HARVARD UNIVERSITY
Graduate School of Arts and Sciences

DISSERTATION ACCEPTANCE CERTIFICATE

The undersigned, appointed by the

Department of Molecular and Cellular Biology

have examined a dissertation entitled

"Optimizing Nanodiscs for Membrane Proteins and Mapping a Translation Initiation Complex"

presented by Diego Baptista,

candidate for the degree of Doctor of Philosophy and hereