligand interaction experiment if a sufficient excess of ligand is used, so that saturation of the protein is achieved before the end of the titration and the final few injections correspond to the ligand being injected into the buffer without any binding occurring.

Other sources of heat may arise from "indirectly" coupled events such as buffer ionization that occurs if there is a net proton flux associated with the protein–ligand interaction. A change in protonation will occur whenever there is a shift in pK_a of groups in the protein or ligand as a result of complex formation: obviously a common feature where ionizable groups participate in the interaction providing favorable or unfavorable contributions (*see* **Notes 8** and **14**). Similarly, there are more "directly" coupled events such as the displacement of the majority or all solvent from the protein–ligand interface. This produces heat effects because water and ions that are solvating the surfaces of protein and ligand end up displaced from the interface into bulk solvent upon formation of the complex where they have different interaction partners. This solvation environment effect can contribute heat to the signal.

Despite all this complexity, the measured ITC heat signal is still a direct reporter for the extent of protein–ligand interaction under the conditions of measurement (temperature, buffer, pH, ionic strength, etc.) and thus can be fit to yield an association constant (and free energy) for the process. Meanwhile the amplitude of the ITC signal (ΔH) will vary with conditions depending on the nature and size of these additional coupled events. It is then difficult to relate the measured thermodynamic enthalpy and entropy from an ITC measurement to high-resolution structural information that may be available for the interacting components or the complex. These structures are static snapshots that do not inform on the dynamics of the system nor detail other participants in the binding event, such as the solvent, both of which contribute to the energetics of binding seen by ITC.

3.8 Measurement and Temperature: The Heat Capacity for Binding

ITC experiments are performed at a specific temperature but modern instrumentation can measure over a wide range of conditions (~4–70 °C) giving scope to probe the temperature dependence of binding. There is normally a significant decrease in the constant pressure heat capacity (ΔC_P) between free components and their complexes in protein–ligand interactions that is largely a product of changes in the solvation of the interacting surfaces. This ΔC_P of binding makes both the enthalpy and entropy temperature dependent since:

$$\Delta H = \int_{T_1}^{T_2} \Delta C_P \mathrm{d}T \tag{5}$$

$$\Delta H_{T_2} = \Delta H_{T_1} + \Delta C_P \left(T_2 - T_1\right) \tag{6}$$

$$\Delta S = \int_{T_1}^{T_2} \frac{\Delta C_P}{T} \mathrm{d}T \tag{7}$$

$$\Delta S_{T_2} = \Delta S_{T_1} + \Delta C_P \ln\left(\frac{T_2}{T_1}\right) \tag{8}$$

So in addition to enthalpy and entropy depending on solution conditions, such as buffer and pH used, the values will also differ depending on the temperature of measurement. This frustrates further any interpretation of enthalpy and entropy as isolated values from single ITC experiments. Indeed, the values of enthalpy for biomolecular interactions are generally not large ($\pm$<10 kcal/mol) and when combined with a significant ΔC_P of binding there may be experimental temperatures where ΔH is close to zero. Thus, if an ITC experiment initially produces no evidence of binding, it is always prudent to repeat the measurement at a higher or lower temperature in case the initial choice was such a situation.

These large changes in enthalpy and entropy (even to the extent of inverting sign) tend to have a much more minor effect on the binding free energy in a phenomenon of enthalpy–entropy compensation. Thus, as one term changes to increase binding affinity, the other changes in a compensating direction that opposes the increase (*see* Eq. 4). There are many sources of this compensation that are widely discussed in the literature [11, 12]. They reflect aspects of the experimental and analytical limitations of ITC measurements as well as contributions from the chemistry of the interacting components and the ever present yet unquantified effects of solvent. Unfortunately, the consequence of this compensation is often to frustrate any ligand design that attempts to be guided by thermodynamics or structure alone.

From Eq. 6, it is evident that ΔC_P can be determined simply from the slope of a plot of measured enthalpy against temperature. For example, the value determined for lysozyme binding tri-acetyl glucosamine is ~ −0.1 kcal/mol/K in acetate buffer pH 5. Ironically, despite the enthalpy being difficult to interpret, the value of ΔC_P itself does have some predictive power since it seems to scale with the surface area of the interacting interface. Alternatively, an abnormally large ΔC_P or a positive value can be a sign of binding being coupled to folding or unfolding of one or both interacting components. These changes are of particular interest in studying binding of IDPs [13].

3.9 Same "Color" of Heat, But Different Kinetics

Although the heat produced during interactions is the sum of many changes in the system, these processes can have different kinetics (rates of heat release). The input of differential power in ITC instrumentation is typically damped to smooth out high-frequency

fluctuations (noise) in the feedback circuitry. Time constants of 3–10 s or slower, depending on ITC model, are typically the maximum that can be obtained for "instant" heat signals (either electrical signals or dilution experiments; *see* **Note 15**). Processes that are slower than this instrumental response, which might include conformational rearrangements, polymerization, or aggregation, will lead to broadening of the ITC peak shape. This can even generate very unusual peak profiles with opposing endothermic and exothermic phases in the same injection once the slower events become uncoupled from the faster mechanical and dilution heats that are inherent to every injection. The binding itself can also involve multiple steps, such as complex formation that then triggers a conformational change, and these may also have disparate kinetic time windows and result in unusual peak profiles.

More generally, there is the potential for changes in the rate of heat release during every ITC titration because each injection changes the concentrations of reactants, and the equilibrium between bound and unbound will change following each injection with kinetics that are determined by contributions from forward and reverse rate constants (*see* Eq. 2). As the effective equilibrium concentration of free protein decreases during sequential addition of ligand because of binding, then the bimolecular rate component will also decrease leading to a reduction in the overall rate of relaxation of the equilibrium. Where this process leads to slowing beyond the time constant of the ITC there will be a broadening of the ITC peak shape reflecting this kinetic limitation. This phenomenon leads to the possibility of extracting kinetic on and off rate constants from ITC data during analysis in addition to the standard thermodynamic binding parameters [6, 14, 15]. It is certainly not uncommon to observe this kinetic broadening of ITC peak shapes as the heat effect of adding ligand decreases when the experiment approaches saturation, but the ability of this analysis to yield reliable kinetics will depend on the experimental design, the magnitude of the individual rate constants, and the configuration of the ITC instrument.

3.10 Not Just Protein–Ligand Interactions

The nature of ITC as a label-free nonoptical method gives it distinct advantages over many other techniques under challenging conditions (such as high absorbance solutions or crude cell extracts) or for more complex binding interactions. Competition (displacement) binding experiments can be easily performed to confirm whether different binding partners share the same site. Where ligands compete for an identical site, competitive titrations can increase the upper and lower range of affinities that ITC can measure. An economical way to perform this type of experiment for a high-affinity interaction (low nM or tighter) is to measure a weaker binding ligand in an initial titration and leave this end point in the ITC cell. The syringe is then reloaded with the tighter binding

ligand that displaces the first ligand in a second titration (*see* **Note 16**). There are effects on both the observed K_d of the second experiment (bringing it experimentally into a *c*-value window suitable for measurement) and on the ΔH since the ITC measures the net heat of both processes, displacement and binding, which frequently have opposing signs. These multiple signatures of displacement binding are useful in applying this methodology to ligand screening. In this case, titrating an established ligand against a protein that has been premixed with one or more potential competing compounds gives changes in both ΔH and K_d where there is competition.

ITC can also be used to quantify processes of reversible equilibrium self-association, such as dimerization, using simple buffer dilution experiments. For systems with a moderate (μM) K_d for forming a dimer (or higher order species), an injection into buffer and the resulting change in concentration produces some levels of dissociation. This produces a series of injection peaks (typically endothermic) that gradually decrease in magnitude since they result in progressively smaller concentration changes from the stock solution in the syringe as the concentration of protein in the cell increases. All the protein can be recovered after this titration into buffer and the data can be fit to give a K_d and ΔH for the process.

These two "atypical" applications of ITC rely to some extent on the technique's ability to measure heat nonspecifically as the sum of all events occurring during injection and mixing. This nonspecificity makes interpretation of ΔH of a single isolated experiment very difficult. But it also enables ITC to be deployed in a wide range of applications in protein chemistry and in the broader biological context. Noteworthy among these in the context of this chapter is ITC's potential use in studying enzyme kinetics, where ligand binding is followed by catalytic turnover, as has been reviewed elsewhere [16, 17].

4 Notes

1. These details refer to the Malvern Panalytical iTC200 instrument. Other models of ITC with different cell geometries, volumes, and syringe types are available as are instruments from other manufacturers such as TA Instruments.
2. There can be some ambiguity in the use of the term "ligand" (meaning "to bind") across various techniques. Classically, a receptor ligand would often be much smaller in size than the protein itself. In ITC, it is common to refer to the component being titrated from the syringe as the ligand (even if it is a protein, as in a protein–protein interaction, or physically larger than the other component). This cell/syringe terminology will

be used here. In SPR and other surface-based sensor technologies, the ligand is designated as the binding component immobilized on the sensor surface (irrespective of size) while the other partner in the interaction is then termed the analyte.

3. If measurements are being made in solvent with a significantly different heat capacity, e.g., buffers containing high salt concentration, organic solvents such as DMSO, or cryo preservatives such as glycerol, then the reference cell should be filled with the same solvent.
4. Heat is being applied to both ITC cells during measurement with a very small excess positive differential power being applied to the sample cell. Since the cells are well insulated in the ITC instrument, this energy input leads to a very small increase in the absolute temperature of the cells over the duration of the measurement. However, this is very far from a level significant for measurement accuracy and probably smaller than temperature variations seen in many other techniques. The temperature of measurement reported should be that set in the instrument control software and used by the ITC during its initial equilibration phase.
5. The integration of the excess peaks in differential power seen in ITC data by defining baseline regions before and after the injections can be problematic especially where the amplitude of the heat effect is close to background noise levels. The analysis software supplied with ITC instruments can generally do a reasonable job in this regard but may require "manual adjustment" where there is drift, stepping, transient glitches, or spurious points in the data. Other programs are available, such as NITPIC, which have been developed to try and remove any requirement for user input in this process [18]. Users are encouraged to "experiment" with different manual and automated approaches. However, if the experimental result and subsequent conclusions are in any way dependent on the way in which the program or user has defined the baselines, then it is clear that the experiment should be repeated and preferably under conditions giving better signal amplitude that will eliminate any dependence on this element of the analysis process.
6. ITC instruments use a total cell fill mode, so that the cell volume and protein concentration define how much material is available. During the sequential injections of ligand, some of the cell volume is displaced out into the filing tube and this is prevented from re-equilibrating with cell material by the Teflon plug on the syringe needle that sits in the neck of the cell. As a consequence, the protein concentration in the cell will decrease with each injection that increases the ligand concentration. However, since the injection typically takes several seconds,

then some of the protein that will be displaced can still participate in the binding and produce measured heat before it is eventually ejected. The ITC instrument software accounts for these complex effects and so it is best to use the concentrations of protein and ligand that are tabulated in spreadsheet for export of the data and fitting elsewhere.

7. It is common practice to include the small "pre-injection" in ITC experiments eliminating premixed protein and ligand arising from diffusion during the equilibration phases of the instrument as well as some unmixed ligand. In the iTC200, this might be 0.5 μL while subsequent injections are 2 μL. The concentration of ligand after this pre-injection is assumed to be correct but the heat produced by the injection is reduced due to some level of premixing and so it is not included as a data point when fitting the titration. However, there is also the possibility that the heat of this first injection is underestimated due to mechanical backlash in the stepper motor that is used to drive the syringe when it reverses direction following the loading process [19]. Solutions are drawn into the syringe from the tip during loading and titrated back into the cell by driving the system in the reverse direction. Thus, it is critical to make a small movement in the direction of ligand ejection (a "downward" movement in the iTC200 software) immediately after loading but before putting the syringe into the ITC cell. Without this procedure, the pre-injection will take up the backlash and give a smaller volume and incorrect ligand concentration after injection 1.

   The use of approximately 20 injections during ITC titrations is generally engrained in the field as a "standard methodology" but there is evidence that the number of points in the data set may be less important in constraining fitted parameters than the amplitude of the heat measured for each injection. It has been suggested that ten injections or even fewer may be a more optimal experimental strategy, although despite some considerable weight of literature in this area, this has not been widely adopted [20] (and references therein). Advice would generally be to initially use 20 injections to account for possible complexity in the shape and features of the ITC binding data but that for established systems with well characterized binding behavior then a reduced number of larger injection volumes should be considered.

   An alternative approach to sequential injections and baseline recovery is to perform a single very slow injection of ligand from the syringe [21]. The ITC signal is then contained in a single differential power deflection that can be integrated and analyzed to yield binding parameters. This single injection methodology (SIM) has potential for some additional time

saving in higher throughput situations since it does not involve repeated phases of baseline re-equilibration. However, it is essential that the binding kinetics (k_{obs}) is faster than the rate of change in ligand concentration during the single injection which limits the time economy of the method. In light of the relatively low throughput of ITC in standard or SIM mode and the modest gains in experimental time, it is perhaps not surprising that SIM is not very commonly employed.

8. Buffers will have a characteristic heat (enthalpy) of ionization in the equilibrium; $BH \leftrightarrow B + H^+$, and any net proton flux inherent to protein–ligand interactions will be provided or absorbed in this associated buffering process [22]. The magnitude of the ionization enthalpy also determines the sensitivity of the buffer system to changes in temperature; through mass action effects, the ionization equilibrium is driven in the direction of endothermic heat absorption as temperature is increased. Most buffers become more ionized and the pH decreases, in the worst case for TRIS buffer, by ~0.3 pH units for +10 °C.
9. The ProtParam program uses protein sequence to sum the molar extinction coefficients of aromatic amino acid content along with minor contributions from cystine (cysteine residues oxidized in disulfides). It thus reports values for the oxidized and reduced forms of the protein. This approach is based on two publications where extinction coefficients from solvent exposed residues were used (denatured proteins or short unstructured peptides) [23] or where average values were obtained from a collection of folded proteins [24]. The values are slightly different, reflecting the sensitivity of the aromatic absorption to environment. ProtParam currently uses values from the folded protein data set. However, it is clear that the extinction coefficients obtained in this way are not without a potential error of a few percent or so and other methods may be considered depending on the sequence, oxidation state, etc. (*see* Chapter 2).
10. Before measuring first check the quoted pathlength accuracy of the cuvette or instrument device being used since this could be another source of uncertainty in the absolute value measured. The spectrometer should be blanked carefully using the exact solvent that the protein is in. If the protein has been dialyzed then the final dialysis buffer should be used. There can be problems where DTT is present, often at mM levels, since its oxidized state has absorbance with maxima 283 nm exactly overlapping the aromatic absorbance region of proteins. This can produce large errors where oxidation has proceeded at differing rates in a protein sample compared to buffer stock solutions as the buffer is no longer a suitable blank.

Samples for measurement should have an absorbance between 0.1 and 1.5 at the measurement pathlength (optimally absorbance 0.87) or should be diluted. Pipettes used for any dilutions should be reliable and properly calibrated. Alternatively, the dilution can be done with a balance assuming 1 g/mL density. It is also important to measure the full spectrum of the protein from 240 nm to at least 360 nm or longer wavelengths. Examining the spectrum gives important additional information about the sample. A 280/260 nm ratio >1.6 is characteristic of pure proteins while lower values can indicate nucleic acid or nucleotide contributions as contaminants or bound ligands. At longer wavelengths >320 nm, the protein should not absorb, unless it contains an additional prosthetic group. Thus, absorbance above zero in this range is indicating scattering from aggregated material or other particles in the sample. Scattering has a reciprocal dependence on wavelength raised to power between 2 and 4 and so at lower wavelengths, such as where the protein is being quantified, the apparent absorbance will be even higher. If scattering absorbance is observed >320 nm, it can be corrected for at 280 nm by plotting the log of absorbance observed above 320 nm and fitting this to a linear function. The extrapolated log of absorbance at 280 nm should be converted to absorbance and subtracted from the measured value. If this correction indicates more than a few percent of measured absorbance, then it could be a cause of some concern indicating that the sample contains significant levels of aggregated and presumably inactive materials.

11. The SI unit of heat is the Joule. Unsurprisingly, in the area of calorimetry, there is some attachment to the older unit of calories as has been used here. NB: 1 cal = 4.184 J.
12. The background control heat associated with the ITC experiment can be determined in a separate experiment injecting ligand into buffer. Solutions used should be identical to those of the binding experiment. The diluted ligand can be recovered from the ITC at the end of the experiment and recycled. The background end point heat has contributions from the mechanical and instrumental heat associated with an injection (seen when injecting water into water) and from the heat of dilution of the ligand associated with its ~100-fold drop in concentration and any subtle solvent differences between the protein and ligand solutions. Solvent effects can be minimized by dialyzing extensively both components against buffer. Heats of dilution will vary from ligand to ligand because they are inherent to its solvation and chemistry. In addition, each injection of a particular ligand increases its concentration in the cell so that the effective dilution factor decreases during the

titration. This can be seen on close inspection of control heats as a small linear change in the peak integral with injection number. In this case these values should be fitted to a linear function which is used to correct the experimental data prior to fitting.

13. It is highly informative to examine the signal from an ITC instrument injecting water from the syringe into water in the cell as this will give users an idea of the optimal signal-to-noise that can be expected when there is no binding or dilution heats. Changes in this signal may be diagnostic of mechanical problems (bent syringe needle) or difficulty with total fill operation that may arise when the cell requires cleaning. In addition, the amplitude of the water–water injections will depend on the volume of the injection, the speed with which the injection is made and the temperature of measurement in the ITC. This reflects the fact that the syringe itself is located outside of the calorimeter cell at room temperature and that the final temperature equilibration between the syringe component and the solution in the cell takes place as the solution passes down the needle into the core of the instrument and the measurement cell. The larger the volume injected, the faster this occurs and the further from room temperature is the temperature of measurement then the larger will be the background heat from this process. Having a reference data set of such effects will allow the selection of optimal ITC parameters for real binding experiments.
14. Calorimetry provides a unique and direct method for determining the net proton flux associated with an interaction if measurements are made in buffers of differing ionization enthalpy but under identical conditions of pH and ionic strength and temperature. In such cases, the observed ITC enthalpy is simply plotted against the buffer ionization enthalpy to yield a slope equal to the net number of protons exchanging during the interaction with the intercept equal to the enthalpy of the interaction in the absence of coupled buffer effects.
15. Determining the time constant (instrument response function) of an ITC instrument is relatively easy as the sample cell has electrical calibration heaters installed to facilitate calibration of the differential power signal. These heaters produce a defined offset in differential power and this heat is input much faster than the time constant of the instrument. Therefore, the change in signal to the new differential power level will occur as an exponential function that can be fit to yield the time constant of the ITC.

A more realistic approach is to inject a solution with a high heat of dilution (such as 1% v/v ethanol or methanol) into water in the cell. This method includes the effects of injection speed and stirring on the measurement. The final portion of the heat peak produced as the signal recovers to baseline should fit an exponential that is the time constant of the ITC. It is likely that injection and stirring speed will influence the values obtained, and it has been noted that cell cleanliness seems to surprisingly have a significant effect [6]

16. On removing the syringe to reload while leaving the cell contents in situ, it is important to recalculate the concentrations of protein and ligand present. During the initial titration, material is eliminated from the cell and does not re-equilibrate as discussed in **Note 3**. However, the removal of the syringe to reload will remix the contents of cell and loading tube giving a new set of starting conditions for a subsequent titration.

References

1. Sophianopoulos AJ, Vanholde KE (1964) Physical studies of muramidase (lysozyme). II. pH-dependent dimerization. J Biol Chem 239:2516–2524
2. Turnbull WB, Daranas AH (2003) On the value of c: can low affinity systems be studied by isothermal titration calorimetry? J Am Chem Soc 125(48):14859–14866. https://doi.org/10.1021/ja036166s
3. Kantonen SA, Henriksen NM, Gilson MK (2018) Accounting for apparent deviations between calorimetric and van't Hoff enthalpies. Biochim Biophys Acta, Gen Subj 1862 (3):692–704. https://doi.org/10.1016/j.bbagen.2017.11.020
4. Kantonen SA, Henriksen NM, Gilson MK (2017) Evaluation and minimization of uncertainty in ITC binding measurements: heat error, concentration error, saturation, and stoichiometry. Biochim Biophys Acta, Gen Subj 1861(2):485–498. https://doi.org/10.1016/j.bbagen.2016.09.002
5. Brautigam CA (2015) Fitting two- and three-site binding models to isothermal titration calorimetric data. Methods 76:124–136. https://doi.org/10.1016/j.ymeth.2014.11.018
6. Dumas P, Ennifar E, Da Veiga C, Bec G, Palau W, Di Primo C, Pineiro A, Sabin J, Munoz E, Rial J (2016) Extending ITC to kinetics with kinITC. Methods Enzymol 567:157–180. https://doi.org/10.1016/bs.mie.2015.08.026
7. Le VH, Buscaglia R, Chaires JB, Lewis EA (2013) Modeling complex equilibria in isothermal titration calorimetry experiments: thermodynamic parameters estimation for a three-binding-site model. Anal Biochem 434 (2):233–241. https://doi.org/10.1016/j.ab.2012.11.030
8. Vega S, Abian O, Velazquez-Campoy A (2015) A unified framework based on the binding polynomial for characterizing biological systems by isothermal titration calorimetry. Methods 76:99–115. https://doi.org/10.1016/j.ymeth.2014.09.010
9. Zhao H, Piszczek G, Schuck P (2015) SEDPHAT--a platform for global ITC analysis and global multi-method analysis of molecular interactions. Methods 76:137–148. https://doi.org/10.1016/j.ymeth.2014.11.012
10. Freiburger L, Auclair K, Mittermaier A (2015) Global ITC fitting methods in studies of protein allostery. Methods 76:149–161. https://doi.org/10.1016/j.ymeth.2014.12.018
11. Olsson TS, Ladbury JE, Pitt WR, Williams MA (2011) Extent of enthalpy-entropy compensation in protein-ligand interactions. Protein Sci 20(9):1607–1618. https://doi.org/10.1002/pro.692
12. Chodera JD, Mobley DL (2013) Entropy-enthalpy compensation: role and ramifications in biomolecular ligand recognition and design. Annu Rev Biophys 42:121–142. https://doi.org/10.1146/annurev-biophys-083012-130318

13. Vega S, Abian O, Velazquez-Campoy A (2016) On the link between conformational changes, ligand binding and heat capacity. Biochim Biophys Acta 1860(5):868–878. https://doi.org/10.1016/j.bbagen.2015.10.010
14. Burnouf D, Ennifar E, Guedich S, Puffer B, Hoffmann G, Bec G, Disdier F, Baltzinger M, Dumas P (2012) kinITC: a new method for obtaining joint thermodynamic and kinetic data by isothermal titration calorimetry. J Am Chem Soc 134(1):559–565. https://doi.org/10.1021/ja209057d
15. Vander Meulen KA, Horowitz S, Trievel RC, Butcher SE (2016) Measuring the kinetics of molecular association by isothermal titration calorimetry. Methods Enzymol 567:181–213. https://doi.org/10.1016/bs.mie.2015.08.012
16. Demarse NA, Killian MC, Hansen LD, Quinn CF (2013) Determining enzyme kinetics via isothermal titration calorimetry. Methods Mol Biol 978:21–30. https://doi.org/10.1007/978-1-62703-293-3_2
17. Transtrum MK, Hansen LD, Quinn C (2015) Enzyme kinetics determined by single-injection isothermal titration calorimetry. Methods 76:194–200. https://doi.org/10.1016/j.ymeth.2014.12.003
18. Keller S, Vargas C, Zhao H, Piszczek G, Brautigam CA, Schuck P (2012) High-precision isothermal titration calorimetry with automated peak-shape analysis. Anal Chem 84(11):5066–5073. https://doi.org/10.1021/ac3007522
19. Mizoue LS, Tellinghuisen J (2004) The role of backlash in the "first injection anomaly" in isothermal titration calorimetry. Anal Biochem 326(1):125–127. https://doi.org/10.1016/j.ab.2003.10.048
20. Tellinghuisen J (2012) Designing isothermal titration calorimetry experiments for the study of 1:1 binding: problems with the "standard protocol". Anal Biochem 424(2):211–220. https://doi.org/10.1016/j.ab.2011.12.035
21. Markova N, Hallen D (2004) The development of a continuous isothermal titration calorimetric method for equilibrium studies. Anal Biochem 331(1):77–88. https://doi.org/10.1016/j.ab.2004.03.022
22. Cooper A, Johnson CM (1994) Introduction to microcalorimetry and biomolecular energetics. Methods Mol Biol 22:109–124
23. Gill SC, von Hippel PH (1989) Calculation of protein extinction coefficients from amino acid sequence data. Anal Biochem 182(2):319–326
24. Pace CN, Vajdos F, Fee L, Grimsley G, Gray T (1995) How to measure and predict the molar absorption coefficient of a protein. Protein Sci 4(11):2411–2423. https://doi.org/10.1002/pro.5560041120

Chapter 6

Measuring the K_D of Protein–Ligand Interactions Using Microscale Thermophoresis

Shih-Chia Tso and Chad A. Brautigam

Abstract

Microscale thermophoresis (MST) has become a widely used technique to determine the K_D or EC_{50} of protein–ligand interactions. The method exploits the tendency of macromolecules to migrate along a thermal gradient (i.e., thermophoresis). Differences in thermophoresis as a function of the liganded state of a macromolecule can be measured and assembled into a binding curve that can be analyzed to yield K_D. In this protocol, we outline a simple experiment designed for new MST users, with the goal of using readily available, inexpensive materials to plan, execute, and analyze an MST experiment.

Key words Microscale thermophoresis, Protein–ligand interactions, Protein–protein interactions, K_D, Affinity measurement

1 Introduction

Since it became available in a commercially distributed instrument, microscale thermophoresis (MST) has enjoyed rapid growth and widespread adoption. Although many kinds of experiments can be conducted in an MST instrument [1], most MST studies aim to characterize the interaction(s) between a fluorescently labeled receptor (termed "B*" herein) and an unlabeled ligand (called "A" henceforth; the 1:1 complex between the two is "AB*"). In such studies, a mixture of the receptor and ligand is placed into a glass capillary tube (Fig. 1a). A portion of the tube near its center is illuminated with visible light that stimulates the fluorophore associated with the receptor. Fluorescence emitted from the capillary is monitored via a detector configured as a confocal microscope. Next, an infrared (IR) laser is actuated, rapidly forming a temperature gradient at the monitored part of the capillary. Through a process called "thermophoresis," the labeled receptor molecules display a net movement along this gradient, changing the net fluorescence recorded (Fig. 1b). Thermophoresis is usually "positive"; i.e., the molecules move from hotter to colder areas of the

Tina Daviter et al. (eds.), *Protein-Ligand Interactions: Methods and Applications*, Methods in Molecular Biology, vol. 2263, https://doi.org/10.1007/978-1-0716-1197-5_6,

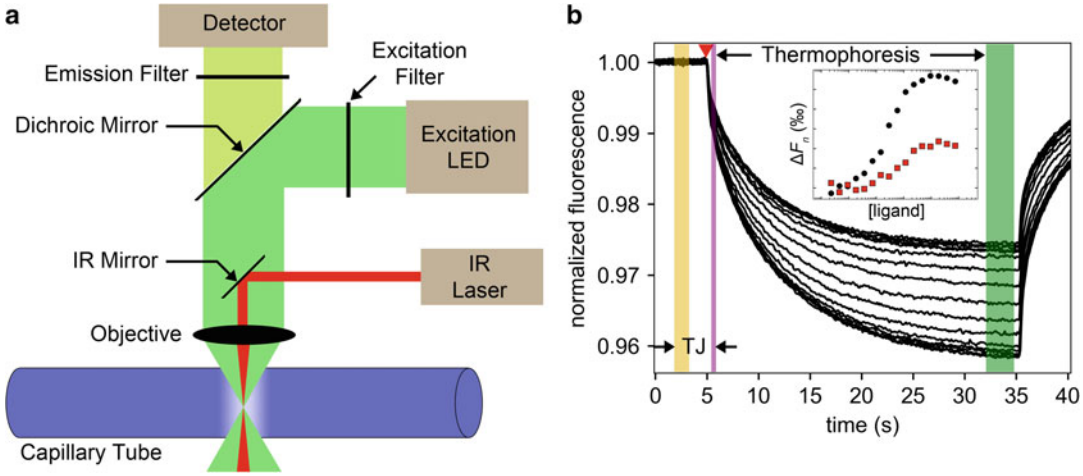


Fig. 1 Microscale thermophoresis. (**a**) A schematic of an MST experiment (not to scale). The fluorescent sample (blue), which is housed in a glass capillary tube, is illuminated (green) by an LED with an appropriate excitation filter in line. This light is directed to and focused on the sample using a mirror and an objective. Note that only a small portion of the capillary is illuminated. IR radiation (red), originating from a laser, is also focused on the sample. Fluorescence emission is passed through a filter and impinges on a solid-state detector. (**b**) MST fluorescence traces. All of the traces have been normalized to an initial value of 1.0. Fluorescence is tracked as a function of time; the IR laser is activated approximately 5 s into the experiment, marked by an inverted red triangle. In this chapter, we focus on three areas of comparison: before IR illumination, just after IR illumination, and just before IR extinguishment, marked by gold, purple, and green, respectively. "Thermophoresis" compares the purple and green areas, while temperature jump or "T-Jump" compares the gold and purple areas

capillary tube, manifesting as a diminution in fluorescence. Because the net observed thermophoresis is expected to be linearly dependent on the liganded state of the receptor, this characteristic can be monitored over many capillaries having different concentrations of the ligand, and these signals can be transformed into a binding curve (Fig. 1b, inset) suitable for analyses designed to yield the K_D of the interaction (in more complex situations, two K_Ds or the EC_{50} of the interaction can be estimated).

Although thermophoresis theory, developed in this context by Duhr and Braun [2–7], will not be treated comprehensively here, the reader should be familiar with some basic facts about the method and some terminology. First, thermophoretic phenomena are governed by the Soret coefficient, S_T, where

$$\frac{c_{hot}}{c_{cold}} = e^{-S_T \Delta T} \tag{1}$$

with c_{hot} and c_{cold} representing the concentrations of the monitored molecules in the hot (IR-illuminated) and cold areas of the capillary, respectively [5], and ΔT representing the temperature change. S_T is thought to be a function of several molecular properties,

including size, charge, and hydration entropy [2]. The quantity measured in MST experiments is a normalized fluorescence, F_n:

$$F_n = \frac{F_s}{F_r} = 1 + \left(\frac{\partial F}{\partial T} - S_T\right)\Delta T \tag{2}$$

where F_s is the fluorescence in a "subject" region of the fluorescence trace, F_r is the fluorescence in a "reference" region, and $\partial F/\partial T$ is the response of the fluorophore to the temperature change [7]. As depicted in Fig. 1b, for example, for "Thermophoresis," F_r is the average fluorescence in the purple zone and F_s is the average fluorescence in the green zone. F_n is usually multiplied by 1000 for numerical convenience. In this framework, we define F_{n,B^*} and F_{n,AB^*} as the F_n values for pure labeled receptor, B*, and the ligand–receptor complex, AB*, making the assumption that

$$F_n = \left(\frac{B^*}{B^*_{tot}}\right)F_{n,B^*} + \left(\frac{AB^*}{B^*_{tot}}\right)F_{n,AB^*} \tag{3}$$

where we establish the convention throughout this chapter that the italicized component name stands for its respective molar concentration. The binding curve, therefore, represents many such F_n values assembled as a function of A_{tot} (Fig. 1b, inset). It can be analyzed to yield K_D using standard mass-action and mass-conservation considerations [7, 8].

Because many of the factors needed to describe the Soret coefficient are not known *a priori*, a successful MST experiment involves thoughtful planning and careful pilot experimentation. With the introduction of a (usually extrinsically) labeled receptor, control experiments are needed to rule out nonspecific binding effects. Finally, the rigorous treatment of parameter confidence intervals and correlations can be achieved through the use of modern analysis software. The goal of this chapter is to cover all of these aspects of MST experimentation and analysis. All steps of the thermophoresis experiment will be presented, including labeling the protein with a fluorophore, optimizing experimental parameters, performing the experiment, data analysis, and data presentation.

The chapter will focus on the interaction of the readily available proteins α-chymotrypsin (α-CT) and soybean trypsin inhibitor (SBTI) to produce a protocol that can be followed in virtually any laboratory equipped with an MST instrument. Here, the "ligand," "A," is the inhibitor, which is in fact a 20-kDa protein. Although the differences between this protocol and one using a small molecule as the ligand could be minimal, we have striven to note where such differences can occur and offer suggestions to optimize such experiments. Notably, two molecules of α-CT ("B*") bind to one of SBTI. However, under the conditions presented here, the binding curve has the appearance of a 1:1 interaction and may be treated as such.

2 Materials

Prepare all buffers in ultrapure H_2O, which is attained by filtering deionized H_2O through purification filters until it has a resistivity reading of ≥18 MΩ cm at room temperature. All buffers and proteins should be stored at 4 °C when not in use. Follow all institutional practices on waste disposal, particularly of the glass capillaries used in the MST experiments. Use appropriate personal protective gear (gloves, eyewear, etc.) throughout.

2.1 Buffers/Detergent

1. 10× phosphate-buffered saline (PBS): Place 80 g of NaCl, 2 g of KCl, 17.8 g of $Na_2PO_4 \cdot 2H_2O$, and 2.45 g of KH_2PO_4 in a 1 L graduated cylinder, and fill to 950 mL. Adjust the pH to 7.4 using 12N HCl or 10 M NaOH as necessary. Bring the final volume to 1.0 L with H_2O. Pass through a 0.22-μm filter and store.
2. 1× PBS: We find that sometimes dilution from the 10× stock alters the final pH of the 1× PBS. Therefore, we find it prudent to check the pH upon dilution as follows. Add 100 mL of 10× PBS to 850 mL of H_2O. Check the pH and adjust as necessary using 12N HCl or 10 M NaOH. Bring the total volume to 1 L with 0.22-μm filtered H_2O.
3. 10% (v/v) Tween-20 (supplied by NanoTemper Technologies GmbH when ordering capillaries, see below).
4. PBS–Tween: Mix 5 μL of 10% Tween-20 with 995 μL of 1× PBS (*see* **Note 1**).

2.2 Proteins

1. α-chymotrypsin (α-CT): Dissolve approximately 20 mg of the protein with 1 mL of 1× PBS. Place 500 μL into a 0.22-μm centrifuge filter and centrifuge at a maximum of 5000 × *g* for 5 min (*see* **Note 2**). Dilute the protein 100-fold (i.e., use a dilution factor, *f*, of 100) and measure its absorbance at 280 nm (A_{280}) using a spectrophotometer (*see* **Notes 3–5**). Using an extinction coefficient (ε_{280}) of 51,840 M^{-1} cm^{-1}, calculate the concentration of the stock using the following formula:

$$\alpha - CT = \frac{A_{280} \times f}{\varepsilon_{280} \times d} \tag{4}$$

   where *d* is the path length of the cuvette used. Make 200 μL of 20 μM α-CT, 0.05% (v/v) Tween-20 by mixing the proper volume of protein with 1 μL of 10% (v/v) Tween-20 and the remaining volume of 1× PBS.
2. Soybean trypsin inhibitor (SBTI): Dissolve approximately 20 mg of the protein with 1 mL of 1× PBS. Follow the same procedure as for α-CT to measure the concentration of the

protein, but using a dilution factor f of 10 and ε_{280} of 18,450 M^{-1} cm^{-1}. Make 100 μL of 40 μM SBTI, 0.05% (v/v) Tween-20.

3. Carbonic anhydrase isoform II (CAII): Dissolve approximately 2 mg of the protein in 250 μL of PBS. As above, measure the concentration spectrophotometrically, using a dilution factor of 100 and ε_{280} of 55,100 M^{-1} cm^{-1}. Make 100 μL of 40 μM CAII, 0.05% (v/v) Tween-20.

2.3 Dye Labeling

1. 40 mM Cyanine-5 *N*-hydroxysuccinimide ester (Cyanine-5 NHS): Dissolve 1 mg of Cyanine-5 NHS in 37.5 μL of 100% DMSO. Dispense the stock into 1 μL aliquots in PCR tubes and freeze at −80 °C. *See* **Note 6**.
2. A small-scale desalting gravity column, such as a PD MiniTrap G-25 column (GE Healthcare).

2.4 Thermophoresis Supplies

1. Standard- and Premium-coated capillaries, NanoTemper Technologies GmbH.
2. PCR tubes.

2.5 Software/Computer

1. PC computer equipped with Windows 7 or higher.
2. NT.Control or MO.Control Acquisition software (*see* **Note 7**).
3. PALMIST [8, 9] version 1.5.8 or higher: Available as a zipped archive for free at http://biophysics.swmed.edu/MBR/software.html.
4. GUSSI [10] version 1.4.2 or higher: Available as a zipped archive for free at http://biophysics.swmed.edu/MBR/software.html.
5. Complete the analysis environment by creating a folder called C:\sedfit. Next, extract the PALMIST archive into C:\sedfit. Do the same for the GUSSI archive. The C:\sedfit folder should now have subfolders called "PALMIST" and "GUSSI." The respective executables can be found in these folders. It is convenient to make desktop shortcuts to these executable or to pin them to the system taskbar.

2.6 Instrumentation

1. A NanoTemper NT.115 MST instrument equipped with a red filter set.

3 Methods

Keep all protein solutions on ice unless noted. All dye-containing solutions should be shielded from light when not in use.

3.1 Labeling the Receptor

1. Thaw one tube of Cyanine-5 NHS on ice. Add 9 μL of 1× PBS and mix by pipetting up and down (*see* **Note 8**). There may be signs of dye precipitation at this point; they can be safely disregarded.
2. Take 1 μL of the diluted dye and add it to the 200 μL of 20 μM α-CT. Mix gently by pipetting up and down. Incubate at room temperature for 30 min in the dark. During the last 10 min of the incubation period, transfer the tube to a room-temperature microcentrifuge and centrifuge it at maximum speed (usually about 15,000 × *g*) for 10 min.
3. While the labeling reaction incubates, equilibrate the small-scale desalting column with 9 mL of 1× PBS by allowing the solution to flow through the column by gravity.
4. Carefully remove the labeling reaction from the microcentrifuge and apply 190 μL of the supernatant to the top of the equilibrated column, allowing the solution to fully enter the column's bed. Next, apply 210 μL of 1× PBS, and allow it to enter the bed also.
5. Situate a 1.5-mL microcentrifuge tube under the column, and then apply 600 μL of 1× PBS to the column, collecting all eluent. Gently mix the eluent.
6. Using a spectrophotometer, scan the eluent from 250 to 800 nm (no dilution is necessary; *see* **Note 9**). Determine the *α-CT** (i.e., the concentration of the labeled protein) and the efficiency of the labeling (ϕ) using the following formulas:

$$\alpha - CT^* = \frac{A_{280} - 0.03A_{640}}{\varepsilon_{280} \times d}$$
$$\phi = \frac{A_{640}}{\varepsilon_{640} \times d \times \alpha - CT^*} \quad (5)$$

   where ε_{640} is 250,000 M^{-1} cm^{-1}. Starting from a 20-μM stock, we usually obtain 4–6 μM of the protein; using these conditions with α-CT, we obtained a final concentration of 4.4 μM labeled at 13% efficiency (*see* **Note 10**).
7. Make 1 mL of a 100 nM working stock of the labeled protein (*see* **Note 11**) by diluting the appropriate volume with 1× PBS, and make the final Tween-20 concentration of this solution 0.05% (v/v).

3.2 Optimizing Capillaries, LED Power, and Buffer Conditions

1. Mix 50 μL of the α-CT* working stock and 50 μL of PBS–Tween (this is the "0 μM SBTI" sample). In a second tube, mix 50 μL of the α-CT* working stock and 50 μL of 40 μM SBTI (this is the "20 μM SBTI" sample). Allow both mixtures to incubate in the dark at room temperature for 30 min.
2. Fill three Standard and three Premium capillary tubes by dipping one of each kind of capillary into the 0 μM SBTI sample.

The tubes "self-fill" by capillary action. Also, fill one Standard and one Premium capillaries with the 20 μM SBTI sample. Arrange the eight capillaries in the instrument's capillary tray in the following order, starting at capillary position 1:

(a) 0 μM SBTI, Standard Capillary
(b) 0 μM SBTI, Premium Capillary
(c) 20 μM SBTI, Standard Capillary
(d) 20 μM SBTI, Premium Capillary
(e) 0 μM SBTI, Standard Capillary
(f) 0 μM SBTI, Standard Capillary
(g) 0 μM SBTI, Premium Capillary
(h) 0 μM SBTI, Premium Capillary

3. Insert the capillary tray into the instrument (*see* **Note 12**), making sure that the red filter set is selected. If using NT. Control, proceed to **step 4**. In MO.Control, perform a "Binding Test," selecting "Auto-detect" for LED power and "Low" MST power. The test will conduct a pre-run capillary scan (i.e., it will pass LED light through all capillaries sequentially), a brief MST experiment (e.g., Fig. 1b) on all capillaries, and a post-run capillary scan (*see* **Note 13**). It will examine the capillary scans for overall intensity, scan shape, and intensity trends; it examines the MST Scans for signs of aggregation, and raises warnings if any problems occur. If a low- or high-intensity warning is sounded, the LED power can be adjusted (also *see* **Note 14**). If a scan-shape warning is raised, it likely indicates protein sticking to the capillaries, and that defect must be addressed before moving forward (*see* **Note 15**). An intensity-trend warning can be ignored at this stage; however, if there is a noticeable, systematic difference in the peak intensities between the Standard and the Premium capillaries, the capillary type having the highest intensity should be used henceforth. Note which LED power was selected by the software in this step. The results of the "Binding Test" should be ignored. Repeat this step with "Medium" and "High" MST powers, initiating a new "Binding Check" in both instances. Proceed to **step 7**.
4. Using NT.Control, perform a capillary scan on the eight capillaries at 50% LED power.
5. Examine the capillary scan (Fig. 2). In our case, all of the peaks in the scan have raw fluorescence intensity values of about 500. This is a desirable outcome; according to the manufacturer, the peak values should be between 200 and 2000. We usually aim for 400–500, and the LED power can be adjusted to meet this goal (*see* **Note 14**). Alternatively, the concentration of the labeled protein can be modified accordingly. Other features of

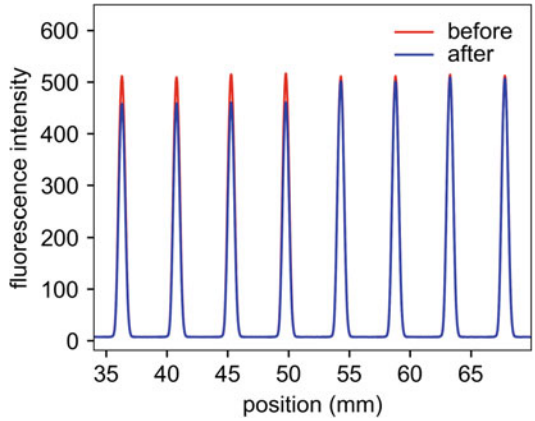


Fig. 2 Preliminary capillary scans. Capillaries are scanned in reverse order but displayed in the correct order, such that Capillary 1 (in this instrument) is located between 35 and 40 mm, while Capillary 8 is between 65 and 70 mm. Colors that represent the scan before and after the procedure described in **step 6** of Subheading 3.2 are described in the inset legend

the capillary scan should be noted. For example, are the peak shapes symmetric and monomodal? A dip in the middle of the peaks indicates protein adhesion to the interior surface of the capillaries, which must be addressed (*see* **Note 15**). Are all of the peaks of the same magnitude? If not, the capillaries having the highest peak magnitude should be used.

6. Perform a capillary scan on the first four capillaries alone, and assign pseudo-concentrations to these in descending order (e.g., 20,000, 10,000, 5000, and 2500). Next, perform an MST experiment with 50% LED power (or whatever LED power provides the optimal fluorescence reading), 5 s of pre-IR time, 30 s of IR-on time, and 5 s of IR-off time. Execute this three times sequentially, with 20%, 40%, and then 60% MST power (MST power is the power supplied to the IR laser).
7. Start PALMIST (*see* **Note 16**) and load the .ntp file (*or .moc file*) into the software using the File menu. In the .ntp/.moc browser, select one of the MST powers (ultimately, this should be repeated for all three powers), and then Load. Examine the fluorescence traces, which will appear in the upper graph in the software window (Fig. 3). One expects a smoothly shaped curve. Deviations from this expectation usually indicate aggregation and disqualify the samples for analysis until the problem can be ameliorated (*see* **Note 17**).
8. Assess whether there is a binding signal. Click on the "Thermophoresis" button on the lower-right part of the software window. An example is shown for the 60% MST data collected using NT.Control (Fig. 3). The right two data points are the 0-μM SBTI samples, and the left two have 20 μM SBTI. One

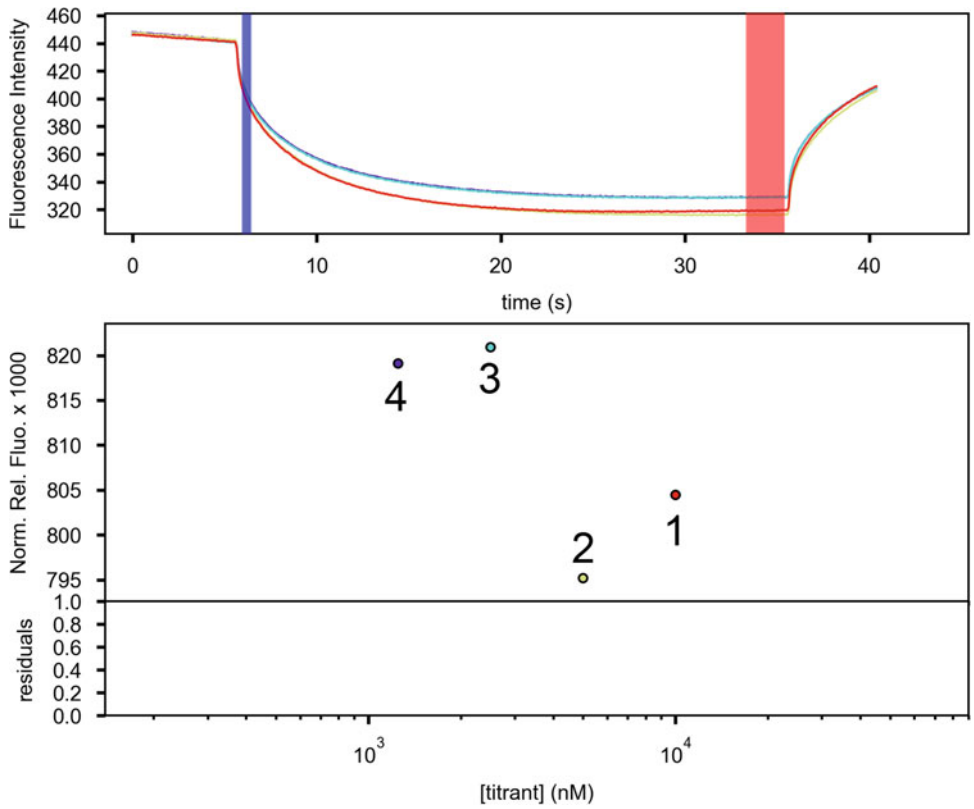


Fig. 3 Initial MST experiments. The fluorescence traces are shown in the upper panel. The colors indicate the identity of the capillary, proceeding along the visual spectrum from Capillary 1 (red) to Capillary 4 (violet). The blue- and red-shaded areas demarcate the areas of the traces that are used to calculate thermophoresis. The lower panel shows the thermophoresis values; the capillary identities are numbered

can readily observe that there will potentially be 20–25 units of thermophoretic signal available for the analysis, which is generally enough. This result indicates that the experiment can potentially succeed (*see* **Note 18**).

9. Perform another scan on all eight capillaries (this was done automatically if using MO.Control). This scan is important to establish that no sticking has occurred post-MST or over time (*see* **Note 13**). In our example (Fig. 2), we found that there was no sticking over time, but that the first four capillaries experienced discernible photobleaching; because our experiment was conducted using NT.Control, only these four capillaries experienced prolonged LED exposures. This fact should not affect the experiment.

3.3 Optimizing MST Power

1. Prepare a 1:1 serial dilution of SBTI by first labeling 16 PCR tubes with the numerals 1–16. In Tube 1, place 20 μL of 40 μM SBTI. Into the other 15 tubes, pipette 10 μL of PBS–Tween

(*see* **Note 19**). Next, withdraw 10 μL of SBTI from Tube 1 and mix it with the buffer in Tube 2 by pipetting up and down several times. Next, mix 10 μL from Tube 2 with the contents of Tube 3, etc., until Tube 15 is reached. Take 10 μL out of Tube 15 and discard it. Tube 16 is the 0-μM control, and thus is not part of the dilution series. Finally, add 10 μL of the 100-nM stock of labeled α-CT to each of the 16 tubes. Incubate in the dark at room temperature for 30 min.

2. Fill 16 of the preferred style of capillaries (we used Standard capillaries, as there was no difference between Premium and Standard) with the incubated samples. Each PCR tube should have about 10 μL remaining after filling; store them in the dark at room temperature. Place the capillaries into the capillary tray in numbered order, and then insert the tray into the MST instrument.
3. If using MO.Control, initialize a new experiment in the "Expert Mode" module. Provide the α-CT concentration and the desired LED level to the software (we used 50% in this example; this was optimized in **step 3** of Subheading 3.2). Set the MST power to "Low." Initiate a new dilution series and provide the top concentration of SBTI as 20,000. Manually edit the concentration for Capillary 16 to be 0.4. For the time periods of the various experimental phases, *see* **step 6** of Subheading 3.2. Start the experiment and allow it to complete. The software will automatically conduct a post-run capillary scan (*see* **Note 13**). Repeat this entire step two more times (with the same capillaries in place) but using "Medium" and "High" MST powers. Proceed to **step 5**.
4. If using NT.Control, perform a capillary scan using the LED power optimized in Subheading 3.2 and insert the appropriate concentrations for all capillaries (in nM units; Tube 1 is 20,000 in this example), except for Tube 16; divide the value in Tube 15 by approximately 3 and use this number (we inputted 0.4). Remember to input the proper concentration of the labeled α-CT, which is 50. Using the instrument parameters described in **step 6** of Subheading 3.2, collect three sequential data sets using 20%, 40%, and 60% MST power. Collect a capillary scan after each experiment (*see* **Note 13**).
5. Read the resulting data file into PALMIST, examining the 20% MST-power experiment first. Press the "Cold Fluo." button, which presents a graph of the fluorescence before the IR laser has been actuated (Fig. 4a). Is there a strong trend? If so, it is possible that the fluorescence trend and not any thermophoretic phenomenon should be analyzed (*see* **Note 20**). In the example case, there is a weak but noisy upward trend; this is

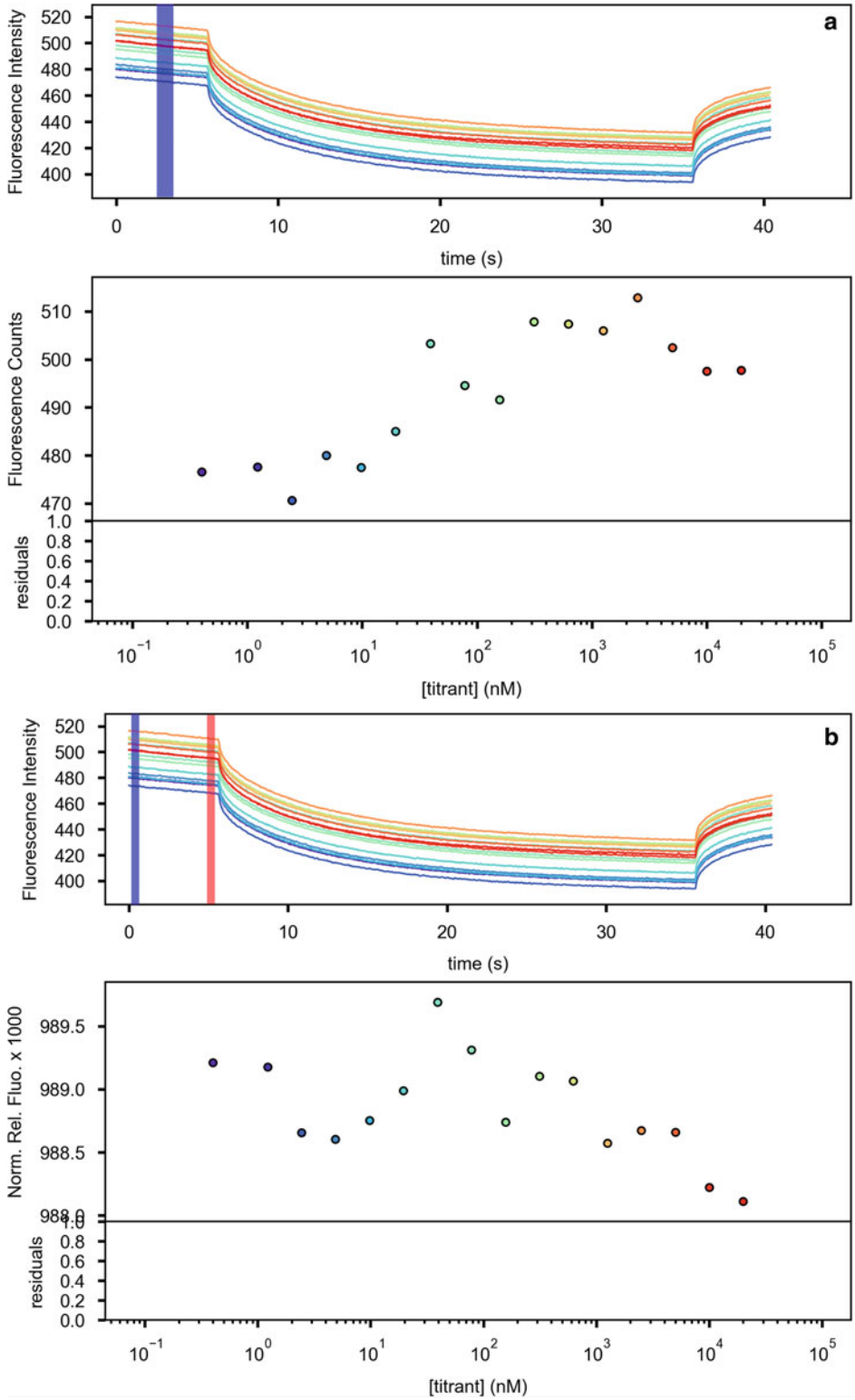


Fig. 4 PALMIST windows from the preliminary experiment. (**a**) The "Cold Fluorescence" signal. (**b**) The "Differential Bleach" signal. Coloration is as described in Fig. 3

not likely to deleteriously affect the analysis or force the analysis of the fluorescence alone, so the protocol proceeded.

6. Press the "Differential Bleach" button. This presents another pre-IR phenomenon that is related to the rate of bleaching of the illuminated fluorophore. Again, if there is a strong trend here, it is preferable to analyze this signal over thermophoresis (*see* **Note 21**). In the example case, although photobleaching is evident, no trend in the differential bleach signal is present (Fig. 4b), allowing analysis of post-IR phenomena.
7. Press the "Thermophoresis" button. This compares the fluorescence just after IR illumination and that just before IR extinguishment. Examine the data; a strong trend here can be analyzed. However, before using this signal for analysis, one should perform the next step.
8. Press the "T-Jump" button. This compares the fluorescence before IR actuation to that just after. In our experience, a strong trend here should be analyzed over that in thermophoresis alone, as K_D derived from T-Jump has yielded better agreement with K_Ds derived from other biophysical techniques [8]. In the example case, the thermophoresis signal is difficult to interpret and noisy, but the T-Jump data are essentially monotonic and sigmoid, as expected (Fig. 5a). T-Jump will be used for this example. If no trend is present here but there is one in the "Thermophoresis" mode, select Thermophoresis and proceed.

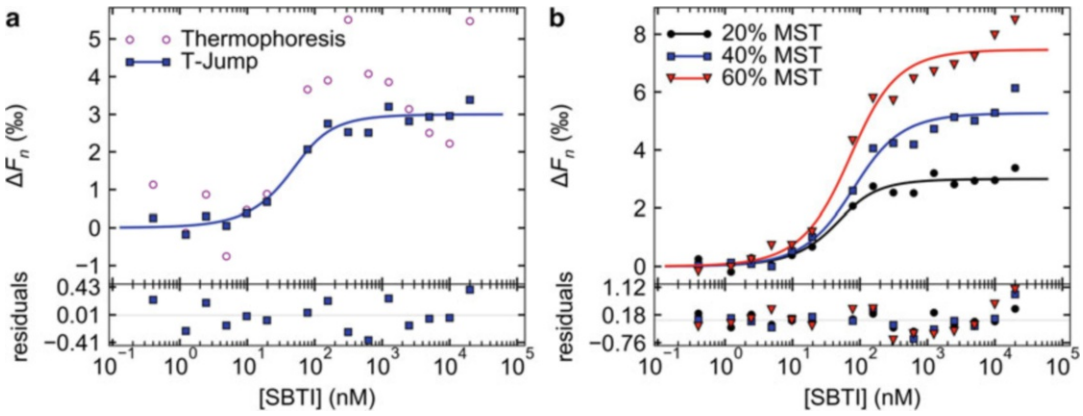


Fig. 5 Data from the preliminary experiment. (**a**) Comparison of the 20% MST data. The fit line (blue) is shown only for the T-Jump data. (**b**) Comparison of the fits for the 20%, 40%, and 60% MST-power data. See inset legend for coloration. The fitted value of F_{n,B^*} has been subtracted from all data to facilitate the comparison

Table 1
Signal-to-noise values for the pilot experiment in Subheading 3.3 (see main text for definitions of symbols)

MST power	$F_{n,B*}$	$F_{n,AB*}$	r.m.s.d.	$\widehat{F}$
20	953.7	956.7	0.212	14.15
40	924.0	929.3	0.324	16.35
60	894.8	902.2	0.442	16.74

9. Press "Fit." Define the signal-to-noise ratio $\widehat{F}$ as

$$\widehat{F} = \frac{\left|F_{\mathrm{n,AB}*} - F_{\mathrm{n,B}*}\right|}{\mathrm{r.m.s.d.}} \tag{6}$$

where $F_{\mathrm{n,AB}*}$ and $F_{\mathrm{n,B}*}$ are the refined thermophoresis (or T-Jump) values for the AB complex and the labeled protein alone, respectively, and r.m.s.d. is the root-mean-squared deviation.

10. Repeat **steps 6** and **7** of this section for the 40% and 60% MST-power data (*see* **Note 22**). For future experiments, choose the best signal-to-noise ratio; in this case (Fig. 5b), it was for the 60% MST-power experiment (Table 1; *see* **Note 23**).

3.4 Performing the MST Experiment

1. Fill 16 of the preferred style of capillaries with the reserved solutions from **step 2** of Subheading 3.3. Place in the capillary tray, perform a capillary scan, and collect a single set of data using the preferred LED power and MST power defined in Subheadings 3.2 and 3.3, and the timing parameters delineated in **step 6** of Subheading 3.2. In this case, make sure to name the experiment; an example is "Replicate #1."
2. If using the NT.Control software, actuate a post-run capillary scan. Optionally, you may save this scan. The post-run capillary scan was conducted automatically if using MO.Control. Check to ensure that the peak shapes have a symmetric, monomodal appearance (*see* **Note 13**).
3. Make the same serial dilution as described in **step 1** of Subheading 3.3. After a 30-min incubation in the dark, repeat **steps 1** and **2** of this section, naming the new experiment appropriately.
4. Repeat **step 3** of this section. This is the final replicate; i.e., the experiment has now been performed in triplicate.
5. Negative control: Make the serial dilution described in **step 1** of Subheading 3.3, but using 40 μM CAII instead of SBTI. CAII is a good choice for this control because it is an unrelated

protein of similar molecular weight (30,000 Da compared to 22,000 Da for SBTI). Use the same incubation period, parameters, and style of capillaries to collect a data set as in **steps 1** and **2** of this section, again naming the experiment appropriately.

3.5 Data Analysis

1. Load one replicate and the negative-control experiment into PALMIST (Fig. 6) and select "T-Jump" to confirm that there is no strong trend with the negative-control protein. In this case, there is a weak positive trend in the negative-control data, but we deemed it to be inconsequential for the result.
2. Load the three replicate data sets into PALMIST and press T-Jump (Fig. 7a) and examine for outlying data. In our example, the highest concentration of Experiment 1 appears to be an outlier; we excluded it by left-clicking on it. Click on the "Use Averages" checkbox at the upper right (resulting data points are in Fig. 7b). This averages all replicates (*see* **Note 24**).
3. Press "Predict" in the program's main menu at the top of the window. It projects PALMIST's default guesses onto a gray line in the binding curve graph. All that is necessary at this point is a crude match in overall appearance between the line and the data. If the curve is not matched, the fitted parameters at the right of the window can be adjusted followed by further "Predicts" to arrive at good initial guesses.
4. Press "Fit." By default, PALMIST will optimize the K_D, the $F_{n,B*}$, and the $F_{n,AB*}$ (i.e., the checkboxes next to those three parameters are checked; *see* **Note 25**). After a short pause, PALMIST will display a black fit line and the optimized parameters (Fig. 7b). Importantly, 68.3% confidence intervals are displayed in square brackets. These are the result of a rigorous

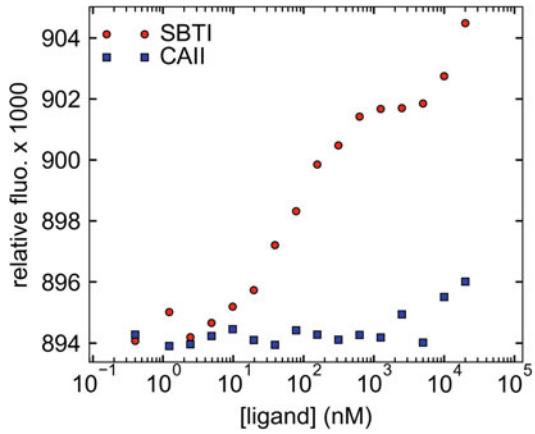


Fig. 6 T-Jump comparison for SBTI and CAII. Colors and markers are described in the inset legend

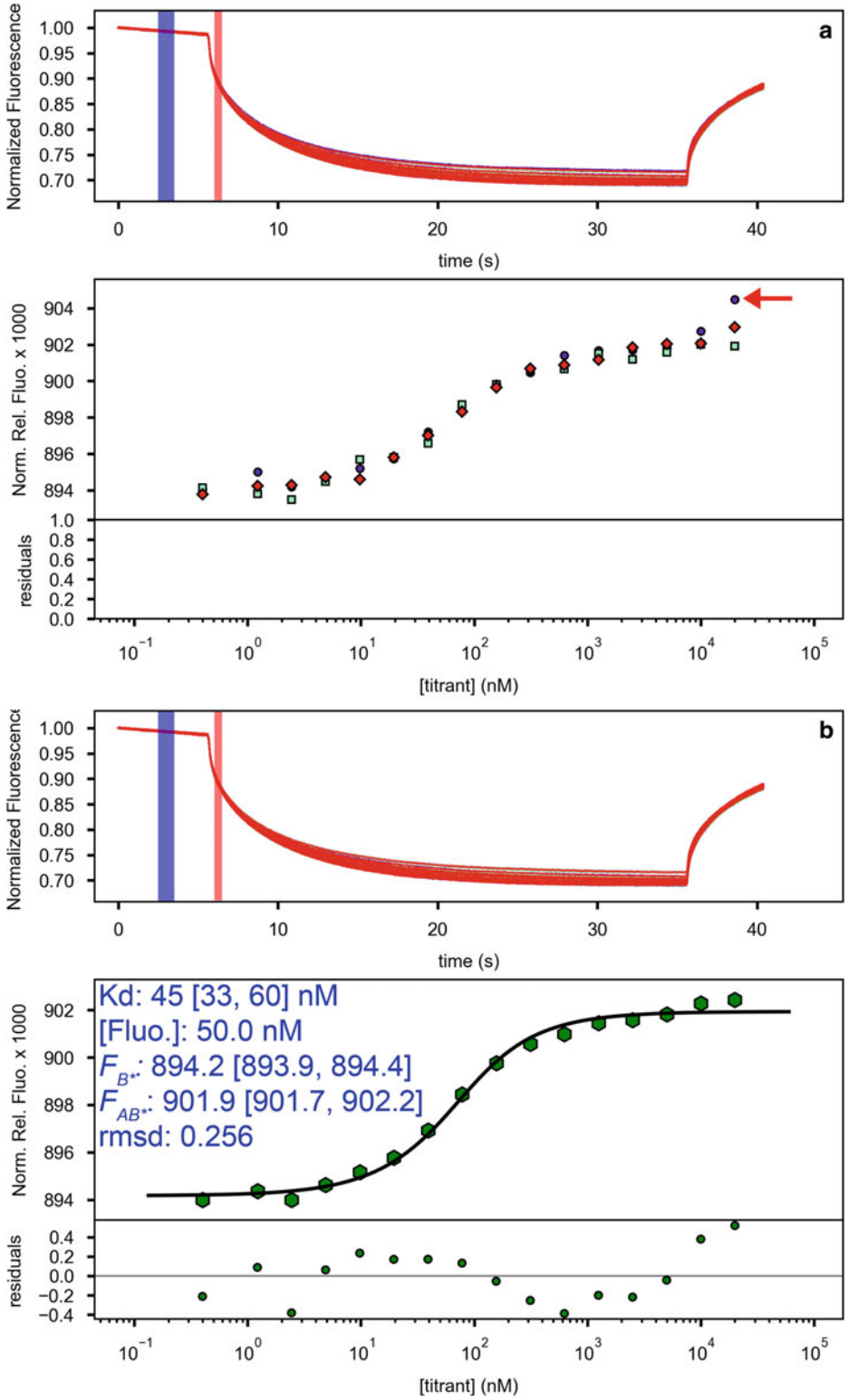


Fig. 7 MST data and fitting. (**a**) The initial view in PALMIST after loading the three replicates. Colors are violet, Replicate #1; cyan, Replicate #2; and red, Replicate #3. The highest-concentration data point in Replicate #1 (marked with a red arrow) was deemed to be an outlier and was excluded from further analysis. (**b**) After averaging and fitting. The green data points represent the average of the three replicates (except for the top concentration, where the first replicate has been excluded). The black line is the fit line. The fitted values have been rendered larger for legibility

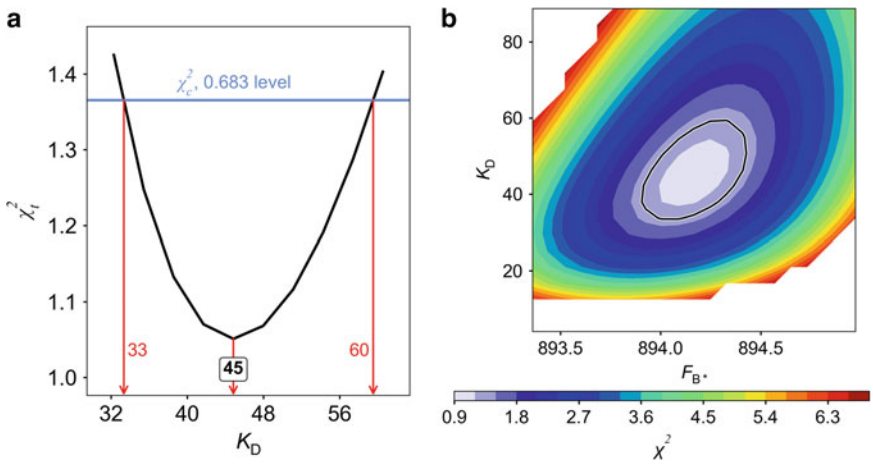


Fig. 8 Confidence intervals. (**a**) The ESP trace for the SBTI/α-CT interaction. The boxed value represents the optimized K_D. This quantity is varied and fixed in the procedure, allowing all other parameters to refine, and the χ^2 values at each step are recorded. The red arrows represent the boundaries of the interval, which are respectively labeled. Further information on the procedure is found in Refs. 8 and 11. (**b**) Here, the same process is carried out for two parameters simultaneously; the plot is contoured with χ^2 mapped to the colors as depicted in the legend. The black line represents the boundaries of the 68.3% confidence interval

search of error space (an "error-surface projection" [8, 11]) and are more accurate than the results of most fitting programs (*see* **Note 26**). If 95% confidence intervals are desired, one can select "0.95" from the "Confidence Level" section of the "Confidence Intervals" menu followed by another fitting session.

5. If desired, examine the results of the error-surface projection (Fig. 8a) by selecting "View ESP Plots" from the "Fit Docs" menu.
6. If desired, examine the two-dimensional error-surface projection of two selected parameters by selecting a pair of parameters from the upper right of the ESP view window (Fig. 8b). Uncorrelated parameters should have a symmetric oval appearance in this graph; diagonality indicates parameter correlations that should be noted.

3.6 Saving and Documenting the Analysis

1. Save the analysis by selecting "Save PALMIST State." This saves all aspects of the analysis into a ".palmpkl" file that can be recalled at a later time.
2. Optionally, it is convenient to document the result in a PDF file. Choosing "Write PDF Report" from the File menu saves a PDF file with a graph of the result, a summary of the best-fit parameters, a copy of the log detailing all steps taken since loading the data, and all ESP graphs.

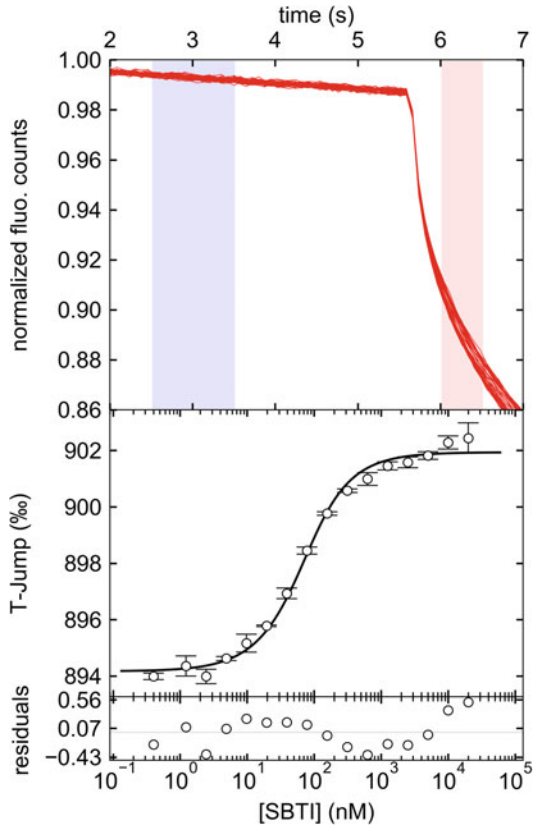


Fig. 9 Final GUSSI output. The upper panel shows the fluorescence traces; the three replicates are colored red, blue, and gold, respectively. Light blue and pink areas show the regions of the traces that were compared for the construction of the binding curve. In the middle panel is the binding curve, with circles representing the averaged data and a solid, black fit line. All data points have "error bars" showing ±1 standard deviation of the respective means. The bottom panel shows the residuals between the data and the fit line

3. Prepare a figure by choosing "Export and Start GUSSI" from the File menu. PALMIST will write a file with a .dat extension, and then GUSSI will start automatically with the file loaded. Clicking on the "Averaged" button at the right and choosing "Normalize Fluorogram" from the "Axes" menu simplifies the presentation considerably. Additional adjustments to markers, axis labels, and fluorescence-trace limits/colors result in Fig. 9 (*see* **Notes 27** and **28**).

4 Notes

1. This buffer will be used as the dilution buffer in creating dilution series (*see* Subheading 3.3, **step 1**). It must therefore

support the solubility of the ligand through a large concentration range. In the case of a small-molecule ligand, solubilizing agents (e.g., DMSO) are sometimes required, and we recommend that the concentrations of such agents be kept low (e.g., <5% DMSO) so that the protein receptor is minimally affected by them. Importantly, as formulated in this chapter, any additives included in this buffer but not in the protein buffer will be *halved* when the protein is added. Thus, if the receptor is thought to tolerate 5% DMSO, 10% could be included in this buffer.

2. We ordinarily prefer highly purified proteins for MST, i.e., >95% pure as judged by a sodium dodecyl sulfate polyacrylamide gel stained with Coomassie Blue. Indeed, with the proteins used in this protocol, our usual practice is to purify them using gel-filtration chromatography. However, we have noticed no significant difference in the results using the unpurified proteins, as described in this protocol.
3. Although spectrophotometers are available with very short path lengths and very small volumes, we do not recommend their use for this protocol. Experience has demonstrated that the best accuracy is attained by using a cuvette-based spectrophotometer.
4. The extinction coefficients were calculated from the amino-acid sequences of the respective proteins using the method of Pace and coworkers [12]. This method is encoded in the webserver ProtParam (www.expasy.org/protparam). It requires knowledge of the amino-acid sequence of the protein.
5. We recommend taking a full spectrum of the protein solutions from 250 to 400 nm. The reading at 333 nm should be doubled and subtracted from that at 280 nm to arrive at the A_{280} corrected for scattering of aggregates in solution [12].
6. Many other dyes are available; we have had success with Alexa 488 as well. In addition to NHS esters, maleimide versions of dyes may be used to label cysteine residues. Also, there are dyes that are able to bind specifically to histidine tags. The instrument's manufacturer, NanoTemper Technologies, sells a number of dyes specially designed for use in MST.
7. Both pieces of software are adequate for this protocol. Most MST instruments in the field have MO.Control, but we prefer NT.Control because of its increased flexibility for advanced users. We present instructions for both pieces of software in this chapter.
8. This dilution step may not be needed. Because of the high efficiency with which α-CT can be labeled, it is necessary to dilute the dye to prevent disruption of the SBTI/α-CT

interaction. The experimenter may need to adjust the dye concentration to achieve the desired labeling efficiency.

9. For this protocol, we used a cuvette that holds 150 μL and did not recover the measured protein solution.
10. The goal of labeling is to achieve sub-stoichiometric labeling (i.e., $\phi < 1.0$). Greater labeling efficiencies increase the likelihood that multiple lysines on the protein's surface have been labeled, which can hamper the interaction.
11. The optimal concentration of the protein is well below K_D, e.g., $0.1K_D$ or lower. However, this is difficult to know *a priori*. We have designed this experiment with a final *α-CT* of 50 nM; in this case, $K_D \sim$ *α-CT*. Although this is not ideal, Monte Carlo simulations of typical MST data demonstrate that accurate parameters can still be attained under this circumstance [8].
12. Do not use a cell phone within 5 m of the instrument. The electronic noise can cause artifacts in the fluorescence traces.
13. Aberrations in post-run capillary scans indicate that the protein is likely unstable at the higher temperature induced by the IR laser. Lower MST powers or adding stabilizing agents (e.g., glycerol) to the buffer should be explored.
14. Very high LED powers, i.e., >75%, are best avoided, as they may result in significant photobleaching of the fluorophore.
15. Protein sticking to the capillaries can be addressed by adding detergent (as in the Tween already included in this study) or a "carrier protein" to passivate sticky surfaces. Many researchers consider bovine serum albumin (BSA) to be an excellent carrier protein, but we have observed artifacts when using it [8]. If studying interactions other than the one presented hereinabove, we have found SBTI to have excellent properties for this purpose.
16. PALMIST's methods are documented in the literature [8, 9], and the software comes with a detailed manual describing its usage.
17. High-speed centrifugation, addition of detergents, and additional purification are candidate treatments for aggregation.
18. If no signal is observed, it is advisable to try other settings, e.g., "T-Jump," which compares fluorescence before the IR laser actuation to that just after. Also, it is possible to try higher MST powers or labeling with other dyes. If studying two proteins, the other binding partner could alternatively be labeled.
19. It is imperative that the buffer composition of the entire dilution series must be the same; i.e., the buffer containing the

ligand in Tube 1 must be identical to that pipetted into Tubes 2–16.

20. Strong trends in pre-IR fluorescence could be due to several phenomena. The most common is a fluorescence increase with ligand concentration, potentially indicating that the ligand has passivated sticky plasticware and thus there is more of it to detect in the high-ligand-concentration samples. The manufacturer recommends a control experiment in which mixtures with the highest and lowest ligand concentrations are centrifuged in a benchtop microcentrifuge at its maximum speed for 10 min, followed by the addition of equal volumes of a solution containing 4% (w/v) sodium dodecyl sulfate and 40 mM dithiothreitol. After incubation of these solutions at 95 °C for 10 min, they are placed into capillary tubes and a capillary scan is performed; if the difference has disappeared, this suggests that the interaction of protein and ligand was responsible for the difference and thus it is valid to use the flourescence signal for the analysis.
21. Data with a strong signal (e.g., $\widehat{F} \geq 10$; *see* Eq. 6) in the differential bleaching of the fluorophore should not be analyzed for using T-Jump or thermophoresis. Instead, the differential bleach signal can be analyzed if it can be observed reproducibly [13].
22. For this analysis, Tube 10 appeared to be an outlier, and data derived from this sample were excluded from these sample analyses. A binding-curve data point can be excluded by left-clicking on it in PALMIST.
23. Differences in opinion exist among MST practitioners concerning which MST power to use; some would see the 20% MST-power data in Table 1 and find this to be adequate. Their notion is that the lowest MST power possible should be used to minimally perturb the system. However, in systems in which T-Jump will be used for analysis, we almost always use the MST power that results in the highest $\widehat{F}$, while confirming that the MST traces are smooth, i.e., have no signs of aggregation. Because T-Jump analyzes data just after actuation of the IR laser, the effects of the high ΔT are not as deleterious as they are to thermophoresis.
24. By averaging the data, statistics on the standard deviation of F_n at each concentration are obtained. These can be used as weights in the fitting by clicking on "Use Weighted Fitting" in the panel at the right side of PALMIST. However, we do not recommend doing so; Monte Carlo simulations indicate that fitting outcomes are slightly worse when using such a fitting scheme (C.A.B., unpublished results).

25. It is strongly advised never to refine the fourth parameter in this type of analysis, i.e., concentration of the labeled protein. Its value is too correlated with the other parameters.
26. Most programs show only the square roots of the diagonal elements of the variance–covariance matrix multiplied by a factor as the "errors" in the parameters. These are usually unrealistically low.
27. As emphasized in the upper part of Fig. 9, the T-Jump signal can be very small. However, it is often very reproducible, as evidenced by the narrow error bars in the lower part of the figure.
28. GUSSI is richly featured, and it is extensively documented in its accompanying manual.

References

1. Jerabek-Willemsen M, André T, Wanner R et al (2014) MicroScale thermophoresis: interaction analysis and beyond. J Mol Struct 1077:101–113
2. Duhr S, Braun D (2006) Why molecules move along a temperature gradient. Proc Natl Acad Sci U S A 103:19678–19682
3. Baaske P, Wienken CJ, Reineck P et al (2010) Optical thermophoresis for quantifying the buffer dependence of aptamer binding. Angew Chem Int Ed 49:2238–2241
4. Wienken CJ, Baaske P, Rothbauer U et al (2010) Protein-binding assays in biological liquids using microscale thermophoresis. Nat Commun 1:100
5. Jerabek-Willemsen M, Wienken CJ, Braun D et al (2011) Molecular interaction studies using microscale thermophoresis. Assay Drug Dev Technol 9:342–353
6. Lippok S, Seidel SAI, Duhr S et al (2012) Direct detection of antibody concentration and affinity in human serum using microscale thermophoresis. Anal Chem 84:3523–3530
7. Seidel SAI, Dijkman PM, Lea WA et al (2013) Microscale thermophoresis quantifies biomolecular interactions under previously challenging conditions. Methods 59:301–315
8. Scheuermann TH, Padrick SB, Gardner KH et al (2016) On the acquisition and analysis of microscale thermophoresis data. Anal Biochem 496:79–83
9. Tso SC, Chen Q, Vishnivetskiy SA et al (2018) Using two-site binding models to analyze microscale thermophoresis data. Anal Biochem 540-541:64–75
10. Brautigam CA (2015) Calculations and publication-quality illustrations for analytical ultracentrifugation data. Methods Enzymol 562:109–134
11. Bevington PR, Robinson DK (1992) Data reduction and error analysis for the physical sciences. WCB/McGraw-Hill, Boston, MA
12. Pace CN, Vajdos F, Fee L et al (1995) How to measure and predict the molar absorption coefficient of a protein. Protein Sci 4:2411–2423
13. Fung HYJ, Fu S-C, Brautigam CA et al (2015) Structural determinants of nuclear export signal orientation in binding to exportin CRM1. elife 4:e10034

Chapter 7

Quartz Crystal Microbalance with Dissipation Monitoring (QCM-D): Preparing Functionalized Lipid Layers for the Study of Complex Protein–Ligand Interactions

Holly L. Birchenough and Thomas A. Jowitt

Abstract

Quartz crystal microbalance with dissipation monitoring (QCM-D) is one of the most widely used techniques for the deposition of lipid layers and provides a useful tool for protein–ligand analysis. By using functionalized lipids, for example, with nitrilotriacetic acid (NTA) or biotin, one can couple a molecule to the surface to investigate ligand interactions. Using lipid layers in this way allows for the analysis of complex binding events such as conformational changes, fibrillation, and hierarchical clustering on the surface, which is difficult to interpret with conventional surface sensor techniques. Deposition of lipids and subsequent molecular interactions are easily monitored using both the frequency and the dissipation, which have distinct features in bilayer formation and make QCM-D the ideal technique to use. Here we describe the formation of biotinylated lipid bilayers using two different types of lipids and the subsequent addition of avidin, which can then be used as a basis for linking biotinylated molecules to the surface. These protocols can be adapted to use other lipid moieties and linking chemistries.

Key words Lipid bilayers, QCM-D, Functionalized lipids, Biosensors

1 Introduction

The utilization of lipid supports on sensor surfaces has increased dramatically over the last few years. One reason for this is the relative ease in which one can make a well-ordered layer that can be tailored to suit the needs of the experiment. These can be bilayers [1–5], self-assembled monolayers [6, 7], or tethered bilayers, using the integration of other molecules as supports distancing the layer from the surface [8, 9]. The adaptability of the approach has led to exciting new developments in bio-nanotechnology [10, 11] and in the design of novel characterization methods for investigating molecular interactions, membrane protein–lipid interactions, and antimicrobial peptides [3, 12–14]. For the study of protein–ligand interactions, the layers can be adapted with the addition of lipids with functional head groups

Tina Daviter et al. (eds.), *Protein-Ligand Interactions: Methods and Applications*, Methods in Molecular Biology, vol. 2263, https://doi.org/10.1007/978-1-0716-1197-5_7,

that can be used to attach molecules. These can be, for example, biotin for avidin attachment (and thus biotinylated-protein attachment) or nitrilotriacetic acid (NTA) for the attachment of hexahistidine-labeled proteins. These layers can then be used for further protein–ligand interaction analysis. Lipid layers, as well as being adaptable, also have the potential to deliver a surface that is less chemically reactive, therefore reducing nonspecific background binding.

Quartz crystal microbalance with dissipation monitoring (QCM-D) is a biosensor technique well suited to the analysis of lipid bilayer deposition and the subsequent analysis of complex interactions. It is a mass sensing technique also capable of measuring the physical properties of molecules/complexes associated with the sensor surface [15, 16]. Sensors consist of a thin quartz crystal disc, often coated with SiO_2, sandwiched between a pair of electrodes. Application of a voltage via the electrodes to the quartz causes oscillation of the crystal at a specific resonant frequency. Addition of mass at the sensor surface leads to a detectable decrease in the oscillation frequency (f) of the crystal producing measurements of mass change in real time. If the layer produced by mass addition at the sensor surface is thin and rigid, then the decrease in f is proportional to the mass of the layer. In most instances, certainly with biological samples, addition of mass at the sensor surface will also include coupled water, and analysis of the "softness" of layers produced provides information on their structural properties. These structural aspects can be measured simultaneously to mass changes by measuring the dampening or dissipation (D) of the oscillation of the crystal. "Soft" or viscoelastic layers increase the dampening of the signal, resulting in an increase in dissipation readings.

QCM-D is particularly amenable for the study of lipid bilayer formation as it is able to detect the addition and loss of water during bilayer deposition [17–19]. The stages of bilayer formation on solid supports have been well characterized and are illustrated in Fig. 1, which demonstrates the multistage nature of lipid deposition [17–20]. Formation of lipid bilayers on QCM-D SiO_2 sensors follows the method pioneered by McConnell [21] and has been extensively studied previously; therefore, a well-established protocol for production is available [17, 18, 22]. With the incorporation of functional lipids in the lipid bilayer, further molecules can be attached. In this way, layers of molecules can be added to the surface so that complex interactions can be investigated. Indeed, the surfaces are large enough that the sensors can be removed and the protein/lipid layer can be extracted for western blot or silver stain analysis, which provides invaluable information about what is left on the surface after complex interactions have taken place.

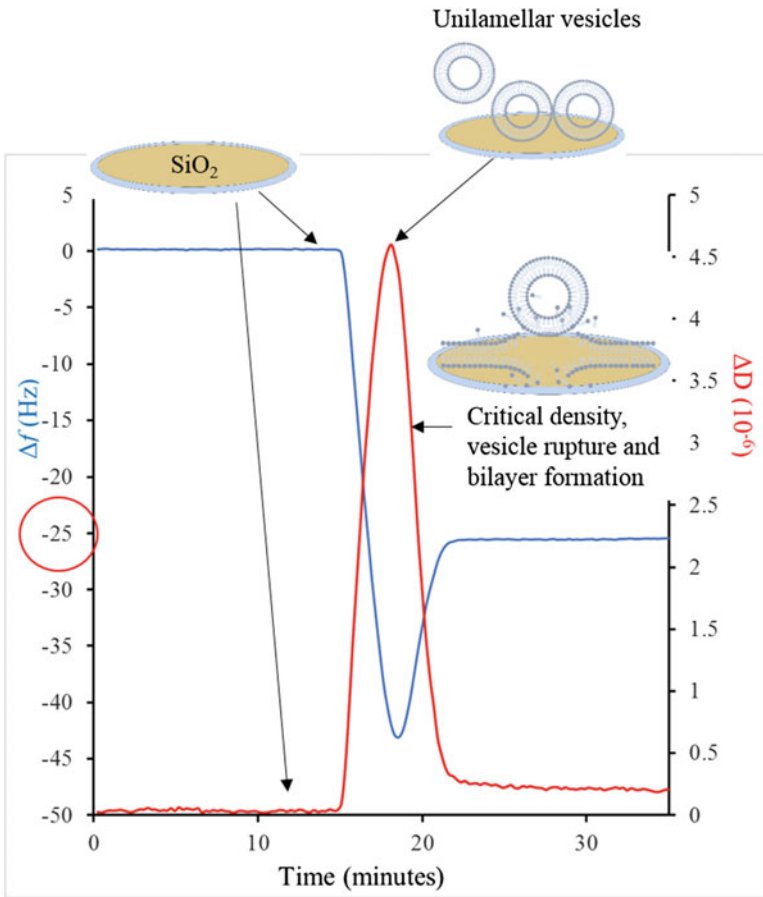


Fig. 1 Lipid bilayer formation on QCM-D SiO_2 sensors. Unilamellar vesicles first attach to the surface intact, creating a large response in frequency and dissipation due to the presence of water within the vesicles. Once critical coverage is reached, surface tension causes the liposomes to fuse and break apart, releasing water, with a subsequent increase in frequency and decrease in dissipation. Note that the final level of a fully formed bilayer is always ~−25 Hz (ringed in red)

Here we describe the method for the production of functionalized lipid bilayers on QCM-D SiO_2 sensor surfaces, which have the capability of further immobilizing molecules of interest for analysis. We describe two methods of functionalization: first, using a biotinylated lipid for direct immobilization of biotin-binding avidin/streptavidin to the lipid layer; and second, using a lipid containing a disulfide bond within a head group that can be reduced to expose a free sulfhydryl for the immobilization of avidin/streptavidin via maleimide–biotin linkers (Fig. 2). This extended approach allows for the addition of spacer arms to increase the flexibility of the immobilized protein. Once functionalized bilayers have been

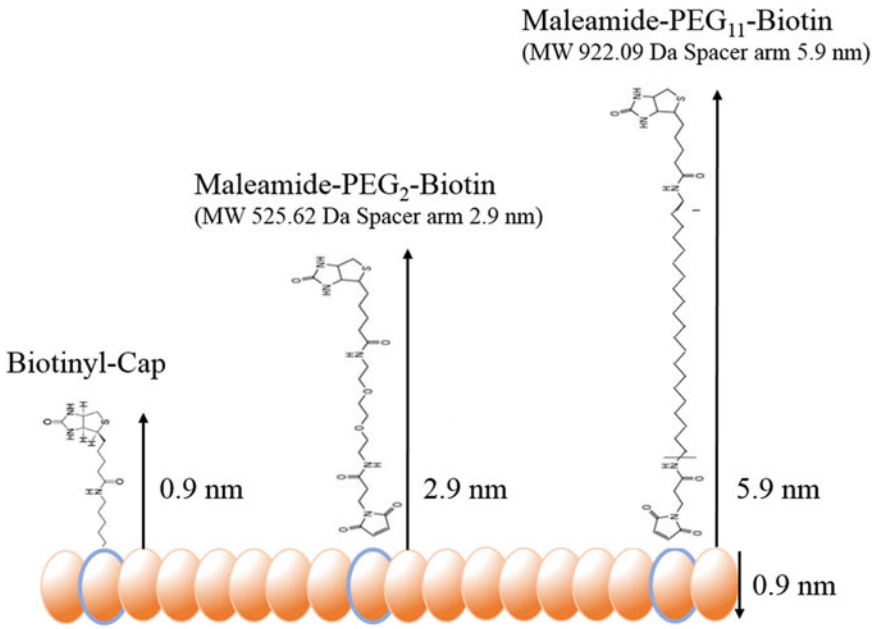


Fig. 2 The Biotinyl-Cap lipids have a 0.9-nm head group, while the PDP PE lipids can be used to attach longer spacer arms. These maleimide PEG_2 and PEG_{11} linkers provide spacers of 2.9 nm and 5.9 nm, respectively, which can prevent unwanted interactions of the molecules with the lipid surface

formed, the surface can be used to monitor more complex interactions. Figure 3 shows how we used these layers to study the interaction of a heparin-binding protein with end-biotinylated dp-24 heparin. Following bilayer formation with 2 mol% Biotinyl-Cap PE lipids, streptavidin was added to the layer followed by heparin. The addition of heparin gives a characteristically large change in dissipation caused by the viscoelastic nature of the highly solvated heparin chains arranged perpendicular to the surface as compared to the small dissipation change caused by streptavidin, which forms a compact layer. Subsequent interaction of a heparin-binding protein causes a collapse of the heparin layer caused by heparin cross-linking. Interaction of the protein to the heparin is seen by the change in frequency, which is the result of mass addition, but the sudden decrease in dissipation means that there is a significant change in surface viscoelasticity caused by a collapse of the heparin chains. This phenomenon can only be interpreted using dissipation measurements from QCM-D as the collapse of the heparin layer is hidden when using other purely mass sensing techniques, such as surface plasmon resonance (SPR), and this highlights the unique advantages of this methodology.

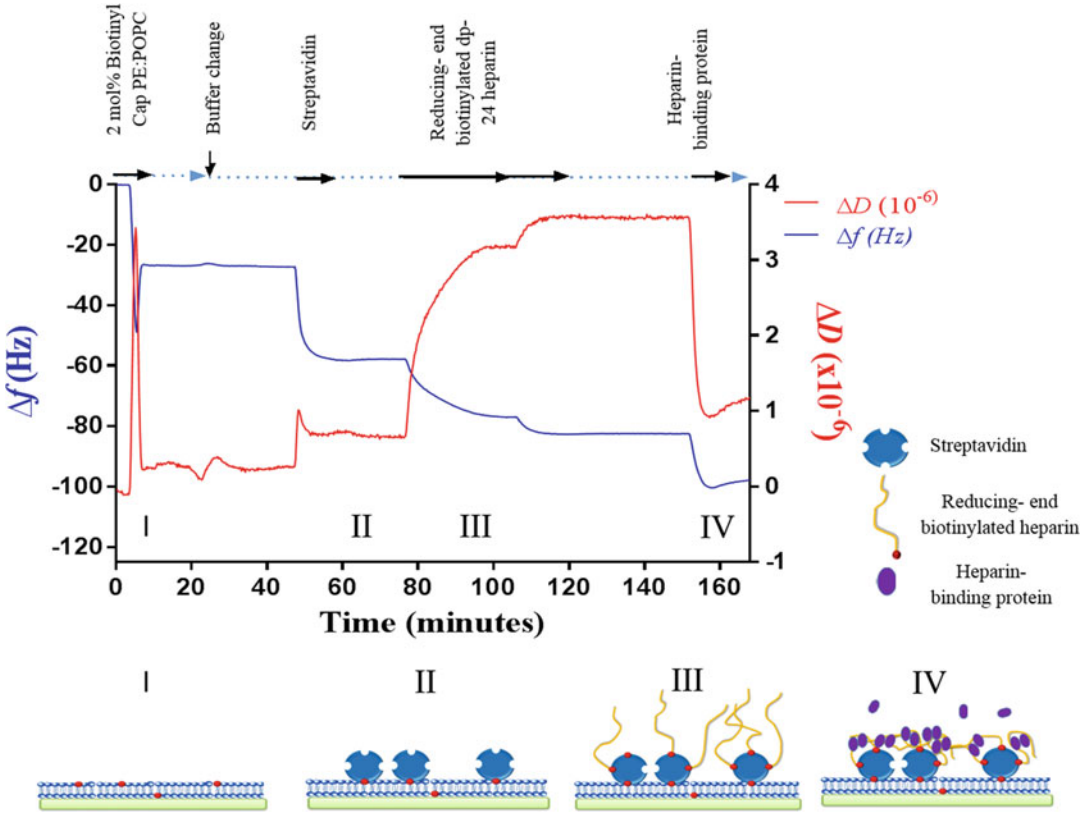


Fig. 3 Investigating the effect of binding a heparin-binding protein to a heparin layer. (I) POPC lipids with 2 mol% Biotinyl-Cap lipids are injected onto a SiO_2 QCM-D sensor prior to (II) the addition of streptavidin. After washing, biotinylated heparin is injected twice over the surface (III), which creates a highly hydrated viscoelastic layer. This layer collapses when a heparin-binding protein is injected (IV), which is seen as a decrease in the dissipation. This level of detail in binding events is impossible to interpret through conventional mass sensing techniques

2 Materials

Prepare solutions using ultrapure water (prepared by purifying deionized water, to attain a sensitivity of 18 MΩ cm at 25 °C). Prepare and store all reagents at room temperature (unless indicated otherwise). Follow all waste disposal regulations when disposing waste materials.

All reagents, unless otherwise stated, were purchased from Thermo Fisher Scientific (Loughborough, UK) or Sigma-Aldrich Corporation (Gillingham, UK). Lyophilized lipids were purchased from Avanti Polar-Lipids (Alabaster, USA) with the exception 2-oleoyl-1-palmitoyl-*sn*-glycero-3-phosphocholine (POPC), which was purchased from Sigma-Aldrich Corporation. UV Ozone cleaner ProCleaner Plus purchased from Bioforce Nanosciences (Salt Lake City, USA). Polycarbonate membranes for lipid extrusion were purchased from GE Healthcare Whatman (Little

Chalfont, UK). Mini-extruder set was purchased from Avanti Polar-Lipids (Alabaster, USA). QCM-D measurements were performed on a Q-sense E1 with a standard flow module or Omega automated system with the 8-channel standard flow module (Q-sense Ltd., Gothenburg, Sweden). QCM-D SiO_2 coated sensor crystals were also purchased from Q-sense Ltd. All QCM-D measurements were undertaken in flow mode (continuous solution delivery to the measurement chamber) using a peristaltic pump in the case of the E1.

2.1 Buffers

- **100 mM $CaCl_2$.** Weigh 1.1 g of $CaCl_2$ (anhydrous) and add to a 500-mL graduated cylinder or beaker. Add ultrapure water to 100 mL and mix using a stir bar (*see* **Note 1**).
- **10 mM HEPES, 150 mM NaCl, 2 mM $CaCl_2$, pH 7.4.** Add approximately 100 mL ultrapure water to a 1-L graduated cylinder or beaker with a stir bar. Weigh 2.38 g of HEPES and transfer to the cylinder. Weigh 8.77 g of NaCl and transfer to the cylinder. Add ultrapure water to a volume of 900 mL and mix. Add 20 mL of 100 mM $CaCl_2$ and mix. Adjust to pH 7.4 using NaOH and then add ultrapure water to a final volume of 1 L. Filter using a 0.2-μm membrane and degas. Buffer can be stored at room temperature for 1 week (*see* **Note 2**).

2.2 Proteins

- **1 mg/mL avidin/streptavidin.** Weigh 1 mg of avidin/streptavidin into a 1.7-mL tube. Add 1 mL of 10 mM HEPES, 150 mM NaCl, and 2 mM $CaCl_2$ (pH 7.4) and vortex to mix. Syringe filter at 0.2 μM. Store at 4 °C for up to 1 week. Alternatively, aliquot and freeze at −20 °C.

2.3 Reagents

- **200 mM EZ-Link maleimide-PEG_2-biotin stock.** Weigh 25 mg of EZ-Link maleimide-PEG_2-biotin into a 1.7-mL tube. Add 238 μL of DMSO and mix by pipetting. Aliquot and freeze at −20 °C. Aliquots will be stable for approximately 3 months (*see* **Note 3**).
- **200 mM EZ-Link maleimide-PEG_{11}-biotin stock.** To 25 mg EZ-Link maleimide-PEG_{11}-biotin, add 136 μL of DMSO and mix by pipetting. Aliquot and freeze at −20 °C. Aliquots will be stable for approximately 3 months.
- **250 mM tris(2-carboxyethyl)phosphine (TCEP).** Weigh 71 mg of TCEP and dissolve in 1 mL of ultrapure water in a 1.7-mL tube. Mix by pipetting and aliquot. Aliquots can be stored at −20 °C for approximately 6 months.
- **2% sodium dodecyl sulfate (SDS) (w/v).** Weigh 2 g of SDS and dissolve in 100 mL of ultrapure water. Store at room temperature.

3 Method

Bilayers of 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) were produced with head group-functionalized lipid (1,2-dipalmitoyl-*sn*-glycero-3-phosphoethanolamine-*N*-(cap biotinyl) (Biotinyl-Cap PE) or 1,2-dipalmitoyl-*sn*-glycero-3-phosphoethanolamine-*N*-[3-(2-pyridyldithio)propionate] (PDP PE) (18:1)). Here we describe the preparation of a 2-mol% head group-functionalized lipid layer. The percentage of head group-functionalized lipid can be adjusted, but percentages above 5% produce redundancy in the occupancy of the head groups.

3.1 Preparation of Multilamellar Vesicles

1. 10 mg/mL POPC stock suspension. Weigh 10 mg of lyophilized POPC within a glass bijoux with airtight lid. Dissolve in 1 mL of chloroform and swirl to homogenize.
2. 1 mg/mL head group-functionalized lipid stock suspension. Weigh 1 mg of lyophilized Biotinyl-Cap PE or PDP PE within a glass bijoux with airtight lid. Dissolve in 1 mL of chloroform and swirl to homogenize (*see* **Note 4**).
3. For a 2% head group-functionalized lipid bilayer, mix 27.4 μL of Biotinyl-Cap PE stock suspension or 23.7 μL of PDP PE stock suspension with 96.8 μL of POPC stock suspension within a flat bottomed glass bijoux (*see* **Note 5**).
4. Remove chloroform under a continuous gentle nitrogen stream to produce a thin lipid film. Ensure that all chloroform is removed by drying under vacuum overnight.
5. Rehydrate desiccated lipids in 1 mL of 0.2 μM filtered 10 mM HEPES, 150 mM NaCl, and 2 mM $CaCl_2$ (pH 7.4) (final molar concentration of 1.3 mM). Incubate at room temperature for a minimum of 30 min (*see* **Note 6**).

3.2 Lipid Extrusion to Produce Unilamellar Vesicles (See Note 7)

1. Assemble mini-extruder with a 50-nm-pore polycarbonate membrane as per the manufacturer's instructions. Ensure that filter supports have been pre-wetted with ultrapure water or buffer (*see* **Note 8**).
2. Pre-wet assembled extruder parts by passing 1 mL of buffer or ultrapure water through the extruder using gastight syringes. Discard the buffer.
3. Load hydrated lipids into gastight syringes and extrude (pass from one syringe to the other) a minimum of 19 times (*see* **Notes 9** and **10**).
4. Store unilamellar vesicle stock at 4 °C until use, and use within a week.

3.3 Cleaning and Preparation of Qsense SiO_2 Sensors (See Note 11)

1. Clean and prepare sensors immediately prior to use.
2. Sonicate sensors in 2% SDS (w/v) for approximately 15 min to remove bound particles. Rinse both sides of sensors with copious amounts of ultrapure water. Hold the sensor by the outer edge O-ring using tweezers and ensure that none of the surface or electrodes are touched (*see* **Note 12**).
3. Rinse surfaces with ethanol and dry under nitrogen stream to ensure that no streaks are present. Liquid should be chased off the crystal rather than being evaporated on it (*see* **Note 13**).
4. Place sensor in UV Ozone cleaner with SiO_2 surface facing upward. Clean for 30 min and use immediately (*see* **Note 14**).

3.4 Lipid Bilayer Formation Using QCM-D

For this method, measurements were taken using the Q-sense E1 system with external peristaltic pump as sample feeder. Adaptations to the following method can be made for use with alternative QCM-D systems.

3.4.1 Mounting Sensor in the E1 System

1. Remove flow module from chamber platform. Unscrew the flow module and remove the contact block (upper casing). Ensure that the chamber void is dry and O-ring is lying flat in its bed (*see* **Note 15**).
2. While holding the sensor by its outer edge and using tweezers, place the sensor on the chamber O-ring with the SiO_2 surface facing downward toward the chamber wall. The electrode should be facing to the left and match up with the electrode shape on the module.
3. Screw the contact block back onto the flow module ensuring not to disturb the sensor. Place in the chamber platform with the electrode pins facing downward. Replace the lid of chamber platform.
4. Connect pump and tubing to the flow module.

3.4.2 Lipid Bilayer Deposition and Measurement Procedure

1. Open the Qsoft software. Activate temperature control and set desired temperature to 20 °C. Wait for temperature equilibration (approximately 5–10 min) (*see* **Note 16**).
2. Place inlet tubing into running buffer (filtered and degassed 10 mM HEPES, 150 mM NaCl, and 2 mM $CaCl_2$ (pH 7.4) aliquoted into appropriate vessel). Run the pump at 100 μL/min and fill flow module with running buffer (*see* **Note 17**).
3. Start data acquisition and find the specific resonances 3, 5, 7, 11, and 13. Ideally wait for all resonances to be found. Viscoelastic modeling requires only three resonances, which do not need to be in order. Wait for reading to equilibrate (f and D signal is a stable plateau) which will take approximately 10–15 min (*see* **Note 18**).

4. Prepare 0.13 mM unilamellar lipid solution for bilayer deposition. Dilute 100 μL of unilamellar vesicle stock with 900 μL of running buffer in a 1.7 mL tube and mix by pipetting.
5. Temporarily stop the pump (stopping the pump between injections is not necessary in the Omega system as flow is continuous). Move the inlet tube to the unilamellar lipid solution in the 1.7-mL tube. Ensure that no air bubbles are present within the tubing. Change the pump speed to 25 μL/min and restart the pump.
6. Lipid bilayer deposition should give the characteristic signal as shown in Fig. 1 when measuring in the seventh overtone (*see* **Note 19**). Fully formed bilayers should produce final frequency and dissipation values of approximately −25 Hz and 0.1×10^{-6}, respectively (*see* Fig. 1).
7. Once bilayer has formed and signals have stabilized, stop the pump. Move the inlet tube back into the running buffer ensuring that no bubbles are present in the tubing. Change the pump speed to 100 μL/min and restart the pump. Allow the running buffer to flush through the system to ensure removal of free lipid (approximately 10 min). If preparing a Biotinyl-Cap PE surface, proceed to **step 14** (*see* **Notes 20** and **21**).
8. Prepare 10 mM TCEP by diluting 40 μL of TCEP stock in 960 μL of running buffer in a 1.7-mL tube.
9. Stop the pump. Move the inlet tube to the TCEP solution in the 1.7-mL tube. Ensure that no air bubbles are present within the tubing. Change the pump speed to 50 μL/min and restart the pump. Pump TCEP solution over the lipid surface for a minimum of 10 min or until the response has plateaued.
10. Stop the pump. Move the inlet tube back into the running buffer ensuring that no bubbles are present in the tubing. Change the pump speed to 100 μL/min and restart the pump. Allow the running buffer to flush through the system to ensure removal of free TCEP (approximately 10 min).
11. Prepare biotin linkers. Dilute 5 μL of 200 mM EZ-Link maleimide-PEG_2-biotin stock/200 mM EZ-Link maleimide-PEG_{11}-biotin stock in 995 μL of running buffer to give 1 mM final concentration.
12. Stop the pump. Move the inlet tube to the linker solution in the 1.7-mL tube. Ensure that no air bubbles are present within the tubing. Change the pump speed to 50 μL/min and restart the pump. Pump linker solution over lipid surface for a minimum of 10 min or until the response has plateaued.
13. Stop the pump. Move the inlet tube back into the running buffer ensuring that no bubbles are present in the tubing.

Change the pump speed to 100 μL/min and restart the pump. Allow the running buffer to flush through the system to ensure removal of free linker (approximately 10 min).

14. Prepare 10 μg/mL avidin/streptavidin solution. Dilute 10 μL of 1 mg/mL avidin/streptavidin stock with 990 μL of running buffer in a 1.7-mL tube and mix by pipetting.
15. Temporarily stop the pump. Move the inlet tube to the avidin/streptavidin solution in the 1.7-mL tube. Ensure that no air bubbles are present within the tubing. Change the pump speed to 50 μL/min and restart the pump. Wait for *f* and *D* signal to plateau to ensure saturation of surface (*see* **Note 22**).
16. Temporarily stop the pump. Move the inlet tube back into the running buffer ensuring that no bubbles are present in the tubing. Change the pump speed to 100 μL/min and restart the pump. Allow the running buffer to flush through the system to ensure removal of free avidin/streptavidin (approximately 15 min).
17. Surface is now functionalized and ready for the immobilization of biotinylated molecule and further interaction analysis. Repeat **step 16** with molecules of interest at appropriate concentrations.

4 Notes

1. If $CaCl_2$ used to prepare buffer is nonanhydrous/partially hydrated, ensure that the added molecular weight of water content is included in calculating mass needed for 100 mM solution. $CaCl_2$ is not always required for lipid bilayer production using zwitterionic lipids; however, we have found that using $CaCl_2$ produces consistency in results with varied lipid mixtures.
2. Buffers can be degassed in three ways. Vacuum: place buffer in a flask with a side arm attached to a vacuum pump. Add a stir bar to the flask and seal the flask at the top using a rubber stopper. Place the flask on a stir plate, and under low vacuum, stir buffer at medium speed for a minimum of 1 h. Vacuum and sonication: it is the same as the previous one; however, the stir bar/plate can be omitted and the flask under low vacuum can be placed in a sonicator water bath. Sonicate for approximately 1 h. Helium sparging: place a sparging frit to the end of a helium line and place into the buffer. Turn on the helium at a very low pressure for approximately 5 min. Always ensure that filtering is performed *prior* to degassing as the filtration process will aerate your solution.

3. EZ-Link Maleimide-PEG_n-biotin reagent is moisture sensitive. Upon use, allow the reagent to reach room temperature inside the packaging with desiccant before opening to avoid moisture condensation.
4. Stock suspensions of lipids in chloroform can be stored under nitrogen within airtight glass containers at −20 °C for approximately 3 months. To observe if evaporation has occurred during storage, make a mark on the outside of the glass container at the chloroform meniscus. Evaporation will lead to a change of lipid concentration within the solution, and volumes of lipids required in subsequent steps will need to be adjusted accordingly.
5. Alternative percentage concentrations of functionalized lipids can be used. However, concentrations over 5% do not lead to greater mass coverage of avidin and lead to redundancy of the linkers. Molar percentage greater than 10% can be detrimental to bilayer formation.
6. If alternative lipids are to be used, consideration must be given to the phase transition temperature (T_m) of the lipid. Rehydration must be performed at temperatures above the T_m of the lipid.
7. Unilamellar vesicles can also be produced from rehydrated multilamellar vesicle solutions using sonication. Investigation would be required to achieve the ideal sonication strength and length for individual sonicators to produce correctly sized unilamellar vesicles. Consideration of the T_m of the base lipid (nonfunctionalized lipid) may also limit the use of sonication. Sonication must be performed at temperatures above the T_m; therefore, sonicating on ice to prevent overheating and evaporation may not be achievable.
8. If alternative lipids are to be used, consideration must be given to the T_m of the base lipid (nonfunctionalized lipid). Extrusion should be performed at a temperature above the T_m of the base lipid. This can be achieved by placing the mini-extruder on a hot plate and allowing approximately 15 min for the extruder to reach the required temperature before use.
9. Use the same syringe used to take up buffer in the previous step to load lipid sample. It is also important to always pass lipids through the membrane an odd number of times. This ensures that any unwanted particulates are filtered by the membrane and are not in the final extruded lipid solution. Lipid suspension should change from cloudy to clear following extrusion.
10. Unilamellar vesicle size can be checked using dynamic light scattering. Vesicles should be ~80 nm and homogeneous when extruded through a 50-nm pore size filter. Vesicles considerably below this size should be discarded. If vesicles are larger,

further extrusion steps or a longer period of sonication can be performed. Potentially, extrusion could also be performed with a 30-nm-pore polycarbonate membrane, which we have found to produce vesicles of approximately 50–60 nm. Lipid suspension may also benefit from freeze-thawing (4× cycles of placing the sample in liquid nitrogen and then 50 °C water bath) posthydration and prior to extrusion/sonication.

11. If cleaned and stored correctly, SiO_2 Qsense sensors can be reused 10–30 times.
12. We would recommend using reverse action tweezers. The tweezers should have a round end and smooth gripping surfaces to avoid damage to sensors. Hold the sensors from underneath to prevent washing contaminants from the tweezers onto the sensor surface.
13. Liquid may remain on the edge of the sensor under the tweezers. This can be wicked off with a clean, lint-free tissue.
14. Sensors should be used for lipid deposition immediately after UV Ozone cleaning. This is due to the preparation step increasing the charge on the sensor surface. Sensors can be stored dry at the post ethanol clean step but would require ethanol rinse and nitrogen dry again to ensure no dust on the surface. Proceed with UV Ozone clean.
15. If liquid is present in the chamber void or under the O-ring, this can be dried with a lint-free tissue.
16. If alternative lipids are to be used, then the temperature needs to be set above the lipid T_m. Once bilayer has formed, the temperature can be returned to 20 °C; however, this will affect the fluidity of the formed bilayer.
17. To avoid disturbances in f and D, it is vital to use the same stock of 10 mM HEPES, 150 mM NaCl, and 2 mM $CaCl_2$ (pH 7.4) for running buffer and all subsequent sample dilution steps.
18. Drift should be around 2 Hz/h (f) and 0.2×10^{-6} (D). Noise levels should be lower than 0.6 Hz (peak-to-peak) and 0.15×10^{-6} (peak-to-peak), respectively, when measured within a period of 2 min. If the sensor is failing to equilibrate, this could indicate dirty tubing, unfiltered/old buffer, or a dirty sensor. Buffer should be made fresh. System can be flushed with 2% SDS, ultrapure water, ethanol, and a final rinse with ultrapure water. Repeat cleaning and preparation procedure for Qsense sensors. If following these steps the signal still fails to stabilize during equilibration, then the sensor may be faulty and will need to be replaced. Alternatively, the flow module may require cleaning, which can be performed as per the manufacturer's instructions.

19. If lipid deposition does not show the responses demonstrated here, this indicates that a full lipid bilayer may not have been formed. A large decrease in f (and high D), which plateaus, indicates intact liposomes that have not popped to form a bilayer. This is due to liposomes that are too large, forgetting to add calcium or dirty sensor surfaces. Further extrusion or sonication would be required (*see* **Note 10**). It is also possible that the surface charge has been lost and so the sensor cleaning and preparation steps should be repeated. A decrease in f to approximately −25 Hz without the dip indicates that the vesicles are popping as soon as they reach the surface and no liposome critical coverage has been achieved before popping. This can give incomplete bilayers and is due to vesicles that are too small. Unilamellar vesicle preparation should be repeated. If the return in frequency to −25 Hz is slow (inverse peak is shifted to the right) and D signal is high, this could indicate that the lipid layer contains some vesicles that have not popped. This is due to a nonhomogeneous vesicle preparation that includes larger liposomes. Further extrusion or sonication would be required (*see* **Note 10**).
20. As a measure to check completeness of layer coverage and passivation, 50 μg/mL BSA in running buffer can be flowed over the surface at 50 μL/min for 10 min. A passivated layer should not show changes in f or D.
21. Once bilayer has formed, the running buffer can be changed at this point, e.g., to exclude $CaCl_2$. A change in running buffer will cause a change in f and D and so a period of equilibration will be required to achieve a stable signal again. If proceeding with the maleimide linker protocol, then ensure that running buffer is sulfhydryl-free.
22. It is possible to control the level of biotinylated molecule to be subsequently immobilized to the surface by controlling the amount of avidin/streptavidin immobilized to the lipid surface. This can be achieved by reducing the time and/or concentration of avidin/streptavidin introduced to the surface.

Acknowledgments

Figures were produced using BioRender with thanks to Maryline Fresquet and The Wellcome Trust for funding.

References

1. Dubacheva GV, Araya-Callis C, Geert Volbeda A, Fairhead M, Codée J, Howarth M, Richter RP (2017) Controlling multivalent binding through surface chemistry: model study on streptavidin. J Am Chem Soc 139:4157–4167. https://doi.org/10.1021/jacs.7b00540
2. Sumino A, Dewa T, Takeuchi T, Sugiura R, Sasaki N, Misawa N, Tero R, Urisu T, Gardiner AT, Cogdell RJ, Hashimoto H, Nango M (2011) Construction and structural analysis of tethered lipid bilayer containing photosynthetic antenna proteins for functional analysis. Biomacromolecules 12:2850–2858. https://doi.org/10.1021/bm200585y
3. Forbes S, McBain AJ, Felton-Smith S, Jowitt TA, Birchenough HL, Dobson CB (2013) Comparative surface antimicrobial properties of synthetic biocides and novel human apolipoprotein E derived antimicrobial peptides. Biomaterials 34:5453–5464. https://doi.org/10.1016/j.biomaterials.2013.03.087
4. Beutel O, Roder F, Birkholz O, Rickert C, Steinhoff HJ, Grzybek M, Coskun Ü, Piehler J (2015) Two-dimensional trap for ultrasensitive quantification of transient protein interactions. ACS Nano 9:9783–9791. https://doi.org/10.1021/acsnano.5b02696
5. Meléndrez D, Jowitt T, Iliut M, Verre AF, Goodwin S, Vijayaraghavan A (2018) Adsorption and binding dynamics of graphene-supported phospholipid membranes using the QCM-D technique. Nanoscale 10:2555–2567. https://doi.org/10.1039/c7nr05639g
6. Barbosa JN, Martins MCL, Freitas SC, Gonçalves IC, Águas AP, Barbosa MA (2010) Adhesion of human leukocytes on mixtures of hydroxyl- and methyl-terminated self-assembled monolayers: effect of blood protein adsorption. J Biomed Mater Res A 93:12–19. https://doi.org/10.1002/jbm.a.32480
7. Ballav N, Terfort A, Zharnikov M (2009) Fabrication of mixed self-assembled monolayers designed for avidin immobilization by irradiation promoted exchange reaction. Langmuir 25:9189–9196. https://doi.org/10.1021/la9007476
8. Budvytyte R, Mickevicius M, Vanderah DJ, Heinrich F, Valincius G (2013) Modification of tethered bilayers by phospholipid exchange with vesicles. Langmuir 29:4320–4327. https://doi.org/10.1021/la304613a
9. Budvytyte R, Valincius G, Niaura G, Voiciuk V, Mickevicius M, Chapman H, Goh HZ, Shekhar P, Heinrich F, Shenoy S, Lösche M, Vanderah DJ (2013) Structure and properties of tethered bilayer lipid membranes with unsaturated anchor molecules. Langmuir 29:8645–8656. https://doi.org/10.1021/la401132c
10. Albrecht T (2011) Nanobiotechnology: a new look for nanopore sensing. Nat Nanotechnol 6:195–196. https://doi.org/10.1038/nnano.2011.52
11. Yusko EC, Johnson JM, Majd S, Prangkio P, Rollings RC, Li J, Yang J, Mayer M (2011) Controlling protein translocation through nanopores with bio-inspired fluid walls. Nat Nanotechnol 6:253–260. https://doi.org/10.1038/nnano.2011.12
12. Hall K, Lee TH, Mechler AI, Swann MJ, Aguilar MI (2014) Real-time measurement of membrane conformational states induced by antimicrobial peptides: balance between recovery and lysis. Sci Rep 4:1–9. https://doi.org/10.1038/srep05479
13. Hirst DJ, Lee TH, Swann MJ, Aguilar MI (2013) Combined mass and structural kinetic analysis of multistate antimicrobial peptide-membrane interactions. Anal Chem 85:9296–9304. https://doi.org/10.1021/ac402148v
14. Huang CJ, Cho NJ, Hsu CJ, Tseng PY, Frank CW, Chang YC (2010) Type i collagen-functionalized supported lipid bilayer as a cell culture platform. Biomacromolecules 11:1231–1240. https://doi.org/10.1021/bm901445r
15. Dixon MC (2008) Quartz crystal microbalance with dissipation monitoring: enabling real-time characterization of biological materials and their interactions. J Biomol Tech 19:151–158
16. Reviakine I, Johannsmann D, Richter RP (2011) Hearing what you cannot see and visualizing what you hear: interpreting quartz crystal microbalance data from solvated interfaces. Anal Chem 83:8838–8848. https://doi.org/10.1021/ac201778h
17. Richter RP, Brisson AR (2005) Following the formation of supported lipid bilayers on mica: a study combining AFM, QCM-D, and ellipsometry. Biophys J 88:3422–3433. https://doi.org/10.1529/biophysj.104.053728
18. Richter RP, Bérat R, Brisson AR (2006) Formation of solid-supported lipid bilayers: an integrated view. Langmuir 22:3497–3505. https://doi.org/10.1021/la052687c
19. Reimhult E, Zäch M, Höök F, Kasemo B (2006) A multitechnique study of liposome adsorption on Au and lipid bilayer formation

on SiO2. Langmuir 22:3313–3319. https://doi.org/10.1021/la0519554

20. Zhu L, Gregurec D, Reviakine I (2013) Nanoscale departures: excess lipid leaving the surface during supported lipid bilayer formation. Langmuir 29:15283–15292. https://doi.org/10.1021/la401354j
21. McConnell HM, Watts TH, Weis RM, Brian AA (1986) Supported planar membranes in studies of cell-cell recognition in the immune system. Biochim Biophys Acta 864:95–106. https://doi.org/10.1016/0304-4157(86)90016-X
22. Cho NJ, Frank CW, Kasemo B, Höök F (2010) Quartz crystal microbalance with dissipation monitoring of supported lipid bilayers on various substrates. Nat Protoc 5:1096–1106. https://doi.org/10.1038/nprot.2010.65

Part III

Screening for Ligand Binding

Chapter 8

Indirect Detection of Ligand Binding by Thermal Melt Analysis

Joseph Shaw and Christopher Stubbs

Abstract

A thermal shift assay (TSA) involves measuring the effect of a compound on the thermal stability of a protein as an indirect measure of ligand binding. In this chapter, we provide a protocol for a conventional TSA with recombinant/purified proteins using differential scanning fluorimetry (DSF), followed by a protocol for a Cellular Thermal Shift Assay (CETSA®), which measures the soluble cellular protein remaining after a transient heat shock of live cells to detect intracellular ligand binding.

Key words TSA, Thermal shift assay, Ligand binding, Target engagement, DSF, Differential scanning fluorimetry, CETSA, Cellular Thermal Shift Assay

1 Introduction

Binding of a ligand to the native, folded state of a protein results in increased protein stability. This often results in an increase in the melting temperature of the protein, and it is this effect that is measured in a thermal shift assay (TSA).

Differential scanning fluorimetry (DSF) is a widely used biophysical technique that can be used to perform a TSA [1–3]. In a DSF experiment, purified protein (typically recombinant) is mixed with an environmentally sensitive dye (e.g., SYPRO® Orange), whose fluorescence increases with the concentration of unfolded protein in the mixture. The protein/dye mixture is then subjected to a temperature ramp in an RT-PCR instrument, where dye fluorescence intensity is recorded as a function of temperature, resulting in a protein melt curve. The midpoint of the melt curve is defined as the protein melting temperature (T_m, Fig. 1a). In a DSF TSA, the experiment is performed in the absence and presence of ligands, and a thermal shift ΔT_m is calculated (by subtracting the T_m of the protein alone from the T_m in the presence of the ligand). Ligands that induce a ΔT_m of ≥ 1 °C are usually considered as "hits."

Tina Daviter et al. (eds.), *Protein-Ligand Interactions: Methods and Applications*, Methods in Molecular Biology, vol. 2263, https://doi.org/10.1007/978-1-0716-1197-5_8,

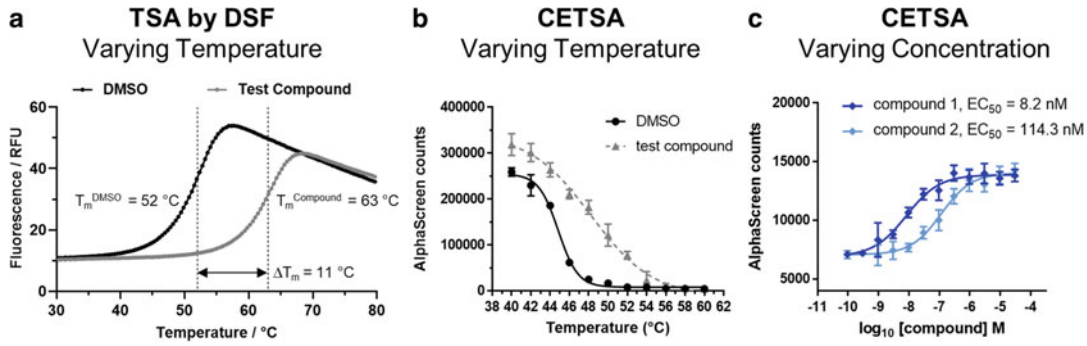


Fig. 1 Different experimental setups for thermal shift assays (TSA) on purified proteins by differential scanning fluorimetry (DSF), and a Cellular Thermal Shift Assay (CETSA). (**a**) In DSF, each sample is ramped through increasing temperatures over time. The melting temperature (T_m) is defined as the midpoint of the melt curve, and the thermal shift (ΔT_m) is the difference between the T_m of the protein in the presence of compound and the protein alone. (**b**) CETSA can be performed using a similar setup to the DSF experiment, except a different sample must be used for each heat shock temperature. (**c**) CETSA can also be performed at a single, optimized temperature where multiple concentrations of compound are tested in parallel

Ligand-induced thermal stabilization is also known to occur within the context of the intracellular environment, which can be detected in a Cellular Thermal Shift Assay (CETSA®) [4, 5]. Here, ligands are incubated with live cells before subjecting cells to a transient heat shock, after which the remaining soluble protein is quantified. CETSA® experiments can be performed with a single concentration of compound tested across multiple temperatures (Fig. 1b) or in a dose-response format at a single temperature optimized for the protein of interest (Fig. 1c). This chapter discusses high-throughput CETSA® (CETSA® HT), which uses AlphaScreen® technology [6] to quantify the remaining soluble protein after heat shock. In this format, a pair of antibodies directed at orthogonal epitopes on the target protein are used to form a ternary complex that can be detected by suitable donor and acceptor AlphaScreen® beads [7].

We have developed CETSA HT assays for several targets using a standardized process of antibody screening and optimization and quantitatively assessed intracellular binding to a target using a standardized, semi-automated workflow [8, 9]. Successful generation of a CETSA HT assay for a particular target is dependent on the availability of three key reagents:

1. A suitable cell line as a source of cellular target protein.
2. Multiple available antibodies to the target protein, with a minimum requirement of two different antibodies directed to two different epitopes and generated in two different species.
3. A well-validated tool compound known to bind the target protein in cells.

There are also two key scientific observations as part of assay development:

1. The observation of sufficient AlphaScreen signal from optimized antibody pairs to quantify target protein from the cell line of interest.
2. The observation of a sufficiently large change in the thermal stability of the target protein upon binding of the tool compound in cells.

2 Materials

All solutions should be prepared using ultrapure water (deionized water purified to achieve a resistivity of 18.2 MΩ/cm at 25 °C) and analytical-grade reagents. Reagents should be stored as recommended by their suppliers. SYPRO® Orange (and other fluorescent dyes) and AlphaScreen® reagents must be stored in the dark. Protein reagents should generally be thawed quickly and then kept on ice until use.

2.1 Optimization of Conditions and Screening of Compounds by DSF Experiments

1. DSF buffer: target dependent, e.g., 50 mM HEPES-Na (pH 7.5), 100 mM NaCl (*see* **Note 1**).
2. Purified protein target of interest: commercially available or purified in-house. Typically stored at −80 °C at 10–100 μM, aliquoted to minimize freeze–thaw cycles.
3. Liquid dispenser, e.g., D300 liquid dispenser (Tecan).
4. 5000× SYPRO® Orange (*see* **Note 2**).
5. DMSO (100% v/v).
6. 384-well PCR plate, e.g., 4titude FrameStar 384 well skirted PCR plate compatible with real-time PCR amplification and detection instrument and thermosealing tape (4titude).
7. A real-time PCR amplification and detection instrument, e.g., LightCycler480 II (Roche).

2.2 Screening of Compounds by DSF

1. As in Subheading 2.1.
2. Test compounds as 10–100 mM stocks in DMSO (100% v/v).

2.3 Quantitative AlphaScreen Assay to Measure Soluble Target Protein for CETSA

1. Source several mouse-derived and rabbit-derived antibodies directed to the target protein. Antibodies should be chosen to maximize selective recognition of the desired target (*see* **Note 3**).
2. 5× SureFire Lysis Buffer (PerkinElmer).
3. AlphaScreen compatible white, shallow 384-well microplate, e.g., ProxiPlate-384 Plus (PerkinElmer).

4. ImmunoAssay Buffer, 10× (PerkinElmer).
5. Anti-Mouse IgG Alpha Donor beads (PerkinElmer).
6. Anti-Rabbit IgG (Fc specific) AlphaLISA Acceptor beads (PerkinElmer).
7. AlphaScreen enabled plate reader, e.g., EnVision (PerkinElmer).

2.4 Thermal Melt Analysis in Live Cells by CETSA

1. Cell culture flasks, e.g., T175 flask (ThermoScientific).
2. PCR strip tubes, e.g., MicroAmp® 8-Tube Strip, 0.2-mL PCR tubes (ThermoFisher).
3. Strip caps for PCR strip tubes, e.g., MicroAmp® 8-Cap Strip, Clear (ThermoFisher).
4. PCR strip-compatible microcentrifuge, e.g., MiniStar Silverline (VWR).
5. Multi-block thermal cycler, e.g., SimpliAmp™ Thermal Cycler (ThermoFisher).
6. Electronic adjustable tip spacing multichannel equalizer pipette, e.g., E1-ClipTip™ (ThermoFisher).
7. Pipette Tips, e.g., 125-μL tips for ClipTip™ 384 (ThermoFisher).
8. Centrifuge capable of spinning microplates, e.g., Heraeus Multifuge 1S centrifuge.
9. A well-validated tool compound known to bind the target protein in cells (*see* **Note 4**).

2.5 Quantification of Thermostable Target Protein by AlphaScreen

1. As in Subheading 2.3.
2. AlphaScreen enabled plate reader, e.g., EnVision (PerkinElmer).

2.6 Screening of Compounds at a Single Optimized Temperature by CETSA HT

1. 384-well PCR plate, e.g., 4titude FrameStar 384 well skirted PCR plate compatible with real-time PCR amplification and detection instrument and thermosealing tape (4titude).
2. Liquid handling device capable of dispensing nanoliter volumes, e.g., Echo 555 acoustic dispenser (LabCyte) or D300 liquid dispenser (Tecan).
3. Liquid dispenser capable of dispensing microliter volumes, e.g., Multidrop Combi and small cassette (ThermoScientific).
4. 384-well PCR machine, e.g., LightCycler480 II (Roche).
5. Liquid handing platform such as the Bravo liquid handler (Agilent).
6. Tips, 10 μL, 384 in rack, compatible with the liquid handling.
7. As in Subheading 2.5.

3 Methods

Perform all steps at room temperature unless otherwise indicated. Ensure appropriate waste disposal throughout and appropriate safety measures when using automated liquid handling devices.

3.1 Optimization of Conditions for DSF Experiments

The concentrations of protein and SYPRO® Orange dye must be optimized for the protein of interest in order to maximize reproducibility while minimizing reagent consumption. Figure 2a depicts the plate layout and exemplar results for the optimization experiment described below.

1. Prepare protein solutions at 1.25 μM, 2.5 μM, 5 μM, and 10 μM in DSF buffer by adding 300 μL of 10 μM protein in DSF buffer to a 0.5-mL polypropylene tube. Mix 150 μL of this solution with 150 μL of DSF buffer in a fresh tube. Take 150 μL of the resulting solution and mix with 150 μL of DSF buffer in another tube. Repeat once more.
2. To a PCR instrument compatible 384-well microplate, e.g., 4titude FrameStar 384 well skirted PCR plate, add 10 μL of 1.25 μM protein to wells A1–A12, 2.5 μM protein to wells B1–B12, 5 μM protein to wells C1–C12, and 10 μM protein to wells D1–D12 (*see* **Note 5**).
3. Briefly centrifuge the plate at 300 × *g* to ensure that the solution is at the bottom of the wells.
4. Add DMSO and SYPRO® Orange to the PCR plate using a liquid dispenser, e.g., D300 (Tecan), as shown in Table 1.
5. Seal the plate with the optical sealing film and centrifuge the plate at 300 × *g* for 1 min.
6. Precool the real-time PCR instrument by running a method to hold the block temperature at 20 °C for 5 min with a blank plate (*see* **Note 6**).
7. When the precooling method finishes, immediately insert the prepared PCR plate and run a method that will hold the temperature at 20 °C for 30 s and then run a temperature gradient from 20 to 95 °C at 1 °C/min. Measure the fluorescence throughout the gradient (at least 1 acquisition per 1 °C) using filters compatible with SYPRO® Orange ($\lambda_{ex} = 470$ nm, $\lambda_{em} = 570$ nm, *see* **Note 7**).
8. Visually inspect the melt curves to choose the optimal protein/dye combination for screening, considering reproducibility and the signal/background ratio (*see* Fig. 2b for more detail and *see* **Note 8**).
9. Assess across-plate reproducibility of screening conditions for DSF experiment. If protein is not limiting, prepare 4 mL of

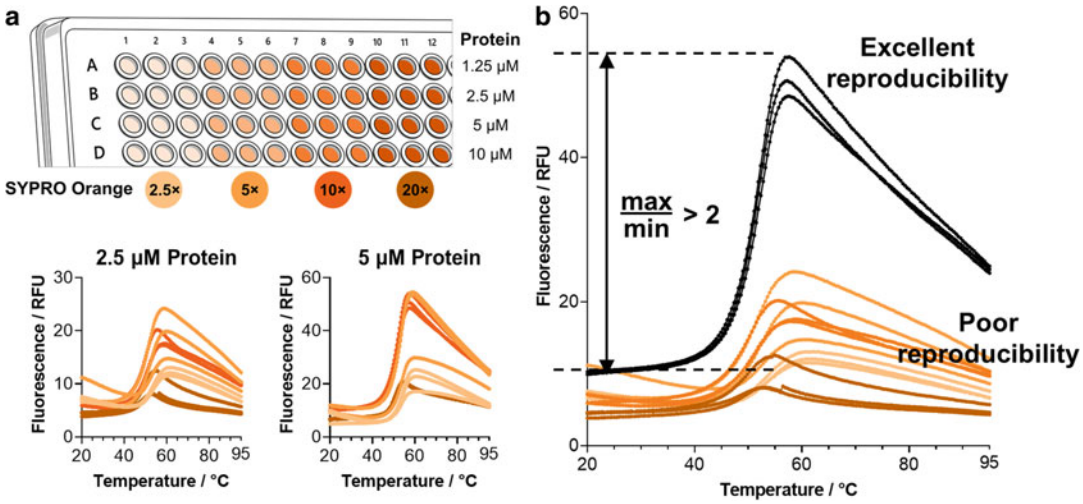


Fig. 2 Optimization of conditions for the DSF experiment. (**a**) The optimization matrix is designed such that protein is titrated down columns and SYPRO® Orange is titrated across rows. Each condition has three replicates. Real data for two protein concentrations are shown, where darker color indicates higher SYPRO® Orange concentration. (**b**) In black, the curves for 5 μM protein and 10× SYPRO® Orange from the optimization experiment in (**a**) are shown. These curves show excellent reproducibility and a high signal/background (max/min) ratio. The other curves shown are those from the 2.5 μM protein samples. These show much worse reproducibility and would be less suitable for screening

Table 1
DSF optimization matrix. Components to be added to the relevant columns of rows A–D in the 384-well PCR plate

	Columns 1–3	Columns 4–6	Columns 7–9	Columns 10–12
Protein solution	10 μL	10 μL	10 μL	10 μL
5000× SYPRO® Orange	5 nL	10 nL	20 nL	40 nL
DMSO	100 nL	100 nL	100 nL	100 nL
Final well contents	2.5× SYPRO® 1.04% (v/v) DMSO	5× SYPRO® 1.09% (v/v) DMSO	10× SYPRO® 1.19% (v/v) DMSO	20× SYPRO® 1.38% (v/v) DMSO

protein at the concentration determined in **step 8** and add 10 μL to all wells of the 384-well PCR plate. If protein is limiting, prepare 120–240 μL of protein in DSF buffer at the concentration determined in **step 8** and add to 10–20 wells distributed across a 384-well PCR plate (e.g., A1, A9, A17, A24, D4, D21, F17, G7, H1, H12, H13, I12, I13, I24, J17, K7, M4, M21, P1, P9, P17, and P24).

10. Seal the plate with optical sealing film and centrifuge the plate at 300 × *g* for 1 min.

11. Add SYPRO® Orange at the optimal concentration determined in **step 8** and 100 nL DMSO to every well on the plate.
12. Seal the plate with optical sealing film and centrifuge the plate at 300 × *g* for 1 min.
13. Place the plate into the precooled real-time PCR instrument and ramp the temperature from 20 to 95 °C at 1 °C/min while measuring the fluorescence using filters compatible with SYPRO® Orange as in **step 7**.
14. Determine the melting temperature of the protein for each well (*see* **Note 9**) and then calculate the mean melting temperature and its standard deviation. For a robust DSF screening assay, the standard deviation should be ≤0.3 °C. If the standard deviation of the mean melting temperature is significantly greater than this, try different buffers and/or cofactors. If the melting temperature is <35 °C, it may help to keep the plate chilled as much as possible during handling and, if possible, start the temperature ramp from a lower temperature.

3.2 Screening of Compounds by DSF

1. Prepare the screening plate using liquid handling, e.g., an ECHO 555 acoustic dispenser to add 100 nL of DMSO to Columns 1 and 23 and 100 nL of 10 mM positive control (in DMSO) to Columns 2 and 24 of a 384-well PCR plate. Fill the remaining wells of the plate with 100 nL of 10 mM test compounds in DMSO (up to 320 compounds). If a positive control is unavailable, just add 100 nL DMSO to use these wells as extra DMSO controls.
2. Seal the plate and centrifuge at 300 × *g* for 1 min.
3. Prepare 4 mL of protein in DSF buffer at the concentration determined in **step 8** (Subheading 3.1) and add 10 μL to all wells.
4. Add the volume of 5000× SYPRO® Orange determined in **step 8** in Subheading 3.1 to all wells on the plate using a liquid dispenser, e.g., D300 (Tecan).
5. Seal the plate and centrifuge at 300 × *g* for 1 min.
6. Place the plate into the real-time PCR instrument and ramp the temperature from 20 to 95 °C at 1 °C/min while measuring the fluorescence using filters compatible with SYPRO® Orange.
7. Determine the melting temperature of the protein in each swell (*see* **Note 9**).
8. Calculate the thermal shift (ΔT_m) of the test compounds by subtracting the mean melting temperature of the DMSO only wells from the melting temperature of the protein in the presence of the test compound. A hit is typically defined as a compound inducing a $\Delta T_m \geq 1$ °C or ≥3× the standard deviation of the DMSO only melting temperature, whichever is greater.

3.3 Quantitative AlphaScreen Assay to Measure Soluble Target Protein for CETSA

1. Dilute 0.6 mL of 5× SureFire Lysis Buffer in 2.4 mL of dH_2O to give 3 mL of 1× SureFire Lysis Buffer. Harvest cells (*see* **Note 10**) and pellet before resuspending in 1× SureFire Lysis Buffer to a final cell density of 1×10^7 cells/mL. Incubate for 1–2 h at room temperature or >2 h at 4 °C to ensure homogeneous cell lysis (*see* **Note 11**).
2. Dispense 3 μL/well of cell lysate into an AlphaScreen compatible white, shallow 384 well microplate, e.g., ProxiPlate-384 Plus.
3. Dilute 5 mL of 10× ImmunoAssay Buffer in 45 mL of dH_2O to give 50 mL of 1× ImmunoAssay Buffer. Prepare microcentrifuge tubes with 500 μL of 1× ImmunoAssay Buffer. To each tube, add 0.5 μL (1:1000) of one mouse-derived anti-target antibody and 0.5 μL (1:1000) of one rabbit-derived anti-target antibody, prepared in a matrix such that each mouse-derived antibody is tested in combination with each rabbit-derived antibody. Include a buffer only control.
4. Dispense 3 μL/well of each antibody combination into separate wells of cell lysate in the ProxiPlate-384 plate, ensure mixing by brief centrifugation at 300 × *g* for 10 s, and then seal the plate and incubate for 1 h at room temperature.
5. Dilute 0.3 mL of 10× ImmunoAssay Buffer in 2.7 mL of dH_2O to give 3 mL of 1× ImmunoAssay Buffer. Under subdued light, add 72 μL of anti-Mouse IgG Alpha Donor beads to give a concentration of 120 μg/mL and 18 μL of anti-rabbit IgG (Fc specific) AlphaLISA Acceptor beads to give a concentration of 30 μg/mL. Dispense 3 μL/well of AlphaScreen reagents onto the cell lysates and antibody mix in the ProxiPlate-384, ensure mixing by brief centrifugation at 300 × *g* for 10 s, and then seal the plate and incubate for 4–16 h at room temperature.
6. Analyze AlphaScreen signal using a suitable plate reader, e.g., EnVision. Excite at 680 nm for 180 ms and measure emission at 570 nm for 550 ms.
7. For the antibody pair(s) that give AlphaScreen signal above the buffer control in **step 6**, repeat **steps 2–6** with variations in the dilution of the mouse-derived and rabbit-derived antibody and the cell density of the lysate to achieve a maximum AlphaScreen signal, which falls within a linear range with respect to cell number (*see* **Note 12**). Dilutions of the mouse antibody at 1:200, 1:500, 1:1000, 1:2000, 1:5000, and 1:10,000 should be titrated against the same dilutions of the rabbit antibody. Once optimal antibody dilutions have been established, cell density should be titrated from 2×10^7 cells/mL to 2×10^6 cells/mL (*see* **Note 13**).

3.4 Thermal Melt Analysis in Live Cells by CETSA

Volumes are provided for testing of one compound as a thermal melt curve but can be scaled as appropriate.

1. Harvest cells to 4 mL at the optimal cell density as determined in **step 7** in Subheading 3.3. Split the cells into two separate pools of 2 mL. To one pool of cells, add tool compound (*see* **Note 4**) to a final concentration of 10 μM. To the other pool, add an equivalent volume of DMSO.
2. Aliquot compound-treated cells into four tubes of each eight-tube PCR strip, 20 μL/tube, across 12 PCR strips. Aliquot DMSO-treated cells into the other four tubes of each eight-tube PCR strip, 20 μL/tube, across 12 PCR strips. Apply lids to the PCR tubes. Each eight-tube PCR strip should contain four tubes of compound-treated cells and four tubes of DMSO-treated cells, and there should be 12 full eight-tube PCR strips (Fig. 3).
3. Briefly clarify the PCR strips using a PCR strip microcentrifuge, place in a PCR rack, and incubate under tissue culture conditions (37 °C, 5% CO_2) to allow time for compound target engagement. Incubations are typically performed for 1–2 h.
4. Using a multi-block thermal cycler, e.g., SimpliAmp PCR machine, heat shock each PCR strip for 3 min at a certain temperature followed by 1 min at 20 °C (*see* **Note 14**). Load the first three PCR strips into the three separate blocks of the PCR machine and perform the first three heat shocks at 37 °C, 40 °C, and 42 °C. Remove and store at 4 °C, then load the next three PCR strips, and perform the next three heat shocks at 44 °C, 46 °C, and 48 °C. Repeat to perform heat shocks at 50 °C, 52 °C, and 54 °C and 56 °C, 58 °C, and 60 °C (Fig. 3).
5. Following completion of all heat shocks, ensure that all samples are incubated at 4 °C for a minimum of 5 min.
6. Open the PCR tubes and add 20 μL/tube of 2× SureFire Lysis Buffer. Briefly clarify the tubes using a PCR strip microcentrifuge and incubate for 10 min at room temperature.
7. Using a 96-well to 384-well convertible multichannel pipette, e.g., E1 ClipTip, mix each sample by aspirating 10 μL and dispensing back into the PCR tube, before aspirating 10 μL and dispensing 3 μL into an AlphaScreen compatible white, shallow 384-well microplate, e.g., ProxiPlate-384 Plus.

3.5 Quantification of Thermostable Target Protein by AlphaScreen

1. Dilute 0.5 mL of 10× ImmunoAssay Buffer in 4.5 mL of dH_2O to give 5 mL of 1× ImmunoAssay Buffer. Add primary antibodies (mouse and rabbit) to the dilution optimized in **step 7** in Subheading 3.3.
2. Under subdued light, add to the primary antibody mix 120 μL of anti-Mouse IgG AlphaDonor beads to give a concentration

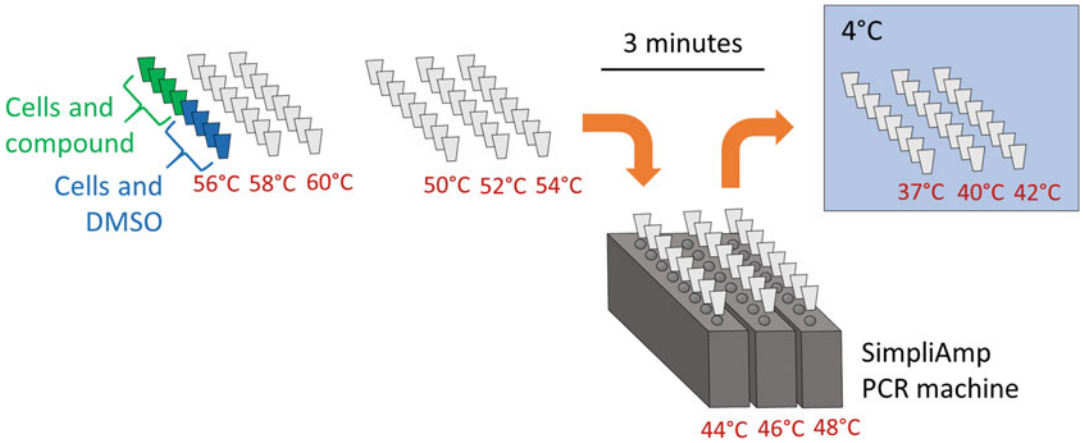


Fig. 3 Experimental setup to perform a CETSA thermal melt analysis. Each PCR strip contains both compound-treated cells and control cells. Following incubation under tissue culture conditions, each PCR strip is heat shocked at a single temperature for 3 min. Using the SimpliAmp PCR machine, three PCR strips can be heat shocked at a time at three different temperatures. Samples can be stored at 4 °C until all heat shocks have been performed before analyzing AlphaScreen signal. In this way, samples can be heat shocked across a range of temperatures to evaluate the thermal melting behavior of a protein in the presence or absence of compound (Fig. 1b)

of 120 μg/mL and 30 μL anti-rabbit IgG (Fc specific) AlphaLISA Acceptor beads to give a concentration of 30 μg/mL.

3. Under subdued light, add 6 μL/well of the combined antibody solution to the cell lysate in the AlphaScreen compatible 384-well plate, e.g., ProxiPlate-384.
4. Apply a plate seal.
5. Ensure mixing of antibodies with cell lysate by performing a brief centrifugation at 300 × *g* for 10 s.
6. Under subdued light, incubate plates for 4–16 h at room temperature.
7. Analyze AlphaScreen signal using a compatible plate reader, e.g., EnVision. Excite at 680 nm for 180 ms and measure emission at 570 nm for 550 ms.

3.6 Screening of Compounds at a Single Optimized Temperature by CETSA HT

Volumes are provided for execution of one 384-well plate for compound screening but can be scaled as appropriate. The optimal target temperature for the heat shock will be determined from the thermal melt analysis performed in Subheading 3.4. The optimal temperature is the lowest available temperature, which gives a suitable assay window between low AlphaScreen signal in cells treated with DMSO control and high AlphaScreen signal in cells treated with the tool compound. This is typically where 80–90% of AlphaScreen signal has been lost in the DMSO control melt curve (Fig. 1b).

1. Dispense compounds into a 384-well PCR plate using a liquid handling device, e.g., ECHO 555 acoustic dispenser (*see* **Note 15**) or a D300 dispenser (Tecan). Dispense compounds at 66.6× desired final concentration; for example, dispense 150 nL of a 2 mM DMSO stock of compound to achieve a final screening concentration of 30 μM.
2. Harvest cells to the optimal density (determined in **step 7**, Subheading 3.3), ensuring single-cell suspension (*see* **Note 16**), and dispense onto compounds in the 384-well PCR plate using a liquid dispenser, e.g., Multidrop Combi set to dispense 10 μL/well (*see* **Note 15**).
3. Apply a lid to the PCR plate.
4. Ensure mixing of the cells with the compounds by performing a brief centrifugation at 300 × *g* for 10 s (*see* **Note 17**).
5. Incubate the plate under tissue culture conditions (37 °C, 5% CO_2) to allow time for compound target engagement. Incubations are typically performed for 1–2 h.
6. Remove the plate from the incubator and replace the lid with Thermosealing tape (*see* **Note 14**).
7. Load the plate into the 384-well PCR instrument, e.g., LightCycler 480 II to perform the heat shock. Set a single cycle to hold at the target temperature as determined by analysis of data from Subheading 3.4 for 3 min with a ramp rate of 4.8 °C/s, followed by a hold at 20 °C for 1 min with a ramp rate of 2.5 °C/s.
8. Remove the plate from the 384-well PCR instrument and remove the Thermosealing tape.
9. Dilute 2 mL of 5× SureFire Lysis Buffer in 8 mL of dH_2O to give 10 mL of 2× SureFire Lysis Buffer. Lyse the cells by addition of 2× Lysis Buffer using a liquid dispenser, e.g., Multidrop Combi to dispense 10 μL/well (*see* **Note 15**). Ensure mixing of cells with Lysis Buffer by performing a brief centrifugation at 300 × *g* for 10 s. Wells will now contain 20 μL of cell lysates in 1× Lysis Buffer.
10. Progress to **step 11** within 1 h or store at 4 °C.
11. Thoroughly mix the cell lysate solution using a liquid handling platform, e.g., Bravo liquid handler to perform 10 repetitive cycles of aspirating and dispensing 7 μL. Transfer 3 μL of the 20 μL lysate to an AlphaScreen compatible 384-well plate, e.g., ProxiPlate-384, using the liquid handler (*see* **Note 18**). Aspirating and dispensing from the PCR plate is performed at 1.5 mm from the bottom of the well. Dispensing into the AlphaScreen 384-well ProxiPlate-384 plate is performed at 2 mm from the bottom of the well.
12. Quantify thermostable target protein following the steps in Subheading 3.5.

4 Notes

1. The components of the DSF buffer will depend on the target of interest and its requirements. A typical DSF buffer might consist of 50 mM HEPES-Na (pH 7.5) and 100 mM NaCl. It is advisable to avoid buffers whose pK_a changes significantly with temperature (e.g., Tris) to minimize the change in pH during the temperature ramp. When using SYPRO® Orange (or similar dyes) for DSF, it is imperative that all buffers are free of detergents as they will significantly increase the background fluorescence.
2. The fluorescence intensity of SYPRO® Orange increases with decreasing dielectric constant, and such is the case when it binds to the exposed hydrophobic cores of proteins upon thermal unfolding. Unfortunately, as a consequence of this, SYPRO® Orange is not suitable for protein formulations containing detergent (e.g., membrane proteins). In these situations, potential alternative dyes are CPM (*N*-[4-(7-diethylamino-4-methyl-3-coumarinyl)phenyl]maleimide [10]) or DCVJ (9-(2,2-dicyanovinyl)julolidine [11]), which work by different mechanisms. CPM reacts with cysteine residues within the protein of interest as it unfolds, yielding an increase in fluorescence. DCVJ is a fluorescent molecular rotor whose fluorescence is sensitive to viscosity and consequently aggregation as the protein unfolds. If these dyes are investigated, the same optimization protocol described herein should be employed.
3. An AlphaScreen antibody pair can be constructed using antibodies from any two different species, provided suitable anti-species conjugated AlphaScreen donor and acceptor beads are available (PerkinElmer). In practice, the majority of antibodies directed to most target proteins are raised in mouse or rabbit, and the most efficient route to identify suitable antibody pairs is to screen mouse-derived and rabbit-derived antibodies. The protocol is described in terms of screening mouse- and rabbit-derived antibodies but can be adjusted for other species if required. The number of antibodies that are screened will depend on the number available and any degree of prior knowledge of the target protein that may allow antibodies to be rationally selected.
4. To validate the thermal shift of a cellular protein upon target engagement, a well-validated tool compound is required. This should ideally be a potent binder of the target protein, with suitable physicochemical properties for cellular permeability and strong evidence to support binding to the target in cells. The compound should be tested at a concentration that

achieves maximal inhibition but avoids off-target cellular toxicity following 1–2 h incubation with cells, e.g., 10 μM. If such a tool compound is not available for a target, compounds can still be profiled by CETSA to look for novel binders, but the results should be treated with caution in the absence of a control that demonstrates exactly what change in thermal stability is induced upon intracellular binding.

5. To expedite the addition of protein to the PCR plate, prepare protein stock solutions in a deep 384-well polypropylene microplate. This allows the use of a multichannel pipette to add 10 μL of the protein solution to the PCR plate.
6. Between runs, the temperature of the block in the LightCycler will be higher than 20 °C, so precooling the instrument can increase reproducibility for proteins that melt at temperatures <35 °C.
7. Choose the pair of emission/excitation filters that matches the SYPRO® Orange maxima best. For the LightCycler 480 II, this would be 465/480 nm. It may be necessary to optimize the integration time and/or acquisitions per °C to achieve the desired ramp rate. The ramp rate need not be 1 °C/min (we have successfully used up to 2 °C/min to reduce acquisition time), but it is essential that all experiments being compared use identical parameters.
8. The optimal protein/dye concentration pair is target dependent. If the protein has a low melting temperature (<35 °C), it may be helpful to keep the plate cool as much as possible during its preparation. If the melt curve is poor under all tested concentrations of protein and dye, it may be necessary to consider different buffers for the DSF experiment or investigate alternative protein constructs/sources.
9. The melting temperature is defined as the midpoint of the protein melt curve. The simplest way to determine this is using the T_m Calling function in the LightCycler software, which will calculate and plot the negative of the first derivative of the melt curve. The minimum of this curve is the melting temperature. For large numbers of compounds, we use the TSA package in the Genedata® Screener software, but there are alternative solutions that are freely available [3, 12].
10. If required, it is possible to screen several cell lines at this stage, especially where expression levels of the target protein are unknown. It is also possible to perform this step using purified recombinant protein at 2 nM–5 nM, though any antibody pairs identified will later need to be confirmed against cell lysates.
11. In most cases, we do not find it necessary to add protease inhibitors to the SureFire Lysis Buffer, even for extended incubations of cell lysates up to 48 h.

12. A robust CETSA HT assay requires optimization of the antibody dilution, AlphaScreen reagent dilution, and cell density to achieve a minimum signal of around 15,000 AlphaScreen counts.
13. Once an antibody pair has been identified and optimized, the signal should be tested in the absence and presence of the tool compound (10 μM) to ensure that compound binding does not disrupt antibody recognition.
14. When handling cells in PCR tubes or a PCR plate, avoid touching the bottom of the tubes or wells to maintain consistent temperatures of the cell suspension prior to and after the heat shock.
15. To ensure compatibility with the Roche LightCycler 480 II, 384-well PCR plates have a particular footprint, which differs from most standard microtiter plates. The design of these plates makes them incompatible for certain dispensers. For use with the Echo 555 acoustic dispenser, an adapter is required (LabCyte). For use with the Multidrop Combi dispenser, PCR plates must be loaded backward (A1 in the bottom right corner) to ensure correct dispensing into each well.
16. CETSA HT assays typically require cells to be prepared to high cell densities in the region of 2×10^6 cells/mL to 2×10^7 cells/mL. To avoid cell damage, minimize the time cells are left at such high densities prior to seeding and ensure regular mixing to prevent cells settling. High cell densities can also cause cells to clump together, which can adversely affect results. To ensure a single-cell suspension, gently draw the solution of cells up into a syringe and filter through a 40-μm filter prior to dispensing cells with a liquid dispenser, e.g., Multidrop Combi.
17. Following brief centrifugation of PCR plates containing cells and compound, it is likely that cells will form a pellet at the bottom of the well. Across several different targets, we have not observed this to influence the compound responses in CETSA HT and find that cells do not require resuspension or mixing prior to, or throughout, the compound incubation.
18. We have found thorough mixing of the cell pellet with the Lysis Buffer to be essential to observe consistent AlphaScreen® signal between wells. Consistent signal is also dependent upon accurate volumes of liquid transfer from the PCR plate to the ProxiPlate-384 Plus.

References

1. Lo MC, Aulabaugh A, Jin G et al (2004) Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. Anal Biochem 332(1):153–159. https://doi.org/10.1016/j.ab.2004.04.031
2. Mashalidis EH, Sledz P, Lang S, Abell C (2013) A three-stage biophysical screening cascade for fragment-based drug discovery. Nat Protoc 8(11):2309–2324. https://doi.org/10.1038/nprot.2013.130
3. Niesen FH, Berglund H, Vedadi M (2007) The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. Nat Protoc 2(9):2212–2221. https://doi.org/10.1038/nprot.2007.321
4. Martinez Molina D, Jafari R, Ignatushchenko M et al (2013) Monitoring drug target engagement in cells and tissues using the cellular thermal shift assay. Science 341(6141):84–87. https://doi.org/10.1126/science.1233606
5. Jafari R, Almqvist H, Axelsson H et al (2014) The cellular thermal shift assay for evaluating drug target interactions in cells. Nat Protoc 9(9):2100–2122. https://doi.org/10.1038/nprot.2014.138
6. Eglen RM, Reisine T, Roby P et al (2008) The use of AlphaScreen technology in HTS: current status. Curr Chem Genomics 1:2–10. https://doi.org/10.2174/1875397300801010002
7. Almqvist H, Axelsson H, Jafari R et al (2016) CETSA screening identifies known and novel thymidylate synthase inhibitors and slow intracellular activation of 5-fluorouracil. Nat Commun 7:11040. https://doi.org/10.1038/ncomms11040
8. Shaw J, Leveridge M, Norling C et al (2018) Determining direct binders of the androgen receptor using a high-throughput cellular thermal shift assay. Sci Rep 8(1):163. https://doi.org/10.1038/s41598-017-18650-x
9. Shaw J, Dale I, Hemsley P et al (2019) Positioning high-throughput CETSA in early drug discovery through screening against B-Raf and PARP1. SLAS Discov 24(2):121–132. https://doi.org/10.1177/2472555218813332
10. Alexandrov AI, Mileni M, Chien EY et al (2008) Microscale fluorescent thermal stability assay for membrane proteins. Structure 16(3):351–359. https://doi.org/10.1016/j.str.2008.02.004
11. Menzen T, Friess W (2013) High-throughput melting-temperature analysis of a monoclonal antibody by differential scanning fluorimetry in the presence of surfactants. J Pharm Sci 102(2):415–428. https://doi.org/10.1002/jps.23405
12. Schulz MN, Landstrom J, Hubbard RE (2013) MTSA--a Matlab program to fit thermal shift data. Anal Biochem 433(1):43–47. https://doi.org/10.1016/j.ab.2012.10.020

Chapter 9

The Use of Acoustic Mist Ionization Mass Spectrometry for High-Throughput Screening

Helen Plant, David Murray, Hannah Semple, Gareth Davies, Ian Sinclair, and Geoffrey A. Holdgate

Abstract

It is clear from the analysis of the distribution of approved drug targets that enzymes continue to be a major target class for the pharmaceutical industry. The application of high-throughput screens designed to monitor the activity of these enzyme targets, and the ability of test compounds to modulate this activity, is still the predominant hit finding approach in the industry. The widespread use of enzyme activity-based screens has led to the development of several useful guidelines for the development and validation of robust and reliable assays. Key learnings for the development, validation, and implementation of acoustic mist ionization mass spectrometry for high-throughput enzyme assays are described.

Key words Mass spectrometry, Acoustic mist ionization, High-throughput screening, 384-well, Enzyme, Substrate, Label free, Plate reader

1 Introduction

The purpose of high-throughput enzyme activity screening is to allow monitoring of any particular transformation of substrate to product and allow detection of molecules that modulate the target enzyme activity. In general, high-throughput activity assays for enzymes have been developed using two main approaches. These are the detection of substrate depletion and the detection of product formation. The development of sensitive enzyme assays suitable for high-throughput screening (HTS) requires consideration of several factors: the biochemical reagents required, choice of technology, and the desired mode(s) of action to be identified. Reagent consideration encompasses the identification of relevant enzyme and substrate forms, cofactors, buffers, and additives essential to build a robust biological system. The choice of detection technology will involve aligning the biological system with the appropriate assay methodology to provide acceptable sensitivity, throughput, and cost. The desired modes of action that the assay should identify

Tina Daviter et al. (eds.), *Protein-Ligand Interactions: Methods and Applications*, Methods in Molecular Biology, vol. 2263, https://doi.org/10.1007/978-1-0716-1197-5_9,

also must be considered during the assay development phase. Typically, this requires designing biologically relevant assays, where possible utilizing full-length enzyme and physiological substrates, in which the concentration of the different enzyme forms present in the assay is approximately balanced [1, 2].

Historically, many high-throughput enzyme assays have been based on the use of spectrophotometric measurements, including absorbance [3], fluorescence [4, 5], or luminescence technologies [6]. Recently mass spectrometry (MS) [7, 8] has emerged as an increasingly applied technology for monitoring enzyme reaction progression. Direct measurement of substrate depletion and product formation using label-free detection, combined with a low artifact hit rate compared to traditional assays, makes this an attractive screening technology. Simple biochemical assay builds of enzyme plus substrate also makes MS amenable to automated assay plate generation. Recent breakthroughs in technology development [9] allowing lower sample volumes combined with higher processing speeds now make acoustic mist ionization mass spectrometry (AMI-MS) suitable for high-throughput screening of hundreds of thousands of compounds [10]. This chapter describes key information and considerations for developing and screening robust AMI-MS assays at scale.

2 Materials

Acoustic mist ionization mass spectrometry (AMI-MS) is a novel technology for directly introducing the contents of an assay well into a mass spectrometer at high-throughput. AMI-MS is a label-free method preferred for directly measuring the substrate to product conversion achieved by enzyme assays, due to the advantages afforded in the fidelity of the readout and the greatly reduced susceptibility to assay technology artifacts. The AMI-MS platform currently exists only as a prototype instrument, which we have been developing and using within Hit Discovery department at AstraZeneca and is now used in business-as-usual screening activities.

2.1 Instrumentation

AMI-MS requires that samples are introduced into the mass spectrometer via acoustic dispensing. The current prototype combines a Waters Xevo G2-XS quadrupole time-of-flight (qTOF) mass spectrometer fitted with a universal ion source coupled to a transducer from Echo® 555 integrated with a motorized XY stage (Fig. 1).

2.2 Plates

It is essential to use 384-well polypropylene source microplates that are Echo®-qualified, such as 384 PP plates catalog number P-05525 from Labcyte (*see* **Note 1**).

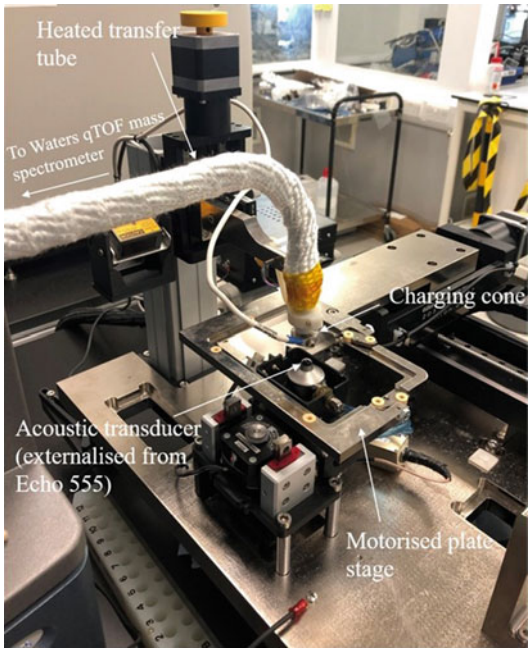


Fig. 1 Setup of the prototype instrument showing the acoustic transducer sitting below the plate stage, which sprays a droplet mist of the contents of the assay well into a transfer tube, which leads directly to the TOF MS

2.3 Compound Dispensers for Assay-Ready Plate Production

In HTS at AstraZeneca, we use Echo® 555 acoustic dispensers (Labcyte, CA, USA) to dispense nanoliter volumes of compounds into plates. Alternatives to this include ATS (EDC Biosystems, CA, USA), which also uses acoustic-based dispensing, or the Mosquito (TTP, UK), which uses tip-based dispensing.

2.4 High-Throughput Platform

2.4.1 Assay Plate Production

Multidrop™ Combi reagent dispensers (Fisher Thermo Scientific UK Ltd., Loughborough, UK) have been used for these assays on Agilent Biocel automation platforms (Santa Clara, California, USA). Certus Flex dispensers (Fritz Gyger, Gwatt, Switzerland) have also been tested to allow flexibility for transfer between the available automation (*see* **Note 2**).

2.4.2 Assay Plate Read

To allow integration of the plate production and read process, the commercial instrument will be compatible with the Access™ Laboratory Workstation and the Access™ Dual Robot System (Labcyte). This will allow automation of the movement of plates to the read position.

2.5 Buffers/ Reagents

1. Assay buffer: A simple biochemical assay buffer that is suitable for both maintaining the activity of the target enzyme and supporting the electrospray process should be selected (*see* **Note 3**). Buffers that have been shown to be effective for

these screens include Tris–HCl, tricine, and ammonium phosphate, at a maximum concentration of 20 mM.

2. Surfactants: These can both positively and negatively influence detection of the substrate and product in the assay (*see* **Note 3**). Detergents such as triton, DDM, CHAPS, and BSA have been successful in AMI-MS screens.
3. Cleaning solution: 100% (v/v) ethanol.

2.6 Software

1. Microsoft® Excel, or other spreadsheet-based analysis packages.
2. Genedata Screener® (Genedata), or similar high-throughput data analysis software.

3 Method

3.1 Instrument Setup

The schematic diagram (Fig. 2a) shows the principal components involved in the mist ejection and the charging and ionization processes.

3.1.1 Spray

Using an acoustic transducer, 20–80 nL aliquots of reaction mix are fired, creating a chaotic mist droplet distribution that is sprayed directly into the mass spectrometer (*see* **Note 4**). A custom csv file is used to control the transducer.

3.1.2 Charging

The charging voltage will typically be +2.5 to 3 kV. This is optimum for its two functions: to charge the liquid surface and to entrain and focus the charge droplets leaving the surface of the liquid.

3.1.3 Polarity Switching

When charging the well to fire positively charged droplets, the well becomes increasingly negatively charged. This prevents charge separation in the well and so prevents more positive droplets from being fired. Polarity switching is therefore needed to return the liquid in the well back to a neutral state to enable firing of further positively charged droplets.

3.1.4 Target Enhancement

This needs to be optimized to each particular screen around a set mass that is appropriate for both product and substrate in the reaction (Fig. 2b) (*see* **Note 5**). This generally increases sensitivity around four- to fivefold versus a non-enhanced acquisition. The user needs to be aware that masses outside or at the edge of the enhancement window will be understated or undetectable.

3.1.5 Buffer Considerations for Acoustic Settings

For simple aqueous buffers, default screening run protocols from the acoustic mist system will be applicable. The default settings will have typical parameters of a desolvation temperature (transfer tube) of 300–330 °C and a cone gas flow that is between 30 and 80 L/h

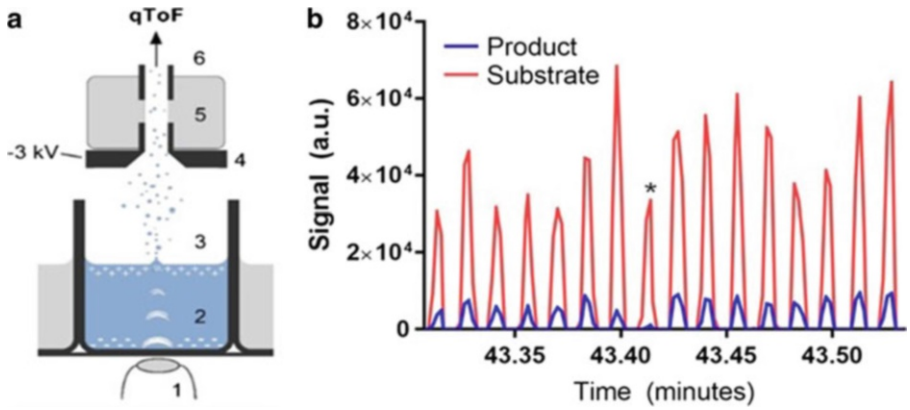


Fig. 2 (**a**) Diagram showing the principal components involved in the mist ejection and charging and ionization processes. The instrument is composed of an acoustic transducer (1) that emits sound waves into a sample in a microtiter plate (2), which is located on a moving XY-stage. High voltage is applied to a charging cone (4) directly above the transducer, and this induces a charge separation in the sample. A mound (3) is formed on the meniscus, and micrometer-sized charged droplets are sprayed off directly through an insulating piece (5) into a heated transfer tube (6), leading to the source of a mass spectrometer. (**b**) Raw AMI-MS data for the extracted ions of substrate and product for an enzyme assay plate from individual samples at a rate of 0.5 s per sample. An identified inhibitor is highlighted with an asterisk

(tuned for optimum sensitivity on the local instrument). Spray is a function of viscosity and surface tension, so additives in the assay will change these properties. For example, the addition of surfactants or organic solvents will reduce the surface tension and will therefore require less acoustic power to generate a mist of the same droplet size. For assays with a significant content of these additives, it may be necessary to run protocols utilizing lower acoustic energy for a screen.

3.1.6 Detection

Ion detection and isotope ratio measurement accuracy is key to the system's ability to quantify the output of the enzyme reaction. The MS system has an analog to digital converter (ADC) that requires software settings to provide a baseline measurement for the detection of a "real" ion event. There is also a gain control in the voltage applied to the detector plates. As part of the routine qualification of instrument performance, a series of concentrations of a single sample should be acquired. The ratio of carbon-12 to carbon-13 should be measured and plotted against ion counts. Accuracy is lost at both high and low ion counts. This information informs the user of the useable range of operation for the assay.

3.2 Assay Development for AMI-MS

Detailed assay development for acoustic mist ionization mass spectrometry assays is important for producing a robust assay for use in HTS due to the sensitivity of these assays to minor alterations in conditions. There are many variables that can elicit a significant change in the sensitivity or robustness of an assay, which would alter

the output of a screen. In this way, the setup of the assay should be well considered and validated to ensure reliability throughout a screen and produce the desired output. This method will describe the considerations necessary for setting up an effective AMI-MS assay for use in HTS.

AMI-MS assays require use of a simple biochemical assay buffer such as Tris or tricine, as many reagents interfere with signal detection. For example, salts interfere with signal detection as they can cause ion suppression or reduce the sensitivity of the detection. Even the acid or base used to set the pH of such buffers can be important. Inorganic acids or bases such as sodium hydroxide can have detrimental effects on acoustic firing; hence, the use of ammonium hydroxide as a base and acetic or formic acid is recommended for setting buffer pH (*see* **Note 6**).

Assay conditions should be optimized to ensure that the signal measured is directly proportional to the rate of reaction. Ensure that enzyme progress curves and the initial rates generated from them are measured in the linear phase with respect to time and enzyme concentration, respectively, and that a proportional decrease in signal is reflective of inhibition and in no way compromises the sensitivity of the assay. Where possible, the assay should be run using a substrate concentration(s) around K_m, to allow a balanced probability of detecting all modes of inhibition [11]. Measuring assay parameters such as K_m can be technically difficult in AMI-MS if an internal standard is not available, due to the variability of the data produced (Subheading 3.6). This variability is due to differences in firing events from well to well; hence, in the absence of an internal standard, normalization cannot be performed. In this case, K_m should be measured under the same assay conditions but using an alternative detection system, e.g., LC-MS, and this value should be used to establish the concentration of the substrate to use in the assay.

Another variable that should be optimized prior to an AMI-MS HTS is the volume of ejection into the spectrometer. This can affect the quality of the output data, especially the variability (Table 1). A larger ejection volume increases the read time and so the final volume selected will be a compromise between data quality and throughput.

One key advantage of AMI-MS over other screening technologies is the ability to measure the conversion of multiple substrates to products in one assay. The advantage of measuring multiple endpoints is that a choice of substrate to product conversion may allow multiplexing of the measurement of initial rate for single-step multisubstrate reactions. It also provides a route for rate measurements for the two half-reactions carried out by enzymes catalyzing sequential reactions using independent active sites. This dual endpoint assay can be set up by running the assay and then reading the plates twice in AMI-MS, enhancing for the different masses for the

Table 1
Varying ejection volume for a deiminase target AMI-MS assay. Increasing ejection volume improves data quality and reduces data variability, as shown by the lower standard deviation of the vehicle controls and a lower % coefficient of variance (% CV). % conversion is (Product/(Product + Substrate)) × 100

Ejection volume (nL)	Vehicle controls mean (% conversion)	Vehicle controls standard deviation	% CV vehicle controls
3	10.4	1.8	17.4
5	10.9	1.6	14.6
10	11.1	0.8	7.0

two reads. In this way, information about the mode of inhibition of a compound can be obtained during the primary screen. It is necessary to ensure that both sides of the reaction are able to proceed even in the presence of an inhibitor of the other side; hence, the substrates for both sides must be included in the assay setup.

3.3 Screening

Biochemical assays for AMI-MS have a very simple setup, comprising addition of enzyme to an assay-ready compound plate, followed by substrate addition, incubation at room temperature, and then addition of a stop reagent, which is usually an acid such as formic or acetic acid.

There is a range of methods for preparing compound assay plates for screening, from using manual handheld pipettes through to fully automated systems, providing that the concentrations of compound and vehicle are within acceptable limits for assays. Within AstraZeneca, a system of assay-ready plates (ARP) is used for both high-throughput and concentration-response screening, utilizing automated compound storage combined with acoustic dispensing, to ensure the highest quality and reproducibility possible [12]. Test compounds are held in long-term automated storage within a climate-controlled environment. After plating into master plates, an Echo® 555 acoustic dispenser (Labcyte, CA, USA) is used to pre-dispense nanoliter volumes of compounds into assay plates prior to the addition of reagents. This removes the need for extra pre-dilution steps of compounds dissolved in 100% (v/v) DMSO (*see* **Note 7**). Unless otherwise stated, all test compounds are prepared in 100% (v/v) DMSO at 10 mM.

The volume of compound transferred to ARPs is assay dependent. For primary screening, each compound is tested once at a single concentration (10 μM–100 μM, depending on the compound type). For concentration-response (CR) screening, each compound is dispensed to create single 10-point curves. To ensure

a consistent final screening concentration of DMSO, wells are backfilled, where necessary, with the appropriate volume of DMSO.

All ARPs have specific wells assigned for assay controls, and the positioning of these control compounds is dependent on the assay format used (*see* **Note 8**). Control compounds allow analysis of both inter- and intra-plate variation, calculation of Z' values [13], and for setting the normalization window.

The assay protocol used is usually a 10 μL enzyme mixture addition followed by a 10 μL substrate addition and finally a 30 μL acid step to stop the assay. Of course, while these volumes in 384-well plates bring a disadvantage in that they result in large assay volumes (50 μL), which increases assay reagent costs, they are required to achieve efficient firing (*see* **Note 9**).

There are many variables that appear to influence firing from these plates and hence the quality of the data produced. The meniscus formed by various dispensers or whether plates are centrifuged can cause problems with firing. Use of a Multidrop™ Combi, or similar dispenser, on a slow or medium speed for the final addition step of stop reagent to the assay is recommended. This reduces bubble formation in the well and appears to produce more consistent liquid surfaces. Centrifugation after both the enzyme and substrate additions is possible if the final addition step is added with a Multidrop™ Combi dispenser. Assay plates should always be stored at room temperature as using cold plates reduces effective firing.

An advantage of AMI-MS technology is the ability to read an assay while it is still progressing. For example, measurement of real-time assay progression can be obtained by adding enzyme and substrate to a plate and then repeatedly reading on the AMI-MS to obtain a significantly larger volume of kinetic data.

3.4 Automated Screening

3.4.1 Automation of Assay Plate Production

The simple nature of these assays makes automation for production of assay plates straightforward. The combination of three dispensers and a robot arm to transfer plates between dispensers at the appropriate time has successfully been applied to automate this process for a range of assays. When running AMI-MS at high throughput, the assay should be stopped to ensure that incubation times are consistent, usually using an acid such as formic acid or acetic acid. This enables the read to take place any time after the assay is performed, and often the plates can be re-read over a week after plate production, providing there is no breakdown of products occurring (*see* **Note 10**).

3.4.2 Automation of Assay Plate Read on the AMI-MS

The Access™ Dual Robot System (Labcyte) will allow incorporation of both assay plate production and plate read into one automated system, as it allows the movement of plates from the plate stacker to the read position. This will provide flexibility of running

either or both production and read processes as automated procedures.

3.5 Instrument Cleaning and Maintenance

The main reason for cleaning is due to the constant flow of air through the system, entering via the charging cone. The charging cone therefore becomes contaminated, which can change the field, leading to variability in sensitivity and edge effects on the plate. Cleaning processes for the AMI-MS include removal of the charging cone and sonication in 100% (v/v) ethanol. This process is recommended to be carried out on a weekly basis. The primary maintenance required for the Echo® is around the acoustic coupling fluidics. As this involves recirculated water, evaporation occurs so that regular refilling is necessary. Also, as this is open to the atmosphere, microbial contamination can be an issue. At each water change (recommended weekly), the fluidics need to be backflushed and a fresh bottle of water fitted. Preventative maintenance on all instruments as per the manufacturer's schedule should also be performed.

3.6 Data Acquisition

Plate reading in AMI-MS has a low failure rate; however, around 0.1% of wells do fail to fire. This can give a false hit in the analysis, due to apparent inhibition, if the substrate appears to be low, or zero, when the ratio is calculated (*see* **Note 11**).

Well-to-well variability can be seen with AMI-MS assays due to the high sensitivity of the technology. In this way, a small change in firing volume can have large consequences on assay output. Firing of the droplets from the Echo® system is variable, which means that wells can appear to have different levels of substrate or product, due to the variation in the amount of liquid fired. For this reason, it is recommended where possible to use an internal standard in the assay to allow normalization across wells and plates (Fig. 3) (*see* **Note 12**).

The hit rate tends to be low for AMI-MS assays as there are few artifacts compared to traditional technologies such as fluorescence intensity where, for example, false positives may be identified in the form of fluorescent compounds. Since AMI-MS only detects the substrate and product of interest using mass detection, potential artifacts may be limited to compounds that cause mass interference. Mass interference is an issue caused by utilization of an internal standard in the assay, although this does bring advantages in the quality of the screening output (Fig. 3) (Table 2). Mass interference occurs if a compound, a fragment of a compound, or a contaminant of a compound being tested, is the same mass as the internal standard. This leads to such a compound appearing active when the ratio of product/internal standard is calculated, as the internal standard layer will appear large and therefore the ratio will be small. This manifests as a low product result (well appears inhibited) when in reality the compound is simply interfering with the ratio

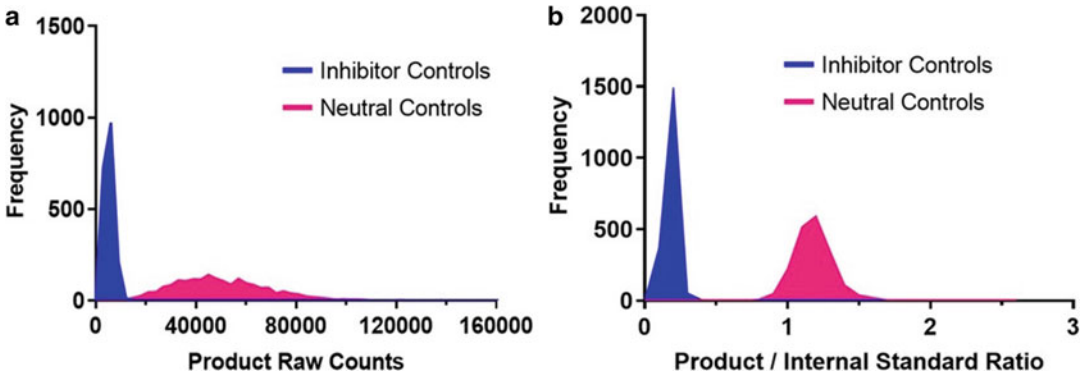


Fig. 3 Histogram showing data for inhibitor ($n = 1920$) and vehicle ($n = 1920$) controls. Raw data (**a**) are compared to data normalized to an internal standard (**b**), which is present in each well. Normalizing data to the internal standard improve standard deviation and % coefficient of variance (% CV) of the controls

Table 2
Statistics and parameters shown are for a methyl transferase target AMI-MS assay

Normalization	Neutral controls mean	Neutral controls standard deviation	% CV neutral controls	Inhibitor controls mean	Inhibitor controls standard deviation	% CV inhibitor controls
Raw product	49116.8	19460.6	39.6	3784.3	1791.6	47.3
Product/ internal standard	1.1	0.1	12	1.1	0.1	12

calculation. This issue is exacerbated when testing compounds in a concentration-response format as an increasing concentration of compound mimics the increasing concentration of internal standard and hence appears as a decreasing product when the ratio is calculated. A concentration-response curve is therefore seen, which is simply a false positive (Fig. 4).

3.7 Data Analysis

Currently, raw MS files are exported and converted into .csv files for import into separate data analysis software. The data structure is in a plate map format, with one .csv file per plate, and one plate map per target mass (*see* **Note 13**). For AMI-MS assays, when using an internal standard, a ratio of product divided by internal standard is calculated. If an internal standard is not available, analysis may be more difficult, as discussed in **Note 12**. In this case, a ratio of substrate to product is calculated. In both cases, the % effect is converted to a *Z*-score (*see* **Note 14**) that is used to identify active compounds. Ratios must be used for this analysis as product and substrate vary from well to well due to differential firing, as described above.

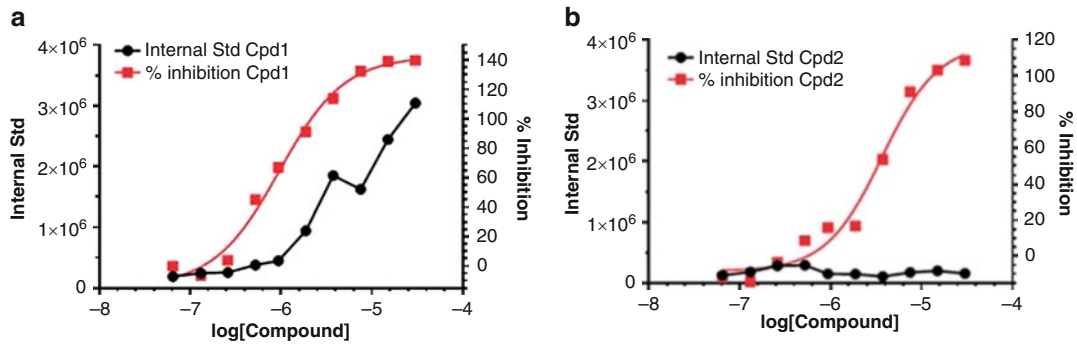


Fig. 4 Mass interference in an AMI-MS methyltransferase screen. (**a**) Test compound 1 interferes with the mass of the internal standard compound. Increasing concentrations of compound result in an increased measured level of internal standard due to the test compound (or contaminant in the test compound) being of the same mass as the internal standard. When the product/internal standard ratio is calculated, a false concentration-response curve results. (**b**) Test compound 2 does not interfere with the mass of the internal standard, resulting in a constant level of internal standard measurement independent of test compound concentration. A test compound inhibiting enzyme activity generates a concentration-response curve, as less product is measured with an increasing compound concentration: a decreasing level of product against a constant level of internal standard gives an increasing calculated % inhibition

4 Notes

1. It is essential to use 384-well polypropylene source microplates that are Echo®-qualified, such as 384 PP plates catalog number P-05525 from Labcyte. During manufacture, these plates are generated using a molding process that results in a higher consistency of plate flatness and base thickness, which supports consistent acoustic dispensing. Alternative plate types are not suitable for firing from an Echo®. The use of 1536-well plates is not currently possible with this technology. Since the rate of charging is inversely proportional to the surface area, a 1536-well plate charges four times more quickly than a 384-well plate and hence the speed of polarity switching approaches the scan speed of the instrument. Improvements in instrument sensitivity that would support the use of 1536-well plates are in progress, and this plate format may be supported in the future.
2. For 384-well plates, the use of 0.3/0.1 mm microvalves for the Certus Flex is recommended for the highest accuracy and precision combined with medium dispensing speed.
3. Several common buffers and detergents have been tested to understand their ionization potential and ion suppression effects in both positive and negative ion modes. Some buffers such as 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) caused severe signal suppression, whereas a combination of tris(hydroxymethyl)-aminomethane (Tris) and Triton X-100 as the buffer–detergent system produced clean mass

spectra of substrates/products at concentrations ranging from 1 to 10 μM. Use of a surfactant can influence the droplet formation in the acoustic event, which may promote the ionization process but can also cause ion suppression, hence reducing detection.

4. All data have been collected using a Waters Xevo G2-XS quadrupole time-of-flight (qTOF) mass spectrometer fitted with a universal ion source. One of the features of the Waters mass spectrometer employed is the ability to enhance sensitivity for a small mass range, nominally 200–300 Da.
5. For example, MS acquisition from 500 to 700 Da with a scan time of 100 ms reduces the amount of data collected and produces enough data points across a well ejection.
6. If salts are required for the enzyme reaction, the MS interference problem can be solved in two ways. The simplest is by adding a large volume of stop reagent at the end of the assay, as this dilutes the salts down prior to the AMI-MS read. Alternatively, the salts can be precipitated out of the solution in the stop step. For example, $MgCl_2$ can be precipitated by using the appropriate amount of ammonium phosphate in the assay buffer, which precipitates the salt on addition of ammonium hydroxide during addition of the stop reagent.
7. Users should not see any difference between using assay-ready plates and plates prepared using more conventional methods where a solution of compound is diluted stepwise prior to adding to the plate. It is important that the DMSO concentration is 1% (v/v) or lower as DMSO in higher concentrations is potentially inhibitory.
8. It is important to consider carefully where to place control wells on the plates. It is quite common to see control wells placed on the outermost columns, but this is not recommended as these wells will be affected by any edge effect, which in turn will lead to incorrect normalization of the data. Control wells are placed in the middle of the plate to minimize any edge effects. Ideally, controls could be dispersed across more of the plate, but this is difficult to achieve with more traditional liquid handling equipment and adds complexity, although these layouts are employed for some concentration-response screens.
9. A volume of 50 μL in standard 384-well plates is necessary as this improves firing in the AMI-MS. A 10 μL + 10 μL addition for the enzyme and substrate is generally necessary to ensure the two components mix effectively due to the large surface area of the well base. To reduce reagent costs, these ratios could be altered. It is possible to use lower volume additions, e.g., 2 μL of enzyme plus 2 μL of substrate, if a centrifuge step is added. This allows mixing of the components. Following this, a

stop mixture should be added to give a final volume of 50 μL for the read. Using a low-volume reaction mixture and having a large dilution to 50 μL final volume is advantageous when using salts, as discussed in **Note 6**. Low-volume 384-well polypropylene plates are a viable alternative, which would allow lower volumes of reagents to be used. Currently, these plates are expensive, so they may only be beneficial if the cost of reagents outweighs the extra costs of the plates.

10. When storing assay plates to be read or re-read after more than a day, they should be stored at ambient temperature rather than cold. Storing assay plates at 4 °C appears to reduce the ability to fire on the MS. If plates are stored cold, they must be allowed to fully equilibrate to ambient temperature prior to reading. Plates should be stored in a stack or with a lid on to reduce evaporation. Water may be added to the plates if the volume has been reduced by evaporation when in storage.
11. It was decided to mask any well where the measured substrate value is 0, but other low reading wells are taken through to confirmation or IC/EC_{50} screening to ensure that potential hits are not missed.
12. Use of an internal standard in any AMI-MS assay greatly improves ease of analysis as data can be referenced to and normalized against this standard. Anything that does not interfere with the assay, such as a modified substrate or product or any derivative of these, can be used as an internal standard. One example is the use of stable isotopically labeled standards, which are chemically identical to reagents used in the screen, such as deuterium-labeled products. If these can be obtained, they can have positive effects on the quality of a screen output and so should be investigated early during assay development. The use of an internal standard which can be added to the acid in the stop step can also be considered. As seen in Table 2, a product to standard ratio can be calculated, which mitigates for the intra-well variation of product formation. This results in lower standard deviations and % coefficient of variance of both the vehicle and inhibitor controls.
13. The communication between the MS system and the acoustic mist system is such that each scan performed on the MS is annotated with the acoustic mist well location. This embedding of location within the raw data of the MS allows a single, combined spectrum for each well on a plate to be produced in an automated postprocessing algorithm (prototype in house software) using the raw data. This embedding of well location in the raw data also enables use of commercial applications such as Genedata Expressionist and MassLynx for data parsing. A full spectrum is produced for each well and is queried to

produce the area under the spectral peak of the mass of interest. This number is then recorded in the .csv output file for the mass of interest to give a single plate map of data for each mass.

14. Robust *Z*-score is calculated from % effect and is used to identify active compounds. Robust *Z*-score is calculated using the equation below:

$$\text{Robust } Z\text{-score} = \frac{x - m}{\text{RSD}} \quad (1)$$

where x is the raw data value of the well to be standardized, m is the median of the chosen control well group, and RSD is the robust standard deviation for the chosen control well group.

References

1. Copeland R (2003) Mechanistic considerations in high-throughput screening. Anal Biochem 320:1–12
2. Williams K, Scott J (2009) Enzyme assay design for high-throughput screening. Methods Mol Biol 565:107–126
3. Lavery P, Brown MJ, Pope AJ (2001) Simple absorbance-based assays for ultra-high throughput screening. J Biomol Screen 6:3–9
4. Rogers M (1997) Light on high-throughput screening: fluorescence-based assay technologies. Drug Discov Today 2:156–160
5. Zaman GJ, Garritsen A, de Boer T, van Boeckel CA (2003) Fluorescence assays for high-throughput screening of protein kinases. Comb Chem High Throughput Screen 6:313–320
6. Gokhin S, Auld D (2017) Chapter 6 Understanding luminescence-based screens. In: Bittker JA, Ross NT (eds) High throughput screening methods: evolution and refinement. The Royal Society of Chemistry, London. ISBN 978-1-78261-471-4
7. Rohman M, Wingfield J (2016) High-throughput screening using mass spectrometry within drug discovery. Methods Mol Biol 1439:47–63
8. de Rond T, Danielewicz M, Northern T (2015) High throughput screening of enzyme activity with mass spectrometry imaging. Curr Opin Biotechnol 31:1–9
9. Sinclair I, Stearns R, Pringle S, Wingfield J, Datwani S, Hall E, Ghislain L, Majlof L, Bachman M (2016) Novel acoustic loading of a mass spectrophotometer: toward next-generation high-throughput ms screening. J Lab Autom 21(1):19–26
10. Sinclair I, Bachman M, Addison D, Rohman M, Murray DC, Davies G, Mouchet E, Tonge ME, Stearns RG, Ghislain L, Datwani SS, Majlof L, Hall E, Jones GR, Hoyes E, Olechno J, Ellson RN, Barran PE, Pringle SD, Morris MR, Wingfield J (2019) Acoustic mist ionization platform for direct and contactless ultrahigh-throughput mass spectrometry analysis of liquid samples. Anal Chem 91(6):3790–3794
11. Acker M, Auld D (2014) Considerations for the design and reporting of enzyme assays in high-throughput screening applications. Perspect Sci 1(1–6):56–73
12. Griffith D, Northwood R, Owen P et al (2012) Implementation and development of an automated, ultra-high-capacity, acoustic, flexible dispensing platform for assay-ready plate delivery. J Lab Autom 17(5):348–358
13. Zhang JH, Chung TD, Oldenburg KR (1999) A simple statistical parameter for use in evaluation and validation of high throughput screening assays. J Biomol Screen 4:67–73

Chapter 10

Ligand Discovery: High-Throughput Binding: Fluorescence Polarization (Anisotropy)

Geoffrey A. Holdgate and Paul E. Hemsley

Abstract

High-throughput assays based on fluorescence polarization (or fluorescence anisotropy) technology have often been employed for primary hit-finding in drug discovery. These binding assays provide a homogeneous format and consistent performance and offer advantages over some other optical methods. Developments in assay design and improvements in fluorescent probes have enabled the application of the technique to even complex biological systems. Here we describe the practical considerations for development of FP assays applied in high-throughput screening, including fluorophore selection, assay design, data analysis, and approaches for detecting compound interference.

Key words Fluorescence Polarization, Fluorescence Anisotropy, High-throughput Screening, Drug Discovery, Dye, Ratiometric

1 Introduction

Screening large compound libraries to identify small molecule modulators of biological targets is a well-established and commonly used procedure in drug discovery [1–3]. The choice and configuration of the primary screen technology is influenced by many factors, including the biological target, the range of desired modes of action of hit molecules, access to tools and reagents, the follow-up cascade of assays available, and any previous experience with screening technologies applied to the actual or similar targets. Typically, primary screening methods are then followed up with orthogonal approaches designed to remove false hits arising from primary assay interference or to demonstrate functional activity of compounds identified as binding to the target [4]. Thus, drug discovery often involves the use of both binding assays and activity assays, and these can often be used in combination to provide confidence in the primary screen output. One technology for identifying compounds that bind to the biological target is fluorescence polarization (or fluorescence anisotropy) [5, 6]. This method relies

Tina Daviter et al. (eds.), *Protein-Ligand Interactions: Methods and Applications*, Methods in Molecular Biology, vol. 2263, https://doi.org/10.1007/978-1-0716-1197-5_10,

on the inverse relationship between molecular rotation and the degree of polarization of a fluorophore excited by linear polarized light [7]. As the apparent molecular weight of the fluorescent probe molecule is increased by binding to the target protein, the polarization of the subsequent emitted light increases. This phenomenon has been exploited to configure assays for several different biological targets in a single molecule binding (protein–ligand interaction) assays through enzyme activity (substrate turnover to product) and even to live-cell formats [8, 9]. There are several advantages of FP assays that make them particularly applicable to high-throughput screening, including solution-based measurement, the ability to measure kinetics, the avoidance of radioisotopes, the simplicity of the assay format, the lack of filtration or separation steps, and the relatively inexpensive reagents required. Additionally, as FP is a ratiometric measurement (it is independent of the actual light intensity and hence the fluorophore concentration because it is defined as a ratio of two components, both of which are themselves proportional to the concentration), it is relatively insensitive to artifacts such as compound absorption or inner filter effects that may constrain other light-based technologies. However, the technique can suffer from some issues experienced by other methods such as light scattering and autofluorescence [10]. The technology is readily scalable to both 384- and 1536-well formats, and with improvements in the understanding of issues associated with fluorescent probes (including peptides, small molecule drugs, and cytokines) used in FP assays, such as the potential for depolarization due to flexibility in the attachment of the dye, sometimes referred to as the "propeller effect," suitable probes utilizing dyes such as fluorescein, rhodamine, BODIPY, Cy, and Alexa fluors, without long aliphatic linkers, can often be synthesized [11]. In this chapter, we will predominantly use the terminology for fluorescence polarization, using anisotropy in equations, where this is simpler.

2 Materials

Buffer systems (*see* **Note 1**) will vary depending upon the binding interaction studied. Ideally, buffer systems should be as simple as possible, containing the minimum number of additives required to support the functional biochemical interaction. Buffers should be chosen such that a suitable buffering capacity is available to minimize changes in pH during the reaction of interest (*see* **Note 2**). Often, buffers such as Tris or HEPES (*see* **Note 3**) and additives such as inorganic salts containing magnesium or sodium ions (*see* **Note 4**), as well as reducing agents including DTT and TCEP (*see* **Note 5**), are suitable for FP assays. Subcritical micellar concentrations (CMCs) of detergents are also frequently used (*see* **Note 6**).

All solutions should be prepared using ultrapure water (prepared by purifying deionized water to achieve a sensitivity of 18 MΩ cm at 25 °C) and analytical-grade reagents.

1. 1 M HEPES (stock): Weigh 238 g and dissolve in 900 mL of ultrapure water in a beaker. Add a stir bar and stir until fully dissolved. Monitor the pH and add a solid pellet of NaOH until dissolved. Continue to add and dissolve NaOH pellets one at a time until the pH approaches the desired quantity. Fine-tune the pH by addition of small aliquots of 100 mM solution of NaOH. Add ultrapure water to a final volume to 1 L. Sterile filter the solution (*see* **Note 7**) and store for up to 3 months in the dark at room temperature prior to use.
2. 4 M sodium chloride (stock): Weigh 234 g and dissolve in 1 L of ultrapure water. Filter the solution, sterilize (*see* **Note 7**), and store for up to 6 months in the dark at room temperature prior to use.
3. 1 M magnesium chloride: Manufacturer-supplied as a solution, and it has a long shelf life when stored at room temperature in the dark.
4. 1 M dithiothreitol (DTT stock): Weigh 1.54 g and dissolve in 10 mL of ultrapure water. Filter the solution, sterilize (*see* **Note 7**), and store in aliquots of 100 μL for up to 6 months in the dark at −20 °C prior to use.
5. Detergents: Manufacturer-supplied and stored according to their instructions.
6. High-molecular-weight binding partner: proteins (*see* **Note 8**), manufacturer-supplied or purified in-house, are typically stored in aliquots of 10–20 μL at high concentration, often around 10 μM, at −80 °C.
7. Fluorescently labeled low-molecular-weight binding partners: manufacturer-supplied or labeled in-house (*see* **Note 9**).
8. 1 mM fluorescein (stock): Dissolve 1 mg of fluorescein in 3 mL of ultrapure water.

3 Methods

3.1 Solution Handling

Care must be taken when daily removing aliquots of each buffer component to avoid contamination. Working solutions are made as follows: half the required volume of ultrapure water is added to a sterile plastic vessel followed by addition of the required volume of each of the assay components to give the correct final concentration. Subsequently, the required dilution is made by addition of a further quantity of ultrapure water. This protocol is followed to prevent buffer components from precipitating if they are added

together prior to addition of the ultrapure water. For fluorescence polarization assays, it is extremely important to avoid the light scattering effects of dust particles [12]. Such scattering is highly polarized and can have a dramatic effect on the quality of the analysis, and so all solutions are filtered before use (*see* **Note 10**).

3.2 Instrumentation and Calibration

For FP experiments, detection of emitted light must be measured in two planes—parallel and perpendicular to the plane of excitation polarization. Fluorescent plate readers with two detectors (Fig. 1), such as a BMG Pherastar™, may be used. In this case, the sample is excited by polarized light, which is selected by specific polarizing filters, and the instrument allows simultaneous dual emission, permitting the detection of different polarization vectors. One detector measures parallel polarized emitted light and one measures perpendicular polarized emitted light. Other instruments may use one detector, making two measurements from the sample, whereby the polarizing filter is turned by 90° between measurements. Experimentally, the degree of polarization is determined from these measurements of fluorescence intensities parallel and perpendicular with respect to the plane of linearly polarized excitation light and may be expressed either in terms of fluorescence polarization (P) or anisotropy (r).

These values are calculated as follows:

$$P = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + I_{\perp}} \tag{1}$$

$$r = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + 2I_{\perp}} \tag{2}$$

where $I_{\parallel}$ = intensity of light detected in the parallel detector and $I_{\perp}$ = intensity of light detected in the perpendicular (antiparallel) detector (*see* **Note 11**).

Two detector instruments should be calibrated to ensure that there is no difference in their photon counting efficiency. A suitable approach for doing this is to use a 1 nM aqueous solution of fluorescein to calibrate the instrument such that it records a value of 27 mP, which is the theoretical value for this molecule [13, 14]. Fluorescein stock solution is diluted to 1 μM by adding 10 μL into 10 mL of ultrapure water and subsequently further diluted to 1 nM by adding 10 μL into 10 mL of assay buffer. This solution is added to a well in a black microtiter plate such that the volume matches the volume used in the assay. The plate is placed into the instrument and calibrated using the software to the target mP value (*see* **Note 12**).

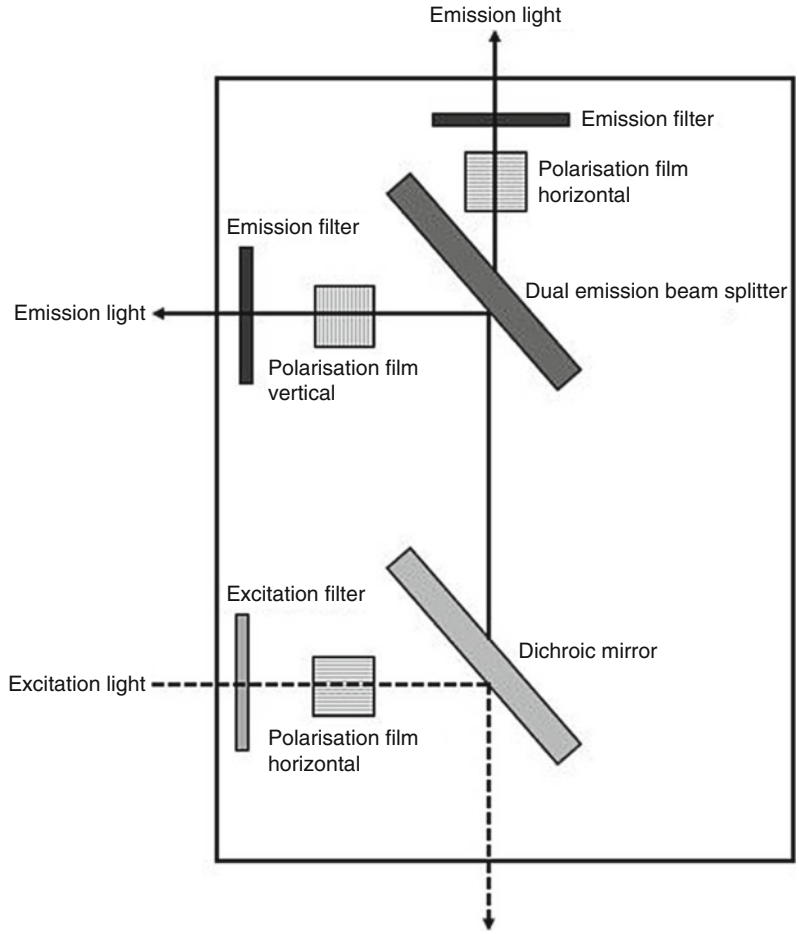


Fig. 1 Schematic representation of a dual detector instrument

3.3 Experimental Design and Assay Development

FP measurements report the status of a binding interaction and may be used in several different ways to provide information about the behavior of small molecule drugs. One of the primary uses is to identify compounds that can displace a fluorescent probe from a target protein, and it this type of approach that will be described below. However, many of the experimental approaches toward assay development in this format will be equally appropriate for the development of other formats.

3.3.1 Measurement of Background Fluorescence

Instruments report fluorescence intensity in both planes as well as the calculated ratiometric FP value (*see* **Note 13**). It is important to measure the fluorescence properties of the buffer that has been optimized because significant background fluorescence should be subtracted from each detector/plane prior to the calculation of the polarization value. Most instruments do not do this, and they simply return values of individual fluorescence values and the calculated polarization value. For this reason, it is often easier to ensure

that background fluorescence is not significant, rather than to correct for it. Two approaches to keep background fluorescence as low as possible are to use black microtiter plates (*see* **Note 14**) and to use red-shifted fluorophores (*see* **Note 15**).

Background fluorescence that is less than 5% of the ligand at its final assay concentration can be ignored. If background fluorescence cannot be removed due to the requirement for a critical assay component, then the best approach is to collect individual fluorescence values from each detector/plane, subtract the background fluorescence value, and use these corrected values in the polarization calculation (*see* Subheading 3.3.4).

Another reason for measuring background fluorescence in the buffer is to determine whether it is equal in each detector/plane. A significant difference in the two values is indicative of the presence of a polarizing component in the buffer. Often, this may be indicative of particulates that are known to produce highly polarized scattered light, the presence of which may be remedied by filtering the buffer.

3.3.2 Selection of a Binding Probe

Fluorescent probe selection is, arguably, the most difficult step in the development of a good FP assay. Protein targets that are well known in the literature will often have known small molecule binders, which can be chemically modified to produce a fluorescent binding probe. A good probe will have all of the properties described below:

1. A red-shifted absorbance wavelength. It is crucial that compounds do not interfere in the assay either by fluorescing or quenching at the wavelengths used. This is because FP is a technology that lends itself very well to measuring the effects of compounds in a homogeneous mixture without separation steps. For high-throughput screening, the Alexa 647® dye is often used as the wavelengths used are longer than most of the compounds in the screening collection. A way of measuring and dealing with fluorescent interference is described in Subheading 3.3.4.
2. High fluorescence yield. The probe should have a high fluorescence quantum yield so that it may be detectable in low concentrations.
3. Suitable binding affinity. A good probe should bind with an affinity that is neither too high nor too low. Low-affinity probes require a relatively high concentration of target protein to achieve the level of binding required to get a high polarization signal. This is problematic in terms of protein reagent consumption and results in sensitivity issues due to the lower limit on the measurable IC_{50} that is introduced by the target protein concentration (*see* **Note 16**). A high-affinity probe

introduces issues in terms of the limit of fluorescence detection since the fluorescence of a very high-affinity probe will not be detectable at concentrations around its K_d in many instruments. Often, a suitable probe has a K_d of between 1 and 10 nM. Such a probe allows reduction of the amount of target protein required and allows for the detection of relatively high-affinity displacing compounds.

3.3.3 Measurement of the Time to Equilibrium

Before a classic receptor/ligand equilibrium binding experiment can be completed, the incubation period required for equilibrium to be achieved must be empirically determined. The time taken to reach equilibrium at room temperature, which is used for high-throughput measurements, is dependent on the concentrations of the ligand and receptor. By using a constant label concentration and the lowest measurable receptor concentration, it is possible to determine the time the system takes to reach equilibrium as follows:

1. Ensure that the background from the buffer is <5% of that of the ligand.
2. Add the ligand, at a concentration close to its K_d, to the plate and take several readings to determine the polarization of the free ligand.
3. To the same well, add an amount of receptor approximately equal to 1/20 of the expected K_d. Since the rate of association is dependent on the receptor concentration, this combination of low ligand and receptor concentrations represents the longest time necessary to reach equilibrium.
4. Follow the increase in polarization over time and determine the time at which the polarization values reach a plateau, representing the approach to equilibrium. Depending upon the binding kinetics, the time to equilibrium may be many hours, but for practical purposes, a suitable ligand for high-throughput screening should reach equilibrium faster than about 5 h. This time should then be used as the incubation period for the binding experiments.

3.3.4 Effect of Binding on Total Fluorescence

Once a buffer system and probe have been identified, it is important to measure the effect of binding on the total fluorescence of the ligand, since binding may cause a reduction in the amount of fluorescence detected due to quenching effects:

1. Ensure that the background from the buffer is <5% of that of the ligand.
2. Add the ligand, at a concentration close to its K_d, to ten wells in a black microtiter plate and dilute to the required assay volume with assay buffer.

3. Read the plate to determine both the total fluorescence and the polarization of free ligand.
4. Add the ligand to another ten wells on a black microtiter plate and, to these wells, also add receptor to a final concentration approximately equal to tenfold above the expected K_d. Dilute to the required assay volume with assay buffer.
5. Incubate for sufficient time to allow the binding reaction to reach equilibrium (as determined in Subheading 3.3.3) and measure total fluorescence and polarization. If total fluorescence changes by more than approximately threefold, a correction may be required. The values obtained in the experiments above are used to correct the anisotropy (or polarization) value obtained, employing the equation below [15]:

$$r_c = \frac{\{[(r - r_f)/(r_b - r)](Q_f/Q_b)(r_b)\} + r_f}{1 + [(r - r_f)/(r_b - r)(Q_f/Q_b)]} \quad (4)$$

where r_c is the corrected anisotropy, r is the measured anisotropy, r_f is the anisotropy of the ligand alone, r_b is the anisotropy of the ligand bound to the receptor ([receptor] = 10 × K_d), Q_f is the total fluorescence of the ligand alone, and Q_b is the total fluorescence of the ligand bound to the receptor ([receptor] = 10 × K_d).

This correction will be required for all subsequent readings and is applicable for data that refer to anisotropy, but polarization values may be substituted using Eq. 3 (*see* **Note 11**).

This effect may be modeled, as shown in Fig. 2 (*see* **Note 17**), using values of 25 mA for r_f, 300 mA for r_b, 20,000 for Q_f, and varying levels of quench affecting Q_b.

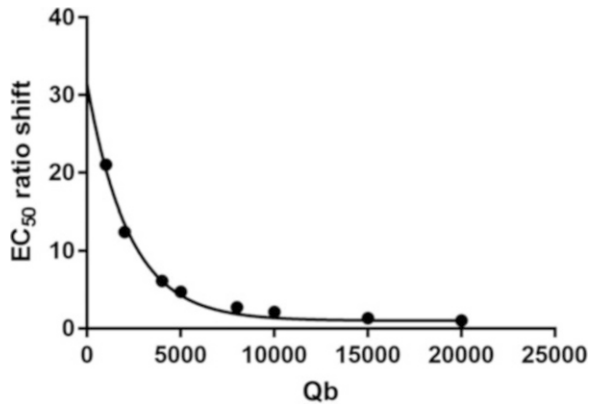


Fig. 2 Effect of fluorescence quenching on the measured IC_{50}. Values of 25 mA for r_f, 300 mA for r_b, 20,000 for Q_f, and varying levels of quench affecting Q_b have been used

It should be noted that this correction is only needed for severely quenched ligands, and, in these circumstances, the correction is very sensitive in that small changes in measured anisotropy can lead to large changes in corrected anisotropy and for this reason measured anisotropy values must be measured extremely carefully (*see* **Note 17**).

If possible, it is preferable to work with ligands that are quenched by less than 75% on binding. If the quenching is more severe, then a correction may be applied, but the variability of the correction at low anisotropy values means that the measurements must be made extremely carefully. Hence, it is advisable to employ a relatively unquenched ligand rather than to treat high levels of quench mathematically.

3.3.5 Measurement of Binding of the Probe to the Target

To design experiments that demonstrate whether test compounds can displace the probe, it is important to have a situation, in the absence of test compound, where the probe is almost fully bound to the target. This is so that displacement by test compound will result in a change of anisotropy that may be measured. To arrange for this situation and to allow subsequent calculation of test compound K_i values, the affinity of the probe must be measured (*see* **Note 18**).

To do this, a simple titration of target protein is made in the presence of several different concentrations of the probe (*see* **Note 19**).

The target protein is diluted to ~800 nM in assay buffer followed by a serial dilution in assay buffer to give a range of target protein concentrations. A twofold serial dilution is ideal in this respect so that there are a large number of points on the concentration-response curve (*see* **Note 20**). A control representing zero target protein should also be made using assay buffer alone.

The fluorescent probe is diluted depending upon the storage conditions of the probe. Probe concentrations of 60, 20, 6, 2, and 0.6 nM are good starting points. If the probe is stored in DMSO, then any intermediate dilution steps required should be carried out in DMSO before subsequent dilution to 2× the final assay concentration using assay buffer (*see* **Note 21**). Equal volumes of target protein and probe are mixed such that there is a concentration-response curve for target protein at each probe concentration. The observed K_d values will change with probe concentration where the probe concentration is higher than its K_d. As the probe is diluted to a value below the K_d concentration, the measured K_d will approach a constant value, representing the true K_d (Fig. 3).

This experimental format can also be used to measure the kinetics of probe binding. This involves reading the plate at regular intervals over time. This provides an understanding of how long it takes to reach equilibrium (*see* Subheading 3.3.3), which is

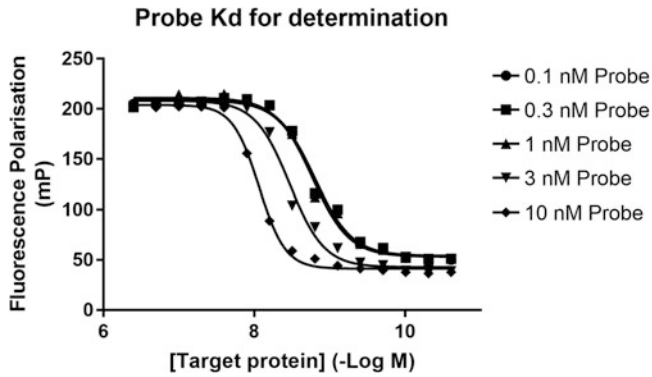


Fig. 3 Typical dose-response curve. In the example above, the lines overlie at probe concentrations of 0.1, 0.3, and 1 nM probe and give a K_d value of 1 nM. At higher probe concentrations, the curve is shifted to higher apparent K_d values because more protein is required to bind the higher levels of probe due to depletion

particularly important for high-affinity probes. Further readings taken following establishment of equilibrium result in an appreciation of the stability of the reagents during the assay.

3.3.6 Reagent Stability

For high-throughput applications, where reagents need to be stable for extended periods, it is useful to assess the stability of the probe and target protein. To do this, target protein is made and stored for various amounts of time prior to the addition of probe. The plate is incubated for a time necessary for the binding to reach equilibrium and the signal measured. Repeat experiments should be undertaken where the probe is stored for various amounts of time prior to addition of target protein. The result of these experiments indicates the time for which the assay components are stable and may be prepared in the plate before the start of the binding experiment (*see* **Note 22**).

3.4 Measuring the Effect of Test Compounds

Compounds that displace the labeled probe from the target protein cause a decrease in the observed polarization. The effect of compounds may be observed by the addition of a single concentration of test compound (in primary screening), or by creating a concentration response for test compound covering a wide range of concentrations and differing by half-log steps. The output of such a concentration-response experiment is often analyzed by fitting a four-parameter logistic equation:

$$\Upsilon = \text{Min} + \frac{(\text{Max} - \text{Min})}{\left\{1 + 10^{(\log \text{IC}_{50} - X).h}\right\}} \quad (5)$$

where Υ is the measured signal, Min is the anisotropy value found with fully displaced probe (*see* **Note 23**), Max is the anisotropy value found with fully bound probe, IC_{50} is the concentration of test compound leading to 50% probe displacement, and h is the Hill slope (*see* **Note 24**).

Compounds may then be rank-ordered by IC_{50} as a measure of the effectiveness of probe displacement, which reflects the affinity of the test compound, although to assess true binding affinity, the calculations below must be used.

3.4.1 Calculation of Affinity from IC_{50}

It is important to note that the method of Cheng and Prusoff [16] to calculate a compound K_i is not appropriate for fluorescence polarization assays because these assays are established in such a way that the assumption $[L]_t = [L]_f$, i.e., the concentration of the total probe added is equivalent to the free concentration, is not valid due to significant depletion of the free probe by binding to protein in order to generate the anisotropic signal.

Consequently, to account for the effect of probe depletion, the measured IC_{50} value should be converted to K_i values using alternative approaches, such as the Munson–Rodbard equation [17] (*see* Eq. 6 below) or the method of Wang [18].

$$K_i = \frac{IC_{50}}{1 + \frac{L_t(y_0+2)}{[2K_d(y_0+1)]} + y_0} - K_d\left[\frac{y_0}{y_0+2}\right] \tag{6}$$

where y_0 is the initial bound/free ratio for the labeled probe, L_t is the total concentration of the labeled probe, and K_d is the dissociation constant for the labeled probe. When y_0 is small, the equation reduces to the familiar Cheng–Prusoff form.

A useful resource is the IC_{50}-to-K_i calculator, which is a web-based tool for converting IC_{50} to K_i values for inhibitors of enzyme activity and ligand binding (*see* **Note 25**).

It can be informative to utilize these calculations to demonstrate that the assay is performing as expected, as it is possible to change IC_{50} values by altering the [ligand] and/or [protein] in the assay. When the assay is working as expected, these equations can be used to illustrate that under different conditions, different IC_{50} values do correct to yield a constant K_i value.

3.4.2 Identifying Compound Interference

Sometimes, problems resulting from interference by test compounds, which are typically added in great excess compared to the probe concentration, can occur in these assays, and examples of these are discussed below.

Autofluorescence displayed by the compound will result in a spuriously low anisotropy value because the vast excess of this compound, whether it binds the target or not, will be free in solution. This low value tends to lead to false-positive results, as

the compound appears to have displaced the probe, even though it may not have done so. In a high-throughput fluorescence polarization screen, Turconi et al. [19] found that the most common cause of compound interference was autofluorescence. Autofluorescent compounds may easily be identified by measuring total fluorescence intensity. Wells where the total intensity is statistically significantly greater than that of control wells should be flagged as potential artifacts. It may be possible to correct for this effect where the level of compound fluorescence is relatively low, but a preferred approach is to employ a probe carrying a label that is fluorescent at a wavelength that is significantly different from the wavelengths where compounds in the collection may interfere.

Compounds may also interfere if they aggregate or precipitate during the assay. Such aggregates or precipitates are often highly anisotropic due to light scattering. This results in very high polarization values that could mask a true displacement effect, leading to false-negative results.

Another potential issue arises from inner filter effects if compounds absorb the excitation or emitted light (*see* **Note 26**). However, as fluorescence polarization is a ratiometic technique, this effect is often well tolerated, and checks can be made by carrying out absorbance scans on compounds where this effect may be suspected.

4 Notes

1. Many biochemical interactions are affected by changes in pH. The H^+ concentration in vitro must therefore be controlled by adding a suitable buffer to the medium, without affecting the function of the system. A buffer keeps the pH of a solution constant by taking up protons when released during reactions, or by releasing protons when they are consumed by reactions.
2. The buffer capacity represents the amount of H^+ or OH^- ions that can be neutralized by the buffer. The buffer capacity is related to the buffer concentration and the pH at which it is used relative to the buffer pK_a value. Generally, buffers may only be reliably used within a pH range of one pH unit above or below the pK_a.
3. HEPES is a member of the list of 12 buffers produced by Norman Good [20] and has favorable qualities for biochemical assay development. HEPES has a pK_a of 7.55, at 20 °C, which is close to physiological pH. Tris is also often used but has poor buffering capacity below pH 7.5. Care is needed to ensure that the pH is measured at the temperature at which the buffer will be used, as pK_a varies with temperature.

4. Metal ions are chosen where they are needed as cofactors and are added at concentrations to mimic the physiological environment of the target. Experiments conducted during assay development determine the optimum concentration of cofactors.
5. Reducing agents are used to reproduce the reducing environment within cells and to prevent disulfide bond formation and potential aggregation of cytosolic proteins.
6. Detergents are added to help prevent protein adherence to plastic as well as to solubilize test compounds and to prevent their aggregation. Nonionic detergents (Tween 20, Triton X-100, BRIJ, Octyl β-D-glucopyranoside, etc.) are often preferred, as they are less harsh and are nondenaturing compared with ionic detergents. Thus, they are often employed when maintaining protein structure is important. The choice of the optimal detergent and the concentration for use is made during assay development.
7. Buffer components may be prepared by sterile filtration (through a 0.22 μm filter) or by autoclaving. Heat-labile components should not be autoclaved. The literature has mixed opinions on whether HEPES can be safely autoclaved.
8. The high-molecular-weight binding partner should be physiologically relevant and as pure as possible to minimize the presence of nonspecific binding sites. Knowledge of the number and nature of binding sites on each molecule is useful, and analysis is simplified when there is a single site.
9. The fluorescent low-molecular-weight binding partner can often be identified from the literature. Alternatively, chemical fluorescence labeling of known tool compounds can be undertaken. Reference 8 contains several examples of the generation of FP probes using several different fluorescent dyes. Fluorescence labeling of tool compounds will frequently yield a molecule that will still bind to the target protein but with different affinity compared to the parent molecule. For high-throughput screening applications, such tool compounds are preferentially labeled with dyes such as Alexa647® because the excitation and emission wavelengths are in the far-red part of the electromagnetic spectrum, where interference due to compound absorption is significantly reduced. The Alexa Fluor family of fluorescent dyes is a series of dyes invented by Molecular Probes (now part of Thermo Fisher Scientific and sold under the Invitrogen brand name). The dyes can be conjugated directly to a range of biomolecules.
10. Buffer is drawn into a syringe, an acrodisc (or similar 0.22-μm filter) is added, and the syringe contents are filtered into a sterile container.

11. The polarization and anisotropy values contain the same information. The main difference is that the value $I_{\parallel} + 2.I_{\perp}$ is equal to the total fluorescence of the fluorophore and accounts for emitted light that is perpendicularly polarized in the third dimension and is therefore not captured by any detector. Anisotropy may often be preferred because it is normalized by the total intensity and so simplifies some of the equations. P has physically possible values ranging from −0.33 to 0.5, although measured values in biochemical studies typically range from 0.01 to 0.3 or 10 to 300 mP (mP = $P \times 1000$). Modern instrumentation allows very precise measurements ($P \pm 0.002$ or ±2 mP), which leads to a relatively wide dynamic range.

    It is possible to convert between anisotropy and polarization as follows:

$$P = 3r/(2 + r) \text{ and } r = 2\text{P}/(3 - P) \quad (3)$$

12. Instruments such as the Pherastar will, at the same time, optimize read height and fluorescence gain for each detector to return the target mP value.
13. Ratiometric methods are based on the use of a ratio between two fluorescence intensities. This approach allows correction of potential interference due to artifacts, bleaching, and issues with variation in changes in focus or laser intensity.
14. Black plates help to avoid strong fluorescence signals affecting adjacent wells, sometimes referred to as a halo effect.
15. The fluorescence profile of more than 70,000 samples across spectral regions commonly utilized in HTS was analyzed by Simeonov et al. [21]. It was demonstrated that red-shifting the spectral window by as little as 100 nm was accompanied by a dramatic decrease in autofluorescence.
16. There is a practical limit to the IC_{50} that can be measured in any FP assay. This limit is equal to [target protein]/2.
17. In practice, the effect of binding driven quenching is an exponential function (Fig. 2).

    When measuring compound antagonism, the corrected anisotropy shifts make no difference to the measured IC_{50} value. However, the correct binding K_d for the ligand binding to the receptor must be known to correctly convert a measured IC_{50} value to the desired K_i value.
18. Competition experiments between the probe and test compounds reveal compounds which can displace the probe, and concentration responses allow the measurement of an IC_{50} value. To calculate the true affinity of the test compound requires knowledge of the probe K_d value and application of a relevant equation (*see* **Note 25**).

19. For dilution of the target protein, it is very useful if the stock concentration is at a relatively high concentration so that additional components in the storage buffer may be diluted out.
20. As a guide, if the probe can be detected at a concentration of 0.3 nM, it is only necessary to dilute the target to around 0.2 nM.
21. Carrying out intermediate dilutions in DMSO avoids decreasing the solvent concentration by the addition of buffer until the final dilution, which helps to maintain compound solubility. The final DMSO concentration is dependent on the effect of DMSO on the assay, but it is useful to keep it at or below 1% (v/v). It is also important to remember that the addition of competitor compounds will introduce a further amount of DMSO.
22. At this stage, it is also prudent to test the tolerance of the assay to different concentrations of DMSO, linked to **Note 21** above.
23. Often, a compound known to fully displace the probe is used for the measurement of this Min value.
24. The Hill slope refers to the steepness of the curve and often reflects the stoichiometry of the interaction.
25. IC_{50}-to-K_i: A web-based tool for converting IC_{50} to K_i values for inhibitors of enzyme activity and ligand binding [22]—https://bioinfo-abcc.ncifcrf.gov/IC50_Ki_Converter/index.php.
26. Inner filter effects can be classed as a primary inner filter effect—defined as the decrease in the intensity of the excitation light because of optical absorption in the excitation region, or secondary inner filter effect when the fluorescence intensity decreases because of the absorption in the emission region. For dilute solutions (where absorbance <0.01), the effect is negligible.

References

1. Powell D, Hertzberg RP, Macarrón R (2016) Design and implementation of high-throughput screening assays. Methods Mol Biol 1439:1–32
2. Janzen W (2014) Screening technologies for small molecule discovery: the state of the art. Chem Biol 21:1162–1170
3. Macarrón R, Banks MN, Bojanic D et al (2011) Impact of high-throughput screening in biomedical research. Nat Rev Drug Discov 10:188–195
4. Mayr L, Bojanic D (2009) Novel trends in high-throughput screening. Curr Opin Pharmacol 9:580–588
5. Burke TJ, Loniello KR, Beebe JA, Ervin KM (2003) Development and application of fluorescence polarization assays in drug discovery. Comb Chem High Throughput Screen 6:183–194
6. Owicki J (2000) Fluorescence polarization and anisotropy in high throughput screening: perspectives and primer. J Biomol Screen 5:297–306

7. Huang X, Aulabaugh A (2016) Application of fluorescence polarization in HTS assays. Methods Mol Biol 1439:115–130
8. Lea W, Simeonov A (2011) Fluorescence polarization assays in small molecule screening. Expert Opin Drug Discovery 6:17–32
9. Gradinaru C, Marushchak D, Samim M et al (2010) Fluorescence anisotropy: from single molecules to live cells. Analyst 135:452–459
10. Pope A, Haupts U, Moore K (1999) Homogeneous fluorescence readouts for miniaturized high-throughput screening: theory and practice. Drug Discov Today 4:350–362
11. Johnson I, Spence M (2010) The molecular probes handbook: a guide to fluorescent probes and labeling technologies 11th edition. Life Technologies, Carlsbad, California
12. van de Hulst H (1981) Light scattering by small particles. Dover Publications Inc., New York
13. Buchli R, VanGundy RS, Hickman-Miller HD et al (2005) Development and validation of a fluorescence polarization-based competitive peptide-binding assay for HLA-A*0201–a new tool for epitope discovery. Biochemistry 44:12491–12507
14. Hall M, Yasgar A, Peryea T et al (2016) Fluorescence polarization assays in high-throughput screening and drug discovery: a review. Methods Appl Fluoresc 4(2):022001
15. Dandliker W, Hsu M, Levin J et al (1981) Equilibrium and kinetic inhibition assays based upon fluorescence polarization. Methods Enzymol 74. Pt C:3–28
16. Cheng Y, Prusoff W (1973) Relationship between the inhibition constant (K1) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction. Biochem Pharmacol 22(23):3099–3108
17. Munson P, Rodbard D (1988) An exact correction to the "Cheng-Prusoff" correction. J Recept Res 8(1–4):533–546
18. Wang Z (1995) An exact mathematical expression for describing competitive binding of two different ligands to a protein molecule. FEBS Lett 360(2):111–114
19. Turconi S, Shea K, Ashman S et al (2001) Real experiences of uHTS: a prototypic 1536-well fluorescence anisotropy-based uHTS screen and application of well-level quality control procedures. J Biomol Screen 26(5):275–290
20. Good N, Winget GD, Winter W et al (1966) Hydrogen ion buffers for biological research. Biochemistry 5:467–477
21. Simeonov A, Jadhav A, Thomas CJ et al (2008) Fluorescence spectroscopic profiling of compound libraries. J Med Chem 51:2363–2371
22. Cer RZ, Mudunuri U, Stephens R, Lebeda FJ (2009) IC_{50}-to-Ki: a web-based tool for converting IC_{50} to Ki values for inhibitors of enzyme activity and ligand binding. Nucleic Acids Res 37(Web Server issue): W441–W445

Chapter 11

Fragment Screening by NMR

Ben J. Davis

Abstract

This chapter describes the use of NMR to screen a fragment library as part of a fragment-based lead discovery (FBLD) campaign. The emphasis is on the practicalities involved in fragment screening by NMR, with particular attention to the use of 1D ligand-observed ^{1}H NMR experiments. An overview of the theoretical considerations underlying the choice of method and experimental configuration is given, along with a discussion of steps that can be taken in order to minimize the risk of experimental artifacts often associated with the identification of low-affinity interactions.

Key words NMR, Fragment screening, Drug discovery, FBLD, Fragment-based lead discovery, Biophysics

1 Introduction

Since its inception in the early 2000s, fragment-based lead discovery (FBLD) has become a widely used method for hit identification in early-stage drug discovery [1–6]. "Fragment-based lead discovery" refers to the identification of low-molecular-weight molecules, which bind to the macromolecular target of interest, and the subsequent evolution of these molecules into classical highly potent ligands and drugs [7–11]. Several marketed drugs have been developed using FBLD methods [12, 13], and at least 40 more compounds are currently in clinical trials [14], demonstrating the widespread applicability and utility of the approach.

The essential premise of FBLD is to use a robust assay to screen a library of fragments and thus to identify which of those fragments interact with a defined molecular target (Fig. 1). The key feature of FBLD is that the initially screened fragments are smaller, and therefore tend to be of lower affinity, than is the case for most other hit identification methods. The initial screen must be able to identify these weak interactions, and this imposes a number of constraints on the assay used: it must be sensitive, often being required to work with a K_D in the high μM or low mM range; the

Tina Daviter et al. (eds.), *Protein-Ligand Interactions: Methods and Applications*, Methods in Molecular Biology, vol. 2263, https://doi.org/10.1007/978-1-0716-1197-5_11,

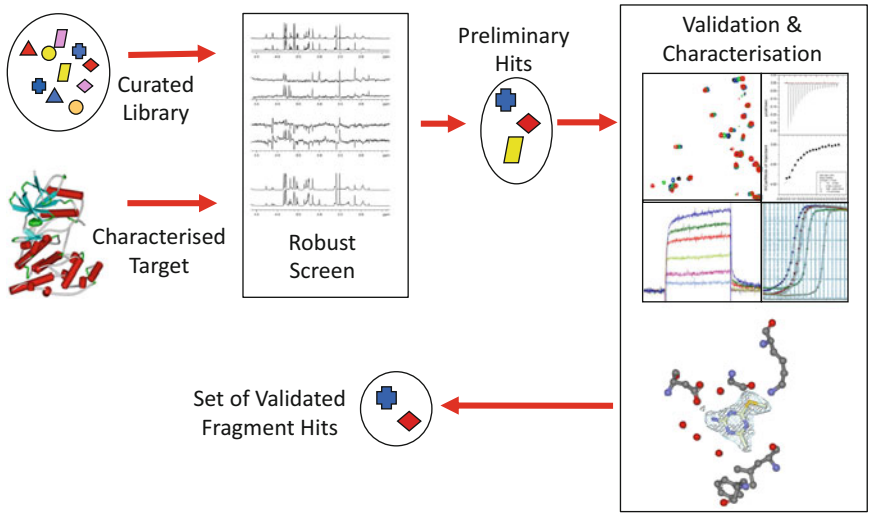


Fig. 1 An overview of the fragment-based ligand discovery process. FBLD uses a robust assay (typically a biophysical assay) to screen a curated library of low-molecular-weight compounds (fragments) for binding to a pure, characterized molecular target. Putative hits identified by this primary screen are then validated using orthogonal methods in order to exclude false positives. The binding of these validated hits is characterized further by biophysical methods, yielding a set of validated characterized fragment hits

assay must be reliable, with few false-negative results; and the assay must be robust, able to avoid (or at least to identify) artifacts that give rise to false-positive results [15]. As a result, biophysical methods have tended to dominate fragment-based screening (FBS), with NMR often regarded as the "gold standard" method for identifying and characterizing low-affinity intermolecular interactions [16, 17].

NMR is often regarded as a technique restricted to expert users, both in terms of data acquisition and data analysis. However, this chapter will attempt to describe fragment screening by NMR for the nonspecialist, focussing on ligand-observed ^{1}H-detected NMR screening experiments, with the intention of making this approach more accessible. There will be an emphasis on the practicalities and requirements specific to using NMR as an assay method to identify low-affinity interactions. The theory underlying the common NMR experiments used in FBS will be touched upon, but an extensive discussion of these experiments lies outside of the scope of this chapter; many excellent reviews have been written, which discuss these experiments in detail [16, 18–20], and the reader is urged to consult these for further information.

NMR is a solution-based technique, which can be used in near-physiological conditions; solid-state NMR is also widely used, but not typically in FBS. NMR does not require any chemical modification of protein or fragment, such as the introduction of a fluorophore or a heterogeneous phase such as a solid support. As a result,

intermolecular interactions can be robustly identified with few artifacts arising from the detection method. This is particularly important when screening for low-affinity interactions since these can be readily masked or distorted by modification of the target's environment. Additionally, NMR observes almost all species present in solution, allowing verification of identity and behavior of the compounds, protein and buffer and further reducing the risk of errors which can give rise to erroneous results.

NMR detects intermolecular interactions by observing one or more spectral parameters (such as chemical shift, relaxation rates, or the transfer of magnetization). These parameters are modulated by interactions that occur when a bimolecular complex is formed, giving rise to spectra whose features differentiate binding from nonbinding species and thus can be used to identify ligands. Numerous experiments have been proposed to identify intermolecular interactions by detecting various effects, although a relatively small set has found widespread use [16, 21]. As always, care must be taken so as not to be misled by experimental artifacts, but in general, these widely used experiments are sensitive, reliable, and robust.

In the context of FBS, most NMR experiments are categorized by the primary species observed—typically either the small molecular components ("ligand-observed NMR binding experiments") or macromolecular species (usually "protein-observed NMR binding experiments," although nucleic acids can also be readily studied by using different experiments). As mentioned previously, various NMR experiments are used to identify intermolecular interactions; the most common of these are summarized in Fig. 2 and described briefly below. The different NMR experiments are best suited to different types of target, and it is necessary to consider carefully which approach will be taken before embarking on an NMR-based fragment screen.

Protein-observed NMR (PO-NMR) was the first approach used to experimentally identify fragments binding to proteins [22] and remains widely used. Effects other than ligand-binding that may perturb the protein spectrum (such as changes in pH or the binding of DMSO [15]) must be excluded, but if these controls are performed, then PO-NMR remains one of the most reliable NMR methods used for FBS. Various spectra of the protein can be acquired in order to detect ligand-induced chemical shift perturbations (changes in frequency of spectral peaks), ranging from simple ^{1}H 1D spectra (under highly favorable circumstances) through the commonly used ^{15}N–^{1}H HSQC spectra (suitable for smaller proteins below 20–25 kDa) to the more rarely employed TROSY-type spectra suitable for application to larger systems of up to 50–80 kDa (or, under some circumstances, even higher [23, 24]). In addition to identifying intermolecular interactions, PO-NMR can be used to identify both the K_D (via titration of the ligand onto the protein)

Experiment	Category	Physical basis	Reference
Heteronuclear correlation spectra	Protein observed. Typically 15N-1H correlation, although 13C-1H is also used.	Perturbation of chemical environment of observed group by bound ligand	14
Saturation Transfer Difference (STD)	Ligand observed ^{1}H	Direct transfer of magnetisation from protein to bound ligand	17
Water-LOGSY	Ligand observed ^{1}H	Phase modulated water mediated transfer of magnetisation	18
Relaxation filtered	Ligand observed. ^{1}H and ^{19}F nuclei are widely used	Rapid T2 or T1ρ relaxation of ligand in bound state	19

Fig. 2 Properties of common NMR ligand-binding experiments

and the ligand-binding site (either by using an existing assignment of protein resonances or via comparison with the perturbations induced by a known ligand). ^{15}N–^{1}H correlation spectra are, however, sometimes prone to widespread perturbations caused by even small alterations of secondary or tertiary structure making precise localization of the binding site difficult. In such cases, ^{13}C–^{1}H correlation spectra can be more useful as they typically respond only to more localized effects [24]. Generally, PO-NMR requires large amounts of protein labeled with stable isotopes (^{13}C or ^{15}N), and in practice, this limits the widespread application to relatively small proteins expressed to high levels in bacterial systems.

A number of ligand-observed NMR (LO-NMR) experiments have been developed; the most widely used of these are the saturation transfer difference (STD) [25], water–ligand-observed gradient spectroscopy (water-LOGSY) [26], and relaxation-filtered (T2 or T1ρ filtered) [27] sequences. These experiments modulate the magnetization of the ligand in the bound state; this modulation is then transferred to the free state, where it is detected as a change in the intensity of the ligand spectrum. This requirement for modulation transfer from free to bound state limits these experiments to systems where the ligand is in fast exchange between the free and bound states, i.e., the off-rate should be much greater than the frequency difference between the two states. For FBS, this is rarely an issue because fragments are typically weakly bound with rapid binding and unbinding, but it is an important limitation that does need to be considered.

Since the spectral modulation is passed from the free state to the bound state, and this modulation persists over time, for low-affinity ligands such as fragments a "signal amplification" effect occurs when using LO-NMR. In this case, because of the rapid

binding and unbinding, a significant fraction of the free population of the ligand has been bound during the course of the experiment and so is labeled with a signal indicative of binding. The observed signal is substantially larger than the bound fraction of the ligand might indicate [28, 29], and the sensitivity of these experiments is dramatically increased when a large molar excess of the ligand over the protein is used. Also, since a large molar excess of the ligand is used, relatively little protein is required, resulting in reduced protein demands compared to PO-NMR. Additionally, since only the free population of the ligand is observed, there is no limit on the size of the protein that can be used; indeed, these experiments work best with large, slowly tumbling macromolecules. These factors have contributed strongly to the prevalence of LO-NMR as a fragment screening technique [17].

A limitation of LO-NMR is that these experiments only indicate an interaction between the protein and the ligand; unlike PO-NMR, no information is obtained as to the location of the binding site (or even whether the interaction is specific to a single site or nonspecific). A simple solution to this is to perform a competition experiment using a potent ligand known to bind to the site of interest; if the fragment is displaced by the competitor, the signal in the LO-NMR experiment will be reduced or abolished [16, 30]. However, this requires prior knowledge of a potent competitor that binds to the site of interest; while this is the case for many proteins, a significant number remain where this is not applicable.

A third class of NMR experiments widely used for FBS is ^{19}F-observed NMR. While this usually refers to a ligand-observed experiment, the considerations and requirements are sufficiently different from ^{1}H observed LO-NMR that ^{19}F LO-NMR is often considered as a separate technique. ^{19}F-observed NMR has been reviewed extensively elsewhere [21, 31, 32], and only a brief overview will be given here. The chemical shift of a ^{19}F nucleus is exquisitely sensitive to changes in the chemical environment; consequently, the large spectral width together with the low number of resonances in each fragment allows the use of mixtures containing large numbers of compounds and the simple analysis of these spectra using automated software. Thus, the differential relaxation rates of ^{19}F in the free and bound states of a fragment can be maximally exploited to rapidly identify ligands. However, the high sensitivity of ^{19}F to changes in the chemical environment also results in a high false-positive rate, and this can reduce the reliability of the data from the screen. This can be mitigated through the use of orthogonal methods (such as PO-NMR or SPR), but this in turn has implications for the amount of time and resource required for the FBS.

Because both LO-NMR and PO-NMR use large amounts of protein compared to other FBS techniques such as SPR or DSF

[33], it is common practice to screen mixtures of compounds in the same experiment. The size of the mixture varies from 4 to 10 compounds (for ^{1}H LO-NMR) to as high as 30 (for PO-NMR or ^{19}F-NMR). Where the compounds are observed directly, it is usually possible to directly distinguish ligands from nonbinding molecules based on their 1D NMR spectra if care is taken in order to minimize spectral overlap within the mixture. In the case of PO-NMR, deconvolution of the mixture (e.g., by subsequent analysis of subsets of molecules) or the application of orthogonal methods must be applied to identify which specific ligand is responsible for the observed chemical shift perturbations. One potential issue when screening mixtures of compounds is the potential for compound–compound interactions. Although rare, these interactions do occur [15], and so ligands identified directly from mixtures should be tested as single compounds before confirmation as a putative hit from the NMR screen.

Having determined which set of NMR experiments are most suitable for the target in question, several key steps need to be undertaken before the screen can begin. As discussed previously, the focus of this chapter is on ligand-observed fragment screening by NMR (LO-NMR) using ^{1}H-detected experiments since this is the most commonly used experimental configuration. Many excellent reviews of PO-NMR and ^{19}F-NMR FBS have been published, and these cover details, which are outside of the scope of this chapter. However, some features of NMR FBS are common to all approaches. (1) A screening library of fragments is required; in addition to the typical principles used to design a fragment library [34–36], these fragments need to be compatible with the NMR screening method to be applied. The fragments also need to be characterized in terms of solubility, stability in both DMSO and aqueous solutions [15], and PAINS-like behavior [37]. (2) Characterization of the protein and any competitor ligands is vital in order to avoid erroneous results caused by unusual behavior of these molecules. NMR screening experiments are relatively time consuming (often taking several days to run a complete screen), and so the prior assessment and optimization of the stability of the protein (*see* Chapter 1) and competitor ligands is key.

After identification of suitable conditions, the NMR screen can then be completed using the methods discussed previously. The screening data are then analyzed, either manually or in a semiautomated manner, and a set of preliminary fragment hits identified.

An important final step, which lies outside of the scope of this chapter, is the validation of these initial fragment hits using orthogonal methods. Often this validation step will be performed using an alternative biophysical method (such as SPR, MST, or X-ray crystallography). However, PO-NMR is also often used as a validation method orthogonal to an initial screen performed using LO-NMR or ^{19}F-NMR. This case will be discussed further below.

2 Materials

2.1 NMR Spectrometer and Accessories

Although the exact materials required for FBS by NMR will vary dramatically depending on the target, certain features will remain constant. The most central of these is, obviously, the NMR spectrometer. Any modern NMR spectrometer should be capable of running the experiments described in Subheading 1, although the exact implementation will vary between vendors and local installations. Ideally, the NMR spectrometer should be of 500 MHz or higher field strength, with a cryogenic probe to enhance sensitivity and shorten the time for each experiment and automated sample handling. If ^{19}F-NMR is to be used, a probe capable of ^{19}F acquisition with ^{1}H decoupling is strongly recommended.

A hand centrifuge is strongly recommended in order to allow spinning of the NMR sample tubes. These are widely available from laboratory equipment suppliers. Numerical labeling of the centrifuge rotor buckets is advised in order to allow ready identification of samples after spinning.

2.2 Fragment Library

The second key requirement is a suitable fragment library. A number of publications describe the generation of such libraries [34–36]; additionally, many chemical vendors sell preselected fragment libraries. Some libraries are also available from academic groups or institutions. A recent survey of commercial fragment libraries found that a relatively small number of vendors were commonly used to obtain fragment libraries [38]. Regardless of the source of the library, it is important to consider the suitability of the fragment library for the NMR screening technique to be used.

For LO-NMR, buffer components may overlap extensively with resonances (particularly aliphatic resonances) from the fragments, confounding analysis of the LO-NMR experiments. If the protein is stable in a phosphate buffer, this can be used to reduce buffer overlap issues; alternatively, deuterated buffers can be used to good effect. If possible, it is advantageous to have one or more aromatic protons on the fragments since these are rarely overlapped with buffer resonances. Compounds with no observable protons (typically no C–H bonds) should be excluded from a fragment library intended for use with ^{1}H LO-NMR. Although probably unavoidable, it is important to be aware that compounds with aliphatic methyl groups may cause issues with artifactual direct irradiation in the STD experiment, while compounds with exchangeable protons will give rise to artifactual positive signals in the water-LOGSY experiment. All compounds should also be soluble in water to at least 250–500 μM.

PO-NMR is less sensitive to the chemical nature of the fragment than is LO-NMR. However, the solubility requirement is rather more stringent since it is advisable to titrate any putative

hits to a point where a degree of saturation of the binding curve is observed (typically 3–4 times K_D). This can often require compounds to be soluble to the low mM range.

The requirements for compounds for a ^{19}F-NMR FBS are rather different from those discussed above. Obviously, the compound must contain at least one ^{19}F atom; typically, only one C–F or –CF3 group per molecule is preferred in order to reduce spectral overlap. –CF2 groups can be used, although fluorine coupling can be a problem in this instance. However, compounds are typically screened at 25–50 μM, and as a result the solubility requirement is significantly less demanding than is the case for LO- or PO-NMR.

2.3 Protein

For all NMR FBS, relatively large amounts of pure protein are required. Assuming a 2000 compound fragment library and a 25-kDa protein, approximately 1 mg (^{19}F-NMR), 30 mg (^{1}H LO-NMR), and more than 200 mg (PO-NMR) are required. For PO-NMR, labeling with a stable isotope (usually ^{13}C or ^{15}N) is typically required. This protein should be as homogeneous and pure as possible and importantly should be stable for the duration of the NMR screen (often several days).

2.4 Tool Compounds

If possible, a potent competitor molecule should be identified, which binds to the site of interest. Although not absolutely required, observation of displacement of weak binding fragments by a potent competitor increases the confidence level of the fragment screen. This competitor is typically a compound known from the literature, although proteins, peptides, or other tight binding molecules can also be used effectively. The K_D for this competitor molecule should be μM or better.

If available, a low-affinity molecule with a K_D of 100 μM or greater can be used to identify conditions where weak binding can be reliably identified. This molecule can be a fragment of a larger known ligand, or another molecule such as a substrate or small peptide.

2.5 Analysis Software

Several software tools are available which assist in the analysis of NMR binding experiments. The most widely used are Fragment-Based Screening (Bruker, https://www.bruker.com), MNova Binding (Mestrelab, http://mestrelab.com), and ACD/Labs (https://www.acdlabs.com). Each of these can assist in the organization and analysis of the extensive data generated during an NMR FBS campaign, with visualization and guided spectral interpretation being common features. However, it is also possible to analyze the NMR data acquired manually using freely available NMR tools, although this requires careful and consistent use of a database in order to track the results from the screen.

3 Methods

An overview of FBS using LO-NMR is shown schematically in Fig. 3. Certain steps (such as the characterization of the fragment library) need to be done only once, with periodic checks of compound stability. Other steps (such as characterizing the protein or ensuring that binding can be observed with a control ligand) must be repeated with each target or screen. These steps will be discussed in detail below.

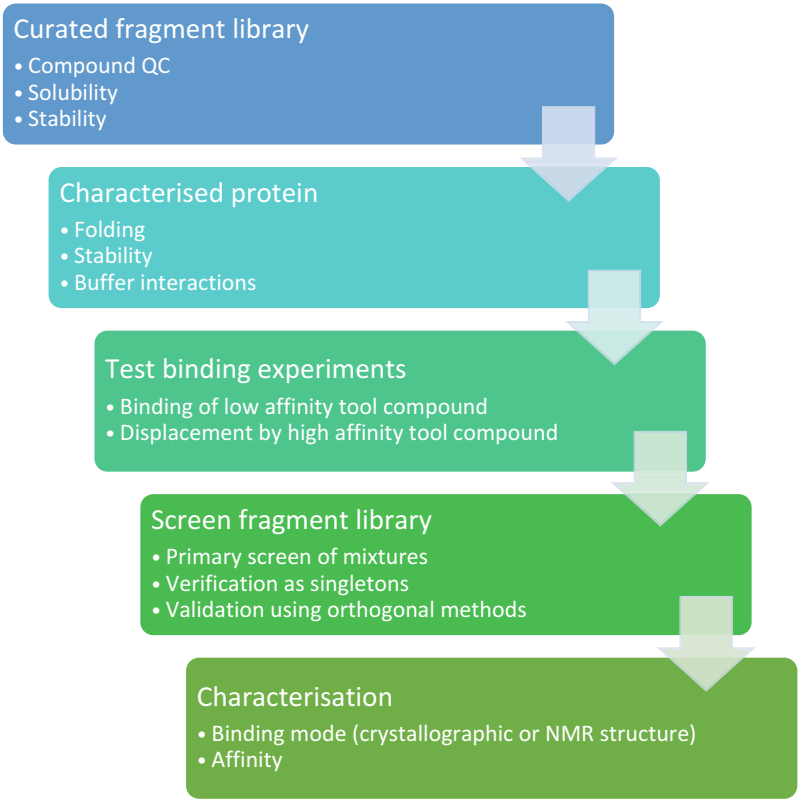


Fig. 3 An overview of the NMR fragment-based screening process. A curated fragment library is prepared following experimental verification of potential fragments, with particular emphasis on compound QC, solubility, and stability. The protein target is characterized to confirm fold and stability, and interactions with the buffer identified. If possible, the NMR binding experiments are then tested to confirm that binding of a known low-affinity binding (tool) compound can be observed and that this compound is displaced by a high-affinity tool compound. The curated library is then screened in mixtures, followed by verification of individual (singleton) compounds; these compounds are validated by orthogonal methods and further are characterized to determine affinity and structural data

3.1 Characterization of the Fragment Library

Before using a fragment library for screening, it is important to characterize the fragments thoroughly—both in order to verify the compound identity and also to reduce the risk of artifacts associated with poor compound behavior (such as instability or insolubility) in the assay. This characterization can be performed in a number of ways, but the focus here will be on NMR characterization (since the library is to be screened by NMR). Additionally, the library should be quality controlled periodically (typically every 1–2 years) in order to identify degradation or precipitation of the stock solutions.

3.1.1 Sample Preparation

Stock solutions of all compounds intended for inclusion in the library should be made up at 200 mM in d6-DMSO. This high concentration enables cocktails of compounds to be prepared while keeping the total concentration of d6-DMSO at levels that are tolerated by most proteins (*see* Subheading 3.2). Samples of each compound should then be made in buffered aqueous solution; typically, a concentration of 500 μM compound is used in QC buffer (such as 20 mM sodium phosphate (pH 7.5), 2% final concentration of d6-DMSO, 10% D_2O, 50 μM DSS). A large number of samples will typically be required in order to characterize the entire fragment library (*see* **Note 1**).

3.1.2 Data Acquisition

1D ^{1}H NMR spectra should be acquired on each aqueous sample, using a robust solvent suppression method (such as excitation sculpting [39]) in order to remove the solvent signal (*see* **Note 2**). If the compound contains ^{19}F nuclei, a ^{19}F 1D NMR spectrum should be acquired at this point. Additionally, it is useful to acquire a water-LOGSY spectrum [26] on this sample in order to identify compounds that readily self-associate [34].

Since fragment screening by NMR typically takes 1–2 days of acquisition time, it is important to confirm that the fragments are stable under "typical" aqueous conditions for this length of time. This is simply done by repeating the data acquisition after 24 h using the same samples. Examination of the two sets of spectra rapidly reveals samples that are not stable under these conditions.

3.1.3 Data Analysis

For each compound, the 1D ^{1}H NMR spectrum should be analyzed in order to confirm that the spectrum is consistent with the chemical structure. If the spectrum is not clearly consistent, it may be necessary to acquire a 1D ^{1}H spectrum of the compound stock in DMSO and/or to acquire LCMS data in order to check the identity of the compound (even commercially supplied compounds are not always what they are supposed to be). Additionally, the 1D ^{1}H NMR spectrum acquired after 24 h should be analyzed in order to identify compounds that are not stable for extended periods in aqueous solution. Compounds that fail either of these tests should not be included in the fragment library.

In addition to verifying the structure and stability of each compound, it is also necessary to confirm that the compound is soluble at the expected concentration. This is readily done by comparison of the compound resonance integrals with the integral of an internal standard (such as 50 μM DSS (*see* **Note 3**)). Compounds that are not soluble at the required concentration should not be incorporated into the screening library.

A final step in characterization of the compounds to be included in the fragment library is to analyze the water-LOGSY spectrum acquired on each compound sample. The water-LOGSY spectrum is sensitive to association with slowly tumbling macromolecules and as such can be used as an indicator of compound self-association to form microaggregates. Although many factors influence the intensity and magnitude of the water-LOGSY spectrum, self-association will dominate the observed spectrum in a sample containing only compound and buffer. If the compound is self-associating, a positive water-LOGSY spectrum is observed and the compound should be excluded from the library; the formation of microaggregates can give rise to many artifactual signals and complicates analysis of the water-LOGSY spectra acquired during the fragment screening procedure.

After data analysis and subsequent selection of compounds suitable for incorporation in the fragment library, it is useful to store the reference spectra for each compound in a spectral database. The procedure for doing this will obviously depend on the analysis software used, but most modern NMR analysis packages include a spectral database functionality.

3.2 Characterization of the Target Protein

Characterization of the protein is a key part of establishing the NMR fragment screening conditions. The protein should be as pure as possible (>95% by SDS-PAGE), and ideally, the identity of the protein should be confirmed (for example, by intact mass spectrometry or peptide-mass fingerprinting [40, 41]). Samples of the protein in a suitable buffer should be prepared (*see* **Note 4**), and ^{1}H 1D NMR spectra of the protein should be acquired in order to confirm that the protein is folded correctly. ^{1}H 1D or ^{15}N–^{1}H HSQC spectra should be used to identify interactions with buffer components such as metal ions or detergents [42].

Acquisition of NMR spectra should be repeated at intervals to confirm that the protein is stable under these conditions for the expected duration of the fragment screening. It is also advisable to examine the protein by SDS-PAGE and/or intact mass spectrometry after the final experiment to confirm that no degradation has occurred. If multiple batches of the protein are to be used, 1D ^{1}H NMR spectra can be used to confirm the batch-to-batch consistency of both protein and buffer.

3.3 Development of NMR Fragment Screening Assay

Although many NMR binding experiments have been described in the literature, this chapter will focus on the application of the commonly used set of 1D ^{1}H NMR, STD, water-LOGSY, and relaxation-filtered 1D with solvent suppression experiments as discussed above. These ligand-observed NMR binding experiments should be implemented and tested as described elsewhere—*see* **Note 2** and Ref. 16.

Depending on the details of the protein and spectrometer, one or more of these experiments may be excluded. For example, the STD experiment has relatively low signal/noise, and is consequently often the longest of the acquisitions, and thus may not be suitable to be run in a reasonable time on a lower field instrument with a room temperature probe. The water-LOGSY experiment has also been observed to give rise to anomalous spectra for some protein:buffer combinations and may not be suitable to be acquired in all circumstances [42]. In general, however, it is usually better to acquire more experiments rather than fewer. For each sample, during assay development, trial screening, and screening of the full fragment library, the same set of experiments should be acquired.

The simple ^{1}H 1D with solvent suppression should always be acquired since this experiment allows inspection of almost all components present in the assay. As such, it is advisable to acquire a 1D ^{1}H spectrum with sufficient signal/noise to observe the protein resonances since this enables identification of situations where the protein has denatured, refolded, or precipitated during the experiment.

3.3.1 Binding of a Low-Affinity Tool Compound

The first step in developing the NMR fragment screening assay is to establish a positive control, i.e., to confirm that binding of a known low-affinity ligand (tool compound) can be observed using the LO-NMR experiments (*see* **Note 5**). Binding in this case is taken as a positive signal in the STD spectrum, a shift toward a more positive signal in the water-LOGSY spectrum and a reduction in signal in the relaxation-filtered spectrum. If binding is not observed for a known low-affinity ligand, the first parameter to change is to increase the concentration of protein in the sample; if this does not result in observation of binding, the concentration of the low-affinity ligand should be increased. Additional parameters (such as STD or water-LOGSY mixing times) can also be modified, although it is also important at this point to verify that the low-affinity ligand is binding to the target protein through the use of orthogonal methods such as protein-observed NMR or another biophysical technique.

If no low-affinity tool compound is available, the trial screen (*see* Subheading 3.3.4) should be run using "default" parameters, and the hit rate analyzed. Conditions that give hit rates of 2% or less should be modified as discussed for situations where binding of the low-affinity tool compound is not observed.

3.3.2 Displacement of the Tool Compound by a Potent Competitor

Once binding of a low-affinity ligand has been observed, a potent competitor can be used to displace the ligand. The concentration of competitor required for full displacement of the low-affinity ligand should be determined by titration.

If no potent competitor molecule is available, the screen can still be acquired. However, the rate of false positives is likely to be significantly higher than is the case where a competition step is possible. This will increase the overall hit rate and will also increase the burden on subsequent activities such as singleton validation and validation using orthogonal biophysical techniques. Where no competitor ligand is available, the stringency of subsequent validation assays must be high in order to reduce the number of false positives diluting the pool of fragment hits (*see* Subheading 3.8).

3.3.3 Stability of Binding and Competition

Having determined the concentration of competitor required in order to observe displacement of the low-affinity tool compound, a number of samples (typically 4 or 5) containing protein and the low-affinity tool compound should be prepared. The competitor should then be added to different samples over time in order to observe the stability of binding of the low-affinity ligand and degree of displacement by the competitor ligand. Typically, time intervals of 0, 4 h, 8 h, 24 h, and 48 h are used, although this depends on the total expected duration of the NMR-observed fragment screen.

If stable binding is not observed, it may be necessary to optimize buffer conditions (or possibly modify the protein construct) in order to increase the stability of the system. It is also possible to screen the fragment in library in subsections such that each experiment duration is relatively short; although this increases the overall duration and workload of the screen, the investment is often worthwhile for the increased quality of the data.

If stable competition is not observed, the stability of the competitor molecule should be examined—this is the most commonly observed cause of variable degrees of competition. Degradation over time of the competitor molecule can readily lead to decreased levels of displacement, and once identified, this is readily overcome by preparation of fresh samples of the competitor.

3.3.4 Trial Screen

After identifying conditions where the binding of low-affinity fragments can be readily observed, along with displacement by a potent competitor molecule, it is productive to run a "trial" screen of 100–200 fragments. This screen is run in order to identify any issues regarding the stability of protein or fragments to optimize experimental parameters and to determine an approximate hit rate for the target. Issues identified with the trial screen should be rectified before proceeding to the full fragment screen.

Samples for the trial screen are prepared as relatively small cocktails of 4–8 compounds per mixture, and the LO-NMR experiments acquired on each sample. The competitor molecule is then added to each sample, and the LO-NMR experiments reacquired on the samples containing competitor. Analysis of the trial screen is then performed (*see* Subheading 3.5), and the number of putative hits (which show binding and displacement) is calculated.

Typically, a hit rate of 2–10% is observed for most targets. If the hit rate is below 2%, conditions and/or experimental parameters should be optimized using a low-affinity tool compound; at this point, a preliminary fragment hit can be used as a low-affinity tool, although further biophysical validation is advised if this is to be the case in order to reduce the risk of being misled by experimental artifacts. If the hit rate is above 10%, it may be worthwhile increasing the stringency of the screen (for example, by reducing the concentration of protein and/or fragments) so as to identify only apparently higher affinity hits and avoid creating a large pool of fragments requiring additional downstream validation.

It is important to consider at this point where the putative hits are located in the course of the screen. If the rate of identified binders is higher at the beginning of the screen than the end, this can be a strong indication that the system is not stable and further experimental optimization is required. Similarly, an increased rate of noncompetitive binders toward the end of a screen can be an indication of protein degradation or other instability.

The trial screen is also useful for determining the number of compounds that should be present in each mixture in the screen. If the hit rate is high, relatively few compounds (for example, 4–6) should be included in each mixture in order to reduce the number of samples that contain more than one putative hit. If the hit rate is low, the number of compounds per mixture can be increased (for example, to 8–12).

3.4 Screening the Library

After running the trial screen, the following points should be confirmed:

1. Low-affinity fragments can be observed to bind.
2. These fragments are (in some or all cases) displaced by a potent competitor.
3. The degree of binding and competition is stable over time.
4. The hit rate is constant over time.
5. The mixture size is appropriate for the observed hit rate.

If these points are valid, the full library should be screened under these conditions.

Samples should be prepared (*see* **Note 1**), and NMR experiments acquired under automation. Details of the experimental acquisition will obviously vary according to both the specific

spectrometer and the type of sample changer used, and the manufacturer's instructions should be followed for these details. A 1D ^{1}H NMR spectrum with solvent suppression should be acquired on each sample, along with the required LO-NMR binding experiments.

After acquisition of the first set of LO-NMR experiments, a small volume of competitor ligand should be added to the required concentration as discussed previously. The set of LO-NMR experiments should then be acquired again, allowing comparison of the observed binding before and after the displacement step. If no potent competitor is available, this step can be excluded, although subsequent validation steps will be required to be stringent (*see* Subheading 3.8).

3.5 Data Analysis

The signal in each LO-NMR experiment reflects different aspects of the sample and of the binding of any ligands present. A brief overview of these aspects will be discussed below, followed by a discussion of the analysis of these data.

3.5.1 1D ^{1}H NMR

As mentioned, the ^{1}H 1D NMR experiment contains information on almost every component present in the sample, the exceptions being molecules that contain no observable resonances (such as phosphate buffer, metal ions, or deuterated solvents) and those macromolecules that are tumbling so slowly as to give rise to signals that are too broad to be observed (such as protein or compound aggregates). Of particular importance are the resonances from the ligands, the protein, and the buffer components.

Resonances from the ligands should be compared with the reference spectra acquired for the isolated compounds. Minor perturbations in the spectra of the compounds are typically the result of small differences in the pH or ionic strength of reference and mixture samples. However, significant differences between the observed and reference spectra are often indications of compound degradation (either of the DMSO stocks themselves or subsequent to their being diluted in aqueous solution) or of interactions between compounds present in the mixture. Such differences can significantly complicate analysis or give rise to artifactual results. Importantly, ligand resonances should not change significantly between the "before" and "after" addition of competitor; loss of ligand signal between these two conditions is typically associated with compound precipitation and should not be interpreted as displacement (particularly in the STD experiment, see below).

Inspection of resonances from the protein, particularly within the methyl envelope (between 0.7 and 1.1 ppm) and any resolved shifted aliphatic resonances (below 0.7 ppm) can readily identify differences for samples where the protein has precipitated or denatured. This typically occurs as a function of the one or more compounds present in a mixture and is one of the most common causes of false positives and negatives in screening experiments.

3.5.2 STD

Signal in the STD spectrum arises from the transfer of saturation from the protein directly to bound ligands; this saturation then persists after dissociation and causes the build-up of a population of ligand with saturated resonances, which is manifested as a positive signal in the saturation transfer difference experiment [25]. The signal in the STD spectrum therefore reflects the bound population of the ligand and indicates which molecules are binding to the protein. After displacement by a potent competitor, the STD signal for a ligand that is binding specifically to the protein will be substantially reduced.

However, artifactual signal can occur in the STD following direct saturation of ligand resonances, particularly where the ligand contains aliphatic groups with chemical shifts that are relatively close to the saturation frequency (within 1 ppm or so, depending on the shape used for selective saturation of the protein). In this case, a large STD signal is typically seen for the ligand (particularly for the aliphatic resonances that are directly saturated), and this STD signal is not perturbed by the addition of a potent competitor molecule.

3.5.3 Water-LOGSY

In the water-LOGSY experiment, magnetization is transferred from excited water molecules directly to the compound [26]. If the compound is free in solution, these water molecules form the hydration shell, and the rapid tumbling of the compound and water molecules gives rise to a negative signal. If the compound is bound to a protein, the primary water molecules giving rise to signal are those present in and around the ligand-binding site on the protein. Since this system is tumbling slowly, a signal with a positive sign is observed; this positive signal persists into solution after dissociation from the protein.

The resultant signal is therefore a function of the free unbound population (with negative signal that depends on the nature of the hydration shell of the free molecule) and the population that has been bound to the protein (with positive signal that depends on the presence of waters associated with the ligand in the bound state). This means that compounds that bind may still give rise to a negative signal if the contribution from the free state is large and/or that from the bound state is small.

In our experience, the most reliable interpretation of water-LOGSY spectra relies on perturbation of the free and bound populations following the addition of a potent competitor. Displacement of the fragment from the binding site by the competitor results in an increase in the unbound population, which gives rise to a more negative signal regardless of the magnitude of the contributions from the free and bound states of the ligand. This competition step is not strictly required for interpretation of the water-LOGSY spectra, but does significantly increase the confidence level of the result.

3.5.4 Relaxation-Filtered 1D

For slowly tumbling molecules, magnetization relaxes rapidly via the T2 or T1ρ routes, while for rapidly tumbling molecules, these relaxation rates are significantly slower. This differential relaxation is used in the relaxation-filtered 1D to selectively attenuate the signal of molecules bound to a slowly tumbling protein compared to those molecules that are free in solution. Since the magnetization does not recover after dissociation, the relaxation-filtered 1D therefore reflects the unbound population of the ligand. After displacement by a potent competitor molecule, the unbound population of the compound is increased and the signal in the relaxation-filtered 1D increases as a result.

3.5.5 Combined Data Analysis

Since each experiment contains data reflecting different aspects of the sample and of the interactions between protein and compounds, it is useful to analyze all acquired data in order to identify any potential ligands. We have found a simple empirical grouping system to be useful when analyzing binding data from multiple LO-NMR experiments, where "class 1" refers to a compound showing binding and displacement in all three acquired experiments, "class 2" refers to a compound showing binding and displacement in any two of the three experiments, and "class 3" refers to a compound showing binding and displacement in only one of the three experiments (Fig. 4).

However, when considering this combined analysis of all acquired data, it is important to recognize that many phenomena can give rise to false-positive or negative results, and as such consistent behavior across multiple experiments is an indicator of increased confidence rather than a prerequisite for classification of a compound as a putative ligand. Ligands with high confidence levels should be prioritized for subsequent validation steps, but if resources allow, then all putative ligands should be characterized in order to find as complete a set as possible of fragments that bind to the protein target.

3.6 Analysis Software

Several software packages are available that assist in analyzing the data from a LO-NMR FBS (*see* Subheading 2.5). These software tools apply the analysis principles discussed above in an automated or semiautomated manner to assist the user in analyzing the large amount of data generated during a screening campaign. The reader is advised to contact the software vendors directly for more details on the availability and use of these tools.

3.7 Singleton Validation

Compounds identified as putative ligands from the analysis of samples containing mixtures of compounds should be verified as "singletons." A sample is prepared containing the compound of interest along with protein and buffer as determined previously. LO-NMR experiments are then acquired as described above, competitor molecule added, and the set of LO-NMR experiments

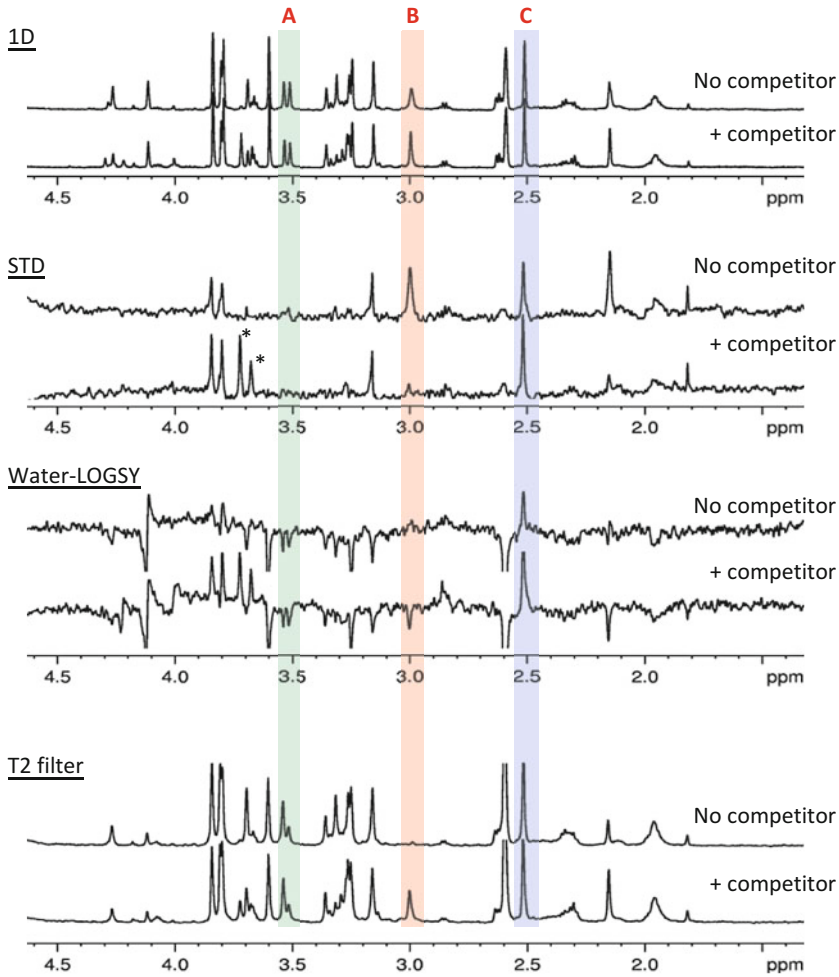


Fig. 4 Example spectra and data analysis. LO-NMR experiments acquired on 10 μM Hsp90 with a mixture of 12 compounds (each at 500 μM). Spectra were acquired before (upper) and after (lower) the addition of 100 μM PU3 competitor [44]. Resonances from the methoxy groups of PU3 are clearly visible in the STD spectrum after addition of PU3 (marked with *), confirming addition of the competitor. Three compounds are highlighted (A, B, and C). All three compounds are visible in the 1D NMR spectrum before and after addition of competitor, indicating that no precipitation or degradation of the compounds has occurred. Compound A shows no binding response in any of the LO-NMR experiments (no signal in the STD spectrum, negative signal in the water-LOGSY spectrum, and positive signal in the T2-filtered spectrum) and is classified as "not binding." Compound B shows binding and displacement by PU3 in all of the LO-NMR experiments (positive signal in the STD experiment, which is attenuated on addition of competitor; no net signal in the water-LOGSY experiment, which shifts to negative after addition of the competitor, indicating an increase in the unbound population; no signal in the T2-filtered experiment, which increases on addition of competitor). Compound B is therefore classified as a "class 1" hit. Compound C shows binding in the STD and water-LOGSY experiments, but these signals are not affected by addition of PU3. The compound is therefore classified as a "noncompetitive" hit, binding either nonspecifically or at a site that is unaffected by competitor binding

acquired. Analysis of the data is then performed, and compounds prioritized for further validation depending on the confidence levels of the singleton LO-NMR data.

3.8 Orthogonal Biophysical Validation

After the primary screen and the singleton validation steps, a set of putative fragment ligands is identified. However, these compounds may still contain a number of false positives, and further validation steps are strongly advised before embarking on a medicinal chemistry campaign.

Typically, an orthogonal biophysical technique is used as a validation step at this point. A number of biophysical techniques can be used at this point, the most common being SPR, X-ray crystallography, or protein-observed NMR [17]. A crystal structure of the ligand bound to the protein target is often regarded as a prerequisite for further progression of a fragment, although a range of methods exist that allow fragment evolution in the absence of a crystal structure [43]. However, discussion here will be limited to validation via protein-observed NMR (PO-NMR) since other techniques lie firmly outside of the remit of this chapter.

In order to validate the putative ligand via PO-NMR, the compound is titrated onto the isotope-labeled protein; perturbations of the NMR spectrum that occur in a dose–response manner are taken to be indicative of binding (*see* **Note 6**). If possible, separate samples should be used with a constant concentration of DMSO in order to reduce the potential for false positives resulting from the concomitant titration of DMSO alongside the compound. Alternatively, a separate DMSO titration can be performed against protein to serve as a control for a series of putative ligands. Additionally, if any of the compounds are charged, then a control experiment where a simple acid or base is titrated onto the protein is strongly advised in order to identify possible pH-related artifacts.

Further validation of a putative ligand can be obtained via examination of the pattern of observed chemical shift perturbations (CSPs), particularly where ^{15}N–^{1}H or ^{13}C–^{1}H correlation spectra are acquired. A CSP pattern that is localized to a region of the protein (where sequence specific assignments are available), or which is similar to that observed for a known ligand or substrate (where sequence specific assignments are not available), is a strong indicator that the putative ligand interacts with a defined binding site on the protein. ^{15}N–^{1}H correlation spectra typically show more widespread CSP patterns than do ^{13}C–^{1}H spectra, owing to the propagation of chemical shift perturbations along and across secondary structure elements, but both types of spectra can be reliably used to distinguish true ligands from false positives.

4 Notes

1. Fragment screening by NMR typically requires the preparation of large numbers of samples, often 50–100 or more at a time. While these can be prepared manually, automated (such as a liquid handling robot) or semiautomated approaches (such as a multichannel pipette) significantly reduce the manual workload required. A number of experimental tips are useful to consider when using multichannel pipettes to prepare samples for NMR.
   (a) It is advised to prepare the samples in deep-well 96-well plates to allow thorough mixing and to transfer the prepared samples to NMR tubes positioned in empty 96 position pipette tip racks; these racks hold the sample tubes securely and in the correct format for loading with a multichannel pipette (Fig. 5a). Sample racks can also be printed readily using a 3D printer.
   (b) Gel-loading tips are useful when loading NMR tubes with a multichannel pipette to avoid the formation of air bubbles close to the top of the sample tube.
   (c) Gentle spinning of the sample tubes in a hand centrifuge improves the quality of the shimming by ensuring uniform sample depth and removing bubbles; labeling of the rotor buckets allows ready sample tracking (Fig. 5b). A hand centrifuge is preferred to a benchtop centrifuge in order to reduce the risk of breaking the NMR tubes.
2. Commonly used acquisition parameters and details of LO-NMR experiments:
   (a) Excitation sculpting [39] has proven to be a robust method of solvent suppression suitable for use under automation.
   (b) A 2 s relaxation delay in the 1D ^{1}H NMR experiment is sufficient to allow approximate quantitation of the ligand concentration with respect to an internal standard.
   (c) In our experience, a T2 relaxation filter using a CPMG train is reliable and generic and is preferred over a T1ρ filter using a spinlock.
   (d) Mixing delays of 2.2 s (STD), 2 s (water-LOGSY), and 400 ms (T2 relaxation filter) are suitable for most LO-NMR FBS situations. Longer CPMG filters are advised only when fragments are binding with low affinity ($K_D \geq 1$ mM).
3. Approximate compound concentration can be readily determined from a 1D ^{1}H NMR experiment acquired with 50 μM DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid) present in

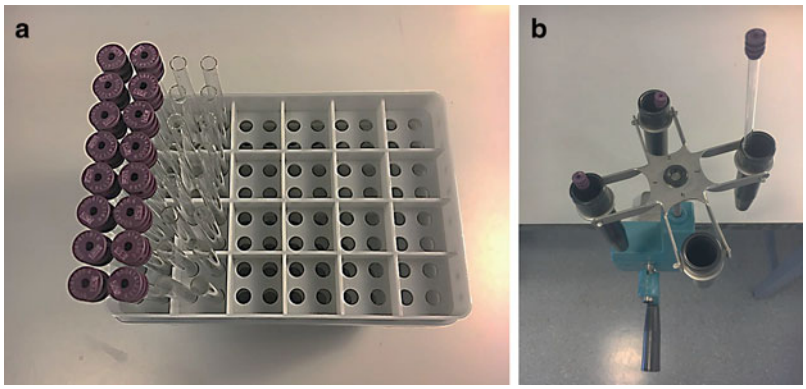


Fig. 5 Practical tips for preparation of large numbers of NMR samples. (**a**) 5-mm NMR tubes can be placed into a pipette tip box; 200-μL tips are typically a suitable size. NMR samples can then be loaded readily using a multichannel pipette with gel-loading tips. (**b**) A hand centrifuge should be used in order to spin the NMR samples after preparation. This improves the homogeneity of the sample depth and thus the shimming under automation. Short NMR tubes can simply be placed in the rotor buckets as shown, but care must be taken to avoid breakages if several long NMR tubes are spun simultaneously

the buffer. Since DSS has nine equivalent protons, the integrated peak at 0.0 ppm should be calibrated to 450 μM. Compound resonances can then be compared directly to this calibrated integral. The measured concentration is approximate (since a relatively short relaxation delay is used), some variation will be observed depending on the relaxation rate of the observed resonance), and care should be taken when integrating strongly coupled peaks when using excitation sculpting. However, the method is simple, fast, readily applicable, and sufficient to identify compounds that show limited solubility under the experimental conditions.

4. There are a number of specific requirements for a buffer suitable for ^{1}H-observed LO-NMR FBS. Most importantly, the protein should be folded, stable, and active in the chosen buffer. However, if possible, high concentrations of additives such as glycerol or detergent should be avoided since these will dominate the NMR spectra and hinder observation of the ligand resonances. A high buffering capacity is advised since the samples will often contain high concentrations of fragments, which can perturb the final pH of the sample; as such, it is advised to use relatively high concentrations (50 mM or higher) of a buffer, which has a p*K*a close to the required pH. Where possible, phosphate buffer is preferred since it does not contain any observable resonances, although it is not compatible with some proteins and can cause issues at high molarity in some cryoprobes. High salt concentrations (typically greater than 250 mM) are also be an issue with cryoprobes, although this can be ameliorated by using 3-mm

sample tubes. If the protein absolutely requires the presence of glycerol or detergent, it is advised to consider screening using ^{19}F LO-NMR or ^{15}N–^{1}H HSQC PO-NMR rather than ^{1}H LO-NMR.

5. Typical initial screening conditions are 500 μM fragment and 10 μM protein in a suitable buffer containing 90% H_2O/10% D_2O. This molar ratio typically gives an observable signal in STD, water-LOGSY, and relaxation-filtered 1D experiments. If relatively hydrophobic fragments are to be screened, a lower concentration may be required owing to solubility limits. Protein concentration should be increased in situations where lower affinity interactions are expected. 10% D_2O is used to ensure reliable locking under automation when mixtures of fragments are used since this can result in high concentrations of d6-DMSO present in the sample that could otherwise confuse the lock system.
6. The most widely used experiment for PO-NMR is the ^{15}N–^{1}H HSQC spectrum, which is rapid, sensitive and requires only inexpensive ^{15}N labeling of the protein. Numerous variants of the ^{15}N–^{1}H HSQC exist; the two most applicable for FBS validation are the ^{15}N–^{1}H SOFAST-HMQC (which is rapid but has a narrow excitation range than the HSQC) and the ^{15}N–^{1}H TROSY (which is relatively slow and insensitive but allows spectra to be acquired on significantly larger proteins than the HSQC experiment). The ^{13}C–^{1}H HSQC or HMQC spectra are also a robust option for validation of putative ligands since, in general, ^{13}C chemical shift perturbations are less prone to long-range effects (and so are more localized around the ligand-binding site) than is the case for ^{15}N. However, this typically requires relatively expensive ^{13}C labeling of the protein, which can apply practical limitations on the number of samples that can be screened using this approach. It is also worth considering the use of simple ^{1}H spectra—either 1D or 2D homonuclear TOCSY spectra. In favorable circumstances (notably proteins of less than 20 kDa where good chemical shift dispersion is observed and the inherent linewidth of the protein is not excessive), these can be highly informative and provide robust validation of a putative ligand without the requirement for specific isotopic labeling.

References

1. Erlanson DA (2012) Introduction to fragment-based drug discovery. Top Curr Chem 317:1–32. https://doi.org/10.1007/128_2011_180
2. Erlanson DA, Fesik SW, Hubbard RE, Jahnke W, Jhoti H (2016) Twenty years on: the impact of fragments on drug discovery. Nat Rev Drug Discov 15(9):605–619. https://doi.org/10.1038/nrd.2016.109
3. Fischer M, Hubbard RE (2009) Fragment-based ligand discovery. Mol Interv 9 (1):22–30. https://doi.org/10.1124/mi.9.1.7

4. Schulz MN, Hubbard RE (2009) Recent progress in fragment-based lead discovery. Curr Opin Pharmacol 9(5):615–621. https://doi.org/10.1016/j.coph.2009.04.009
5. Hubbard RE, Murray JB (2011) Experiences in fragment-based lead discovery. Methods Enzymol 493:509–531. https://doi.org/10.1016/B978-0-12-381274-2.00020-0
6. Davis BJ, Roughley SD (2017) Fragment-based lead discovery. In: Goodnow RA (ed) Annual reports in medicinal chemistry, vol 50. Academic Press, Cambridge, pp 371–439. https://doi.org/10.1016/bs.armc.2017.07.002
7. Chessari G, Woodhead AJ (2009) From fragment to clinical candidate--a historical perspective. Drug Discov Today 14(13–14):668–675. https://doi.org/10.1016/j.drudis.2009.04.007
8. Howard SP, Abell C (eds) (2015) Fragment-based drug discovery. RSC Publishing, London
9. Erlanson DA, Jahnke W (eds) (2015) Fragment-based drug discovery : lessons and outlook. Wiley-VCH, Weinheim
10. Czechtizky W, Hamley P (eds) (2015) Small molecule medicinal chemistry : strategies and technologies. John Wiley, New Jersey
11. Magee TV (2015) Progress in discovery of small-molecule modulators of protein-protein interactions via fragment screening. Bioorg Med Chem Lett 25(12):2461–2468. https://doi.org/10.1016/j.bmcl.2015.04.089
12. Bollag G, Tsai J, Zhang J, Zhang C, Ibrahim P, Nolop K, Hirth P (2012) Vemurafenib: the first drug approved for BRAF-mutant cancer. Nat Rev Drug Discov 11(11):873–886. https://doi.org/10.1038/nrd3847
13. Souers AJ, Leverson JD, Boghaert ER, Ackler SL, Catron ND, Chen J, Dayton BD, Ding H, Enschede SH, Fairbrother WJ, Huang DC, Hymowitz SG, Jin S, Khaw SL, Kovar PJ, Lam LT, Lee J, Maecker HL, Marsh KC, Mason KD, Mitten MJ, Nimmer PM, Oleksijew A, Park CH, Park CM, Phillips DC, Roberts AW, Sampath D, Seymour JF, Smith ML, Sullivan GM, Tahir SK, Tse C, Wendt MD, Xiao Y, Xue JC, Zhang H, Humerickhouse RA, Rosenberg SH, Elmore SW (2013) ABT-199, a potent and selective BCL-2 inhibitor, achieves antitumor activity while sparing platelets. Nat Med 19(2):202–208. https://doi.org/10.1038/nm.3048
14. Erlanson DA (2018) Fragments in the clinic: 2018th edition. Practical Fragments. http://practicalfragments.blogspot.com/2018/10/fragments-in-clinic-2018-edition.html
15. Davis BJ, Erlanson DA (2013) Learning from our mistakes: the "unknown knowns" in fragment screening. Bioorg Med Chem Lett 23(10):2844–2852. https://doi.org/10.1016/j.bmcl.2013.03.028
16. Davis B (2013) Screening protein-small molecule interactions by NMR. Methods Mol Biol 1008:389–413. https://doi.org/10.1007/978-1-62703-398-5_14
17. Erlanson D (2016) Poll results: affiliation, metrics, and fragment-finding methods. Practical Fragments. http://practicalfragments.blogspot.com/2016/10/poll-results-affiliation-metrics-and.html
18. Aguirre C, Cala O, Krimm I (2015) Overview of probing protein-ligand interactions using NMR. Curr Protoc Protein Sci 81:17.18.11–17.18.24. https://doi.org/10.1002/0471140864.ps1718s81
19. Fejzo J, Lepre C, Xie X (2003) Application of NMR screening in drug discovery. Curr Top Med Chem 3(1):81–97
20. Hajduk PJ, Meadows RP, Fesik SW (1999) NMR-based screening in drug discovery. Q Rev Biophys 32(3):211–240
21. Jordan JB, Poppe L, Xia X, Cheng AC, Sun Y, Michelsen K, Eastwood H, Schnier PD, Nixey T, Zhong W (2012) Fragment based drug discovery: practical implementation based on ^{19}F NMR spectroscopy. J Med Chem 55(2):678–687. https://doi.org/10.1021/jm201441k
22. Shuker SB, Hajduk PJ, Meadows RP, Fesik SW (1996) Discovering high-affinity ligands for proteins: SAR by NMR. Science 274(5292):1531–1534
23. Chimenti MS, Bulfer SL, Neitz RJ, Renslo AR, Jacobson MP, James TL, Arkin MR, Kelly MJ (2015) A fragment-based ligand screen against part of a large protein machine: the ND1 domains of the AAA+ ATPase p97/VCP. J Biomol Screen 20(6):788–800. https://doi.org/10.1177/1087057115570550
24. Wiesner S, Sprangers R (2015) Methyl groups as NMR probes for biomolecular interactions. Curr Opin Struct Biol 35:60–67. https://doi.org/10.1016/j.sbi.2015.08.010
25. Mayer M, Meyer B (1999) Characterization of ligand binding by saturation transfer difference NMR spectroscopy. Angew Chem Int Ed 38(12):1784–1788. https://doi.org/10.1002/(SICI)1521-3773(19990614)38:12<1784::AID-ANIE1784>3.0.CO;2-Q
26. Dalvit C, Fogliatto G, Stewart A, Veronesi M, Stockman B (2001) WaterLOGSY as a method for primary NMR screening: practical aspects

and range of applicability. J Biomol NMR 21 (4):349–359

27. Hajduk PJ, Meadows RP, Fesik SW (1997) Discovering high-affinity ligands for proteins. Science 278(5337):497–499
28. Becker W, Bhattiprolu KC, Gubensäk N, Zangger K (2018) Investigating protein-ligand interactions by solution nuclear magnetic resonance spectroscopy. Chemphyschem 19 (8):895–906. https://doi.org/10.1002/cphc.201701253
29. Meyer B, Peters T (2003) NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors. Angew Chem Int Ed 42(8):864–890. https://doi.org/10.1002/anie.200390233
30. Dalvit C, Fasolini M, Flocco M, Knapp S, Pevarello P, Veronesi M (2002) NMR-based screening with competition water–ligand observed via gradient spectroscopy experiments: detection of high-affinity ligands. J Med Chem 45(12):2610–2614. https://doi.org/10.1021/jm011122k
31. Jordan JB, Whittington DA, Bartberger MD, Sickmier EA, Chen K, Cheng Y, Judd T (2016) Fragment-linking approach using (19)F NMR spectroscopy to obtain highly potent and selective inhibitors of β-secretase. J Med Chem 59 (8):3732–3749. https://doi.org/10.1021/acs.jmedchem.5b01917
32. Arntson KE, Pomerantz WCK (2016) Protein-observed fluorine NMR: a bioorthogonal approach for small molecule discovery. J Med Chem 59(11):5158–5171. https://doi.org/10.1021/acs.jmedchem.5b01447
33. Renaud J-P, Chung C-w, Danielson UH, Egner U, Hennig M, Hubbard RE, Nar H (2016) Biophysics in drug discovery: impact, challenges and opportunities. Nat Rev Drug Discov 15:679. https://doi.org/10.1038/nrd.2016.123
34. Baurin N, Aboul-Ela F, Barril X, Davis B, Drysdale M, Dymock B, Finch H, Fromont C, Richardson C, Simmonite H, Hubbard RE (2004) Design and characterization of libraries of molecular fragments for use in NMR screening against protein targets. J Chem Inf Comput Sci 44(6):2157–2166. https://doi.org/10.1021/ci049806z
35. Chen IJ, Hubbard RE (2009) Lessons for fragment library design: analysis of output from multiple screening campaigns. J Comput Aided Mol Des 23(8):603–620. https://doi.org/10.1007/s10822-009-9280-5
36. Ray PC, Kiczun M, Huggett M, Lim A, Prati F, Gilbert IH, Wyatt PG (2017) Fragment library design, synthesis and expansion: nurturing a synthesis and training platform. Drug Discov Today 22(1):43–56. https://doi.org/10.1016/j.drudis.2016.10.005
37. Baell JB, Holloway GA (2010) New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. J Med Chem 53(7):2719–2740. https://doi.org/10.1021/jm901137j
38. Erlanson DA (2018) Poll results: library vendors. http://practicalfragments.blogspot.com/2018/12/poll-results-library-vendors.html
39. Stott K, Stonehouse J, Keeler J, Hwang T-L, Shaka AJ (1995) Excitation sculpting in high-resolution nuclear magnetic resonance spectroscopy: application to selective NOE experiments. J Am Chem Soc 117(14):4199–4200. https://doi.org/10.1021/ja00119a048
40. Henzel WJ, Watanabe C, Stults JT (2003) Protein identification: the origins of peptide mass fingerprinting. J Am Soc Mass Spectrom 14 (9):931–942. https://doi.org/10.1016/S1044-0305(03)00214-9
41. El-Aneed A, Cohen A, Banoub J (2009) Mass spectrometry, review of the basics: electrospray, MALDI, and commonly used mass analyzers. Appl Spectrosc Rev 44(3):210–230. https://doi.org/10.1080/05704920902717872
42. Davis BJ, Giannetti AM (2016) The synthesis of biophysical methods in support of robust fragment-based Lead discovery. In: Fragment-based drug discovery lessons and outlook. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, pp 119–138. https://doi.org/10.1002/9783527683604.ch06
43. Erlanson DA, Davis BJ, Jahnke W (2018) Fragment-based drug discovery: advancing fragments in the absence of crystal structures. Cell Chem Biol 26:9–15. https://doi.org/10.1016/j.chembiol.2018.10.001
44. Brough PA, Barril X, Borgognoni J, Chene P, Davies NG, Davis B, Drysdale MJ, Dymock B, Eccles SA, Garcia-Echeverria C, Fromont C, Hayes A, Hubbard RE, Jordan AM, Jensen MR, Massey A, Merrett A, Padfield A, Parsons R, Radimerski T, Raynaud FI, Robertson A, Roughley SD, Schoepfer J, Simmonite H, Sharp SY, Surgenor A, Valenti M, Walls S, Webb P, Wood M, Workman P, Wright L (2009) Combining hit identification strategies: fragment-based and in silico approaches to orally active 2-aminothieno[2,3-d]pyrimidine inhibitors of the Hsp90 molecular chaperone. J Med Chem 52 (15):4794–4809. https://doi.org/10.1021/jm900357y

Part IV

Nucleotide Binding and Hydrolysis

Chapter 12

A Quick Primer in Fluorescence-Based Equilibrium and Pre-steady State Methods for Determining Protein–Nucleotide Affinities

Harland E. Brandon and Hans-Joachim Wieden

Abstract

Biomolecular interactions facilitate the biochemical processes that sustain life. Proteins, RNAs, and ribonucleoprotein complexes perform cellular functions that range from catalyzing the formation or cleavage of bonds to being structural building blocks, both of which are only possible through the interaction with their respective biomolecular partner(s). Having access to the parameters that describe these interactions is important for our understanding of the principles that underlie enzymatic and nonenzymatic processes. Here we describe two fluorescence-based approaches to determine two key parameters, the affinity and the rate of association/dissociation of a protein and a ligand. Considerations are provided to expand the described approach to other experimental systems.

Key words Affinity, Fluorescence, Stopped-flow, Equilibrium binding, Pre-steady state, Nucleotide, GTPase, FRET

1 Introduction

Biological systems are composed of many biomolecules that are required to interact with each other to drive the chemical processes that facilitate life as we know it. Understanding which and how biomolecules interact can provide insight into the mechanism that are at the core of biochemical processes and how they form the molecular pathways that underlie cellular life. Proteins make up a large percentage of the total dry weight of a cell, with 55% of a single *Escherichia coli* cell being protein [1]. Proteins are constantly interacting with various ligands in the cell, some that they act upon, and others that they only transiently interact with. Knowing the affinity of a protein to a particular ligand informs our understanding of how the protein functions mechanistically and how it integrates into the complex reaction pathways of the cell. Nucleotides, the ligand of interest in the following examples, are the building blocks of nucleic acids, along with being important secondary

Tina Daviter et al. (eds.), *Protein-Ligand Interactions: Methods and Applications*, Methods in Molecular Biology, vol. 2263, https://doi.org/10.1007/978-1-0716-1197-5_12,

messengers and energy sources for biochemical process. Many nucleotide-binding proteins serve essential functional roles including translation, cellular import and export, cell signaling, biosynthesis and catabolism, cell motility, to name only a few. While there are enzymes that bind all nucleotides present in the cell, the majority of enzymes bind and utilize either adenosine triphosphate (ATP) or guanosine triphosphate (GTP).

Techniques to measure the interaction between a protein and its ligands are becoming ever more important. Here we provide examples to measure the kinetic parameters for enzymes and nucleotides; these techniques are highly versatile and can also be adapted for use with ligands that are not nucleotides. Besides their critical role in drug development, these techniques are also being used, for example, by the biotechnology industry as they continue to develop new protein-based products for various applications [2–4]. Several experimental approaches exist that can be used to determine the affinities, including isothermal calorimetry (ITC), nuclear magnetic resonance (NMR), surface plasmon resonance (SPR), and radiography-based techniques. Historically, fluorescence-based techniques, including the methods described here, have been used widely due to their great versatility with respect to the samples, sensitivity, and data obtainable. Determining which technique should be used depends on the system being examined and should be well thought out before proceeding.

Here we describe two examples for measuring the dissociation constant (K_d) for a protein and a ligand via fluorescence-based techniques, both using the intrinsic fluorescent properties of the protein and/or a fluorescently labeled ligand or protein. Primarily this chapter will focus on assays that use two fluorophores for Förster/fluorescence resonance energy transfer (FRET) but can be adapted for single fluorophore studies as well. FRET is the nonradiative transfer of energy between an excited donor fluorophore and acceptor fluorophore in a proximity-dependent manner. As such, FRET is a distance-dependent measurement between the donor and acceptor fluorophores that increases the signal-to-noise ratio by separating the excitation and emission wavelengths. Conversely, FRET does require two fluorophores (or a fluorophore and quenching dye) that may alter the dynamics of the biomolecule they are attached to and/or not all biomolecules are amendable for being labeled with a fluorophore. Natural fluorophores (such as tryptophan and tyrosine in proteins) can be used in conjunction with synthetic fluorophores to create a FRET pair where only the ligand needs to be labeled. Fluorescently modified analogs of certain ligands can be purchased, or produced, such as Mant-labeled ((2′-(or-3′)-*O*-(*N*-Methylanthraniloyl) nucleotides used in this chapter to measure FRET between intrinsic tryptophan residues in a protein (here HflX) and the Mant-fluorophore on the nucleotide (here Mant-GDPNP) for K_d measurements [5, 6].

While we will present examples of nucleotide binding in this protocol, other ligands can be measured by adjusting the protocol accordingly, including larger ligands such as proteins or RNAs that may interact with the protein of interest. It is to be noted that additional factors may make data analysis difficult or impossible to deconvolute and as such proper controls are required to ensure the signal change observed reports on the appropriate reaction. In the following protocols, the tryptophan and tyrosine residues of HflX are excited at 280 nm and serve as donor dye(s) that can transfer energy to the Mant group located on the nucleotide (acceptor dye) via FRET when bound to the protein.

2 Materials

All buffers prepared in house were made using ultrapure water (Millipore) and filtered through 0.45 μm pore size Whatman™ mixed cellulose ester membrane filters (GE Healthcare Life Sciences).

2.1 General

1. HflX (or other protein of interest): purified as described in Shields et al. [5].
2. Guanosine triphosphate (GTP; or other ligand of interest): 30 mM stock (Sigma Aldrich Cat. # G8877; bought as a solid).
3. Mant-GTP (or other ligand of interest): 10 mM stock (Jena Biosciences).

2.2 Equilibrium Nucleotide Binding

1. Guanosine triphosphate (GTP; nonfluorescently labeled nucleotide): 30 mM stock. (Sigma Aldrich Cat. # G8877; bought as a solid).
2. Reaction buffer: 50 mM Tris–HCl pH 7.5 4 °C; 70 mM KCl; 30 mM NH_4Cl; 7 mM $MgCl_2$; filtered and degassed prior to use (*see* **Note 1**).
3. Cuvette (Starna Cells Rectangular Quartz Fluorometer Micro Cell with 3 mm Open Top; Cat. No: 3-3.45-Q-3).
4. Cuvette adaptor (StrataCells FCA3).
5. Long gel loading pipet tips—helpful for efficient mixing of solution after each ligand addition.
6. Ultrapure water—filtered and degassed prior to use.
7. Acetone (Certified ACS; 99% purity).

2.3 Pre-steady State Nucleotide Binding

1. Mant-GTP (Fluorescently labeled ligand).
2. GTP (nonfluorescently labeled ligand required for dissociation chase experiments described in [5, 6]).

3. Reaction buffer: 50 mM Tris–HCl pH 7.5 4 °C; 70 mM KCl; 30 mM NH_4Cl; 7 mM $MgCl_2$; filtered and degassed prior to use (*see* **Note 1**).
4. Ultrapure water—filtered and degassed prior to use.
5. 30% ethanol—filtered and degassed prior to use.
6. 3 mL syringes (BD Luer-Lok™ Tip 3 mL syringe).
7. Disposable needles (BD PrecisionGlide™ needle; 20G × 1½ in.).

3 Methods

The method of choice depends on the experimental question to be addressed and the equipment available. Subheading 3.1, equilibrium nucleotide binding, is a technique that allows the determination of the affinity (K_d) between a protein and ligand and can be performed using a standard fluorometer. The pre-steady state approach described in Subheading 3.2 allows to determine the rate constants describing the association (k_{on}) and dissociation (k_{off}) reactions for a particular protein–ligand interaction pair (or k_1 and k_{-1} in Eq. 1), which can ultimately be used to calculate the respective K_d for this interaction (Eq. 2). As a binding event involves two molecules interacting with each other, the rate of association (v_{on}) is dependent on the concentration of both molecules (Eq. 3). This becomes particularly important when fitting pre-steady state data with two or more observed rates, for only the rate that is concentration dependent (the rate increases with an increasing concentration of either protein or ligand) is indicative of a second-order binding event.

$$\text{Protein} + \text{Ligand} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \text{Protein} \cdot \text{Ligand} \tag{1}$$

$$K_d = k_{-1}/k_1 \tag{2}$$

$$v_{on} = k_1 \times [\text{Protein}] \times [\text{Ligand}] \tag{3}$$

Additionally, pre-steady state approaches allow the dissection of multiple-step binding events, such as a nucleotide binding step followed by a subsequent conformational change in the protein upon nucleotide binding [5, 7]. For reactions that have additional steps, the rate constants for each step are required to determine the K_d. For example, in Eq. (4) there are four rate constants (k_1, k_{-1}, k_2, k_{-2}) that are required to determine the K_d using Eq. (5). An altered conformation of the protein of interest is denoted as Protein* in Eq. (4). For more information on determining the K_d for reactions with more than two rate constants see Structure and

Mechanism in Protein Science by Alan Fersht [8], or other similar enzyme kinetics texts.

$$\text{Protein} + \text{Ligand} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \text{Protein} \cdot \text{Ligand} \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} \text{Protein}^* \cdot \text{Ligand} \quad (4)$$

$$K_d = (k_{-1} \times k_{-2})/(k_1 \times k_2) \quad (5)$$

3.1 Equilibrium Nucleotide Binding

3.1.1 Experimental Procedure

1. Start up the fluorimeter, in particular the light source (e.g., xenon lamp), 15 min before the first experiment (*see* **Note 2**).
2. Adjust the excitation and emission slit widths to allow for an optimal spectral bandwidth and intensity (*see* **Note 3**).
3. Set all the parameters in the software for doing an emission scan. Excitation at 280 nm, measure emission from 305–450 nm, and, for example, 1 s acquisition time and 1 nm step size (*see* **Note 4**).
4. Clean the cuvette with water followed by acetone. Use compressed air to (carefully) dry the acetone from the cuvette quickly.
5. Determine the volume of buffer to make a 200 μL protein solution (here containing 1 μM HflX) and add that volume to the cuvette (*see* **Note 5**).
6. Perform an initial emission scan of the buffer alone.
7. Add HflX to reaction buffer by pipetting it along the corner of the cuvette. Use long gel loading pipet tips to wash protein down into the cuvette and mix.
8. Initiate another emission scan.
9. Add a volume of nucleotide (or ligand of choice) containing solution to the cuvette in the same manner as the protein addition in **step 7** (*see* **Note 7**).
10. Equilibrate the reaction solution at room temperature for 1 min (adjust this step accordingly if measurements are made at temperatures different from RT) before performing the emission scan. The reaction is scanned after 1 min because the binding reaction (in most cases) will have reached equilibrium.
11. Repeat **steps 9** and **10** for each addition of nucleotide (ligand).

3.1.2 Data Analysis

1. Correct the fluorescence intensity for dilution of the protein sample and subtract the background fluorescence (**step 6**) from the emission scan at each nucleotide concentration (*see* **Notes 8** and **9**).
2. Plot all corrected emission scans as shown (*see* Fig. 1a).

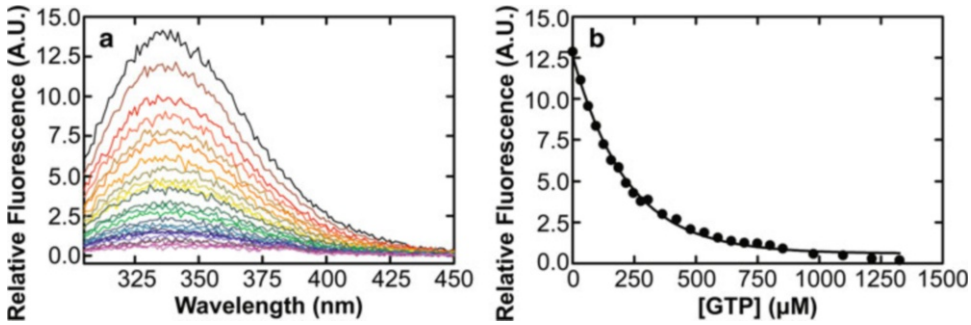


Fig. 1 (**a**) Equilibrium fluorescence titration of 1 μM HflX with increasing concentrations of GTP. Tryptophan and tyrosine residues were excited at 280 nm, and the fluorescence emission was measured from 305 to 450 nm. (**b**) The fluorescence signal at the emission maximum (338 nm in this example) was plotted against the nucleotide concentration. The resulting curve is fit with a one-site binding equation (hyperbolic function) to determine the affinity

3. Determine the fluorescence maximum for the protein, and plot the fluorescence intensity at the maxima against the nucleotide concentration as shown (*see* Fig. 1b), (*see* **Note 10**).
4. Fit the data with the appropriate function to determine the binding parameters, including K_d, for the proposed kinetic mechanism. Here we used a simple hyperbolic binding equation (Eq. 6) that was fit to the data using GraphPad Prism (GraphPad Software). Where RF is the relative fluorescence, F is the initial fluorescence, F_{max} is the fluorescence when it plateaus, and [nt] is the concentration of nucleotide.

$$\mathrm{RF} = F + (F_{max} \times [\mathrm{nt}])/(K_d + [\mathrm{nt}]) \quad (6)$$

For data that deviates from the simple binding equation, refer to more in-depth kinetic analysis detailed in Structure and Mechanism in Protein Science by Alan Fersht [8] to determine the K_d. Furthermore, data that does not fit a simple binding equation can provide insight into the mechanism of ligand binding.

Additionally, if the free ligand is being significantly depleted during the measurement, thus decreasing the concentration of ligand, then the data should be fit with a quadratic binding equation (Eq. 7). Where ΔFl is the change in fluorescence from the initial value, [P] is the total concentration of protein (HflX), and B is the signal amplitude ($B = \mathrm{Fl}_{max} - \mathrm{Fl}_{initial}$).

$$\Delta\mathrm{Fl} = 0.5 \times (B/[\mathrm{P}]) \times \left(K_D + [\mathrm{P}] + [\mathrm{nt}] - \left((K_D + [\mathrm{P}] + [\mathrm{nt}])^2 - 4 \times [\mathrm{P}] \times [\mathrm{nt}]\right)^{\frac{1}{2}}\right) \quad (7)$$

3.2 Pre-steady State Nucleotide Binding

3.2.1 Experimental Design

1. Reactions are set up as depicted in Fig. 2 and listed in Table 1.

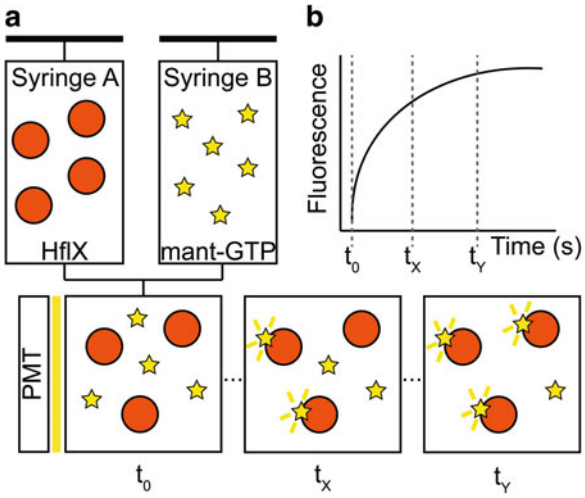


Fig. 2 Graphical depiction of the stopped-flow apparatus experimental setup. (**a**) Two syringes are filled with either reaction component, in this case the protein HflX in syringe A and Mant-labeled nucleotide (GTP) in syringe B. Over time, Mant-GTP binds to HflX allowing FRET to occur, and the overall fluorescence signal to increase. (**b**) The resulting time course is fit with an exponential function

Table 1
Reaction design for a Mant-nucleotide binding assay. Values are those used for the data presented in Fig. 3

Syringe A	Reaction Concentration (μM)[a]	Volume (μL)	Syringe B	Reaction Concentration (μM)[a]	Volume (μL)
Protein (HflX) [HflX] = 100 μM	2	*X*	Ligand (Mant-GDPNP)	(10–50 μM)[b]	*X*
Other[c]			Other[c]		
Buffer			Buffer		
Total[d]		300 μL	Total[d]		300 μL

[a]Reaction concentration is double the desired final concentration upon mixing of the two mixtures in the stopped-flow. It should be noted that some stopped-flows allow more than two syringes to be mixed sequentially and, as such, one would need to adjust the reaction concentrations accordingly for the degree of dilution upon mixing (*see* **Note 14**)
[b]This experiment is repeated for several ligand concentrations as binding reactions are concentration dependent
[c]Depending on the protein/ligand, other factors can be included into the reaction (*see* **Note 20**)
[d]Reaction volume can be altered depending on the number of measurements for each experiment desired and is also influenced by the volume of the cuvette

*3.2.2 Experimental Procedure (See **Note 11**)*

1. Ensure the desired long-pass wavelength filters to remove excitation light are inserted in front of the detectors (*see* **Note 12**).
2. Turn on the xenon lamp to warm up for 15 min before performing the first experiment (*see* **Note 13**).
3. Drain the 30% ethanol solution from stopped-flow syringes.
4. Wash syringes, mixer, and cuvette with water three times followed by washing with reaction buffer three times each.
5. Set the excitation monochromator to the desired excitation wavelength and confirm optimal slit width.
6. Prepare syringe A and B reaction mixtures separately (*see* **Note 14**).
7. Incubate reaction mixtures at 37 °C for 15 min in a water bath (*see* **Note 15**).
8. Centrifuge reactions at 17,000 × *g* for 2 min to remove any precipitate or particles before loading into the stopped-flow (*see* **Note 16**).
9. Draw reaction mixture A up into a 3 mL disposable syringe using a needle (avoid frothing).
10. Remove needle, insert the syringe into the respective port, and press slowly the reaction mixture A into stopped-flow syringe A (avoid capturing an air bubble here). Make sure the stopped-flow is in the load position before injecting reaction mixture (*see* **Note 17**).
11. Repeat **steps 7** and **8** for reaction mixture B.
12. Adjust the syringe drive so that the plunger of each stopped-flow syringe is in contact with the motor drive.
13. Switch the stopped-flow into the fire position.
14. Set the duration and number of measurements to be carried out in the software.
15. Collect data and adjust duration of measurement as needed (*see* **Note 18**). Data should appear similar to that in Fig. 3a.

3.2.3 Data Analysis

1. Fit each recorded time-dependent signal change (referred to as a fluorescence trace) with an exponential function. Selecting the correct equation to fit the data is critical and depends on the reaction components and proposed kinetic mechanism. For a simple one-step binding mechanism (Eq. 1), a one-exponential equation (Eq. 8) is fit to the data using TableCurve 2D (Systat Software) to determine the apparent rate (k_{app1}). Here we used a two-exponential equation (Eq. 9) after initial one-exponential fits were not describing the data properly due to an additional kinetic step detected in our experimental system. For more details on evaluating which equation should be used to fit the data, *see* **Note 19**. In both exponential

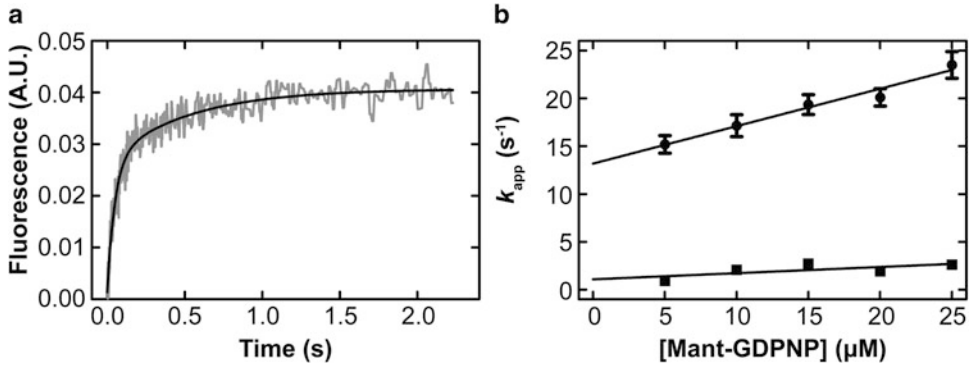


Fig. 3 Fluorescently labeled nucleotide (Mant-GDPNP) association to HflX carried out using a stopped-flow apparatus. Binding of the Mant-GDPNP is observed as a high FRET signal. Tryptophan and tyrosine residues in HflX are excited at 280 nm and FRET occurs between these residues and the Mant group covalently attached to the nucleotide. (**a**) Characteristic Mant-GDPNP association time course (grey line) to HflX fit with a two-exponential function (black line). (**b**) GDPNP concentration dependence on the exact rate constants from association. One rate is concentration-dependent (squares) and one is concentration-independent (circles). The concentration-dependent rate is the rate of association for that step is a second-order reaction dependent on the concentration of either reactant

equations (Eqs. 8 and 9), *Fl* is the relative fluorescence at time t, k_{app} is the apparent rate constant, A_1 is the signal amplitude, and Fl_∞ is the final fluorescent signal. The two-exponential equation has an additional exponential term to account for the second apparent rate and the corresponding amplitude change (denoted as A_2 and k_{app2}).

$$Fl = Fl_\infty + A_1 \ \exp\left(-k_{app1} t\right) \tag{8}$$

$$Fl = Fl_\infty + A_1 \ \exp\left(-k_{app1} t\right) + A_2 \ \exp\left(-k_{app2} t\right) \tag{9}$$

2. Average all traces that have similar values for each variable from the individual fits. Aim to average as few traces as possible (up to 10 is a typical number for a smaller signal change).
3. Repeat experiment for additional ligand concentrations to ensure the rate is concentration-dependent, as a binding reaction should be (*see* Fig. 3b and *see* **Note 21**).
4. Rate constants obtained from the fit of the respective signal changes can be used to calculate the K_d describing the protein–ligand interaction (*see* Subheading 3).

4 Notes

1. The techniques described here are amendable to various buffer solutions. It is recommended that buffer alone controls are performed to ensure nothing in the buffer is causing a

fluorescence change over the course of the experiment. Furthermore, buffers should not be stored in plastic containers as leaching of fluorescent softeners can occur.

2. Fluorescence is detected at 90° from the incident light source.
3. The excitation and emission slit widths should be set so as to not overexpose your sample and the photomultiplier tubes (PMTs) to light, respectively. Proteins and ligands can become photobleached over the course of an experiment, reducing the amount of emitted light over time, thus it is advised that the amount of incident light is decreased to a level that provides a measurable output (*see* **Note 6**). Additionally, the larger the excitation slit widths are, the larger the wavelength range around the excitation wavelength is that is allowed to pass through and excite the sample. Thus, depending on the range of excitation wavelengths, there may be other fluorophores excited that could contribute to the observed fluorescence output. An example of this is the excitation of tryptophan and tyrosine in proteins, both of which can be excited at 280 nm, but tryptophan primarily excited at 295 nm. This can be beneficial as it allows selective excitation of tryptophan residue (opposed to also exciting the more abundant tyrosine) in the protein of interest (if the protein only has one tryptophan, is a variant with only one tryptophan, or only has one tryptophan within the proximity of the ligand binding site) and the measurement of the distance between a specific donor tryptophan residue and acceptor fluorophore based on the FRET efficiency. This is a specific example, but the logic can be applied to other FRET pair systems to measure distances on a nanoscale.
4. The excitation wavelength is dependent on the donor fluorophore used in the experiment. Here we used FRET between tryptophan/tyrosine residues in our protein of interest (HflX) excited at 280 nm and measure the fluorescence emission of the Mant group covalently attached to the bound nucleotide. Other donor fluorophores can be used and the excitation wavelength adjusted accordingly. Additionally, the emission scan should not overlap the excitation wavelength (e.g., exciting at 420 nm and measuring emission from 350 to 500 nm) and emission is generally measured starting 10–15 nm from the excitation wavelength (also *see* **Note 3** for slit width considerations). This is to prevent overexposing the PMT, causing damage to the instrument.
5. The protein concentration used in the experiment is dependent on the intensity of the fluorescence signal but will determine the lowest possible K_d that can be measured. As a general rule, this lower boundary is equal to the concentration of protein used in the experiment as this ensures that the measured

fraction-bound will be at 50%. *See* **Note 6** for rationale behind the intensity of the fluorescence signal.

6. A measurable output is such that the baseline fluctuation in signal is not equal to or greater than the increase/decrease in fluorescent signal from the experiment. This is valid for both equilibrium and pre-steady state experiments.
7. The concentrations of nucleotide or ligand in general to add over the course of the titration should be based on an approximate K_d if known, otherwise should span several orders of magnitude. Additions do not need to be of the same volume at each point as long as the volume added is noted (*see* **Note 8**). The number of additions (data points) one should perform for each titration curve is dependent on the quality of the data such that a defined curve can be fit with Eq. (6) (*see* Subheading 3.1.2, **step 15**). Typically, 15–20 ligand concentrations that are distributed equally above and below the anticipated K_d will provide an accurate initial measurement of the K_d.
8. A dilution of the protein occurs with each addition of ligand which will result in a decrease in fluorescent signal as a function of protein concentration. When analyzing the data this must be taken into account. This is done by calculating the adjusted concentration of the protein after each addition and normalizing to protein concentration (*see* Eq. 10). The measured fluorescence intensity is multiplied by the correction factor (CF). V_I is the initial volume of the reaction and V_D is the diluted volume. The CF will need to be calculated for each dilution performed and measured.

$$CF = V_D / V_I \tag{10}$$

9. The background fluorescence should be subtracted from each subsequent measurement to correct for any trace fluorescence from the buffer alone. This is typically carried out after correcting for the dilution change with each addition of ligand (*see* **Note 8**).
10. The fluorescence maximum depends on the physical properties of the protein being studied. Therefore, the wavelength chosen to be plotted should be consistently the highest from several trials.
11. Many steps in this protocol are specific to the KinTek SF-2004 stopped-flow apparatus and appropriate changes should be considered when using other stopped-flow devices. Consult the user manual or manufacturing company for specific operational procedures to the apparatus.
12. Long-pass wavelength filters allow only light of wavelengths greater than the indicated wavelength to pass through. Light of

shorter wavelengths is attenuated and does not enter the detector (photomultiplier tube). This is beneficial for FRET experiments whereby one excites at a lower wavelength and wants to record the emitted light from the acceptor fluorophore. Therefore, having a long-pass filter separating the donor excitation and acceptor emission wavelengths will prevent any scattered light of the incident wavelength or FRET donor emission from contributing to the fluorescence signal detected by the instrument. Other filters can be used such as band-pass filters to detect only a specific range of light depending on experimental conditions. For a good example of this *see* Gzyl and Wieden [9].

13. The xenon arc lamp needs to reach operating temperature prior to the first measurement to achieve a stable signal (typically 15 min prior to the first measurement).
14. Each syringe holds one reaction solution, generally keeping the two molecules of interest separate before mixing. Upon mixing, the solutions dilute each other by half if both syringes are the identical volume, thus the final concentration of each reactant in the cuvette will be half that of the individual reaction solutions before mixing. Therefore, it is important to prepare reaction solutions that are double the desired reaction concentration. Some stopped-flow instruments have three or more syringes so that three different reactants can be rapidly or sequentially mixed without having to incubate two with each other prior, and as such, one must be careful with the dilution upon mixing three syringes and the desired concentrations of each reactant in the cuvette. Additionally, the syringes can be of different sizes having different diameters such that upon moving the drive piston an unequal volume of solution in either syringe is mixed resulting in a greater dilution than 1:1 which can be beneficial depending on the experiment. Protein refolding experiments tend to utilize stopped-flow experimental setups with different syringe sizes such as in Visconti et al. [10].
15. The incubation temperature is dependent on the particular experiment and/or molecular system in the experiment and thus can be adjusted to suit the respective experimental conditions.
16. It is important to centrifuge each reaction solution at maximal speed in a tabletop microcentrifuge before loading into the stopped-flow to remove any precipitate. Precipitates can clog the lines of the stopped-flow and/or interfere with subsequent experiments if not cleaned properly. Furthermore, any precipitate that is not removed and part of a mixed sample will scatter light during detection, thus interfering with the measurement.

17. When loading the reaction solutions into the syringes, it is important to avoid introducing bubbles into the syringe. Bubbles can scatter light and/or can compress upon the motor drive firing thus preventing the total reaction volume from being mixed and causing the liquid column after the flow was stopped to move, ultimately creating noise in the measurements. Therefore, large bubbles should be removed by displacing the reaction solution from the stopped-flow syringe back into the disposable syringe trapping the bubbles in the disposable syringe. Additionally, it is important to move the reaction solution between the stopped-flow and disposable syringes several times before starting the experiment to make sure the reactant solutions are well mixed (e.g., after centrifuging to remove precipitates in the previous step).
18. Ideally, the fluorescence signal should plateau upon reaching equilibrium, and thus it is important to measure until a stable signal is reached (plateau). This will improve the fitting step. Furthermore, most of the data points do not need to be in the plateau and if they are, the time recorded for each shot can be reduced so that more of the data points fall within the section of the signal that shows the fastest change. It is recommended to do single shots until the optimal recording time and point distribution is determined for each experiment (for fast reactions it is often beneficial to use exponential data distributions with increasing separation between each recorded point).
19. When fitting an equation to the data set, it is important to check that the resulting fit adequately describes the signal. As general rule, this entails that the experimental data is equally distributed above and below the fit curve. A good fit should have equivalent data points on either side of the line corresponding to the fit and no systematic deviation should be observed (e.g., the first part of the data is below and the later part above the fit). If not, additional exponential (or linear) terms can be added to the used equation. However, additional exponential terms indicate additional kinetic steps (or processes) in the reaction and must be reflected in the kinetic mechanism. It may be necessary to titrate one of the reaction components to determine which of the rates is concentration dependent, i.e., the rate of binding. Fundamentally, a protein–ligand binding event should be a one-step process and fit with a one-exponential equation; however, it is possible that additional kinetic steps can be observed in these pre-steady state measurements as a result of a change in either fluorophore's position and/or local environment (*see* Subheading 3). For example, in Shields et al., we observed a second phase in the nucleotide binding experiments that, through subsequent titration of the Mant-nucleotide, was shown to be

concentration independent and was subsequently attributed to a conformational change within the protein (HflX) following nucleotide binding [5]. As such, pre-steady state kinetics not only allows for the determination of rate constants but also provides insight into the kinetic mechanism and structural dynamics of the protein being studied.

20. Additional reaction components can be added to either reaction solution noting the dilution effect described in **Note 14**.
21. Titration of nucleotide (or any ligand) for determining the rate of association yields a graph of the apparent rate at each concentration of ligand tested. The slope of the line is equal to k_1 and the y-intercept is equal to k_{-1}. The value for k_{-1} from the y-intercept can be used to validate the k_{-1} value determined from the dissociation experiment (e.g., via nucleotide chase [5, 6]).

Acknowledgments

This work was supported by H.-J.W.'s Alberta Innovates—Technology Futures (SC60-T2), Canadian Institutes of Health Research (MOP246392), and Natural Sciences and Engineering Research Council of Canada (05199-2016). H.E.B. receives funding from the Natural Sciences and Engineering Research Council of Canada Postgraduate Scholarship-Doctoral.

References

1. Delgado FF, Cermak N, Hecht VC, Son S, Li Y, Knudsen SM, Olcum S, Higgins JM, Chen J, Grover WH, Manalis SR (2013) Intracellular water exchange for measuring the dry mass, water mass and changes in chemical composition of living cells. PLoS One 8(7): e67590. https://doi.org/10.1371/journal.pone.0067590
2. Austin HP, Allen MD, Donohoe BS, Rorrer NA, Kearns FL, Silveira RL, Pollard BC, Dominick G, Duman R, El Omari K, Mykhaylyk V, Wagner A, Michener WE, Amore A, Skaf MS, Crowley MF, Thorne AW, Johnson CW, Woodcock HL, McGeehan JE, Beckham GT (2018) Characterization and engineering of a plastic-degrading aromatic polyesterase. Proc Natl Acad Sci U S A 115 (19):E4350–E4357. https://doi.org/10.1073/pnas.1718804115
3. Hu JH, Miller SM, Geurts MH, Tang W, Chen L, Sun N, Zeina CM, Gao X, Rees HA, Lin Z, Liu DR (2018) Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. Nature (London) 556 (7699):57–63. https://doi.org/10.1038/nature26155
4. Aigner TB, DeSimone E, Scheibel T (2018) Biomedical applications of recombinant silk-based materials. Adv Mater 30(19):1704636. https://doi.org/10.1002/adma.201704636
5. Shields MJ, Fischer JJ, Wieden H-J (2009) Toward understanding the function of the universally conserved GTPase HflX from *Escherichia coli*: a kinetic approach. Biochemistry 48 (45):10793–10802. https://doi.org/10.1021/bi901074h
6. Fischer JJ, Coatham ML, Eagle Bear S, Brandon HE, De Laurentiis EI, Shields MJ, Wieden H-J (2012) The ribosome modulates the structural dynamics of the conserved GTPase HflX and triggers tight nucleotide binding. Biochimie 94(8):1647–1659. https://doi.org/10.1016/j.biochi.2012.04.016
7. Rosler KS, Mercier E, Andrews IC, Wieden H-J (2015) Histidine 114 is critical for ATP hydrolysis by the universally conserved ATPase

YchF. J Biol Chem 290(30):18650–18661. https://doi.org/10.1074/jbc.M114.598227

8. Fersht A (2017) Structure and mechanism in protein science, vol 9. World Scientific, Singapore. https://doi.org/10.1142/10574
9. Gzyl KE, Wieden H-J (2017) Tetracycline does not directly inhibit the function of bacterial elongation factor Tu. PLoS One 12(5): e0178523. https://doi.org/10.1371/journal.pone.0178523
10. Visconti L, Malagrinò F, Toto A, Gianni S (2019) The kinetics of folding of the NSH2 domain from p85. Sci Rep 9(1):4058. https://doi.org/10.1038/s41598-019-40480-2

Chapter 13

Measurement of Nucleotide Hydrolysis Using Fluorescent Biosensors for Phosphate

Simone Kunzelmann

Abstract

Assays for the detection of inorganic phosphate (P_i) are widely used to measure the activity of nucleotide hydrolyzing enzymes, such as ATPases and GTPases. The fluorescent biosensors for P_i, described here, are based on fluorescently labeled versions of *E. coli* phosphate-binding protein (PBP), which translates P_i binding into a large change in fluorescence intensity. In comparison with other P_i-detection systems, these biosensors are characterized by a high sensitivity (sub-micromolar P_i concentrations) and high time resolution (tens of milliseconds), and they are therefore particularly well suited for measurements of phosphate ester hydrolysis in real time. In this chapter, it is described how the P_i biosensors can be used to measure kinetics of ATPase and GTPase reactions, both under steady state and pre-steady state conditions. An example protocol is given for determining steady state kinetic parameters, K_m and k_{cat}, of the ATP-dependent chromatin remodeler Chd1, in a plate reader format. In addition, the measurement of P_i release kinetics under pre-steady state conditions is described, including a detailed experimental procedure for a single turnover measurement of ATP hydrolysis by the ABC-type ATPase SufBC using rapid mixing.

Key words Phosphate detection, Kinetics, ATPase, GTPase, Phosphate release, Enzyme mechanism, Single turnover, Multi-turnover, Stopped-flow

1 Introduction

Inorganic phosphate is the product of numerous cellular reactions catalyzed by enzymes called phosphohydrolases. These include several families of nucleoside triphosphatases, e.g., ATPases and GTPases, and different types of phosphatases that cleave inorganic phosphate (P_i) from nucleic acid, protein, or small metabolite substrates. Fluorescent phosphate-binding protein is widely used as a probe to monitor P_i release from these reactions in real time.

A fluorescent version of the *Escherichia coli* phosphate-binding protein (MDCC-PBP) was first developed as a P_i biosensor in the laboratory of Martin Webb [1], initially with the aim to study the role of P_i release from myosin in muscle fibers [2, 3]. MDCC-PBP consists of a single cysteine mutant of periplasmic

Tina Daviter et al. (eds.), *Protein-Ligand Interactions: Methods and Applications*, Methods in Molecular Biology, vol. 2263, https://doi.org/10.1007/978-1-0716-1197-5_13,

phosphate-binding protein (PBP A197C) that is covalently modified with a diethylamino coumarin fluorophore via reaction of the cysteine thiol with a maleimide group on the fluorescent molecule. PBP has a two-lobed structure with the phosphate-binding site located in between the two lobes (*see* Fig. 1a). P_i binding induces a large rigid body movement between the two sub-domains leading to closure of the binding cleft. This brings the reporter fluorophore into a different environment which, by change of specific interactions with the protein [4], causes a large increase in fluorescence quantum yield (approximately eightfold) [5] and thus in measured fluorescence intensity (*see* Fig. 1b). MDCC-PBP binds phosphate rapidly ($k_{on} = 1.36 \times 10^8\ M^{-1}\ s^{-1}$) and tightly ($K_d \sim 0.1\ \mu M$, values at pH 7.0 and 22 °C) [1], properties which make it a very sensitive probe for fast changes in P_i concentrations.

There are several more recent variants of the original P_i biosensor. The first, Rho-PBP, has two cysteines (one on each lobe, A17C and A197C) modified with tetramethylrhodamine iodoacetamide [6]. The P_i-induced fluorescence change is based on a stacking interaction of the two rhodamines [7, 8] which is altered upon P_i binding. Rho-PBP has very similar P_i binding strength and dynamics to MDCC-PBP and is used in the same assay types. The main differences between Rho-PBP and MDCC-PBP are the optical properties of the fluorescent probe. Tetramethylrhodamine is more photostable than diethylamino coumarin and has higher excitation and emission wavelengths, properties that are advantageous when using high intensity light sources (e.g., lasers) or where optical interference with other assay components is an issue (e.g., small molecule screening). Another variant, Rho-PBPw (weak binding), is based on Rho-PBP with an additional mutation (I27G) that weakens the binding of P_i [9]. This version, as well as a genetically encoded variant of PBP [10], has potentially different applications, which are not described here.

Based on their design principle, the above P_i biosensors are classified as reagentless biosensors, which are built from a protein scaffold that specifically binds the molecule of interest and a reporter group (fluorophore) that generates an optical readout. Reagentless biosensors have been developed for the detection of a number of small molecule metabolites, e.g., ADP and GDP [11–13], ATP [14], the ATP/ADP ratio [15, 16], glucose [17, 18], or glutamate [17, 19], and many of them are based on periplasmic binding protein scaffolds [17–20]. Other detection systems exist, such as coupled-enzyme assays (e.g., [21, 22]) or dye-based assays, for example P_i-detecting phosphomolybdate assays (e.g., malachite green) [23]. However, reagentless biosensors offer the advantages of being simple, consisting of only one component, and having a fast response time which is limited only by the rate of binding and conformation change of the scaffold, generally being in the order of sub-seconds rather than up to minutes, e.g., for

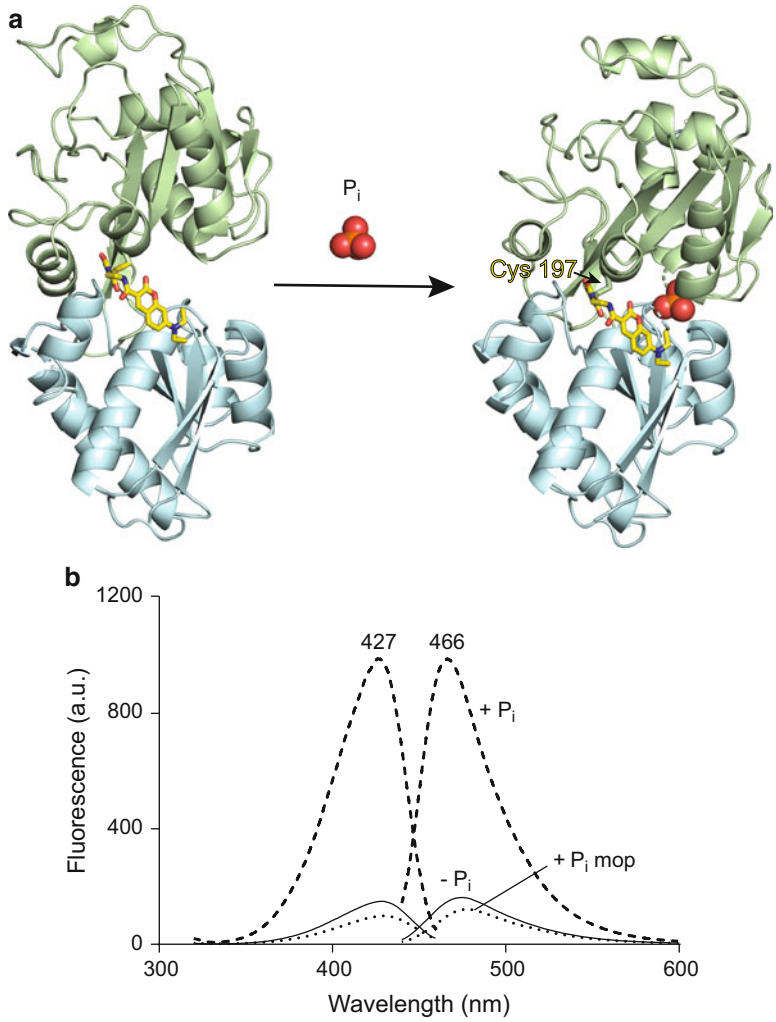


Fig. 1 Phosphate biosensor MDCC-PBP. (**a**) Conformation change of MDCC-PBP upon P_i binding. The structure of *E. coli* phosphate-binding protein is shown in the open, P_i-free conformation (left, PBP T141D, PDB 1OIB, [65]) and the closed, P_i-bound conformation (right, MDCC-PBP, PDB 1A54, [4]). The P_i bound structure is showing the actual structure of MDCC-PBP with the covalently attached fluorophore (yellow). In the P_i-free structure, the MDCC is just placed manually in the structure figure (light yellow) to illustrate the change in overall position of the fluorophore. (**b**) Fluorescence excitation and emission spectra of 2 μM MDCC-PBP in 10 mM PIPES pH 7.0. Fluorescence spectra are shown in the absence of additional phosphate (continuous line), after addition of 20 μM phosphate (dashed line) and after addition of P_i mop (400 μM 7-MEG and 1U/ml PNPase, dotted line). P_i mop treatment reduces the fluorescence intensity, indicating that there is a small amount of P_i contamination in the MDCC-PBP preparation

malachite green. Therefore, they are ideal tools for measuring the kinetics of enzyme reactions in real time, in steady state as well as pre-steady state experiments.

Steady state kinetic assays allow determination of k_{cat}, K_m, and the kinetic efficiency k_{cat}/K_m, as well as the study of activation and inhibition of an enzyme. Biosensors with very fast response times, like PBP, also provide the time resolution needed for pre-steady state, kinetic experiments using flash photolysis, or stopped-flow techniques, which can give a more detailed insight into the mechanism of the enzyme reaction.

In this chapter, I describe how to use phosphate-binding protein to measure the kinetics of ATPase and GTPase reactions. Protocols are given for determining steady state kinetic parameters, K_m and k_{cat}, in a plate reader format in the absence and presence of an activator. In addition, I describe examples of direct measurement of P_i release under pre-steady state conditions using stopped-flow and illustrate what information about the enzyme mechanism can be deduced from this type of data.

2 Materials

2.1 Phosphate Biosensor

The coumarin version of the biosensor, MDCC-PBP (*see* **Note 1**), is sold as ready-to-use phosphate sensor by Thermo Fisher Scientific (PV4406/7), but the biosensor used for the experiments in this chapter was purified and labeled in the laboratory, which is a straight forward procedure. Rho-PBP is not currently available commercially. The latest expression vectors for both biosensors are available from Addgene (http://www.addgene.org/): plasmids pET22b_PstS_1 for MDCC-PBP and pET22b_PstS_2 for Rho-PBP [9]. Detailed protocols for purification, labeling, and characterization of the biosensors can also be found on the Addgene webpage. Note that the expression system has changed to IPTG induction [9] since the original publications [1, 5, 6].

2.2 Reaction Buffer

The most critical factor for successful assays of P_i is to avoid contamination with P_i, which is a very widespread impurity (*see* **Note 2**).

All buffer components should be analytical grade. Water used for buffer preparation and rinsing should be used directly from the distillation apparatus or Milli-Q system without storage to avoid contamination. We have worked successfully with double-distilled water, Milli-Q, and commercial water (HPLC grade). Disposable plasticware is used where possible and everything is rinsed with P_i-free water before use. If glassware cannot be avoided, it is soaked in 25% HCl overnight and rinsed with water. The pH of final (diluted) buffers should not be adjusted using a pH electrode, if possible only adjust pH of stocks. If it cannot be avoided, either pH

paper is used or two solutions are prepared in parallel and the volume of acid/base needed for the adjustment of one is measured, then the same volume is added to the other, P_i-free, buffer.

Other than inorganic phosphate, most buffer components are compatible with P_i detection by phosphate-binding protein, including common reducing agents (DTT, TCEP, β-mercaptoethanol). Most nucleotide hydrolyzing enzymes depend on Mg^{2+} (or Mn^{2+}) as a cofactor. Therefore 1 to 10 mM $MgCl_2$ is generally included in the buffer for these reactions.

Adding BSA or detergent can help to prevent sticking of proteins to the surface of tubes, plates, and cuvettes and therefore stabilizes enzymes at very low concentration. We generally use BSA (5–20 μM) for enzyme solutions at concentrations of 0.5 μM or lower.

2.3 Nucleotide Solutions

Nucleotide solutions are generally the biggest source of P_i contamination due to the spontaneous hydrolysis of phosphodiester bonds in water. The stability of nucleoside triphosphates depends on the nature of the nucleotide. Purine nucleotides (G and A) are less stable than pyrimidine nucleotides (C and T).

For ATPase reactions, we use adenosine 5′-triphosphate disodium salt hydrate (BioXtra, >99% (HPLC)) from Sigma-Aldrich. This ATP source normally contains between 0.2 and 0.5% (mol/mol) inorganic phosphate and can be used at up to 500–1000 μM in PBP-based assays. GTP, guanosine 5′-triphosphate sodium salt hydrate (>95% by HPLC), can also be purchased from Sigma-Aldrich, but at lower purity than ATP. Both nucleotides can be further purified by anion-exchange chromatography if needed for a particular assay [11].

Solutions are made up at 20–100 mM in equimolar Tris–HCL (pH 8.0–9.0). If necessary, the pH is adjusted with NaOH using pH paper. All nucleotide solutions are stored at −20 °C or −80 °C (purified nucleotides).

2.4 P_i Standard

The fluorescence response of MDCC-PBP or Rho-PBP is calibrated with an inorganic phosphate standard. Here we use a Centripur® phosphate standard solution, 1000 mg/ml PO_4^{3-} (Merck), corresponding to 10.53 mM phosphate ions.

2.5 P_i Mop

To remove P_i contaminations from multi-well plates and the stopped-flow instrument, a P_i mop [21] system is used which consists of 7-methylguanosine (M0627, Sigma-Aldrich) and purine nucleoside phosphorylase (N8264, Sigma-Aldrich) (*see* **Note 3**).

PNPase (purine nucleoside phosphorylase) catalyzes the phosphorolysis of 7-methylguanosine (7-MEG) to form ribose-1-phosphate and guanine, thereby removing free phosphate from the solution. With this system, P_i concentrations can be lowered to about 0.1 μM [1].

7-MEG is dissolved in water at 20 mM and stored at −20 °C. PNPase is dissolved in water at 1000 U/ml, shock-frozen in small aliquots, and stored at −80 °C. Repeated freeze–thaw cycles reduce the activity of PNPase and should be avoided.

For cleaning instruments and cuvettes, a solution of 200 μM 7-MEG and 1 U/ml PNPase in P_i-free buffer is prepared. The P_i mop can also be used to pretreat nucleotide and protein solutions prior to kinetics experiments, but in this case one has to carefully choose the PNPase concentration so that the removal of P_i by the mop is much slower than the P_i release from the system under study. This is normally only done for stopped-flow experiments.

2.6 Instrumentation

For steady state kinetic assays:

1. Fluorescence plate reader equipped with monochromators for excitation and emission light or suitable filter set for MDCC or tetramethylrhodamine fluorescence, ideally with temperature control (here we use a CLARIOstar reader from BMG Labtech).
2. 96- or 384-well plates for fluorescence (black plates), here we use 384-well plates, low volume, NBS (3820, Corning) (*see* **Note 4**).

For pre-steady state kinetics:

1. Stopped-flow instrument, setup for fluorescence detection. We use a HiTech SF61 DX2 stopped-flow instrument equipped with a Xe/Hg lamp (TgK Scientific, UK).
2. Long-pass filter suitable for MDCC or tetramethylrhodamine fluorescence detection (Schott GG 455 for MDCC-PBP or Schott OG570 for Rho-PBP, both from UQG Optics).

2.7 Data Fitting Software

1. For steady state experiments, MARS software (BMG Labtech) is used to determine initial rates by linear regression (alternatively this can be done using Microsoft Excel, GraphPad Prism, or other software).
2. Prism 7 (GraphPad) is used to run nonlinear least-square fits and to prepare all plots.
3. For pre-steady state experiments, Kinetic Studio (HiTech) or Prism 7 (GraphPad) is used for nonlinear least-square fitting using exponentials.
4. KinTek Explorer (KinTek Corporation) software [24] is used for all simulations.

3 Methods

3.1 Steady State Kinetic Assay to Determine K_m and k_{cat}

In steady state kinetics experiments, one measures the initial rate at which substrate (S) is turned over by a small amount of enzyme ($[E]_0 \ll [S]$), most commonly at different concentrations of substrate to determine K_m and k_{cat}. These assays are also used to study the mechanism of activation or inhibition by other molecules and the apparent affinities of activators and inhibitors.

Here, it is described how to use phosphate-binding protein to measure K_m and k_{cat} of a P_i-producing enzyme using a fluorescence plate reader. I describe general considerations for the experimental design and procedure and present a detailed example of characterizing the ATPase activity of Chd1, an ATP-dependent chromatin remodeler, and its activation by DNA. Formation of P_i in real time is measured using MDCC-PBP. K_m and k_{cat} for ATP hydrolysis by Chd1 are determined in the absence and presence of activating DNA.

3.1.1 General Principles

To measure initial rates, enzyme and substrate are mixed in the presence of MDCC-PBP and the fluorescence signal is monitored over time (*see* Fig. 2a).

A linear fit to the data within the steady state phase gives the rate in fluorescence units per time interval. Using a calibration with a P_i standard (*see* Fig. 3), the rate of fluorescence change is converted into the rate of P_i produced.

To determine K_m and k_{cat} for an enzyme–substrate reaction, initial rates are measured using a range of substrate concentrations. The dependence of the initial rate, V, on substrate concentration [S] is described by the Michaelis–Menten equation, which is in its original form:

$$V = \frac{V_{max} \cdot [S]}{K_m + [S]} \quad \text{with} \quad V_{max} = k_{cat} \cdot [E]_0 \tag{1}$$

Often it is useful to calculate and plot the specific rate ν, which is the initial rate normalized to the enzyme concentration $\nu = V/[E]_0$ (*see* Fig. 2b). The specific rate ν is given by

$$\nu = \frac{k_{cat} \cdot [S]}{K_m + [S]} \tag{2}$$

Presenting the data on a plot of specific rate, ν, versus substrate concentration, [S], according to Eq. (2), enables the direct visualization of k_{cat} (ν at saturating concentration), as well as the comparison of multiple datasets, which have been measured at different enzyme concentrations. It is common practice to vary enzyme concentration as appropriate, to accurately measure reactions with

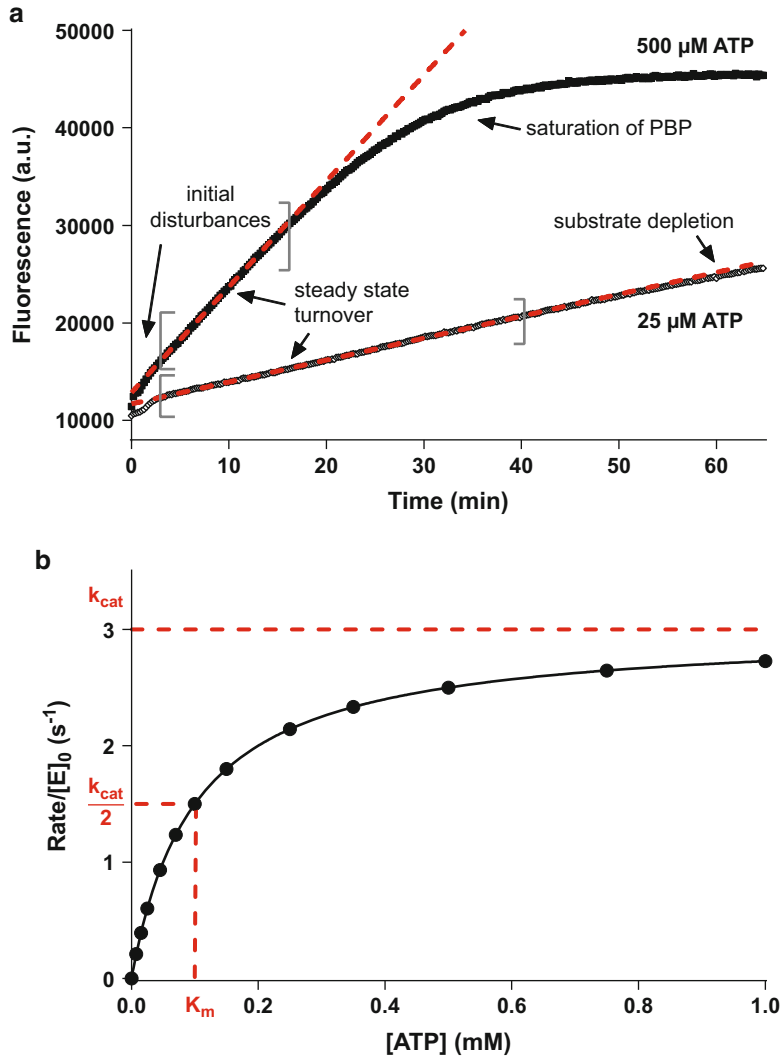


Fig. 2 Steady state rate measurement using the P_i biosensor. (**a**) Example of experimental fluorescence traces obtained after mixing enzyme and nucleotide (ATP) in the presence of MDCC-PBP and monitoring fluorescence intensity over time. The reaction rates are determined by linear regression using the data points within the grey brackets. (**b**) Plot of the specific reaction rate ($\nu = V/[E]_0$) versus substrate concentration (simulated data). Data were generated using the parameters $K_m = 100$ µM and $k_{cat} = 3\ s^{-1}$

very different turnover rates, for example, in the case when comparing enzyme mutations that significantly alter the rate of substrate turnover or comparing activity in the absence and presence of an activator (*see* Fig. 4c).

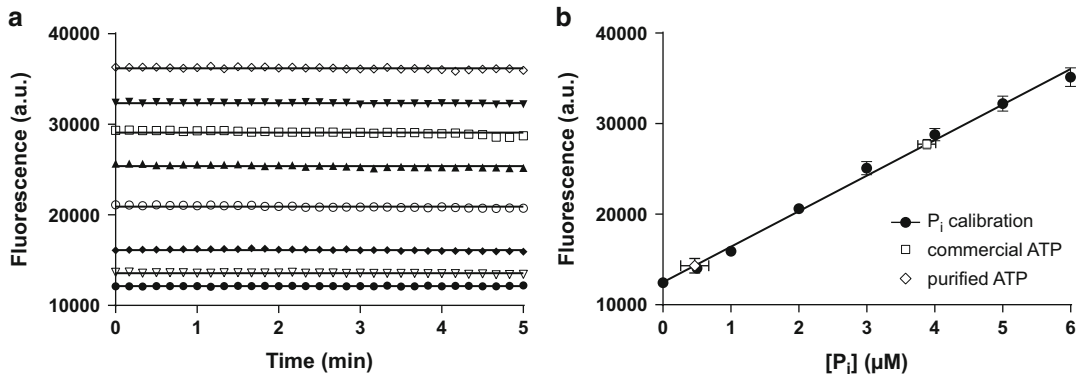


Fig. 3 Calibration of the MDCC-PBP fluorescence signal with phosphate standard solutions in a fluorescence plate reader (*see* Subheading 3.1.3). (**a**) Time-dependent fluorescence of 15 μM MDCC-PBP after addition of 0, 0.5, 1, 2, 3, 4, 5, and 6 μM phosphate (from bottom to top), illustrating the stability of the signal over time. (**b**) Plot of the average fluorescence from traces in (**a**) versus P_i concentration (filled circles). The line is the result of linear regression analysis with slope = 3900 ± 100 μM^{-1} and intercept = 12,500 ± 400. Open symbols show the fluorescence in the presence of 1 mM commercial ATP (square) or purified ATP (diamond), as described in Subheading 2.3. From the calibration curve, it is estimated that they contain 3.9 ± 0.1 μM (0.39% of ATP) and 0.5 ± 0.2 μM (0.05%) phosphate, respectively. Data are shown as mean ± SD ($n = 3$)

3.1.2 Experimental Design and Optimization

Biosensor Concentration

To determine the concentration of MDCC-PBP required, one must consider the amount of P_i to be measured and the amount of P_i contamination in the reagents. Generally, between 10 and 20 μM MDCC-PBP is appropriate, allowing detection of P_i with a linear response that extends until saturation of approximately half of the PBP (5–10 μM, *see* Fig. 3b). If very small amounts of P_i are to be measured, the PBP concentration can be reduced to low micromolar concentrations in order to gain sensitivity but additional care must be taken to minimize P_i contamination from reagents. In contrast, if P_i contamination is significant or higher P_i concentrations are to be measured, the MDCC-PBP concentration can be increased.

Concentration of Substrate

For accurate measurements of K_m for a particular substrate, a concentration range from <10-fold below the K_m up to at least 5- to 10-fold above the K_m is recommended. The choice of highest substrate concentration may be limited by P_i contamination of the substrate stock, especially if nucleoside triphosphates are used as substrates (*see* Subheading 2.3). Reagent stocks can be tested for P_i contamination using the phosphate biosensor prior to the experiments (see below). The lowest measurable substrate concentration depends on the detection limit of P_i under the conditions used.

Enzyme Concentration

As mentioned above, in steady state kinetics the enzyme concentration is much lower than the lowest substrate concentration used ($[E]_0 \ll [S]$), in order to ensure that the system reaches steady state before significant substrate turnover is detected.

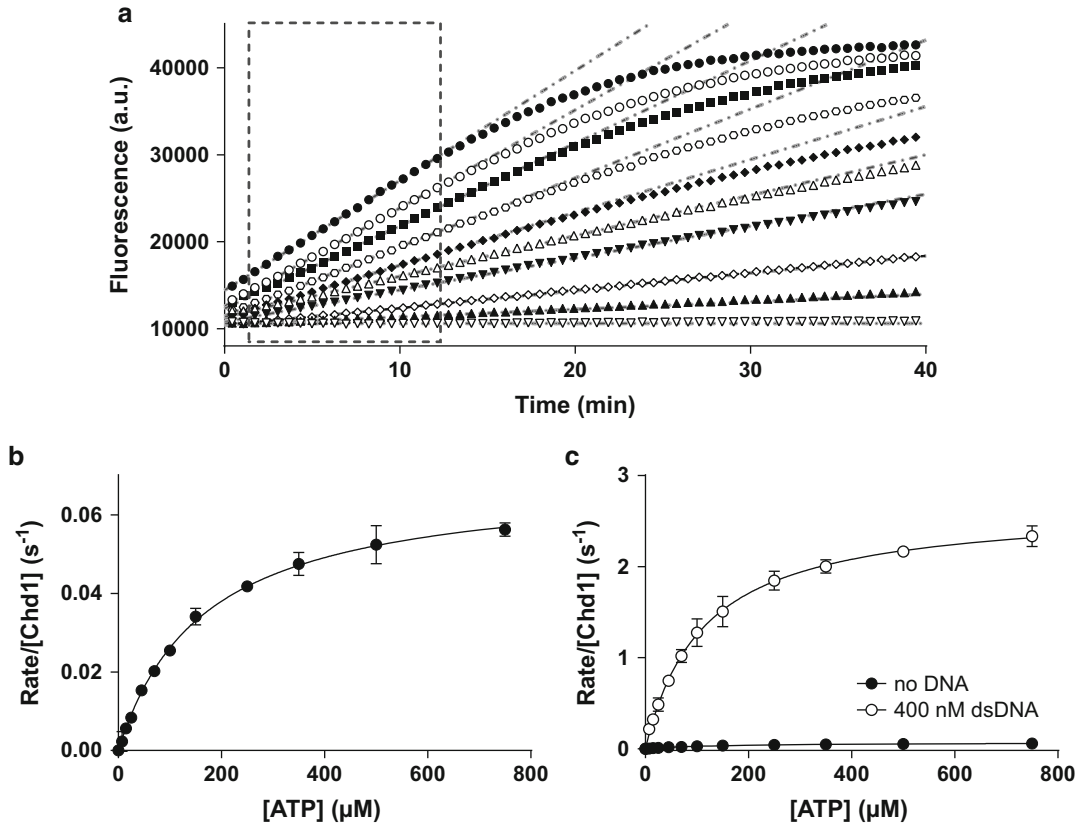


Fig. 4 ATPase activity of Chd1 in the absence and presence of dsDNA. (**a**) Time courses of MDCC-PBP fluorescence change after mixing Chd1 and ATP in the presence of MDCC-PBP. The reactions contained 15 μM MDCC-PBP, 0.1 μM Chd1, and 0–750 μM ATP (bottom to top) as described in Subheading 3.1.3. Note that at higher ATP concentrations, the curves start at higher fluorescence levels which is partly due to some P_i already formed during the dead time (while mixing and starting data acquisition) and partly due to P_i contamination of the ATP. Initial rates of fluorescence change were determined by linear regression to the data points between 2 and 12 min (grey box) and converted to P_i release rates (nM P_i/s) as described in the text. The fits are shown as dash-and-dotted lines. (**b**) Plot of the initial rates, normalized to the enzyme concentration, versus ATP concentration. Data were analyzed using the Michaelis–Menten equation (Eq. 2). The fitting results are $K_m = 166 \pm 10$ μM and $k_{cat} = 0.069 \pm 0.002$ s^{-1}. (**c**) Plot of the specific initial rates versus ATP concentration for dsDNA stimulated Chd1 ATPase (open circles) in comparison to the data without DNA from (**b**) (closed circles). The reactions contained 15 μM MDCC-PBP, 2 nM Chd1, and 400 nM 30 bp DNA as described in Subheading 3.1.3. Data analysis using the Michaelis–Menten equation (Eq. 2) yields $K_m = 112 \pm 6$ μM and $k_{cat} = 2.66 \pm 0.05$ s^{-1}. dsDNA stimulates ATPase activity by increasing k_{cat} about 40-fold whereas K_m is only slightly lowered. Data in (**b**) and (**c**) are represented as mean ± SD ($n = 3$)

The rate of the reaction (V) is directly proportional to the enzyme concentration (*see* Eq. 2). Thus, the enzyme concentration can be used to control the rate of the reaction. In practice, the assay works well if a few micromolar P_i are produced over 10 min at the highest substrate concentration and the maximal measurement time is ~20–60 min. If the enzyme concentration is too low and

the reaction is much slower than this, unwanted signal drifts could impair the data quality. If the reaction is too fast, then the initial rate may be underestimated, and this depends on the sampling rate of the plate reader.

With a new reaction system, one can vary the enzyme concentration to optimize the assay in advance. Testing the linearity of the reaction rate with enzyme concentration is also good practice, for example, to exclude problems with enzyme stability at low enzyme concentration. However, nonlinearity does not necessarily point to an artifact and could instead be a physiological feature of the system, for example, due to the enzyme activity being regulated by self-assembly [25].

Fitting of Initial Rates

As a rule of thumb, to capture the initial rate prior to substrate depletion and significant product inhibition, one should only use measurements made before 10% of the substrate is depleted. In practice, there are other, often assay-specific, reasons that define the range over which the product formation is linear with time. Figure 2a shows example traces of P_i formation during ATP hydrolysis by Chd1 with two different ATP concentrations. The signal appears unstable for the first minute of the reaction, likely due to the settling of the meniscus in the wells and/or temperature equilibration. Subsequently, the signal reaches a linear steady state phase. At high substrate concentration, the signal plateaus due to saturation of MDCC-PBP. At low substrate concentration, the signal deviates from linearity much later in time but at lower fluorescence level, likely due to substrate depletion. While 10% turnover can be used as an upper limit to fit initial rates, visual inspection is required to judge the linear range to be fit (grey brackets in Fig. 2a).

3.1.3 Example Protocol: Steady State Kinetic Assay—Chd1 ATPase

Chd1 is an ATP-dependent chromatin remodeler that uses ATP hydrolysis to move and position nucleosomes along DNA [26, 27]. The intrinsic ATPase activity of Chd1 is low, but ATP hydrolysis is strongly stimulated by the binding of double-stranded DNA (dsDNA) or nucleosomes [28, 29], which relieves an intramolecular autoinhibition [28]. MDCC-PBP is used here to measure the K_m and k_{cat} for the basic and dsDNA-stimulated ATPase of Chd1.

Buffer and Concentrations

The assay is performed in buffer containing 30 mM Tris–HCl pH 7.5, 50 mM KCl, 5 mM $MgCl_2$, and 1 mM TCEP (using components with minimal P_i, *see* Subheading 2.2) at 25 °C (*see* **Note 5**). The buffer is prepared freshly and kept at room temperature throughout the experiments.

The final concentrations of assay components are:

1. 15 μM MDCC-PBP.
2. 100 nM and 2 nM Chd1 (in the absence and presence of DNA, respectively) (*see* **Note 6**).

3. 7.5–750 μM ATP (*see* **Note 7**).
4. 0 or 400 nM dsDNA oligonucleotide (*see* **Note 8**).
5. 5 μM BSA.

The assay is run in 20 μl volume reactions in Corning 384 low-volume plates (*see* **Note 9**) in a BMG CLARIOstar plate reader. Stock solutions of the reaction components are prepared at two or fourfold the final concentration, adding 10 or 5 μL, respectively, to the wells. The enzyme stock (Chd1) contains 20 μM BSA to stabilize the enzyme at low concentrations.

Calibration

Calibration of the MDCC-PBP signal is best performed at the beginning of the experiments (*see* Fig. 3). Since the fluorescence signal of the calibration run does not change over time, it can be used to set up the optics in the plate reader protocol. When the assay is run for the first time, the P_i contamination in the ATP stock solution is ideally also tested.

1. Set the temperature of the plate reader to 25 °C.
2. Set up the method in the plate reader software:
   (a) Use fluorescence intensity—plate mode.
   (b) Choose plate type: Corning 384 low-volume flat bottom.
   (c) Excitation wavelength: 430–8 (430 nm with 8 nm slit width).
   (d) Dichroic: automatic (447.5 nm).
   (e) Emission wavelength: 465–8 (465 nm with 8 nm slit width).
   (f) Select top optics, settling time 0.2 s, no. of flashes 40, cycle time 10 s, number of cycles 30.
   (g) Shaking frequency 500 rpm, mode double orbital, time 15 s before first cycle.
3. Prepare 60 μM MDCC-PBP solution in reaction buffer. Make enough solution for the calibration and the experiments you want to run.
4. Prepare 20 μM BSA solution.
5. In 8-strip PCR tubes, prepare a concentration series of P_i solutions (0, 1, 2, 4, 6, 8, 10, 12 μM) from the 10.53 mM P_i standard stock solution (and, optional, 2 mM ATP solution to test for P_i).
6. In a 384-well plate, pipette 5 μl MDCC-PBP solution into 10 wells and add 5 μl BSA.
7. With a multichannel pipette, add 10 μl P_i standard solution (and ATP test solution) to the wells, mix by pipetting up and down a couple of times, and place in the plate reader.

8. Set the gain and focal height:
   (a) Run gain and focal height adjustment on the well with the highest P_i concentration.
   (b) Set gain, so that the output signal is well above background, but far lower than saturation (in this example gain 1300, gives ~35,000 RUs).
   (c) Run the focal height adjustment (gain adjustment switched off) for several wells to check that there is no large variation between wells (<±0.1 mm) and set to an average (here 7.0 mm).
9. Run measurement (about 5 min).

To analyze data:

10. Plot the fluorescence signal versus time for all P_i concentrations (and ATP test solutions), and calculate the average of the fluorescence signal over the time points leaving out the initial time points if the signal is not stable (*see* Fig. 3a).
11. Plot the averaged fluorescence (F) versus P_i concentration to generate the standard curve. Fit a standard curve by linear regression (*see* Fig. 3b):

$$F = \text{slope} \times [P_i] + \text{intercept} \quad (3)$$

    The slope gives the fluorescence units per μM P_i ($\Delta F/\Delta[P_i]$) and is used to calculate how much P_i is produced from fluorescence transients.
12. Calculate the concentration of P_i in the ATP samples from the standard curve above (*see* Eq. 3).

Determine k_{cat} and K_m of the Basal and dsDNA-Activated ATPase of Chd1

1. Set up plate reader temperature and acquisition method as described in the previous section, but change cycle time to 13 s (minimum time to read 24 wells) and no. of cycles to 300.
2. In 8-strip PCR tubes, prepare a concentration series of ATP (*see* **Note 7**) (0, 15, 30, 50, 90, 140, 200, 300, 500, 700, 1000, 1500 μM).
3. Prepare a solution of 400 nM Chd1 plus 20 μM BSA, and another solution of 8 nM Chd1 plus 1.6 μM dsDNA and 20 μM BSA.
4. In a 384-well plate, pipette 5 μl 60 μM MDCC-PBP solution into 24 wells.
5. Add 5 μl Chd1 solution to the first 12 wells and 5 μl Chd1 plus 30 bp DNA to the other 12 wells.
6. With a multichannel pipette, add 10 μl ATP solution into the wells, mix by pipetting up and down a couple of times. Place in the plate reader and immediately start data acquisition. Read for 30 min.

To analyze data:

7. Plot the fluorescence signal versus time for all ATP concentrations (*see* Fig. 4a, for example, data for the experiment in the absence of DNA). Determine the initial rate of fluorescence change ($V_F = \Delta F/\Delta t$) at each ATP concentration by linear regression using a data range where the fluorescence increase is linear (boxed region in Fig. 4a) (*see* **Note 10**).
8. Correct the rate of fluorescence change V_F for background by subtracting the value at zero ATP.
9. Calculate the initial rate V using the corrected V_F and the slope from the calibration above (Eq. 3)

$$V = \frac{V_F}{\text{slope}} = \frac{\Delta F}{\Delta t} \times \frac{\Delta[P_i]}{\Delta F} \tag{4}$$

and the specific rate, dividing by the enzyme concentration.

$$\nu = \frac{V}{[E]_0} \tag{5}$$

10. Plot the specific rate ν versus ATP concentration (*see* Fig. 4b, c). Run a nonlinear least-square fitting using the modified form of the Michaelis–Menten equation (*see* Eq. 2) to determine K_m and k_{cat}.

Testing the Linearity of Measured ATPase Rates with Enzyme Concentration

When establishing a steady state assay with a new enzyme, it is good practice to check that the rate varies linearly with the enzyme concentration and thereby also to determine the optimal enzyme concentrations for the assay.

1. Set up plate reader temperature and acquisition method as in previous sections.
2. Prepare 60 μM MDCC-PBP, 1.5 mM ATP, and a concentration series of Chd1 plus 20 μM BSA (0, 0.1, 0.2, 0.4, 0.8, 1.6 μM).
3. In a 384-well plate, pipette 5 μl MDCC-PBP and 5 μl Chd1 solution.
4. Start the reaction by adding 10 μl ATP solution and record data for 30 min.

To analyze data:

5. Plot the fluorescence over time and determine the initial rate, V, at different enzyme concentrations, as described before (*see* Fig. 5a).
6. Plot the initial rate V versus enzyme concentration and analyze by linear regression (*see* Fig. 5b). The data should be well described by a linear function with an intercept close to zero.

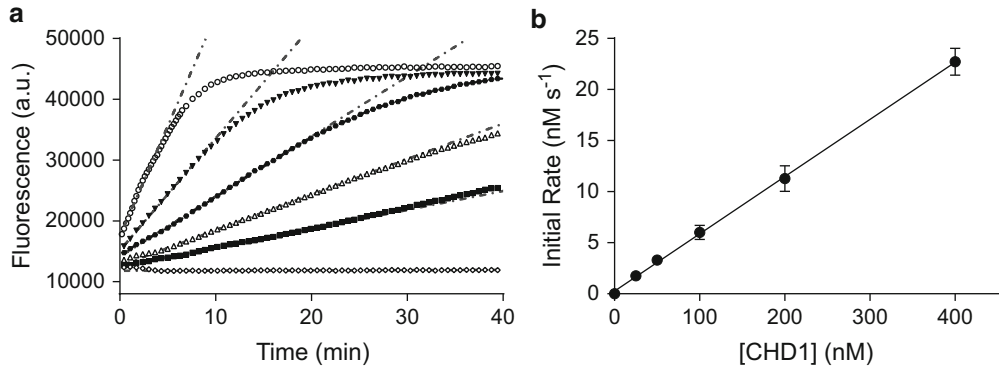


Fig. 5 Testing linearity of the ATPase rate with enzyme (Chd1) concentration (**a**) Time-dependent fluorescence change after mixing 750 μM ATP and 0, 25, 50, 100, 200, and 400 nM Chd1 (from bottom to top) in the presence of 15 μM MDCC-PBP. Data were analyzed by linear regression to determine the initial rates. The fitting range has been adapted for the different Chd1 concentrations using 0–4 min for 400 nM, 0–8 min for 200 nM, 5–18 min for 100 nM, and 8–25 min for the other concentrations. (**b**) Plot of the initial rates versus Chd1 concentration. Linear regression analysis yields the slope, 0.056 ± 0.006 s^{-1}. The value corresponds well to the specific rate obtained at 750 μM ATP in the experiment in Fig. 4b

3.2 P_i Release Kinetics Under Pre-steady State Conditions Using Stopped-Flow

Steady state kinetic assays give information about the efficiency and specificity of an enzyme (from the ratio of k_{cat}/K_m), as well as, of course, the rate constant of the rate-limiting step (k_{cat}). However, they provide limited insight into the mechanism of an enzyme reaction, such as the number of steps and intermediates involved, and the rate constants of each individual step, which are accessible only through measurement of pre-steady state kinetics. Even the minimal reaction scheme for enzyme-catalyzed ATP (or GTP) hydrolysis comprises four steps: ATP binding, to an enzyme (E), ATP cleavage, dissociation of P_i, and dissociation of ADP (Scheme 1). Product dissociation often takes place in the order of P_i first, then ADP (as shown in Scheme 1), but can also occur the other way around, or in random order.

ATPase (or GTPase) activity is usually coupled to another process, e.g., motility, transport, or signaling, via conformational changes during the ATPase cycle. The actual kinetic pathway is more complex as it includes these conformational changes as separate steps and also binding events with other molecules, e.g., the protein or DNA track for translocating motor proteins or effector and regulators for signaling proteins.

In order to obtain information about individual steps, one has to monitor the first cycle (first turnover) of the enzyme reaction before it reaches the steady state as after that the rate of all following turnovers will only depend on the rate-limiting step(s) of the reaction. This means that experiments are performed at high enzyme concentrations and measured on the time scale within

Scheme 1 ATP binding, to an enzyme (*E*), ATP cleavage, dissociation of P_i, and dissociation of ADP

which the first turnover occurs, usually milliseconds to seconds, by using flash photolysis or rapid mixing techniques like stopped-flow. While steady state kinetics experiments always provide the same information, irrespective of the substrate or product that is monitored, in pre-steady state assays, different methods for detecting substrates, intermediates, or products are combined to dissect the individual steps in the pathway of an enzyme reaction. The use of a phosphate biosensor enables the direct measurement of when and how fast P_i is released.

MDCC-PBP has been widely applied to study the kinetic mechanisms of nucleotide hydrolyzing enzymes, in particular for motor proteins like myosins [30–35], kinesins [36–38], or DNA and RNA helicases [39–42], but also for many others, e.g., GTPases of the Ras and dynamin superfamilies [43, 44] or GTPases involved in protein synthesis at the ribosome [45–47]. While many of these examples, as well as the description in this chapter, focus on the use of MDCC-PBP to specifically measure the P_i release step, the biosensor has also been used to monitor nucleotide hydrolysis in general and relate hydrolysis to mechanical events, such as force generation in muscle fibers [2, 3] or DNA translocation. Interesting examples include the measurement of translocation rates of helicases as well as the coupling ratio that is the number of nucleoside triphosphates hydrolyzed per base translocated [48–50].

3.2.1 General Principle

To measure the transient kinetics of P_i release, e.g., for an ATPase or GTPase reaction, a stopped-flow instrument is used to rapidly mix the enzyme and substrate (within <2 ms) in the presence of the P_i biosensor, and the fluorescence change is recorded over the subsequent milliseconds or seconds. It is critical that one considers the rate of P_i binding to the P_i biosensor, which should be much faster (>10-fold) than the reaction under study in order to measure P_i release kinetics accurately. At 22 °C, the observed rate constant for P_i binding to MDCC-PBP is higher than 300 s^{-1} at P_i or MDCC-PBP concentrations >2 μM [1], and rate constants above 1000 s^{-1} have been measured at >50 μM (Martin Webb, personal communication) (*see* **Note 11**). Hence, at 10 μM MDCC-PBP, a commonly used concentration for these experiments, P_i release rate constants up to 30 s^{-1} can be easily measured and up to 100 s^{-1} is measurable if the MDCC-PBP concentration is increased to >50 μM (*see* **Note 12**). Since MDCC-PBP binds P_i tightly and is

present in excess of the measured P_i, it can be assumed that all of the P_i released by the enzyme under study will be bound by MDCC-PBP. MDCC-PBP therefore acts not only as a sensor but also as a phosphate trap, making the P_i release step quasi-irreversible.

Transient kinetics experiments can be performed in two different ways, under either single or multi-turnover conditions. In *single turnover experiments*, the enzyme is in excess of the substrate, ideally at concentrations high enough to ensure fast enzyme–substrate binding ($k_{+1} \times [E] \gg k_{+2}$) and substrate saturation by the enzyme ($[E] \gg K_{d1}$), so that binding is close to completion before the subsequent ATP/GTP cleavage (k_{+2}, all rate constants refer to the numbering in Scheme 1) (*see* **Note 13**).

Figure 6a, b show a simulation of single turnover experiments performed with a large excess of enzyme. Three scenarios are possible in this case (*see* **Note 14**): (1) P_i release is much faster than chemical cleavage ($k_{+2} \ll k_{+3}$), (2) cleavage is much faster than P_i release ($k_{+2} \gg k_{+3}$), or (3) both steps have similar rate constants ($k_{+2} \approx k_{+3}$). The first two cases give rise to single-exponential P_i release kinetics (red trace, Fig. 6a), with a rate constant k, that corresponds to the slower step, either cleavage ($k = k_{+2}$) or P_i release ($k = k_{+3}$) described by Eq. 6:

$$[P_i] = [ATP]_0\left(1 - e^{-kt}\right) \tag{6}$$

In the third case, $k_{+2} \approx k_{+3}$, the P_i release trace shows a significant lag phase governed by the rate constant of the faster step followed by an exponential increase corresponding to the slower step (*see* Fig. 6b). The data follow a double-exponential equation, which describes the kinetics of two consecutive, unimolecular steps:

$$[P_i] = [ATP]_0\left(1 + \frac{k_{slow}\, e^{-k_{fast}\, t} - k_{fast}\, e^{-k_{slow} t}}{k_{fast} - k_{slow}}\right) \tag{7}$$

k_{fast} and k_{slow}, correspond to k_{+2} and k_{+3}, for the case of P_i release kinetics described here. Based solely on P_i release data, it is not possible to identify the faster step (cleavage or P_i release) in either of the cases above. (Note that k_{fast} and k_{slow} are interchangeable in Eq. 7). Therefore, the rate constant of chemical cleavage, k_{+2}, is usually measured in a separate experiment, for example, using the quench-flow technique, where the total concentration (free plus enzyme-bound) of product, ADP/GDP or P_i is monitored (*see* Fig. 6a, b) [42, 43].

In *multi-turnover experiments*, the enzyme is mixed with an excess of substrate and the reaction is measured for the first few turnovers of the enzyme, so one can observe possible transient phases before the reaction reaches a steady state. The substrate concentration is ideally high enough for enzyme–substrate binding to be fast and for the substrate to saturate the enzyme. While single-

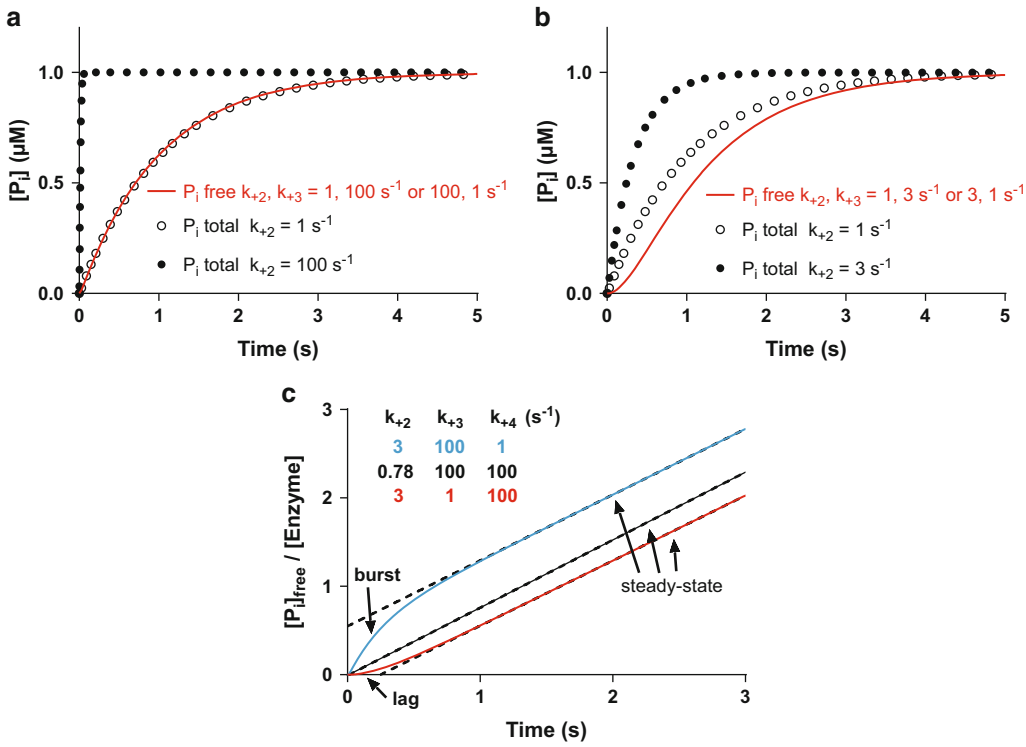


Fig. 6 Simulation of P_i release kinetics under single- and multi-turnover conditions. Simulations were performed according to Scheme 1 (*see* Subheading 3.2), assuming ATP cleavage and P_i release are irreversible (k_{-2}, $k_{-3} = 0$) (**a**) Single-turnover kinetics with one rate-limiting step (ATP cleavage (step 2) or P_i release (step 3)). Time courses of free P_i (red line, as measured with MDCC-PBP) and total P_i (symbols, quench-flow experiment) were simulated for hydrolysis of 1 μM ATP by 10 μM enzyme with $k_{+1} = 10\ \mu M^{-1}\ s^{-1}$, $k_{-1} = 5\ s^{-1}$ and k_{+2}, $k_{+3} = 1$, 100 s^{-1} or 100, 1 s^{-1} as indicated in the figure. Under these conditions ATP binding is fast, ~105 s^{-1} ($k_{+1} \times [E] + k_{-1}$) and nearly saturated ($[E] = 20 \times K_{d1}$) and will not rate-limit subsequent steps. The stopped-flow trace, monitoring released P_i, shows a single exponential with a rate constant (1 s^{-1}) that corresponds to the slower of the two steps, 2 or 3. The quench-flow experiment, showing free plus enzyme-bound P_i, directly monitors the cleavage reaction (step 2) and shows if cleavage (data overlapping with P_i release trace) or P_i release (data much faster than P_i release trace) is the rate-limiting step. (**b**) Single-turnover kinetics when both, ATP cleavage and P_i release, are rate-limiting. Simulation parameters were as in (**a**) but k_{+2}, $k_{+3} = 1$, 3 s^{-1} or 3, 1 s^{-1} as indicated in the figure. The P_i release trace (red) is biphasic, showing an initial lag governed by the faster rate constant followed by an exponential increase with the slower rate constant. The stopped-flow P_i release experiment cannot distinguish between release and hydrolysis, so again, the quench-flow experiment monitoring total amount of P_i from hydrolysis complements this information (**c**) P_i release kinetics under multi-turnover conditions (1 μM enzyme and 50 μM ATP), for cases where ATP cleavage and ADP release (blue), only ATP cleavage (black) or cleavage and P_i release (red) are rate limiting. Data were simulated with $k_{+1} = 10\ \mu M^{-1}\ s^{-1}$, $k_{-1} = 5\ s^{-1}$, $K_{d4} = 10$ μM, and k_{+2}, k_{+3} and k_{+4} as shown in the figure. The kinetics look very similar if the values for k_{+2} and k_{+3} are swapped. Note that the data are plotted as the ratio of P_i to enzyme concentration, so a value of one corresponds to the first turnover. All simulations were performed using KinTek Explorer software

turnover experiments only provide information about the steps preceding the formation of the species being detected (in this case all steps prior to and including release of free P_i, **step 3**) multi-turnover kinetics can also give insight into the succeeding steps (in this case ADP release, **step 4**). For example, a reduction of the reaction rate after the first turnover (burst kinetics) indicates that there is a slow, rate-limiting step following the measured intermediate state (*see* Fig. 6c, blue line).

Figure 6c displays simulated data for three different scenarios in a multi-turnover experiment. In the first case, there is only one rate-limiting step, either ATP cleavage (**step 2**) or P_i release (**step 3**), and the other steps, **steps 3** or **2** and ADP release (**step 4**), are fast. P_i release data then show an approximately linear increase without a detectable, initial transient phase (*see* Fig. 6c, black line). The slope is the steady state rate constant, k_{ss}, which corresponds to the specific rate, ν, in steady state experiments (*see* Eq. 2), and, under conditions of saturating substrate concentrations, to k_{cat}. In the second case, if a step following P_i release is also rate-limiting, e.g., ADP dissociation (**step 4**), a transient burst in P_i concentration is observed, as described above (blue line in Fig. 6c). In case a rate-limiting step precedes P_i release (**steps 1** or **2**), data are characterized by a lag before reaching steady state (red line in Fig. 6c). This behavior is described by Eq. 8:

$$\frac{[P_i]}{[E]} = A_{burst/lag} \times \left(1 - e^{-k_{burst/lag}t}\right) + k_{ss} \times t \tag{8}$$

The amplitude $A_{burst/lag}$ is positive for burst kinetics and negative if there is a lag. The steady state rate constant, k_{ss}, the rate constant of the burst or lag, $k_{burst/lag}$, and the amplitude, $A_{burst/lag}$, are all functions of the rate constants of the individual steps in a proposed mechanism. A simple example where there are only two rate-limiting steps is described in reference [51]. For more complex mechanisms, data might need to be analyzed using kinetic simulations, e.g., using KinTek Explorer (*see* Subheading 2.7) or DynaFit [52].

It is important to note that P_i release is normally only one measurement in a series of transient kinetics experiments that need to be combined to define the mechanism of an enzyme reaction. These are, for example, nucleotide-binding experiments using intrinsic or extrinsic fluorescence, quench-flow experiments to measure ATP/GTP cleavage and biosensors for other products, e.g., ADP/GDP [11–13, 16].

It is difficult do give standard instructions for the design and analysis of transient kinetic experiments, since it depends very much on the mechanism and properties of the system under study. Here, we describe an example of measurement of the P_i release kinetics of SufBC and interpretation of the data, also including previously published data on nucleotide binding and hydrolysis.

3.2.2 Example Protocol: P_i Release Under Single-Turnover Conditions—SufBC ATPase

SufC is an ABC-type ATPase involved in iron sulfur cluster assembly under oxidative stress or Fe starvation in bacteria (including human pathogens like *Staphylococcus aureus*, *Mycobacterium tuberculosis*), archaea, and the plastid organelle of algae and plants [53].

The kinetic mechanism of SufC ATPase alone and the complex with its regulatory protein SufB has been studied using tryptophan fluorescence measurements and fluorescently labeled ATP and ADP analogues, mant-ATP, and mant-ADP [54, 55]. For the SufBC complex, rate constants were determined for mant-ATP and ATP binding (k_{+1} = 0.55 μM^{-1} s^{-1} and 0.13 μM^{-1} s^{-1}), mant-ATP cleavage (k_{+2} = 0.088 s^{-1}), mant-ADP and ADP binding (k_{-4} = 0.91 μM^{-1} s^{-1} and 0.27 μM^{-1} s^{-1}), and mant-ADP dissociation (k_{+4} = 0.038 s^{-1}), suggesting that ATP cleavage and ADP dissociation are rate-limiting [55]. This was confirmed in a later study, where P_i release and ADP release were measured directly using MDCC-PBP and an ADP biosensor (MDCC-ParM), respectively [11]. A detailed protocol for the single-turnover P_i release measurements is described here.

Buffer and Concentrations

Experiments are performed in buffer containing 50 mM Tris–HCl pH 7.6, 100 mM KCl, 5 mM $MgCl_2$, and 2 mM DTT (using components with minimal P_i, *see* Subheading 2.2) at 20 °C. The buffer is prepared freshly on the day and kept at room temperature throughout the experiments (*see* **Note 15**). The final concentrations of assay components are 2 μM ATP, 10 μM SufBC, and 10 μM MDCC-PBP.

In the stopped-flow experiments, the two reactants (ATP and SufBC) are mixed in a 1:1 ratio, so all solutions are prepared at 2× the final concentration (*see* **Note 16**). The stop syringe is set to 100 μl, so the experiments need 50 μl volume per solution per push plus 150 μl for priming.

Stopped-Flow Setup and Cleaning

1. Switch on the lamp at least 20 min before the first measurement to allow for signal stabilization.
2. Switch on all the other parts of the stopped-flow instrument.
3. Set the water bath to 20 °C.
4. Put a 455 nm long-pass filter in front of the photomultiplier and set the entrance and exit slits of the monochromator to 2.5 mm (*see* **Note 17**).
5. The stop syringe is set to 100 μl.

   In order to remove any contaminating phosphate from the stopped-flow instrument, it is cleaned out with P_i mop solution (*see* Subheading 2.5) prior to the experiments.
6. Prepare ~10 ml solution of 200 μM 7-MEG and 1 U/ml PNPase in P_i-free buffer.

7. Fill the whole stopped-flow system (drive syringes, mixer and cuvette, stop syringe, following the instrument manual) with the P_i mop solution and incubate for 20 min.
8. Clean out the system thoroughly with P_i-free water.

P_i Calibration

1. Prepare 1.8 ml of 2× MDCC-PBP (20 μM) solution.
2. Prepare a concentration series of 2× P_i standard solutions: 1 ml of 1, 2, 3, and 4 μM P_i.
3. Rinse drive syringes, lines, cuvette, and stop syringe with buffer.
4. Load the two drive syringes (here C and D) with MDCC-PBP solution and buffer, respectively, and prime the system (three pushes with both syringes to completely exchange the solutions in the two lines and the cuvette).
5. Set the excitation wavelength to 436 nm.
6. Set the photomultiplier voltage, so that the signal is about 10–25% of the maximum to allow detection of an increase in fluorescence in the presence of P_i (around 250–300 V in our stopped-flow instrument) (*see* **Note 17**). Test if you are on the 436 nm Hg line by moving the monochromator wavelength in 1 nm steps in both directions and see where the signal is maximal. Use this wavelength and readjust photomultiplier voltage.
7. Record at least three traces at 0 P_i, acquiring data for 10 s.
8. Exchange the buffer in syringe D for 1 μM P_i solution and prime: In order to save MDCC-PBP, instead of priming by pushing both syringes, the PBP syringe (C) is blocked off (closed valve) and 150 μl of P_i solution from the syringe (D) is pushed through the cuvette manually (in flush mode).
9. Record at least three traces, acquiring data for 10 s. Repeat **steps 8** and **9** for all other P_i solutions.

   To analyze data:

10. Calculate the average fluorescence signal over time for all P_i concentrations leaving out the initial time points where the signal is not stable (similar to the steady state experiments in Fig. 3a).
11. Plot the averaged fluorescence (F) versus P_i concentration to generate the standard curve. Fit a standard curve by linear regression (*see* Fig. 3b). The slope gives the fluorescence units per μM P_i ($\Delta F/\Delta\,[P_i]$).
12. P_i calibration data are not strictly needed for the analysis of single-turnover data, but are necessary for the interpretation of multi-turnover experiments (*see* **Note 18**).

P_i Release Kinetics

1. Prepare 600–800 μl solution containing 20 μM SufBC and 20 μM MDCC-PBP.
2. Prepare 600–800 μl 4 μM ATP.
3. Flush the stopped-flow system thoroughly with buffer, before loading the reaction solutions.
4. Load the drive syringes with SufBC/MDCC-PBP and ATP and prime with three pushes of both syringes.
5. Keep all the settings (excitation wavelength, slits, PM voltage) as in the calibration.
6. Record traces on different time scales (e.g., 1, 10, 100 s) to find the optimal time range for the measurement.
7. Record at least three traces in the appropriate time scales (here 10 and 100 s).

To convert the fluorescence traces into P_i concentration changes:

8. Subtract the minimum fluorescence value from all other values in a trace to set the initial fluorescence to zero.

The P_i concentration is calculated either by using the result from the calibration curve (**step 9**) or by setting the end value of the P_i release trace to the used ATP concentration (**step 10**) (*see* **Note 18**).

9. Divide all offset corrected data points by the slope of the calibration curve.

   or

10. Divide all offset corrected data points by the end point of the trace (average of the last few data points) and multiply by the total ATP used (here 2 μM).
11. At the beginning of the trace, an artificial increase or decrease of the fluorescence signal is often observed due to re-equilibration of P_i binding to PBP due to different concentrations of contaminating P_i in the solutions. These data points are deleted before data analysis.
12. While calibration is necessary for the interpretation and analysis of multi-turnover data, for single-turnover experiments conversion of fluorescence data into P_i concentration is not required, but rate constants can directly be obtained from curve fitting to the raw data (arbitrary fluorescence units).

Data Interpretation and Analysis

P_i release kinetics during a single turnover of ATP hydrolysis by SufBC are shown in Fig. 7. The increase in $[P_i]$ was analyzed using a single-exponential curve (*see* Eq. 6), which gives a rate constant of 0.11 s^{-1} (*see* **Note 19**). A very similar rate constant has been previously determined for the cleavage of a fluorescent analogue, mant-ATP ($k_{+2}^{mATP} = 0.088\ s^{-1}$) using the quench-flow technique

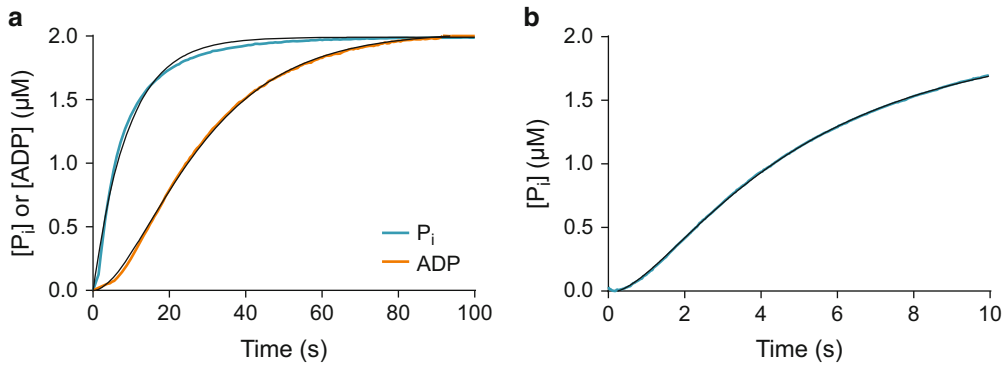


Fig. 7 P_i and ADP release in a single turnover of ATP hydrolysis by SufBC. (**a**) Time course of P_i release (blue line) and ADP release (orange line) after mixing 2 μM ATP and 10 μM SufBC in the presence of 10 μM MDCC-PBP and 20 μM MDCC-ParM (ADP biosensor), respectively. P_i release kinetics were analyzed by single-exponential curve fitting (black line) (*see* Eq. 6), giving a rate constant, $k = 0.11\ s^{-1}$, which corresponds to ATP cleavage (k_{+2}) (as described in main text). Curve fitting of ADP release data is performed using Eq. 7. The rate constants correspond to ATP cleavage, $k_{+2} = 0.093\ s^{-1}$ and ADP release, $k_{+4} = 0.053\ s^{-1}$. (**b**) P_i release data (blue line) on a shorter time scale with a double-exponential fit (black) using Eq. 7. The lag phase, characterized by the rate constant $k = 1.1\ s^{-1}$, most likely reflects ATP binding under these conditions (see main text)

[55]. This suggests that P_i release reflects the kinetics of ATP cleavage ($k_{+2} = 0.11\ s^{-1}$) and the subsequent P_i release is much faster and not rate-limiting ($k_{+3} \gg k_{+2}$). In addition to P_i release, ADP release kinetics was also measured using a similar type of biosensor based on an ADP-binding protein, MDCC-ParM [11]. ADP release kinetics was analyzed using Eq. 7. They show a pronounced lag with a rate constant ($0.093\ s^{-1}$) that fits very well with the ATP cleavage rate obtained from P_i release data ($k_{+2} = 0.11\ s^{-1}$). The following ADP release is about twofold slower ($k_{+4} = 0.053\ s^{-1}$). Slow ADP release is in-line with previous data, where dissociation of mant-ADP has been measured by displacement with unlabeled ADP ($k_{+4}{}^{mADP} = 0.038\ s^{-1}$) [55].

A more careful look at the P_i release on a shorter time scale (*see* Fig. 7b) shows a lag before the increase in [P_i]. The rate constant of the lag phase, $1.1\ s^{-1}$, obtained from fitting using Eq. 7, can be attributed to the ATP binding, which, from previous data, is expected to occur with an observed rate constant of about $1.3\ s^{-1}$ (calculated from (SufBC) $\times\ k_{+1} = 10\ \mu M \times 0.13\ \mu M^{-1}\ s^{-1}$, assuming k_{-1} is close to zero [55]).

3.3 Extensions and Modification of the Assay

The P_i-detection assay descsribed above can be adapted to measure the kinetics of enzymes that generate a phosphorylated product (including polyphosphates), rather than P_i itself. In this case, a secondary enzyme must be added, which converts the phosphorylated product into P_i for detection by the P_i biosensor.

For example, we previously developed an MDCC-PBP–based assay to study the triphosphohydrolase activity of the HIV-1

restriction factor SAMHD1 [56]. By coupling the reaction with yeast exopolyphosphatase (PPX1), the primary product, triphosphate, is converted into pyrophosphate (PP_i) and P_i, which is then detected by MDCC-PBP. Similarly, the combination of inorganic pyrophosphatase (PPase) and MDCC-PBP can be applied to measure the formation of PP_i in a number of enzymatic reactions. This method has been used, for example, to measure PP_i release from protein prenyltransferases [57] or DNA and RNA polymerases [58–61]. Finally, real-time kinetics of nuclease reactions were measured using MDCC-PBP in combination with alkaline phosphatase or T4 polynucleotide kinase as secondary enzyme, which cleave off the terminal phosphate from the newly generated nucleic acid products [62].

The key to successfully applying these coupled-enzyme assays is that the reaction of the secondary enzyme is much faster than the reaction under study and therefore is not rate-limiting. This is particularly critical under pre-steady state conditions and requires careful experimental design and controls, see for example [59].

3.4 Summary

Fluorescent biosensors for phosphate, MDCC-PBP, and Rho-PBP are very useful tools to measure the kinetics of nucleotide hydrolysis by enzymes, such as ATPases and GTPases. The assay principle is simple and is based on direct detection of the product, so apart from product depletion, the reaction being studied is not modified. The biosensors allow real-time measurements of P_i release at high time resolution, down to milliseconds. Other P_i detecting systems are much more limited, having response times of seconds (coupled enzyme assays) or even minutes (malachite green), so the latter can only be used in a discontinuous format.

The ability to measure P_i release in real time has greatly added to our understanding of the mechanism of motility proteins like myosin, kinesin, or DNA helicases. In addition, due to the simplicity, the P_i biosensor assays have also found application in screening for inhibitors. It will be interesting to see how the usage of phosphate biosensors will be extended in the future, for example, in the fields of single-molecule enzymology or drug discovery.

4 Notes

1. In this chapter, all assays are described using MDCC-PBP, which is the original biosensor variant, is easier to prepare, and therefore is most commonly used for steady state and pre-steady state kinetics. However, the tetramethylrhodamine version, Rho-PBP, can be used interchangeably in most assay types, just by adjusting the excitation and detection wavelengths (excitation maximum 553 nm, emission maximum 575 nm [6]).

2. How critical the P_i background is for an assay depends largely on the sensitivity needed. For measurements of several micromolar P_i, and using >15 μM PBP, even 2 μM P_i contamination can be tolerated. Keeping the P_i concentration under 2 μM is easily achieved when following the instructions given here. The use of a P_i mop is generally not necessary. For very sensitive assays where hundreds of nM or even less P_i need to be measured and lower PBP concentrations are used more care has to be taken in choosing and preparing the buffers and nucleotide solutions. It might also be necessary to include the P_i mop (*see* Subheading 2.5) in some of the reagents.
3. Not all PNPase variants accept 7-methylguanosine as a substrate. The given source works well in our experiments.
4. Multi-well plates, in particular those with a nonprotein-binding coating, can also have significant levels of P_i contamination. For the assay we present here this is not a problem. However, if needed one can reduce P_i contamination by rinsing the plates with P_i-free water or even pretreat them with P_i mop.
5. When running assays at near ambient temperature, it is not necessary to take special precautions to minimize temperature equilibration times. Buffers and diluted solutions are just kept at room temperature (not on ice) before pipetting into the plates. However, when measuring at higher temperatures, e.g., 30 °C or 37 °C, solutions (and plates) have to be pre-equilibrated before starting the experiments or temperature will increase during the initial minutes of the reaction. At higher temperatures in nonsealed plates, evaporation can also be a problem, particularly when small volumes and long incubation times are used.
6. All experiments described here were performed with an N- and C-terminally truncated construct of *S. cerevisiae* Chd1 (residues 118-1274) [29]. The protein was expressed from pET49b in *E. coli* and purified by GSH-affinity, Ni-NTA affinity, and size exclusion chromatography.
7. Since the ATPase K_m of Chd1 is relatively high (>100 μM) and therefore high ATP concentrations are needed, we used purified ATP for the experiments.
8. A 30 bp double-stranded DNA was generated by annealing two complementary oligonucleotides with the following sequences: 5′-GTA ACC ACG CGT ATA GAA ACG GGA CTC ATT-3′ and the reverse complement.
9. The assay also works well with down to 20 μl in Corning 384 plates with square wells (not small volume). If enough material is available, the assay can be more robust and accurate with larger volumes (50–200 μl) and/or 96-well plates.

10. In this example, using the same data range for linear regression analysis for all ATP concentrations works well. However, sometimes it is necessary to use different ranges at high and low concentration. For example, if there is larger noise on the data or some artifacts at the beginning of the reaction like a lag, it can be better to fit the low concentration data over a longer time range. The MARS analysis software also has a "maximum of slope" function, which can be used to find the initial rates. In this case, it is important to choose the right width for slope calculation (at least 10 data points, better 20 or more).
11. P_i binding to MDCC-PBP occurs in two steps: initial, loose binding, followed by a conformation change that closes the P_i binding cleft [5]. The kinetics of P_i binding show a hyperbolic dependence of the observed rate constant of binding on the concentration of the excess compound (P_i or MDCC-PBP). The rate constant of cleft closure (reached at high concentrations) will therefore pose an upper limit on the rate constant of P_i release that can be measured. At 5 °C the dissociation constant for the initial loose binding step is 4.5 μM and the rate constants for closing and opening are 317 s^{-1} and 4.5 s^{-1}, respectively [5]. At 20 °C it is more difficult to resolve the rate constant for cleft closure, as it is reaching the limit of time resolution of stopped-flow instruments, but more than 1000 s^{-1} has been measured at concentrations above 50 μM (Martin Webb, personal communication).
12. If the rate of P_i binding to PBP is tenfold higher than the rate constant of the reaction under study, the error in the measured rate constant is less than 10%, even if the short lag phase due to the P_i binding rate is ignored in the data fitting. However, P_i release rates only two-threefold slower than P_i binding can also be measured, if data are analyzed more carefully with kinetic modeling software, including the rate of P_i binding to PBP.
13. It is ideal to have the cleavage (**step 2**) and/or P_i release (**step 3**) kinetically isolated from the binding since the traces can be analyzed by single- or multi-exponential curve fitting to extract the rate constants for these steps. In practice, it is often difficult to reach high enough enzyme concentrations for a clear separation of the binding step. In this case, the true rate constants will be underestimated in simple-exponential fits. For more accurate rate constants, data can be analyzed using kinetic simulation software, e.g., Dynafit or KinTek Explorer.
14. To simplify the mechanism, we assume here that chemical cleavage and P_i release are both irreversible. Regarding P_i release, this is the case for most enzymes, since P_i affinity is generally very low. In addition, free P_i is trapped by the binding to MDCC-PBP making the reaction quasi-irreversible. In

contrast, chemical cleavage is often reversible (e.g., for myosin [63, 64]), and this should be considered when interpreting transient kinetic data of a novel enzyme.

15. Using cold solutions can lead to the formation of gas bubbles when warmed up in the stopped-flow instrument (in particular at temperatures above 25 °C), which causes disturbances in the signal. For this reason, it is best to prepare all solutions with buffer at room temperature and not keep premade solutions on ice.
16. Ideally, MDCC-PBP is included in both solutions, enzyme and ATP, at the final concentration to minimize artifacts, e.g., a small increase or decrease in fluorescence at the beginning of the reaction caused by re-equilibration of P_i binding due to different P_i concentrations in the two solutions. However, to save material, for example, in the P_i calibration or when an ATP concentration series is used, it is often preferable to use MDCC-PBP only in the enzyme solution.
17. The slit width and photomultiplier voltage setting depend on the instrument, lamp age, and concentration of PBP used.
18. Using the calibration data to calculate the P_i concentration changes can sometimes deviate significantly from the expected concentration of released P_i that is the initial nucleotide concentration. Therefore, if there is a defined end value, like in single-turnover kinetics, the final P_i concentration is often set to the ATP concentration used, assuming that all ATP has been converted to free P_i at the end of the reaction. This is a valid assumption for most systems. If the reaction is not observed until the complete substrate turnover, e.g., in multi-turnover experiments, one has to rely on the calibration data. One problem with calibration in the stopped-flow is that absolute fluorescence values (as opposed to amplitudes of fluorescence changes) are often not very reproducible and due to the fast binding of P_i to MDCC-PBP only end values can be analyzed. One way around this problem is to measure the calibration by mixing MDCC-PBP plus different [P_i] with the P_i mop and follow the slow kinetics of P_i removal [48]. Other sources of errors include errors in nucleotide concentration, concentration of the P_i solutions, or the fact that the calibration is normally not measured in exactly the same conditions (in the presence of enzyme and ATP) as the actual P_i release trace.
19. It has to be noted that the P_i release data in Fig. 7a are described better with a double-exponential fit than a single exponential, suggesting that the mechanism is more complex, for example, including a conformation change after ATP binding [54, 55].

Acknowledgments

Expert advice provided by Dr. Martin Webb (The Francis Crick Institute) in the preparation of this chapter is gratefully acknowledged. I would also like to thank Dr. Elizabeth Morris (The Francis Crick Institute) and Dr. Martin Webb for comments on the manuscript.
This work was supported by the Francis Crick Institute which receives its core funding from Cancer Research UK (FC001999), the UK Medical Research Council (FC001999), and the Wellcome Trust (FC001999).

References

1. Brune M, Hunter JL, Corrie JE, Webb MR (1994) Direct, real-time measurement of rapid inorganic phosphate release using a novel fluorescent probe and its application to actomyosin subfragment 1 ATPase. Biochemistry 33(27):8262–8271
2. He ZH, Chillingworth RK, Brune M, Corrie JE, Trentham DR, Webb MR, Ferenczi MA (1997) ATPase kinetics on activation of rabbit and frog permeabilized isometric muscle fibres: a real time phosphate assay. J Physiol 501 (1):125–148
3. He ZH, Chillingworth RK, Brune M, Corrie JET, Webb MR, Ferenczi MA (1999) The efficiency of contraction in rabbit skeletal muscle fibres, determined from the rate of release of inorganic phosphate. J Physiol 517 (3):839–854
4. Hirshberg M, Henrick K, Haire LL, Vasisht N, Brune M, Corrie JE, Webb MR (1998) Crystal structure of phosphate binding protein labeled with a coumarin fluorophore, a probe for inorganic phosphate. Biochemistry 37 (29):10381–10385
5. Brune M, Hunter JL, Howell SA, Martin SR, Hazlett TL, Corrie JE, Webb MR (1998) Mechanism of inorganic phosphate interaction with phosphate binding protein from *Escherichia coli*. Biochemistry 37(29):10370–10380
6. Okoh MP, Hunter JL, Corrie JE, Webb MR (2006) A biosensor for inorganic phosphate using a rhodamine-labeled phosphate binding protein. Biochemistry 45(49):14764–14771
7. Förster T, König E (1957) Absorptionsspektren und Fluoreszenzeigenschaften konzentrierter Lösungen organischer Farbstoffe. Z Elektrochem 61:344–348
8. Selwyn JE, Steinfeld JI (1972) Aggregation equilibria of xanthene dyes. J Phys Chem 76:762–774
9. Solscheid C, Kunzelmann S, Davis CT, Hunter JL, Nofer A, Webb MR (2015) Development of a Reagentless biosensor for inorganic phosphate, applicable over a wide concentration range. Biochemistry 54(32):5054–5062
10. Gu H, Lalonde S, Okumoto S, Looger LL, Scharff-Poulsen AM, Grossman AR, Kossmann J, Jakobsen I, Frommer WB (2006) A novel analytical method for in vivo phosphate tracking. FEBS Lett 580 (25):5885–5893
11. Kunzelmann S, Webb MR (2009) A biosensor for fluorescent determination of ADP with high time resolution. J Biol Chem 284 (48):33130–33138
12. Kunzelmann S, Webb MR (2010) A fluorescent, reagentless biosensor for ADP based on tetramethylrhodamine-labeled ParM. ACS Chem Biol 5(4):415–425
13. Kunzelmann S, Webb MR (2011) Fluorescence detection of GDP in real time with the reagentless biosensor rhodamine-ParM. Biochem J 440(1):43–49
14. Vancraenenbroeck R, Webb MR (2015) A fluorescent, Reagentless biosensor for ATP, based on Malonyl-coenzyme a Synthetase. ACS Chem Biol 10(11):2650–2657
15. Berg J, Hung YP, Yellen G (2009) A genetically encoded fluorescent reporter of ATP:ADP ratio. Nat Methods 6(2):161–166
16. Brune M, Corrie JE, Webb MR (2001) A fluorescent sensor of the phosphorylation state of nucleoside diphosphate kinase and its use to monitor nucleoside diphosphate concentrations in real time. Biochemistry 40 (16):5087–5094
17. de Lorimier RM, Smith JJ, Dwyer MA, Looger LL, Sali KM, Paavola CD, Rizk SS, Sadigov S, Conrad DW, Loew L et al (2002) Construction

of a fluorescent biosensor family. Protein Sci 11 (11):2655–2675

18. Salins LL, Ware RA, Ensor CM, Daunert S (2001) A novel reagentless sensing system for measuring glucose based on the galactose/glucose-binding protein. Anal Biochem 294 (1):19–26
19. Helassa N, Durst CD, Coates C, Kerruth S, Arif U, Schulze C, Wiegert JS, Geeves M, Oertner TG, Torok K (2018) Ultrafast glutamate sensors resolve high-frequency release at Schaffer collateral synapses. Proc Natl Acad Sci U S A 115(21):5594–5599
20. Dwyer MA, Hellinga HW (2004) Periplasmic binding proteins: a versatile superfamily for protein engineering. Curr Opin Struct Biol 14 (4):495–504
21. Nixon AE, Hunter JL, Bonifacio G, Eccleston JF, Webb MR (1998) Purine nucleoside phosphorylase: its use in a spectroscopic assay for inorganic phosphate and for removing inorganic phosphate with the aid of phosphodeoxyribomutase. Anal Biochem 265(2):299–307
22. Webb MR (1992) A continuous spectrophotometric assay for inorganic phosphate and for measuring phosphate release kinetics in biological systems. Proc Natl Acad Sci U S A 89(11):4884–4887
23. Fiske CH, SubbaRow Y (1925) The colorimetric determination of phosphorus. J Biol Chem 66:375–400
24. Johnson KA, Simpson ZB, Blom T (2009) Global kinetic explorer: a new computer program for dynamic simulation and fitting of kinetic data. Anal Biochem 387(1):20–29
25. Kunzelmann S, Praefcke GJ, Herrmann C (2005) Nucleotide binding and self-stimulated GTPase activity of human guanylate-binding protein 1 (hGBP1). Methods Enzymol 404:512–527
26. Lusser A, Urwin DL, Kadonaga JT (2005) Distinct activities of CHD1 and ACF in ATP-dependent chromatin assembly. Nat Struct Mol Biol 12(2):160–166
27. Stockdale C, Flaus A, Ferreira H, Owen-Hughes T (2006) Analysis of nucleosome repositioning by yeast ISWI and Chd1 chromatin remodeling complexes. J Biol Chem 281 (24):16279–16288
28. Hauk G, McKnight JN, Nodelman IM, Bowman GD (2010) The chromodomains of the Chd1 chromatin remodeler regulate DNA access to the ATPase motor. Mol Cell 39 (5):711–723
29. Patel A, McKnight JN, Genzor P, Bowman GD (2011) Identification of residues in chromodomain helicase DNA-binding protein 1 (Chd1) required for coupling ATP hydrolysis to nucleosome sliding. J Biol Chem 286 (51):43984–43993
30. Bickham DC, West TG, Webb MR, Woledge RC, Curtin NA, Ferenczi MA (2011) Millisecond-scale biochemical response to change in strain. Biophys J 101 (10):2445–2454
31. De La Cruz EM, Ostap EM (2009) Kinetic and equilibrium analysis of the myosin ATPase. Methods Enzymol 455:157–192
32. Heeley DH, Belknap B, White HD (2002) Mechanism of regulation of phosphate dissociation from actomyosin-ADP-Pi by thin filament proteins. Proc Natl Acad Sci U S A 99 (26):16731–16736
33. Lionne C, Brune M, Webb MR, Travers F, Barman T (1995) Time resolved measurements show that phosphate release is the rate limiting step on myofibrillar ATPases. FEBS Lett 364 (1):59–62
34. White HD, Belknap B, Webb MR (1997) Kinetics of nucleoside triphosphate cleavage and phosphate release steps by associated rabbit skeletal actomyosin, measured using a novel fluorescent probe for phosphate. Biochemistry 36(39):11828–11836
35. Muretta JM, Rohde JA, Johnsrud DO, Cornea S, Thomas DD (2015) Direct real-time detection of the structural and biochemical events in the myosin power stroke. Proc Natl Acad Sci U S A 112(46):14272–14277
36. Gilbert SP, Webb MR, Brune M, Johnson KA (1995) Pathway of processive ATP hydrolysis by kinesin. Nature 373(6516):671–676
37. Klumpp LM, Hoenger A, Gilbert SP (2004) Kinesin's, second step. Proc Natl Acad Sci U S A 101(10):3444–3449
38. Krzysiak TC, Gilbert SP (2006) Dimeric Eg5 maintains processivity through alternating-site catalysis with rate-limiting ATP hydrolysis. J Biol Chem 281(51):39444–39454
39. Cao W, Coman MM, Ding S, Henn A, Middleton ER, Bradley MJ, Rhoades E, Hackney DD, Pyle AM, De La Cruz EM (2011) Mechanism of Mss116 ATPase reveals functional diversity of DEAD-box proteins. J Mol Biol 409(3):399–414
40. Gyimesi M, Sarlós K, Kovács M (2010) Processive translocation mechanism of the human Bloom's syndrome helicase along single-stranded DNA. Nucleic Acids Res 38 (13):4404–4414
41. Jeong YJ, Kim DE, Patel SS (2002) Kinetic pathway of dTTP hydrolysis by hexameric T7 helicase-primase in the absence of DNA. J Biol Chem 277(46):43778–43784

42. Toseland CP, Martinez-Senac MM, Slatter AF, Webb MR (2009) The ATPase cycle of PcrA helicase and its coupling to translocation on DNA. J Mol Biol 392(4):1020–1032
43. Kunzelmann S, Praefcke GJ, Herrmann C (2006) Transient kinetic investigation of GTP hydrolysis catalyzed by interferon-gamma-induced hGBP1 (human guanylate binding protein 1). J Biol Chem 281 (39):28627–28635
44. Nixon AE, Bruñe M, Lowe PN, Webb MR (1995) Kinetics of inorganic phosphate release during the interaction of p21ras with the GTPase-activating proteins, p 120-GAP and Neurofibromin. Biochemistry 34 (47):15592–15598
45. Goyal A, Belardinelli R, Maracci C, Milón P, Rodnina MV (2015) Directional transition from initiation to elongation in bacterial translation. Nucleic Acids Res 43 (22):10700–10712
46. Savelsbergh A, Katunin VI, Mohr D, Peske F, Rodnina MV, Wintermeyer W (2003) An elongation factor G-induced ribosome rearrangement precedes tRNA-mRNA translocation. Mol Cell 11(6):1517–1523
47. Tomšic J, Vitali LA, Daviter T, Savelsbergh A, Spurio R, Striebeck P, Wintermeyer W, Rodnina MV, Gualerzi CO (2000) Late events of translation initiation in bacteria: a kinetic analysis. EMBO J 19(9):2127–2136
48. Dillingham MS, Wigley DB, Webb MR (2000) Demonstration of unidirectional single-stranded DNA translocation by PcrA helicase: measurement of step size and translocation speed. Biochemistry 39(1):205–212
49. Donmez I, Patel SS (2008) Coupling of DNA unwinding to nucleotide hydrolysis in a ring-shaped helicase. EMBO J 27(12):1718–1726
50. Sarlós K, Gyimesi M, Kovács M (2012) RecQ helicase translocates along single-stranded DNA with a moderate processivity and tight mechanochemical coupling. Proc Natl Acad Sci U S A 109(25):9804–9809
51. Fersht A (1985) Enzyme structure and mechanism. Freeman, New York, p 144
52. Kuzmic P (1996) Program DYNAFIT for the analysis of enzyme kinetic data: application to HIV proteinase. Anal Biochem 237 (2):260–273
53. Takahashi Y, Tokumoto U (2002) A third bacterial system for the assembly of iron-sulfur clusters with homologs in archaea and plastids. J Biol Chem 277(32):28380–28383
54. Eccleston JF, Petrovic A, Davis CT, Rangachari K, Wilson RJ (2006) The kinetic mechanism of the SufC ATPase: the cleavage step is accelerated by SufB. J Biol Chem 281 (13):8371–8378
55. Petrovic A, Davis CT, Rangachari K, Clough B, Wilson RJ, Eccleston JF (2008) Hydrodynamic characterization of the SufBC and SufCD complexes and their interaction with fluorescent adenosine nucleotides. Protein Sci 17 (7):1264–1274
56. Arnold LH, Kunzelmann S, Webb MR, Taylor IA (2015) A continuous enzyme-coupled assay for triphosphohydrolase activity of HIV-1 restriction factor SAMHD1. Antimicrob Agents Chemother 59(1):186–192
57. Pais JE, Bowers KE, Stoddard AK, Fierke CA (2005) A continuous fluorescent assay for protein prenyltransferases measuring diphosphate release. Anal Biochem 345(2):302–311
58. Hanes JW, Johnson KA (2007) A novel mechanism of selectivity against AZT by the human mitochondrial DNA polymerase. Nucleic Acids Res 35(20):6973–6983
59. Hanes JW, Johnson KA (2008) Real-time measurement of pyrophosphate release kinetics. Anal Biochem 372(1):125–127
60. Li A, Gong S, Johnson KA (2016) Rate-limiting pyrophosphate release by HIV reverse transcriptase improves Fidelity. J Biol Chem 291(51):26554–26565
61. Li J, Johnson KA (2016) Thumb site 2 inhibitors of hepatitis C viral RNA-dependent RNA polymerase allosterically block the transition from initiation to elongation. J Biol Chem 291(19):10067–10077
62. Eriksson J, Langel U (2016) Quantitative microplate assay for real-time nuclease kinetics. PLoS One 11(4):e0154099
63. Trentham DR, Eccleston JF, Bagshaw CR (1976) Kinetic analysis of ATPase mechanisms. Q Rev Biophys 9(2):217–281
64. Webb MR, McDonald GG, Trentham DR (1978) Kinetics of oxygen-18 exchange between inorganic phosphate and water catalyzed by myosin subfragment 1, using the 18O shift in 31P NMR. J Biol Chem 253 (9):2908–2911
65. Yao N, Ledvina PS, Choudhary A, Quiocho FA (1996) Modulation of a salt link does not affect binding of phosphate to its specific active transport receptor. Biochemistry 35(7):2079–2085

Part V

Binding Nucleic Acids

Chapter 14

Gel-Based Analysis of Protein–Nucleic Acid Interactions

James A. W. Stowell, Terence T. L. Tang, Maximilian Seidel, and Lori A. Passmore

Abstract

Electrophoretic mobility shift assays (EMSAs) are among the most frequently used and straightforward experiments for studying protein–nucleic acid interactions. EMSAs rely on the principle that protein–nucleic acid complexes have reduced electrophoretic mobility in a native gel matrix compared to free nucleic acid due to their larger size and reduced negative charge. Therefore, bands for the protein–nucleic acid complexes are shifted in a gel and can be distinguished from free nucleic acids. EMSAs remain a popular technique since they do not require specialist equipment and the complexes formed are easily visualized. Furthermore, the technique can be adapted to enable various aspects of protein–nucleic acid interactions to be investigated, including sequence specificity, estimated binding affinity, and binding stoichiometry.

Key words EMSA, Electrophoretic mobility shift assay, Band shift assay, Protein–DNA interactions, Protein–RNA interactions, Affinity measurement

1 Introduction

Protein–nucleic acid interactions coordinate many fundamental cellular processes, including DNA replication, transcription, RNA processing, and translation. The electrophoretic mobility shift assay (EMSA), or gel shift assay, is a straightforward but sensitive method of characterizing protein–nucleic acid interactions. Although EMSAs are typically used for qualitative purposes, they can provide quantitative estimates of dissociation constants (affinity measurements), binding stoichiometry, and sequence and structural specificity [1].

Gel-based detection of protein–nucleic acid complexes was first described for the DNA-binding lactose operon regulatory network [2, 3]. The method anticipates that under electrophoretic conditions, larger protein-bound nucleic acids are retarded within a native gel matrix, whereas unbound nucleic acids have higher electrophoretic mobility [4]. In a vertical electrophoresis apparatus, free nucleic acid is generally found at the bottom of the gel towards

Tina Daviter et al. (eds.), *Protein-Ligand Interactions: Methods and Applications*, Methods in Molecular Biology, vol. 2263, https://doi.org/10.1007/978-1-0716-1197-5_14,

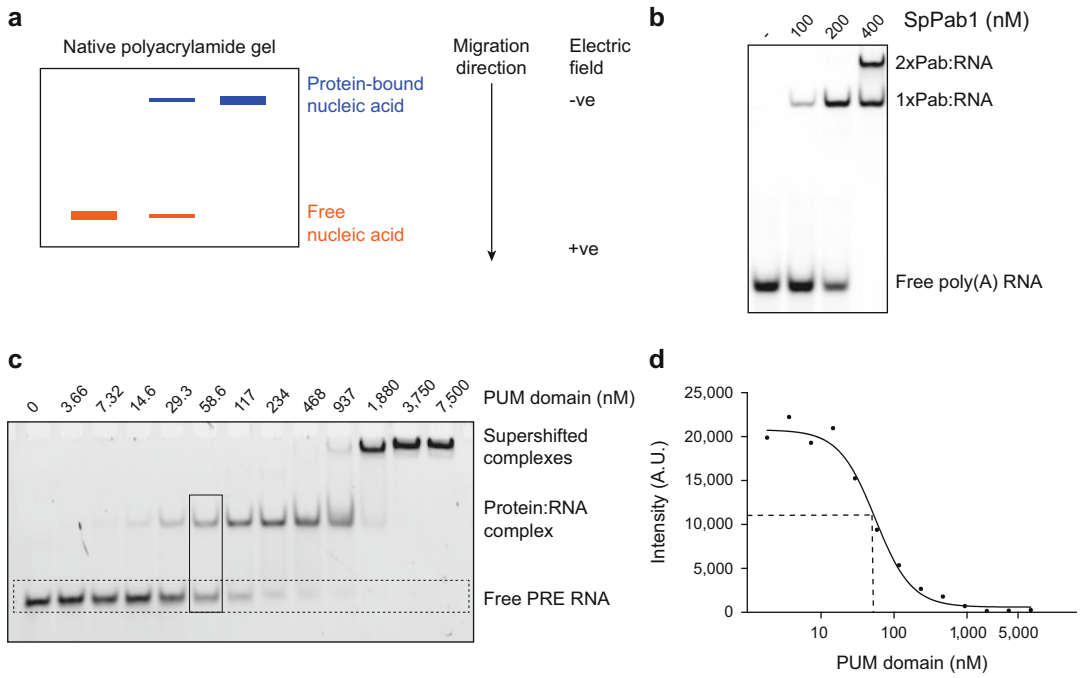


Fig. 1 Overview of electrophoretic mobility shift assays (EMSAs). (**a**) Gel shift assays make use of the fact that nucleic acids migrate towards the anode. Complex formation with binding proteins will lead to a reduction in electrophoretic mobility based on size and charge. (**b**) *Schizosaccharomyces pombe* (Sp) poly(A) binding protein (SpPab1) incubated at the indicated concentrations with 200 nM RNA substrate containing a 5′ 20-mer "upstream" region (CAGCUCCGCAUCCCUUUCCC) followed by a 3′ poly(A) tail of 30 adenosines. Higher order structures with multiple SpPab1 molecules are visualized as "supershifts" [9]. (**c**) Semiquantitative EMSA to estimate the dissociation constant (K_d) for the interaction between the SpPuf3 PUM domain and an RNA substrate containing a Pumilio response element (PRE). Indicated concentrations of protein were incubated with 1 nM 5′ 6-FAM labeled RNA and scanned using a Typhoon imager (GE). Solid box shows lane where roughly half the substrate is bound which equated to the estimated K_d. Supershifted complexes are uncharacterized protein:RNA oligomers or protein:protein interactions. (**d**) Densitomery analysis of the boxed dashed area of the gel in (**c**). *See* Subheading 3 for details

the anode and the protein-bound nucleic acids are found "shifted" towards the top (Fig. 1a).

The results of an EMSA experiment are usually analyzed by detection of the nucleic acid. Thus, the detection limit of the EMSA is defined by the method of readout. Autoradiography of radiolabeled ([^{32}P]) nucleic acids is the most sensitive method of detection, allowing concentrations of 0.1 nM or less to be used [4]. Radiolabeling does not introduce artificial structures which sometimes interfere with binding. On the contrary, fluorescent or chemiluminescent labels or dyes, while less sensitive, provide a safer and more convenient alternative for nucleic acid detection [5–9].

In addition to being a fast and sensitive technique, EMSAs are compatible with a wide range of nucleic acid and protein structures and sizes. Furthermore, EMSAs have been widely adapted and

combined with other techniques such as Western blotting [10] and high-throughput sequencing [11] to extract additional information. Nonetheless, EMSAs have some limitations as complex formation is not in true chemical equilibrium. Furthermore, EMSAs do not provide information regarding the binding site on either the protein or nucleic acid. Thus, observations from EMSAs are most often verified using other complementary techniques, such as fluorescence polarization (Chapter 10), microscale thermophoresis (Chapter 6), surface plasmon resonance (Chapter 17), or structural methods.

Here, we outline the materials and steps required for EMSAs. These include preparation of the native polyacrylamide gel, preparation of protein–nucleic acid samples, electrophoresis, and imaging. We further discuss several common adaptations of EMSAs, including semiquantitative estimation of affinity, supershift assays, and competitive EMSAs.

2 Materials

RNase and DNase contamination should be avoided. Therefore, it is important to work on clean benches and to protect the samples from contamination by wearing gloves at all times (*see* **Note 1**). Furthermore, all solutions should be made up and diluted using fresh ultrapure water (MilliQ) with analytical grade reagents. Solutions can also be treated with diethylpyrocarbonate (DEPC) and autoclaved to avoid nuclease contamination (*see* Subheading 2.2). Filter-sterilize all solutions with a 0.22 μm filter to remove precipitates or particulate contaminants. Take special precautions (according to local safety procedures) when handling toxic or radioactive materials.

2.1 Native Polyacrylamide Gel

1. 10× TBE (Tris-Borate EDTA) stock: Weigh 108 g tris (hydroxymethyl)aminomethane (Tris) base, 9.3 g ethylenediaminetetraacetic acid (EDTA), and 55 g boric acid. Dissolve in water and top up to 1 l. Autoclave the resulting solution (121 °C, 15 min). The solution can be stored at room temperature indefinitely, but should be discarded if there is any visible precipitation.
2. Isopropanol: Supplied at ≥99.9% purity (HPLC grade).
3. Native polyacrylamide gel mix: 8% gel stock (*see* **Note 2**). Mix 50 ml of 10× TBE and 100 ml of 40% (w/v) 19:1 acrylamide: bis-acrylamide solution (preferably gas-stabilized, *see* **Note 3**). Make up to 500 ml with MilliQ water and degas by vacuum filtration. This solution can be made fresh or stored as an unpolymerized stock at 4 °C for several months. We routinely make an unpolymerized gel stock solution when performing multiple experiments, using the desired volume when required.

4. Ammonium persulfate (APS): 10% (w/v) solution. Weigh 5 g ammonium persulfate. Dissolve in MilliQ water and top up to 50 ml. Freeze aliquots at −20 °C (*see* **Note 4**).
5. *N*,*N*,*N*,*N*′-Tetramethyl-ethylenediamine (TEMED): Supplied as a liquid. Can be stored at room temperature, or at 4 °C to reduce vapor.
6. 1× TBE running buffer: Dilute the 10× TBE stock tenfold in MilliQ water.
7. 10× EMSA loading dye: 0.025% (w/v) Orange G and 20% (v/v) glycerol (*see* **Note 5**).
8. Electrophoresis chamber: Use a vertical electrophoresis apparatus with corresponding glass plates, spacers, and well-forming combs. Clamps and/or a gel-casting stand are required for gel preparation. This protocol uses the Mini-PROTEAN Tetra cell with 1.0 mm thick backing plates and combs (Bio-Rad) (*see* **Note 6**).
9. Silanization solution: 5% (v/v) dimethyldichlorosilane (≥99.5%; Sigma-Aldrich catalog no.: 440272) or chlorotrimethylsilane (≥99.5%; Sigma-Aldrich catalog no.: 386529) in heptane if silanization of glassware is carried out (*see* **Note 7**).
10. Power supply: Minimum 100 V, 25 mA capacity.
11. Tapered or round gel-loading pipette tips are useful but not essential.

2.2 Protein and Nucleic Acid Preparation

The exact requirements for reagents will depend on the properties of the proteins and nucleic acids to be studied. Optimal conditions for each EMSA experiment must therefore be determined for every study. Here, we focus on the general requirements of each sample and the "standard" conditions that we have outlined in the methods.

1. Protein: The protein sample should be highly pure (ideally >95% as assayed by SDS-PAGE) and in a buffer with conditions where the protein is known to be stable (usually around pH 7–8). The protein should either be freshly purified or flash frozen and stored at −80 °C (*see* **Note 8**).
2. TE buffer: 10 mM Tris pH 8.0, 0.5 mM EDTA. For long-term storage of nucleic acids, *see* **Note 9**.
3. DEPC-treated water: Treat MilliQ water with 0.1% (v/v) DEPC for at least 2 h at 37 °C. Autoclave the resulting solution (121 °C, 15 min) to inactivate trace DEPC.
4. DNA or RNA preparation: DNA samples should be highly purified and stored in either MilliQ water or TE buffer (*see* **Note 9**). Because of the relatively low cost and commercial availability, we use chemically synthesized DNA. RNA samples

should be highly purified. RNAs can be prepared either by in vitro transcription of a DNA template [12] or by commercial chemical synthesis. In vitro transcribed RNAs should be purified by denaturing polyacrylamide gel electrophoresis and by gel extraction, either by crush and soak [13] or electroelution [14] after excising the correct band.

5. 10× EMSA buffer: 20 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) pH 7.5, 100 mM sodium chloride (NaCl), 2 mM magnesium acetate (MgAc), 0.1 mM tris(2-carboxyethyl)phosphine (TCEP). The composition of the EMSA buffer will depend on the experiment and nature of the interaction (*see* **Note 8**).

2.3 Detection Methods

Additional materials depend on the method of detection.

1. Fluorescent stain: SYBR Safe (DNA, Thermo Fisher) or SYBR Green II (RNA, Thermo Fisher) (*see* **Note 10**), visible with a blue light or UV transilluminator.
2. Fluorescent label: Fluorescent labels can be incorporated during chemical synthesis of DNA or RNA. Alternatively, nucleic acids can be fluorescently labeled with a homemade or commercially available kit (*see* **Notes 11**, and **12**). An example is the 5′- or 3′-end EndTag labeling kit (Vector Laboratories) used with a conjugable fluorophore (e.g., fluorescein maleimide, Vector Laboratories catalog number SP-1502-12).
3. Radioactive label: γ-^{32}P dNTP/γ-^{32}P rATP for end labeling, α-^{32}P NTPs for internal labeling. Radioactive reagents should be handled according to local safety procedures. The nucleic acid can be radioactively end-labeled by a relatively inexpensive protocol using T4 polynucleotide kinase [15].
4. Gel dryer (Bio-Rad) or similar platform.
5. Amersham Typhoon FLA Imager (GE) or similar platform with multiple lasers/filters (*see* **Note 13**) if using fluorescence or phosphorimaging and quantification.
6. Coomassie blue stain (or equivalent) if visualization of protein bands is required.

3 Methods

Carry out all procedures at room temperature unless otherwise specified.

3.1 Preparation of Polyacrylamide Gel

1. Ensure that the glass plates and combs are clean. If not, clean, rinse with 70% (v/v) ethanol and allow components to thoroughly dry before continuing. For larger gel-casting systems, it

may be helpful to silanize the glass surface to allow easier removal of gels from the plates. To do so, wipe the silanization solution with lint-free paper over clean and dry glass plates and leave the solution to dry in a fume hood (*see* **Note 7**).

2. Assemble the gel caster according to manufacturer's specifications (*see* **Notes 6** and **14**).
3. For a mini-gel, add 35 μl 10% APS and 3.5 μl TEMED to 7 ml of native polyacrylamide gel mix (*see* **Note 3**) in a sterile 15 ml conical bottom centrifuge tube. Mix well by gentle inversion. For more than one gel or larger gels, reagent volumes can be scaled up correspondingly.
4. Slowly pour the resulting mixture into the assembled gel-casting apparatus. Immediately insert a comb between the glass plates, taking care not to introduce any air bubbles (*see* **Notes 15** and **16**).
5. Leave the gels to polymerize at room temperature for at least 2 h and at most overnight (*see* **Note 17**). Once polymerized, gels can be kept up for up to 1 week at 4 °C covered with paper towels soaked in 1× TBE buffer and sealed in plastic wrap.

3.2 Sample Preparation

1. In a typical experiment, the concentration of nucleic acid is kept constant and the concentration of protein is varied (*see* **Notes 18** and **19**). It is important to include a negative control with no protein. A positive control with a protein which is known to bind the nucleic acid can also be included.
2. Prepare 10× protein stocks in a dilution buffer (*see* **Notes 8** and **20–22**) in 1.5 ml microcentrifuge tubes. This can be a dilution series (*see* Subheading 3.5) or a smaller range of concentrations (*see* Subheadings 3.6 and 3.7).
3. Prepare a master mix of nucleic acid, EMSA buffer, and loading dye in a 1.5 ml microcentrifuge tube (*see* **Note 23**). For example, for nine 10 μl binding reactions, assemble a master mix for 10 reactions (to account for dead volume when pipetting) as follows: 10 μl 10× RNA stock, 10 μl 10× EMSA buffer, and 10 μl loading dye in 90 μl total volume. This can be scaled accordingly.
4. In 0.2 or 0.5 ml tubes, pipette 1 μl of the 10× protein stock, followed by 9 μl of the master mix. Mix by gently pipetting up and down.
5. Incubate the sample at room temperature for at least 1 h to allow the interaction to reach equilibrium (*see* **Note 24**).

3.3 Polyacrylamide Gel Electrophoresis

1. Rinse the glass plates containing the completely polymerized polyacrylamide gel with MilliQ water to remove gel debris.
2. If the gel is to be run at 4 °C, prechill 1× TBE buffer (*see* **Note 25**).
3. In a suitable electrophoretic chamber, assemble the gel plates and add 1× TBE buffer. The electrophoretic apparatus should be filled such that the upper chamber is full and the bottom of the glass plate is submerged in the bottom chamber. Reassemble the electrophoretic apparatus if any leaks occur at this stage. Buffer in the bottom chamber can be filled to cover most of the gel to act as a heat sink, minimizing gel heating. Remove the gel comb and any debris from the wells.
4. If desired, the gel can be pre-run at 100 V for at least 30 min (*see* **Note 26**).
5. Carefully add the samples to the center of the bottom of the wells of the gel (*see* **Note 27**). The loaded tip should be near the bottom of the well, and the sample should be slowly expelled, without introducing any air bubbles.
6. Electrophorese the sample at 100 V at the desired temperature. Ensure that the electrophoretic chamber and buffer do not exceed the desired temperature. The negatively charged nucleic acid and any stably bound proteins will move through the gel towards the anode (*see* **Note 28**).
7. When the Orange G dye front reaches the bottom of the gel plate (~60 min) (*see* **Note 29**), turn off the power supply.

3.4 Gel Imaging

1. Disassemble the gel apparatus. Carefully pry the gel plates apart using either a metal spatula or a plastic wedge. The gel should remain on one of the two glass plates.
2. Rinse the gel with deionized water.
3. How the gel is handled at this stage will depend on the detection method to be used (*see* **Note 30**). For colorimetric or fluorescent stains, the gel should be stained. When the nucleic acid has been directly labeled by fluorescent dyes or radioactive isotopes, no staining is required. We use fluorescent labels or fluorescent stains due to their reasonably high sensitivity.
4. If using radiolabeled nucleic acids: The resulting gels must be exposed to a phosphorimaging screen. If the sample is sufficiently radioactive and the detection method sufficiently sensitive, the gel can be covered with plastic wrap and imaged directly on the phosphorimaging screen. Alternatively, the gel can be dried prior to imaging. To do so, the gel is sandwiched between a sheet of filter paper and plastic wrap and dried in the gel dryer, with the plastic wrap facing upwards. Drying should be complete to avoid cracking of the gel. This process typically takes 2 h to complete.

5. If using fluorescence stains: Dissolve or dilute the stain in 1× TBE according to manufacturers' instructions. For SYBR Green II (Invitrogen), for each gel, 2 μl of the stain is diluted 10,000× in 20 ml 1× TBE buffer. Carefully transfer each gel to a separate clean and opaque container containing the stain, and proceed with staining and destaining according to manufacturer's instructions. For SYBR Green II, the gel is transferred to the diluted stain in a clean plastic container and allowed to stain for 10 min on a benchtop orbital shaker. The gel is then washed in 1× TBE for 10 min.
6. Carefully transfer the gel to an appropriate imaging stage and proceed with imaging (*see* **Note 31**) (Fig. 1b). If fluorescent detection methods are used, an appropriate combination of stimulation lasers and emission filters should be selected. For SYBR Safe and 6-FAM, we use an excitation wavelength of 473 nm and a filter of 510 nm for detection of fluorescence emission.
7. If protein staining is required after nucleic acid imaging, the gel can be carefully removed and stained by a protein-specific stain such as Coomassie blue according to manufacturers' instructions. An image of the protein in the gel can be overlaid on the previously obtained image of nucleic acid in image analysis software. If the protein band and shifted nucleic acid band coincide, then the shifted band likely corresponds to a protein–nucleic acid complex.

3.5 Variation: Semiquantitative Estimation of Interaction Affinity

By titrating the protein against a fixed and low concentration of nucleic acid, the affinity of the protein–nucleic acid interaction can be estimated. This is carried out using densitometry of the band corresponding to free nucleic acid. The estimate is valid only if the method of detection is proportional to nucleic acid concentration. It must be noted that the observed dissociation constant is only an estimate as the measurement is not carried out under true solution equilibrium conditions (*see* **Note 32**). We therefore normally use EMSAs to qualitatively assay the differences between different protein constructs/point mutants and nucleic acid substrates.

1. Prepare a twofold dilution series of protein in sample buffer; the maximum concentration of the protein should ideally be ~100× greater than the expected dissociation constant of the interaction. To avoid pipetting errors, choose a suitable volume (~50 μl). Prepare the highest protein concentration in dilution buffer in double this volume (i.e., 100 μl). Prepare a series of tubes with 50 μl of dilution buffer. Serially dilute the protein by taking 50 μl from the tube with the higher concentration, mixing by pipetting up and down thoroughly, and drawing up 50 μl to dilute in the next tube.

2. Prepare a master mix solution (*see* Subheading 3.2). The final concentration of nucleic acid probe should be low enough to avoid ligand depletion (<10% K_d) (*see* **Note 33**) but high enough to have good signal for detection. We typically use at least 1–10 nM fluorescently labeled probe in a 10 μl reaction (*see* **Note 19**).
3. Assemble the reaction, allow it to reach equilibrium (*see* Subheading 3.2; *see* **Note 24**), perform the EMSA, and image the gel.
4. Open the image of the gel in image analysis software such as *ImageJ* [16]. Change the contrast of the gel so that the bands corresponding to free nucleic acid and protein-bound nucleic acid are clearly visible (*see* **Note 34**).
5. Box each lane and measure the intensity of the band corresponding to free nucleic acid by densitometry in the image analysis software (Fig. 1c).
6. In a graphical analysis software such as *GraphPad Prism* (GraphPad Software, Inc.), plot the intensity of the band against protein concentration (Fig. 1d). The axis of protein concentration can be changed to a logarithmic scale to aid visual analysis. The concentration of protein at which half of the nucleic acid is bound corresponds to the estimated dissociation constant (*see* **Note 35**).

3.6 Variation: Supershift of Ternary Protein –Protein–Nucleic Acid Complexes

Using EMSAs, ternary interactions can also be detected. If a prey protein (P2) interacts with a bait protein (P1)-nucleic acid complex, it will cause a further reduction in electrophoretic mobility, termed as "supershift." The supershift assay can be used to demonstrate various aspects of the interaction. For example, the interaction of P2 with the P1–nucleic acid complex can be tested with P2 and nucleic acid alone. If P2 alone does not cause a shift in nucleic acid mobility, then it must interact with P1 only, or at a composite P1–nucleic acid binding site. P2 could alternatively disrupt the P1–nucleic acid interaction (*see* **Note 36**). Moreover, if P2 is an antibody against P1, the identity of the protein P1 can be verified. Finally, the supershift assay can also be used to assess the stoichiometry of a complex. For example, the ability of a protein to multimerize on a nucleic acid substrate can be tested by varying the length of the nucleic acid or the protein concentration.

1. Carry out a titration of P1 against the nucleic acid, as in Subheading 3.5 above. Select a concentration of P1 where the free nucleic acid band has completely disappeared.
2. Keeping the concentration of P1 constant, titrate an increasing concentration of the putative binding protein P2 against the protein–nucleic acid complex. It is important to keep one sample with no protein as the negative control.

Alternatively, if the stoichiometry of only one protein against the nucleic acid substrate is tested, titrate the protein directly against the nucleic acid. Multimerization becomes visible at higher protein concentrations (Fig. 1b).

3. Carry out the EMSA and image the gel as described in Subheadings 3.1–3.4.

3.7 Variation: Dual-Color Competition EMSA

EMSAs can be carried out under competitive conditions with more than one type of nucleic acid. The additional nucleic acid (N2) can be unlabeled, or labeled differently from N1 ("two-color") and can be used to assess the nucleic acid binding specificity of the protein (*see* **Note 37**). In this protocol, we detail a method for a dual-color RNA EMSA [17].

1. This method requires the nucleic acids to be differentially labeled (*see* **Note 37**). We routinely synthesize one substrate with a 5′-FAM "blue" label and the other with a 5′-Alexa 647 "red" label.
2. The protein is titrated against fixed concentrations of two nucleic acids (N1 and N2) (*see* **Note 38**). Perform a series of protein dilutions to sample a particular range of concentrations. In the experiment shown in Fig. 2, a range from 50 nM to 2 μM was used.
3. Prepare a master mix solution of two different RNA concentrations, for example, 10 nM and 100 nM each RNA (20 nM and 200 nM total RNA) (*see* **Note 39**).
4. Assemble 10 μl binding reactions with the protein dilutions and master mix (*see* Subheading 3.2). Incubate the mixtures for at least 1 h to reach equilibrium.
5. Perform the EMSA as before (*see* Subheading 3.3).
6. Scan the gel (*see* Subheading 3.4) using the two nonoverlapping excitation wavelengths and emission filters (Typhoon FLA Scanner, GE). Save each channel as a separate image file.
7. Using Adobe Photoshop or another suitable image-processing software, convert each image to 8-bit grayscale and set false color either using the duotone mode or the channel mixer in RGB. Ensure the resulting image is an RGB image and overlay as a separate, partially transparent layer on the other false color image (Fig. 2).

4 Notes

1. General-use laboratory benches (such as those used for plasmid and protein purifications) should be cleaned using 70% (v/v) ethanol followed by an RNase inactivating solution such as RNaseZAP™ (Invitrogen) to reduce the risk of RNase or

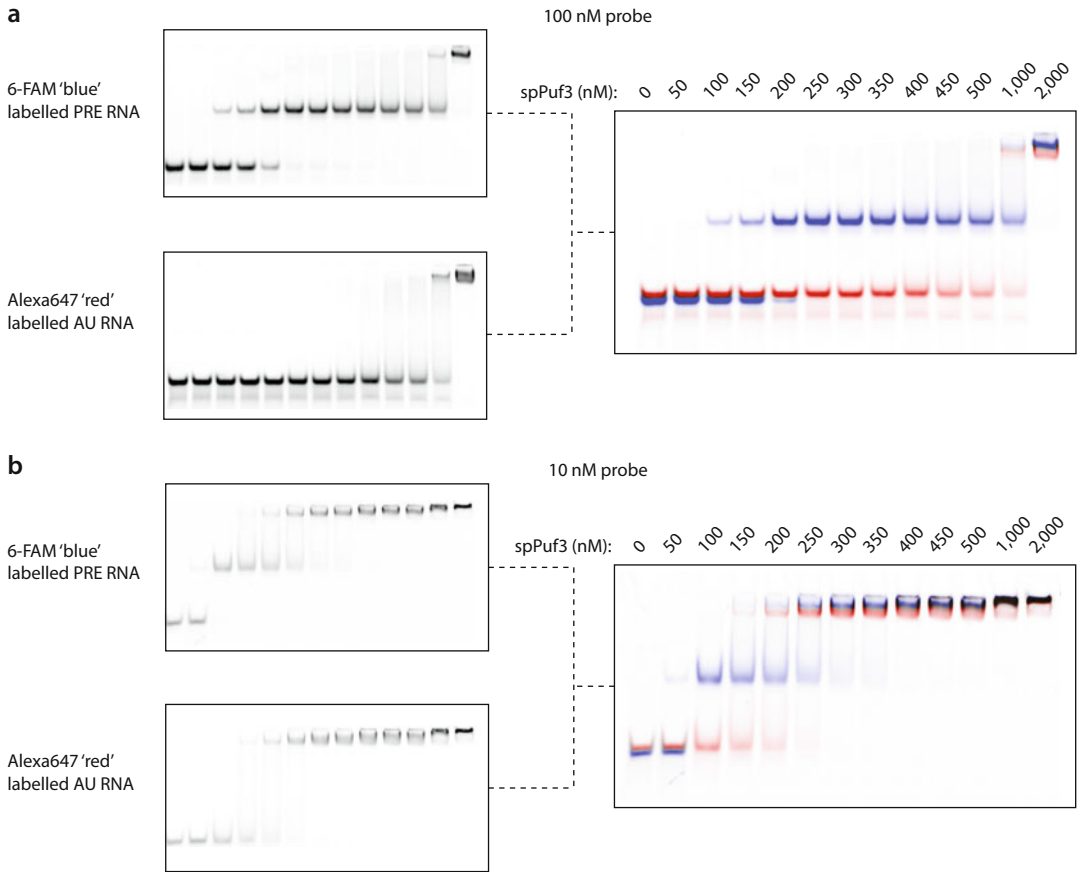


Fig. 2 Dual-color competition EMSA. EMSAs were performed with two differentially labeled substrates (PRE: AACUGUUCCUGUAAAUACGCCAG $[A]_{30}$ or AU: AAUCAUCCUUAUUUAUUACCAUU $[A]_{30}$) to examine substrate specificity. SpPuf3 PUM domain was incubated with 5′ 6-FAM Pumilio response element (PRE) substrate and 5′ Alexa647 AU substrate at the indicated protein concentrations [17]. (**a**) EMSA performed with 100 nM each substrate. Scans at different excitation and emission wavelengths are shown on the left with overlaid false color image on the right. (**b**) The same EMSA performed in (**a**) but with 10 nM each substrate

DNase contamination. It is also possible to bake glassware and tubes to remove RNases but this is not generally necessary.

2. Depending on the required resolution and size of macromolecules, the final acrylamide concentration should be adjusted accordingly. We use 8% polyacrylamide for RNAs of ~10–50 bases. Larger proteins and longer nucleic acids should be resolved on a lower percentage polyacrylamide gel. Smaller substrates may resolve better on higher percentage gels.
3. Unpolymerized acrylamide is a potent neurotoxin and should always be handled with caution. To minimize the risk of inhalation, we use a premade 40% acrylamide:bis-acrylamide (19:1 ratio) solution as a working stock and avoid powdered

acrylamide. Gas-stabilized ultrapure acrylamide:bis-acrylamide solution (such as National Diagnostics AccuGel) is readily available from commercial sources. This also minimizes the risk of degradation products, such as acrylic acid, interfering with electrophoresis.

4. Ammonium persulfate is unstable but can be stored frozen at −20 °C for 6 months. Alternatively, a small volume of APS can be made fresh and kept at 4 °C for up to a week.
5. A colored loading dye allows sample visualization during pipetting and gel loading and increases the density of the sample, causing it to sink to bottom of the wells. We routinely use Orange G since bromophenol blue can migrate at the same position as free nucleic acid in the gel, interfering with visualization. If the dyes nonspecifically bind proteins or nucleic acids, samples can be run with a loading buffer containing only glycerol and without a dye. If glycerol is incompatible with the sample, Ficoll 400 (~1–2% w/v final concentration) can be used as an alternative. In the absence of loading dye, a lane can be kept free for the addition of a dye to monitor migration.
6. The dimensions of gels cast using this system are 8.3 × 7.3 cm. For greater resolving power, larger gels can be used. Many other suitable apparatus options exist, including those that accommodate precast native PAGE gels.
7. Preparation of the silanization solution and application of the solution to glassware should be carried out under a fume hood as fumes from both the silanizing compound and the organic solvent are toxic.
8. The choice of buffer for the interaction analysis is an important consideration. Parameters to consider include: salt concentration (ionic strength), pH (accounting for buffer pK_a), additional metal ions, and additives. Protein–nucleic acid interactions can be dependent on macromolecular charge, which in turn is sensitive to salt concentration and pH. We usually use solutions which approximate physiological salt concentrations and pH. Additional metal ions, such as divalent cations or potassium (for example, in the case of G-quadruplex formation [18]), may be important for mediating protein–nucleic acid interactions. In some cases, EDTA in the gel and running buffer can disrupt interactions—if this is the case, EDTA can be omitted and/or additional magnesium can be included. Additives such as detergents or reducing agents may be important to solubilize macromolecules or maintain them in a near-physiological state. If proteins tend to adhere to surfaces, a low concentration of surfactant can also be added. Finally, a polyanionic additive such as tRNA or heparin can be used (start with 0.01 mg/ml) to reduce nonspecific binding.

9. RNA is less stable than DNA and should not be repeatedly freeze–thawed. To ensure long-term stability, we store our RNA stocks at −80 °C aliquoted in TE buffer.
10. Where possible, non-intercalating fluorescent dyes should be used to label nucleic acids. We use SYBR Safe stain for DNA as studies have demonstrated that SYBR Safe is less mutagenic than ethidium bromide.
11. The site of the fluorescent label can be 5′, 3′ or internal. The choice of labeling site will depend on the characteristics of the nucleic acid; this should be chosen to minimize interference with RNA structure or protein binding.
12. If labeling is carried out after chemical synthesis or in vitro transcription of the nucleic acid, the efficiency of labeling must be assessed. This can be determined by comparing the molarity of the fluorescent label (by absorbance of the fluorophore) to the molarity of the nucleic acid (by absorbance at 260 nm). This is carried out to ensure that there is sufficient fluorescently labeled nucleic acid for later detection.
13. Typhoon FLA Imagers (GE) enable the detection of fluorescently stained, fluorescently labeled, or radioactively labeled (phosphor imaging screen) nucleic acids at high spatial resolution. The compatibility of the Typhoon lasers/emission filters and the selected fluorophore should be confirmed prior to beginning the experiment.
14. Once the gel-casting apparatus has been assembled, it can be checked for leakage. Pour 7 ml of isopropanol between the glass plates and monitor the meniscus level; if the meniscus falls over time, there is a leak and the gel caster should be reassembled. Pour away the isopropanol if there are no leakages. Ensure that the plates are dry before continuing with gel polymerization.
15. The choice of comb depends on spacer width, number of samples, and sample volume. We typically use Bio-Rad Mini-PROTEAN glass plates with a 1 mm spacer and combs with either 10 or 15 wells. These correspond to maximum sample volumes of roughly 26 μl and 44 μl, respectively.
16. Air bubbles introduce undesirable smears in the gel and should be avoided. In our experience, pouring the gel mixture to the top of the apparatus so that there is spill over and then inserting the comb at an angle reduces the chance of bubbles.
17. In our experience, although gels will appear to polymerize within ~10 min, gels run more uniformly and with sharper bands if left for at least an hour, presumably since the cross-linked matrix is more homogenous.

18. Reagents such as protein and nucleic acids are normally added from stocks at 10× concentration to minimize the contribution of buffer carryover to the EMSA reaction. If greater volumes of reagent are added, the 10× EMSA buffer can be adjusted accordingly to account for buffer carryover.
19. The nucleic acid concentration should be optimized depending on detection of the nucleic acid and the experiment type. For fluorescently labeled nucleic acids and detection by a laser scanner such as the Typhoon FLA Scanner (GE), approximately 0.01 pmol fluorophore suffices for detection. Furthermore, the amount of nucleic acid should not exceed the maximum detection limit (for example, pixel saturation) on the detector. Many nucleic acid:protein interactions have K_ds less than 10 nM. In this scenario, fluorescently labeled RNAs are not optimal substrates for EMSAs because of ligand depletion (*see* **Note 33**). Instead, sub-picomolar radiolabeled nucleic acid substrates are preferred.
20. Protein concentration should be accurately determined, for example, by measuring the absorbance at 280 nm (A_{280}) and calculating the protein concentration using a theoretical extinction coefficient. Measurements should be made under denaturing conditions and compared with the native protein to see if there is a substantial difference in the calculated concentration. If possible, a UV spectrum of the sample should be taken. Absorbance in the region beyond which proteins normally absorb (>320 nm) indicates the presence of light scattering due to aggregation or particulates, which would lead to overestimation of the absorbance at 280 nm due to unaggregated protein and hence its concentration. Samples with significant absorbance at 320–340 nm region should be centrifuged to remove any aggregates. If the sample does not contain any tryptophan residues, which are the main source of protein absorbance at 280 nm, consider using a colorimetric assay such as the Bradford method.
21. It is preferable to have a protein stock at a high concentration (>100 μM) as long as the protein does not aggregate. This is because a greater concentration range can be sampled and the A_{280} measurement is more accurate (*see* **Note 20**).
22. Protein dilution buffer composition is sample dependent (*see* **Note 8**). To minimize carryover of buffer components into the binding reaction, our default dilution buffer contains 20 mM HEPES pH 7.5, 100 mM NaCl, and 0.5 mM TCEP.
23. To minimize the adsorption of protein to plastic tubes, we use low-binding plasticware such as protein Lo-Bind tubes (Eppendorf). This also ensures that the concentration of the protein is consistent in the assays.

24. The time required to reach equilibrium will depend on the nature of the protein–nucleic acid interaction and the reaction conditions. Failing to attain equilibrium may lead to misleading results and failure to reproduce observations. At high protein concentrations, the observed rate constant is dominated by the on-rate (k_{on}) [19]. In practice, k_{on} is often 10^5 to 10^6 M^{-1} s^{-1}. At low protein concentrations, however, the observed rate constant is instead dominated by the off-rate (k_{off}) [19]. For a high affinity interaction with a K_d of 1–10 nM, k_{off} can be in the range of 0.001 to 0.0001 s^{-1} and the complex has a half-life ($t_{1/2}$) of ~10 min to >100 min. Since the time taken to reach close to equilibrium is $5 \times t_{1/2}$, it can thus take hours to reach equilibrium with low protein concentrations. For an interaction with a binding affinity of 100 nM, the complex has a $t_{1/2}$ of 1–10 min and equilibrium should be obtained in 5–50 min. It is thus safe to assume that for interactions with affinities in this region, equilibrium will be reached after a 60-min incubation. Also *see* **Note 32**.
25. Running the gel at low temperatures can be beneficial for two reasons. Firstly, heat dissipates more evenly from the gel into the surrounding cold buffer, ameliorating localized heating which causes uneven bands. Secondly, since the dissociation rate of the interaction is a function of temperature, running the gel at ~4 °C will increase the $t_{1/2}$ for the interaction and potentially lead to a better estimate of the K_d (*see* **Notes 24** and **32**).
26. We do not observe major differences between pre-run PAGE gels and loading the samples directly. However, pre-running the gel can remove excess ammonium and persulfate ions and other impurities, such as acrylic acid, that could interfere with complex formation and gel running.
27. We prefer to use tapered "gel-loading" tips to ensure that the sample is evenly distributed along the bottom of the well. Small sample volumes are advantageous as they result in sharper bands at the end of electrophoresis, but the bottom of the well must be evenly covered. For typical gels, this is approximately 5 μl.
28. While the nucleic acid alone will likely move through the polyacrylamide matrix, the mobility of protein (or protein–nucleic acid complex) through the gel will primarily depend on two factors. Firstly, the molecular weight (and shape) of the protein affects how the complex migrates through the gel. The larger the protein, the slower it will move through the polyacrylamide matrix. Secondly, because the assay is carried out near physiological pH, the isoelectric point (pI) and thus net charge of the protein will also affect its mobility. Only proteins

which have a net negative charge under the experimental conditions will migrate through the gel on their own. If the protein is stably bound to nucleic acid, it is more likely to migrate through the gel due to the negative charge of nucleic acids. Since many nucleic acid binding proteins have a net positive charge at physiological pH, they are less likely to run far into gel on their own.

29. The mobility of the Orange G dye front will depend on the percentage of the polyacrylamide gel. In our experience, the Orange G dye front migrates similarly to a ten-nucleotide single-stranded nucleic acid in an 8% polyacrylamide gel. Orange G is often preferred as other commonly used dyes may co-migrate with the nucleic acid and lead to shadows in imaging.
30. Colorimetric stains such as methylene blue or crystal violet require no special equipment, but have low sensitivity (and thus more nucleic acid and protein will be required for the assay) and require a destaining step. Fluorescent stains such as ethidium bromide or SYBR Safe (Invitrogen) are often mutagenic and require UV excitation, but have greater sensitivity and so are often preferred if the nucleic acid is not directly labeled.

    The alternative detection method is to directly label the nucleic acid. Nucleic acids can be chemically synthesized with a fluorescent label such as 6-FAM or Alexa dyes, or unlabeled nucleic acids can be labeled with in-house protocols or commercially available kits. Fluorescent labels are sensitive and their fluorescence is directly proportional to molarity, allowing semi-quantitative analysis of interactions. Alternatively, nucleic acids can be directly labeled with radioisotopes such as ^{32}P. Radioactive labeling is advantageous in that it does not introduce artificial structures that influence binding and is the most sensitive. However, the use of radioactive labeling requires training and precautions for radioactive safety.
31. The fragility of the gel will depend on the percentage of acrylamide used. The lower the acrylamide percentage, the more fragile the gel. To minimize the chance of gel tearing, the tools and glass plates used to manipulate the gels should be kept wet by deionized water at all times. If the gel is too fragile, it can also be stained and directly imaged on the glass plate with a suitable holder, but the background noise from the glass plate itself will likely be higher, potentially hindering further analysis.
32. Even if equilibrium has been attained in the binding reaction, this will be perturbed when applied to an electric field (i.e., during electrophoresis). In our hands, EMSAs often lead to an underestimation of the true affinity, since complexes can

dissociate during electrophoresis. This will depend on the particular complex under study since there are examples of EMSA experiments that agree well with both equilibrium binding experiments and kinetic measurements. Thus, a true equilibrium binding experiment or kinetic measurements should be used in tandem with EMSA analysis.

33. The estimation of K_d requires the protein concentration added to the reaction (P_{tot}) to be approximately equal to the free protein concentration after equilibrium is reached (P_{free}). If the nucleic acid concentration is close to the K_d of the interaction, depletion of the protein will underestimate the resulting affinity. For example, if the K_d is 100 nM and the same concentration of nucleic acid probe is added to the reaction, any protein added to the reaction at and above this concentration will begin to associate, deplete the "free" protein concentration and invalidate the assumption required for K_d estimation [19].
34. Image contrast and brightness adjustments must only be made linearly and applied to an entire image or plate. Nonlinear adjustments are strongly discouraged and are unsuitable for semiquantitative analysis of protein–nucleic acid interactions, as the measured intensity is no longer proportional to the molarity of nucleic acid.
35. There are multiple ways of fitting the data. If the data are not plotted logarithmically, they can be fitted with a single-site binding equation (a form of the Langmuir isotherm) or a quadratic binding function. If plotted logarithmically, the data can also be fitted with a four-parameter logistic function. All of these equations will contain terms that correspond to the dissociation constant, where binding is half maximal.
36. To test whether P2, a protein known to bind to the P1–N1 protein–nucleic acid complex, competes with P1 binding to nucleic acid N1, carry out a titration of P1 against the nucleic acid as in Subheading 3.5 above. Select a concentration of P1 where the free nucleic acid band has completely disappeared. Titrate P2 against fixed concentrations of P1 and N1. A lack of supershift indicates that P2 cannot bind simultaneously to a P1–N1 complex. If the band corresponding to free N1 appears, then P2 interaction with P1 precludes P1 binding to N1. If a small shift in electrophoretic mobility occurs, then it may indicate that the P2–N1 interaction is mutually exclusive with the P1–N1 interaction.
37. Because two different nucleic acids are used in the assay, they must be differentially labeled. For example, the two nucleic acids can be labeled with different fluorophores with nonoverlapping excitation and emission spectra. Theoretically, this

assay could be performed with three differentially labeled substrates: Alexa Fluor 405, Alexa Fluor 568, and Alexa Fluor 790. Alternatively, one nucleic acid can be labeled with a fluorophore or radioactive isotope ("hot"), and the other nucleic acid kept unlabeled ("cold"). This latter protocol can be used if there is no way of scanning each channel separately.

38. Alternatively, a fixed concentration of P1 and labeled N1 can be used, and the concentration of nucleic acid N2 can be titrated. As the concentration of N2 is increased, P1 may be titrated away from the P1–N1 complex. Labeled N1 can thus be detected as a free nucleic acid band.
39. The nucleic acid concentration used will depend on the particular experiment. We often perform these assays with relatively high nucleic acid concentrations to mirror that used in activity assays. In Fig. 2, the two different concentrations illustrate how the differences in specific binding depend on both probe and protein concentration.

Acknowledgments

We thank Michael Webster for providing purified SpPuf3. This work was supported by a Herchel Smith PhD Studentship from the University of Cambridge (to T.T.L.T.); the European Union's Horizon 2020 research and innovation programme (ERC grant 725685) (to L.A.P.); and Medical Research Council grant MC_U105192715 (L.A.P.).

References

1. Pagano JM, Clingman CC, Ryder SP (2011) Quantitative approaches to monitor protein-nucleic acid interactions using fluorescent probes. RNA 17(1):14–20. https://doi.org/10.1261/rna.2428111
2. Fried M, Crothers DM (1981) Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. Nucleic Acids Res 9(23):6505–6525
3. Garner MM, Revzin A (1981) A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the *Escherichia coli* lactose operon regulatory system. Nucleic Acids Res 9(13):3047–3060
4. Hellman LM, Fried MG (2007) Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. Nat Protoc 2(8):1849–1861. https://doi.org/10.1038/nprot.2007.249
5. Berger R, Duncan MR, Berman B (1993) Nonradioactive gel mobility shift assay using chemiluminescent detection. BioTechniques 15(4):650–652
6. Hsieh YW, Alqadah A, Chuang CF (2016) An optimized protocol for electrophoretic mobility shift assay using infrared fluorescent dye-labeled oligonucleotides. J Vis Exp 117:54863. https://doi.org/10.3791/54863
7. Jing D, Agnew J, Patton WF et al (2003) A sensitive two-color electrophoretic mobility shift assay for detecting both nucleic acids and protein in gels. Proteomics 3(7):1172–1180. https://doi.org/10.1002/pmic.200300438
8. Rye HS, Drees BL, Nelson HC, Glazer AN (1993) Stable fluorescent dye-DNA complexes in high sensitivity detection of protein-DNA interactions. Application to heat shock transcription factor. J Biol Chem 268(33):25229–25238

9. Webster MW, Chen YH, Stowell JAW et al (2018) mRNA Deadenylation is coupled to translation rates by the differential activities of Ccr4-not nucleases. Mol Cell 70 (6):1089–1100 e1088. https://doi.org/10.1016/j.molcel.2018.05.033
10. Harbers M (2015) Shift-Western blotting: separate analysis of protein and DNA from protein-DNA complexes. Methods Mol Biol 1312:355–373. https://doi.org/10.1007/978-1-4939-2694-7_36
11. Blackwell TK, Weintraub H (1990) Differences and similarities in DNA-binding preferences of MyoD and E2A protein complexes revealed by binding site selection. Science 250 (4984):1104–1110
12. Beckert B, Masquida B (2011) Synthesis of RNA by in vitro transcription. Methods Mol Biol 703:29–41. https://doi.org/10.1007/978-1-59745-248-9_3
13. Sambrook J, Russell DW (2006) Isolation of DNA fragments from polyacrylamide gels by the crush and soak method. CSH Protoc 2006(1). https://doi.org/10.1101/pdb.prot2936
14. Zarzosa-Alvarez AL, Sandoval-Cabrera A, Torres-Huerta AL, Bermudez-Cruz RM (2010) Electroeluting DNA fragments. J Vis Exp 43. https://doi.org/10.3791/2136
15. Struhl K (2001) Reagents and radioisotopes used to manipulate nucleic acids. Curr Protoc Mol Biol. Chapter 3:Unit3 4. https://doi.org/10.1002/0471142727.mb0304s09
16. Rueden CT, Schindelin J, Hiner MC, DeZonia BE, Walter AE, Arena ET, Eliceiri KW (2017) ImageJ2: ImageJ for the next generation of scientific image data. BMC Bioinformatics 18 (1):529. https://doi.org/10.1186/s12859-017-1934-z
17. Webster MW, Stowell JA, Passmore LA (2019) RNA-binding proteins distinguish between similar sequence motifs to promote targeted deadenylation by Ccr4-not. Elife 8:e40670. https://doi.org/10.7554/eLife.40670
18. Bhattacharyya D, Mirihana Arachchilage G, Basu S (2016) Metal cations in G-Quadruplex folding and stability. Front Chem 4:38. https://doi.org/10.3389/fchem.2016.00038
19. Hulme EC, Trevethick MA (2010) Ligand binding assays at equilibrium: validation and interpretation. Br J Pharmacol 161 (6):1219–1237. https://doi.org/10.1111/j.1476-5381.2009.00604.x

Chapter 15

Biophysical Studies of the Binding of Viral RNA with the 80S Ribosome Using switchSENSE

Emma Schenckbecher, Guillaume Bec, Taiichi Sakamoto, Benoit Meyer, and Eric Ennifar

Abstract

Translation initiation, in both eukaryotes and bacteria, requires essential elements such as mRNA, ribosome, initiator tRNA, and a set of initiation factors. For each domain of life, canonical mechanisms and signals are observed to initiate protein synthesis. However, other initiation mechanism can be used, especially in viral mRNAs. Some viruses hijack cellular machinery to translate some of their mRNAs through a noncanonical initiation pathway using internal ribosome entry site (IRES), a highly structured RNAs which can directly recruit the ribosome with a restricted set of initiation factors, and in some cases even without cap and initiator tRNA. In this chapter, we describe the use of biosensors relying on electro-switchable nanolevers using the switchSENSE® technology, to investigate kinetics of the intergenic (IGR) IRES of the cricket paralysis virus (CrPV) binding to 80S yeast ribosome. This study provides a proof of concept for the application of this method on large complexes.

Key words Kinetics, switchSENSE, Ribosome, RNA, Biophysics

1 Introduction

Due to its key role in the cell, translation machinery has been the subject of intense studies, especially since access to high-resolution structures has been facilitated by X-ray crystallography and cryo-electron microscopy studies. This includes studies focusing on the IRES-mediated initiation such as found in viral RNAs [1–3]. A plethora of studies investigated interactions between various viral IRES (internal ribosomal entry site) and the ribosome [4, 5], especially through the work on the intergenic (IGR) IRES of the cricket paralysis virus (CrPV) which is now well characterized. Numerous approaches are available to characterize interactions between such macromolecular complexes, whether from a kinetic [6, 7], structural [8, 9], or biochemical point of view [10, 11]. Among all new emerging techniques, switchSENSE® technology appears as very innovative in the solid-support immobilization field [12–14]. By

Tina Daviter et al. (eds.), *Protein-Ligand Interactions: Methods and Applications*, Methods in Molecular Biology, vol. 2263, https://doi.org/10.1007/978-1-0716-1197-5_15,

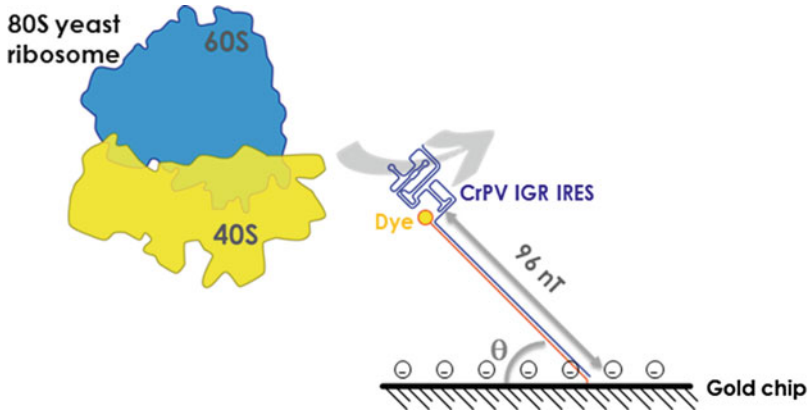


Fig. 1 Principle of switchSENSE® technology applied to the CrPV IGR IRES—ribosome interaction. The 96-nuclotide DNA attached to the chip and labeled with a dye is in red, whereas the complementary RNA sequence tethered to the CrPV IGR IRES is in blue. The potential applied to the chip brings the RNA/DNA hybrid closer from the surface

measuring analytes adsorption on a layer of actuated surface-bound fluorescent probe, this combination of biophysical approaches can be used for several applications, as binding kinetics and affinity, protein diameter, conformational change, or even nuclease and polymerase activity.

The originality of the switchSENSE® technology lies in the DNA strand bearing a fluorescent dye at one extremity, attached on its opposite end to a gold-quenching surface. The complementary strand could be used alone or chemically coupled to an interaction partner (a protein). Hybridization of the DNA strands generates a rigid negatively charged electro-switchable biosensor, also referred as nanolever (Fig. 1). Two principal measurement modes are accessible: (1) a static mode ("proximity sensing"), when the nanolever is repelled from the gold surface by an applied negative charge, and analyte binding can be measured by a change in fluorescence signal of the nanolever (2) a dynamic mode ("switching mode") where binding is detected through changes in the kinetics of the nanolever oscillation under an alternating electric field. While the technique relies on the formation of a double-stranded nanolevers with defined sequences, additions or extensions to the complimentarily strand allow a wide variety of adaptions such as: length and type of DNA [15]; RNA/DNA hybrid [16]; attachment of protein or capture molecules such as biotin, streptavidin, and NTA [12]. Nanolevers with different sequences can be supplied labeled with different dyes attached (depending on the configuration of the instrument) enabling internal referencing between an active ligand-bound nanolever and the nanolever alone. Hence this technique is applicable to the investigation of a range of biomolecular interactions.

In this study, we describe an original use of this technology to study kinetics binding of CrPV IGR IRES with the 80S yeast ribosome. We showed that this method, originally dedicated to smaller complexes, is also well adapted to the study of large macromolecular complexes.

2 Materials

2.1 Instruments and Accessories

1. DRX 2400 or DRX^2 instrument (Dynamic Biosensors).
2. Nanodrop ND-1000 spectrophotometer (Thermo Fisher) or equivalent.
3. switchCONTROL (version 1.6.20.2393), switchBUILD (version 1.12.0.21), and switchANALYSER (version 1.9.0.31) software (Dynamic Biosensors).
4. Biochip for large complexes studies (MPC-96-2-Y1-S, Dynamic Biosensors, *see* **Note 1**) bearing a 96-mer DNA probe (nanolever) attached to the surface at the 5′ end.
5. 1.5 and 10 mL autosampler vials with septa caps and insets.

2.2 Buffers

1. Passivation solution: as supplied by manufacturer (SOL-PAS-1-5, Dynamic Biosensors).
2. Regeneration solution: as supplied by manufacturer (SOL--REG-12-1, Dynamic Biosensors).
3. 10× Auxiliary Buffer: 100 mM sodium phosphate, pH 7.0, 400 mM NaCl, 0.5% (v/v) Tween 20, 0.5 mM EDTA.
4. Running buffer: 20 mM Na cacodylate, pH 7.0, 7 mM $MgCl_2$, 100 mM KCl, 0.05% (v/v) Tween 20 (*see* **Note 2**).
5. Ribosome buffer: 50 mM Na cacodylate, pH 7.0, 7 mM $MgCl_2$, 30 mM KCl, 70 mM NH_4Cl, 1 mM DTT, 5% (w/v) glycerol.
6. Ultrapure RNase-free water (homemade, used for the preparation of all buffers).

2.3 Ligand Preparation

1. CrPV IGR IRES—cNL-B96 RNA sequence, complementary to the 96-mer DNA probe onto the chip (*see* **Note 3**); stored at 1 mg/mL in 10 mM 2-(*N*-morpholino)ethanesulfonic acid, 200 mM NaCl at −20 °C.

2.4 Analyte Preparation

1. 80S *S. cerevisiae* ribosome purified from JD1370 strain following a protocol adapted from Ben Shem et al. ([17]; aliquots, flash-frozen in liquid nitrogen at 10 μM and stored at −80 °C or freshly prepared, in Ribosome buffer).
2. Amicon Ultra-15 Centrifugal Filter with 100 kDa molecular weight cutoff (EMD Millipore).

3 Methods

3.1 Ribosome Preparation

1. After purification, buffer exchange 80S against Running buffer with Amicon concentrator. We generally exchange step by step a final volume of 15 mL Running buffer for an initial volume of less than 1 mL of sample.
2. Determine the 80S concentration using a NanoDrop spectrophotometer. For an accurate concentration, perform three independent twentieth dilutions of the stock solution (i.e., 1 μL stock solution + 19 μL buffer), and measure the absorbance at 260 nm ($\varepsilon_{260} = 50{,}000{,}000\ M^{-1}\ cm^{-1}$).
3. Dilute the stock solution with the 1× Running buffer appropriately to make a maximal concentration of 2 μM in 500 μL. Store at 4 °C.

3.2 Designing the switchSENSE® Experiment

3.2.1 Experimental Considerations

DRX 2400 instrument was not originally designed for analysis of complexes as large as ribosomes. Here, the size of the DNA probe attached to the chip has been doubled, compared to standard conditions, in order to deal with the large size of 80S ribosome. Furthermore, measurements are done in "proximity sensing mode" and not in "switching mode" (*see* **Note 4**) in order to avoid steric constrains. Here we are taking advantage of an increase of fluorescence of the nanolever when the IRES RNA is bound by the ribosome.

3.2.2 Experimental Workflow Building

1. Open the switchBUILD software to design the experiment by the creation of successive programming blocks.
2. In the first block, choose the biochip (MPC-96-2-Y1-S), channel number to be analyzed, Auxiliary Buffer (P40) and Running Buffer (×140). Alternatively a custom buffer can be added manually.
3. A "Passivation" step is added automatically as a second block for all experiments (*see* **Note 5**).
4. To add a new measure block click "+" and select "Split Shot Kinetics."
5. In the new experimental panel in the "Properties" window, select "Conjugate hybridization" in the immobilization method and "static mode" in the measurement mode pull-down menus.
6. Input the name of ligand (CrPV IGR IRES—cNL-B96 oligonucleotide) and its concentration (*see* **Note 6**) and name of the analyte (yeast ribosomes) and its mass.
7. To determine the experimental parameters for analyte injection, either press "PRESETS" and select an appropriate model system or use the sliders for the different kinetic parameters:

dissociation constant K_d; association rate constant k_{on}; and dissociation rate constant k_{off}. Alternatively, the parameter values can be directly entered into the text boxes.

8. Press "Auto Generate" and the highest starting analyte concentration, number of dilutions, association/dissociation times, and flow rates will be automatically filled with the appropriate values determined by the program (*see* **Note 7**). These values can be altered manually, for example, by increasing the time for dissociation or increasing the number of concentrations tested. The graph of predicted results automatically updates. In the specific case of our experiment, we chose the following parameters: 100, 33, and 11 nM for ribosome concentration, 5 min at 50 μL/min for the association step, and 20 min at the same rate for dissociation.
9. Input the analysis temperature; select the electrode number to be measured for both association and dissociation steps (one of the six available in each channel). Electrode 3 at 25 °C was chosen for the experiment.
10. The regeneration step button must be checked (default value), after each concentration for the association part, and chose the dissociation event only at the last concentration (*see* **Note 8**).
11. It is highly recommended to perform blank runs using Running Buffer for both association and dissociation steps. To do this, check that the "with blank run" button is checked for both steps.

    Finally, a "Standby" block is dragged into sequence at the end of the experiment with the "Reset surface button" checked (*see* **Note 9**).
12. Save the switchBUILD assay file in an appropriate folder but keep the program open.

3.3 Performing the switchSENSE Experiment

1. Open the "Autosampler" window in the "Kinetics" block of the switchBUILD experiment file. This displays the volumes, concentrations, and positions in the autosampler rack, of all the solutions required for the experiment (*see* **Note 7**). In our example, three concentrations and buffer (as a blank run) are tested (100, 33, 11, and 0 nM of ribosome in 250 μL of 1× Running buffer) at a flow of 50 μL/min after hybridization with a solution of CrPV IGR IRES—cNL-B96 RNA at 380 nM in 120 μL.
2. Fill the positions in autosampler as displayed with ligand, analyte, Ultrapure RNase-free water, Passivation, and Regeneration solutions and an empty vial for waste.
3. Dilute the Auxiliary Buffer 10× tenfold with Ultrapure RNA-free water: 20 mL stock plus 180 mL water. Insert the appropriate inlet pipes into the 1× Auxiliary and Running buffers.

4. Open the switchCONTROL software.
5. Load the biochip, align and prime the system with buffer.
6. Set the temperature of the autosampler tray and chip if necessary.
7. In the "Taskflow" window, open the folder and add the saved assay from **step 12** (Subheading 3.2). The "Name" and "Project" windows can be filled in order to classify and identify data files more easily.
8. Start the experiment by clicking on the "Start/stop" button.

3.4 switchSENSE® Data Analysis

1. Open the switchANALYSIS software and open a "NEW" analysis.
2. Load the appropriate dataset of the whole experiment.
3. In this dataset, select the kinetic data.
4. Drag and drop these kinetic data in the right part of the window.
5. Click on "CREATE NEW ANALYSIS" in the bottom right part of the window. A new window opens up.
6. Select "Kinetics." All kinetic data are shown on the left part with all association and dissociation events being listed as successive blocks (one block corresponding to one concentration of 80S, Fig. 2).
7. The blank experiment for association and dissociation should be subtracted by dragging and dropping blank experiments (0 M concentrations) on the corresponding association or dissociation blocks for each 80S concentration. This results in the addition of a new line in the bottom of each block (left part of the window), and the displayed curve is now corrected from background.
8. Remove all non-corrected data from the right part of the window by clicking on the garbage icon on the right.
9. Click on "AUTO-LOAD" in the left bottom part of the window. All background-subtracted data are now displayed on the main right window (Fig. 3).
10. If necessary, manually select the proper window for data analysis for each 80S concentration in order to remove artifacts generated by air bubbles following the beginning of each injection.
11. Click on "FIT ANALYSIS" on the bottom right part of the window. Results are shown in the main right window (Fig. 4).
12. A publication-quality image can be made by clicking on the camera icon in the top right part of the window. Results will be copied in the clipboard as a PNG file. Raw datasets could also be exported by clicking on the "EXPORT" icon in the bottom right part of the window.

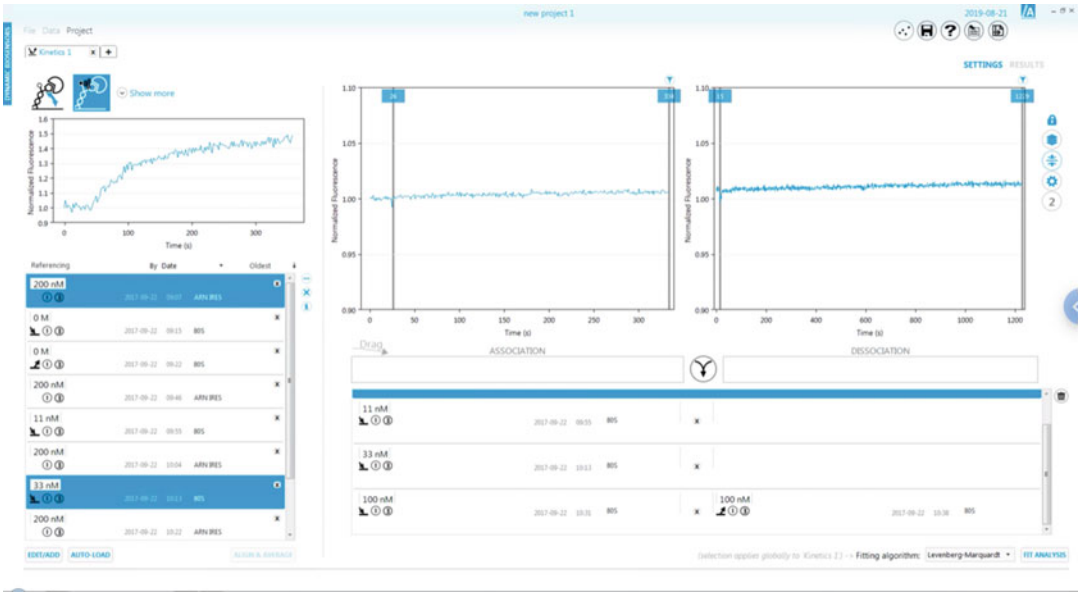


Fig. 2 Screenshot of the switchANALYSIS program after selection of kinetic data. Blocs are listed on the left part, with a down arrow icon for association, an up arrow icon for dissociation, and no arrows for hybridization experiments

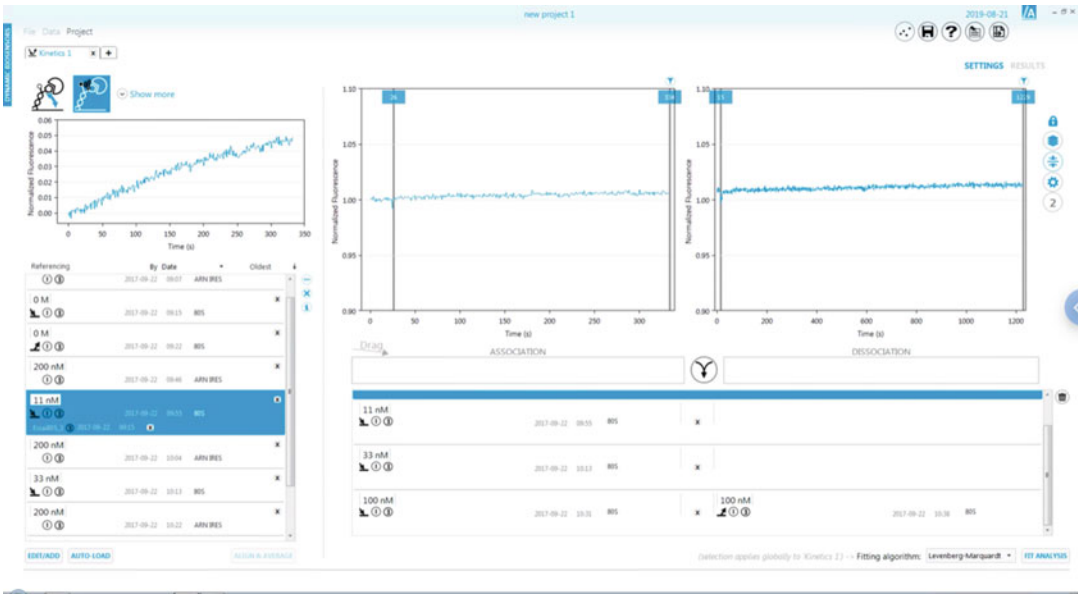


Fig. 3 Screenshot of the switchANALYSIS program following the background correction

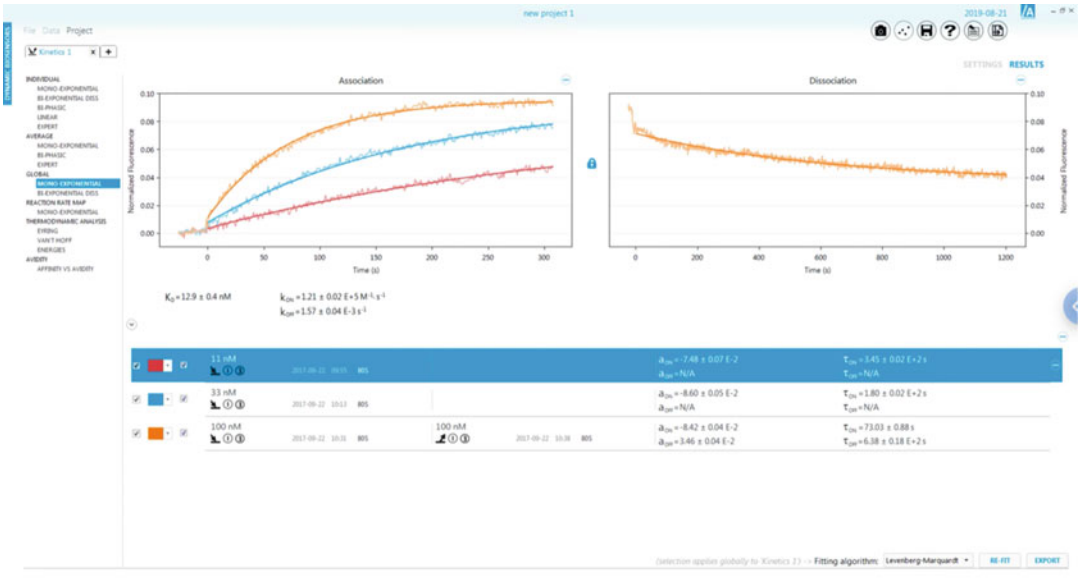


Fig. 4 Screenshot of the switchANALYSIS program showing binding kinetics of the 80S yeast ribosome interaction with CrPV IGR IRES at 25 °C. Data are shown for association at three concentrations of 80S at 11 nM (red), 33 nM (blue), and 100 nM (orange) and for dissociation only at the highest concentration, monitored by changes in normalized "fluorescence up." Fits are shown as solid lines

4 Notes

1. Any standard chip bearing a 96-nucleotide DNA probe would be suitable for switchSENSE® studies of large complexes such as the ribosome. Here a yellow probe was used due to specifications of our DRX instrument. A MPC-96-2-Y1-S chip with NLA and NLB electrodes is used here. As a consequence, cNLA is always included (at a 500 nM concentration) in any hybridization solution in order to protect NLA electrodes. This would be dispensable with a MPC-96-1-Y1-S chip.
2. Buffer restriction for proximity sensing experiments are less restrictive than for the switching mode (up to ~1.5 M in salts vs up to ~150 mM).
3. CrPV IGR IRES—cNLB96 oligonucleotide RNA sequence. (5′**CAACAAATATTAATACGACTCACTATAGCAAAAA UGUGAUCUUGCUUGUAAAUACAAUUUUGAGAG GUUAAUAAAUUACAAGUAGUGCUAUUUUUGUA UUUAGGUUAGCUAUUUAGCUUUACGUUCCAGG AUGCCUAGUGGCAGCCCCACAAUAUCCAGGAAG CCCUCUCUGCGGUUUUUCAGAUUAGGUAGUCG AAAAACCUAAGAAAUUUACC**UGCUACAUUUCAAG AUACCGAAGACGCCAAAAAC<u>AUCAGGAACUACAGGG UGCCCUACUUGCUCUCGGAGGUACUGUAACUAAU</u>

CAGCGUUCGAUGCUUCCGACUAAUCAGCCAUAUC
AGCUUACGACUA3′) is a T7 transcript, containing the CrPV IGR IRES sequence in bold and the sequence commentary to the attached nanolever underlined, obtained from PCR products, and purified by phenol–chloroform extraction followed by a dialysis in Amicon concentrator against miliQ H_2O.

4. In the proximity sensing mode, a constant voltage (−0.1 V in our conditions) is applied on the chip, maintaining the nanolever at a constant angle in the 1× Running Buffer. Any event affecting this angle and/or the distance of the fluorophore with the quenching surface of the biochip (protein binding, polymerization, dissociation) could lead to a signal interpretable for the experiment.
5. This automatized standard procedure (using a commercially available kit solution) is used to avoid any unspecific binding of analytes on the biosurface.
6. The time required for hybridization of the complementary nanolever on the DNA anchor on the chip is directly correlated to its concentration; this can be adjusted depending on the hybridization quality.
7. Association and dissociation time, as well as flow rate, depend on the predicted kinetic values given in the switchBUILD program. Consequently, volumes and concentrations of the analyte are also directly correlated to the considered kinetic parameters. Thanks to the very stable baseline provided by the technology, very long dissociation times are accessible to measurement. Consequently, very slow off-rates can be accurately obtained from switchSENSE®, which is one of the main advantages over other biosensor technologies.
8. It is possible to insert a dissociation step after association for each concentration. However, this will significantly increase duration of the overall experiment and will decrease the chip lifespan.
9. This automatized standard procedure removes analytes and ligands from the surface and defines conditions suitable for chip storage.

Acknowledgments

Special thanks to Claire Batisse, who provided the yeast strain for 80S purification. This strain was generously given by Yusupov lab, which was elaborated in the first place by Dinman lab. The authors would like to thank Dynamic Biosensors for the encouraging collaboration. Finally, thank you to Philippe Dumas for his constant support, Yaser Hashem, Stefano Marzi, and Angelita Simonetti for fruitful discussion.

References

1. Balvay L, Soto Rifo R, Ricci EP et al (2009) Structural and functional diversity of viral IRESes. Biochim Biophys Acta 1789 (9-10):542–557. https://doi.org/10.1016/j.bbagrm.2009.07.005
2. Kieft JS (2008) Viral IRES RNA structures and ribosome interactions. Trends Biochem Sci 33 (6):274–283. https://doi.org/10.1016/j.tibs.2008.04.007
3. Thompson SR (2012) Tricks an IRES uses to enslave ribosomes. Trends Microbiol 20 (11):558–566. https://doi.org/10.1016/j.tim.2012.08.002
4. Filbin ME, Kieft JS (2009) Toward a structural understanding of IRES RNA function. Curr Opin Struct Biol 19(3):267–276. https://doi.org/10.1016/j.sbi.2009.03.005
5. Johnson AG, Grosely R, Petrov AN, Puglisi JD (2017) Dynamics of IRES-mediated translation. Philos Trans R Soc Lond Ser B Biol Sci 372(1716). https://doi.org/10.1098/rstb.2016.0177rstb.2016.0177
6. Bugaud O, Barbier N, Chommy H et al (2017) Kinetics of CrPV and HCV IRES-mediated eukaryotic translation using single-molecule fluorescence microscopy. RNA 23 (11):1626–1635. https://doi.org/10.1261/rna.061523.117rna.061523
7. Petrov A, Grosely R, Chen J et al (2016) Multiple parallel pathways of translation initiation on the CrPV IRES. Mol Cell 62(1):92–103. https://doi.org/10.1016/j.molcel.2016.03.020
8. Muhs M, Hilal T, Mielke T et al (2015) Cryo-EM of ribosomal 80S complexes with termination factors reveals the translocated cricket paralysis virus IRES. Mol Cell 57 (3):422–432. https://doi.org/10.1016/j.molcel.2014.12.016
9. Murray J, Savva CG, Shin BS et al (2016) Structural characterization of ribosome recruitment and translocation by type IV IRES. Elife 5:e13567. https://doi.org/10.7554/eLife.13567
10. Kerr CH, Ma ZW, Jang CJ et al (2016) Molecular analysis of the factorless internal ribosome entry site in cricket paralysis virus infection. Sci Rep 6:37319. https://doi.org/10.1038/srep37319
11. Pestova TV, Lomakin IB, Hellen CU (2004) Position of the CrPV IRES on the 40S subunit and factor dependence of IRES/80S ribosome assembly. EMBO Rep 5(9):906–913. https://doi.org/10.1038/sj.embor.7400240
12. Knezevic J, Langer A, Hampel PA et al (2012) Quantitation of affinity, avidity, and binding kinetics of protein analytes with a dynamically switchable biosurface. J Am Chem Soc 134 (37):15225–15228. https://doi.org/10.1021/ja3061276
13. Langer A, Hampel PA, Kaiser W et al (2013) Protein analysis by time-resolved measurements with an electro-switchable DNA chip. Nat Commun 4:2099. https://doi.org/10.1038/ncomms3099
14. Rant U, Pringsheim E, Kaiser W et al (2009) Detection and size analysis of proteins with switchable DNA layers. Nano Lett 9 (4):1290–1295. https://doi.org/10.1021/nl8026789
15. Kroener F, Heerwig A, Kaiser W et al (2017) Electrical actuation of a DNA origami Nanolever on an electrode. J Am Chem Soc 139 (46):16510–16513. https://doi.org/10.1021/jacs.7b10862
16. Clery A, Sohier TJM, Welte T et al (2017) switchSENSE: A new technology to study protein-RNA interactions. Methods 118-119:137–145. https://doi.org/10.1016/j.ymeth.2017.03.004
17. Ben-Shem A, Garreau de Loubresse N, Melnikov S et al (2011) The structure of the eukaryotic ribosome at 3.0 a resolution. Science 334 (6062):1524–1529. https://doi.org/10.1126/science.1212642

Chapter 16

Biolayer Interferometry: Protein–RNA Interactions

Stephen R. Martin, Andres Ramos, and Laura Masino

Abstract

RNA-binding proteins often contain multiple RNA-binding domains connected by short flexible linkers. This domain arrangement allows the protein to bind the RNA with greater affinity and specificity than would be possible with individual domains and sometimes to remodel its structure. It is therefore important to understand how multiple modules interact with RNA because it is the modular nature of these proteins which specifies their biological function. This chapter is concerned with the use of biolayer interferometry to study protein–RNA interactions.

Key words Biolayer interferometry, RNA, Kinetics

1 Introduction

1.1 Protein–RNA Interactions

Posttranscriptional gene regulation consists of a ubiquitous and essential network of protein–RNA-based cellular processes that expands genomic diversity, and it is essential in the development and function of complex organisms. Not surprisingly, misfunction of the different RNA regulation steps has been associated with a range of pathologies, including different cancers, neurodevelopmental and neurodegenerative diseases, immunopathologies, and viral infection [1, 2].

RNA regulation is mediated by between 1000 and 2000 RNA-binding proteins, a few hundred of which have been validated functionally [3]. In contrast, a human cell typically contains 10 to 20,000 different mRNAs [4], each binding many different RNA-binding proteins. Most RNA-binding proteins recognize large and diverse sets of RNA targets, including both mRNAs and noncoding RNAs. An accurate recognition of these targets is essential not only to define the set of genes regulated by a protein but also to define gene expression programs in different cellular locations at specific times [1, 2]. Understanding, at a mechanistic level, how RNA-binding proteins recognize the RNA targets is one of the main challenges in gene regulation.

Tina Daviter et al. (eds.), *Protein-Ligand Interactions: Methods and Applications*, Methods in Molecular Biology, vol. 2263, https://doi.org/10.1007/978-1-0716-1197-5_16,

Despite the large number of RNA-binding proteins, recognition of the RNA targets is mediated by a relatively small number of different RNA-binding domains, which are present in multiple copies of the same or different domains within one RNA-binding protein [5]. In these proteins, RNA binding is mediated by a combinatorial action of more than one RNA-binding domain. In this context, diversity of recognition stems from both the specificity of a domain in a given protein and from a range of different inter-domain coupling modes. Inter-domain coupling can increase affinity and specificity, reshape the RNA structure, and provide new opportunities for regulation [6].

While a global survey of RNA-binding domains shows different sizes and RNA recognition properties, the most common of these domains are less than 100 amino acids in size. For many of these domains, a structural characterization of a "canonical" binding mode is available, together with, in some cases, the description of a few of the structural and RNA-binding variations on this mode [5, 6]. However, information on the kinetics of binding and inter-domain coupling is available only in a small number of systems. This information is essential to model the binding of these domains in the cellular environment.

As an example of the RNA-binding domains discussed above, the KH domain is a small (~70 amino acids) αβ fold found in a number of RNA regulatory proteins important in development, function, and disease. The domain binds to single-stranded nucleic acids with a varying degree of affinity and specificity. RNA binding is mediated by the interaction of the nucleic acid backbone with a negatively charged GxxG loop [7]. The details of this interaction are different in different domains, but binding of the loop orients the nucleobases towards a hydrophobic groove in the protein, for sequence specific recognition [7]. In addition, individual KH domains can interact using a variety of surfaces. Importantly, despite the importance of KH-containing proteins in human health, kinetic and mechanistic information on multi-domain binding has only recently started to become available. This is partly due to the difficulties in obtaining high-quality data on the kinetics of protein–RNA interactions at different affinities in the same experimental system.

1.2 Biolayer Interferometry

Biolayer interferometry (BLI) is a label-free method that enables real-time analysis of biomolecular interactions occurring on 8 or 16 biosensors in 96- or 384-well plates [8–10]. White light travels down the biosensors and is reflected back to spectrometers from two places: an internal reference layer and the interface between the solvent and molecules immobilized on the sensor tip. This results in an interference pattern, and the instrument measures the maximum wavelength of the pattern.

The interaction of a binding partner with a molecule immobilized on the biosensor tip gives an increase in the distance between the internal reference layer and material attached to the biosensor. This results in a wavelength shift in the maximum of the interference pattern which is monitored in real time.

BLI can be used to analyze binding interactions of small molecules, proteins, antibodies, nucleic acids, viruses, or whole cells. It can determine specificity, binding kinetics, and affinity and perform quantitation assays. The tips of the biosensors are derivatized with a range of different surface chemistries and can be used to analyze macromolecules with different tags (e.g., His-tag, GST-tag, biotin). In the case of protein–nucleic acid interactions, biotinylated nucleic acids can be immobilized on streptavidin sensors and binding of the protein partner can be recorded. One significant advantage of BLI is that only molecules binding to or dissociating from the biosensor will change the interference pattern and generate an instrument response. Unbound molecules and changes in the refractive index of the solvent have no effect on the interference pattern. A further advantage is that the measurement is nondestructive and samples are recoverable.

1.3 Kinetic Theory

Although all the Octet instruments come with built-in software for curve analysis, it is of course advisable to fully understand the kinetic theory that underpins the technique. In addition, if in-house software for kinetic analysis is available, then complex instrument response curves can be downloaded and analyzed using more sophisticated approaches than those available with the instrument.

In the simplest case, the kinetic analysis of the biosensor data is based on the idea that the interaction between the soluble protein reactant (P) and an immobilized nucleic acid (N) may be described by the following scheme:

$$\mathrm{P} + \mathrm{N} \underset{k_{\mathrm{off}}}{\overset{k_{\mathrm{on}}}{\rightleftarrows}} \mathrm{PN}$$

where k_{on} and k_{off} are the association and dissociation rate constants (units $\mathrm{M}^{-1}\ \mathrm{s}^{-1}$ and s^{-1}, respectively). Under conditions where the extent of the reaction is governed by reaction kinetics rather than mass transport considerations, the differential equation for such a system is:

$$\frac{\mathrm{d}[\mathrm{PN}]}{\mathrm{d}t} = k_{\mathrm{on}}[\mathrm{P}][\mathrm{N}] - k_{\mathrm{off}}[\mathrm{PN}] \tag{1}$$

Substituting $[\mathrm{N}] = [\mathrm{N}_0] - [\mathrm{PN}]$ (where $[\mathrm{N}_0]$ is the total (unknown) concentration of binding sites on the sensor) gives:

$$\frac{\mathrm{d}[\mathrm{PN}]}{\mathrm{d}t} = k_{\mathrm{on}}[\mathrm{P}]([\mathrm{N_o}] - [\mathrm{PN}]) - k_{\mathrm{off}}[\mathrm{PN}] \tag{2}$$

In most cases, the protein concentration will remain at its initial value ($[P_0]$) throughout the reaction because the total concentration of binding sites on the sensor is vanishingly small compared with the protein concentration in the well. Therefore:

$$\frac{d[PN]}{dt} = k_{on}[P_o]([N_o] - [PN]) - k_{off}[PN] \tag{3}$$

If there is no nonspecific binding of the protein to the sensor, then the instrument response must be directly proportional to [PN], and this equation may therefore be rewritten as:

$$\frac{dR}{dt} = k_{on}[P_o](R_{max} - R) - k_{off}R \tag{4}$$

where R denotes the response at time t and R_{max} is the maximal response that would be obtained if all available binding sites on the sensor were saturated (i.e., when $[PN] = [N_0]$). Integration of this equation gives:

$$R = \frac{k_{on}[P_o]R_{max}\left\{1 - e^{-t(k_{on}[P_o]+k_{off})}\right\}}{k_{on}[P_o] + k_{off}} \tag{5}$$

Assuming that the maximum possible response (R_{max}) and the equilibrium response at the end of the association phase (R_{eq}) must be proportional to $[N_0]$ and [PN], respectively, one may write:

$$\frac{k_{on}}{k_{off}} = \frac{[PN]}{[P][N]} = \frac{[PN]}{[P_o]([N_o] - [PN])} = \frac{R_{eq}}{[P_o]\left(R_{max} - R_{eq}\right)} \tag{6}$$

$$\text{and } R_{eq} = \frac{k_{on}[P_o]R_{max}}{k_{on}[P_o] + k_{off}} \tag{7}$$

Substitution in Eq. 5 then gives:

$$R = R_{eq}\left\{1 - e^{-t(k_{on}[P_o]+k_{off})}\right\} \tag{8}$$

and further substituting $k_{obs} = k_{on}[P_o] + k_{off}$ gives:

$$R = R_{eq}\left\{1 - e^{-tk_{obs}}\right\} \tag{9}$$

The time dependence of the biosensor response in the association phase is then expressed in terms of a pseudo-first-order rate constant k_{obs} and R_{eq}, the response at equilibrium. Values for k_{on} and k_{off} can then, in favorable cases, be obtained as the slope and y-axis intercept of a plot of k_{obs} versus $[P_0]$, and the equilibrium dissociation constant (K_d) can be calculated as k_{off}/k_{on}.

A value for the K_d can also be obtained from the variation of the R_{eq} value with protein concentration. Dividing Eq. 7 by k_{on} and substituting $K_d = k_{off}/k_{on}$ gives:

$$R_{eq} = \frac{[P_o]R_{max}}{[P_o] + K_d} \tag{10}$$

Nonlinear regression analysis of R_{eq} values obtained at a series of protein concentrations should therefore yield estimates of K_d and R_{max}.

A much simpler expression applies to the protein dissociation that results when buffer is substituted for the protein solution as the liquid covering the biosensors (i.e., $[P_o] = 0$ in Eq. 4):

$$R = R_o e^{-tk_{off}} \tag{11}$$

where R_o is the biosensor response prior to the start of the dissociation. R_o will not always be equal to R_{eq} as the association curves, particularly those recorded at low added protein concentrations, will not necessarily have reached equilibrium at the end of the association phase. Analysis of the dissociation phase can therefore, in favorable cases, give an additional independent measure of the dissociation rate constant k_{off}.

In relatively rare cases, it is possible to extract a self-consistent set of kinetic and thermodynamic parameters using the three approaches outlined above in Eqs. 9, 10, and 11. It is more often the case that not all relevant parameters can be determined. For example, low-affinity interactions generally have high k_{off} values and their study necessarily requires the use of high protein concentrations. In such cases, the association and dissociation phases are likely to be very fast and because the instrument only records data every 0.2 s it will not be possible to extract rate constants using Eq. 9 or Eq. 11. In favorable cases, it may still be possible to determine a value for the K_d using Eq. 10 as R_{eq} values can generally be obtained from the "top-hat" instrument response curves that are observed when the association rate equals the dissociation rate and the overall response is flat. In the case of high affinity interactions, the dissociation rate is likely to be very slow, and it is often not possible to determine a value for k_{off} using Eq. 11 or from the intercept of a plot of k_{obs} vs $[P_0]$. The determination of a K_d for a high affinity interaction necessarily requires the use of very low protein concentrations and this can be problematic because the reaction will take a long time to reach equilibrium and extracting reliable R_{eq} values using Eq. 9 may be difficult.

2 Materials

2.1 Instrumentation

Instruments are available from ForteBio (part of the Sartorius group, https://www.sartorius.com/en/products/protein-analysis). The Octet RED96 system that we use is an 8-channel instrument that is ideally suited for the characterization of protein–nucleic acid interactions (*see* **Note 1**).

2.2 Consumables

Two 96-well microplates are required for every assay with the Octet RED96: a sample plate for the experiment and a plate for prehydrating the sensors. The sample volume for the plates is 200 μl per well and for the Octet RED96 the plates must be Greiner catalogue number 655209.

As noted above, a wide range of biosensors is available from ForteBio (Sartorius group). For most of our work, we have used streptavidin-coated biosensors to capture 5′ biotinylated oligonucleotides.

2.3 Reagents

1. Experimental buffer: The manufacturer's recommended buffer for routine measurements is PBS (or HBS) containing 0.1 mg/ml BSA and 0.002% (v/v) Tween 20. The buffer should be the same for all samples and in all experimental steps (*see* **Note 2**).
2. Biotinylated oligonucleotides: These can be purchased from various suppliers (e.g., Dharmacon and Integrated DNA Technologies) (*see* **Note 3**).
3. Protein samples: These should be of the highest possible purity and concentrations need to be accurately determined.

3 Methods

3.1 Standard Binding Experiment

This section describes the method that is most often used in the determination of equilibrium dissociation constants and kinetic constants for the interaction of proteins with immobilized oligonucleotides.

1. Soak the biosensors in experimental buffer using a sensor rack and a 96-well microplate. The biosensors should be prehydrated for at least 20 min before initiating the measurement.
2. Fill three columns in the sample microplate with experimental buffer.
3. Fill a column in the sample microplate with the biotinylated oligonucleotide at the same concentration in each well (typically around 0.5 μg/ml) (*see* **Note 4**).
4. Fill a column in the sample microplate with protein at different concentrations. Ideally, the concentration range should be from $0.1 \times K_d$ to $10 \times K_d$. In order to give adequate coverage of this range, it is frequently necessary to repeat the measurement with a different set of protein concentrations.
5. The typical standard binding experiment involves five steps in which sensors must be dipped into the different columns of the 96-well plate (*see* Fig. 1). Program the computer to perform the following steps:

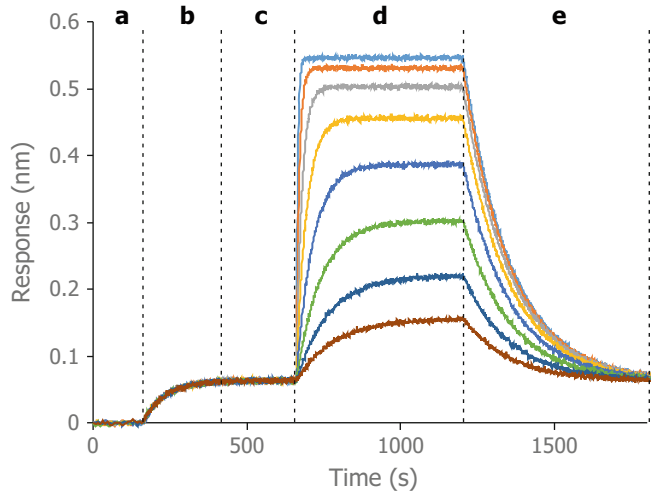


Fig. 1 Computer simulation of the five steps required in a "typical" BLI experiment (*see* text for details). The curves were simulated with $k_{on} = 4 \times 10^5\ M^{-1}\ s^{-1}$, $k_{off} = 0.006\ s^{-1}$, and protein concentrations ranging from 4 to 500 nM

(a) Baseline 1: The sensors are dipped in the first column of buffer for equilibration. The length of this step is typically 200 s.

(b) Loading: The sensors are dipped in the column containing the biotinylated oligonucleotide. In experiments designed to determine a K_d using Eq. 10, it is important that all sensors should give the same response in this step (*see* **Note 5**).

(c) Baseline 2: The sensors are dipped in the second column of experimental buffer to remove any unbound RNA. These may be the same wells as those used in **step** (**a**) but we would advise using a separate column of wells (*see* **Note 6**).

(d) Association: The sensors are dipped in the column containing the protein (*see* **Note 7**).

(e) Dissociation: The sensors are dipped in the third column of experimental buffer. It is advisable to use a separate column of buffer and not the same as those used in **steps** (**a**) and (**c**) in order to avoid any cross contamination (*see* **Note 8**).

At the end of the experiment, samples in the 96-well plate can be recovered or reused for subsequent experiments, provided the reagents have not deteriorated and little or no evaporation has occurred, as this would lead to a concentration increase. Although we routinely discard the sensors at the end of the experiment, they can be used for additional cycles of measurements provided that the

bound molecule completely dissociates in **step** (**e**). In the case of incomplete dissociation, the sensors can sometimes be regenerated and used in subsequent experiments (*see* **Note 9**).

Two common problems that may be encountered are nonspecific binding of the protein to the sensor surface and instrument drift, which may be a problem in those experiments which are designed to run for a long time. In initial experiments, we routinely include two controls. We check for nonspecific binding of the protein to the sensor surface using **steps** (**a**)–(**e**) above but using a reference biosensor with no oligonucleotide loaded in **step** (**b**).

Instrument drift problems can be identified by using a reference biosensor with no immobilized oligonucleotide in **step** (**b**) and no protein present in **step** (**d**).

The inclusion of BSA and Tween 20 will generally reduce nonspecific binding but will not always eliminate it completely (*see* **Notes 10** and **11**). In principle a small amount of nonspecific binding can be corrected for by including reference sensors where no oligonucleotide is bound in **step** (**b**) but because the amount of nonspecific binding will depend on the protein concentration, this requires a separate reference sensor for each protein concentration used.

3.2 Ternary Complexes

More complex experimental designs are, of course, possible and are often very informative. For example, it is possible to study the formation of ternary complexes where two molecules can bind to different sites on an immobilized molecule [11]. This would be described by the following scheme:

$$\begin{array}{ccc} N + P1 & \rightleftarrows & P1N \\ + & & + \\ P2 & & P2 \\ \uparrow\downarrow & & \uparrow\downarrow \\ P2N + P1 & \rightleftarrows & P1P2N \end{array}$$

where N is the immobilized nucleic acid and P_1 and P_2 are the first and second binding partners. In this case, the protocol would be:

1. Use the method described in Subheading 3.1 to assess the binding of the individual proteins to form the binary complexes P_1N and P_2N.
2. Set up the 96-well microplate for the ternary complex binding experiment by filling the following columns:

   Columns 1–3: Experimental buffer.

   Column 4: Biotinylated oligonucleotide.

   Columns 5–6: First binding partner P_1 at a fixed saturating concentration (at least 20 times the K_d for formation of the binary complex P_1N).

Column 7: First binding partner P_1 at the same fixed concentration and varying concentrations of the second partner P_2 (ideally in the range $0.1 \times K_d$ to $10 \times K_d$ for P_2 binding to P_1N, if known).

3. Program the computer to perform the following steps:

   Steps (a)–(c) Baseline 1, Loading, Baseline 2: As in **step 5** in Subheading 3.1.

   Step (d) Association 1: The sensors are dipped in column 5, containing P_1. In this step, the binary complex P_1N is formed.

   Step (e) Association 2: The sensors are dipped in column 7, containing P_1 and varying concentrations of P_2. In this step, the ternary complex P_1P_2N is formed.

   Step (f) Dissociation 1: The sensors are dipped in column 6, containing P_1 but no P2. In this step, the dissociation of P_2 is measured.

   Step (g) Dissociation 2: The sensors are dipped in buffer column 3, to measure the dissociation of P_1.

3.3 Competition

It is also possible to perform various different types of competition experiments. If two protein molecules (say P1 and P2) compete for the same site on an immobilized molecule, then it is possible to study species P2 displacing bound species P1 (or vice versa), providing that the two species give a significantly different response when bound (*see* **Note 12**). This would be described by the following scheme:

$$N + P1 \rightleftarrows P1N$$
$$N + P2 \rightleftarrows P2N$$

where N is the immobilized nucleic acid and P_1 and P_2 are the first and second binding partners. In this case, the protocol would be:

1. Use the method described in Subheading 3.1 to assess the binding of the individual proteins to form the binary complexes P_1N and P_2N.
2. Set up the 96-well microplate for the experiment by filling the following columns:

   Columns 1–3: Experimental buffer.

   Column 4: Biotinylated oligonucleotide.

   Columns 5–6: First binding partner P_1 at a fixed saturating concentration (at least 20 times the K_d for formation of the binary complex P_1N, or higher).

Column 7: First binding partner P_1 at the same fixed concentration and varying concentrations of the second partner P_2, ideally in the range $0.1 \times K_d$ to $10 \times K_d$ for P_2 binding to N, if known. This concentration range might need to be changed if P_1 and P_2 bind with very different affinities.

3. Program the experiment to perform the following steps:

   Steps (a)–(c) Baseline 1, Loading, Baseline 2: As in **step 5** in Subheading 3.1.

   Step (d) Association 1: The sensors are dipped in column 5, containing P_1. In this step the binary complex P_1N is formed.

   Step (e) Association 2: The sensors are dipped in column 7, containing P_1 and P_2. In this step, some of P1 will dissociate and some of P2 will bind.

   Step (f) Dissociation 1: The sensors are dipped in column 6, containing only P_1. In this step, the P2 should dissociate and P1 should rebind.

   Step (g) Dissociation 2 (optional): The sensors are dipped in buffer column 3, to measure the dissociation of P_1.

3.4 Data Analysis

In recent years, we have employed the analytical approach described in Subheading 1.3 to determine kinetic (k_{on} and k_{off}) and thermodynamic (K_d) parameters for the interaction of different proteins with their DNA and RNA target sequences [11–16] and also showed that it is possible, starting from those data, to build a kinetic model for the interaction that provides information on mRNA regulation and RNA remodeling [15, 16].

IMP1/ZBP1 is a multifunctional RNA-binding protein that regulates mRNA metabolism, transport, and translation during development and in cancer [17]. It contains six putative RNA-binding domains (two RRM and four KH) organized in three two-domain units. Interestingly, the binding of different RNA targets is mediated by the two KH di-domains, KH1KH2 and KH3KH4, in a target-dependent fashion. For example, binding of the c-Myc oncogene mRNA in highly proliferating cells requires the KH1KH2 di-domain [18], while the interaction of IMP1 with the β-actin mRNA in neurons requires only the KH3KH4 di-domain [19].

In a recent study, we examined the interaction of KH1KH2 with an oligo recapitulating an IMP1-binding site (CACAGCAUA CAUCCUGUCCGUC), which we named MYCRNA [16]. An important tool to dissect this interaction has been a KH domain mutant where nucleic acid binding is eliminated by the mutation of the two variable amino acids in the hallmark GxxG loop to Aspartate (GxxG-to-GDDG) [12, 13]. This mutation does not affect the structure or the stability of the domain and allows one to examine

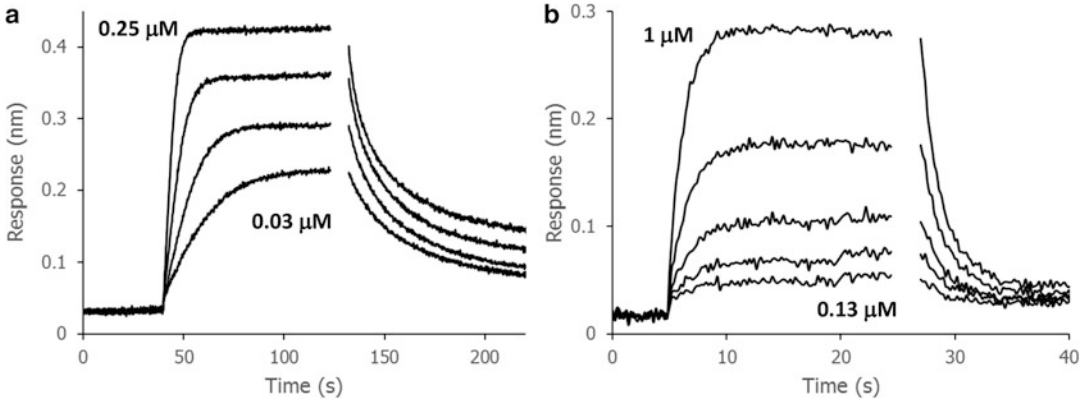


Fig. 2 BLI data for the interaction of IMP1 constructs with MYCRNA. (**a**) Wild-type KH1KH2: serial dilutions from 0.25 μM (0.25, 0.13, 0.06, 0.03 μM). $k_{on} \sim 1 \times 10^6\ M^{-1}\ s^{-1}$, $k_{off} \sim 0.047\ s^{-1}$, $K_d \sim 47$ nM. (**b**) KH1KH2(DD): serial dilutions from 1 μM (1, 0.5, 0.25, 0.13 μM). $k_{on} \sim 2.7 \times 10^5\ M^{-1}\ s^{-1}$, $k_{off} \sim 0.48\ s^{-1}$, $K_d \sim 1.76$ μM

RNA binding of the individual KH domains within an intact IMP1 KH di-domain structural context. We have used three protein constructs: wild-type KH1KH2, KH1KH2(DD) (the KH2 KO), and KH1(DD)KH2 (the KH1 KO). BLI experiments using immobilized MYCRNA exposed to different concentrations of IMP1 allowed us to obtain the equilibrium dissociation constants (K_d) as well as the kinetic parameters for the interactions of two of these constructs. Typical experimental data for the wild-type and the KH1KH2(DD) construct are shown in Fig. 2 along with the kinetic and thermodynamic parameters for the interactions.

In a different study, we again used BLI to investigate the interaction of β-actin mRNA with a KH3KH4 di-domain from the chicken orthologue of IMP1, Zipcode binding protein 1 (ZBP1) [15]. The IMP1 protein is conserved from Drosophila to human, in particular within the KH domains [17]. ZBP1 has the same RNA-binding properties as the human protein and is often used as a proxy to study the IMP1–RNA interaction in vitro. As discussed above, the interaction with the β-actin mRNA is mediated by the KH3KH4 di-domain [19], which recognizes the 28-nucleotide β-actin 3′ UTR Zipcode RNA element (ACCGGACU GUUACCAACACCCACACCCC) (*see* Fig. 3). In order to study the KH3KH4 interaction, we used wild-type protein plus two GxxG-to-GDDG ZBP1constructs, KH3KH4(DD) (the KH4 KO) and KH3(DD)KH4 (the KH3 KO). This is similar to what was discussed above for the KH1KH2–RNA interaction. The equilibrium interaction was studied by using immobilized 28-nucleotide Zipcode RNA exposed to different concentrations of ZBP1 KH3(DD)KH4 and KH3KH4(DD). The equilibrium dissociation constants for the Zipcode RNA:KH3KH4(DD) and RNA:KH3(DD)KH4 complexes were found to be ~1.5 μM and ~0.9 μM, respectively. Although the affinities of the two domains are similar, the kinetic constants are somewhat different.

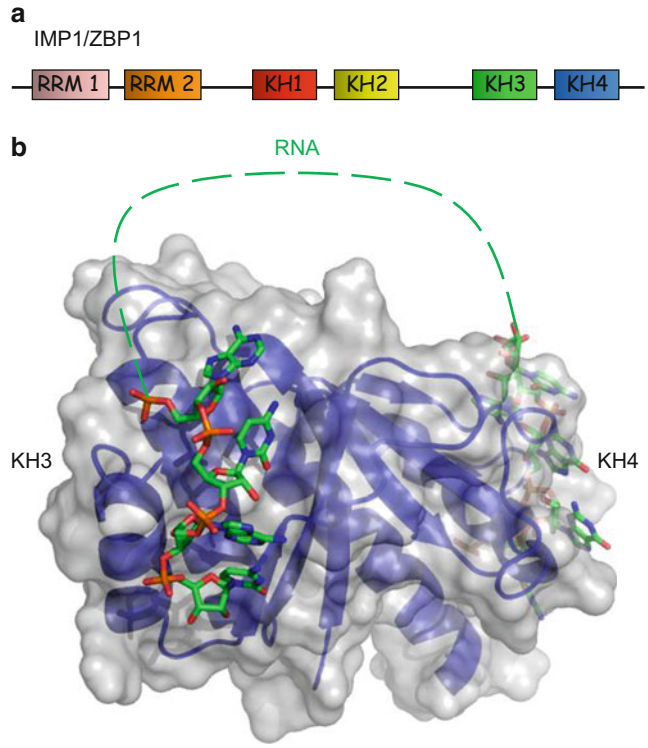
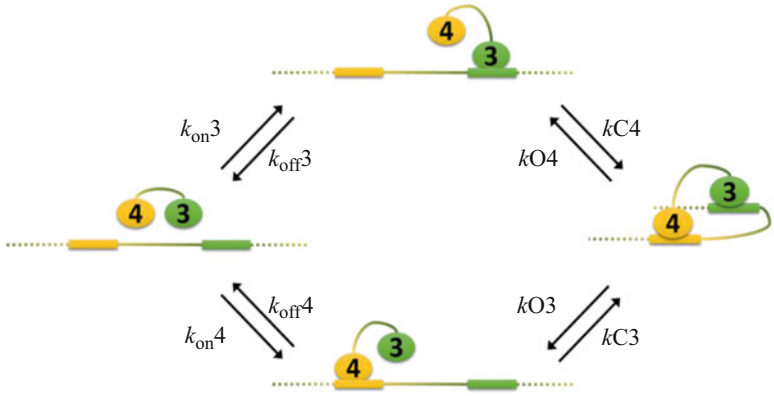


Fig. 3 RNA binding by the protein regulator IMP1/ZBP1. (**a**) Domain organization of IMP1. (**b**) Inter-domain arrangement and RNA binding by the ZBP1 KH3KH4 di-domain structural unit. The surface representation of the bound KH3KH4 protein (grey) and the ribbon representation of the protein backbone (blue) are shown. The two bound cognate RNA sequences from the well-characterized β-actin mRNA target (CACA for KH3 and CGGAC for KH4) are displayed using a stick representation colored by atom type. A dashed line has been traced to represent the connection between these two sequences, which does not make contact with the protein. The image has been built by superimposing the NMR structures of the KH3KH4DD–CACA and KH3DDKH4–CGGAC complexes

The association rate constant for KH3(DD)KH4 (1.4×10^5 M^{-1} s^{-1}) is five times faster than for KH3KH4 (DD) (3.0×10^4 M^{-1} s^{-1}). Conversely, the dissociation rate constant for the KH3KH4(DD):RNA complex (0.046 s^{-1}) is three times slower than that for the KH3(DD)KH4 complex (0.13 s^{-1}).

The wild-type construct (KH3KH4) in which both domains can engage in the interaction binds to the RNA with an association rate constant (1.6×10^5 M^{-1} s^{-1}) that is similar to that for KH4 but the dissociation rate constant is very much smaller (0.0033 s^{-1}). The ratio of these constants gives a K_d of ~20 nM, indicating that the coupling of KH3 and KH4 binding is relatively weak, increasing the affinity of the individual interactions by only a factor of ~50 (*see* **Note 13**).



$k_{on}3$	$3.0 \times 10^4\ M^{-1}s^{-1}$
$k_{off}3$	$0.046\ s^{-1}$
$k_{on}4$	$1.4 \times 10^5\ M^{-1}s^{-1}$
$k_{off}4$	$0.13\ s^{-1}$
$kC3$	$\sim 2\ s^{-1}$
$kO3$	$0.046\ s^{-1}$
$kC4$	$\sim 9.3\ s^{-1}$
$kO4$	$0.13\ s^{-1}$

$k_{on} = 1.6 \times 10^5\ M^{-1}s^{-1}$, $k_{off} = 0.00033\ s^{-1}$, and $k_d = 20$ nM.

Fig. 4 Kinetic model for the interaction of KH3KH4 constructs from ZBP1 with a 28-nucleotide Zipcode RNA (ACCGGACUGUUACCAACACCCACACCCC). $k_{on}3$ and $k_{off}3$ were determined from experiments with KH3KH4 (DD), $k_{on}4$ and $k_{off}4$ were determined from experiments with KH3(DD)KH4, k_{on} and k_{off} were determined from experiments with wild-type KH3KH4. The remaining constants were estimated as described in the text

On the basis of the experiments described above, we proposed a model (Fig. 4) in which either domain of KH3KH4 can associate with its cognate sequence on the Zipcode to form a 1:1 complex. Each of the two possible complexes formed in this way can then proceed through a "ring-closure" step, in which the remaining unbound domain binds to its cognate RNA sequence [20]. Alternatively, a second KH3KH4 protein could bind to the unoccupied cognate sequence (*see* **Note 14**). The second scenario leads to the formation of a 2:1 protein–RNA complex, whereas the first leads to RNA remodeling.

Both pathways for formation of the closed complex involve a bimolecular step followed by what is in effect a conformational change. For such a mechanism, the equilibrium dissociation constant (K_d) for formation of the closed complex is given by [21]:

$$K_d = \frac{K_d A . K_d B}{1 + K_d B} \tag{12}$$

In the case of the upper pathway $K_d A = k_{off}3/k_{on}3$ and $K_d B = kO4/kC4$

$$K_d = \frac{K_d A . {}^{kO4}/_{kC4}}{1 + {}^{kO4}/_{kC4}} = \frac{K_d A . kO4}{kC4 + kO4} \tag{13}$$

$$kC4 = \frac{kO4(K_d A - K_d)}{K_d} \tag{14}$$

The data available from our measurements allowed us to calculate the $kC4/kO4$ ratio using Eq. 14 but not the absolute values.

However, a value of $kC4$ (~9.3 s^{-1}) was obtained from this equation by making the reasonable assumption that $kO4$ is, in fact, the same as $k_{off}4$. In the case of the lower pathway $K_dA = k_{off}4/k_{on}4$ and $K_dB = kO3/kC3$. Assuming, as above, that $kO3$ is the same as $k_{off}3$, a value of $kC3$ (~2 s^{-1}) was calculated from the following equation:

$$kC3 = \frac{kO3(K_dA - K_d)}{K_d} \tag{15}$$

At the low protein concentrations used in our experiments, both pathways for formation of the closed complex include a conformational change step that is very much faster than the initial bimolecular binding step. Under these conditions, the rate expressions for reactions occurring exclusively by the upper and lower pathways in Fig. 4 are given by Eqs. 16 and 17, respectively (*see* **Note 15**).

$$k_{obs} = k_{on}3[\text{Protein}] + \frac{k_{off}3.kO4}{kC4 + kO4} \tag{16}$$

$$k_{obs} = k_{on}4[\text{Protein}] + \frac{k_{off}4.kO3}{kC3 + kO3} \tag{17}$$

For a reaction occurring exclusively by the upper pathway, the k_{on} and k_{off} values would be 3×10^4 M^{-1} s^{-1} and 0.00063 s^{-1}. For a reaction occurring exclusively by the lower pathway, the k_{on} and k_{off} values would be 1.4×10^5 M^{-1} s^{-1} and 0.0029 s^{-1}. These latter values are very close to the values observed in our experiments with the wild-type KH3KH4 di-domain (1.4×10^5 M^{-1} s^{-1} and 0.0033 s^{-1}), suggesting that the lower pathway dominates in both the association and dissociation steps. For both pathways, the overall dissociation constant for formation of the closed complex is ~20 nM, as it must be (*see* **Note 16**).

4 Notes

1. The new Octet RED96e system is an enhancement to the Octet RED96 instrument that permits assays to be performed over a slightly wider temperature range (15–40 °C), allowing for kinetic measurement of unstable proteins. An evaporation cover for microplates also results in minimal sample evaporation for up to 12 h. The Octet RED384 is a 16-channel instrument that provides analytical performance similar to the 8-channel Octet RED96 systems.
2. The BSA and Tween 20 are included to minimize nonspecific binding and are not always needed. In practice many different buffer systems can be used, and in some cases the buffer requires additional additives. For example, in the measurements with oligonucleotides, we routinely use 10 mM sodium

phosphate, pH 7.4, 50 mM NaCl, 0.5 mM TCEP, with 0.5 mg/ml BSA, 0.002% (v/v) Tween-20, and RNAse inhibitor (RNAsin, Promega) at 40–100 u/ml.

3. Oligonucleotides should be de-protected by following the manufacturer's instructions, lyophilized, and resolubilized in the appropriate buffer. Final oligonucleotide concentrations are then calculated from absorption spectroscopy by using the Beer-Lambert law ($A = c\,\varepsilon\,l$, where A is the absorption, c the concentration in $\mathrm{mol{\cdot}L^{-1}}$, ε the extinction coefficient in $\mathrm{L{\cdot}mol^{-1}{\cdot}cm^{-1}}$, and l the pathlength in cm).
4. The lowest concentration of immobilized oligonucleotide that gives enough signal in the protein association step should be selected as overloading the biosensor may lead to overcrowding and steric hindrance. It is generally the case that slow loading for a long time is preferable to fast loading in a short time.
5. It can sometimes happen that not all sensors give the same response in this step. In this case, the measured association or dissociation amplitudes can be normalized for the different loading levels. With some oligonucleotides, the response during the loading phase can be very small, making it difficult to ensure equal loading of all the biosensors. In some cases, we have found that using a different salt concentration in **step** (**b**) increases the size of the response so that equal loading can be confirmed. The sensors then need to be returned to the experimental buffer in **step** (**c**).
6. The baseline signal after loading in **step** (**b**) should be stable, that is there should be no leaching of the bound RNA. This is almost always the case with streptavidin (SA) biosensors but may not be with other sensor types. If leaching does occur then it is generally the case that reducing the concentration in the loading step reduces the extent of the leaching.
7. In the ideal case, the length of the association phase should be long enough to allow all response curves to approach close to equilibrium but this is not always possible, particularly in studies of high affinity interactions which require the use of low concentrations and therefore slow binding kinetics. When repeating the measurement with a different set of protein concentrations in order to cover the appropriate concentration range, the loading level reached in **step** (**b**) must be the same in each measurement as the instrument response, but not the kinetics of the response, is directly proportional to the loading level. If this is not the case, the instrument response (signal amplitude of the association or dissociation phase) can be normalized for the loading level (*see* **Note 5**).
8. The manufacturers recommend that the duration of the dissociation phase should be long enough to give at least 5% dissociation. Although this may be reasonable in some cases, it is

advisable to remember that the determination of the dissociation rate from such limited dissociation is based primarily on the assumption that the response following *complete* dissociation will be identical to that recorded prior to the association step (i.e., the response in **step (c)** in Fig. 1). Any instrument drift or residual nonspecific binding can therefore have a significant effect on the dissociation rate constant determined.

9. Regeneration and subsequent reuse of biosensors offers the user considerable cost savings. The immobilized molecule must be stable under the regeneration conditions employed (generally high or low pH, high salt concentration or added detergent) and must retain binding capacity over several regeneration cycles. In addition, bound molecules must of course be completely removed by the regeneration process.
10. In rare cases, the amount of nonspecific binding observed with a loaded biosensor can be greater than that seen on an unloaded one. The presence of a slow phase in the association step that never reaches equilibrium followed by incomplete dissociation is an indication that there are problems with nonspecific binding of the protein to the sensor surface.
11. SSA biosensors are super streptavidin sensors. They have higher density of streptavidin on the surface compared to the SA sensors. This allows for a higher binding signal, and interactions over a larger surface area of the biosensor will be specific, thus reducing nonspecific binding. Note however that SSA biosensors are much more expensive than SA.
12. This approach is particularly useful if one of the proteins gives such a poor response that it is not possible to obtain a dissociation constant using the standard protocol described in Subheading 3.1.
13. The coupling is defined as weak because in the case of perfect coupling, the affinity of the KH3KH4 construct would be equal to the product of the affinities of the RNA:KH3KH4 (DD) and RNA:KH3(DD)KH4 complexes (~1.5 μM and ~0.9 μM), i.e., ~1.4 pM.
14. The alternative binding pathway, i.e., binding of a second protein to the same RNA, would require a significantly higher affinity for the two interactions because the concentrations of protein and RNA used in our experiments are low compared with the K_ds for the binding of the individual domains.
15. These equations only apply if the conformational change is very much faster than the bimolecular step over the range of protein concentrations being examined. If the bimolecular step is faster than the conformational change under all conditions, then there would be two kinetic phases with the fast phase varying linearly with protein concentration and the slow process

varying hyperbolically with protein concentration. For the upper pathway, the observed rates of the fast and slow processes would be given by:

$$k_{\mathrm{obs}}(\mathrm{F}) = k_{\mathrm{on}}3[\mathrm{Protein}] + k_{\mathrm{off}}3$$

$$k_{\mathrm{obs}}(\mathrm{S}) = \frac{k\mathrm{C}4[\mathrm{Protein}]}{K_{\mathrm{d}}3 + [\mathrm{Protein}]} + k\mathrm{O}4$$

16. The Gibbs free energy (ΔG) of a reaction depends only on the free energy of the products (the final state) minus the free energy of the reactants (the initial state). The ΔG of a reaction is therefore independent of the path (or molecular mechanism) of the transformation. The free energy change, and therefore the K_{d}, must be the same for both pathways.

References

1. Licatalosi DD, Darnell RB (2010) RNA processing and its regulation: global insights into biological networks. Nat Rev Genet 11:75–87
2. Morris AR, Mukherjee N, Keene JD (2010) Systematic analysis of posttranscriptional gene expression. WIREs Syst Biol Med 2:162–180
3. Gerstberger S, Hafner M, Tuschl T (2014) A census of human RNA-binding proteins. Nat Rev Genet 15:829–845
4. Marinov KG, Williams BA, McCue K et al (2014) From single-cell to cell-pool transcriptomes: stochasticity in gene expression and RNA splicing. Genome Res 24:496–510
5. Gronland GR, Ramos A (2018) The devil is in the domain: understanding protein recognition of multiple RNA targets. Biochem Soc Trans 45:1305–1311
6. Lunde BM, Moore C, Varani G (2007) RNA-binding proteins: modular design for efficient function. Nat Rev Mol Cell Biol 8:479–490
7. Nicastro G, Taylor IA, Ramos A (2015) KH–RNA interactions: back in the groove. Curr Opin Struct Biol 30:63–70
8. Ciesielski GL, Hytönen VP, Kaguni LS (2016) Biolayer interferometry: a novel method to elucidate protein-protein and protein-DNA interactions in the mitochondrial DNA replisome. Methods Mol Biol 1351:223–231
9. Lou X, Egli M, Yang X (2018) Determining functional aptamer-protein interaction by biolayer interferometry. Curr Protoc Nucleic Acid Chem 67:7.25.1–7.25.15
10. Sultana A, Lee JE (2015) Measuring protein-protein and protein-nucleic acid interactions by biolayer interferometry. Curr Protoc Protein Sci 79:19.25.1–19.25.26
11. Cukier CD, Hollingworth D, Martin SR et al (2010) Molecular basis of FIR-mediated *c-myc* transcriptional control. Nat Struct Mol Biol 17:1058–1064
12. Hollingworth D, Candel AM, Nicastro G et al (2012) KH domains with impaired nucleic acid binding as a tool for functional analysis. Nucleic Acids Res 40:6873–6886
13. Candel AM, Hollingworth D, Nicastro G et al (2012) KH domains with impaired nucleic acid binding as a tool for functional analysis. FEBS J 279(Suppl 1):478–478
14. Nicastro G, García-Mayoral MF, Hollingworth D et al (2012) Noncanonical G recognition mediates KSRP regulation of let-7 biogenesis. Nat Struct Mol Biol 19:1282–1286
15. Nicastro G, Candel AM, Uhl M et al (2017) Mechanism of β-actin mRNA recognition by ZBP1. Cell Rep 18:1187–1199
16. Dagil R, Ball NJ, Ogrodowicz RW et al (2019) IMP1 KH1 and KH2 domains create a structural platform with unique RNA recognition and re-modelling properties. Nucleic Acids Res 47:4334–4348

17. Yisraeli JK (2005) VICKZ proteins: a multi-talented family of regulatory RNA-binding proteins. Biol Cell 97:87–96
18. Wächter K, Köhn M, Stöhr N, Hüttelmaier S (2013) Subcellular localization and RNP formation of IGF2BPs (IGF2 mRNA-binding proteins) is modulated by distinct RNA-binding domains. Biol Chem 394:1077–1090
19. Patel VL, Mitra S, Harris R et al (2012) Spatial arrangement of an RNA zipcode identifies mRNAs under post-transcriptional control. Genes Dev 26:43–53
20. Chao JA, Patskovsky Y, Patel V et al (2010) ZBP1 recognition of β-actin zipcode induces RNA looping. Genes Dev 24:148–158
21. Eccleston JF, Martin SR, Schilstra MJ (2008) Rapid kinetic techniques. Methods Cell Biol 84:445–477

Chapter 17

Analysis of Protein–DNA Interactions Using Surface Plasmon Resonance and a ReDCaT Chip

Clare E. M. Stevenson and David M. Lawson

Abstract

The recognition of specific DNA sequences by proteins is crucial to fundamental biological processes such as DNA replication, transcription, and gene regulation. The technique of surface plasmon resonance (SPR) is ideally suited for the measurement of these interactions because it is quantitative, simple to implement, reproducible, can be automated, and requires very little sample. This typically involves the direct capture of biotinylated DNA to a streptavidin (SA) chip before flowing over the protein of interest and monitoring the interaction. However, once the DNA has been immobilized on the chip, it cannot be removed without damaging the chip surface. Moreover, if the protein–DNA interaction is strong, then it may not be possible to remove the protein from the DNA without damaging the chip surface. Given that the chips are costly, this will limit the number of samples that can be tested. Therefore, we have developed a *Re*usable *D*NA *Ca*pture *T*echnology, or ReDCaT chip, that enables a single streptavidin chip to be used multiple times making the technique simple, quick, and cost effective. The general steps to prepare the ReDCaT chip, run a simple binding experiment, and analysis of data will be described in detail. Some additional applications will also be introduced.

Key words SPR, Protein, DNA, Reusable, Binding, Affinity, Kinetics

1 Introduction

SPR is an ideal technique for the study of protein–DNA interactions. It offers several advantages over more traditional in vitro methods such as electrophoretic mobility shift assays or nitrocellulose filter binding assays. It can provide real-time monitoring of binding and dissociation events, enabling the identification of binding and the measurement of affinity and kinetic parameters [1–3]. Furthermore, the technique is quick, sensitive, uses low amounts of protein, and is highly automatable.

A traditional approach to study protein–DNA interactions by SPR is to tether the DNA (the "ligand") via a biotin tag to a streptavidin (SA)-coated chip and then to flow the protein (the "analyte") over the top to monitor the interaction. However,

Tina Daviter et al. (eds.), *Protein-Ligand Interactions: Methods and Applications*, Methods in Molecular Biology, vol. 2263, https://doi.org/10.1007/978-1-0716-1197-5_17,

once the biotinylated DNA has been captured on the chip surface, the streptavidin–biotin interaction is so tight that it cannot be removed without damage to the chip. In addition, if the protein of interest binds tightly to the DNA, it may not be possible to remove the protein without damage to the bound DNA or to the chip surface. In practice, this means a new chip would be required to study each protein or DNA combination, thereby making this approach very costly.

To enable the routine and cost-effective use of SPR for the study of any protein-oligonucleotide pair, previous work [4–6] was extended, and this chapter will describe an indirect capture method that uses a ReDCaT chip [7]. This ReDCaT chip can be repeatedly used; any DNA sequence can be bound and then removed without compromising the chip, allowing multiple samples to be tested in a high throughput and automated manner using a single chip. The method can be used not only to identify binding events but also to obtain quantitative affinity and kinetic data [7, 8]. This chapter will describe, with technical details, how this method is implemented and how the results are analyzed.

2 Materials

2.1 Buffers and Reagents

Prepare all solutions using ultrapure water and analytical grade reagents. Prepare and store at room temperature and follow all waste disposal regulations (*see* **Note 1**).

1. Running buffer: 10 mM HEPES, pH 7.4, 150 mM NaCl, 3 mM Ethylenediaminetetraacetic acid (EDTA), 0.05% (v/v) Tween 20. Weigh 2.383 g HEPES (238.3 g/mol), 8.766 g NaCl (58.4 g/mol), 0.877 g EDTA (292.2 g/mol) and add to a graduated glass beaker. Add water to a volume of approximately 750 ml and then add 0.5 ml Tween 20. Mix and adjust pH to 7.4 using 5 M NaOH then make up to 1000 ml with water.
2. 0.5 M NaCl: Weigh 0.292 g NaCl and make up to 10 ml with water.
3. Regeneration solution: 1 M NaCl and 50 mM NaOH. Weigh 5.844 g NaCl and 0.2 g NaOH pellets (40.0 g/mol) and make up to 100 ml with water.

2.2 Instrument and Chip and Protein Sample (See Note 2)

1. Instrument and Chip: SPR Biacore T200 (GE Healthcare) and Series S SA chip (GE Healthcare).
2. Protein samples to be tested: Dilute the purified protein samples in running buffer to 1.0 and 0.1 μM concentration (*see* **Note 3**). For an initial binding check in duplicate, 100 μl of each concentration will be sufficient.

Table 1
DNA sequences. Examples of the DNA that would be required to prepare the ReDCaT chip and to test one sequence of DNA

Oligo	Length (bases)	Sequence 5′ to 3′
ReDCaT linker	20	Biotin- GCAGGAGGACGTAGGGTAGG
Test DNA Forward	36	ACTCCAATACTTGAACTCTCAATCTTTACGTGCCGT
Test DNA Reverse	56	ACGGCACGTAAAGATTGAGAGTTCAAGTATTGGAGT CCTACCCTACGTCCTCCTGC

2.3 DNA

The DNA should be ordered from the supplier as a desalted sample, with standard purification on the smallest synthesis scale.

1. ReDCaT linker: The 20-nucleotide ReDCaT DNA linker (*see* **Note 4**) is ordered with a biotin label added to the 5′ end. The sequence of the ReDCaT linker is shown in Table 1. Dilute the ReDCaT linker to 100 nM concentration in running buffer, and this can be stored in aliquots at −80 °C until required.
2. Test DNA: Forward (F) and reverse (R) strands (*see* **Note 5**). The forward strand comprises just the sequence of DNA to be tested, whilst the reverse strand includes an additional 20 nucleotides attached to the 3′ end, which is the complement to the ReDCaT linker (sequence 5′ CCTACCC TACGTCCTCCTGC 3′). Examples of forward and reverse DNA samples are shown in Table 1. Dilute the F and R DNA samples to 100 μM in running buffer or water. Mix 55 μl F strand and 45 μl R strand and heat to 95 °C for 10 min and allow to cool. This gives a stock at 45 μM for the double-stranded DNA (dsDNA; *see* **Note 6**). Prior to use, dilute to 1 μM (22 μl dsDNA stock +978 μl running buffer). This should be done for all the DNA samples to be tested. Store both the dsDNA stocks and the dilutions at −80 °C until required.

3 Methods

The method uses a biotinylated single-stranded (ss) DNA ReDCaT linker, which is permanently bound to a standard SA chip. Subsequently, a dsDNA oligomer containing the sequence of interest and bearing an overhang that is complementary to the linker is captured on the chip through hybridization. At the end of the experiment, the captured oligonucleotide is stripped from the linker by denaturation to regenerate the chip. Figure 1 shows an illustration of the ReDCaT method and a representative sensorgram. The technical details to run the method are presented below.

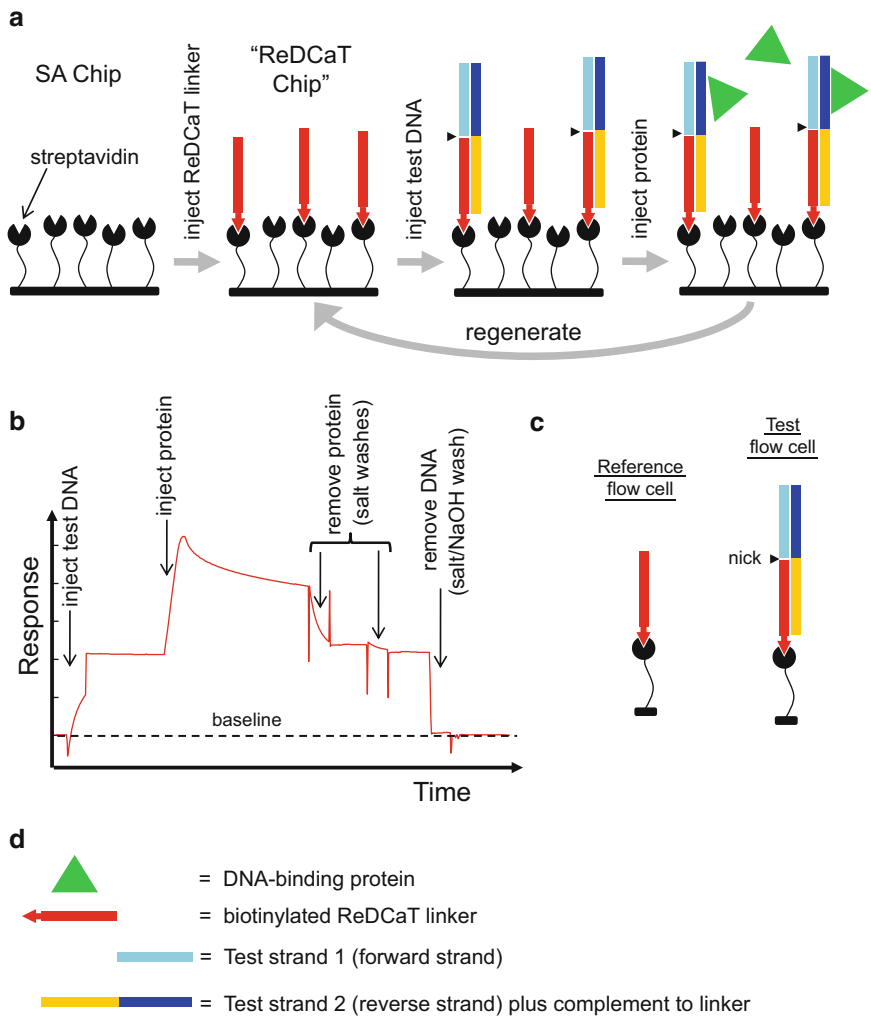


Fig. 1 The ReDCaT methodology. (**a**) Procedure for creating, using, and regenerating the ReDCaT chip, specifically illustrating the events taking place in the test flow cell. (**b**) A typical sensorgram for the test flow cell of the ReDCaT chip showing the responses observed during its use and regeneration. Note that the response returns to the original baseline after stripping off the test DNA. (**c**) The composition of bound DNA in the reference and test flow cells prior to injecting protein in the ReDCaT experiment. (**d**) Key to the macromolecular components illustrated in the other three panels. Figure modified from Fig. 2 originally published in Nucleic Acids Res 41 (14):7009–7022

For all experiments, the SPR instrument is docked with a Series S streptavidin chip and primed with running buffer. This instrument creates four flow cells on the chip. The method described below uses only two of these flow cells. Flow cell 1 will be the reference (FC_{ref}) and flow cell 2 the test (FC_{test}). For this instrument and chip combination, flow cells 3 and 4 remain available for use in other experiments (*see* **Note 7**). In addition, the first time a chip is docked, a normalization with 70% glycerol should be carried

out. This can be selected when the method is run. For all experiments, tubes should be loaded with the appropriate amount of solution and placed in the rack as detailed in the Biacore T200 Control Software.

3.1 Preparation of the ReDCaT Chip

1. To remove any unconjugated streptavidin, both flow cells (FC_{ref} and FC_{test}) are washed using three injections of 1.0 M NaCl, 50 mM NaOH for 60 s, each followed by buffer for 60 s (all at 10 μl/min).
2. Then inject the ReDCaT linker (100 nM) at 5 μl/min over flow cell 2 (FC_{test}) only to give a relatively high immobilization level (approx. 500 RU) (*see* **Note 8**). This biotinylated ssDNA ReDCaT linker is now permanently immobilized via the streptavidin, thereby creating the ReDCaT chip. This chip can be used repeatedly for many experiments and can be stored (when not in use) in running buffer at 4 °C. We have found that the ReDCaT chips will keep for over 1 year and still be functional.

3.2 A General Protocol to Screen for Protein:DNA Interactions

The flow rate is set at 30 μl/min unless stated otherwise.

1. Inject test DNA (1.0 μM) over flow cell 2 (FC_{test}) at a flow rate of 10 μl/min for 60 s, followed by buffer for 60 s.
2. Then inject the test protein at the required concentration (or buffer-only control) over both flow cells (FC_{ref} and FC_{test}) for 60 s, followed by buffer for 60 s.
3. If required, inject a solution of 0.5 M NaCl for 60 s over both flow cells (FC_{ref} and FC_{test}), followed by buffer for 60 s. This step is optional, but is useful to examine the effect of NaCl on the protein–DNA interaction (*see* **Note 9**).
4. Inject regeneration solution over both flow cells (FC_{ref} and FC_{test}) at a flow rate of 10 μl/min to remove the test DNA (together with any remaining bound protein) to leave only the ReDCaT linker bound to the chip. The response should return to the level prior to **step 1** (*see* Fig. 1b). Then flow buffer over both flow cells for a further 60 s.

Steps 1–4 can be repeated as many times as required (e.g., with different DNA and protein samples), in an automated fashion without any further user intervention. This is called a "cycle" and each cycle takes approximately 15 min. Ideally the first cycle should be a conditioning cycle where DNA is captured on the second flow cell (FC_{test}) and test protein is run over both flow cells and then both flow cells are regenerated. It is also recommended to run a negative control, which consists of a piece of DNA at the same length as the test DNA, but with a randomized sequence (*see* **Note 10**). All samples should be tested at least twice and at more than one protein concentration, and buffer-only samples should be used

for double referencing. When the experiment is completed, the regenerated chip can be removed from the instrument and stored in buffer at 4 °C until required again.

3.3 Analysis of Results

To readily compare results across multiple cycles, the data can be normalized by correcting for different capture levels, lengths of DNA, and different protein samples. In SPR, binding events are recorded as Response Units (RU), and it is important to compare this response to the expected theoretical response for the interaction. This is done using the following equation (*see* **Note 11**):

$$R_{\max} = \frac{Mwt_{\text{Protein}}}{Mwt_{\text{DNA}}} \cdot R_L \cdot n \cdot 0.78 \quad (1)$$

where $R_{\max}$ is the theoretical maximal response at saturation, Mwt is the molar mass, R_L is the DNA capture response, and n is the binding stoichiometry. Then, the percentage of $R_{\max}$ measured upon protein binding is calculated as follows:

$$\%R_{\max} = \frac{RU_{\max}}{R_{\max}} \cdot 100 \quad (2)$$

where $RU_{\max}$ is the measured maximal response.

3.4 A Worked Example with Explanation of How Some Typical Results Are Analyzed

Table 2 shows typical results that could be obtained from testing the binding of two different proteins against one test and one randomized DNA sequence. For further examples of published results please *see* references [7–16].

In this example, protein A binds well to the test DNA at 85% of the theoretical $R_{\max}$ and there is very little binding to the randomized sequence. For simplicity, in this illustrative example, only one concentration of protein was used, but ideally two or more concentrations should be tested. Examination of the sensorgrams and comparison of percentage $R_{\max}$ values should give an indication that saturation of binding has occurred. These results suggest that the protein is binding to the DNA as a monomer and the binding is sequence-specific as there is very little binding to the randomized sample.

For protein B, the percentage $R_{\max}$ is approaching 200%, suggesting that this protein is binding to the DNA as a dimer. Very little binding is also seen to the randomized sequence.

Analysis of the data in this way allows the amount of captured DNA to be corrected for. Over long experiments, with many cycles, the capture level can sometimes gradually decrease. However, with normalization, all the results can be meaningfully compared, and multiple replicates should give very similar values. Similarly, different lengths of DNA will give different responses for the same capture level, but the results can be directly compared after normalization.

Table 2
Example of the analysis of some typical results. A hypothetical example is given for two different proteins (A and B) each being tested against two different DNA sequences. DNA sequence 1 corresponds to a size of 20 bases and DNA sequence 2 to a size of 36 bases. Each DNA sequence also has the additional 20 bases of the ReDCaT linker. The theoretical R_{max} is estimated using Eq. 1. The percentage of the theoretical R_{max} can then be calculated using Eq. 2

Protein	Protein monomer mass (Da)	DNA	DNA mass (Da)	DNA captured (*RU*)	Theoretical R_{max} (calculated)	Protein bound (*RU*)	% R_{max}
A	20,000	Test 1	18,226	416	356	302	85%
A		Random 1	18,226	410	351	5	1%
B	60,000	Test 2	28,113	452	752	1478	197%
B		Random 2	28,113	444	739	25	3%

3.5 Further Possible Experiments and Uses

The method described gives the technical details to run one cycle with the ReDCaT chip. This can be written as a "method" to run automatically, which enables any number of DNA sequences, DNA lengths, and protein samples and concentrations to be tested in series. This makes the method amenable to a wide range of experiments [7–16].

Initially the method might be used to identify whether a DNA fragment contains a sequence that a particular protein binds to. This might be to verify a binding site identified by another method (like CHIP-seq) or could be used to screen a region of DNA for a potential binding site. Once binding has been identified, the smallest length of DNA that is sufficient for the interaction can be determined. This involves truncating the binding site from each end and testing to see when the interaction is lost. How to design this kind of "footprinting by SPR" experiment is detailed in [7].

The method has also been used to quantify interactions and to obtain kinetic and affinity parameters. Moreover, it can test the effects of site-directed mutations, or it can be used to rank the binding of a single protein to a series of different consensus sequences. For weaker interactions, the test DNA can be retained on the chip and at least five concentrations of protein ranging from 0.1 to 10 times the estimated affinity can be tested in duplicate. For stronger interactions, the DNA with any protein still bound must be removed and new DNA reloaded for each cycle. These experiments should be designed and analyzed using the same principles as for any other affinity or kinetic SPR experiment.

If a ligand is thought to interrupt the interaction, this can also be investigated using the ReDCaT method. The ligand of interest can be added in increasing concentrations and the effect on binding

observed. Alternatively, after the protein has bound to the DNA, a ligand can be injected over the chip to assess whether the interaction is affected.

Most of the interactions studied to date using this method have been protein:DNA interactions, but it is also possible to use the method to study protein:RNA interactions. In this case, the test RNA would have the DNA ReDCaT linker attached creating a RNA:DNA hybrid for the reverse strand.

To summarize, the chip can be used hundreds of times, making the SPR technique for testing protein:DNA interactions attractive. Any protein or DNA sequence can be quickly tested using very small quantities of protein. The results are highly reproducible, and quantitative data can be obtained. Although the ReDCaT method requires the use of an expensive SPR instrument, these are often available through access to shared scientific services at many institutions, with experts available to assist users. Thus, the technique should be considered as a complement to any study that analyzes protein:DNA interactions, irrespective of the expertise of the investigator.

4 Notes

1. A wide range of running buffers can be used, but it is important to pick one that contains the ingredients necessary for the interaction to occur. For example: if the interaction requires the presence of magnesium, this should be added to the running buffer. The buffer given in the method is a good starting point and, in our lab, has proved successful for a wide range of different protein samples.
2. This method has been developed to be implemented using the Biacore T200 SPR instrument and the Series S SA chip (GE Healthcare). However, it could easily be adapted to run with any SPR instrument and compatible streptavidin chip.
3. The protein should be purified and concentrated before dilution into the running buffer. For the first experiment, it is recommended to use two protein concentrations (1 and 0.1 μM). For subsequent experiments, a range of concentrations may be tested. The protein concentration should be calculated assuming the protein is monomeric.
4. The ReDCaT linker used is 20 nucleotides long and is designed to have no secondary structure or any tendency to anneal with itself, as assessed by the Sigma-Aldrich website (www.sigma-gonosys.com/calc/DNACalc.asp). If there is any suspicion that the protein binds to the ReDCaT linker, this should be tested using just the linker or a randomized sequence plus linker. If the linker is found to contain a specific binding site,

a different linker sequence and/or length could be used, but this would need to be optimized with positive and negative controls. However, this ReDCaT linker has been used successfully to test a wide range of different protein samples without any such problems.

5. The test DNA can be of any length. However, we have found that sequences of 20–40 nucleotides in length tend to work well and are cheap to order. Ideally the sequence needs to be long enough to include the whole binding site. If the DNA site is not known and a longer segment of DNA requires testing to locate a binding site, this can be easily fragmented into shorter segments for testing. This is discussed extensively in [7]. The reverse stand is always 20 nucleotides longer as it contains the additional nucleotides to hybridize to the ReDCaT linker attached to the chip.
6. A slight excess of the forward strand is used to ensure all the DNA that binds to the chip will be dsDNA (since the F strand does not contain the additional complement to the ReDCaT linker, it will not bind to the chip surface).
7. In this example, the reference flow cell is kept blank. This is the simplest protocol, and it will reveal any nonspecific binding of the protein to the chip. However, there are other options, including placing the ReDCaT linker (with or without its complement sequence, but no test DNA) on the reference flow cell, or even to immobilize a randomized test DNA sequence via the ReDCaT linker.
8. In the Biacore control software, the user can request that the immobilization step should aim for a required response. This is particularly useful as the instrument will estimate the length of injection time to achieve the required response. If this option is unavailable, multiple short injections should be carried out until the required immobilization level is achieved.
9. Proteins that bind weakly to DNA will be more easily removed than those that bind more tightly by the 0.5 M NaCl wash. If a ligand or inhibitor is known to interrupt the interaction, this could be used instead. It is possible to do the same experiment, but with different ligands or inhibitors as a regeneration solution, to test the effect of these on the interaction.
10. It is important to use a negative control in any analysis. Routinely, one of the test DNA sequences is randomized and used as the negative control. This should be the same length as the test sequence. There are a variety of online tools that will randomize a DNA sequence, such as: http://www.bioinformatics.org/sms2/shuffle_dna.html

11. When the ligand is DNA, it has been suggested that the result needs to be multiplied by a factor of 0.78 because the response associated with nucleic acid binding to the surface is not the same as that for a protein of equivalent mass [1, 3]. Unless it is already known, it is best to assume that the stoichiometry is 1:1 and the concentrations of protein should be calculated based on the molecular weight of the monomer.

Acknowledgments

The work was carried out using the John Innes Centre Biophysical Analysis Facility with funding from the Biotechnology and Biological Sciences Research Council (UK) Institute Strategic Programme Grant BB/P012523/1. We would like to thank Julia Mundy for critically reading the chapter.

References

1. Buckle M (2001) Surface plasmon resonance applied to DNA-protein complexes. Methods Mol Biol 148:535–546. https://doi.org/10.1385/1-59259-208-2:535
2. Majka J, Speck C (2007) Analysis of protein-DNA interactions using surface plasmon resonance. Adv Biochem Eng Biotechnol 104:13–36
3. Stockley PG, Persson B (2009) Surface plasmon resonance assays of DNA-protein interactions. Methods Mol Biol 543:653–669. https://doi.org/10.1007/978-1-60327-015-1_38
4. Buckle M, Williams RM, Negroni M, Buc H (1996) Real time measurements of elongation by a reverse transcriptase using surface plasmon resonance. Proc Natl Acad Sci U S A 93 (2):889–894
5. Katsamba PS, Park S, Laird-Offringa IA (2002) Kinetic studies of RNA-protein interactions using surface plasmon resonance. Methods 26 (2):95–104. https://doi.org/10.1016/S1046-2023(02)00012-9
6. Park S, Myszka DG, Yu M, Littler SJ, Laird-Offringa IA (2000) HuD RNA recognition motifs play distinct roles in the formation of a stable complex with AU-rich RNA. Mol Cell Biol 20(13):4765–4772
7. Stevenson CE, Assaad A, Chandra G et al (2013) Investigation of DNA sequence recognition by a streptomycete MarR family transcriptional regulator through surface plasmon resonance and X-ray crystallography. Nucleic Acids Res 41(14):7009–7022. https://doi.org/10.1093/nar/gkt523
8. Tran NT, Stevenson CE, Som NF et al (2018) Permissive zones for the centromere-binding protein ParB on the *Caulobacter crescentus* chromosome. Nucleic Acids Res 46 (3):1196–1209. https://doi.org/10.1093/nar/gkx1192
9. Bekiesch P, Forchhammer K, Apel AK (2016) Characterization of DNA binding sites of RokB, a ROK-family regulator from S*treptomyces coelicolor* reveals the RokB regulon. PLoS One 11(5):e0153249. https://doi.org/10.1371/journal.pone.0153249
10. Bekiesch P, Franz-Wachtel M, Kulik A et al (2016) DNA affinity capturing identifies new regulators of the heterologously expressed novobiocin gene cluster in *Streptomyces coelicolor M512*. Appl Microbiol Biotechnol 100 (10):4495–4509. https://doi.org/10.1007/s00253-016-7306-1
11. Campilongo R, Fung RKY, Little RH et al (2017) One ligand, two regulators and three binding sites: how KDPG controls primary carbon metabolism in *Pseudomonas*. PLoS Genet 13(6):e1006839. https://doi.org/10.1371/journal.pgen.1006839
12. Fernandes GC, Hauf K, Sant'Anna FH et al (2017) Glutamine synthetase stabilizes the binding of GlnR to nitrogen fixation gene operators. FEBS J 284(6):903–918. https://doi.org/10.1111/febs.14021

13. Greive SJ, Fung HK, Chechik M et al (2016) DNA recognition for virus assembly through multiple sequence-independent interactions with a helix-turn-helix motif. Nucleic Acids Res 44(2):776–789. https://doi.org/10.1093/nar/gkv1467
14. Hauf K, Kayumov A, Gloge F, Forchhammer K (2016) The molecular basis of TnrA control by glutamine synthetase in *Bacillus subtilis*. J Biol Chem 291(7):3483–3495. https://doi.org/10.1074/jbc.M115.680991
15. Traore DAK, Wisniewski JA, Flanigan SF et al (2018) Crystal structure of TcpK in complex with oriT DNA of the antibiotic resistance plasmid pCW3. Nat Commun 9(1):3732. https://doi.org/10.1038/s41467-018-06096-2
16. Wang T, Zhao X, Shi H et al (2018) Positive and negative regulation of transferred nif genes mediated by indigenous GlnR in Gram-positive *Paenibacillus polymyxa*. PLoS Genet 14(9): e1007629. https://doi.org/10.1371/journal.pgen.1007629

Chapter 18

Characterization of Protein–Nucleic Acid Complexes by Size-Exclusion Chromatography Coupled with Light Scattering, Absorbance, and Refractive Index Detectors

Ewa Folta-Stogniew

Abstract

Size-exclusion chromatography (SEC) coupled with multiangle light scattering detection (SEC/MALS) enables determination of the molecular weight, oligomeric state, and stoichiometry of protein–nucleic acid complexes in solution. Often such complexes show anomalous behavior on SEC, thus presenting a challenge in determination of molecular weight and stoichiometry based solely on the elution position from SEC. In contrast to analytical ultracentrifugation, the SEC/MALS analysis is not affected by the shape of the complex. Here we describe the use of SEC/MALS for characterization of the stoichiometry of the complex between the reverse transcriptase (RT) domain from group II intron–maturase from *Eubacterium rectale* and intron RNA, and for monitoring protein dimerization that is driven by interaction between single-stranded DNA upstream of the P1 promoter, known as FUSE and FUSE binding protein-interacting repressor (FIR).

Key words SEC/MALS, Laser light scattering, Size-exclusion chromatography (SEC), Molecular weight, Protein–nucleic acid complex, Stoichiometry, Oligomeric state

1 Introduction

Light scattering is a spectroscopic technique for determination of the molecular weight of biopolymers in solution. Typically, molecular weight of protein complexes is estimated from size-exclusion chromatography (SEC) or from analytical ultracentrifugation (AUC); both techniques are limited by the influence of the shape of the complex on the measurement. Similarly to sedimentation in the analytical ultracentrifuge (AUC), light scattering measurement is an absolute method for determination of molecular weight and does not rely on reference standards.

Tina Daviter et al. (eds.), *Protein-Ligand Interactions: Methods and Applications*, Methods in Molecular Biology, vol. 2263, https://doi.org/10.1007/978-1-0716-1197-5_18,

1.1 Basic Principles of Light Scattering Measurement

The amount of light scattered is directly proportional to the product of the weight-average molar mass (molecular weight), M_w, and the macromolecule (solute) concentration, i.e., static light scattering ~ $M_w \cdot c$. Based on Zimm's formalism, the Rayleigh–Debye–Gans light scattering model for dilute polymer solutions [3] predicts that;

$$\frac{K^* c}{R(\Theta)} = \frac{1}{M_w \cdot P(\Theta)} + 2A_2 c \tag{1}$$

where c is the w/v concentration of a macromolecule (g/l); M_w is its weight-average molar mass (g/mol) also called molecular weight; $R(\Theta)$ is the excess intensity of scattered light at an angle Θ, i.e., the total scattering intensity, I_Θ, corrected for the contribution from the solvent; R_g is the radius of gyration (also referred to as root mean square radius $< r_g^2 >^{1/2}$); A_2 is a second virial coefficient (ml·mol/g^2); K^* is an optical parameter equal to $4\pi^2 n^2 (dn/dc)^2/(\lambda^4 N_A)$; n is the solvent refractive index and dn/dc is the refractive index increment of solute (ml/g); N_A is Avogadro's number; and λ is the wavelength of the scattered light in a vacuum.

The Rayleigh–Debye–Gans light scattering approximation described by Eq. 1 is valid for particles whose maximum dimension is smaller than λ.

The function $P(\Theta)$ describes the angular dependence of scattered light. The expansion of $1/P(\Theta)$ to the first order gives;

$$1/P(\Theta) = 1 + \left(16\pi^2/3\lambda^2\right) R_g^2 \sin^2(\Theta/2) \tag{2}$$

A plot of $K^* c/R(\Theta)$ vs. $\sin^2(\Theta/2)$, a Zimm plot, yields a curve whose intercept gives $(M_w)^{-1}$ and whose slope at low concentration gives the radius of gyration, R_g, which characterizes particle dimensions independently of particle shape. Angular dependence of scattered light is observed for particles that are of a size that corresponds to at least ~1/20th of the incident light. For typically used lasers of 633 or 690 nm, radii smaller than ~12 nm cannot be estimated reliably from static light scattering measurement; thus, the angular dependence of scattered light described by Eq. 2 is negligible for proteins <500 kDa and the data can be even collected at a single angle.

1.2 Light Scattering Coupled with Chromatography

Coupling of light scattering measurement with a fractionation step creates a versatile system for determination of molecular weight, and several approximations can be applied during analysis of light scattering data collected in a chromatographic mode. The second virial coefficient term ($2A_2c$) in Eq. 1 can be neglected when $2A_2cM_w \ll 1$, and such an approximation is valid during a typical analysis of SEC results where the concentration in the eluting peak is usually below ~0.5 mg/ml.

Review articles have been published on the theory and application of static light scattering combined with SEC (SEC/MALS) for determination of the molecular weight of proteins in solution [3–7] and for analysis of protein complexes [5, 8], including protein interactions with DNA [2, 9–14] and RNA [1, 12, 15, 16].

The SEC/MALS analysis utilizes three detectors: an absorbance (UV) detector, a static light scattering (LS) detector and a refractive index (RI) detector, which are placed in series with a SEC column. The UV detector monitors absorbance at a selected wavelength, the RI detector monitors changes in the refractive index, and the LS detector records the excess of scattered light; SEC serves solely as a fractionation step. Since static light scattering provides only the *weight-average* molar mass of the species in solution, the SEC separation plays an integral role in the overall analysis, although the elution from SEC does not need to correlate with the molecular weight of the species being studied. The SEC/-MALS system is validated in a buffer of choice by analyzing protein and nucleic acid standards. The computation of molecular weight is based on the theory that the excess of scattered light is proportional to the product of the molecular weight and concentration [3, 6]. Scattered light is measured by the light scattering detector; concentration is measured by the RI and UV detectors. The responses from the three detectors are processed by the software to calculate the molecular weight of the eluting macromolecule [6]. The accuracy of molecular weight measurement by SEC/-MALS is ~ ±3% [6] and in most cases allows determination of the oligomeric state of the protein and stoichiometry of protein–nucleic acid complexes. In addition, the responses collected by the refractometer and the absorbance detector allow an "online" measurement of extinction coefficient of the eluting material [2, 17], which aids in computation of protein–nucleic acid stoichiometry.

The examples used in this chapter result from the SEC/MALS analyses of two well-characterized protein–nucleic acid complexes with known crystal structures: reverse transcriptase (RT) domain from group II intron–encoded protein from *Eubacterium rectale* (*E.r.* RT domain) [1] and its complex with intron RNA (D4A) and of FUSE binding protein-interacting repressor (FIR) that binds to FUSE DNA [2]. These results presented in Figs. 1, 2 and 3 and summarized in Tables 1, 2 and 3 illustrate two different scenarios observed when studying protein complexes via SEC/MALS: slow dissociating complex that retains molecular weight despite changes in concentration (*E.r.* RT domain; Fig. 1b, c), and fast dissociating complexes that dissociate due to dilution during the fractionation step (FIR; Fig. 2b, c) [18]. Additionally, these examples illustrate clearly that SEC elution volume alone is not a reliable method to determine the mass and oligomeric state of these complexes because their elution from SEC does not correlate with their size

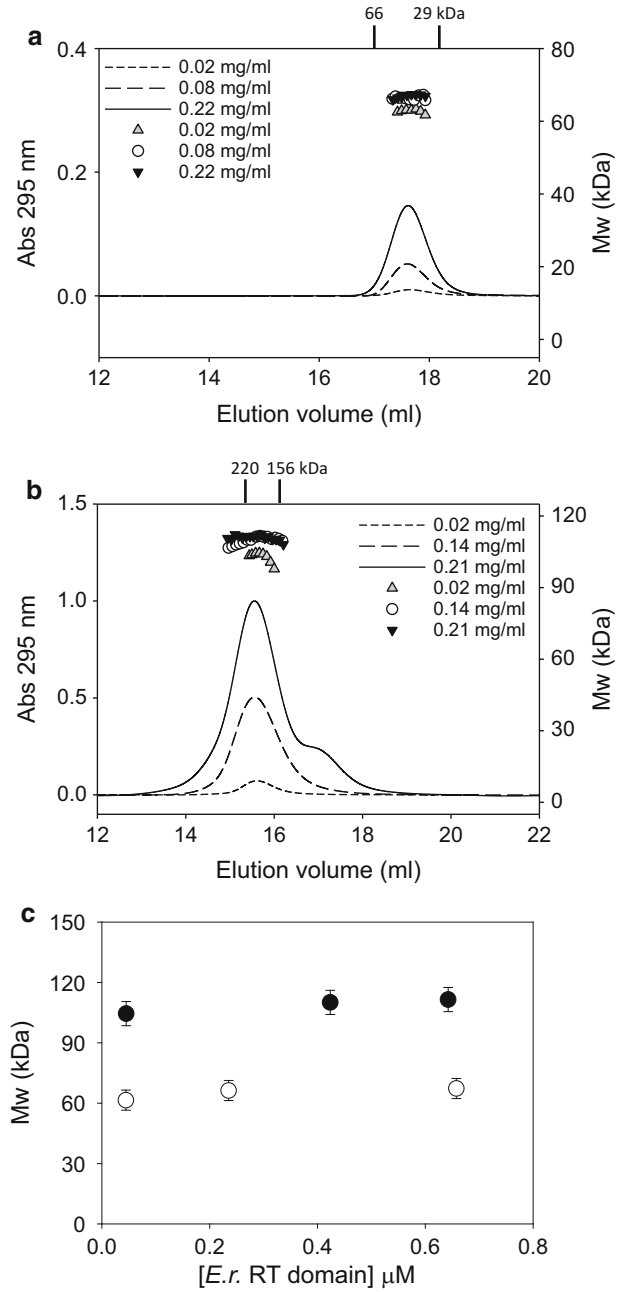


Fig. 1 *E.r.* RT forms a dimer in solution in the absence or presence of D4A RNA. Molecular weight distribution plot from SEC/MALS data for *E.r.* RT (panel **a**) and *E.r.* RT:D4A RNA complex (panel **b**). Lines correspond to UV traces monitored at 295 nm (left axis); concentrations at the apex of the eluting peaks are listed in the legend (in mg/ml); the M_w are plotted as circles, or triangles (right *y* axis). For clarity, only every tenth result of molecular weight measurement across the eluting peak is plotted. Elution position of globular protein standards: beta-amylase (220 kDa) and aldolase from rabbit muscles (156 kDa) are marked in panel (**b**). Weight-average M_ws determined from SEC/MALS analyses are plotted as a function of the concentration at the apex of the eluting peak; filled circles for *E.r.* RT-D4A RNA complex and open circles for *E.r.* RT protein alone (panel **c**) to illustrate that *E.r.* RT forms a dimer that binds one D4A RNA per monomer

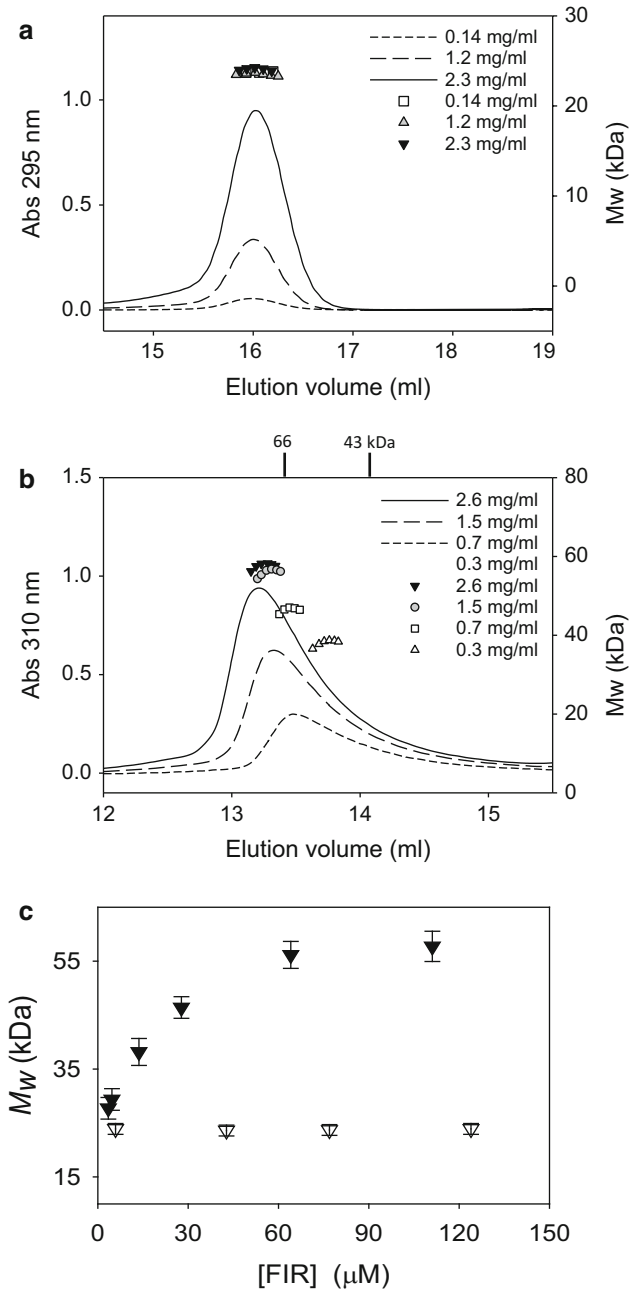


Fig. 2 Dimerization of FIR upon FUSE DNA binding. FIR protein remains monomeric in solution and forms a dimer upon DNA binding. Molecular weight distribution plot from SEC/MALS data for FIR protein (panel **a**) and FIR-FUSE DNA complex (panel **b**). Lines correspond to UV traces monitored at 310 nm (left *y* axis); concentrations at the apex of the eluting peaks are listed in the legend (in mg/ml); the M_ws are plotted as circles, square, or triangles (right *y* axis). For clarity, only every tenth result of molecular weight measurement across the eluting peak is plotted. Elution position of globular protein standards: bovine serum albumin (66 kDa) and ovalbumin (43 kDa) are marked in panel (**b**). Weight-average M_ws determined from SEC/MALS analyses are plotted as a function of the concentration at the apex of the eluting peak; filled triangles for FIR–DNA complex and open triangles for FIR protein alone (panel **c**)

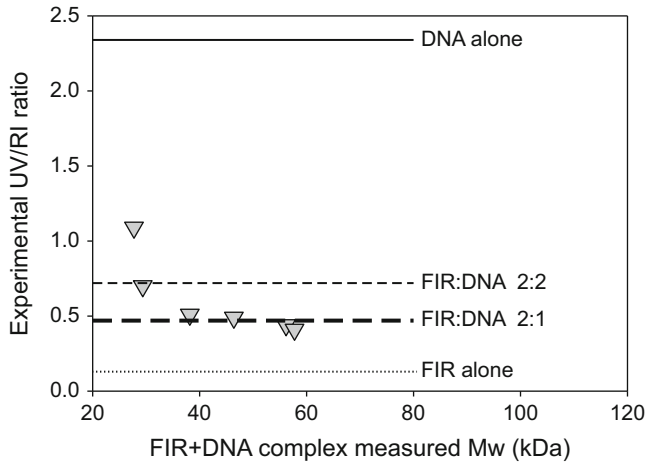


Fig. 3 FIR protein dimerizes upon binding to one FUSE DNA. The ratio of the signals recorded by the absorbance detector and the refractometer, UV/RI ratio, is plotted as a function of the measured M_w to illustrate that upon dimerization, the FIR dimer is associated with one FUSE DNA and forms FIR–DNA complex with 2:1 stoichiometry. The UV/RI ratio measured for FIR protein and FUSE DNA are plotted as dotted and solid lines, respectively; the UV/RI ratios expected for 2:1 and 2:2 FIR:DNA stoichiometries are plotted as dashed lines as indicated, confirming dimerization of FIR on a single FUSE DNA (*see* Table 3 for details)

Table 1
Results of SEC/MALS analysis for *E.r.* RT domain alone (33 kDa) and in complex with D4A RNA (21 kDa); *E.r.* RT forms a dimer in solution in the absence or presence of D4A RNA and binds one D4A RNA per monomer

Sample	Concentration		Observed M_w (kDa)
	mg/ml	μM	
E.r. RT	0.015	0.45	62
	0.078	2.36	66
	0.218	6.61	67
	Average		65
	St Dev		3
E.r. RT + RNA	0.015	0.45	105
	0.141	4.28	110
	0.213	6.45	111
	Average		109
	St Dev		3

due to their elongated shape; the 58 kDa complex of FIR:FUSE DNA elutes ahead of 66 kDa globular standard (Fig. 2b), and 110 kDa complex of *E.r.* RT:D4A RNA elutes ahead of 156 kDa aldolase (Fig. 1b). The signals recorded by the UV detector and refractometer are also used to discriminate between two possible protein–nucleic acid stoichiometries, for which the predicated

Table 2
Results of SEC/MALS analysis for FIR (23.5 kDa) alone and in complex with FUSE DNA (8.5 kDa); FIR remains monomeric in solution and forms a dimer upon binding a single FUSE DNA

Sample	Concentration		Observed M_w (kDa)
	mg/ml	μM	
FIR	0.14	5.9	24
	1.0	43	24
	1.8	77	24
	2.9	124	24
FIR+DNA	0.32	14	38
	0.65	28	46
	1.5	64	56
	2.6	111	58

Table 3
Determination of stoichiometry for *E.r. RT*-D4A RNA and FIR-FUSE DNA complexes from SEC/MALS analysis. Measured and predicated M_w (kDa) and UV/RI ratios for possible stoichiometries for *E.r. RT*-D4A RNA and FIR-FUSE DNA complexes. *E.r.* RT monomer: 33 kDa; UV/RI ratio: 0.95; D4A RNA: 21 kDa; UV/RI ratio: 13. FIR monomer: 23.5 kDa; UV/RI ratio: 0.13; FUSE DNA: 8.5 kDa; UV/RI ratio 2.3

Complexes	Predicted M_w (kDa)	Observed M_w (kDa)	Difference (%)	Predicted UV/RI ratio[a]	Observed UV/RI ratio[b]	Difference (%)
FIR+DNA (2:1) complex	56	58	−3%	0.48	0.41	17%
FIR+DNA (2:2) complex	64	58	10%	0.72	0.41	76%
E.r. RT + RNA (2:2) complex	108	109	−1%	5.6	6.9	−19%

[a] Expected UV/RI ratios are computed as weight averaged values.
[b] UV signal recorded at 295 nm for E.r. RT and at 310 nm for FIR.

masses are within the experimental error of the measured value (as shown for FIR:FUSE DNA complex in Table 3 and Fig. 3).

2 Materials

2.1 SEC/MALS System

2.1.1 Instruments (See Note 1)

1. HPLC pump capable of delivering a flow rate of 0.3 to 1.0 ml/min with low pulsation (Waters Alliance 2695 or Agilent 1200 HPLC).
2. Inline solvent filter with 0.1 μm pore size installed between the pumps and SEC column (*see* **Note 2**).
3. Inline sample filter with 2 μm pore size installed between the injector and SEC column (*see* **Note 2**).

Table 4
Sample and column requirements for SEC/MALS analyses

Column[a]	Optimal amount of protein μg [10^{-6} g]			
	M_w >200 kDa	M_w 40–200 kDa	M_w 10–40 kDa	M_w <10 kDa
Superose 6 HR 10/300	**50**	**50–100**	Not suitable	Not suitable
Superdex 200 HR 10/300	50	**50–100**	**100–200**	Not suitable
Superdex 75 HR 10/300	Not suitable	50–100	**100–200**	Not suitable
Superdex peptide HR 10/300	Not suitable	Not suitable	Not suitable	**400–800**

[a]The Superose/Supredex columns from the HR and GL 10/300 series were extensively tested by the author in a variety of buffer conditions (including buffers supplemented with various detergents or denaturants) as suitable for the SEC/MALS analyses; shown in bold type are the optimal column matches for a given M_w range (as expected for a given oligomeric state)

4. UV detector for "inline" use (*see* **Note 3**).
5. Refractive index detector for "inline" use (Wyatt OptiLab rEX; *see* **Note 4**).
6. Static light scattering detector for "inline" use (Wyatt DAWN EOS or DAWN HELEOS II).
7. Computer and software for data collection; preferably software capable of collecting data from all three detectors simultaneously (Wyatt ASTRA software).

2.1.2 SEC Column

1. High-performance size-exclusion chromatography column appropriate for fractionating the samples analyzed (GE Healthcare Superdex 200, HR 10/300); use Table 4 guidelines regarding column selection; *see* **Note 5**).

2.2 Reagents and Supplies

1. Filters with low-protein-binding properties and pore size of 0.22 or 0.1 μm (*see* **Note 6**).
2. Protein and nucleic acid standards: at least three proteins with M_w spanning the range of the expected M_w of the sample to be analyzed (*see* **Note 7**); for nucleic acids, analyzing the exact piece of DNA or RNA that is used for complex formation is sufficient.
3. Buffer: aqueous buffer compatible with the SEC column requirements; for most media used for SEC, 150 mM salts need to be present to prevent electrostatic interactions with the column's matrix; 0.01% of sodium azide is routinely used as bacteriostatic. The buffer should be filtered through a 0.1 μm filter. Buffers with dithiothreitol (DTT) should be replaced every 24 h. Typically, used phosphate buffer saline (PBS) or other routinely used buffers should be replaced when the noise

in LS signal has increased twofold when compared to the noise recorded when the SEC/MALS system has been initially equilibrated.

2.3 Sample

1. 50–500 μg of protein, nucleic acids, and protein–nucleic acid complex samples (*see* Table 4 for guidelines regarding optimal sample amounts) in a volume that corresponds to ~3% of the total volume of the SEC column (*see* **Note 8**).

3 Methods

3.1 System Setup and Validation

1. The detectors should be connected in the following order: UV, LS, and RI. The RI detector should be last in this series because its cell is fragile and it cannot withstand high pressures. Tubing with an inner diameter of 0.01″ (0.25 mm) should be used throughout the system to minimize the delay volumes between detectors and minimize band-broadening effects (*see* **Note 9**).

3.2 System Equilibration

1. Equilibrate the SEC/MALS system in the buffer of choice; turn on the RI detector and pump the buffer through the RI detector in the "purge" mode for at least 12 h at the flow rate that would be used during data collection.
2. Turn on the LS and UV detectors; the UV detector requires 30 min to warm up the lamp.
3. Stop the purging mode on the RI detector, zero the RI detector, and start monitoring baselines.
4. Check the baseline quality and stability; the noise level in the LS baseline monitored at 90° should not exceed 50 μV on a DAWN HELEOS II detector (with the goal being to keep the noise at less than 20 μV; *see* **Note 10**).

3.3 Validation of SEC/MALS System in the Buffer of Interest

1. Dissolve 500 μg of protein standard in 500 μl of running buffer (use at least three different, individually prepared protein standards; *see* **Note 7**).
2. Filter the solutions of protein standards through a 0.22 μm low-protein-binding filter (*see* **Note 6**).
3. Individually inject 200 μg of each protein standard and collect UV, LS, and RI data.
4. Using software for light scattering data analysis, baseline-correct the signals from all three detectors, determine inter-detector delay, apply band-broadening correction (*see* **Note 9**), normalize the LS detector signals, and calculate M_w.
5. Check whether the computed M_ws are within 5% of expected molecular weights.
6. Inject the nucleic acid standard (*see* **Note 6**) and validate the value for $(dn/dc)_{\text{nucleic acid}}$ (*see* **Note 11**).

3.4 Determination of the Molecular Weight of the Sample Protein, Nucleic Acid, and Protein–Nucleic Acid Complex

1. Filter all samples through a 0.22 μm low-protein-binding filter (*see* **Note 6**).
2. Inject an appropriate amount of sample (*see* Table 4 for optimal amounts of sample) and record signals from all three detectors.
3. Inject all samples in a series of three or more concentrations spanning at least one order of magnitude (*see* **Note 12**).
4. Baseline-correct the signal from all three detectors.
5. Calculate molecular weights for all samples analyzed.

3.5 Determination of Monodispersity

1. Compute $(dn/dc)_{complex}$ as a weight average of $(dn/dc)_{protein}$ and $(dn/dc)_{nucleic\ acid}$ based on proposed stoichiometry (*see* **Note 11**).
2. Use the $(dn/dc)_{complex}$ during processing of the SEC/MALS data using ASTRA software (*see* **Note 11**).
3. Generate a molar mass distribution plot for all samples analyzed, i.e., protein alone, nucleic acid alone, and the complex (examples of such plots for *E.r.*RT domain are shown in Fig. 1a, b and for FIR protein in Fig. 2a, b).
4. Check the distribution of M_w across the eluting peak; for a monodisperse sample, it should vary by no more than ±5% for the middle portion of the peak that is above half-height (thus is not significantly affected by band-broadening effects).
5. Check whether the M_w varies with changes in the concentration of the injected sample by injecting the same volume of a sample that is tenfold diluted from the original sample; for monodisperse samples, the M_w measured for the apex portion of the eluting peak should be concentration independent.

3.6 Determination of Stoichiometry, i.e., Protein to Nucleic Acid Ratio in the Complex

1. Compute the M_w for the possible stoichiometries. If the observed M_w of the complex is concentration independent and within 10% of the M_w for only one of the stoichiometries proposed for the complex, the stoichiometry is computed based on M_w alone (like for *E.r.* RT + D4A RNA complex; Fig. 1b, c).
2. If the observed M_w varies with concentration, plot the M_w as a function of concentration and validate that the observed dependence is close to a plateau (Fig. 2c); if the observed M_w is within 10% of only one of the M_w proposed for the possible stoichiometries, the complex stoichiometry is computed based on M_w alone.
3. If the stoichiometry cannot be determined based on the M_w alone (like for FIR-FUSE DNA complex; Table 3), compute the volume of UV and RI peaks generated during SEC/MALS analyses of the individual components of the complex, i.e., the samples: protein alone and nucleic acid alone. Compute the

UV/RI ratios for the protein and the nucleic acid used for complex formation.

4. Compute the observed UV/RI ratio for the complex from SEC/MALS analyses at different concentrations. Check whether the UV/RI ratios are concentration dependent.
5. Compute the expected UV/RI ratio for the possible complexes as weight-average UV/RI ratios of protein and nucleic acids based on the proposed stoichiometries (results of such analysis for FIR protein complexes are reported in Table 3 and plotted in Fig. 3).
6. Confirm that both the observed M_w and UV/RI ratio are consistent with the proposed stoichiometry (Fig. 3 and Table 3; *see* **Note 12**).

4 Notes

1. The data for the examples shown were collected using an SEC/MALS system consisting of a high-performance liquid chromatography system (HPLC), Waters *Alliance* 2960, or Agilent 1200. The elution from SEC was monitored by a Waters or Agilent photodiode array (PDA) UV/VIS detector, differential refractometer, and static and dynamic, multiangle laser light scattering detector. Two software packages were used for data collection and analysis: the Waters Millennium software controlled the HPLC operation and data collection from the multiwavelength UV/VIS detector, while the Wyatt ASTRA software collected data from the refractive index detector, the light scattering detector, and recorded the UV trace at 280 nm, 295 nm, or 310 nm (*see* **Note 3**) sent from the PDA detector. However, other UV, RI, and LS detectors can be used (a single-angle LS detector is suitable for analysis of macromolecules with molar masses up to ~500 kDa). The UV detector should be capable of extracting signals for various UV wavelengths (*see* **Note 3**).
2. The 0.1 μm "inline" filter placed between pumps and the injector retains any particles that are shed from the HPLC pump's head. An additional "inline" filter (2 μm PEEK frit) with small dead volume is installed between the injector and the SEC column; this filter traps protein aggregates that are formed during the injection step and that may result from exposing the protein sample to high pressure; this filter substantially increases the lifetime of the SEC column. These filters are replaced whenever the system's operating pressure increases by more than 5%.

3. Nucleic acids have high absorbance per mass unit in comparison to protein, so they will easily saturate the 280 nm absorbance signal even at submicromolar concentrations (*see* Table 3 for observed UV/RI ratio of proteins in comparison to nucleic acids). A good starting point is to monitor absorbance simultaneously at 280 nm and 295 nm, but for higher concentrations, or longer nucleic acids, it may be necessary to move the detection even further away from 280 nm to avoid saturation (for FIR–DNA complexes, 310 nm was used).
4. The noise level of the RI signal on a Wyatt Optilab REX instrument should be around $\pm 2 \times 10^{-9}$ RIU. The RI signal is extremely sensitive to temperature, so the Optilab REX is designed to thermostat the internal measurement cell precisely to ± 0.005 °C, but experience has shown that the location of the RI unit and the SEC column in the laboratory should also be considered; for example, placing the unit or the SEC column under air conditioning outflow will generate fluctuations in signals that follow the periodicity of the room temperature. The column should be placed in a temperature control chamber (if available) or isolated by wrapping it in several layers of bubble wrap or other forms of thermal insulation.
5. There are many SEC columns suitable for SEC/MALS systems. The Supredex and Superose HR and GL 10/300 series columns are routinely used by our laboratory; the Superdex Increase columns provide greater resolution. Other commonly used columns are the TosoHaas TSK-GEL series SEC columns. The silica-based TSK-GEL columns provide excellent resolution but are limited in chemical stability to the pH range extending from 2 to 8. Although they provide lower resolution than silica-based columns, the polymer-based TSK-GEL columns are a very good choice for samples with high M_ws. Laboratory-packed columns can also be used when a specific media is required to minimize the interactions with the SEC matrix. The critical factor in a column's compatibility with the SEC/MALS analysis is the quality of the LS baseline after a new column has been conditioned. Every new column sheds particles and the noise in the LS baseline is initially high, but after 24–72 h of continuous pumping, the baseline noise should decrease significantly. A major challenge for SEC column selection is analysis of large assemblies with M_w above 1×10^6 Da (1 MDa), for which field-flow fractionation (reviewed in Ref. 19) would be a preferred method of fractionation; FFF/MALS have been used to study protein–nucleic acid complexes [20].
6. For protein samples in aqueous buffers, the 0.1 or 0.22 μm Durapore®, low-protein-binding membrane (Millipore, Bedford, MA) has proven to be an excellent choice; centrifugal

filter units have low dead volume and are thus preferred over syringe filters.

7. Protein standards that are routinely used for SEC calibration are good choices for validation of SEC/MALS performance (the individual standards, not the premixed sets). The list of 16 commercially available proteins, with M_w ranging from 6.5 kDa to 475 kDa, that our laboratory analyzed by SEC/-MALS system is posted at https://medicine.yale.edu/keck/biophysics/technologies/lightscatter/results/tables.aspx. All proteins included in the list are monodisperse, so data collected for these proteins can be used to normalize the LS detector. For nucleic acid standard, the exact piece of DNA or RNA used for complex formation is sufficient.
8. Preferably, the samples subjected to SEC/MALS analysis are fractions collected from SEC purification that are not concentrated postfractionation to avoid protein aggregation.
9. Band broadening inherent to the SEC/MALS system can be corrected for using software algorithms during data processing. These parameters are computed during processing of SEC/-MALS data collected at the same flow rate for a standard that is known to be monodisperse. Since changes in sample concentration influence its diffusion, this correction should be used with great caution; when sample is analyzed at a broad range of concentrations, the band-broadening correction should be validated at a comparable concentration range.
10. Proper equilibration of the SEC/MALS system is critical for obtaining high precision in the M_w determination; the baselines need to be monitored for at least 60 min to ensure that there is no substantial drift in RI signal, and that the LS baseline is free of noise and stable. Since the response of the LS detector determines the lower limit of detection, low noise is critical when working with a low amount of sample (e.g., 20 μg of a 40 kDa macromolecule). When not in use, the SEC/MALS system should be stored with 20% ethanol continuously recirculating throughout it. For long-term storage, the LS cell should be filled with methanol filtered through a 0.02 μm filter.
11. The error in the M_w measurement of the complex results mostly from the uncertainty in $(dn/dc)_{complex}$. The value of $(dn/dc)_{protein}$ can be estimated with high precision during ASTRA analysis of SEC/MALS data collected for protein standards; starting with the commonly used value of 0.187 ml/g [5, 6, 17, 21, 22], the iterative process leads to a single $(dn/dc)_{protein}$ value within the range of 0.18–0.20 ml/g that provides the best match between the expected and measured M_w for all standards analyzed and taking into

account their expected oligomeric state. A similar approach is used to determine the value of $(dn/dc)_{\text{nucleic acid}}$; values between 0.16 and 0.19 ml/g have been used, with smaller values associated with longer nucleic acids [1, 2, 14, 20, 23]. Since the value of $(dn/dc)_{\text{nucleic acid}}$ varies depending on the length and secondary structure of the nucleic acid and buffer conditions, it is imperative to perform SEC/MALS analysis of the exact DNA or RNA, of known sequence and M_w, that is used for complex formation and validate the refractive index increment for this particular piece of nucleic acid in the buffer conditions used for SEC/MALS analysis.

12. It is critically important to analyze the samples at various concentrations to determine whether the observed M_w is affected by changes in concentration. If sample's amount is limited, the eluting material can be collected as eluting fractions and re-analyzed at lower concentrations. Both M_w and UV/RI ratio should be monitored as a function of concentration to establish whether the complex is stable like observed for *E.r.* RT domain (Fig. 1b, c and Table 1) or experiencing dynamic dissociation process like observed for FIR protein (Fig. 2b, c and Table 2) before reporting stoichiometry; M_w should be within 10% and UV/RI ratio within 20% of predicted values.

References

1. Zhao C, Pyle AM (2016) Crystal structures of a group II intron maturase reveal a missing link in spliceosome evolution. Nat Struct Mol Biol 23(6):558–565. https://doi.org/10.1038/nsmb.3224
2. Crichlow GV, Zhou H, Hsiao HH, Frederick KB, Debrosse M, Yang Y, Folta-Stogniew EJ, Chung HJ, Fan C, De la Cruz EM, Levens D, Lolis E, Braddock D (2008) Dimerization of FIR upon FUSE DNA binding suggests a mechanism of c-myc inhibition. EMBO J 27(1):277–289. https://doi.org/10.1038/sj.emboj.7601936
3. Wyatt PJ (1993) Light scattering and the absolute characterization of macromolecules. Anal Chim Acta 272(1):1–40. https://doi.org/10.1016/0003-2670(93)80373-S
4. Takagi T (1990) Application of low-angle laser light scattering detection in the field of biochemistry: review of recent progress. J Chromatogr A 506:409–416. https://doi.org/10.1016/S0021-9673(01)91596-1
5. Wen J, Arakawa T, Philo JS (1996) Size-exclusion chromatography with on-line light-scattering, absorbance, and refractive index detectors for studying proteins and their interactions. Anal Biochem 240(2):155–166. https://doi.org/10.1006/abio.1996.0345
6. Folta-Stogniew E, Williams KR (1999) Determination of molecular masses of proteins in solution: implementation of an HPLC size exclusion chromatography and laser light scattering service in a core laboratory. J Biomol Tech 10(2):51–63
7. Minton AP (2016) Recent applications of light scattering measurement in the biological and biopharmaceutical sciences. Anal Biochem 501:4–22. https://doi.org/10.1016/j.ab.2016.02.007
8. Mogridge J (2015) Using light scattering to determine the stoichiometry of protein complexes. Methods Mol Biol 1278:233–238. https://doi.org/10.1007/978-1-4939-2425-7_14
9. Gupta K, Curtis Joseph E, Krueger S, Hwang Y, Cherepanov P, Bushman Frederic D, Van Duyne Gregory D (2012) Solution conformations of prototype foamy virus integrase and its stable synaptic complex with U5 viral DNA. Structure 20(11):1918–1928. https://doi.org/10.1016/j.str.2012.08.023

10. Butterer A, Pernstich C, Smith RM, Sobott F, Szczelkun MD, Toth J (2014) Type III restriction endonucleases are heterotrimeric: comprising one helicase-nuclease subunit and a dimeric methyltransferase that binds only one specific DNA. Nucleic Acids Res 42 (8):5139–5150. https://doi.org/10.1093/nar/gku122
11. Deng X, Prakash A, Dhar K, Baia GS, Kolar C, Oakley GG, Borgstahl GE (2009) Human replication protein A-Rad52-single-stranded DNA complex: stoichiometry and evidence for strand transfer regulation by phosphorylation. Biochemistry 48(28):6633–6643. https://doi.org/10.1021/bi900564k
12. Porterfield JZ, Zlotnick A (2010) A simple and general method for determining the protein and nucleic acid content of viruses by UV absorbance. Virology 407(2):281–288. https://doi.org/10.1016/j.virol.2010.08.015
13. Aceytuno RD, Piett CG, Havali-Shahriari Z, Edwards RA, Rey M, Ye R, Javed F, Fang S, Mani R, Weinfeld M, Hammel M, Tainer JA, Schriemer DC, Lees-Miller SP, Glover JNM (2017) Structural and functional characterization of the PNKP-XRCC4-LigIV DNA repair complex. Nucleic Acids Res 45 (10):6238–6251. https://doi.org/10.1093/nar/gkx275
14. Hsiao HH, Nath A, Lin CY, Folta-Stogniew EJ, Rhoades E, Braddock DT (2010) Quantitative characterization of the interactions among c-myc transcriptional regulators FUSE, FBP, and FIR. Biochemistry 49 (22):4620–4634. https://doi.org/10.1021/bi9021445
15. French RL, Gupta N, Copeland PR, Simonovic M (2014) Structural asymmetry of the terminal catalytic complex in selenocysteine synthesis. J Biol Chem 289(42):28783–28794. https://doi.org/10.1074/jbc.M114.597955
16. Aik WS, Lin MH, Tan D, Tripathy A, Marzluff WF, Dominski Z, Chou CY, Tong L (2017) The N-terminal domains of FLASH and Lsm11 form a 2:1 heterotrimer for histone pre-mRNA 3′-end processing. PLoS One 12 (10):e0186034. https://doi.org/10.1371/journal.pone.0186034
17. Wen J, Arakawa T, Wypych J, Langley KE, Schwartz MG, Philo JS (1997) Chromatographic determination of extinction coefficients of non-glycosylated proteins using refractive index (RI) and UV absorbance (UV) detectors: applications for studying protein interactions by size exclusion chromatography with light-scattering, UV, and RI detectors. In: Marshak DR (ed) Techniques in protein chemistry, vol 8. Academic, New York, pp 113–119. https://doi.org/10.1016/S1080-8914(97)80014-2
18. Stevens FJ (1989) Analysis of protein-protein interaction by simulation of small-zone size exclusion chromatography. Stochastic formulation of kinetic rate contributions to observed high-performance liquid chromatography elution characteristics. Biophys J 55 (6):1155–1167. https://doi.org/10.1016/s0006-3495(89)82912-1
19. Giddings J (1993) Field-flow fractionation: analysis of macromolecular, colloidal, and particulate materials. Science 260 (5113):1456–1465. https://doi.org/10.1126/science.8502990
20. Ashby J, Schachermeyer S, Duan Y, Jimenez LA, Zhong W (2014) Probing and quantifying DNA–protein interactions with asymmetrical flow field-flow fractionation. J Chromatogr A 1358:217–224. https://doi.org/10.1016/j.chroma.2014.07.002
21. Kendrick BS, Kerwin BA, Chang BS, Philo JS (2001) Online size-exclusion high-performance liquid chromatography light scattering and differential refractometry methods to determine degree of polymer conjugation to proteins and protein-protein or protein-ligand association states. Anal Biochem 299 (2):136–146. https://doi.org/10.1006/abio.2001.5411
22. Zhao H, Brown PH, Schuck P (2011) On the distribution of protein refractive index increments. Biophys J 100(9):2309–2317. https://doi.org/10.1016/j.bpj.2011.03.004
23. Nishino T, Rago F, Hori T, Tomii K, Cheeseman IM, Fukagawa T (2013) CENP-T provides a structural platform for outer kinetochore assembly. EMBO J 32 (3):424–436. https://doi.org/10.1038/emboj.2012.348

Chapter 19

Analytical Ultracentrifugation for Analysis of Protein–Nucleic Acid Interactions

Andrea Bogutzki and Ute Curth

Abstract

Analytical ultracentrifugation is a powerful tool to characterize interactions of macromolecules in solution. In sedimentation velocity experiments, the sedimentation of interaction partners and complexes can be monitored directly and can be used to characterize interactions quantitatively. As an example, we show how the interaction of the clamp loader subcomplex of DNA polymerase III from *E. coli* and a template/primer DNA saturated with single-stranded DNA-binding protein can be analyzed by analytical ultracentrifugation with fluorescence detection.

Key words Analytical ultracentrifugation, Fluorescence detection, Sedimentation velocity experiments, Protein–DNA interaction, DNA replication, Single-stranded DNA-binding protein (SSB), Clamp loader of DNA polymerase III, Sortase-mediated specific protein-labeling

1 Introduction

1.1 Analytical Ultracentrifugation

In analytical ultracentrifugation (AUC), macromolecules in the absence and presence of their interaction partners can be exposed to a high gravitational field in solution and their sedimentation can be traced by real-time detection. As a result, not only can information about the size, the shape, and the molar mass of the molecules be obtained, but in the case of interacting systems, information about the stoichiometry and the stability of the complex can also be received. There are two generally different methods used in AUC: sedimentation equilibrium and sedimentation velocity experiments. Whereas the former one can be used to determine the molar mass of molecules and complexes independent of their shape, the latter one primarily reveals information about the size and the shape of the particles. Although both methods can be used to quantitatively characterize macromolecular interactions, we will focus here on sedimentation velocity experiments and their application in the investigation of protein–DNA interactions. Basic principles, recent developments, and applications of analytical

Tina Daviter et al. (eds.), *Protein-Ligand Interactions: Methods and Applications*, Methods in Molecular Biology, vol. 2263, https://doi.org/10.1007/978-1-0716-1197-5_19,

ultracentrifugation can be found in several books on this subject that have been published recently, underlining the growing importance of this method [1–4].

Since AUC experiments investigate molecules in solution, measurements are not disturbed by interactions with surfaces, as seen with methods like surface plasmon resonance (SPR) or size-exclusion chromatography (SEC). For example, immobilization of one interaction partner is a prerequisite for SPR, which in turn might result in the occlusion of binding sites. SPR analysis of proteins involved in bacterial DNA replication yielded, for instance, that in the absence of DNA only 2.2 χ subunits of DNA polymerase III bind per immobilized single-stranded DNA-binding protein (SSB) tetramer [5], whereas AUC experiments revealed that, as expected, four χ molecules bind, one to each C-terminus of SSB [6, 7]. An advantage of AUC over SEC is that even weak interactions can be detected, since the complex always sediments in the presence of excess reactants and can therefore reform after dissociation. SEC, however, typically leads to a separation of complex and reactants, precluding re-association of the complex after dissociation and thus the detection of weak interactions. Hence, the interaction of χ and SSB in the absence of DNA could be detected by AUC and SPR [5, 6] but not by SEC [5].

In sedimentation velocity experiments, uniform solutions of the macromolecules of interest are applied and concentration gradients formed by sedimentation that are partially balanced by diffusion are observed as a function of time. From the velocity of the boundary movement, the sedimentation coefficient s can be inferred; the broadening of the sedimentation boundary contains information about the diffusion coefficient D. If s and D are known for a given particle, its molar mass can be calculated. Several methods are available for the measurement of the concentration gradients that form during sedimentation. The absorbance optical system allows the specific and sensitive detection of molecules like proteins, nucleic acids, and extrinsic chromophores that absorb in the UV/Vis range. This method cannot be used, however, when material that significantly absorbs at the detection wavelength is included in the reaction buffer. For example, investigations of proteins or nucleic acids in the presence of nucleotides in the higher micromolar or even millimolar concentration range are impaired by the UV absorbance of these components. Alternatively, the Rayleigh interference optical system can be used that measures differences in the refractive index between sample and the buffer in which the macromolecules are dissolved. The interference signal depends generally on the mass concentration and is therefore sensitive to differences in buffer composition of sample and reference. Therefore, extensive dialysis or size-exclusion chromatography of the sample is required in order to assure that the buffer composition of sample and reference are as similar as possible since even

different concentrations of small ions will result in a signal offset. Whereas the interference optical system is very useful at high concentrations where absorbance is no longer linear, it is not as sensitive as the absorbance optical system at low protein concentrations, where peptide bond absorption can be detected in the far UV.

Here we will focus on the detection by the third optical module that is available in analytical ultracentrifugation, the fluorescence detection system. In our study, we used a Beckman Coulter ProteomeLab XL-I equipped with an Aviv Biomedical fluorescence detection system (FDS), where the sample is excited by light emitted from a laser at a wavelength of 488 nm and the emission is detected through a 505–565-nm bandpass filter [8, 9]. After the rotor loaded with the sample containing cells is mounted into the vacuum chamber, the FDS is installed above the rotor (Fig. 1a). The mode of operation is similar to that of a confocal microscope (Fig. 1b): the excitation light is reflected by a dichroic mirror and focused through a condensing lens into the sample; the same lens is used as an objective lens for the emitted light that passes subsequently through the dichroic mirror and the band pass filter. The emission light is then focused on a pinhole and detected by a photomultiplier tube (PMT). In order to measure the fluorescence intensity of the sample as a function of the radial position, the whole FDS unit is moved along the radial axis by a stepping motor; this type of radial scan can be taken continuously during the whole sedimentation process (Fig. 2a).

AUC samples by default are loaded in two-channel centerpieces (Fig. 1c), where the channels are sector-shaped to prevent collision of the radially sedimenting molecules with the sidewalls, which would otherwise result in convection and disturbance of the concentration gradients. For absorbance and interference measurements, one sector is typically filled with the sample and the other one with the respective buffer to allow for measurements relative to the buffer signal. Fluorescence measurements do not need a reference beam, and thus both sectors can be filled with sample solution. Therefore, in an 8-hole rotor (Fig. 1a), up to 14 samples can be analyzed in one run since the eighth hole has to be reserved for the accommodation of the FDS calibration cell that is required for radial calibration and determination of the angular position of each cell channel [8, 9]. Absorbance and interference measurements require that the light passes through the whole path length of the solution and intensities are detected below the rotor. Fluorescence intensities, however, are detected above the rotor and therefore excitation is performed best in the region directly below the upper window of the AUC cell in order to keep the inner-filter effect as small as possible. Otherwise, this effect, which is a result of reduced light intensity due to the light absorption [10] or scattering of the sample, would result in a nonlinearity between fluorescence intensity and concentration.

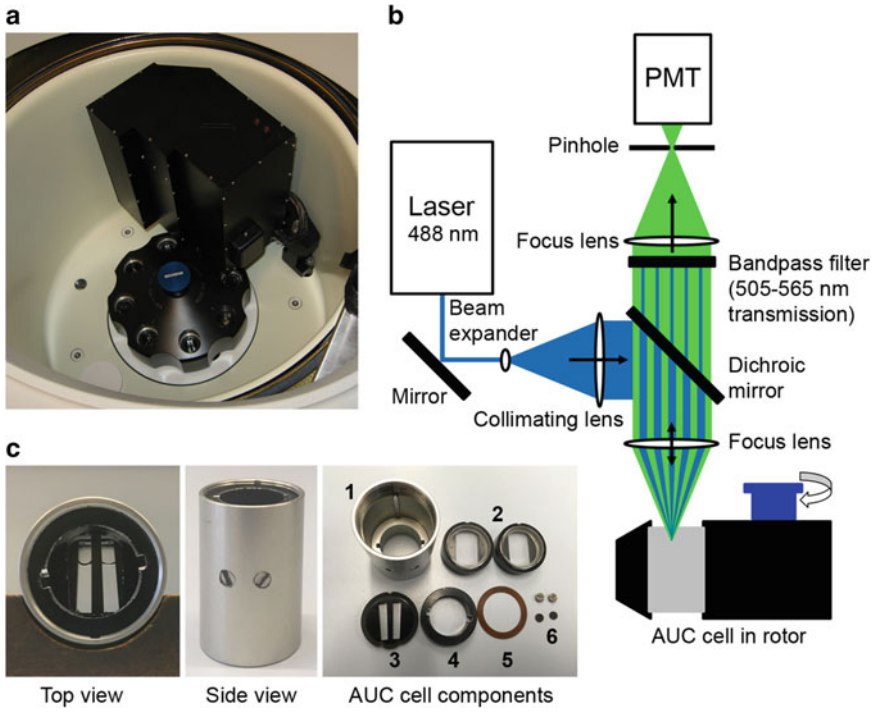


Fig. 1 Analytical ultracentrifugation with fluorescence detection. (**a**) View into the rotor chamber of the analytical ultracentrifuge ProteomeLab XL-I (Beckman Coulter, USA) that is equipped with an An-50 Ti rotor with AUC cells and the fluorescence detection system (FDS, AVIV Biomedical, USA). The FDS is mounted above the rotor and contains the light source, detector, and optical devices. (**b**) Schematic representation of the FDS according to Nelson et al. [9]; for details, *see* text. (**c**) In an assembled AUC cell, the cell housing (1) holds the two-channel centerpiece (3) sandwiched between two window assemblies (2). The cell assembly is closed at the top using a screw ring gasket (5) and a screw ring (4). Filling holes at the front side of the assembled AUC cells are sealed with plug gaskets and housing plugs (6). The housing shown is a special fluorescence cell housing (Nanolytics Instruments, Germany), where the filling holes are shifted 4.5 mm to the top and a 9 mm elevation is included at the bottom of the housing, allowing the placement of a 3-mm centerpiece at the top of the housing without the need of extra spacers

One advantage of fluorescence detection is the high sensitivity, which allows measurements even at concentrations below 1 nM fluorophore and therefore makes it possible to determine binding constants of high-affinity interactions that are not accessible with absorbance or interference detection. Since proteins and nucleic acids are normally not excitable at 488 nm, they have to be extrinsically labeled or proteins alternatively can be expressed as GFP or YFP fusions. It should be checked, however, whether introduction of the label influences the properties of the system. This is most easily done by checking whether application of mixtures of the labeled and unlabeled form of the fluorescing molecule produces different results. Although the requirement for labeling might look like a disadvantage at first sight, the fact that only the specifically labeled molecule is detected allows the investigation of the behavior

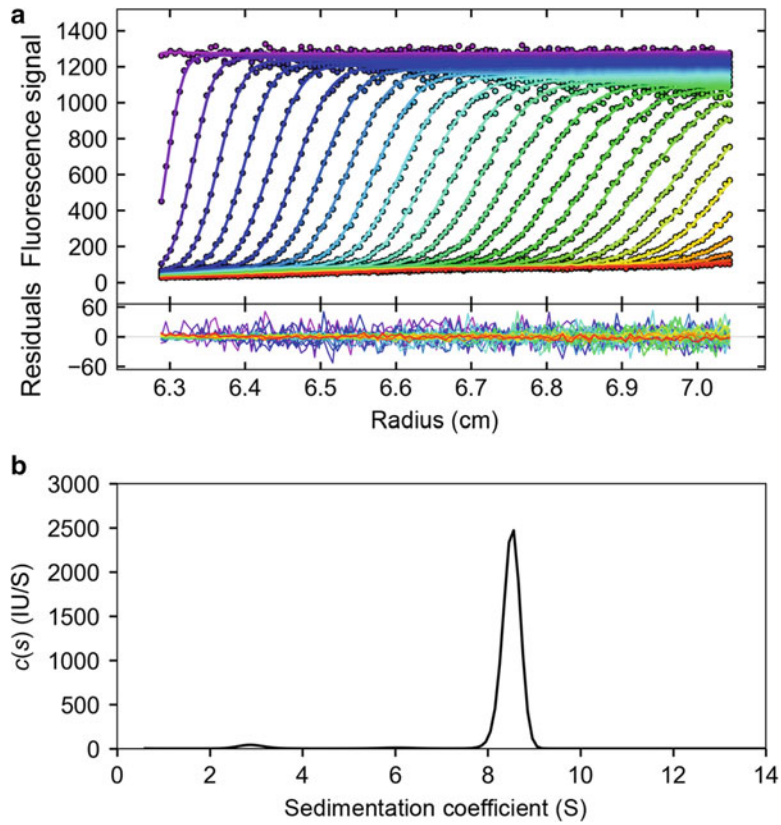


Fig. 2 Analysis of the FITC-labeled clamp loader of pol III by AUC. (**a**) 100 nM FITC-labeled clamp loader was sedimented at 40,000 rpm and 20 °C in buffer A supplemented with 1 μM BSA and 2 mM ATP using an An-50 Ti rotor and fluorescence detection. Scans were recorded every 3 min, and every second scan is shown. Circles represent experimental data (every third data point is shown) and solid lines the best fits using a continuous *c*(*s*) distribution model in SEDFIT. The lower panel shows the residuals of the fit. (**b**) *c*(*s*) distribution as obtained from the data shown in (**a**). The FITC-labeled clamp loader is homogeneous and sediments with an s-value of 8.5 S

of single components even in complex protein or protein/DNA mixtures. Since additives like nucleotides also do not contribute to the fluorescence signal, they can be used without disturbing the measurement.

1.2 Interaction of the Pol III Clamp Loader with SSB-Saturated Template/Primer

At the bacterial replication fork, several protein complexes are involved in the protein–DNA and protein–protein interactions that are required for the exact duplication of the bacterial chromosome. In *E. coli*, the main replication enzyme is DNA polymerase III holoenzyme (pol III). It consists of three subcomplexes: the core polymerase αεθ that holds the actual 5′–3′ DNA polymerase activity and the 3′–5′ exonucleolytical proofreading activity, the

ring-shaped β_2 sliding clamp that confers processivity to pol III, and the clamp loader complex $(\tau/\gamma)_3\delta\delta'\psi\chi$ [11]. The clamp loader is not only responsible for loading the β_2 sliding clamp onto the DNA but also mediates the interaction with pol III core and the DnaB helicase via its τ subunits [12] and with SSB via its χ subunit [5–7, 13]. Whereas τ and γ are both encoded by the *dnaX* gene, the γ variant does not contain the domains required for interaction with the core and the helicase due to translational frameshifting [11]. Therefore, up to three cores can be bound in the holoenzyme depending on the number of τ proteins in the clamp loader complex [14]. Furthermore, the clamp loader interacts with the template/primer allowing ATP hydrolysis and release of β_2 onto the primer of each Okazaki fragment. Since the clamp loader interacts with the template strand only, both RNA and DNA primers can be accepted [15].

The interaction between the χ subunit of the clamp loader and SSB facilitates processive DNA synthesis under elevated salt conditions [5]. We previously found that SSB and SSB/ssDNA complexes have the same affinity for χ under these conditions [7]. In the experiments of Glover and McHenry [13] using surface plasmon resonance (SPR), the χψ subcomplex of the clamp loader was found to interact with SSB with a similar affinity (K_D = 2.7 μM) as determined by AUC for χ alone (K_D = 3.4 μM) [6], whereas the affinity of the τ clamp loader for SSB in the presence of template/primer was found to be 1000-fold higher (K_D = 3 nM). It has been shown that in some cases results obtained by SPR differed significantly from those determined by other methods like isothermal titration calorimetry, AUC, or fluorescence quenching [16]. Since SPR requires the immobilization of one reaction partner to a surface, effects may appear, like occlusion of binding sites, steric hindrance of ligand binding, finite rate of mass transport of the ligand, and, if charged matrices like carboxymethylated dextran are used, nonspecific electrostatic adsorption. As AUC is a first principal method that avoids binding to surfaces, the interaction of τ clamp loader with SSB-saturated template/primer seemed to be a good model system to demonstrate how high-affinity interactions can be investigated using analytical ultracentrifugation with fluorescence detection.

1.3 Specific Fluorescence Labeling

For our experiments, we specifically labeled the clamp loader complex with a single fluorescein probe at the C-terminus of the ψ subunit. As the clamp loader is involved in several interactions, e.g., with SSB, template/primer, DnaB helicase, and the core of pol III, we decided to specifically label a part of the complex that is not known to be involved in any interactions. In order to place a single label, we modified our expression plasmid for the ψ subunit in such a way that at the C-terminus of the resulting protein a *Staphylococcus aureus* sortase (SrtA) recognition site followed by maltose-

binding protein (MBP) was added (Fig. 3a). SrtA is a transpeptidase involved in bacterial cell wall assembly that anchors surface proteins containing an LPXTG recognition motif to the amino group of a pentaglycine on the peptidoglycan cell wall [17]. This enzyme can be used in vitro to covalently link a labeled peptide containing an N-terminal glycine residue to a protein with a SrtA recognition site (Fig. 3a). As the MBP is positioned at the C-terminus of the recognition site, it is cleaved off during the labeling reaction and it can easily be judged by SDS-PAGE whether the labeling is completed (Fig. 3c). Following the labeling reaction, the FITC-labeled clamp loader was isolated by SEC (Fig. 3b, c) and checked for homogeneity by AUC (Fig. 2).

1.4 Template/Primer DNA

The sequences of the oligonucleotides were chosen according to Bloom et al. [18], though omitting the 3′-overhang of the template makes them thereby better comparable to the 105-mer/5-′-fluorescinated 50-mer template/primer used in SPR [13]. However, they were designed to have shorter double-stranded regions since only about ten nucleotides of double-stranded DNA have been shown to be involved in interactions with the clamp loader [15]. We used a template/primer consisting of a 30-mer DNA primer annealed in such a way to a 55-mer DNA template that a 5′-overhang of 25 nucleotides was obtained (Fig. 4a, Table 1) and compared this with a 80-mer/30-mer template/primer that comprised the same 5′-overhang as used in the SPR experiments of Glover and McHenry [13].

To avoid degradation by contaminating nucleases, we used oligonucleotides that contained four phosphorothioate (PTO) bonds at both their 5′- and 3′- ends. A PTO bond is a phosphodiester bond where one nonbridging oxygen is replaced by sulfur (Fig. 4c), and by this, it offers increased stability against nucleolytic attacks. Template/primer constructs were obtained by annealing the respective template with an excess of primer followed by SEC to remove non-hybridized primer (Fig. 4b) and were checked for homogeneity by AUC and gel electrophoresis (Fig. 4d, e).

1.5 Determination of the Affinity of an Interaction by Sedimentation Velocity Experiments

Whenever two molecules interact, the resulting complex will have a higher molecular mass and will usually sediment faster than the two components on their own. If the interaction is fast on the time scale of sedimentation (dissociation rate constant $>10^{-3}$ s^{-1} [19]), the reaction will continue during the sedimentation process, which typically takes several hours. The complex will dissociate and reassociate during sedimentation and will therefore not sediment as a single species in an extra boundary, but together with one of the interaction partners in a so-called reaction boundary. Whether this is the slower (A) or the faster (B) sedimenting interaction partner depends on the composition of the reaction mixture. If an excess of A is present, it can be assumed as a good approximation that the

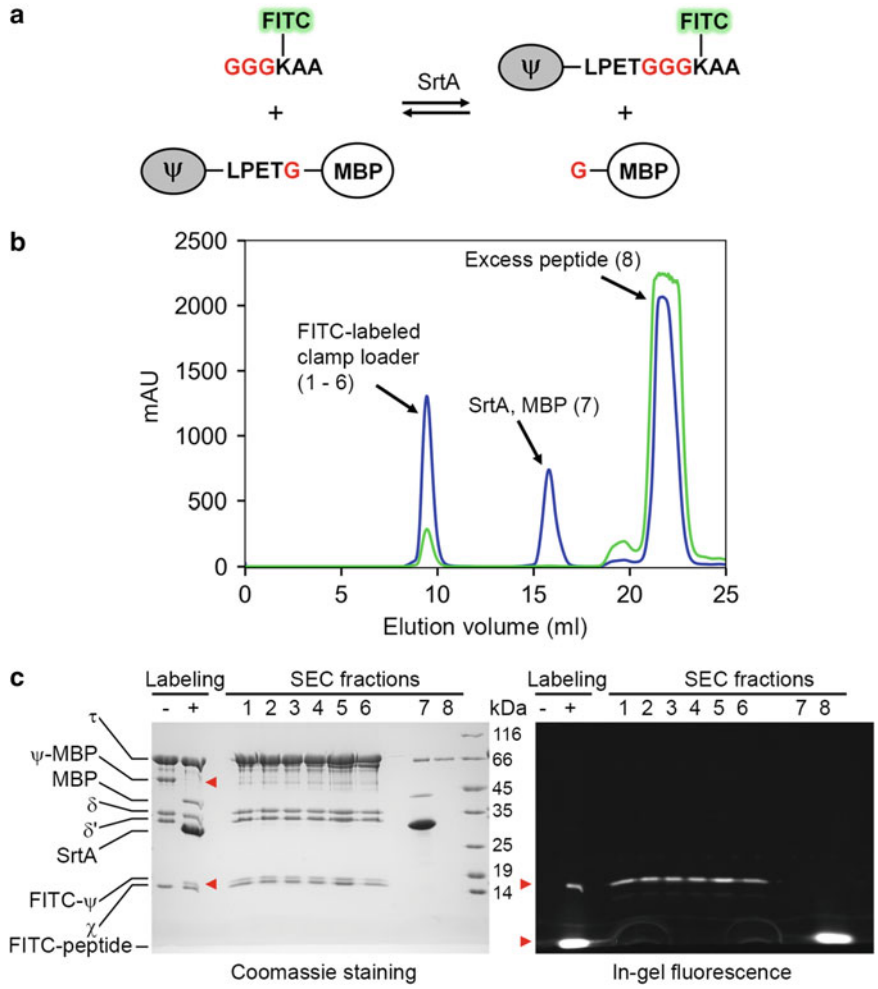


Fig. 3 Site-specific labeling of the pol III clamp loader with FITC. (**a**) SrtA catalyzes the cleavage between T and G in the LPXTG motif and subsequent peptide bond formation between T and the amino group of the fluorescently labeled peptide that has to start with at least one G. The LPETG SrtA recognition sequence and an MBP-tag were added to the C-terminus of the ψ subunit of the clamp loader by cloning. MBP is cleaved off during the labeling reaction. (**b**) Nonreacted FITC-peptide, SrtA, and MBP can be separated from FITC-labeled clamp loader by SEC. Elution was monitored using absorption at 280 nm (blue) and 494 nm (green). Numbers 1–8 indicated in the elution profile represent the samples loaded on the SDS-PAGE in (**c**). (**c**) Samples of the reaction mixture before (−) and after (+) addition of SrtA and FITC-peptide, and aliquots of SEC fractions (1–8 as indicated in **b**) were analyzed by a 12% SDS-PAGE. The left panel shows the gel after Coomassie staining; on the right panel, fluorescence was detected after excitation at 312 nm. For better orientation, ψ-MBP, FITC-ψ, and FITC-peptide are marked with arrowheads

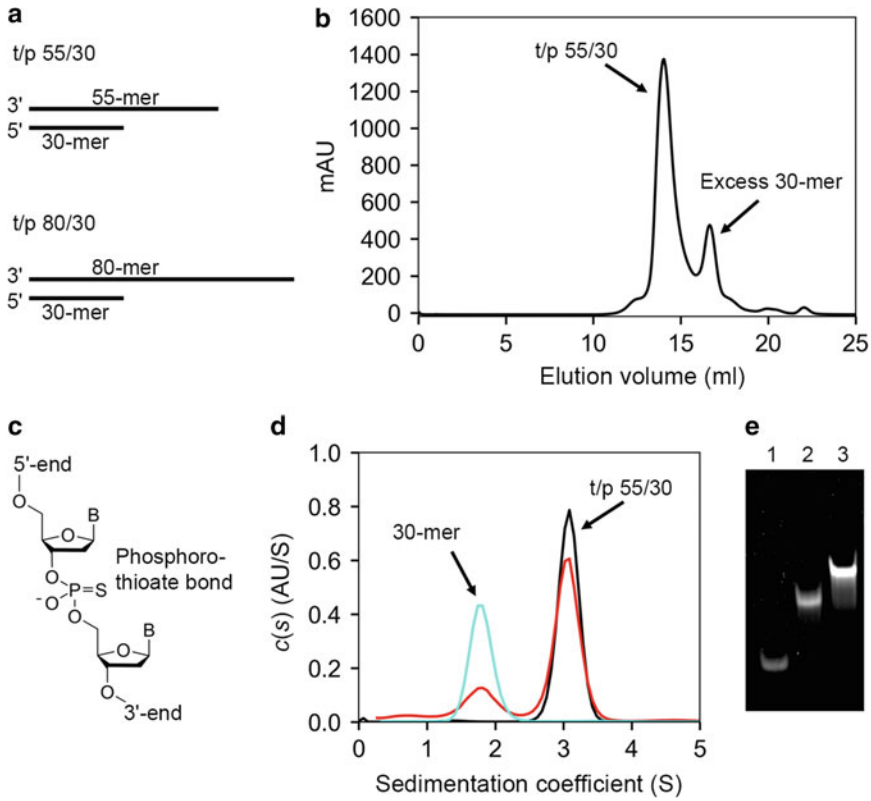


Fig. 4 Preparation of the 55-mer/30-mer template/primer. (**a**) Schematic representation of the template/primer assemblies used. They consist of two different 30-mer DNA primers (*see* Table 1) hybridized to a 55-mer (t/p 55/30) or an 80-mer (t/p 80/30) generating 30 nucleotide double-stranded regions with a 5′-overhang comprising 25 or 50 nucleotides, respectively. (**b**) Excess 30-mer can be separated from hybridized t/p 55/30 by SEC. Elution was monitored using absorption at 280 nm. (**c**) Phosphorothioate bond. (**d**) *c*(*s*) distributions of the sample before SEC (red) and pooled fractions of the two main peaks of the SEC elution profile containing t/p 55/30 (black) and excess 30-mer (cyan). t/p 55/30 is homogeneous and sediments with an s-value of 3.1 S. (**e**) 100 ng of 30-mer (lane 1), 55-mer (lane 2), and t/p 55/30 after SEC (lane 3) were loaded on a 10% TPE-PA gel and stained with SYBR Gold after gel electrophoresis

complex and B sediment together in the reaction boundary in the presence of a constant concentration of slower sedimenting A (constant bath approximation [20, 21]). Another way to describe this process is the effective particle theory (EPT) that takes into account that due to the dissociation of the complex a small fraction of A co-sediments in the reaction boundary and that this fraction of A, total B, and AB sediment together like a single particle in one boundary, whereas the bulk of A sediments with the s-value of free A [22]. Regardless of any model, the observation that the titration of a constant concentration of B with increasing concentrations of A results in a shift of the sedimentation coefficient of B to higher values (Fig. 5a) is a proof of the interaction. With increasing excess of A, the sedimentation coefficient of the reaction boundary s_{rb} will

Table 1
Sequences of the oligonucleotides used. The 30-mer oligonucleotides A and B are complementary to the 3′-end of the corresponding template; therefore, the resulting template/primer assemblies consist of 30 nucleotides of double-stranded DNA and a 5′-overhang. The sequence of t/p 55/30 was chosen according to Bloom et al. [18], though omitting the 3′-overhang of the template. The sequence of t/p 80/30 is based on Glover and McHenry [13] and contains the same 5′-overhang but a shorter double-stranded DNA region comprising 30 instead of 52 nucleotides

	Oligonucleotide sequence (5′ to 3′)
t/p 55/30	
55-mer	TGAGCGTTTTTCCTGTTGCAATGGCTGGCGGTAACAAAGCTTCGGACACTATCCT
30-mer A	AGGATAGTGTCCGAAGCTTTGTTACCGCCA
t/p 80/30	
80-mer	TTACGTTGATTTGGGTAATGAATATCCGGTTCTTGTCAAGATTACTCTTGATGA A GGAAGCTTAGCCTATGCGCCTGGTC
30-mer B	GACCAGGCGCATAGGCTAAGCTTCCTTCAT

increase from the s-value of free B (s_B) to the s-value of the complex (s_{AB}). The s-value s_{rb} reflects the fractional time B spends on average in the complex during sedimentation [19]. Such a titration can also be used to quantitatively determine the binding constant of the interaction. This requires fitting of a binding isotherm to the sedimentation coefficient of the reaction boundary as a function of the concentration of the slower sedimenting interaction partner (Fig. 5b).

For a simple interaction, where one molecule of the slower sedimenting interaction partner A interacts with one molecule of B to form a complex AB, the sedimentation coefficient of the reaction boundary s_{rb} in an excess of A is given both in the constant bath approximation [20] and in EPT [22] by

$$s_{rb} = \frac{[B] \cdot s_B + [AB] \cdot s_{AB}}{[B] + [AB]} = \frac{s_B + s_{AB} \cdot \frac{[A]}{K_D}}{1 + \frac{[A]}{K_D}} \tag{1}$$

where

$$K_D = \frac{1}{K_A} = \frac{[A] \cdot [B]}{[AB]} \tag{2}$$

K_D and K_A are the dissociation and association constants, respectively, and $[A]$, $[B]$, and $[AB]$ are the equilibrium concentrations of A, B, and AB, respectively, which can be calculated from the total concentrations of A (A_0) and B (B_0) as follows:

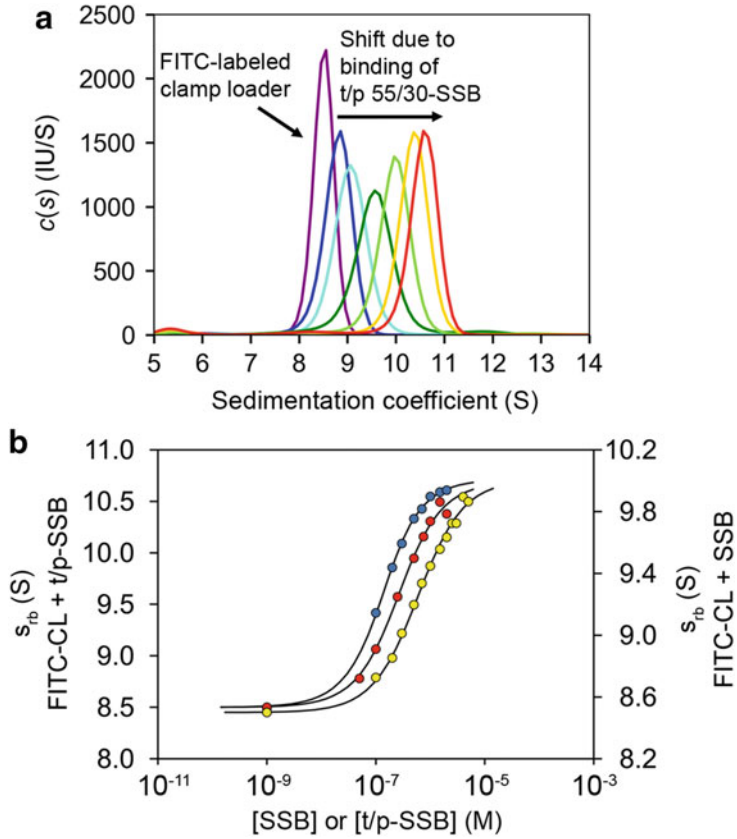


Fig. 5 Interaction of FITC-labeled clamp loader with template/primer-SSB or SSB alone. (**a**) Continuous $c(s)$ distributions of 100 nM FITC-labeled clamp loader (purple) titrated with increasing concentration of t/p 55/30-SSB in buffer A supplemented with 1 μM BSA and 2 mM ATP: 50 nM (dark blue), 100 nM (cyan), 250 nM (green), 0.5 μM (light green), 1 μM (yellow), and 1.5 μM (red). For better clarity, not all titration points are shown. (**b**) Binding isotherms derived from sedimentation coefficients of the reaction boundaries s_{rb} for the interaction of 100 nM FITC-labeled clamp loader (FITC-CL) with t/p 55/30-SSB (from data shown in (**a**), red, $K_D = 0.21$ μM), t/p 80/30-SSB (blue, $K_D = 0.087$ μM), and SSB (yellow, $K_D = 0.55$ μM) were fitted with an A + B ↔ AB model

$$[AB] = \frac{1}{2} \cdot (A_0 + B_0 + K_D) - \sqrt{\left(\frac{1}{2} \cdot (A_0 + B_0 + K_D)\right)^2 - A_0 \cdot B_0} \quad (3)$$

$$[B] = B_0 - [AB], [A] = A_0 - [AB] \quad (4)$$

The sedimentation coefficient of the reaction boundary can be determined from the raw data using the diffusion-corrected differential sedimentation coefficient distribution model (continuous $c(s)$ distribution) of the program package SEDFIT [23]. In this

method, the experimental sedimentation data are matched with the best possible combination of sedimentation patterns from particles of different sizes and in the resulting distribution of s-values the contribution of the diffusional spread is eliminated [19]. This is especially useful if both reaction partners contribute to the detected signal since the diffusion-correction usually allows a separation between the sedimentation coefficients of free A and the reaction boundary. In cases like ours, where the faster sedimenting FITC-labeled clamp loader is detected, but not the slower sedimenting unlabeled SSB-saturated template/primer DNA, only the sedimentation of the reaction boundary can be observed (Fig. 5a). Even in such cases, it is advantageous to use $c(s)$ analysis since determination of s_{rb} by integration of the $c(s)$ distribution in the range from s_B to s_{AB} will omit the contribution of traces of free dye, aggregates, or impurities to the obtained s-value.

In order to assess the affinity of the τ clamp loader to SSB-saturated template/primer, we titrated a constant amount of the faster sedimenting FITC-labeled clamp loader with increasing excesses of unlabeled SSB-saturated template/primer. We used this kind of titration since we wanted to impede the formation of higher complexes, as one SSB tetramer can bind up to four χ subunits [6, 7]. Otherwise, further clamp loader molecules could bind to the DNA/SSB-complex via interactions solely with the SSB C-terminus. Although this interaction is expected to be much weaker, it could interfere with the analysis. Furthermore, we could then investigate the effect of different unlabeled-primer/template sequences on the affinity.

2 Materials

2.1 Buffers and Solutions

All buffer solutions were prepared with chemicals of the highest purity available in ultrapure H_2O; pH was adjusted at room temperature.

1. Buffer A: 20 mM Tris-HCl, 150 mM NaCl, 10 mM $MgCl_2$, 1 mM TCEP, pH 7.5.
2. Buffer B: 20 mM Tris-HCl, 150 mM NaCl, 1 mM TCEP, pH 7.5.
3. Buffer C: 20 mM Tris-HCl, 150 mM NaCl, pH 7.5.
4. ATP stock: 100 mM, pH 7.5.
5. TCEP stock: 0.5 M, pH 7.0.
6. $MgCl_2$ stock: 1 M.
7. $CaCl_2$ stock: 0.2 M.
8. Calibration solution: 100 nM fluorescein, 10 mM Tris-HCl, 100 mM NaCl, pH 8.3.

2.2 Proteins

1. SSB was purified to homogeneity as described before [6]. An extinction coefficient of 113,000 M^{-1} cm^{-1} at 280 nm [24] is used for calculation of molar concentrations, which are given as tetramers.
2. The clamp loader used in this study (subunit composition $\tau_3\delta\delta'\chi\psi$) contains an ψ-MBP variant carrying a SrtA cleavage site (Fig. 3a) for site-specific labeling using a SrtA-mediated reaction [17]. It was purified to homogeneity as described for the $\tau_3\delta\delta'\chi\psi$ clamp loader [25].
3. BSA was purchased in 98% purity, further purified by size-exclusion chromatography, and used as a 100 μM stock solution. It serves as a carrier protein to avoid sticking of FITC-labeled clamp loader (*see* **Note 1**).
4. StrA was purified as described elsewhere [26].

All proteins were flash-frozen in liquid N_2 in aliquots of at most 100 μL and stored at −80 °C until use.

2.3 Oligonucleotides

DNA oligonucleotides (Table 1) used in this study were purchased in HPLC quality. By annealing of the appropriate 30-mer A or B, respectively, to a 55-mer or an 80-mer oligonucleotide, the template/primer constructs t/p 55/30 and t/p 80/30 were obtained that consist of 30 nucleotides of double-stranded DNA with a 5′-overhang comprising 25 or 50 nucleotides, respectively (Fig. 4a).

2.4 Peptide

FITC-peptide NH_2-GGGK(FITC)AA-COOH was purchased and carries a fluorescein isothiocyanate (FITC) probe at the lysine residue. This peptide is used to introduce a fluorescein probe at the C-terminus of the ψ subunit of the clamp loader by a SrtA-mediated labeling reaction (Fig. 3a). An extinction coefficient of 70,000 M^{-1} cm^{-1} at 493 nm is used to determine the molar concentration of the FITC-peptide [27].

2.5 Instrumentation

1. ProteomeLab XL-I analytical ultracentrifuge (Beckman Coulter, USA).
2. A fluorescence detection system with λ_{Ex} = 488 nm and λ_{Em} = 505–565 nm (FDS, Aviv Biomedical, USA).
3. An FDS calibration cell required for radial and angular calibration filled with a solution of 100 nM fluorescein in 10 mM Tris-HCl and 100 mM NaCl, pH 8.3 (*see* **Note 2**).
4. 8-hole An-50 Ti rotor.
5. 3 mm two-channel charcoal-filled Epon centerpieces (e.g., from Beckman Coulter, USA).
6. Standard AUC cell housings (e.g., from Beckman Coulter, USA) or special fluorescence housings (e.g., from Nanolytics

Instruments, Germany) for AUC experiments using absorbance detection or fluorescence detection, respectively.

7. UV/VIS spectrometer such as the Jasco V-560.
8. Protein purification system such as the ÄKTApurifier (GE Healthcare, Germany) equipped with an appropriate size-exclusion chromatography column, e.g., Superdex 200 Increase 10/300 GL column (GE Healthcare, Germany).

2.6 Software

1. SEDFIT [23] is a program package for the analysis of AUC data, was used to generate $c(s)$ distributions from the raw data of the AUC experiments, and is available for free at http://analyticalultracentrifugation.com/download.htm.
2. GUSSI [28] is used for the generation of binding isotherms from $c(s)$ distributions and graph plotting. It is available as a zipped archive for free at https://www.utsouthwestern.edu/labs/mbr/software/.
3. SEDPHAT [29] used to analyze binding isotherms is available for free at http://analyticalultracentrifugation.com/sedphat/download.htm or at https://sedfitsedphat.nibib.nih.gov/software/default.aspx.

All programs should be extracted into the same folder, e.g., C:\sedfit.

3 Methods

3.1 Labeling of the Clamp Loader

1. Thaw the clamp loader-ψMBP on ice.
2. Dissolve an appropriate amount of FITC-peptide in buffer B.
3. Prepare the labeling reaction mixture in a total volume of 500 μL by mixing the components in the following order: Buffer B, 0.5 mM FITC-peptide, 6 mM $CaCl_2$, 10 μM clamp loader-ψMBP, and 75 μM SrtA.
4. Following incubation at 25 °C overnight in the dark, apply the labeling reaction mixture to a Superdex 200 Increase 10/300 GL column (GE Healthcare, Germany), equilibrated with buffer B supplemented with 1 μM BSA (*see* **Note 3**), at a flow rate of 0.5 mL/min to separate FITC-labeled clamp loader from MBP, SrtA, and nonreacted peptide (as shown in Fig. 3b).
5. To determine which fractions to pool, analyze the samples by SDS-PAGE and visualize the proteins by Coomassie staining and fluorescence imaging with excitation at 312 nm (as shown in Fig. 3c).
6. Determine the concentration of purified complex by UV spectroscopy using an extinction coefficient of 298,510 M^{-1} cm^{-1} for the clamp loader at 280 nm [30]. Beforehand, the

absorbance must be corrected for contributions of the FITC dye after measuring the absorbance of the labeled protein at 493 nm:

$$A_{280}^{\text{protein}} = A_{280} - 0.3 \cdot A_{493} \tag{5}$$

7. Calculate the degree of labeling for the FITC-labeled clamp loader from the concentration of the FITC dye divided by the concentration of the protein. The concentration of the attached FITC dye can be calculated from the absorbance at 493 nm using an extinction coefficient of 70,000 M^{-1} cm^{-1}. Using this labeling protocol, the degree of labeling is typically 0.85–0.9.
8. Pool fractions containing high amounts of FITC-labeled clamp loader with a degree of labeling of at least 0.85, flash-freeze aliquots of 10–20 μL in liquid N_2, and store at −80 °C until use. The labeled clamp loader typically has a concentration of 1.5–2 μM.
9. Verify the homogeneity of the FITC-clamp loader preparation by AUC sedimentation velocity experiments in buffer A at 40,000 rpm using the fluorescence detection system (as shown in Fig. 2).

3.2 Template/Primer Preparation

1. Determine the concentrations of oligonucleotides by UV spectroscopy at 260 nm using extinction coefficients calculated from nucleotide composition using IDT SciTools [31]: 290,800 M^{-1} cm^{-1} for 30-mer A, 510,500 M^{-1} cm^{-1} for 55-mer, 281,600 M^{-1} cm^{-1} for 30-mer B, and 765,100 M^{-1} cm^{-1} for 80-mer.
2. Mix 15 μM of template (55-mer or 80-mer) with 25 μM of the appropriate primer (30-mer A or B, respectively) in a total volume of 550 μL of buffer C.
3. Heat the mixture to 90 °C for 10 min and let cool down at room temperature.
4. To separate hybridized template/primer from excess primer, apply the DNA solution to a Superdex 200 Increase 10/300 GL column (GE Healthcare, Germany) equilibrated with buffer C at a flow rate of 0.5 mL/min (as shown in Fig. 4b).
5. Determine the concentrations of template/primer solutions by UV spectroscopy using extinction coefficients at 260 nm calculated from nucleotide composition using IDT SciTools [31]: 703,707 M^{-1} cm^{-1} for t/p 55/30 and 954,221 M^{-1} cm^{-1} for t/p 80/30.

6. Pool fractions containing high amounts of template/primer and store aliquots of 100 μL at −20 °C until use. The concentration of the template/primer solution is typically about 2.5 μM.
7. Verify the homogeneity of the template/primer by AUC sedimentation velocity experiments in buffer C at 50,000 rpm using the absorbance optics at 260 nm (as shown in Fig. 4d) or by gel electrophoresis followed by staining with SYBR Gold (as shown in Fig. 4e).

3.3 Sample Preparation

1. Centrifuge protein solutions after thawing for 20 min at 15,500 × *g* and 4 °C to remove aggregates, which would interfere with the correct determination of protein concentration. Add an appropriate volume of buffer A or buffer A supplemented with 1 μM BSA, respectively, to achieve a concentration of about 15 μM SSB and 1 μM FITC-labeled clamp loader. Thaw 55-mer/30-mer primer/template solution.
2. Measure absorbance spectra of unlabeled protein and DNA samples in the wavelength range of 320 to 220 nm (*see* **Note 4**) or 600 to 220 nm in the case of FITC-labeled clamp loader. Calculate protein and DNA concentrations using the extinction coefficients given above.
3. Prepare template/primer-SSB solution by mixing of 2.2 μM 55-mer/30-mer template/primer with 2.5 μM SSB (1.1-fold excess of SSB over template/primer). As template/primer is stored in a buffer without TCEP and $MgCl_2$, appropriate volumes of TCEP and $MgCl_2$ stock solutions have to be added in order to adjust for composition of buffer A.
4. Prepare 9–14 AUC samples with a total volume of 110 μL by mixing the components in the following order: an appropriate volume of buffer A, BSA to a final concentration of 1 μM, 2 mM ATP, 55-mer/30-mer template/primer-SSB in increasing concentrations, and 100 nM FITC-labeled clamp loader. In the experiment shown in Fig. 5, 55-mer/30-mer template/primer was added in the following concentrations: 0, 50 nM, 100 nM, 250 nM, 500 nM, 750 nM, 1 μM, 1.5 μM, and 2 μM.

3.4 Performing the Analytical Ultracentrifugation Experiments

1. Assemble seven analytical ultracentrifugation cells using standard 3-mm two-channel charcoal-filled Epon centerpieces and special fluorescence cell housings (Fig. 1c) according to the Beckman Coulter rotor manual LXL/A-TB-003F provided with the centrifuge (*see* **Note 5**).
2. Fill each channel with 100 μL of sample (*see* **Note 6**). No reference is needed in fluorescence measurements; therefore, both channels can be loaded with samples (*see* **Note 7**).

3. Balance one cell to the calibration cell and load an An-50 Ti rotor with the cells according to Beckman Coulter rotor manual LXL/A-TB-003F provided with the centrifuge (*see* **Note 8**).
4. Install the rotor in the vacuum chamber of the centrifuge, mount the FDS optics, and start a run at 0 rpm and 20 °C at the centrifuge (*see* **Note 9**). Switch on the AU System Box and start the AOS (AU-Advanced Operating System) program afterward. There you have to select in the "Rotor Setup" menu the "AnTi-8: 8-hole Titanium" rotor, the position of the calibration cell, and "2 channel Charcoal-Epon Velocity" at the seven positions where you placed the sample containing cells.
5. When the vacuum is below 170 mTorr, the laser can be ignited by selecting "Start" and "Start machine." It will take about 5 min for the laser to warm up and lock in the operating state. Successful locking of the laser can be recognized when the "FDS Laser" signal of the "Optical System Status" box turns from cyan to green.
6. Accelerate the rotor to 3000 rpm. After the "Magnet Angle" display turned green, use "Fluorescence focusing" from the "Fluorescence" menu to measure a focus scan of the calibration strip and perform a radial scan at the optimal focus position. Check whether the outer edge of the calibration strip is at the expected position. Otherwise, save the radial calibration, repeat the radial scan, and check the position of the outer edge again; it should now be correct. Perform a focus scan of one of the sample containing channels and save the focus (*see* **Note 10**).
7. Choose "Set gains" from the "Fluorescence" menu and display the signal of all cells. Now adjust the voltage of the photomultiplier (0–100%) and the gain (1–8) in order to get the best signal to noise ratio. The signal can be in the range of 0–4000, which is the upper limit of the photomultiplier. Since the fluorophore concentration in all samples is the same, there should be no need to use different gain settings (*see* **Note 11**).
8. Perform one scan at 3000 rpm in order to check whether your cells are properly filled and whether all settings are correct. Use the "Stop" button in AOS and select "Stop the machine" to switch off the laser and to stop the rotor. Since this will also turn off the diffusion pump of the centrifuge, instantly start a run at 0 rpm at the centrifuge (*see* **Note 9**) and wait for thermal equilibration at least for 1 h after the temperature of 20 °C is displayed at the machine.
9. Switch on the laser by using the "Start" button of AOS and select "Start machine." Create a new method using "Select method" with a rotor speed of 40,000 rpm (*see* **Note 12**), 10 h duration, and 20 °C and set "interval between scans" to

0. After locking of the laser, wait for 30 min (*see* **Note 13**) and start the run by using the "Start" button of AOS and selecting "Start experiment."

10. After the run is finished, redissolve the sedimented material by vortexing the AUC cells and withdraw the solution with a Hamilton syringe. Check on an SDS-PAGE whether protein degradation occurred. Check the integrity of the DNA on a polyacrylamide gel with SYBR Gold staining.

3.5 Data Analysis

1. Load the data for one channel into SEDFIT, define meniscus (*see* **Note 14**), bottom and fitting limits, and choose the "Continuous $c(s)$ Distribution" model for evaluation. Set minimal and maximal s-values and resolution and tick the boxes of the parameters that should be fitted (typically frictional ratio, time-independent noise, and meniscus position). Press "Run" to check whether the chosen settings are correct for your data set and if they are, proceed with pressing "Fit" to adjust the selected parameters. Afterward, check whether the residuals are evenly distributed and whether the calculated curves represent your data properly. Repeat this analysis for all samples.
2. Copy the $c(s)$ distributions and for documentation also screenshots of the fits and the fit parameters to Microsoft Excel. Overlay the $c(s)$ distributions in one diagram and check whether only one peak is observed (*see* **Note 15**) and whether the sedimentation coefficient increases with increasing concentration of SSB-saturated template/primer as in Fig. 5a. If no shift in the sedimentation coefficient is observed, no binding occurs in the concentration range at the buffer conditions used.
3. Start GUSSI and choose the GUSSI module "$c(s)$." Copy a $c(s)$ distribution from Excel or SEDFIT and select "Paste a Distribution" from the "Distributions" menu. Repeat this step for all $c(s)$ analyses you want to examine.
4. In the "Integration" menu of GUSSI, choose "Make Isotherm" → "Hetero" → "EPT sw fast" and select the s-value range of the integration so that all peaks are covered. The "GUSSI's Isotherm Constructor" window will pop up where you will find the sedimentation coefficients and you only have to insert the total concentrations (in μM) of A (slower sedimenting interaction partner) and B (faster one) of the AUC samples that correspond to the loaded $c(s)$ distributions (*see* **Note 16**). When you hit the "Save" button, a *.isotherm file will be saved that contains the total concentrations of A and B in the first and second columns, respectively, and the s-value of the reaction boundary in the third (*see* **Note 17**).
5. Start SEDPHAT and load your file by choosing "Load New AUC Isotherm Data" from the "Data" menu. The

"Experimental Parameters" window will pop up and you have to tick "EPT sw fast" and set the extinction coefficient of A to zero, while choosing an arbitrary extinction coefficient for B (e.g., 1), the only molecule that contributes to the fluorescence signal.

6. Save the *.xp file as suggested and choose from the "Model" menu "A + B ↔ AB Hetero-Association."
7. Open the "Global Parameters" window and give the correct value for the sedimentation coefficient of B (*see* **Note 18**) and estimated values for s_{AB} and log K_A (*see* **Note 19**). Tick the boxes of s_{AB} and log K_A to tell the program that these are the values that should be fitted.
8. Choose "Single Experiment Run" from the "Run" menu to check whether your settings are reasonable. If they are not, change the estimated values, repeat this step, and if the results are satisfying, choose "Single Experiment Fit" from the "Fit" menu. The results of your fit will be displayed (*see* **Note 20**). Check the quality of the fit and look whether the residuals are evenly distributed. The highest s-value measured for the reaction boundary should be close to s_{AB}; if this is not the case, not the whole isotherm is covered and it would be advisable to repeat the experiment at higher concentrations of A. The results of fitting binding isotherms to our data obtained for the interaction of 100 nM FITC-CL with t/p 55/30-SSB, t/p 80/30-SSB, and SSB alone can be seen in Fig. 5b.

4 Notes

1. Especially when working with fluorescently labeled macromolecules at low nanomolar concentrations, the addition of an inert carrier protein to your sample is recommended to avoid sticking of labeled protein to surfaces of, e.g., the AUC cell assembly, pipette tips, and reaction tubes. We used BSA at a concentration of 1 μM (0.066 mg/mL). Alternatively, κ-casein or lysozyme at concentrations of 0.1 mg/mL and 0.2 mg/mL [32], respectively, or low concentrations of nonionic detergents like 0.05% (v/v) Tween 20 can be used.
2. In the AU-FDS calibration cell, the rectangular channel that is positioned in the middle of the centerpiece is named channel A (the so-called calibration strip). Channel A is used for angular location of each cell channel, and the bottom edge of this channel is used for radial calibration. The detection of the signal from the calibration cell is crucial to any measurement with the FDS. Therefore, it is very important to fill channel A with a fluorescing solution that produces a stable signal that can

be detected at the gain settings used; 100 nM fluorescein is recommended as a default. Avoid trapping of air bubbles as they might interfere with the detection of the signal, which can typically lead to the loss of magnet angle or inability to lock the magnet angle at all. Once filled, the calibration cell can be used for several runs without changing the fluorescein solution. Please check regularly before each run whether a bigger air bubble has formed that might interfere with the detection of the signal.

3. As we observed sticking of FITC-labeled clamp loader to the material of the SEC column, we added 1 μM BSA to the running buffer, and by this, a significantly higher yield (two- to threefold) of labeled protein was obtained.
4. It is recommended to measure absorbance spectra instead of absorbance at single wavelengths only. Make sure that your spectra look as expected. Absorbing contaminations and light scattering due to aggregates might interfere with the determination of DNA and protein concentrations using UV spectroscopy. A good estimate to assess the purity and quality of DNA and protein solutions is the absorbance at 320 nm (A320) and the ratios of A260/A280 and A260/A230 and A260/A280, respectively. For DNA, an A260/A280 ratio about 1.8 and an A260/A230 ratio about 2 are expected for pure solutions. If A260/A230 is significantly lower, it may indicate the presence of contaminations that absorb at 230 nm, e.g., residual impurities from oligonucleotide synthesis. Such impurities can be removed by a purification step based on silica adsorption of the DNA. As A260/A280 might be influenced by the pH, it is recommended to measure the absorbance of nucleic acid solutions at a slightly alkaline pH [33]. For tryptophan-containing proteins, an A260/A280 ratio about 0.6 is expected for pure solutions. Higher A260/A280 ratios indicate a contamination of the protein solution with nucleotides or nucleic acids. Often, nucleotides and nucleic acids can be removed from protein preparations by $(NH_4)_2SO_4$ precipitation of the protein. High A320 values might be the result from light scattering of aggregates, which can be removed by high-speed centrifugation or passing the solution through a sterile filter (e.g., PDVF membrane, 0.22 μm pore size).
5. Since fluorescence is detected best in the solution directly beneath the upper window (*see* Subheading 1.1), the use of 12-mm two-channel centerpieces will require an unnecessary high sample volume (400 μL). The usage of 3 mm centerpieces will reduce the required volume by a factor of four, but standard Beckman Coulter AUC cell housings require in this case the placement of a 4.5-mm spacer ring between the screw ring and the upper window assembly to ascertain that the sample

can still be applied through the filling holes. However, lowering the focus height of the FDS to that position would worsen the shadowing effect by the centerpiece and the screw ring [34] and would interfere with the detection of the signal of the calibration strip. Therefore, use of special fluorescence cell housings available from Nanolytics Instruments, where the filling holes are shifted by 4.5 mm to the top of the housing and a 9 mm elevation is included at the bottom of the housing (Fig. 1c), is recommended. These allow the use of 3-mm centerpieces with the FDS without the need for shifting the focus height and for any extra spacers. Alternatively, fluorescence housings can be purchased from SpinAnalytical. However, in this case, the use of additional spacers is necessary since the bottom of the housing is not adjusted.

6. When looking on the screw ring of an AUC cell with the screw holes facing the top, the AOS software of the FDS will name the left channel A whereas the right channel will be named B.
7. It may be worth reserving one channel in the run for a buffer reference to check whether there are fluorescent sedimenting impurities in the reaction buffer.
8. There is no need to place the FDS calibration cell in a special hole of the rotor since the position of the calibration cell can be freely chosen in the "Rotor Setup" menu. However, if you would like to measure the position of the menisci after the run with the absorbance optics of the centrifuge, you should avoid placing the calibration cell in hole 8 since this may interfere with delay calibration. In case you want to use both optics, you have to mount the absorbance optics before the FDS optics.
9. If you just press the "vacuum" button for starting evacuation and temperature equilibration instead of pressing the "start" button, only the vacuum pump, but not the diffusion pump, is turned on. The actual rotor temperature, however, can only be measured if the vacuum is below 100 mTorr; therefore, temperature equilibration might be delayed. Above 100 mTorr, the temperature of the vacuum chamber is measured that might substantially deviate from the rotor temperature.
10. Since the calibration strip is quite shallow and the z-resolution of the FDS is low, the intensity of its focus scan is bell-shaped and the maximum represents the optimal focus position. This position is above the optimal focus position of a sample containing centerpiece. Therefore, it is recommended to perform the actual measurements at the optimal focus position of the sample containing cells. If 3 mm centerpieces are used, a bell-shaped curve is also expected due to the low z-resolution of the FDS. In case when 12 mm centerpieces are used and a decrease

in fluorescence intensity with increasing focus depth is observed, this is a sign of an inner-filter effect (*see* Subheading 1.1), indicating that the fluorophore concentration is too high or that the sample scatters light significantly.

11. If the samples differ significantly in fluorescence intensity, different settings can be used for different cells. Since a change of the PMT voltage will make it impossible to compare the signals from different cells, it is recommended to change the gains only. However, it has to be taken into account that each extra setting requires an extra scan round for those cells, thereby limiting the number of scans that can be taken from an individual cell before the sample is completely sedimented.
12. For AUC, centrifuge settings are quoted in terms of rpm rather than *g*, as the g-force on a sedimenting particle depends on its radial position, and the concentration of the particle is measured as a function of the distance from the center of rotation. Therefore, no single value for the g-force can be given. A rotor speed of 40,000 rpm is optimal for analysis of the interaction of the clamp loader and the SSB-saturated template/primer DNAs we used in this analysis. If you use other proteins or DNAs, it might be necessary to adjust the rotor speed to higher or lower values. The maximal speed of the An-50 Ti rotor is 50,000 rpm. Be aware that sedimentation velocity increases quadratically with rotor speed.
13. We found that the stability of the signal is improved if the laser is turned on 30 min before the start of the experiment. Make sure not to wait any longer since after the system is idle for 30 min, the laser will turn off automatically.
14. In AUC with fluorescence detection, the meniscus position is not as easy to see as in absorbance measurements where due to light diffraction a noticeable optical artifact can be seen. While usually there is no problem to determine the exact meniscus position by fitting in SEDFIT, it is sometimes helpful to measure the meniscus position by an absorbance scan after the actual run. In such cases, it is best to mount the absorbance optics together with the fluorescence optics before the run and to switch after the run, without stopping the rotor, to the Beckman Coulter control software and take absorbance scans from all cells at 320 nm. Since both sectors are filled with samples, intensity mode should be used.
15. If the reaction is slow on the time scale of sedimentation, no reaction boundary will be observed and both interaction partners and the complex will sediment in extra boundaries. Since only one reaction partner is labeled, in this case two peaks are expected in the $c(s)$ distributions: one sedimenting with the s-value of the free labeled macromolecule and the other one

with that of the complex. In this case, the s-values will not change with loading concentrations and binding affinity can be easily estimated by determination of the concentrations of free interaction partner and complex from the area under the peaks in the $c(s)$ distributions. However, it is important to check whether the quantum yield of the labeled molecule is changed by complex formation. Two peaks could also occur in case of a fast reaction if the labeled molecule has a lower s-value than the unlabeled interaction partner. Depending on the concentrations used, peaks of free labeled molecules as well as of the reaction boundary can occur in the $c(s)$ distributions. In this case, the s-value of the reaction boundary is expected to change with loading concentrations. In such a case, it would be easier to label the faster sedimenting partner so that it can sediment in a constant bath of an excess of the slower one.

16. Since SEDPHAT is not able to handle a concentration of 0 in its fitting procedure, a very low concentration of A (e.g., 0.001 μM) has to be entered into the file for the sample where only B was present.
17. There are also other possibilities to produce such *.isotherm files that can be loaded in SEDPHAT. One way is to determine the sedimentation coefficient of the reaction boundary s_{rb} by integration of the peaks of all individual $c(s)$ distributions in SEDFIT by using the "integrate" button and using the same integration range for all distributions. Alternatively, you can calculate s_{rb} yourself in Microsoft Excel by performing numerical integration:

$$s_{rb} = \frac{\int c(s)s \ \ ds}{\int c(s)ds} \tag{6}$$

Afterward, you generate a *.isotherm file in an ASCII editor where in the first column the concentration of A in μM is given, in the second column the concentration of B in μM, and in the third column s_{rb} in Svedberg.

18. Even if the s-value of A is not required for the calculation of s_{rb} in the case where an excess of A is used (*see* Subheading 1.5), the sedimentation coefficients of both interaction partners should be known. The s-value of the free labeled interaction partner B can easily be determined as it is recommended to include free B as a sample in the run that you use for the examination of the interaction. Since the evaluation used here requires that the faster sedimenting molecule is titrated with the slower sedimenting one, the s-value of the free unlabeled partner A has to be determined beforehand in an extra AUC run with absorbance or interference detection. If possible, include the highest and lowest concentrations of A that you

apply in the run where you examine the interaction in order to rule out that oligomerization of A is taking place in the concentration range used.

19. In SEDPHAT, the decadic logarithm of the association constant K_A is varied. Since from Eq. 2.

$$K_A = \frac{1}{K_D} \tag{7}$$

just calculate the decadic logarithm of the inverse of your estimated value for K_D to get a starting value for the fit. A good starting value for s_{AB} would be a value somewhat higher than the highest value measured for s_{rb}.

20. In "Fitting Options" in the "Options" menu, you can choose between the simplex, Marquardt-Levenberg, and simulated-annealing algorithms for the nonlinear least-square fitting [29].

Acknowledgments

We gratefully thank Lidia Litz for excellent technical assistance, Dr. Dietmar J. Manstein for scientific and financial support, Dr. Natalie Naue for valuable discussions, and Dr. Hidde L. Ploegh and Dr. Joachim Orth for providing the SrtA expressing plasmid. This project was supported by a grant from the Deutsche Forschungsgemeinschaft (CU 35/1-1 to U.C.).

References

1. Cole JL (ed) (2015) Analytical ultracentrifugation, Methods in enzymology, vol 562. Academic Press, New York
2. Uchiyama S, Arisaka F, Stafford WF, Laue TM (eds) (2016) Analytical ultracentrifugation: instrumentation, software, and applications. Springer, Tokyo
3. Schuck P, Zhao H (2018) Sedimentation velocity analytical ultracentrifugation: interacting systems. CRC Press, Boca Raton
4. Schuck P, Zhao H, Brautigam CA, Ghirlando R (2016) Basic principles of analytical ultracentrifugation. CRC Press, Boca Raton
5. Kelman Z, Yuzhakov A, Andjelkovic J, O'Donnell M (1998) Devoted to the lagging strand-the χ subunit of DNA polymerase III holoenzyme contacts SSB to promote processive elongation and sliding clamp assembly. EMBO J 17:2436–2449
6. Naue N, Fedorov R, Pich A et al (2011) Site-directed mutagenesis of the χ subunit of DNA polymerase III and single-stranded DNA-binding protein of *E. coli* reveals key residues for their interaction. Nucleic Acids Res 39:1398–1407
7. Witte G, Urbanke C, Curth U (2003) DNA polymerase III chi subunit ties single-stranded DNA binding protein to the bacterial replication machinery. Nucleic Acids Res 31:4434–4440
8. MacGregor IK, Anderson AL, Laue TM (2004) Fluorescence detection for the XLI analytical ultracentrifuge. Biophys Chem 108:165–185
9. Nelson TG, Ramsay GD, Laue TM (2016) Fluorescence detection system. In: Uchiyama S, Arisaka F, Stafford WF, Lau T (eds) Analytical ultracentrifugation: instrumentation, software, and applications, 1st edn. Springer, Tokyo, pp 39–61
10. Gu Q, Kenny JE (2009) Improvement of inner filter effect correction based on determination

of effective geometric parameters using a conventional fluorimeter. Anal Chem 81:420–426

11. Lewis JS, Jergic S, Dixon NE (2016) The *E. coli* DNA replication fork. Enzyme 39:31–88
12. Schaeffer PM, Headlam MJ, Dixon NE (2005) Protein–protein interactions in the eubacterial replisome. IUBMB Life 57:5–12
13. Glover BP, McHenry CS (1998) The chi psi subunits of DNA polymerase III holoenzyme bind to single-stranded DNA-binding protein (SSB) and facilitate replication of an SSB-coated template. J Biol Chem 273:23476–23484
14. Georgescu RE, Kurth I, O'Donnell ME (2011) Single-molecule studies reveal the function of a third polymerase in the replisome. Nat Struct Mol Biol 19:113–116
15. Simonetta KR, Kazmirski SL, Goedken ER et al (2009) The mechanism of ATP-dependent primer-template recognition by a clamp loader complex. Cell 137:659–671
16. Schuck P (1997) Use of surface plasmon resonance to probe the equilibrium and dynamic aspects of interactions between biological macromolecules. Annu Rev Biophys Biomol Struct 26:541–566
17. Mao H, Hart SA, Schink A, Pollok BA (2004) Sortase-mediated protein ligation: a new method for protein engineering. J Am Chem Soc 126:2670–2671
18. Bloom LB, Turner J, Kelman Z et al (1996) Dynamics of loading the beta sliding clamp of DNA polymerase III onto DNA. J Biol Chem 271:30699–30708
19. Brown PH, Balbo A, Schuck P (2008) Characterizing protein-protein interactions by sedimentation velocity analytical ultracentrifugation. Curr Protoc Immunol Chapter 18:Unit 18.15
20. Dam J, Velikovsky CA, Mariuzza RA et al (2005) Sedimentation velocity analysis of heterogeneous protein-protein interactions: Lamm equation modeling and sedimentation coefficient distributions c(s). Biophys J 89:619–634
21. Naue N, Curth U (2012) Investigation of protein-protein interactions of single-stranded DNA-binding proteins by analytical ultracentrifugation. Methods Mol Biol 922:133–149
22. Schuck P (2010) Diffusion of the reaction boundary of rapidly interacting macromolecules in sedimentation velocity. Biophys J 98:2741–2751
23. Schuck P (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. Biophys J 78:1606–1619
24. Lohman TM, Overman LB (1985) Two binding modes in *Escherichia coli* single strand binding protein-single stranded DNA complexes. Modulation by NaCl concentration. J Biol Chem 260:3594–3603
25. Bogutzki A, Naue N, Litz L, Pich A, Curth U (2019) *E. coli* primase and DNA polymerase III holoenzyme are able to bind concurrently to a primed template during DNA replication. Sci Rep 9:14460
26. Ton-That H, Liu G, Mazmanian SK, Faull KF, Schneewind O (1999) Purification and characterization of sortase, the transpeptidase that cleaves surface proteins of *Staphylococcus aureus* at the LPXTG motif. Proc Natl Acad Sci U S A 96:12424–12429
27. Alegria-Schaffer A, Lodge A, Vattem K (2009) Performing and optimizing Western blots with an emphasis on chemiluminescent detection. Methods Enzymol 463:573–599
28. Brautigam CA (2015) Calculations and publication-quality illustrations for analytical ultracentrifugation data. Methods Enzymol 562:109–133
29. Zhao H, Schuck P (2015) Combining biophysical methods for the analysis of protein complex stoichiometry and affinity in SEDPHAT. Acta Crystallogr D Biol Crystallogr 71:3–14
30. Laue MT, Shah BD, Rigdeway TM, Pelletier SL (1992) Computer-aided interpretation of analytical sedimentation data for proteins. In: Harding SE et al (eds) Analytical ultracentrifugation in biochemistry and polymer science. The Royal Society of Chemistry, Cambridge, pp 90–125
31. Owczarzy R, Tataurov AV, Wu Y et al (2008) IDT SciTools: a suite for analysis and design of nucleic acid oligomers. Nucleic Acids Res 36: W163–W169
32. Zhao H, Mayer ML, Schuck P (2014) Analysis of protein interactions with picomolar binding affinity by fluorescence-detected sedimentation velocity. Anal Chem 86:3181–3187
33. Wilfinger WW, Mackey K, Chomczynski P (1997) Effect of pH and ionic strength on the spectrophotometric assessment of nucleic acid purity. BioTechniques 22:474–476, 478–481
34. Zhao H, Casillas E Jr, Shroff H et al (2013) Tools for the quantitative analysis of sedimentation boundaries detected by fluorescence optical analytical ultracentrifugation. PLoS One 8:e77245

Chapter 20

Studying RNA–Protein Complexes Using X-Ray Crystallography

Andrew P. Turnbull and Xiaoqiu Wu

Abstract

A wide range of biological processes rely on complexes between ribonucleic acids (RNAs) and proteins. Determining the three-dimensional structures of RNA–protein complexes is crucial to elucidate the relationship between structure and biological function. X-ray crystallography represents the most widely used technique to characterize RNA–protein complexes at atomic resolution; however, determining their three-dimensional structures remains challenging. RNase contamination can ruin crystallization experiments by degrading RNA in complex with protein, leading to sample heterogeneity, and the conformational flexibility inherent in both protein and RNA can limit crystallizability. Furthermore, the three-dimensional structure can be difficult to accurately model at the typical diffraction limit of 2.5 Å resolution or lower for RNA–protein complex crystals. At this resolution, phosphates, which are electron dense, and bases, which are large, rigid, and planar, tend to be well resolved and easy to position in the electron density map, whereas other features, e.g., sugar atoms, can be difficult to accurately position. This chapter focuses on methods that can be used to overcome the unique problems faced when crystallizing RNA–protein complexes and determining their three-dimensional structures using X-ray crystallography.

Key words X-ray crystallography, Crystallization, RNA-binding protein, RNA-binding domain, RNA–protein complex, RNA–protein interaction, Electrophoretic mobility shift assay, RCrane, COOT

1 Introduction

Many biologically important ribonucleic acids (RNAs) carry out their cellular functions in complexes with RNA-binding proteins (RBPs). RBPs account for between 5 and 10% of the eukaryotic proteome and play fundamental roles in many biological processes such as RNA metabolism, translation, DNA damage repair, and gene regulation. RBPs can bind many different types of RNAs, e.g., transfer RNA (tRNA), ribosomal RNA (rRNA), and messenger RNA (mRNA), with diverse structures ranging from single-stranded RNAs through to complex three-dimensional motifs and typically engage RNA through one or more RNA-binding domains (RBDs). In addition, a recent proteome-wide analysis has identified

Tina Daviter et al. (eds.), *Protein-Ligand Interactions: Methods and Applications*, Methods in Molecular Biology, vol. 2263, https://doi.org/10.1007/978-1-0716-1197-5_20,

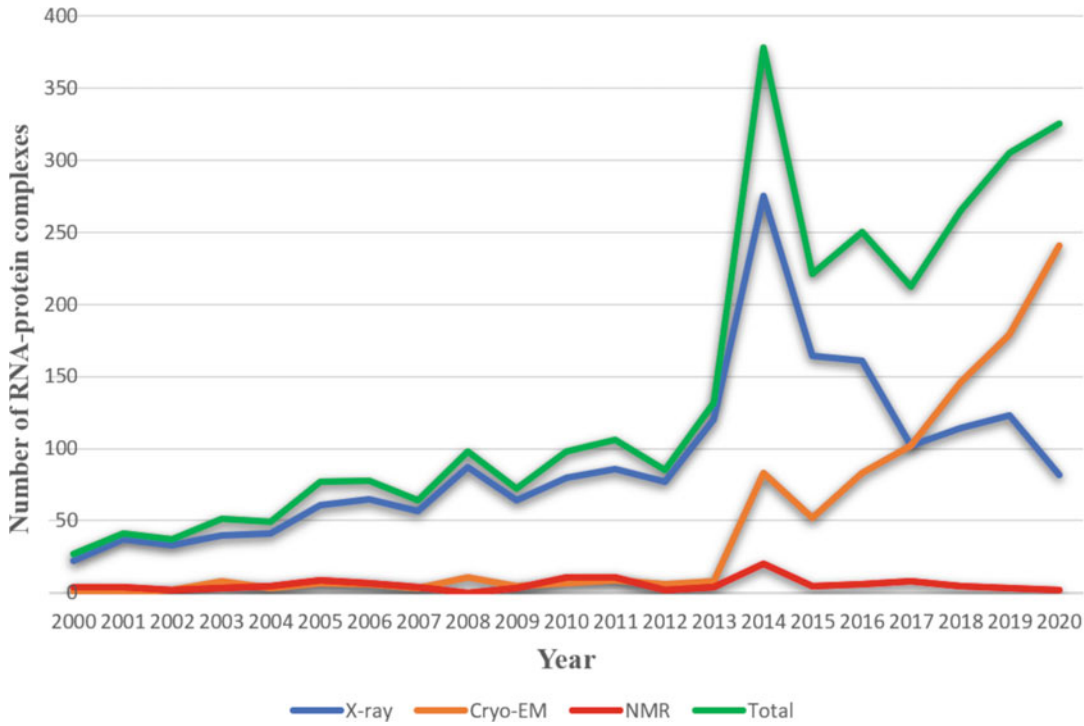


Fig. 1 Number of RNA–protein complexes deposited in the PDB per year between 2000 and 2020, plotted according to experimental technique (X-ray crystallography (X-ray), cryo-electron microscopy (cryo-EM), and nuclear magnetic resonance (NMR))

new classes of RBPs that mediate RNA interactions via disordered regions, protein–protein interaction interfaces, and enzymatic cores [1]. Determining the three-dimensional structures of RNA–protein complexes is critical to elucidate the relationship between structure and biological function. X-ray crystallography represents a powerful technique to determine the three-dimensional structures of RNA with unmodified full-length RBPs or RBDs at atomic resolution. The first RNA–protein complex to be crystallized was *E. coli* glutaminyl-tRNA synthetase in complex with its cognate tRNA [2], which represented a major breakthrough in the field. Since then (as of October 2020), 2971 RNA–protein complex structures have been deposited in the RCSB Protein Data Bank (PDB) [3], the majority of which have been determined using X-ray crystallography (63.6% compared with 32.4% and 4.0% determined using cryo-electron microscopy (cryo-EM) and nuclear magnetic resonance (NMR), respectively; *see* Fig. 1). These structures reveal detailed information about RNA–protein interactions, providing insights into biological function that can be used to design specific drugs targeting RNA molecules or RBPs implicated in disease [4]. However, the number of RNA–protein complex structures remains small compared with the number of RBPs in the PDB.

Crystallizing RNA–protein complexes is particularly challenging and can be hampered by RNase contamination, which can degrade RNA, the conformational flexibility of the protein and the inherent flexibility and negative charge of RNA. In addition, in cases where the RNA–protein complex comprises multiple subunits, the increased complexity can make it difficult to structurally characterize the complex. This chapter focuses on methods that can be used to overcome the unique problems faced when determining the three-dimensional structures of RNA–protein complexes using X-ray crystallography.

2 Materials

1. RNase-free stock reagents (*see* **Note 1**).
2. Ambion RNaseAlert® lab test kit to check buffers, RNA, and protein preps for RNase contamination (*see* Subheading 3.3).
3. Synthetic or in vitro transcribed RNA (*see* Subheading 3.4).
4. Purified, crystallization-grade protein (*see* Subheading 3.7). Protein can be produced in house, e.g., overexpressed in bacteria, insect, or mammalian cell systems, and purified using, e.g., an ÄKTA protein purifier (GE Healthcare Life Sciences), obtained through a collaboration with an academic group or outsourced to a contract research organization (CRO) as a fee-for-service.
5. Spectrophotometer (e.g., Thermo Scientific™ NanoDrop™ or Perkin Elmer LAMBDA™ 465) to accurately determine RNA and protein concentrations and to ensure that protein samples are RNA-free (*see* Subheading 3.5).
6. Temperature-controlled, vibration-free crystallization incubator or temperature-controlled room to maintain a constant temperature for growing crystals during the time course of the crystallization experiment.
7. Crystal screening reagents including sparse matrix screening kits (*see* **Note 2**).
8. 24-well Linbro plates and siliconized glass coverslips or 96-well Society for Biomedical Sciences (SBS) format crystallization plates (e.g., Swissci crystallization plates, which feature conical wells with a lens effect for improved well visualization and easier crystal retrieval; https://www.swissci.com/), with access to a crystallization robot for automated plate preparation (*see* **Note 3**).
9. Cold light source stereo microscope to manually inspect crystallization plates or access to an automated crystallization imager with plate hotel (e.g., Formulatrix Rock Imager®).

10. Cryo loops, cryo tools (e.g., CryoWand and CryoTongs), cryoprotectants (e.g., ethylene glycol and glycerol), and access to liquid nitrogen, for handling, manipulating, cryoprotecting, and flash freezing crystals prior to screening and data collection.
11. Access to an "in-house" X-ray source or dedicated macromolecular synchrotron beamline for screening crystals and data acquisition.
12. Crystallographic programs to process data and refine three-dimensional structures including the CCP4 suite [5] and COOT [6, 7] (*see* **Note 4** for a list of useful crystallographic programs and other resources relevant for RNA–protein crystallography).

3 Methods

3.1 Tools to Predict the Fold of RNA

Predicting the fold of RNA and defining its length and composition for use in RNA–protein crystallization experiments can improve the likelihood of obtaining RNA–protein complex crystals. However, modeling RNA structure is hampered by its flexibility that allows it to adopt a wide variety of secondary and tertiary (three-dimensional) motifs. Various computational tools exist to predict RNA structure (*see* **Note 4**) including:

- *RNA BRICKS* comprises three-dimensional structure motifs of RNA including information about their contacts with other RNA motifs, proteins, metal ions, water molecules, or small molecule ligands. The database provides structure-quality score annotations and tools for RNA three-dimensional structure searches [8].
- *RNA 3D MOTIF ATLAS* is a comprehensive and representative collection of internal and hairpin loop RNA three-dimensional motifs extracted from representative sets of RNA three-dimensional structures [9].
- *RNA FRABASE 2.0* is an engine with database to search for three-dimensional fragments within three-dimensional RNA structures using sequence(s) and/or secondary structure(s) as input [10].

3.2 Modeling RNA–Protein Interactions

When RNA binds to a protein to form a complex, multiple interactions and conformational changes occur in both the RNA and protein. There are a variety of methods for macromolecular three-dimensional structure prediction that can be applied to RNA–protein complexes. Typically, atomic coordinates of experimentally determined RNA–protein complexes provide the basis for

computational analysis and prediction techniques. Computational methods can be used to predict RNA-binding sites on RBPs solved in the apoform (unbound) state, to dock RNA molecules into unbound RBPs, or to study the molecular dynamics of RNA–protein complexes. Such modeling exercises can help to predict the minimal RNA-binding protein fragment or domain to include in crystallography construct design. Several methods are available to predict RNA-binding sites on the structures of unbound RBPs (*see* **Note 4**) including:

- *IDEEPE* predicts RNA–protein-binding sites from RNA sequences by combining global and local convolutional neural networks (CNNs) and has been shown to identify experimentally verified binding motifs [11].
- *NPDOCK* (Nucleic acid–Protein Dock) is a web server for modeling RNA–protein and DNA–protein complex structures [12].

Furthermore, fragment-based approaches are being developed that are capable of docking highly flexible single-stranded RNA bound to a protein based on the protein structure, RNA sequence, and conserved contacts [13]. These predictions can be followed up experimentally to confirm that RNA binds to the RBP or RBD (*see* Subheading 3.8).

3.3 Preparing RNase-Free Solutions

Even minute amounts of RNase contamination can ruin crystallization experiments with RNA by degrading RNA in complex with protein leading to sample heterogeneity. Therefore, it is advisable to use disposable RNase-free certified plasticware throughout protein production to minimize the likelihood of RNase contamination (*see* **Note 1**). Buffers, RNA, and protein preps can be conveniently checked for RNase activity using the Ambion RNaseAlert® lab test kit, which uses a fluorescent cleavable RNA oligonucleotide (excitation/emission maxima 490/520 nm) to detect RNase contamination. The test involves pipetting 5 μL of 10× RNaseAlert® lab test buffer into a tube containing the lyophilized RNA oligonucleotide substrate [14]. Subsequently up to 45 μL of the solution to be tested can be added to the tube and incubated for 30–60 min at 37 °C. Solutions with RNase contamination will cleave the substrate producing a fluorescent signal that is proportional to RNase activity. Most solutions can be tested apart from solutions that interfere with the fluorophore's excitation or emission, e.g., gel loading buffers, or inhibit RNases, e.g., high ionic strength solutions or solutions outside the pH range 4–9. Furthermore, crystallographic protein samples can be incubated with RNA oligonucleotides at room temperature for several days to ensure that there is no RNase activity and to confirm that the integrity of

the RNA is retained. Ideally, more than 95% of the RNA sample should remain intact during the time course of the crystallization experiment.

3.4 RNA Synthesis

RNA for crystallization experiments can be obtained from two sources: chemical synthesis and in vitro runoff transcription using T7 RNA polymerase, which is routinely used to produce milligram quantities of RNA. Short RNAs can be synthesized chemically and purchased from companies such as Integrated DNA Technologies Inc. (https://www.idtdna.com) and IBA LifeSciences GmbH (https://www.iba-lifesciences.com). Furthermore, nucleotide modifications can be site-specifically incorporated during chemical synthesis and can be exploited in structure determination and functional studies. However, chemical synthesis is limited to RNAs up to 30 nucleotides in length—any longer and abortive products accumulate with each cycle of nucleotide addition [15]. Larger RNAs can be generated by in vitro transcription using bacteriophage RNA polymerases [16]. Typically, the RNA is dissolved in Ambion RNAse-free water at 4 mM final stock concentration and stored at −80 °C in single-use aliquots.

3.5 Determining the RNA Concentration

The quantification of nucleic acids is typically preformed using a spectrophotometer by measuring absorbance at 260 nm. The ratio of absorbance at 260 nm and 280 nm can be used to assess the purity of large nucleic acid molecules (both DNA and RNA). A ratio of ~2.0 is generally accepted as "pure" for RNA, whereas a value below 1.8 usually indicates the presence of contaminants. The UV spectrum features a prominent peak at 260 nm and a characteristic trough at 230 nm (*see* Fig. 2). However, the UV spectrum for shorter oligonucleotides is highly dependent on base composition, base order, and sequence length, which influence the final absorbance of the oligonucleotide undergoing quantification. Therefore, oligo-specific conversion factors must be applied in the calculation, otherwise the accuracy of the results can vary by as much as 5–10%. Furthermore, corrections for modifications such as 5′ or 3′ fluorophores must be applied when determining the concentration of RNA since many of these modifications absorb light in the UV/visible regions that can affect quantification results.

To determine stock concentrations of RNA, the Beer-Lambert Law can be used:

$$A_{260} = \varepsilon_{260} c l$$

where A_{260} = absorbance at 260 nm; ε_{260} = molecular extinction coefficient of the oligonucleotide at 260 nm (L/mol cm); c = concentration (mol/L); l = pathlength (cm).

Most manufacturers of synthetic RNA provide the extinction coefficient, which can be substituted into the equation above. Spectrophotometers such as the NanoDrop™ (Thermo Scientific)

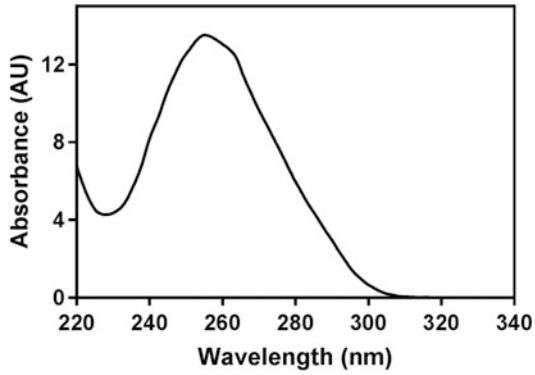


Fig. 2 NanoDrop™ spectrum for a synthetic 20-nucleotide, single-stranded RNA (purchased from IBA LifeSciences). The UV spectrum features a prominent peak at 260 nm and a characteristic trough at 230 nm. $\varepsilon_{260} = 234{,}700\ M^{-1}\ cm^{-1}$; $A_{260} = 13.05$; calculated RNA concentration = 55.6 μM. Data provided by Dr. Tobias Schmidt (Cancer Research UK Beatson Institute, Glasgow, UK)

and LAMBDA™ 465 (Perkin Elmer) offer simple procedures to measure absorbance and accurately determine RNA concentration and purity. In cases where the RNA oligonucleotide contains a modified base that absorbs at 340 nm, baseline correction should be turned off on the NanoDrop™ to obtain more accurate results.

3.6 Construct Design

Many RNA–protein complexes comprise multiple domains connected via flexible peptide linkers. To improve the chances of successfully crystallizing a novel RNA–protein complex, it is important to produce and screen multiple construct variants with carefully selected protein domain boundaries, e.g., constructs for the full-length protein or individual RBD, and to sample a variety of different lengths of RNA [15]. Each construct should retain the key features of the RNA–protein complex including the regions implicated in RNA–protein interactions and other functionality. Ideally, the ability of constructs to bind RNA should be confirmed using orthogonal biochemical and biophysical assays, e.g., the electrophoretic mobility shift assay (EMSA), surface plasmon resonance (SPR), and nuclear magnetic resonance (NMR) (*see* Subheading 3.8). A BLAST search (https://blast.ncbi.nlm.nih.gov) can be used to determine protein domain boundaries by comparing a protein sequence with structurally characterized proteins and protein domains in the PDB. In addition, limited proteolysis followed by mass spectrometry analysis can be used to digest free and RNA-bound protein in order to experimentally identify RNA-binding proteolytic fragments. Subsequently, a series of constructs based on the experimentally determined domain boundaries can be generated and evaluated for expression and solubility.

3.7 Protein Production

Pure RNA and RNase-free protein (greater than 95% purity based on SDS-PAGE analysis) are required for downstream RNA–protein complex formation and crystallization experiments. Care must be taken when producing recombinant RBPs and RBDs for crystallization experiments to ensure that protein samples are homogeneous and that host RNA contaminants are not carried over from the purification. RNA contamination can be removed using a gel filtration and desalting step to separate the larger protein from RNA contaminants and/or anion (MonoQ, Q-Sepharose HP) or cation (MonoS, SP sepharose) exchange chromatography. Anion exchange will bind the RNA contaminants to the resin whereas cation exchange will typically bind the protein with RNA contaminants eluting in the flow through. The 260/280 nm absorbance ratio determined using NanoDrop™ (*see* Subheading 3.5) provides a useful indication whether there is RNA contamination in the protein solution during purification. The A_{260}/A_{280} ratio should be 0.57 for pure proteins, whereas the ratio rises quickly if contamination exists (*see* Fig. 3). Furthermore, it is critical to ensure that proteins are free from RNases, which can degrade RNA when the RNA–protein complex is formed. The Ambion RNaseAlert® lab test kit provides a simple and sensitive RNase contamination assay to evaluate chromatographic fractions (*see* Subheading 3.3). Subsequently, fractions identified as having RNase contamination can be excluded from downstream purification steps.

3.8 RNA Binding Assays

Proteins interact with RNA through electrostatic interactions, hydrophobic interactions, hydrogen bonding, and base stacking in a similar fashion to protein–DNA interactions. Protein–RNA interactions are also influenced by the tertiary structure of the RNA. Therefore, both the RNA and protein(s) must be correctly folded in order to facilitate the correct RNA–protein interactions. RNA is prone to degradation and care must be taken not to introduce RNases into the reaction (*see* **Note 1** and Subheading 3.3). The most common methods for confirming and studying RNA–protein interactions are described below.

3.8.1 EMSA

EMSA (also known as the gel retardation assay or band shift assay) is an in vitro technique that can be used to detect RNA–protein interactions by monitoring the electrophoretic mobility of RNA and to estimate kinetic parameters, e.g., dissociation constants, for RNA–protein complexes [17]. RNA–protein complexes usually remain intact when fractionated by gel electrophoresis, provided that the affinity of the complex is high enough, and migrate with reduced mobility on non-denaturing gels compared with free (unbound) RNA (*see* Fig. 4; for a comprehensive method for running electrophoretic mobility shift assays, *see* ref. 14 and Chapter 14). RNAs can be labeled with radioisotopes, covalent or non-covalent fluorophores, or biotin and detected using

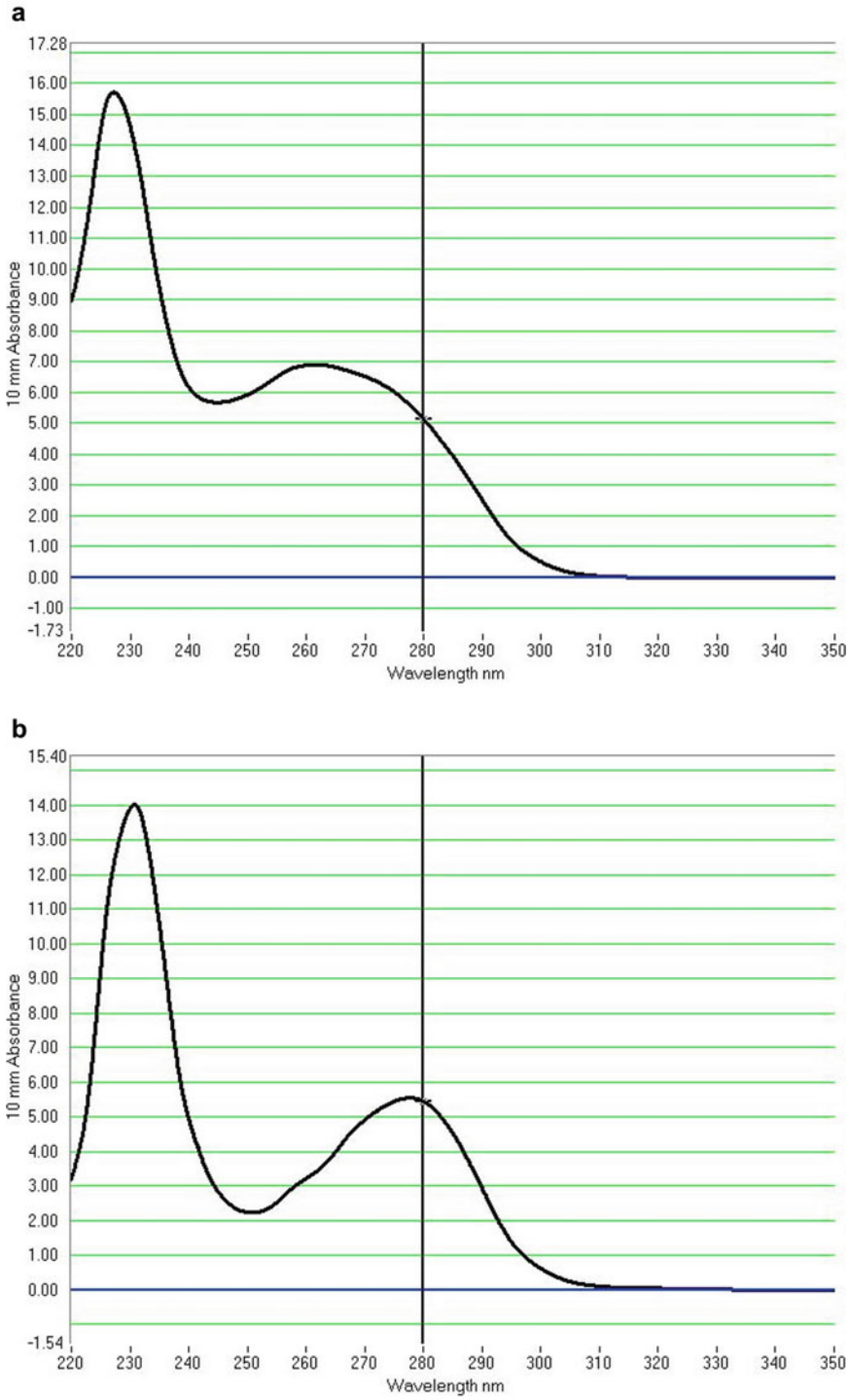


Fig. 3 NanoDrop™ spectrum for different RBP preps (protein molecular weight = 43 kDa). (**a**) 260/280 nm ratio = 1.33. Protein concentration = 6.4 mg/mL. The high 260/280 nm ratio is indicative of RNA contamination, and this apoform protein sample did not crystallize. (**b**) A different prep for the same protein. 260/280 nm ratio = 0.59. Protein concentration = 6.8 mg/mL. The 260/280 nm ratio is much lower compared with the sample in (**a**), indicating that there is little or no RNA contamination. In this case, the purified apoform protein sample crystallized. Data provided by Dr. Paul Owen (Cancer Research UK Therapeutic Discovery Laboratories, London, UK)

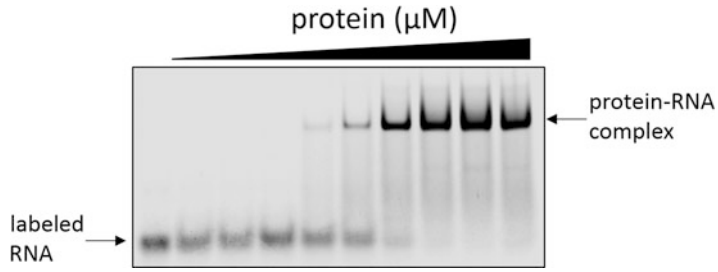


Fig. 4 Electrophoretic mobility shift assay using a custom IR-labeled, 20-nucleotide single-stranded RNA (purchased from IBA LifeSciences) and detected using a LI-COR scanner. The band shift observed at higher protein concentrations indicates that the labeled RNA is forming a complex with the protein. Data provided by Dr. Tobias Schmidt (Cancer Research UK Beatson Institute, Glasgow, UK)

autoradiography, fluorescence imaging, chemiluminescent imaging, and/or chromophore deposition, respectively. Radiolabeling the nucleic acid at the 5′- or 3′-end with ^{32}P represents a widely used, sensitive and inexpensive approach. RNA can also be covalently labeled with a fluorophore in labs where radioisotope use is not permitted, and, with modern instrumentation, the detection of fluorescent probes now rivals that of radioisotope-labeled RNAs. In addition, unlabeled RNAs can be used and detected by post-electrophoretic staining with chromophores or fluorophores that bind to the RNA. First, the RNA probe (*see* **Note 5**) is incubated with a purified protein sample to initiate binding and formation of the complex. The RNA–protein complex migrates through the gel matrix more slowly than the free RNA probe, resulting in a migration shift relative to the non-bound RNA probe that can be detected via the label or post-electrophoretic staining. The polyacrylamide or agarose gel composition and percentage may need to be varied to optimize the separation of protein-bound and free RNA. In addition, specificity can be determined using a competition reaction, where excess unlabeled RNA identical to the labeled RNA is incubated in the binding reaction, resulting in a decrease in the shifted signal when the labeled and unlabeled RNA sequences compete for binding to the same protein. In order to use the electrophoretic mobility shift assay to estimate dissociation constants for RNA–protein interactions, pure RBPs or RBDs are required. Furthermore, the amount of free or protein-bound RNA must be estimated by measuring the intensity of the bands, e.g., in the autoradiograph or using fluorescence imaging. Gel electrophoresis most likely affects the equilibrium between bound and unbound RNA, hence dissociation constant values derived from band shift assays are unlikely to represent absolute dissociation constant values and are best used for comparative studies, e.g.,

assessing relative dissociation constants for different RNA substrates prior to complex formation for crystallization experiments.

3.8.2 Biophysical Characterization of RNA–Protein Interactions

Biophysical characterization methods, e.g., SPR and NMR, can be used to confirm and characterize RNA–protein interactions. SPR represents a popular technique for determining kinetic rate constants (k_a and k_d) and equilibrium constants (K_d) based on mass concentration-dependent changes in refractive index at the sensor chip surface that are monitored in real time. Protein or RNA can be immobilized on the sensor chip surface to detect and monitor RNA–protein interactions. In addition, NMR can be used to determine kinetic parameters and derive structural information. SPR and NMR are covered in more detail elsewhere in this book (*see* Chapters 11 and 17).

3.9 Forming the RNA–Protein Complex

In principle, RNA–protein complexes can be prepared simply by mixing the RNA and protein constituents at the desired ratios and setting up crystallization trials. Alternatively, the individual components of the complex can be mixed at lower concentrations, with the complex subsequently purified using gel filtration chromatography and concentrated prior to setting up the crystallization experiments. For in vitro assembly of large multi-protein complexes, the individual components may need to be added in a specific order, with the assembly of the complex being monitored using native gel electrophoresis and mass spectrometry. Typically, the preferred method is to add RNA directly to the protein sample owing to the high cost of synthesized RNA. K_d is a measure of affinity between ligand and protein. $K_d = [P][L]/[PL]$ where [P], [L], and [PL] correspond to the molar concentrations of protein, ligand, and the protein–ligand complex, respectively. For RNA–protein complexes, K_d represents the concentration of RNA at which the RNA binding site is half occupied, i.e., $[PL] = [P]$. To ensure 90% fractional saturation of the RNA binding site, the free RNA concentration must be in excess of the free protein concentration and, at equilibrium, should not deplete to less than $10 \times K_d$. In practice, the high RNA concentrations used in RNA–protein complex crystallography experiments usually saturate the protein–RNA-binding site. A molar ratio of 1:0.9 RNA:protein is recommended as a starting point for initial crystallization screening [18].

3.10 Setting Up RNA–Protein Crystallization Experiments

The typical RNA–protein complex crystallography cascade is presented in Fig. 5. The first hurdle is obtaining well-ordered, diffraction quality crystals. The protein must be pure and monodisperse, and the RNA–protein complex must be homogeneous and should not noticeably degrade over the time course of the crystallization experiment. The most routine method for setting up crystallization experiments is vapor diffusion using hanging or sitting drops. Crystallization trials are usually carried out using protein

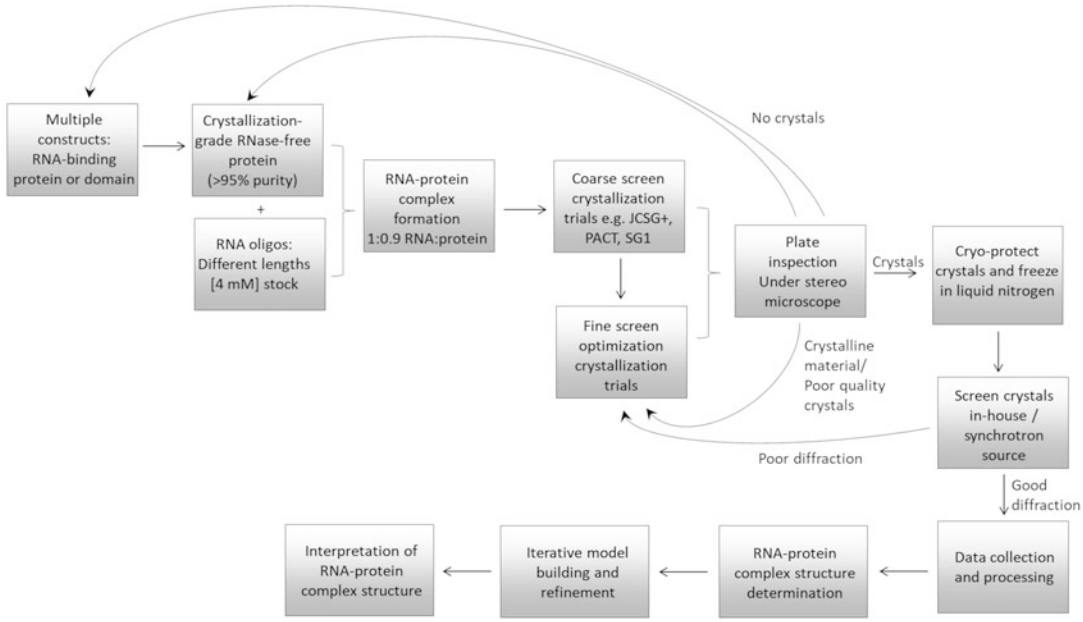


Fig. 5 RNA–protein complex crystallography cascade. Typically, multiple RBP or RBD constructs are produced and purified, ensuring that the samples are highly pure (greater than 95% purity based on SDS-PAGE analysis) and RNase-free. Subsequently, the complex is formed by mixing the RNA (usually at 4 mM stock concentration) with the purified protein sample at a ratio of 1:0.9. Coarse screen crystallization experiments are set up using commercial screens, e.g., JCSG+, PACT, and SG1, stored at a constant temperature in a vibration-free incubator or temperature-controlled room, and inspected under a stereo microscope over a time course ranging from days to weeks. Crystals can be cryoprotected using, e.g., ethylene glycol or glycerol, and flash frozen in liquid nitrogen prior to screening and data collection on an in-house or synchrotron source. Crystalline material, poor quality crystals or crystals with poor diffraction properties can be optimized by setting up fine screen crystallization trials based on preliminary crystallization conditions identified in the coarse screen crystallization experiments. If no crystals are observed, additional constructs can be purified and/or complexed with RNAs of varying lengths. Once a complete data set has been collected, the structure can be solved, e.g., using molecular replacement, built, e.g., using the RCrane plugin in COOT, and the three-dimensional RNA–protein complex structure analyzed to identify key interactions

concentrations ranging between 1 and 20 mg/mL. 24-well Linbro plates are commonly used to manually set up crystallization trials in laboratories without access to a crystallization robot. For hanging drop experiments, typically 0.5 or 1 μL of protein is mixed with an equal volume of reservoir on a siliconized glass coverslip, which is inverted and suspended over the reservoir solution. Alternatively, sitting drop experiments can be performed using micro-bridges (available from Hampton Research) placed in the Linbro plate, which allow for much larger drop sizes to be used compared with hanging drop experiments. The concentration difference between the undersaturated drop and the reservoir at a higher precipitant concentration results in vapor diffusion from the drop until the solution concentration matches that of the reservoir. Consequently,

the concentration of the protein and precipitant in the drop increases and, if the protein becomes supersaturated, crystals may form. Many laboratories now feature dedicated crystallization robots, such as the Mosquito® crystal (SPT Labtech) and NT8® (Formulatrix), that can dispense much smaller (nanoliter) volumes into 96-well SBS standard plates for automated screening of crystallization conditions. Compared with setting up crystallization experiments by hand, crystallization robots significantly improve accuracy and reproducibility, and the ability to dispense nanoliter volumes dramatically reduces protein sample consumption. Furthermore, crystallization robots can be used to rapidly set up commercially available coarse screens covering a broad range of reagents and screen formulations (*see* **Note 3**). For example, JCSG+ and PACT are widely used 96-condition coarse screens that provide a useful minimal crystallization screening strategy (*see* **Note 2**) [19]. In addition, the JBScreen Nuc-Pro (Jena Bioscience GmbH) and Natrix screens (Hampton Research) are designed to screen for preliminary crystallization conditions of RNA, DNA, and protein–nucleic acid complexes. A precipitation pattern ranging from clear to light/heavy precipitate usually indicates that the RNA–protein complex is at a suitable concentration for crystallization experiments: too dilute samples result in clear drops whereas too concentrated samples result in amorphous precipitate. The PCT™ (pre-crystallization test; Hampton Research) provides a useful means of determining the appropriate sample concentration to use prior to setting up crystallization experiments (*see* **Note 6**).

Crystallization conditions for published RNA–protein complexes as of 2007 were analyzed [16]. In general, RNA–protein complex crystals tend to grow at a neutral pH ranging between 6.5 and 7.5, with polyethylene glycol (PEG) representing the most common precipitant. Divalent ions including magnesium, calcium, and manganese may also be important components for RNA–protein crystallization. For example, magnesium ions are often seen to interact with the phosphate backbone of RNA in high-resolution X-ray structures of RNA or RNA–protein complexes. Conditions containing phosphate buffer and high salt concentrations, which can bind to and block RNA-binding sites on proteins, should be avoided since they can perturb RNA–protein complex formation and crystallization.

3.11 Optimizing Crystallization Conditions

Crystallization trays should be inspected periodically under a cold light source stereo microscope to check for the appearance of crystalline material or crystals, over a time course ranging from days to weeks. Preliminary conditions may be identified that produce single crystals that are sufficiently large (typically ≥50 μm in size for synchrotron sources or ≥100 μm for in-house sources) to be screened and evaluated on an X-ray system for diffraction properties. It may be possible to collect a complete data set from a single

crystal and subsequently determine the three-dimensional structure of the RNA–protein complex if the crystal diffracts to a reasonable resolution (preferably around 2.5 Å or higher). However, more commonly crystals do not diffract sufficiently well enough to collect a complete data set or are otherwise of poor quality, e.g., too small, or the crystals are multiple or split. In such cases, the preliminary coarse screen condition(s) must be fine-tuned by varying each crystallization reagent in order to attempt to improve crystal quality. Typically, the protein concentration, RNA:protein ratio, precipitant concentration, additive concentration, buffer pH, and temperature are varied. When several conditions are identified in the initial coarse screens that produce crystals with similar morphology, it is useful to cross-correlate conditions, e.g., pH, precipitant, or additives, to inform the design of follow-up fine (optimization) screens. Alternatively, if crystals with different morphologies are observed, the conditions should be treated independently since the resultant crystals may have different diffraction properties. Liquid handling robots such as the Dragonfly® (SPT Labtech) simplify the process of designing and dispensing optimization screens into 96-well SBS format plates and can dramatically speed up the optimization process.

3.12 Crystal Harvesting, Freezing, and Data Collection

When crystals are observed in the crystallization experiments, it is advisable to seek assistance from a colleague with a crystallographic background to aid crystal mounting and freezing since protein crystals tend to be fragile and can be difficult to handle and manipulate. Crystals must be stabilized and cryoprotected, typically using the reservoir solution supplemented with cryoprotectant, e.g., 20% ethylene glycol or glycerol, prior to harvesting in a cryo loop and flash freezing in liquid nitrogen (*see* **Note 7**). Frozen crystals can be stored long term in a liquid nitrogen dewar in vials or unipucks (*see* **Note 8**), and screened either on an in-house X-ray source or at a synchrotron beamline to evaluate diffraction quality. Typically, data extending to 2.5 Å resolution or higher are desirable for novel RNA–protein complexes so that a three-dimensional model for the complex can be accurately fitted into the electron density map. However, in more challenging cases, data at 3 Å resolution or lower may be sufficient to fit the overall fold of the protein and RNA. Data acquisition involves recording a series of X-ray diffraction images using detectors such as the Rigaku Saturn 944+ charge coupled device (CCD) and Dectris Eiger, which are commonly found on in-house and synchrotron sources, respectively. In-house data can be integrated and scaled to produce an MTZ file, for example, using the StructureStudio™ software platform for data collected on Rigaku systems or, alternatively, MOSFLM [20] and the CCP4 suite of programs (*see* **Note 4**) and its graphical user interface [5, 21]. Synchrotron data are usually processed on-the-fly using Xia2 [22], which is incorporated into the ISPyB (Information System for Protein CrystallographY Beamlines), LIMS (laboratory

information management system) on beamlines at Diamond (Didcot, UK), SOLEIL (Paris, France), EMBL (Hamburg, Germany), and MaxLab (Lund, Sweden) synchrotron sources. Furthermore, there are tools integrated into ISPyB that allow data to be reprocessed using the data reduction pipeline. For example, images with significant radiation damage can be excluded from data processing, the resolution limit of the data set can be adjusted or data can be reprocessed in an alternative space group.

In co-crystallization experiments with protein and nucleic acids (DNA and RNA), it is not always clear whether the crystal contains the intact complex or just a single component. The program RIBER/DIBER [23] (*see* **Note 4**) represents a useful tool to judge the nucleic acid content of a crystal based on the diffraction data prior to attempting to determine its three-dimensional structure. The program provides likelihood estimates of all possible crystal compositions, i.e., protein only, nucleic acid only, or the complex. RIBER is specifically tuned to detect the presence of regular RNA stems in macromolecular crystals and can be used with RNA–protein complex data provided that the resolution extends to 3 Å.

3.13 Determining the Structure

The phase problem represents the major hurdle to obtaining a three-dimensional structure of an RNA–protein complex once a complete data set has been collected. Commonly used techniques to solve protein structures include molecular replacement (MR), heavy-atom derivatization (isomorphous replacement, anomalous scattering, and anomalous dispersion), and direct methods. MR represents one of the most powerful phasing techniques and accounts for over 69% of all X-ray protein structures and 73% of RNA–protein complex structures deposited in the PDB (as of March 2019). MR requires a homologous structure for the protein under investigation, which is used as a search model to calculate initial estimates of the phases of the new structure. The search model can either be a homologous structure from the PDB or a homology model. Automated homology-based modeling servers include Swiss-model [24] and I-TASSER [25] (*see* **Note 4**). In cases where there are up to four molecules in the asymmetric unit of the crystal and the search model is structurally similar to the target protein (30% or higher sequence identity), it is relatively straightforward to use MR to determine the structure using programs such as PHASER [26]. One of the most widely used strategies for determining novel protein structures with limited or no sequence identity ($\leq$30%) to structures deposited in the PDB is selenomethionine incorporation, where methionine is substituted by selenomethionine in the protein during expression, and allows structures to be phased using the single-wavelength anomalous dispersion (SAD) technique. Traditionally, multiple or single isomorphous replacement (MIR or SIR) methods, in which heavy

atoms such as mercury, platinum, or gold are incorporated into the protein under investigation, can be used to calculate initial phase estimates provided that the derivatized crystal is isomorphous with respect to the native crystal. Furthermore, if RNA has been produced using solid-phase synthesis, uracil and cytosine can be substituted with 5′-bromo substituents to facilitate phase determination using multi-wavelength anomalous diffraction (MAD). The PHENIX [27] software suite is a highly automated system for macromolecular structure determination that can rapidly arrive at an initial partial model of a structure without significant human intervention, given moderate resolution and good quality data.

3.14 Model Building Using COOT

COOT is the most widely used crystallographic model building and validation program and is freely available for both academic and industrial users (*see* **Note 4**). COOT can be used for manual and automated model building, model completion, and validation [6, 7]. COOT displays electron density maps and models, e.g., protein and RNA, and allows the models to be manipulated to best-fit the electron density map through real-space refinement. RNA–protein crystals typically diffract to 2.5 Å resolution or lower. At this resolution, phosphates, which are electron dense, and bases, which are large, rigid, and planar, tend to be well resolved and easy to model whereas other features, e.g., sugar atoms, can be difficult to accurately position in the electron density map. In COOT, the interface to build ideal DNA and RNA is launched by selecting the "Ideal RNA/DNA..." option from the "Calculate > Other Modelling Tools" dialog. For a given sequence input by the user, a choice of DNA or RNA, A or B form, single or double stranded, is presented. Mutation of RNA bases in COOT can be performed using "Simple Mutate" from the "Calculate > Model/Fit/Refine. . ." dialog. "Calculate > Model/Fit/Refine. . . > Add Terminal Residue. . ." can be used to extend the nucleic acid chain.

RCrane represents a tool for semi-automated RNA model building into low or intermediate resolution electron density maps and is incorporated into COOT version 0.7 and above [28]. RCrane helps the user to place phosphates and bases into the electron density map and then automatically predicts and builds the detailed all-atom structure of the traced nucleotides (*see* Fig. 6). The RCrane plugin initially produces a low-level trace of the backbone and then calculates the coordinates for all the backbone atoms. RCrane uses three different means of analyzing the RNA structure: backbone conformer suites defined using approximate torsion values, RNA pseudo-torsions that simplify the RNA backbone, and base-phosphate perpendicular (Pperp) distance to determine sugar pucker. The pseudo-torsion and Pperp distance information are used to predict and build the appropriate conformer.

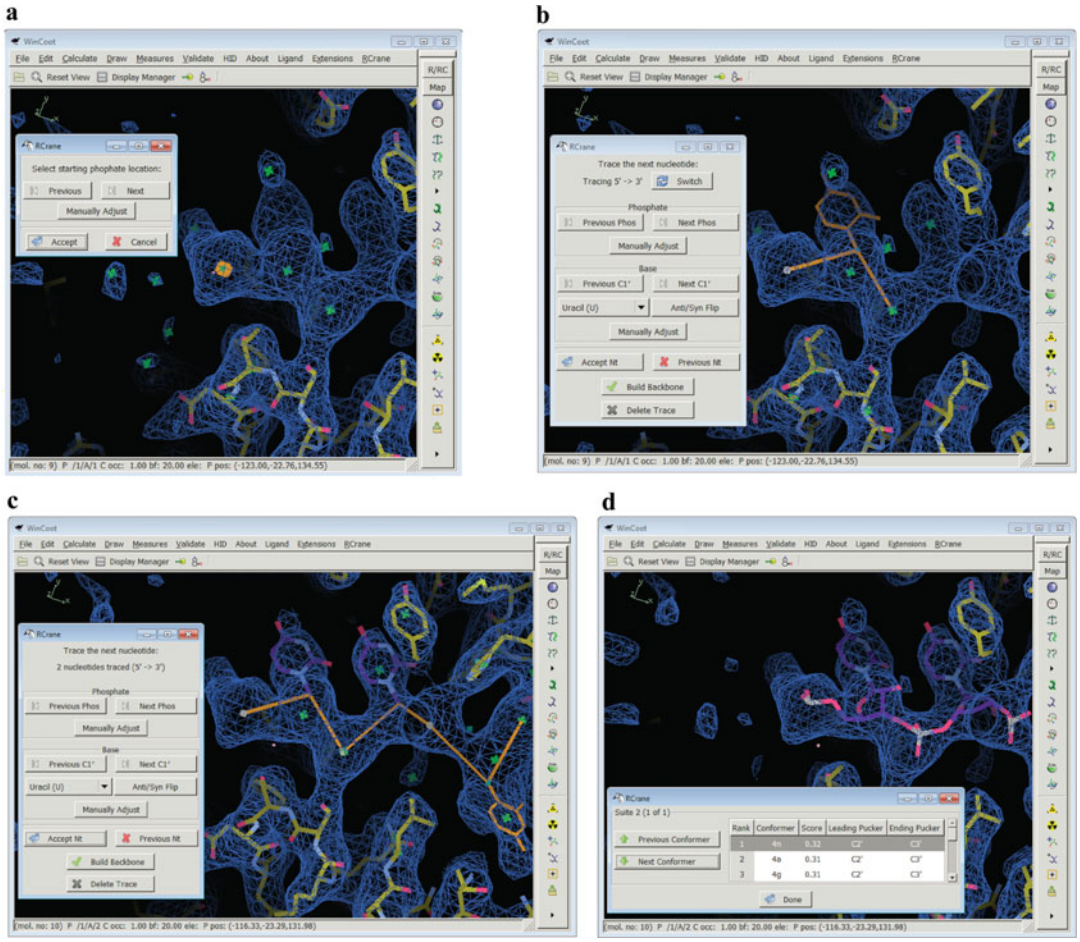


Fig. 6 Building an RNA chain using the RCrane plugin in COOT. The structure of Drosophila DHX36 helicase in complex with poly-uracil (PDB code = 5N94) [31] is used as an illustrative example in the screenshots using COOT version 0.8.9.2. Protein carbon atoms are colored yellow. (**a**) RCrane is launched by clicking on the "Calculate" tab in COOT and selecting "RCrane launch." The pointer in COOT is positioned for the starting phosphate, and "New Trace 5′→3′. . ." is selected from the "RCrane" menu. Peaks of electron density are marked with crosses. The orange cube represents the predicted phosphate position, which can either be accepted or altered using "Next," "Previous," or "Manually Adjust." (**b**) Following step (**a**), a new RCrane window will appear allowing the nucleotide chain to be traced. The current nucleotide position can be accepted using "Accept Nt" or adjusted using "Next Phos," "Previous Phos," or "Manually Adjust." The chain direction can be reversed using "Switch." Additional nucleotides are added to extend the chain. The virtual-bond backbone trace is shown in orange. (**c**) "Build Backbone" calculates the backbone atomic coordinates for the traced nucleotides. (**d**) Conformer suites and scores are listed in rank order. The "Done" button closes the RCrane window

3.14.1 Building an RNA Chain Using RCrane

1. Start COOT.
2. Open the coordinates for the RNA–protein complex using "File> Open Coordinates. . ." from the COOT menu bar, and select the appropriate PDB file.

3. Open the electron density map using "File> Auto Open MTZ. . . ." from the COOT menu bar, and select the appropriate MTZ file. Use "HID> ScrollWheel> Attach scroll wheel to which map? > 1. . . .FWT PHWT . . ." to change the direct electron density map root-mean-square deviation (rmsd) to 1.0 using the mouse scroll wheel (the rmsd contour level is displayed in the top of the graphics window). Use "HID> ScrollWheel> Attach scroll wheel to which map? > 2. . . . DELFWT PHDELWT" to change the difference electron density map rmsd to 3.0 using the mouse scroll wheel.
4. Click on the "Calculate" tab in COOT and select "RCrane launch." An "RCrane" menu appears next to the "Extensions" menu in COOT.
5. Press [CNTRL] + left mouse button to manually move the pointer in COOT to the position of a phosphate.
6. Select "New Trace 5′→3′. . ." from the "RCrane" menu to start an RNA trace.
7. Click "Accept" to accept the phosphate position or select an alternative predicted starting phosphate position using the "Next," "Previous," or "Manually Adjust" buttons. The starting phosphate position is depicted by an orange cube and represents the closest predicted phosphate to the pointer position in COOT. An RCrane window will appear for tracing the nucleotide chain (*see* Fig. 6a).
8. Select the appropriate base type—adenine (A), guanine (G), cytosine (C), or uracil (U)—from the drop-down list in the RCrane window. Click "Accept Nt" to accept the nucleotide position or use the "Next Phos," "Previous Phos," or "Manually Adjust" buttons to cycle through alternative positionings for the nucleotide. Click "Switch" to reverse the tracing, i.e., 3′→5′, if the chain direction is wrong (*see* Fig. 6b).
9. Repeat **step 8** to trace additional nucleotides in the chain (*see* Fig. 6c).
10. Click "Build Backbone" to calculate the backbone atomic coordinates for the traced nucleotides. During minimization, the bases and phosphates are allowed to move slightly while imposing torsion and sugar pucker restraints. An "Overview" popup window will appear with a ranked list of the most likely conformers (*see* Fig. 6d). Typically, the highest ranking conformer is correct. Conformer suites highlighted in orange or red indicate that there is a problem with the nucleotide conformation.
11. Use "Previous Conformer" or "Next Conformer" to calculate backbone coordinates for alternate conformers, which can be

visualized in COOT. Use "Previous Suite"/"Next Suite" to select different nucleotides in the chain.

12. Click "Done" once the conformers have been reviewed. This will close the RCrane window.
13. The RCrane-traced RNA molecule will appear in the COOT display manager as a separate molecule. Use "Edit > Merge Molecules" in COOT to merge the RCrane-traced RNA and protein into a single molecule.
14. The existing RNA chain can be extended in either the 3′ or 5′ direction using the "Extend chain…" option from the RCrane menu.
15. The "Rotamerize existing structure…" option from the RCrane menu can be used to rebuild the RNA structure, e.g., to fix sugar pucker errors or steric clashes.
16. The fit of the model of the RNA–protein complex to the electron density map can be improved using the "Calculate > Model/Fit/Refine… > Real Space Refine Zone" option in COOT.

3.15 Validating the RNA Structure

Various software tools exist to validate geometric parameters and interactions in order to identify errors, e.g., steric clashes, which need to be fixed during the model building and refinement process. For example, the MolProbity structure-validation web service (*see* **Note 4**) can be used to analyze the RNA backbone since it adopts rotameric conformations [29]. MolProbity uses the same Pperp distance technique as RCrane to check sugar puckers ($\geq$2.9Å = C3′-*endo* pucker; <2.9Å = C2′-*endo* pucker).

3.16 Analyzing the RNA–Protein Structure

Following refinement, the RNA–protein interface can be analyzed to provide insights into structure and function. In addition, the PRI HotScore web server (*see* **Note 4**), which uses in silico alanine scanning mutagenesis to calculate interaction scores for RNA–protein complexes, can be used to identify potential "hotspots" in the interface [30] that can be targeted to develop therapeutic agents against medically important RBPs.

3.17 What To Do When It Is Not a Complex…

Following crystallization, data collection, and phasing, inspection of the resultant electron density map may reveal that the data do not correspond to the intact RNA–protein complex but rather to an isolated component of the complex, i.e., the RBP, RBD, or RNA. This can occur if the complex dissociates during the time course of the crystallization experiment and one or more of the uncomplexed components crystallizes. In particular, pH can have a dramatic effect on the affinity and stability of the complex. In addition, unbound RNA, or RNA that has been degraded due to RNase contamination during the time course of the crystallization

experiment, may also crystallize. In cases where the crystal comprises only RNA, MR using, e.g., PHASER [26], and the protein as the search model will fail. The program RIBER/DIBER [23] (*see* **Note 4**) can be used to estimate possible crystal compositions after data collection, i.e., protein only, nucleic acid only, or complex, to confirm that it is likely to comprise the intact RNA–protein complex prior to structure determination. When the crystal does not correspond to the intact complex, any additional crystals that have been observed in different coarse screen crystallization conditions should be evaluated and collected on an X-ray source. A crystal that belongs to a different space group or which has different unit cell parameters may correspond to the intact complex. If no other crystals are available to assess, it may be necessary to repeat the crystallization experiments using a higher affinity RNA substrate (as determined using biophysical techniques) to prevent complex dissociation. By following the procedures outlined in this chapter and carefully considering the design of the RBP/RBD constructs and RNA, and confirming that the RNA and protein form a complex prior to setting up the crystallization experiments, the likelihood of obtaining crystals corresponding to the complex and subsequently determining its three-dimensional structure will be vastly improved.

4 Notes

1. The best way to prepare RNase-free solutions is to start with initial components that are supplied by vendors RNase-free and to use preparation techniques that prevent any possible RNase contamination. For example, whenever possible, RNase-free plasticware should be used. Stock solutions should be prepared using 0.22–0.45 μm-filtered deionized water, which should be RNase-free.
2. Sparse matrix coarse screens are used to scout for initial crystallization conditions. Commercial screens such as PACT and JCSG+ [19] are available from vendors including Molecular Dimensions Ltd. (Sheffield, UK), Jena Bioscience (Jena, Germany), and Hampton Research (Aliso Viejo, CA, USA).
3. Crystallization robots such as the SPT Labtech Mosquito® crystal and Formulatrix NT8® drop setter can be used for setting up crystallization trays in 96-well SBS format. Once preliminary crystallization conditions have been identified, it may be necessary to design follow up optimization screens to improve crystal quality. Access to liquid handling robots such as the SPT Labtech Dragonfly® can greatly speed up the optimization process.

4. Crystallographic programs and other resources relevant for RNA–protein crystallography (all accessed in March 2021):

   CCP4 SUITE. A collection of programs covering most computational aspects of macromolecular crystallography, which is freely available for academic use and can be downloaded from http://www.ccp4.ac.uk/.

   COOT. A macromolecular model building, model completion, and validation program, particularly suitable for protein modeling using X-ray data. COOT is widely used in the crystallographic community for manual and automated model building and is freely available for academic and industrial users from https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/ (Linux distribution) or http://bernhardcl.github.io/coot/wincoot-download.html (WINCOOT, Windows version). WINCOOT can also be installed as an optional component with the CCP4 software suite. Extensive documentation is available on the COOT Wiki page (https://strucbio.biologie.uni-konstanz.de/ccp4wiki/index.php/Coot).

   IDEEPE predicts RNA–protein binding sites from RNA sequences by combining global and local CNNs and identifies experimentally verified binding motifs (https://github.com/xypan1232/iDeepE).

   I-TASSER. A web-based or stand-alone software package for protein structure and function modeling (https://zhanglab.ccmb.med.umich.edu/I-TASSER/).

   MOLPROBITY. A structure-validation web service that provides evaluation of model quality at both the global and local levels for both proteins and nucleic acids (http://molprobity.biochem.duke.edu).

   MOSFLM. A program for integrating crystal diffraction data. Linux, Mac, and Windows distributions are freely available to academic and industrial users (https://www.mrc-lmb.cam.ac.uk/harry/imosflm/ver722/introduction.html).

   NPDOCK. A web server for modeling RNA–protein and DNA–protein complex structures (http://genesilico.pl/NPDock).

   PROTEIN DATABASE (PDB). Archive of experimentally determined structures (https://www.rcsb.org/).

   PHASER. A program for phasing macromolecular crystal structures with maximum likelihood methods, which is available through the PHENIX and CCP4 software suites.

   PHENIX. A software suite for the automated determination of molecular structures using X-ray crystallography and other methods (https://www.phenix-online.org/).

PRI HOTSCORE. Web server for the prediction of amino acid hotspots in RNA–protein interfaces using in silico alanine scanning mutagenesis (https://pri-hotscore.labs.vu.nl).

RCRANE. A plugin that facilitates semiautomated building of RNA structures within COOT (please refer to https://pylelab.org/software; RCrane is included with COOT 0.7 or newer).

RIBER/DIBER. A software suite for crystal content analysis in the studies of protein–nucleic acid complexes (http://diber.iimcb.gov.pl/).

RNA 3D MOTIF ATLAS. A comprehensive and representative collection of internal and hairpin loop RNA three-dimensional motifs extracted from representative sets of RNA three-dimensional structures (http://rna.bgsu.edu/rna3dhub/motifs).

RNA BRICKS. A database of RNA three-dimensional structure motifs and their contacts, both with themselves and with proteins. The database provides structure-quality score annotations and tools for three-dimensional RNA structure searches and comparison (http://iimcb.genesilico.pl/rnabricks/).

RNA FRABASE 2.0. An engine with database to search the three-dimensional fragments within three-dimensional RNA structures using the sequence(s) and/or secondary structure(s) as input (http://rnafrabase.cs.put.poznan.pl/).

SWISS-MODEL. A fully automated protein structure homology modeling server (https://swissmodel.expasy.org/).

5. ^{32}P-labeling represents a commonly used method for making RNAs detectable in the electrophoretic mobility shift assay in labs where radioisotope use in permitted. Labeled RNA probes can be prepared by in vitro transcription using [α-^{32}P]NTPs, whereas oligoribonucleotide or 100–200-nucleotide purified RNA can be end-labeled using [^{32}P]cytidine 3′-5′-bis(phosphate) or [γ-^{32}P]ATP. Alternatively, RNAs can be covalently labeled with a fluorophore in labs where radioisotope use in not permitted. Unlabeled RNAs can be used and detected by post-electrophoretic staining with chromophores or fluorophores that bind to the RNA.

6. Sample concentration is a significant crystallization variable. Samples that are too concentrated can result in amorphous precipitate whereas samples that are too dilute can result in clear drops. The PCT™ (pre-crystallization test; Hampton Research) can be used to determine the appropriate protein concentration for crystallization screening to reduce the number of precipitate and clear results and increase the likelihood of

crystallization. The PCT kit comprises four unique, preformulated, sterile filtered reagents. The protein sample is mixed with the reagents on a glass cover slide and suspended over each reagent in a sealed well in a Linbro plate, and the drops are inspected to determine if the protein concentration is suitable for crystallization screening. If all drops show heavy amorphous precipitate or are all clear, the protein sample can be diluted or concentrated, respectively, and the test repeated.

7. RNA–protein complex crystals tend to decay rapidly when exposed to X-rays; hence, data are collected at cryo-temperatures (100 K) in order to reduce radiation damage. Crystal freezing prevents free radicals that are formed when a crystal is exposed to X-rays from diffusing through the crystal and causing secondary radiation damage. A suitable cryoprotectant must be selected in order to prevent ice formation. Commonly used cryoprotectants include ethylene glycol and glycerol. A useful method to assess the effectiveness of a cryoprotectant is to mix 2 μL of 100% stock cryoprotectant with 8 μL of the crystallization reservoir solution to give a final cryoprotectant concentration of 20%, then freeze 2 μL of this solution in liquid nitrogen in a pipette tip. If the frozen solution appears to be clear in the tip, it should effectively cryoprotect the crystal, whereas if it appears opaque, the cryoprotectant concentration should either be increased or an alternative cryoprotectant should to be selected.
8. The unipuck holds 16 cryo loops and is compatible with most sample mounting robots including the BART sample changer used at the Diamond synchrotron light source, which is capable of storing 37 unipucks and a total of 592 samples.

Acknowledgments

We would like to thank Dr. Paul Owen at Cancer Research UK Therapeutic Discovery Laboratories (London, UK) for providing Fig. 3 and interpreting the data and Dr. Tobias Schmidt at the Cancer Research UK Beatson Institute (Glasgow, UK) for providing the data and analysis for Figs. 2 and 4.

References

1. Hentze MW, Castello A, Schwarzl T et al (2018) A brave new world of RNA-binding proteins. Nat Rev Mol Cell Biol 19:327–341
2. Rould MA, Perona JJ, Söll D et al (1989) Structure of E. coli glutaminyl-tRNA synthetase complexed with tRNA(Gln) and ATP at 2.8 A resolution. Science 246:1135–1142
3. Berman HM, Battistuz T, Bhat TN et al (2002) The protein data bank. Acta Crystallogr D Biol Crystallogr D58:899–907

4. Connelly CM, Moon MH, Schneekloth JS (2016) The emerging role of RNA as a therapeutic target for small molecules. Cell Chem Biol 23:1077–1090
5. Collaborative Computational Project, Number 4 (1994) The CCP4 suite: programs for protein crystallography. Acta Crystallogr D Biol Crystallogr D50:760–763
6. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr D60:2126–2132
7. Emsley P, Lohkamp B, Scott WG et al (2010) Features and development of Coot. Acta Crystallogr D Biol Crystallogr D66:486–501
8. Chojnowski G, Waleń T, Bujnicki JM (2014) RNA Bricks—a database of RNA 3D motifs and their interactions. Nucleic Acids Res 42: D123–D131
9. Petrov AI, Zirbel CL, Leontis NB (2013) Automated classification of RNA 3D motifs and the RNA 3D Motif Atlas. RNA 19:1327–1340
10. Popenda M, Szachniuk M, Blazewicz M et al (2010) RNA FRABASE 2.0: an advanced web-accessible database with the capacity to search the three-dimensional fragments within RNA structures. BMC Bioinformatics 11:231
11. Pan X, Shen HB (2018) Predicting RNA-protein binding sites and motifs through combining local and global deep convolutional neural networks. Bioinformatics 34:3427–3436
12. Tuszynska I, Magnus M, Jonak K et al (2015) NPDock: a web server for protein–nucleic acid docking. Nucleic Acids Res 43(W1): W425–W430
13. Chauvot de Beauchene I, de Vries SJ, Zacharias M (2016) Fragment-based modelling of single stranded RNA bound to RNA recognition motif containing proteins. Nucleic Acids Res 44:4565–4580
14. Rio DC (2014) Electrophoretic mobility shift assays for RNA-protein complexes. Cold Spring Harb Protoc 2014:435–440
15. Oubridge C, Ito N, Teo CH et al (1995) Crystallisation of RNA-protein complexes. II. The application of protein engineering for crystallisation of the U1A protein-RNA complex. J Mol Biol 249:409–423
16. Obayashi E, Oubridge C, Krummel DP et al (2007) Crystallization of RNA-Protein complexes. Methods Mol Biol 363:259–276
17. Yakhnin AV, Yakhnin H, Babitzke P (2012) Gel mobility shift assays to detect protein-RNA interactions. Methods Mol Biol 905:201–211
18. Ke A, Doudna JA (2004) Crystallization of RNA and RNA–protein complexes. Methods 34:408–414
19. Newman J, Egan D, Walter TS et al (2005) Towards rationalization of crystallization screening for small- to medium-sized academic laboratories: the PACT/JCSG+ strategy. Acta Crystallogr D Biol Crystallogr D61:1426–1431
20. Leslie AGW (1992) Recent changes to the MOSFLM package for processing film and image plate data. Jnt CCP4/ESF-EACMB Newslett Protein Crystallogr 26
21. Potterton E, Briggs P, Turkenburg M et al (2003) A graphical user interface to the CCP4 program suite. Acta Crystallogr D Biol Crystallogr D59:1131–1137
22. Winter (2010) xia2: an expert system for macromolecular crystallography data reduction. J Appl Crystallogr 43:186–190
23. Chojnowski G, Bujnicki JM, Bochtler M (2012) RIBER/DIBER: a software suite for crystal content analysis in the studies of protein-nucleic acid complexes. Bioinformatics 28:880–881
24. Waterhouse A, Bertoni M, Bienert S et al (2018) SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res 46(W1):W296–W303
25. Yang J, Yan R, Roy A et al (2015) The I-TASSER suite: protein structure and function prediction. Nat Methods 12:7–8
26. McCoy AJ, Grosse-Kunstleve RW, Adams PD et al (2007) Phaser crystallographic software. J Appl Crystallogr 40:658–674
27. Adams PD, Afonine PV, Bunkóczi G et al (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr D66:213–221
28. Keating KS, Pyle AM (2012) RCrane: semi-automated RNA model building. Acta Crystallogr D Biol Crystallogr 68:985–995
29. Davis IW, Leaver-Fay A, Chen VB et al (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res 35:W375–W383
30. Krüger DM, Neubacher S, Grossmann TN (2018) Protein-RNA interactions: structural characteristics and hotspot amino acids. RNA 24:1457–1465
31. Chen WF, Rety S, Guo HL et al (2018) Molecular mechanistic insights into Drosophila DHX36-mediated G-quadruplex unfolding: a structure-based model. Structure 26:403–415

Part VI

Membrane Binding

Chapter 21

Flow Linear Dichroism of Protein–Membrane Systems

Matthew R. Hicks, Sarah R. Dennison, Adewale Olamoyesan, and Alison Rodger

Abstract

Linear dichroism (*LD*) is the differential absorbance of light polarized parallel and perpendicular to an orientation direction. Any oriented sample will show a signal in its electronic as well as vibrational transitions. Model membrane small unilamellar vesicles or liposomes provide an oriented system when they are subject to shear flow in a Couette or other type of flow cell. Anything, including peptides and proteins, that is bound to the liposome also gives an *LD* signal whereas unbound analytes are invisible. Flow *LD* is the ideal technique for determining the orientation of different chromophores with respect to the membrane normal. To illustrate the power of the method, data for diphenyl hexatriene, fluorene, antimicrobial peptides (aurein 2.5 and gramicidin), are considered as well as another common chromophore, fluorene, often used to increase the hydrophobicity and hence membrane binding of peptides. How *LD* can be used both for geometry, structure analysis and probing kinetic processes is considered. Kinetic analysis usually involves identifying binding (appearance of an *LD* signal), insertion (sign change), often followed by loss of signal, if the inserted protein or peptide disrupts the membrane.

Key words Linear dichroism, Membrane bilayers, Liposomes, Electronic spectroscopy, Infrared spectroscopy

1 Introduction

Although there were 1522 membrane protein structures (though not independent proteins) in the Protein Data Bank (PDB) archive at the end of 2018, this is little more than 0.1% of all PDB structures. Even if we have a crystal or NMR structure of a membrane protein, the data do not necessarily provide information about the protein in an environment of interest. This is most obviously true for membrane-inserting peptides which are often unfolded in aqueous solution and adopt a helical structure in a membrane. A typical example is shown by the circular dichroism (*CD*) spectra in Fig. 1, where the spectrum of aurein 2.5 in water has a small negative *CD* signal at 200 nm characteristic of an unfolded peptide whereas the spectrum in a lipid environment (above the gel to fluid phase transition) has large negative signals at 222 nm and 208 nm and a

Tina Daviter et al. (eds.), *Protein-Ligand Interactions: Methods and Applications*, Methods in Molecular Biology, vol. 2263, https://doi.org/10.1007/978-1-0716-1197-5_21,

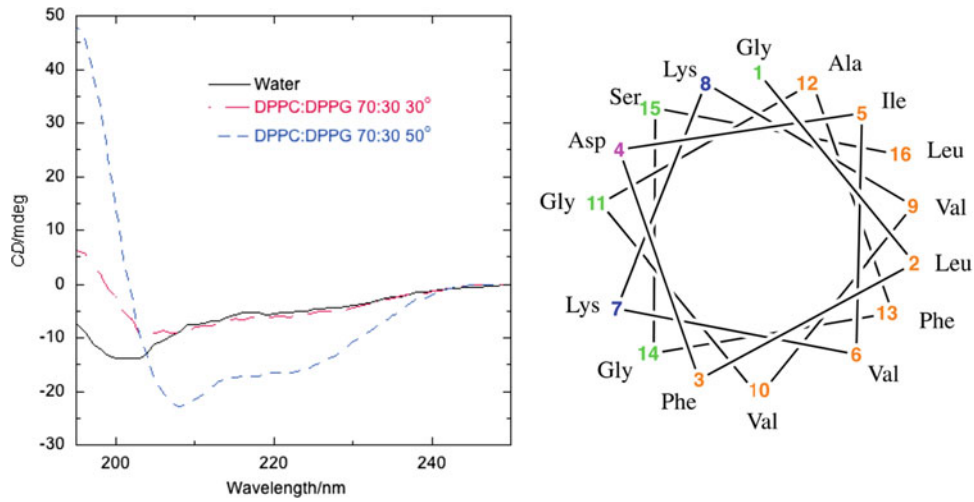


Fig. 1 Circular dichroism spectra of membrane-inserting peptide aurein 2.5 (illustrated as a helical wheel) in water and dipalmitoylphosphatidylcholine (DMPC)/dipalmitoylphosphatidylglycerol (DPPC) (70:30 molar ratio) liposomes below (at 30 °C) and above (at 50 °C) the gel to fluid phase transition. Data collected on a Jasco J-815 spectropolarimeter adapted for *LD* spectroscopy

large positive signal at about 192 nm characteristic of a highly α-helical structure. The gel phase (30 °C) spectrum, where insertion is less effective, suggests that aurein 2.5 is largely unfolded with evidence of some folded structure at 200 nm.

We are therefore almost desperate in our search for techniques that provide information about membrane proteins and peptides in membrane environments. The focus of this chapter is on flow linear dichroism (*LD*) spectroscopy of liposomes. *LD* signal is the difference in absorption, *A*, of light linearly polarized parallel (//) and perpendicular (⊥) to an orientation axis:

$$LD = A_{//} - A_{\perp} \tag{1}$$

When molecules are oriented in a sample, a particular spectroscopic transition can only absorb the light incident on it if its transition moment has a component parallel to the electric field of the light; if the direction of electron motion during the transition is perpendicular to the electric field, no transition could happen. *LD* is therefore positive when a transition's polarization is parallel to the sample orientation direction and negative when it is perpendicular to it. More generally the relationship between the *LD*, the isotropic absorbance, A_{iso}, (the absorbance of an unoriented solution of the same sample), the orientation factor *S* (which is 1 for perfect orientation and 0 for random samples), and the angle the transition moment of the molecule forms with the orientation axis, α, [1–3] is:

$$LD^r = \frac{LD}{A_{iso}} = \frac{3S}{2}\left(3\cos^2\alpha - 1\right) \quad (2)$$

where LD^r is referred to as the reduced *LD* and is concentration and path length independent as long as the *LD* and absorbance are measured on the same sample and the same path length and the Beer-Lambert law is valid. Note that the absorbance must only be of the oriented sample, not unoriented molecules in the sample.

Membranes are locally oriented systems. In order to make them macroscopically oriented for an *LD* experiment, they can be oriented using more-or-less dried films, squeezed gels, or Couette flow of liposomes (Fig. 2) [1–3, 5–9]. Flow orientation of aqueous solutions more closely resembles a biological environment than the other two methods (though dried films are often used for infrared spectroscopy). Liposomes at rest are spherical "balloons" whose walls are a lipid bilayer. When liposomes are subjected to shear flow, they adopt some form of ellipsoidal shape [3, 4, 7, 10], which has a net orientation for the lipids and any other molecules bound to the membrane. As the unique molecular orientation axis of such a system is the membrane normal (Fig. 2), rather than the flow direction, it is convenient to use a slightly different equation [3, 4, 7, 10].

$$LD^r = \frac{LD}{A_{iso}} = \frac{3S}{4}\left(1 - 3\cos^2\beta\right) \quad (3)$$

where β is the angle that the transition moment of interest makes with the membrane normal. Equation 3 is opposite in sign and half the magnitude of Eq. 2 as a consequence of using the membrane normal as the unique axis, rather than the liposome long axis.

When liposomes are oriented in shear flow, anything bound to the membrane also orients. This is illustrated in Fig. 2b where diphenyl hexatriene (DPH) is incorporated into the lipid bilayer parallel to the lipids. The stretched film *LD* at 350 nm is positive so the transition is polarized along the long axis of DPH [11, 12]. Its *LD* in the liposome is negative, consistent with it being inserted parallel to the lipids with a β value of ~0° [4].

In general, the key to interpreting a liposome flow *LD* spectrum is to know the transition polarizations of the molecules bound to the membrane. Some key chromophores for peptides and proteins are illustrated in Fig. 3. The α-helix shows a clear n-π^* transition at about 220 nm in both *CD* and *LD* polarized across the short axis of the helix. So, if the peptide is inserted into the membrane, it gives a positive *LD* signal. The other two transitions which can be accessed for an α-helix are two exciton components of the first π-π^* transition that occur at about 210 nm and 190 nm. A β-sheet, by way of contrast, shows little net *LD* at 220 nm due to two cancelling transition components, and the π-π^* transition is a single one a bit below 200 nm. *CD* can be used to identify the

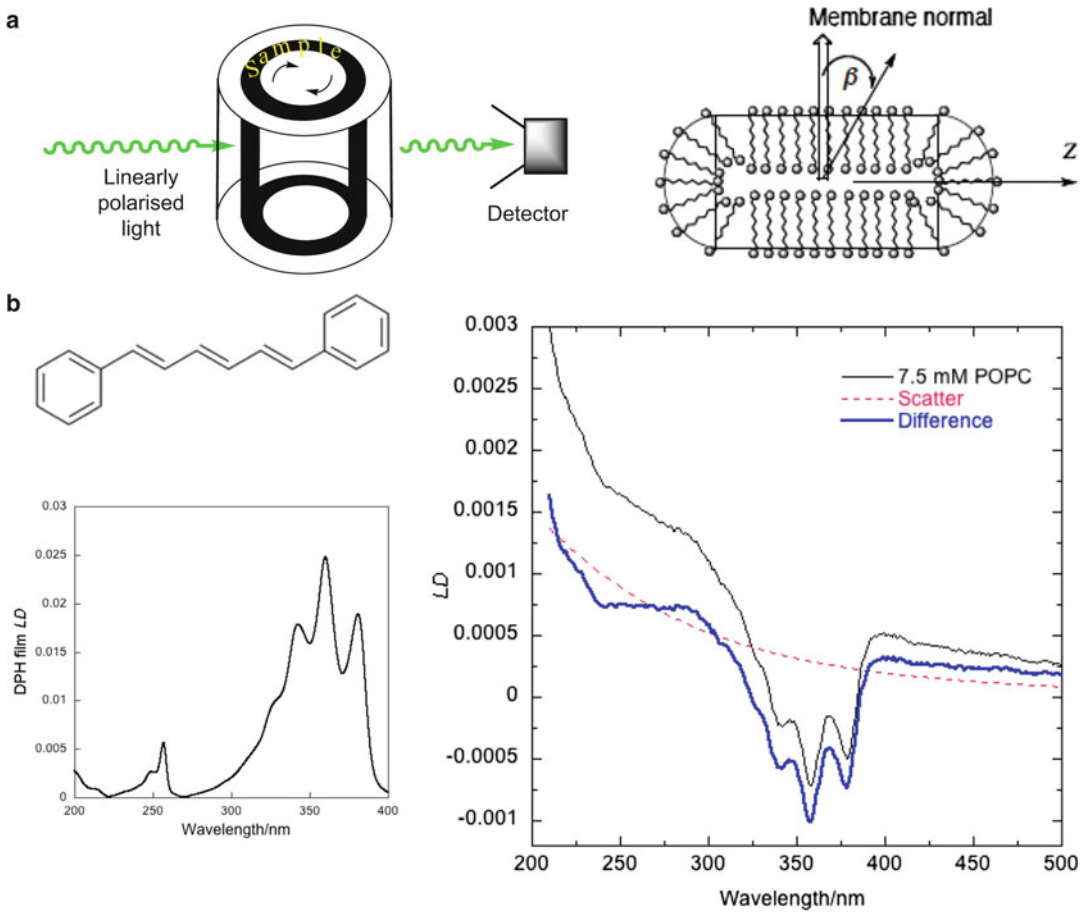


Fig. 2 (**a**) Experimental Couette flow schematic and the predominant local orientation of a lipid bilayer in the flow. *Z* is the macroscopic orientation axis (usually the horizontal direction). β is the angle between the membrane normal and the transition moment polarization. (**b**) Diphenyl hexatriene and its film *LD* together with overlay of the liposome flow absorbance *LD,* which includes a contribution from differential scattering as well as absorption. The scattering or turbidity *LD* is corrected using the method described in [4]. Liposomes are 100 nm POPC liposomes prepared with diphenyl hexatriene dissolved with the lipid (1%) in the membrane. Data collected on a Jasco J-815 spectropolarimeter adapted for *LD* spectroscopy

secondary structure of a peptide. So, if a peptide binds to a liposome, it will have *LD* signals for its transitions whose signs and magnitudes depend on the degree of orientation, *S*, and the orientation on the membrane (Eq. 3). *S* is typically of the order of 0.03.

If a helix is inserted into the membrane, then its 208 nm transition will be negative and the 195 nm and 222 nm transitions positive. However, the 208 nm transition is overlaid by the others and never actually becomes negative. Figure 4 summarizes the possibilities for α-helices and β-strands. Thus, the sign of peptide *LD* indicates its orientation with respect to the membrane normal. Other technique will need to be used to determine whether a

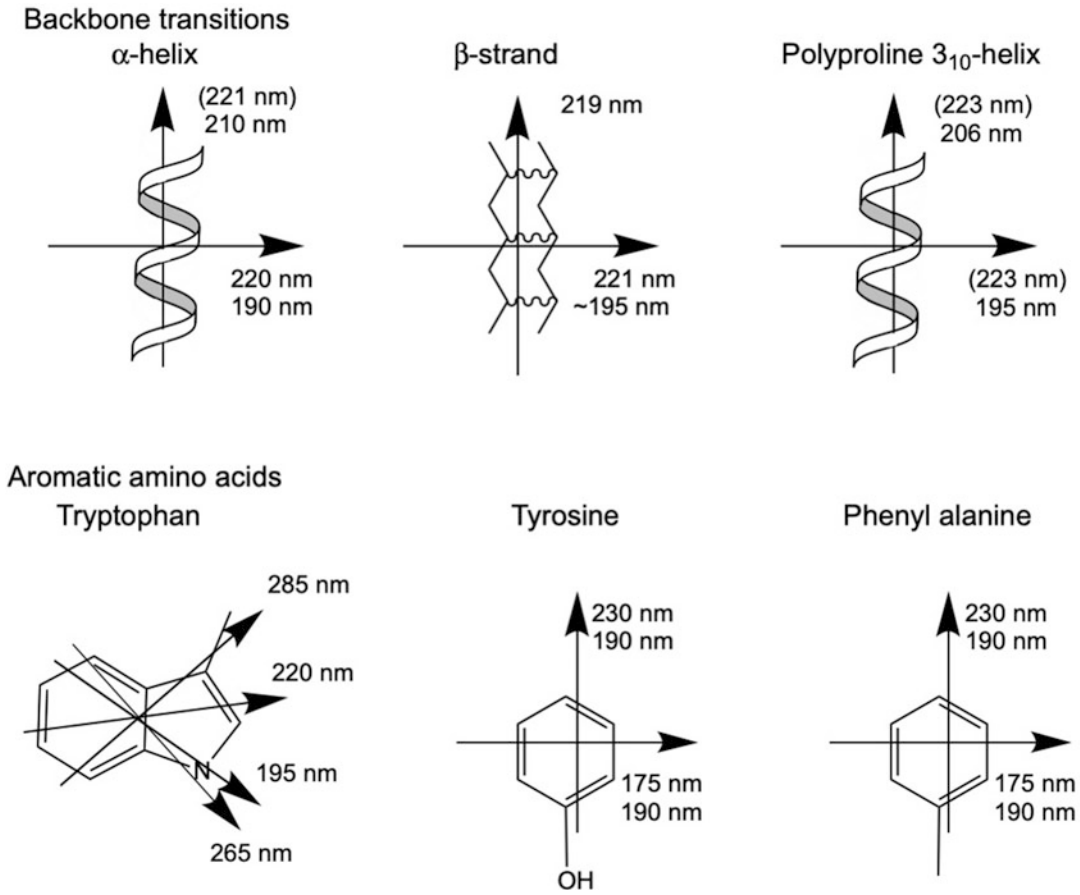


Fig. 3 Transitions polarizations of relevance to the UV spectroscopy of membrane peptides and proteins. Weak transitions are indicated by parentheses [7, 13, 14]

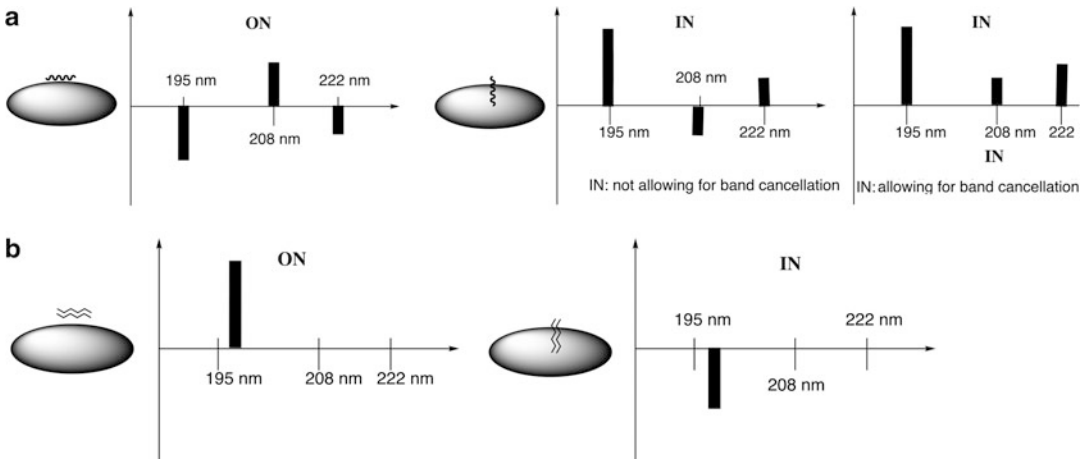


Fig. 4 Schematic of expected *LD* signals for an (**a**) α-helix or (**b**) β-strand lying flat on the surface or inserted into a membrane

peptide is on the membrane surface, in the head group region, or buried. When a ligand or other peptide/protein binds to a membrane protein, additional *LD* signals should be apparent. Liposome flow *LD* experiments can be used to follow the kinetics of membrane binding if it happens slowly enough. Normal Couette flow cells have a deadtime of about 30 s. Rapid injection [15] and channel flow systems [16] can be used to reduce the dead time to 500 μs and 25 μs, respectively, though with the penalty of more sample and worse signal-to-noise relative to Couette flow.

One of the challenges of liposome *LD* spectroscopy is that overlaid on any liposome *LD* spectrum, one collects a scattering contribution as the unilamellar lipid vesicles typically used in *LD* experiments are of the order of the wavelength of light. Since a spectrometer detector simply counts the photons that are incident onto it, photons that do not arrive due to scattering are not differentiated from those missing due to absorbance. The spectrum in Fig. 2b shows the original liposome flow *LD* spectrum of diphenyl hexatriene (DPH) and after correction for scattering.

Overall flow *LD* is the ideal technique for determining the orientation of different chromophores with respect to the membrane normal. Some examples have been given above. To illustrate the method, further examples of antimicrobial peptides, aurein 2.5 and gramicidin are considered below as well as another common chromophore, fluorene, often used to increase the hydrophobicity and hence membrane binding of peptides. How *LD* can be used both for geometry, structure analysis and probing kinetic processes is considered. Kinetic analysis usually involves identifying binding (appearance of an *LD* signal), insertion (sign change), often followed by loss of signal if the inserted protein or peptide disrupts the membrane [17]. One of the great advantages of *LD* is that only oriented molecules give a signal, so peptides that are not membrane bound are invisible. Increased sensitivity and selectivity is sometimes possible by undertaking fluorescence detected *LD* [12, 18].

2 Materials

All solutions should be prepared using ultrapure water (18 MΩ-cm at 25 °C) and analytical grade reagents where available. Follow all waste disposal regulations when disposing of materials. The choice of lipid for any experiment is complex. In our experience, the short lipid DMPC (Fig. 5) forms the least stable liposomes and POPC or naturally occurring mixtures are easier to use.

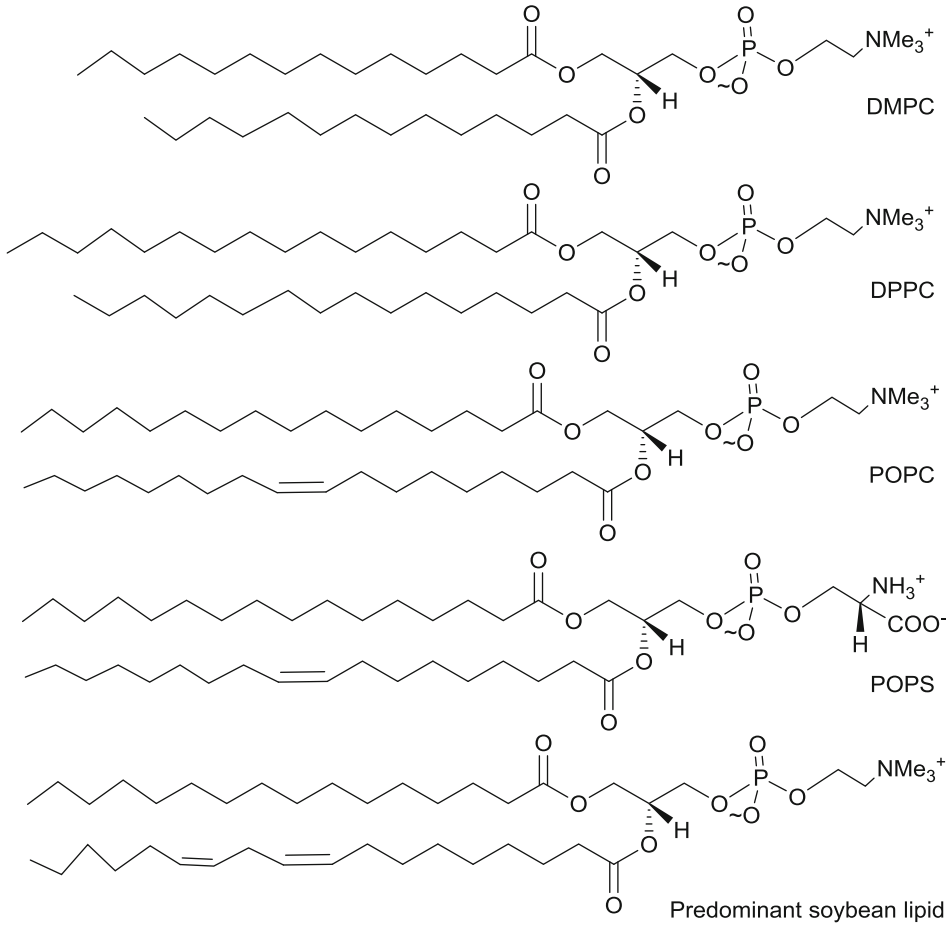


Fig. 5 Examples of lipids (available from Avanti (Alabaster, AL, USA)) that can form bilayers. DMPC, DPPC, POPC, and POPS denote, respectively: 1,2-dimyristoyl-sn-glycero-3-phosphocholine, 1,2-dipalmitoyl-sn-glycero-3-phosphocholine, 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine, and 1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-L-serine

3 Methods

3.1 Small Unilamellar Liposome Preparation

1. An *x* mg/mL lipid small unilamellar vesicle (SUV) stock solution is prepared by dissolving *x* mg of lipid in chloroform (or other solvent of choice). Dry the lipid onto the walls of a round-bottom flask (*see* **Notes 1** and **2**). The analyte(s) may be incorporated at this stage as powder or concentrated solution in a volatile solvent or added later (**step 5**).
2. The film is then resuspended by adding 1 mL of the required buffer (including salts, *see* **Note 3**) to give a turbid suspension at the desired stock concentration.
3. Five cycles of freeze/thawing are then implemented immediately after one another. A convenient freezing method is to use

dry ice/ethanol. Thawing can be done at room temperature or higher (if it is desirable to be above a lipid phase-transition temperature with, e.g., DPPC).

4. Transfer 1 mL of the suspension into the gastight syringes of a LiposoFast extruder assemblage fitted with a 100 nm (or other size) pore polycarbonate membrane. Push the lipid suspension through the membrane 11–15 times on the day of use (*see* **Note 4**).
5. To perform an *LD* experiment, if the analyte was not incorporated at **step 1**, add the desired amount of peptide and any probe molecule to the sample usually in a concentrated solution (in a water miscible solvent) (*see* **Notes 5** and **6**). Typically, one works with at least ~20:1 lipid:peptide ratios.

3.2 *Flow* LD *Data Collection*

1. Liposome *LD* experiments require an instrument that produces alternating beams of parallel and perpendicular light. Most *CD* instruments now come with *LD* as a standard option. "Parallel" is usually horizontal, which is what is needed for a Couette flow cell of the kind illustrated in Fig. 6 (*see* **Note 7**).
2. Carefully load the liposome sample into the cell (e.g., Fig. 6) ensuring the liquid level is well above the observation window as lipidic solutions are prone to forming bubbles, especially on the surface.
3. Choose a wavelength range for data collection. Due to the frequent need to correct for light scattering, data collection

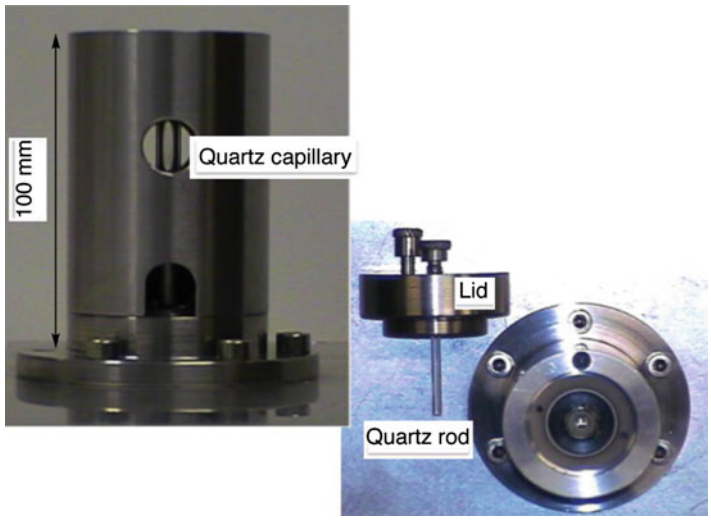


Fig. 6 Microvolume (25–60 μL) outer rotating [8, 19] Couette flow cell showing the outer quartz capillary (3 mm inner diameter) and inner quartz rod (2.5 mm outer diameter). When assembled, the annular gap is 250 μm so the path length is 500 μm

should begin 50–100 nm higher in wavelength than the first sample absorbance signal.

4. A starting set of parameters is 100 nm/min, 1 s response time, and 1 nm bandwidth or equivalent (*see* **Note 8**).
5. Collect a series of *LD* spectra with the Couette cell stationary, rotating at about 1000 revolutions per minute, stationary, and rotating. If the instrument allows it, collect simultaneous *LD* and absorbance spectra. Compare the repeat spectra. Any differences are indicative of a kinetic process such as insertion of the peptide into the membrane or liposome degradation (Fig. 7).
6. If the Couette cell has the same *LD* signal at all orientations (i.e., the quartz is of high quality, *see* **Note 9**), then the simplest baseline is the nonrotating spectrum from **step 5**. Alternative baselines are a rotating Couette cell filled with water or filled with a liposome solution without the analyte.
7. It is often necessary to perform kinetic experiments to see the time dependence of an insertion process and its consequences (Fig. 7). As any scattering signal will also change with time, it is usually necessary to measure full wavelength scans, not single point measurements, as a function of time.

3.3 *Liposome* LD Analysis

3.3.1 Data Processing

1. Begin your data analysis by overlaying plots of your flow *LD* sample spectrum with all the baseline options you have collected. In an ideal world, they all overlay outside the absorbance region (thus have the same scattering signal). Choose the best sample/baseline combination and subtract the baseline (*see* **Note 10**).
2. It may be useful to apply a scattering correction as outlined in [4] and illustrated in Fig. 2.

3.3.2 Data Interpretation

1. Consider where there are obvious bands in the spectrum and try to assign them to chromophores of the system. Usually the lipids can be ignored (*see* **Note 11**). α-helices have three accessible transitions that need to be considered and β-sheets have two with transition polarizations as illustrated in Fig. 3. Peptide flow *LD* spectra are often dominated by side chain signals even in the backbone region (below 240 nm) as aromatic groups are typically present in higher percentages in membrane-binding peptides than in a standard protein. Any chromophores such as the hydrophobic FMOC (fluorenylmethyloxycarbonyl chloride, Fig. 8) added to a peptide to change its properties may also have significant signals (*see* **Note 12**).
2. When the side chain and other chromophore contributions have been identified, the peptide contributions should be apparent. Figure 9 illustrates the *LD* spectra of aurein 2.5

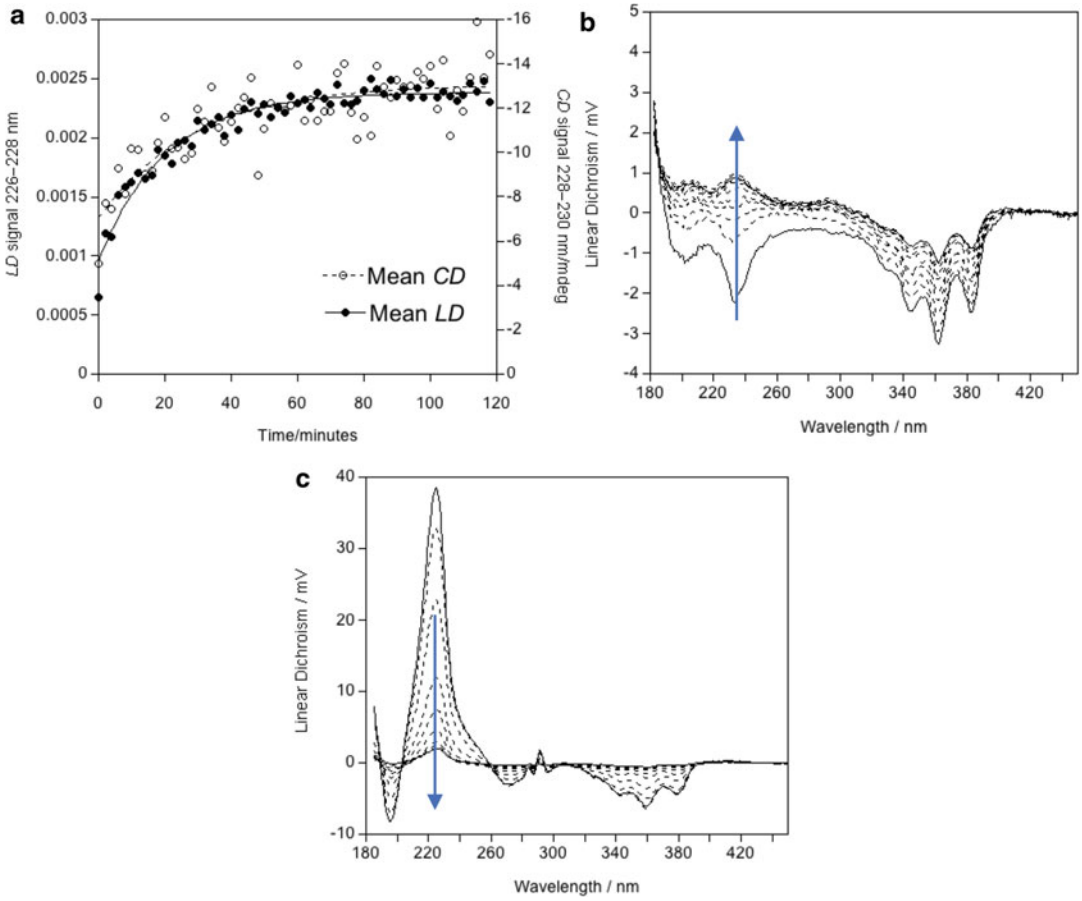


Fig. 7 Spectra of 1 mg/mL lipid, 0.01 mg/mL probe molecule (probe absorbance is from 310 to 350 nm), and 0.05 or 0.1 mg/mL gramicidin in 10% (v/v) TFE, added at time zero, showing that insertion (increase in *LD* signal) and folding (increase in *CD* signal) happen simultaneously. (**a**) *CD* and *LD* spectra averaged over the wavelengths indicated with soybean liposomes [20]. (**b**) Kinetics of insertion of gramicidin D into liposomes made of DPPC liposomes at 30 °C showing reorientation as a function of time (blue arrow indicates increasing time). (**c**) Kinetics of insertion of gramicidin D into DPPC liposomes at 50 °C showing loss of *LD* as a function of time (blue arrow indicates increasing time) indicating loss of liposome structure [21]. Data collected on a Jasco J-815 spectropolarimeter adapted for *LD* spectroscopy with a 500 μm path length at 12-min intervals. Intensity decreases with time

whose *CD* spectrum and sequence are given in Fig. 1. Its only aromatic group is one phenyl alanine of fairly weak absorbance, so the *LD* signal can be attributed to the peptide backbone (*see* **Note 13**).

3. Then consider each peptide transition. It is usually useful to consider *CD* data simultaneously. From the *LD* of Fig. 9, we deduce that at 50 °C (the fluid phase) aurein is α-helical and stably inserted into the membrane parallel to the lipids, but at

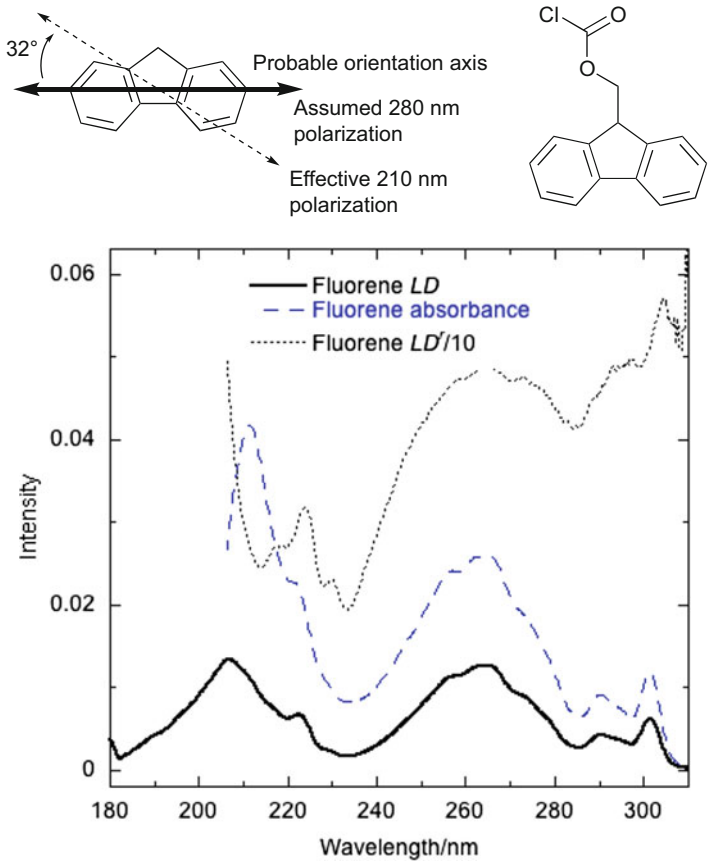


Fig. 8 Probable fluorene chromophore transition polarizations [22], Fmoc (N-(fluorenyl)-9-methoxycarbonyl) structure, and the fluorene chromophore absorbance and film *LD*. LD^r is the ratio of *LD* to Absorbance. Data collected on a Jasco J-815 spectropolarimeter adapted for *LD* spectroscopy

30 °C (the gel phase) it is mainly unfolded and surface bound until it destroys the liposomes (*see* **Note 14**).

4. Often the system is being studied because of some element of membrane bioactivity which occurs as a function of time. In such a case (e.g., Fig. 7), one considers the overall change as a function of time and may also analyze individual wavelength traces.

4 Notes

1. The drying process usually involves drying the solution under reduced pressure using a rotary evaporator and then under high vacuum for at least 2 h.

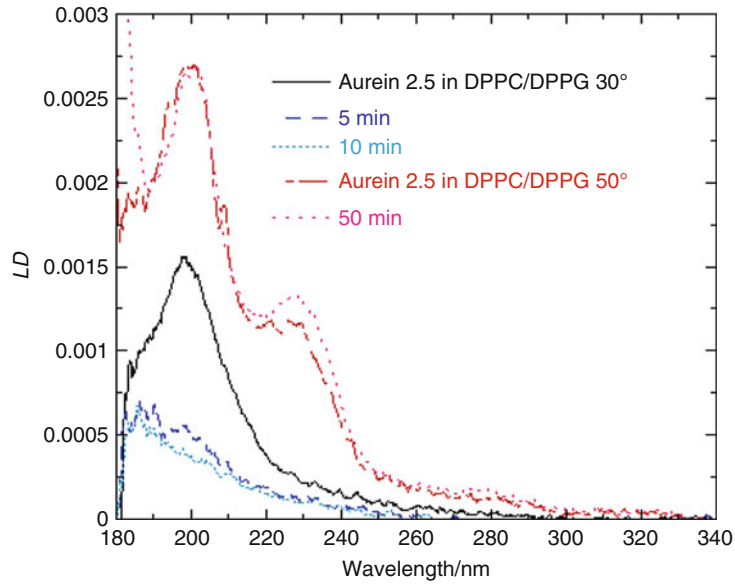


Fig. 9 *LD* spectra corresponding to the *CD* spectra of Fig. 1 at time 0 and later times as indicated. Red and pink spectra are at 50°. Black, blue, and turquoise spectra are at 30°. Data collected on a Jasco J-815 spectropolarimeter adapted for *LD* spectroscopy

2. It is sometimes convenient to add a peptide or small molecule analyte to the starting chloroform solution as this facilitates its incorporation in the membrane.
3. It should be noted that Cl^- absorbs light below about 215 nm, so concentrations higher than 10 mM should be avoided if possible. Although Cl^- and other ions do not have an intrinsic *LD* signal, they can mask the *LD* signal if they absorb too much. Methods to remove Cl^- include dialysis, filtration with appropriate membranes and columns. F^- can often be used to substitute for the ionic strength effect of Cl^-.
4. Dynamic light scattering is a good method to check your liposome size. It generally proves to be slightly higher than the pore size.
5. Approximately 70–80 μL sample volume is usually sufficient in a micro-volume capillary *LD* cell. It is possible to reduce the volume, but ensure the light beam is not incident on the meniscus of the sample as this introduces light scattering and other artifacts.
6. Note that different lipids and analytes have different insertion kinetics (taking from <1 s to overnight to insert). Also some peptides destroy the liposomes (a mode of action of some antimicrobial peptides). So consideration should be given to when to collect spectroscopic data.

7. Wada invented the first Couette flow cell in 1964 [23, 24]. The sample is "endlessly" and stably flowed between two quartz cylinders one of which rotates and one of which is stationary, which enables extended data collection. Couette cells designed to use quartz capillaries and rods and an annular gap of 250 μm reduce the sample volume to less than tens of microliters (Fig. 4) [8].
8. Peptide/liposome samples have broad bands, so increasing the light throughput by choosing 2 nm bandwidth seldom changes the spectral shape. It is usually necessary to average over a number of spectra. As a rule of thumb, the true instrument high tension voltage of the spectropolarimeter should usually be kept below 600 V—otherwise too many of the incident photons are being absorbed or scattered and the *LD* reading will be nonsense. It should be noted that stray light is often introduced in flow *LD* experiments when the power cable is inserted. The photomultiplier tube will happily count these photons but it is not indicative of satisfactory instrument performance because what we need to measure is light passing from the monochromator to the detector. Make sure you check for any instrument error flags.
9. If the stationary baseline of water is always the same (relative to the signal magnitude) when the capillary is reinserted in a different position, then one can assume the baseline is independent of orientation of the capillary.
10. In an ideal world, all three baselines will be identical. However, water, stationary liposomes and liposomes in shear flow seldom have the same scattering. The liposome options usually cancel the scattering contribution to the signal better than the water, though it should be noted that liposomes in stationary cells or without the analyte may be different in size or shape from the rotating sample. Often the only way to get a good baseline correction is to repeat the experiment multiple times and use the baseline overlay outside the absorbance region to choose the best pair of sample and baseline spectra.
11. Although the absorbance of the lipids themselves can usually be ignored, lipids often have light absorbing impurities. In addition, the scattering increases with decreasing wavelength.
12. If literature data are not available, the absorbance spectrum of the chromophores may need to be collected independently. Film *LD* or a calculation may be needed to provide transition polarization assignments. As illustrated in Fig. 8, FMOC has significant approximately long-axis polarized intensity at 260 nm and its 210 nm transition is closer to short-axis polarized. For film orientation we may write:

$$LD^r = \frac{LD}{A_{iso}} = \frac{3S}{2}\left(3\cos^2\alpha - 1\right) \quad (4)$$

13. If $\alpha = 0$ for the 260 nm transition, then the 220 region (which is the overlay of a number of transitions) averages to about 50° from the long axis.
14. A small scattering signal is apparent in the long wavelength part of the spectrum where there is no other absorbance. The spectrum has been zeroed at 340 nm but it should be zeroed at infinity.
15. From the *CD*, we know the 50° C peptide is helical. The 222 nm transition is perpendicular to the helix and has a positive *LD* signal, whereas at 210 nm, parallel to the helix, there is an apparent minimum, which is in fact a negative signal overlaid by neighboring larger positive ones as schematically illustrated in Fig. 8. We therefore conclude that at 50° C, aurein 2.5 is inserted into the membrane parallel to the lipids. After 50 min, a small increase in the orientation is noted. By way of contrast, at 30° C there is little evidence of the 222 nm transition being oriented. The 200 nm region is strongly positive which suggests the *LD* is dominated by unfolded peptides lying on the surface. The random coil polarizations are a less ordered version of the β-strand of Fig. 5 with the 220 nm region having two cancelling transitions and 200 nm region being along the carbonyl bonds [7, 13, 25]. Ten minutes after adding the peptide, the liposomes have been disrupted, the *LD* signal is reduced and represents just a small amount of scattering.

Acknowledgments

Funding from the Engineering and Physical Sciences and Biotechnology and Biological Sciences Research Council is gratefully acknowledged.

References

1. Nordén B (1978) Applications of linear dichroism spectroscopy. Appl Spectrosc Rev 14:157–248. https://doi.org/10.1080/05704927808060393
2. Rodger A, Nordén B (1997) Circular dichroism and linear dichroism. Oxford University Press, Oxford
3. Nordén B, Rodger A, Dafforn TR (2010) Linear dichroism and circular dichroism: a textbook on polarized spectroscopy. Royal Society of Chemistry, Cambridge
4. Dorrington G, Chmel NP, Norton SR, Wemyss AM, Lloyd K, Praveen Amarasinghe D, Rodger AJBR (2018) Light scattering corrections to linear dichroism spectroscopy for liposomes in shear flow using calcein fluorescence and modified Rayleigh-Gans-Debye-Mie scattering. Biophys Rev 10

(5):1385–1399. https://doi.org/10.1007/s12551-018-0458-8

5. Johansson LB-Å, Davidsson Å, Lindblom G, Nordén B (1978) Linear dichroism as a tool for studying molecular orientation in membrane systems. II. Order parameters of guest molecules from LD and NMR. J Phys Chem 82:260–269
6. Matsuoka Y, Nordén B (1983) Linear dichroism studies of flavins in stretched poly(vinyl alcohol) films. Molecular orientation and electronic transition moment directions. J Phys Chem 87:220–225. https://doi.org/10.1021/j100225a009
7. Rodger A, Rajendra J, Marrington R, Ardhammar M, Nordén B, Hirst JD, Gilbert ATB, Dafforn TR, Halsall DJ, Woolhead CA, Robinson C, Pinheiro TJ, Kazlauskaite J, Seymour M, Perez N, Hannon MJ (2002) Flow oriented linear dichroism to probe protein orientation in membrane environments. Phys Chem Chem Phys 4:4051–4057. https://doi.org/10.1039/b205080n
8. Marrington R, Dafforn TR, Halsall DJ, Rodger A (2004) Micro volume Couette flow sample orientation for absorbance and fluorescence linear dichroism. Biophys J 87:2002–2012. https://doi.org/10.1529/biophysj.103.035022
9. Rajendra J, Damianoglou A, Hicks M, Booth P, Rodger P, Rodger A (2006) Quantitation of protein orientation in flow-oriented unilamellar liposomes by linear dichroism. Chem Phys 326:210–220. https://doi.org/10.1016/j.chemphys.2006.02.036
10. Ardhammar M, Mikati N, Nordén B (1998) Chromophore orientation in liposome membranes probed with flow linear dichroism. J Am Chem Soc 120:9957–9958
11. Razmkhah K, Gibson MI, Chmel NP, Rodger A (2014) Oxidized polyethylene films for orienting polar molecules for linear dichroism spectroscopy. Analyst 139:1372–1382
12. Wemyss AM, Chmel NP, Lobo DP, Sutherland JA, Dafforn TR, Rodger A (2018) Fluorescence detected linear dichroism spectroscopy: a selective and sensitive probe for fluorophores in flow-oriented systems. Chirality 30(3):227–237. https://doi.org/10.1002/chir.22795
13. Bulheller BM, Rodger A, Hicks MR, Dafforn TR, Serpell LC, Marshall K, Bromley EHC, King PJS, Channon KJ, Woolfson DN, Hirst JD (2009) Linear dichroism of some prototypical proteins. J Am Chem Soc 131:13305–13314
14. Albinsson B, Kubista M, Thulstrup E, Nordén B (1989) Near-ultraviolet electronic transitions of the tryptophan chromophore: linear dichroism, fluorescence anisotropy, and magnetic circular dichroism spectra of some indole derivatives. J Phys Chem 93:6646–6654
15. Hicks MR, Rodger A, Lin Y, Jones NC, Hoffmann SV, Dafforn TR (2012) Rapid injection linear dichroism for studying the kinetics of biological processes. Anal Chem 84:6561–6566
16. Cheng X, Joseph MB, Covington JA, Dafforn TR, Hicks MR, Rodger A (2012) Continuous-channel flow linear dichroism. Anal Methods 4(10):3169–3173. https://doi.org/10.1039/C2AY25513H
17. Ennaceur SM, Hicks MR, Pridmore CJ, Dafforn TR, Rodger A, Sanderson JM (2009) Peptide adsorption to lipid bilayers: slow rearrangement processes revealed by linear dichroism spectroscopy. Biophys J 96(4):1399–1407
18. Wemyss AM, Razmkhah K, Chmel NP, Rodger A (2018) Fluorescence detected linear dichroism of small molecules oriented on polyethylene film. Analyst 143(23):5805–5811. https://doi.org/10.1039/C8AN01588K
19. Marrington R, Dafforn TR, Halsall DJ, Hicks M, Rodger A (2005) Validation of new microvolume Couette flow linear dichroism cells. Analyst 130:1608–1616. https://doi.org/10.1039/b506149k
20. Hicks M, Damianoglou A, Rodger A, Dafforn T (2008) Folding and membrane insertion of the pore-forming peptide gramicidin occur as a concerted process. J Mol Biol 383(2):358–366
21. Hicks M, Dafforn T, Damianoglou A, Wormell P, Rodger A, Hoffman S (2009) Synchrotron radiation linear dichroism spectroscopy of the antibiotic peptide gramicidin in lipid membranes. Analyst 134:1623–1628
22. Zou Y, Razmkhah K, Chmel NP, Hamley IW, Rodger A (2013) Spectroscopic signatures of an Fmoc-tetrapeptide, Fmoc and fluorene. RSC Adv 3:10854–10858. https://doi.org/10.1039/C3RA41979G
23. Wada A (1964) Chain regularity and flow dichroism of deoxyribonucleic acids in solution. Biopolymers 2:361–380
24. Wada A (1972) Dichroic spectra of biopolymers oriented by flow. Appl Spectrosc Rev 6:1–30
25. Bulheller B, Rodger A, Hirst JD (2007) Circular and linear dichroism of proteins. Phys Chem Chem Phys 9:2020–2035

Chapter 22

Probing Protein–Membrane Interactions and Dynamics Using Hydrogen–Deuterium Exchange Mass Spectrometry (HDX-MS)

Jordan T. B. Stariha, Reece M. Hoffmann, David J. Hamelin, and John E. Burke

Abstract

Cellular membranes are a central hub for initiation and execution of many signaling processes. Integral to these processes being accomplished appropriately is the highly controlled recruitment and assembly of proteins at membrane surfaces. The study of the molecular mechanisms that mediate protein–membrane interactions can be facilitated by utilizing hydrogen–deuterium exchange mass spectrometry (HDX-MS). HDX-MS is a robust analytical technique that allows for the measurement of the exchange rate of backbone amide hydrogens with solvent to make inferences about protein structure and conformation. This chapter discusses the use of HDX-MS as a tool to study the conformational changes that occur within peripheral membrane proteins upon association with membrane. Particular reference will be made to the analysis of the protein kinase Akt and its activation upon binding phosphatidylinositol (3,4,5) tris-phosphate (PIP_3)-containing membranes to illustrate specific methodological principles.

Key words HDX-MS, Hydrogen–deuterium exchange, Structural proteomics, Mass spectrometry, Protein–membrane interactions, Lipid signaling, Protein dynamics

1 Introduction

Hydrogen–deuterium exchange mass spectrometry (HDX-MS) is a biophysical technique that probes conformational protein dynamics through the measurement of the exchange rate of amide hydrogens with deuterated solvent. The basis of the technique relies on the fact that amide hydrogens participate in hydrogen bond networks to form secondary structural elements such as alpha helices and beta sheets. The exchange rate of amide hydrogens is exquisitely sensitive to the presence and dynamics of these secondary structure elements. The differences in secondary structure stability can be used as a probe to examine differences upon binding to ligand, protein, and membrane-binding partners. This can also be used to

Tina Daviter et al. (eds.), *Protein-Ligand Interactions: Methods and Applications*, Methods in Molecular Biology, vol. 2263, https://doi.org/10.1007/978-1-0716-1197-5_22,

examine changes in protein conformation in disease-linked mutations. Importantly, this provides a tool to explore both direct binding interactions and long range allosteric conformational changes induced by a variety of binding partners. The more dynamic a secondary structure element, the faster the exchange rate. Exchange rates within proteins can vary over a large timescale, with disordered regions exchanging within milliseconds to seconds, and extremely stable secondary structure regions exchanging on the order of hours to days [1]. The underlying theoretical basis for how and why amide hydrogens are protected from exchange have been extensively reviewed, and readers are advised to consult the following reviews for more detail [1–10].

The general flowpath for carrying out an HDX-MS experiment begins by incubating the protein of interest with deuterated buffer for a set period of time. During this time, the mass of the protein will increase as hydrogens are exchanged for deuterium throughout the protein. A complete HDX-MS experiment requires the analysis of multiple different time points of deuterium incorporation in order to obtain a comprehensive understanding of protein dynamics over time. The exchange reaction is terminated by addition of a low-temperature acidic quench buffer which dramatically slows the rate of exchange. From there, the protein is digested into peptides using an acid functional protease such as pepsin to enable spatial localization of deuterium exchange. Following digestion, an MS-based mapping approach is utilized which involves separation of peptides by reverse-phase chromatography and measurement of peptide masses using a mass spectrometer. Finally, the deuterium incorporation of each peptide is quantified by examining the shift in the average mass.

HDX-MS has become a powerful tool for defining protein–protein [11, 12], protein–ligand [13], and protein–membrane interactions [14], as well as mapping the effects of posttranslational modifications [15]. A variety of stimuli can be examined using HDX-MS including binding partners, small molecules, ligands, and membrane surfaces. Experiments carried out using membrane stimuli have already provided substantial insight into our understanding of fundamental signaling processes that occur at membranes. Numerous studies to date have implemented the use of HDX-MS to characterize protein–membrane interactions for both integral membrane proteins [16–22] and peripheral membrane proteins [8, 11, 12, 14, 15, 23–34]. HDX-MS provides a method to both probe the conformational changes that occur upon membrane binding and defining membrane binding interfaces.

This chapter will describe the HDX-MS methodology used to study the protein kinase Akt and its interaction with lipid membranes [15]. A brief introduction to Akt will be required to help put the HDX-MS results in context. Akt is composed of two domains: a pleckstrin homology (PH) domain that recognizes and binds to

lipid phosphoinositides and a kinase domain. When the PH domain is not bound to phosphoinositides, Akt is soluble and unphosphorylated at two key phosphorylation sites critical for its activity, T308 and S473. This renders soluble Akt in an autoinhibited, inactive state where the PH domain is engaged in a direct interaction with the kinase domain. When Akt is recruited to membranes containing PIP_3 or $PI3,4P_2$, the PH domain inhibitory interface is disrupted, and it can be phosphorylated at T308 and S473, leading to full activation. The HDX-MS experiments described below investigated the differences in H–D exchange for Akt in the presence and absence of 5% PIP_3 membranes to identify conformational changes that mediate the transition between the autoinhibited and activated form.

2 Materials

2.1 Instrumentation Buffers

Important note: All buffers used on the mass spectrometer (MS) should be *detergent-free*. Use new glassware for making and storing all buffers that are used for the upkeep of the MS and sample preparation to avoid residual detergent. The easiest way to do this is buy ultra-grade reagents, and use in the bottles they are sold in.

1. *Digestion buffer*: 0.05% LC/MS grade trifluoroacetic acid (TFA; pH 2.5). Sterile filter 2 L of deionized water into a detergent-free bottle using a 0.22 μm filter. Add 1 mL 100% LC/MS TFA into the 2 L of filtered deionized water to give a final concentration of 0.05% TFA. Pour off a small volume of the final buffer into a separate container and measure the pH to confirm it is at the desired value of 2.5.
2. *Needle wash 1*: 0.05% LC/MS grade trifluoroacetic acid (TFA; pH 2.5). Use the same protocol as the Digestion Buffer detailed above for making this buffer.
3. *Needle wash 2*: 10:90 HPLC grade acetonitrile (ACN): H_2O + LC/MS grade 0.1% formic acid (FA). Combine 100 mL of ACN with 900 mL of 0.1% FA and filter the mixture through a 0.22 μm filter unit.
4. *Sample loop/pepsin wash*: 0.1% LC/MS grade formic acid (FA) + 1 M guanidine. Measure out 4.78 g of guanidine hydrochloride into a new sterile 50 mL falcon. Add LC/MS grade 0.1% FA up to the 50 mL mark, dissolve the guanidine by vortexing and/or sonication, and filter the mixture through a 0.22 μm filter. Store at −20 °C.
5. *Mobile phase A*: 0.1% LC/MS grade formic acid (FA).
6. *Mobile phase B*: 100% LC/MS grade acetonitrile (ACN).

2.2 HDX-MS Setup Buffers

1. *Quench buffer*: 2 M ≥ 99% purity guanidine +3% LC/MS grade formic acid (FA). Measure out 9.553 g of ≥99% purity guanidine hydrochloride into a detergent-free vessel and bring the total volume up to 32.3 mL. Add 1 mL of 100% FA to bring the total volume to 33.3 mL (final FA concentration of 3%). Keep buffer on ice at all times during the experiment as quenching must be completed at a low temperature (~1 °C) to minimize back exchange.
2. *D_2O buffer*: 10 mM ≥99.5% purity HEPES (pH 7.5), 100 mM ≥99% purity NaCl, 98% D_2O. This buffer needs to be made fresh for each experiment to prevent exchange with H_2O present in ambient air. For making 1 mL of D_2O buffer, combine 20 μL of 5 M NaCl with 20 μL of 0.5 M HEPES (pH 7.5) in a 1.5 mL microcentrifuge tube prior to opening the D_2O. Crack open an ampule of D_2O and use a glass pipette to rapidly transfer the D_2O into the 1.5 mL microcentrifuge tube containing the NaCl and HEPES. Close the tube immediately after addition of D_2O and open only as needed to reduce the amount of contact with H_2O from the environment. Mix by inverting the tube several times. The buffer should be equilibrated to the temperature planned in the H/D exchange experiment.
3. *H_2O buffer*: 10 mM ≥99.5% purity HEPES (pH 7.5), 100 mM ≥99% purity NaCl. For making 1 mL of H_2O buffer, combine 20 μL of 5 M NaCl with 20 μL of 0.5 M HEPES (pH 7.5) in a 1.5 mL microcentrifuge tube. Add 1 mL of deionized water to the tube and mix well. Keep the buffer in the same conditions as the D_2O buffer. As with the D_2O buffer, remake for each experiment.
4. *Lipid buffer*: ≥99% purity 100 mM KCl, ≥99.5% purity 20 mM HEPES (pH 7.5). Combine 1 mL of 5 M KCl and 2 mL of 0.5 M HEPES (pH 7.5). Bring up to a final volume of 50 mL with deionized water and filter through a 0.22 μm filter.
5. *Protein buffer*: The protein buffer used in the exchange reaction will be the same buffer used in the final purification stage for the protein (i.e., the final storage buffer). For the best results, save and store some of the buffer used during purification. It is important to make sure the buffer does not contain any detergent. If multiple proteins are being assessed, and their buffers are different, both buffers will need to be used.

2.3 Description of Mass Spectrometer and Fluidics Setup

The processing of proteins and/or membranes of interest for HDX analysis is carried out using an integrative fluidics system consisting of the following three instruments: an HDx-3 PAL liquid handling robot and climate-controlled chromatography system (LEAP Technologies), a Dionex UltiMate 3000 UHPLC system, and an Impact HD QTOF mass spectrometer (Bruker). These components act in unison to digest and desalt the deuterated proteins of interest, separate the resulting peptides, and deliver the peptides to the mass spectrometer for high-resolution mass analysis.

The HDx-3 PAL liquid handling robot is capable of fully automating the setup of HDX-MS samples; however, we primarily set these samples manually. This provides two major advantages: First, we can setup extremely short pulses of deuterium (<10 s) that is not possible with a robotic approach, and second, we can setup all HDX samples within 1–2 h, limiting any possible problems with protein instability over the many hours to days it would have to sit in the liquid handling robot. For any samples performed with the robot, it is essential that there is testing to verify the long-term stability of the protein of interest and that it can handle storage at 2 °C for the entire length of the HDX experiment without aggregation/degradation.

The major advantage of the HDx-3 PAL system is that it can automate all injection, digestion, desalting, separation, and washing steps. This can also be performed with a manual valve system, which is utilized to direct sample constituents throughout the immobilized flowpath. A diagram of a representative fluidics flowpath, is shown in Fig. 1.

Protein is injected onto the system, and digested over two immobilized pepsin columns (Poroszyme™ Immobilized Pepsin Cartridge, 2.1 mm × 30 mm; Thermo Fisher) held at 10 °C and 2 °C, respectively, at a flow of 200 μL/min for 3 minutes (*see* **Note 1** *for tips on optimal digestion*). The resulting peptides are collected and desalted on a C18 trap column [Acquity UPLC® BEH™ C18 1.7 μm column (2.1 × 5 mm); Waters]. Once digestion and collection/desalting are complete, the C18 trap is put in line with the analytical column. The UltiMate 3000 UHPLC controls the flow and gradient of Mobile Phase A + B over the C18 trap and analytical columns, eluting peptides onto the mass spectrometer, with separation occurring at 40 μL/min ~10,000 psi. Nonlinear-shaped acetonitrile gradients are used to generate elution profiles with an equal spread of peptides over time. The short gradient can lead to a large number of peptides eluting per unit time, complicating the tandem MS/MS analysis required for peptide identification of non-deuterated samples. This is particularly true for proteins >250 kDa. For this reason, any project studying proteins larger than 250 kDa uses two different gradients for MS/MS analysis, including a ~20-min short gradient which is the same as the one used for MS analysis of deuterated samples and a ~60-min long

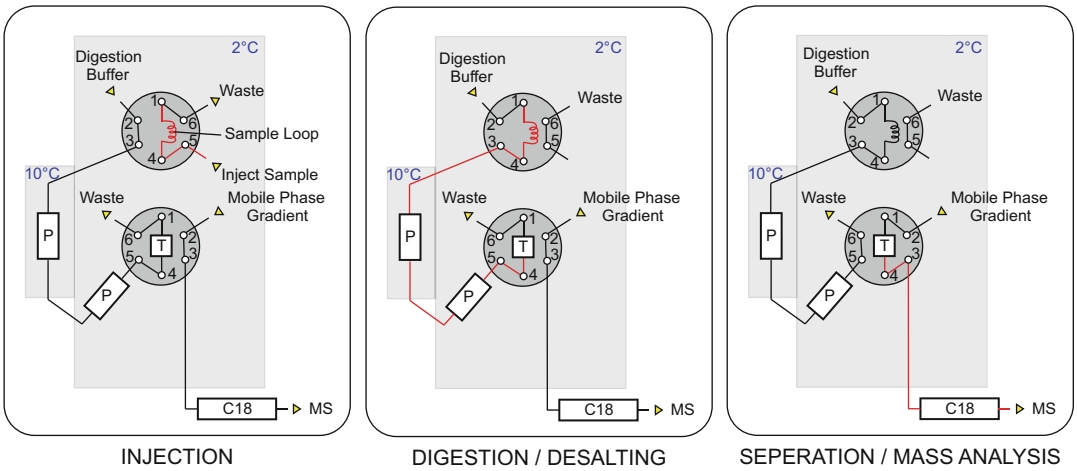


Fig. 1 Representative flowpath for an immobilized HDX-MS experiment. The flowpath of the protein sample is highlighted in red. The valves, reverse-phase columns (T is the 5 mm trapping column, C18 is 100 mm analytical column), sample loop, and one pepsin column (abbreviated as P) are maintained at 10 °C, with a second pepsin column maintained at 2 °C. Sample is initially injected on the system through the injection port into a sample loop (left panel), with this then switched in line with the immobilized pepsin columns, with the digested peptides collected and desalted on a C18 trap column (middle panel). The C18 trap is then switched in line with a mobile phase gradient that will elute and separate peptides according to their hydrophobicity over an analytical column, with the eluted peptides directed at the mass spectrometer (right panel)

gradient. All samples are followed by a ~8 min-blank injection to prevent sample carryover that would hinder data analysis. The different gradients used are shown in Fig. 2a.

3 Methods

3.1 Setting Up the HDX

The study of protein–membrane interactions involves subjecting the protein(s) of interest to various conditions and allowing continuous deuterium labeling to take place over various time intervals. There are many aspects that must be optimized prior to setting up the HDX. One must consider the protein amount required for optimal protein signal; membrane composition/lipid presentation; protein–protein/protein–membrane ratios required for optimal occupancy; deuteration conditions; and quenching conditions (*see* Subheading 4 for more details).

3.1.1 Experiment Design

1. Prior to carrying out any HDX-MS experiment, one must determine the goal(s) of the experiment and design conditions accordingly. In the study of Akt, the major goal was determining the conformational changes that occur upon binding PIP_3-containing membranes.

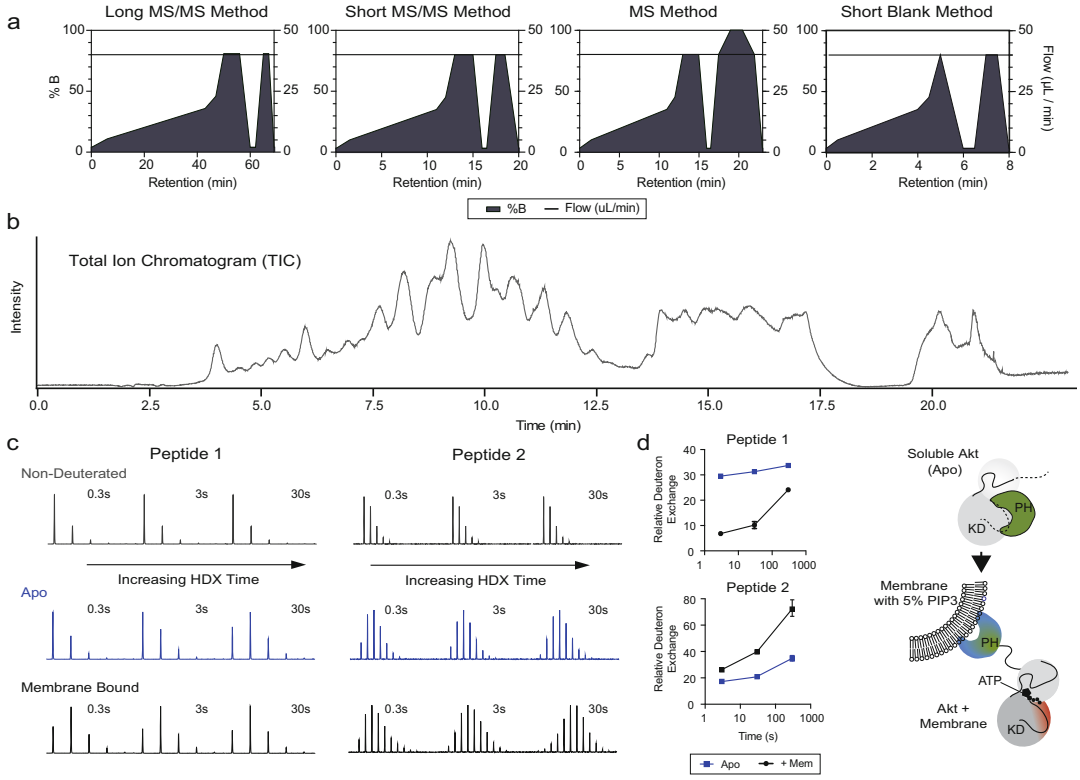


Fig. 2 Layout of HDX-MS experiment. (**a**) Optimized LC methods for MS/MS (Long and Short), MS, and Short Blank runs. The gradients of Mobile Phase A and Mobile Phase B shown have been designed for effective elution and detection of peptides in limited time. Typically the Short MS/MS method (second from left) is run to identify peptides generated during proteolysis; however, for larger proteins or proteins that have incomplete coverage, the Long MS/MS method (far left) can be run which may increase coverage by allowing more time for the spectrometer to select peptides for identification. (**b**) Example UHPLC peptide elution profile for Akt apo using the LC-MS method. Almost all peptides have finished eluting by ~14 min, with the remaining signal being primarily carryover and lipid constituents. (**c**) Example mass spectra data for two Akt peptides showing differences in deuterium incorporation in the presence and absence of PIP_3 vesicles. Peptide 1 is from a region in the PH domain that becomes more protected from exchange in the presence of membrane than in the absence, indicating a putative membrane binding site. Peptide 2 is from a region in the kinase domain that is protected from solvent by the PH domain in the autoinhibited apo state. When the PH domain binds to PIP_3 in the membrane, this region becomes exposed to solvent which increases the exchange rate in that area. (**d**) Deuterium incorporation curves for the two Akt peptides in the presence and absence of membrane. A common concern in HDX data is the shape of the resulting HDX incorporation curves, and either convergence or divergence in differences. The reason for different shapes is dependent on the intrinsic exchange rate of the amides being protected/exposed. The curve for peptide 1 is representative of a region that is highly dynamic in the apo state, with the fast exchanging amides becoming protected upon membrane binding. This leads to the data converging at late time points. The curve for peptide 2 is an example of extremely stable amides in the Apo state that become more dynamic upon membrane binding. This leads to the data diverging at late time points. The regions that these peptides are located in are colored accordingly in the cartoon to the right, where red indicates increased exchange and blue indicates decreased exchange

2. Determine the reaction volume and the final concentration of the protein(s) and lipid(s) that will be used. The concentrations must be carefully optimized to achieve the desired amount of protein in each sample to maximize sensitivity, while at the same time limiting carryover, as well as maximize occupancy and percent D_2O present in the exchange reaction (*see* **Notes 2** and **3**). Experiments on Akt were carried out at a final protein amount of 20 pmol per sample, in a final volume of 50 μL. The final protein concentration of Akt was 400 nM, and the final concentration of lipid vesicles (20% cholesterol, 30% PC, 15% PS, 35% PE, and 5% PIP_3) was 400 μM. Lipid vesicles were generated by extrusion through 0.1 μm filters in lipid buffer.
3. Calculate the final concentration of D_2O in the exchange reaction. In the Akt experiment, the final concentration of D_2O was 78%.
4. Determine the quenching conditions that mediate digestion and denaturation (*see* **Note 4**). The quenching conditions used in the Akt experiment were 20 μL of ice cold 2 M guanidine HCl, 3% formic acid, giving a final concentration of 0.6 M GdHCl and 0.8% formic acid.
5. Determine which D_2O exposure time points you want to assess. Common time points used are 3 s, 30 s, 300 s, and 3000 s exposures, which enable probing a range of amide hydrogen exchange rates. If this range is not sufficient to fully sample the spectrum of amide exchange rates, different temperatures and pH values can be used [35].

3.1.2 Execution of HDX

1. Thaw the appropriate amount of protein and lipids for entire experiment. Once thawed, keep protein on ice and lipids at 20 °C.
2. Mix protein well and spin down in a cooled tabletop centrifuge (4 °C) at 15,000 rpm for 5 min to remove any aggregate. Make sure to check the protein concentrations prior to any experimentation as the protein may not be at the same concentration that it was originally stored at. Adjust calculations as necessary to ensure final protein concentrations are correct.
3. Label and organize 0.6 mL Eppendorf tubes for all samples (indicate condition and time point on label). It is ideal to have a triplicate set of all exchange reactions. Make sure to include tubes for at least two non-deuterated apo samples. One will be used to identify the peptides generated during proteolysis (*see* Subheading 3.3.1); the other will be used as a non-deuterated reference that the analytic software will compare the deuterated samples to.
4. Combine both protein and lipid components in a final volume of 10 μL and incubate at 20 °C for 2 min. For the apo samples

(i.e., samples with single protein and no lipid), add the same volume of lipid buffer as you would lipid to the lipid-containing samples to ensure all samples are in the same buffer (*see* **Note 5**).

5. Add 40 μL D_2O buffer to start the exchange reaction.
6. Incubate the reaction at 20 °C for the correct amount of time (either 3, 30, 300, or 3000 s). For the 3000 s time point, the sample is purged with nitrogen and sealed with parafilm to minimize any problems with oxidation or D_2O exchange with H_2O in ambient air.
7. Add 20 μL quench solution to terminate the exchange reaction. To achieve the 3 s exposures, have the quench volume already drawn up in a pipette. Add the D_2O, mix a specific number of times, immediately add quench using the second pipette, mix a specific number of times, and flash freeze in liquid N_2.
8. Immediately flash freeze the sample tube in liquid N_2 following addition of quench.
9. Make non-deuterated samples in the same fashion but instead of adding D_2O buffer, add H_2O buffer. "Quench" immediately (no exchange reaction occurs in H_2O buffer; therefore no incubation time is required) and flash freeze tube in liquid N_2.
10. Store all samples at −80 °C. Flash frozen samples can be stored for up to a month at this temperature without affecting the amide exchange.

3.2 Operating the Mass Spectrometer

The success of an HDX experiment is highly dependent on taking proper and regular care of the instruments. This includes the following:

3.2.1 Precautions Necessary When Starting a Project

1. *Buffer levels.* All buffers (Mobile Phase, Needle Wash, and Digestion Buffer) used for the LC and RTC PAL fluidics systems must be monitored at all times to avoid running the system dry during an experiment, thus risking the loss of data as well as equipment damage. Buffers should be changed at least every 3 months.
2. *LC purging.* Air bubbles can accumulate within the LC pump system if it is not used for more than a few days, therefore risking pumping air onto the trap and analytical columns. To avoid doing so, the LC fluidics system must be purged, which involves passing Mobile Phases A and B through the pump system into waste at a high flow rate for a period of at least 10 min. This can be done using the liquid chromatography control software, in this case "Chromeleon."
3. *Mass spectrometer calibration.* To ensure maximum accuracy, the mass spectrometer must be calibrated at least twice a week

as well as prior to MS/MS experiments. This is done by directly injecting a solution containing a set of standards of known *m/z* ratio which act as reference points. For the Impact HD QTOF (Bruker™) used in our analysis, the calibration should result in a standard deviation of 0.20 ppm or less compared to reference standards, with this parameter dependent on the instrument used. The instrument should be switched from standby to operate for at least 30 min prior to calibration. For our instrument, this process is controlled using the mass spec operating software called "Otofcontrol."

3.2.2 Starting Up the MS-LC System Prior to Running Samples

1. *Initiating the LC system.* Using Chromeleon, manually set the flow rate to 40 μL/min. The system should initially be in 100% Mobile Phase B, as the C18 analytical column is ideally stored in acetonitrile. From there, ramp the system into 3% Mobile Phase B over the course of 10 min. If the system has not been used in several days, purging should be performed prior to initialization (*see* Subheading 3.2.1, **step 2**).
2. *Blank Methods.* At the start of each day, two blank MS methods (one ~20-min run, followed by a ~8-min run, *see* Fig. 2a) are run to minimize carryover from the previous day, as peptides may accumulate within the immobilized pepsin and analytical columns (*see* **Notes 6** and **7**). The pre-optimized MS-LC methods are loaded in the integrated MS control software, which is Bruker instruments program called "HyStar." The methods loaded in Hystar control the various parameters (MS parameters, data collection) via Otofcontrol and the LC system (gradient specifications) via Chromeleon. A separate method must be loaded in "Chronos," which controls the RTC PAL. The latter controls the valves within the PAL, thus determining the flowpath of the sample through the pepsin column, trap vanguard pre-column, and analytical column (*see* Fig. 1). Prior to initiating the runs, the sample loop is washed with 100 μL 0.1% formic acid. Calibration is performed following the startup methods.

3.2.3 Sample Running

When running samples, the 100 μL Hampton glass syringe used for sample loading and the 0.1% FA used for washing the sample loop are kept on ice so as to minimize back exchange of deuterated samples.

1. Similar to the startup process, the LC (shown in Fig. 2a) and MS methods (*see* **Note 8**) are loaded in both Hystar and Chronos. The MS run will elute the majority of peptides from the sample; the short blank run further elutes residual peptides from the system to reduce carryover between samples (*see* **Notes 6** and **7**). The sample files should be named correctly and saved in the proper folders.

2. *It is important to note that consistency is key when running samples.* Injection of the sample should be done so in a manner that minimizes back exchange:
3. Take the sample out of the −80 °C freezer and immediately place it on ice.
4. Wash the sample loop by injecting 100 μL of 0.1% formic acid.
5. Quickly thaw the sample by rubbing the tube with hands or in a warm water bath, key is consistency between samples.
6. When a small amount of ice remains in the tube, mix the sample with the syringe to finalize the thawing process.
7. Quickly but carefully draw up and dispense the sample into the sample loop. Avoid creating bubbles to minimize putting air on the system.
8. Initiate the run immediately after injection.
9. During the MS method, the total ion current should be monitored throughout all samples, with a similar shape being seen for all samples (for example, TIC trace for Akt shown in Fig. 2b.) Any discrepancies are usually due to problems with carryover (*see* **Notes 6** and **7**) or problems with the fluidics system.

3.3 Data Analysis

Two types of data analyses are to be completed at this point: an MS/MS analysis that will generate a list of peptides for the protein (s) of interest and an HDX data analysis that will measure the deuterium incorporation within the protein(s) across the various time points and conditions.

3.3.1 MS/MS Analysis Using PEAKS7

An MS/MS analysis, which allows for the identification of the peptides generated during proteolysis, must be performed at the beginning of every experiment. This process will confirm the identity of the relevant protein(s), as well as generate a list of peptides that will be tracked for deuterium incorporation throughout the experiment. The peptides produced during proteolysis are fragmented into b/y ions within the mass spectrometer via collision-induced dissociation (CID). The resulting datasets are analyzed using PEAKS7, which uses a data-dependent approach to identify peptides. The tolerances for analysis are a precursor mass ± 10 ppm and a fragment mass ± 0.05 Da. The target database consists of FASTA sequences for the protein(s) of interest, all proteins previously analyzed by the laboratory, as well as common contaminants (pepsin from the immobilized column, chaperones from organisms used for expression), with the decoy database composed of the scrambled sequences. Posttranslational modifications can be incorporated within the search as needed. We typically set the false discovery rate to 0.1%; however, this can be increased to 1% for datasets with minimal coverage. The goal is to attain at least 90%

sequence coverage. For Akt, the final peptide list covered ~95% of the protein (Fig. 3a). The final product of this analysis is a comma separated values (CSV) file, containing the sequence, charge state, ppm mass error, and retention time. The file is modified prior to uploading to HDExaminer to remove peptide hits from other proteins as well as peptides with large ppm errors (*see* **Note 9**).

3.3.2 Measuring Deuterium Incorporation with HDExaminer

HDExaminer (Sierra Analytics) is a commercial software package used for calculating deuterium incorporation that is compatible with multiple MS vendor formats. This software uses the mass centroid as well as envelope-shape information of the isotopic distributions corresponding to each peptide and displays deuterium incorporation values in terms of both % deuteration and number of deuterons (Da). It is essential to manually verify the deuterium incorporation of every peptide at every time point and replicate, to make sure the peak is accurately fit. It is very important to examine the shape of the deuterium incorporation distribution, as it provides deuterium occupancy information corresponding to specific residues within peptides, as well as details on the mechanism of deuterium incorporation (EX2 and EX1 kinetics; [36, 37]). This data is generally lost when only looking at the mass centroid. Readers are advised to consult the following reviews for more details on data analysis [38, 39].

3.3.3 Starting a Project on HDExaminer

Starting a new HDExaminer project requires three key components which are added to the software by dragging the files into the software in the following order:

1. A text file (.txt extension) containing the FASTA sequence for the protein of interest. When performing mutational analysis, the appropriate mutations can be applied to the sequence.
2. The modified peptide list of the protein of interest, generated by PEAKS7. This list is exported from PEAKS7 and modified to include only the peptides from the protein of interest that are within a certain ppm error (average ppm ± 5; *see* **Note 9**).
3. A non-deuterated MS data file for the protein of interest. This software searches the datafile for the presence of peptides identified in PEAKS7. At this step, the initial peptide list is curated to only include peptides that will be trackable throughout the dataset, with this decision based on peptide's signal/noise and level of interference from overlapping peptides (further details described below). The peptides remaining are used as an undeuterated reference point for the calculation of deuterium incorporation for all other samples.

3.3.4 Data Processing of Deuterated Samples

Drawing meaningful and accurate conclusions from HDX-MS experiments requires that the data be manually processed and curated. Data processing is performed in two stages:

1. *Processing data in HDExaminer.* Initial analysis of the non-deuterated file can speed up analysis during later steps by reducing the number of overlapping peptides in areas that are highly concentrated and filling in any essential gaps in the data by searching multiple charge states. The final peptide list for Akt is shown in Fig. 3a. Examples of the raw deuterium incorporation data for two selected peptides in Akt in the presence and absence of membrane are shown in Fig. 2c. Every peptide needs to be scrutinized on HDExaminer to look for the following:t
   (a) Correct charge state;
   (b) Correct retention time (provided by PEAKS7). The selected range can be refined to exclude contaminating peaks. Tightening the range will help remove coeluting peptides, but may also reduce the peptide signal.
   (c) Appropriate selection of isotopic distribution;
   (d) Carryover (*see* **Note 6**).
   (e) Isotopic distributions from co-eluting peptides of similar m/z that overlap/mix with the selected distribution, and may therefore skew the data.
   (f) If more than one protein is present in the sample, ensure that the selected isotopic distribution is from the appropriate protein and not its binding partner. This may occur if the binding partner is present in a much higher concentration.
2. *Further processing of data in excel.* Once the HDExaminer file has been scrutinized and curated, it can be exported as an excel document. This document allows for analysis of the raw data from HDExaminer in order to produce interpretable data. This stage of processing includes the following steps:
3. Generation of the average and standard deviation of deuterium incorporation for all peptides analyzed. Any outliers (high SD, or nonstandard deuterium incorporation curves (e.g., deuterium incorporation decreasing over time)) should be manually inspected again in HDExaminer. When this is finalized, the final data should include the sequence of every peptide, the charge state, the level of deuterium incorporation, and the standard deviation for the triplicate measurement. This data should be included as supplemental data for any publication to allow for detailed review and analysis.

4. Establishing the appropriate comparisons between conditions, as needed. This is done by computing the difference between both the % deuteration and # of deuterons for the conditions that are to be compared. In the Akt study, comparisons were made between the apo and membrane conditions.
5. Determining the thresholds for significance for differences in deuterium incorporation. Our standard threshold requires three specific criteria be met: >7% change in deuterium incorporation at any time point, >0.4 Da difference in deuterium incorporation, and an unpaired student *t*-test value of $P < 0.01\%$ (*see* **Note 10** for more details).

3.3.5 Data Analysis and Presentation of Deuterium Incorporation and Differences Between Conditions

Critical to the proper analysis of the deuterium exchange information is the generation of the appropriate visualization of the data. This processing includes:

1. *Analysis of the H/D incorporation heatmap*. Deuterium incorporation over the entire protein can be viewed as a heatmap, where the level of deuterium incorporation, as indicated by coloring according to the percentage of deuterium exchange, can be viewed at every time point for all conditions tested (Fig. 3b). This data can also be presented in what is termed as a butterfly plot (Fig. 3c), where the %D is graphed according to the sequence, with SD information included. Two conditions can be viewed on a butterfly plot at the same time, which may reveal differences in exchange. Very localized regions of exchange can be determined using overlapping fragments; however, careful interpretation of overlapping fragments is critical. Any differences observed in this analysis must be verified at the single peptide level (*see* below).
2. *Generation of deuterium incorporation timecourses*. The deuterium incorporation of every peptide analyzed should be graphed for all conditions measured, with error bars describing standard deviation included (Figs. 2d and 3f). These graphs are directly generated in HDExaminer and can also be generated in GraphPad Prism Software downstream of the excel analysis.
3. *Analysis of differences in deuterium incorporation*. The differences in deuterium incorporation (either %D or #D) summed across all time points can be visualized over peptide sequence (Fig. 3d). One of the easiest ways to interpret the molecular consequences of the differences in exchange is mapping changes in deuterium incorporation on a structural model of the protein of interest if it is available (Fig. 3e), with the magnitude of the difference in deuterium incorporation indicated by different colors. This can be done by mapping the differences at any time point on a single model, or mapping the difference at each time point on the model (i.e., four different

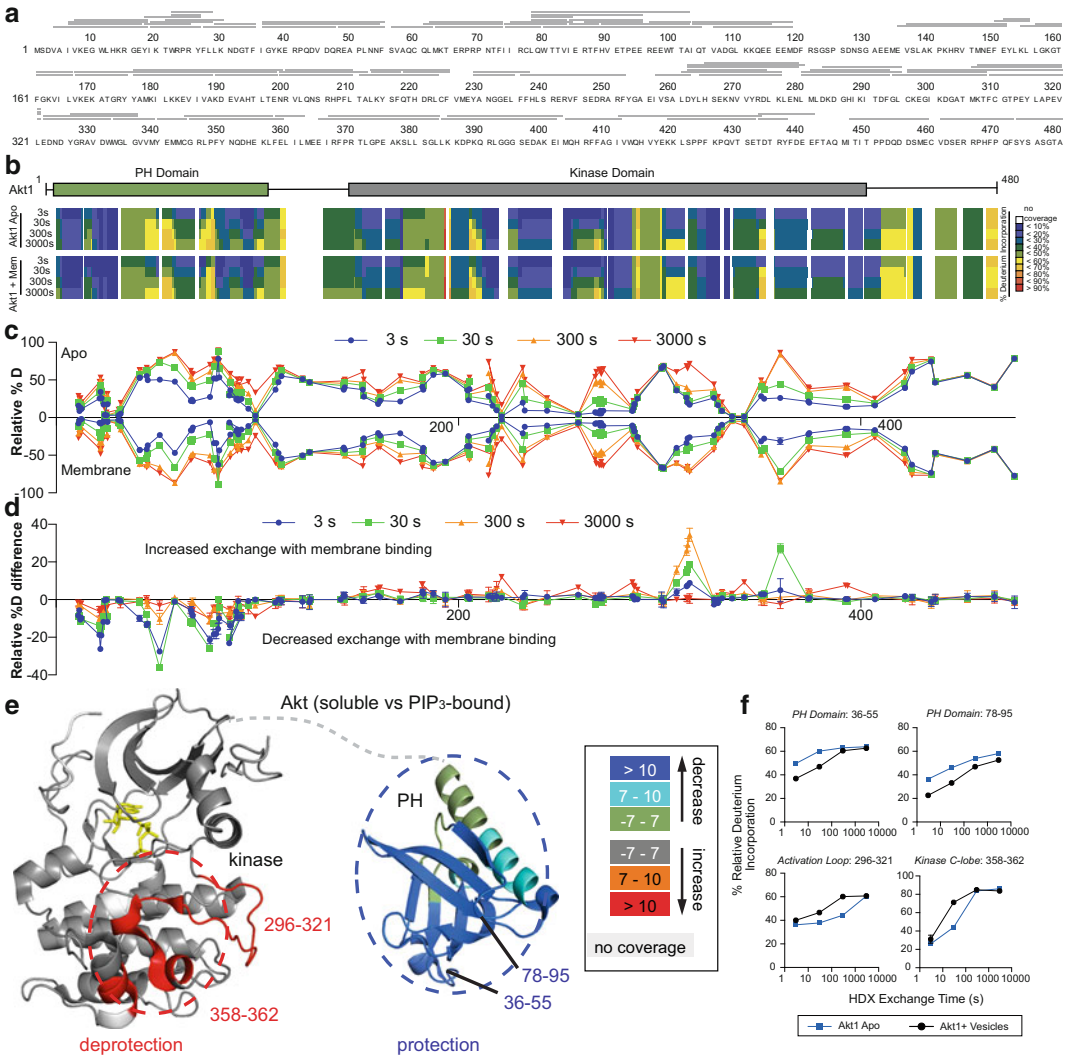


Fig. 3 Data analysis of HDX exchange for Akt binding to membrane vesicles. (**a**) Finalized peptide map for Akt. Each grey line indicates a peptide that was analyzed for its deuterium exchange over the entire timecourse of analysis. (**b**) Heat maps for Akt in the presence and absence of membrane aligned with the domain architecture of Akt. The level of deuterium incorporation is indicated by the legend to the right. (**c**) Butterfly plot for deuterium incorporation of Akt in the presence and absence of membrane. The colors indicated the different time points, with every peptide being represented by a single point graphed according to the central residue. (**d**) Differences in HDX between soluble Akt and membrane-bound Akt graphed, with the different colors representing different time points of exchange. (**e**) Differences in HDX between soluble Akt and membrane-bound Akt mapped on PDB 4EKK and 1UNQ [44]. Changes depicted represent peptides that had % deuteration and # of deuteron differences greater than 7% and 0.4 Da, respectively at any time point and a student *t*-test value of <0.01. The level of difference is indicated in the legend to the right. No significant difference is indicated in green for the PH domain, and grey for the kinase domain. (**f**) Selective deuterium incorporation curves for peptides with either increased or decreased exchange are highlighted. For panels **c**, **d**, and **f**, the errors bars represent SD, with many smaller than the size of the point

colored models for four distinct time points). If only a single model is used, it is highly recommended to include the deuterium incorporation timecourses (Fig. 3f), as this provides essential information on the dynamics of the deuterium incorporation difference.

4 Notes

1. Peptide digestion is achieved using an acid-functional protease such as pepsin, which is active at a pH range of 1–4. In the methods described here, pepsin is immobilized on a column so that it can be incorporated within the fluidics flowpath, enabling online digestion. To maximize the digestion efficiency, the fluidics system was optimized to include two immobilized pepsin columns stored at different temperatures, the first at 10 °C and the second at 2 °C. Storing the columns at different temperatures facilitates a balance between favoring proteolytic cleavage and minimizing back exchange. At 10 °C, proteolytic cleavage is more efficient; however this temperature is less effective at minimizing back exchange. Moreover, amide hydrogen back exchange is decreased at lower temperatures [42], a temperature less favorable for cleavage. There is a second function to having the two columns in series as it increases the back pressure, which has been shown to improve digestion efficiency [43]. In our system, the back pressure of the two pepsin columns should never exceed 2500 psi. Typically the pressure reaches its maximum during the digestion/desalting phase, when the pepsin columns are in line with the C18 trap column (Schematic 1). Sequence coverage can be maximized by optimizing the quench composition (*see* **Note 4**), increasing digestion time by decreasing the flow rate over pepsin, or increasing protein amount.
2. A key factor to consider for obtaining best results is optimizing the protein–membrane interaction conditions. Protein and lipid concentrations, lipid presentation, vesicle size, and membrane compositions are all factors that require careful optimization. Lipids can be presented in membrane mimics, bicelles, nanodiscs, and nanotubes. Membrane compositions and vesicle size must be optimized so as to mimic the cell membrane of interest (plasma membrane, golgi membrane, mitochondrial membrane, etc). These parameters can all be optimized using a variety of biophysical assays such as protein-lipid fluorescence resonance energy transfer (FRET) assays, lipid sedimentation, lipid flotation, surface plasmon resonance (SPR). Final concentrations of both proteins and vesicles should be carefully optimized in order to maximize interaction kinetics while minimizing problematic scenarios associated with high protein

and membrane concentrations (aggregation, instability, etc). Protein amounts should be be maintained within a range that yields adequate signal on the MS and good coverage (usually strive for coverage above 90%), while avoiding overloading the LC system. This range is generally found between 5 and 150 pmol per sample. However, the appropriate protein amount will be completely dependent on the sensitivity of the mass spectrometer being used, protease efficiency, size of the protein, and interaction kinetics. Membrane concentration must also be optimized as proteins may sometimes become unstable when in the presence of membranes at high concentration, with this being tested before setup of HDX samples. This issue can be surmounted by minimizing membrane concentration in the initial protein–membrane incubation, or by incorporating membrane within the D_2O buffer, thereby initiating membrane–protein incubation and deuteration simultaneously [3].

3. Care must be taken to not load excess lipids or detergents onto the LC-MS system during the HDX-MS experiment. Detergents and lipids may accumulate on the analytical or trap columns resulting in increased back pressure and decreased column performance. Lipids and detergents can also lead to significant ion suppression during MS analysis. One solution to this problem is to incorporate additional washing of the reverse phase system in organic solvent, sending the wash product to waste rather than injecting it onto the MS system. It is important to note that if this approach is taken, one must carefully monitor the performance of the LC system (i.e., pressure and retention times) to verify that there is no degradation over time [3].
4. The quenching conditions must be optimized prior to performing an HDX. The nature of HDX experiments requires that the hydrogen–deuterium exchange be dramatically decreased, or "quenched," so as to monitor exchange rates accurately. The quenching solution generally consists of a chaotropic agent, a reducing agent, and an acid. The chaotropic agent, such as guanidine hydrochloride, will denature the protein(s) post hydrogen–deuterium exchange, thus facilitating proteolysis by the acid functional protease. If incomplete digestion is obtained (<80% coverage in initial MS/MS analysis), the amount of denaturant should be optimized, with either increased denaturant amount or increased time of the quench buffer sitting with protein before freezing in liquid nitrogen. The reducing agent should only be incorporated within the quenching solution if the protein contains disulfide bonds. Disulfide bonds will prevent the protein from properly unfolding, therefore reducing protease access to the protein. The

most useful reducing agent is the acid functional compound Tris(2-carboxyethyl)phosphine (TCEP) [34, 40]. The exchange rate of amides with deuterium is an acid–base catalyzed process, with a global minimum at approx. pH 2.5 [41]; therefore, the quench buffer should be at this pH. The concentration of acid in the quench buffer will depend of the buffering capacity of the D_2O buffer, which generally contains Tris or HEPES and the protein buffer(s). That being said, one must check the pH of a mock sample containing all the components in the HDX reaction without protein and D_2O (i.e., a solution containing the exact volumes of quench solution, protein buffer(s), and H_2O buffer) *before* running any samples on the mass spectrometer as to avoid inactivating or ruining the pepsin columns (pepsin becomes inactive at pH >6.5 and fully denatures at pH 8.0). Additionally, the quenching solution is kept cold (~1 °C) throughout the HDX setup to minimize back exchange by dramatically slowing the kinetics of the exchange reaction.

5. Successful HDX experiments heavily depend on accuracy, specificity, and repeatability of sample preparation. Consistency in technique when mixing the D_2O or quench with the sample and when flash freezing the protein in liquid nitrogen will reduce variability in deuterium incorporation between replicates. Additionally, it is important to maintain constant temperature (experiments should be carried out in a climate controlled room) as well as pH throughout the setup and across the various conditions. All comparisons between conditions must use identical buffers. For example, if the protein of interest is in buffer A and its binding partner is in buffer B, all HDX samples should contain the same volume of buffer A and B regardless of whether the protein and/or its binding partner is present in that sample.
6. A consequence of having multiple columns incorporated in the fluidics system is an increased probability of carryover between samples. The amount of carryover will be dependent on the specific protein and the amount being run on the system. As previously mentioned, one of the steps taken to mitigate this issue is to start the system up with a ~20-min blank and a ~8-min short blank to elute peptides that carried over from the previous day. Peptides can also carryover between samples; therefore, to reduce this occurrence, a short ~8-min blank is run between samples. Additionally, every method includes a sample loop/pepsin washing step that occurs simultaneously while peptides elute from the trap. This step is automated using the LEAP PAL liquid handling robot which first injects Sample Loop/Pepsin Wash Buffer into the sample loop and then puts the sample loop in line with the pepsin columns. From there,

Digestion Buffer is run through the sample loop and over the pepsin columns at an increased flow rate of 400 μL/min. The washing method should be optimized to ensure that there is less than 5% carryover of peptides from the previous sample. To monitor the carryover, one should routinely check the Total Ion Current chromatogram (TIC; All MS) of the short ~8-min blank (*see* **Note 7**).

7. When a sample is inputted into Compass MS Data Analysis Software, the data will be presented as a Base Peak Chromatogram (BPC) which normalizes all ions detected to the most intense peak in the mass spectrum and plots them as a function of time. This tends to clean up the spectrum; however, when analyzing the short blank run for peptide carryover, the TIC trace, which displays all the ions detected and plots them as a function of time, is useful to determine if carryover is likely to be a problem (*see* Fig. 2b for an example).
8. Different MS systems will likely have slightly different optimal MS parameters to maximize sensitivity of detection of both precursors and fragments. For our Bruker Imapct HD system, we have found the following ESI and MS parameters to be optimal for sensitivity. Capillary voltage (4500 V), plate offset (500 V), drying N_2 gas (6.5 L/min), Nebulizer gas (0.8 Bar), Dry temp (200 °C). For both MS and MS/MS analysis, data was collected from an m/z range from 150 to 2000. For MS/MS analysis, data-dependent analysis was carried out using a 0.5 s precursor scan, followed by twelve 0.25 s fragment scans.
9. Once exported as a comma separated values (CSV) file, the peptide list must be modified to make sure only peptides from the protein of interest are analyzed. The list is filtered using the ppm error, with any peptide that is >5 ppm from the average standard error being filtered out of the data set. The file is now ready for use in HDExaminer; save the file as a modified peptide list with the .csv extension.
10. It is critical to appropriately determine the threshold for defining a significant change in deuterium incorporation. Using a difference in the # of deuterium incorporated will bias towards long peptides, with using a difference in the % of deuterium incorporated will bias towards short peptides. For this reason, we use a combined threshold of differences in both the number of deuterons incorporated and the percentage of deuterium incorporated. Critically the difference in incorporation must also cross a significance threshold of an unpaired student *t*-test.

Acknowledgments

This work was supported by a Canadian Institutes of Health Research New Investigator award and Open Operating Grant CRN-142393, Cancer Research Society Operating Grants CRS-22641 and CRS-24368, Natural Sciences and Engineering Research Council of Canada Discovery Grants NSERC-2014-05218 and NSERC-2020-04241, and Michael Smith Foundation for Health Research Scholar Award 17686. The authors declare no conflicts of interest with the contents of this chapter.

References

1. Konermann L, Pan J, Liu Y-H (2011) Hydrogen exchange mass spectrometry for studying protein structure and dynamics. Chem Soc Rev 40:1224–1234
2. Masson GR, Jenkins ML, Burke JE (2017) An overview of hydrogen deuterium exchange mass spectrometry (HDX-MS) in drug discovery. Expert Opin Drug Discovery 12:981–994
3. Vadas O, Jenkins ML, Dornan GL et al (2017) Using hydrogen-deuterium exchange mass spectrometry to examine protein-membrane interactions. Methods Enzymol 583:143–172
4. Deng B, Lento C, Wilson DJ (2016) Hydrogen deuterium exchange mass spectrometry in biopharmaceutical discovery and development—a review. Anal Chim Acta 940:8–20
5. Harrison RA, Engen JR (2016) Conformational insight into multi-protein signaling assemblies by hydrogen-deuterium exchange mass spectrometry. Curr Opin Struct Biol 41:187–193
6. Gallagher ES, Hudgens JW (2016) Mapping protein-ligand interactions with proteolytic fragmentation, hydrogen/deuterium exchange-mass spectrometry. Methods Enzymol 566:357–404
7. Guttman M, Lee KK (2016) Isotope labeling of biomolecules: structural analysis of viruses by HDX-MS. Methods Enzymol 566:405–426
8. Vadas O, Burke JE (2015) Probing the dynamic regulation of peripheral membrane proteins using hydrogen deuterium exchange-MS (HDX-MS). Biochem Soc Trans 43:773–786
9. Pirrone GF, Iacob RE, Engen JR (2014) Applications of hydrogen/deuterium exchange MS from 2012 to 2014. Anal Chem 87(1): 99–118
10. Balasubramaniam D, Komives EA (2013) Hydrogen-exchange mass spectrometry for the study of intrinsic disorder in proteins. Biochim Biophys Acta 1834:1202–1209
11. Rostislavleva K, Soler N, Ohashi Y et al (2015) Structure and flexibility of the endosomal Vps34 complex reveals the basis of its function on membranes. Science 350:aac7365
12. Jenkins ML, Margaria JP, Stariha JTB et al (2018) Structural determinants of Rab11 activation by the guanine nucleotide exchange factor SH3BP5. Nat Commun 9:3772
13. Kielkopf CS, Ghosh M, Anand GS et al (2018) HDX-MS reveals orthosteric and allosteric changes in apolipoprotein-D structural dynamics upon binding of progesterone. Protein Sci 282:31068–31374
14. Pulkoski-Gross MJ, Jenkins ML, Truman J-P et al (2018) An intrinsic lipid-binding interface controls sphingosine kinase 1 function. J Lipid Res 59:462–474
15. Lučić I, Rathinaswamy MK, Truebestein L et al (2018) Conformational sampling of membranes by Akt controls its activation and inactivation. Proc Natl Acad Sci U S A 115: E3940–E3949
16. Martens C, Shekhar M, Borysik AJ et al (2018) Direct protein-lipid interactions shape the conformational landscape of secondary transporters. Nat Commun 9:4151
17. Adhikary S, Deredge DJ, Nagarajan A et al (2017) Conformational dynamics of a neurotransmitter:sodium symporter in a lipid bilayer. Proc Natl Acad Sci U S A 114:E1786–E1795
18. Shukla AK, Westfield GH, Xiao K et al (2014) Visualization of arrestin recruitment by a G-protein-coupled receptor. Nature
19. Mehmood S, Domene C, Forest E et al (2012) Dynamics of a bacterial multidrug ABC transporter in the inward- and outward-facing conformations. Proc Natl Acad Sci U S A 109:10832–10836
20. West GM, Chien EYT, Katritch V et al (2011) Ligand-dependent perturbation of the conformational ensemble for the GPCR β2 adrenergic

receptor revealed by HDX. Structure 19:1424–1432
21. Chung KY, Rasmussen SGF, Liu T et al (2011) Conformational changes in the G protein Gs induced by the β2 adrenergic receptor. Nature 477:611–615
22. Zhang X, Chien EYT, Chalmers MJ et al (2010) Dynamics of the beta2-adrenergic G-protein coupled receptor revealed by hydrogen-deuterium exchange. Anal Chem 82:1100–1108
23. Mouchlis VD, Chen Y, McCammon JA et al (2018) Membrane allostery and unique hydrophobic sites promote enzyme substrate specificity. J Am Chem Soc 140:3285–3291
24. Siempelkamp BD, Rathinaswamy MK, Jenkins ML et al (2017) Molecular mechanism of activation of class IA phosphoinositide 3-kinases (PI3Ks) by membrane-localized HRas. J Biol Chem 292:12256–12266
25. Dornan GL, Siempelkamp BD, Jenkins ML et al (2017) Conformational disruption of PI3Kδ regulation by immunodeficiency mutations in PIK3CD and PIK3R1. Proc Natl Acad Sci USA 114:1982–1987
26. Masson GR, Perisic O, Burke JE et al (2016) The intrinsically disordered tails of PTEN and PTEN-L have distinct roles in regulating substrate specificity and membrane activity. Biochem J 473:135–144
27. Vadas O, Dbouk HA, Shymanets A et al (2013) Molecular determinants of PI3Kγ-mediated activation downstream of G-protein-coupled receptors (GPCRs). Proc Natl Acad Sci U S A 110:18862–18867
28. Cao J, Burke JE, Dennis EA (2013) Using hydrogen/deuterium exchange mass spectrometry to define the specific interactions of the phospholipase A2 superfamily with lipid substrates, inhibitors, and membranes. J Biol Chem 288:1806–1813
29. Dbouk HA, Vadas O, Shymanets A et al (2012) G protein-coupled receptor-mediated activation of p110β by Gβγ is required for cellular transformation and invasiveness. Sci Signal 5: ra89–ra89
30. Burke JE, Perisic O, Masson GR et al (2012) Oncogenic mutations mimic and enhance dynamic events in the natural activation of phosphoinositide 3-kinase p110α (PIK3CA). Proc Natl Acad Sci U S A 109:15259–15264
31. Burke JE, Vadas O, Berndt A et al (2011) Dynamics of the phosphoinositide 3-kinase p110δ interaction with p85α and membranes reveals aspects of regulation distinct from p110α. Structure 1993(19):1127–1137
32. Hsu Y-H, Burke JE, Li S et al (2009) Localizing the membrane binding region of Group VIA Ca2+−independent phospholipase A2 using peptide amide hydrogen/deuterium exchange mass spectrometry. J Biol Chem 284:23652–23661
33. Burke JE, Hsu Y-H, Deems RA et al (2008) A phospholipid substrate molecule residing in the membrane surface mediates opening of the lid region in group IVA cytosolic phospholipase A2. J Biochem 283:31227–31236
34. Burke JE, Karbarz MJ, Deems RA et al (2008) Interaction of group IA phospholipase A2 with metal ions and phospholipid vesicles probed with deuterium exchange mass spectrometry. Biochemistry 47:6451–6459
35. Hamuro Y (2017) Determination of equine cytochrome c backbone amide hydrogen/deuterium exchange rates by mass spectrometry using a wider time window and isotope envelope. J Am Soc Mass Spectrom 28:486–497
36. Guttman M, Weis DD, Engen JR et al (2013) Analysis of overlapped and noisy hydrogen/deuterium exchange mass spectra. J Am Soc Mass Spectrom 24:1906–1912
37. Weis DD, Wales TE, Engen JR et al (2006) Identification and characterization of EX1 kinetics in H/D exchange mass spectrometry by peak width analysis. J Am Soc Mass Spectrom 17:1498–1509
38. Engen JR, Wales TE (2015) Analytical aspects of hydrogen exchange mass spectrometry. Annu Rev Anal Chem 8:127–148
39. Wales TE, Eggertson MJ, Engen JR (2013) Considerations in the analysis of hydrogen exchange mass spectrometry data. Methods Mol Biol 1007:263–288
40. Yan X, Zhang H, Watson J et al (2002) Hydrogen/deuterium exchange and mass spectrometric analysis of a protein containing multiple disulfide bonds: Solution structure of recombinant macrophage colony stimulating factor-beta (rhM-CSFbeta). Protein Sci 11:2113–2124
41. Connelly GP, Bai Y, Jeng MF et al (1993) Isotope effects in peptide group hydrogen exchange. Proteins 17:87–92
42. Walters BT, Ricciuti A, Mayne L et al (2012) Minimizing back exchange in the hydrogen exchange-mass spectrometry experiment. J Am Soc Mass Spectrom 23:2132–2139
43. Ahn J, Jung MC, Wyndham K et al (2012) Pepsin immobilized on high-strength hybrid particles for continuous flow online digestion at 10,000 psi. Anal Chem 84:7256–7262
44. Lin K, Lin J, Wu W-I et al (2012) An ATP-site on-off switch that restricts phosphatase accessibility of Akt. Sci Signal 5:ra37–ra37

Index

Tina Daviter et al. (eds.), *Protein-Ligand Interactions: Methods and Applications*, Methods in Molecular Biology, vol. 2263, https://doi.org/10.1007/978-1-0716-1197-5,

D

E

F

G

H

I

K

L

M

N

O

P

Q

R

S

T

U

V

W

X

Z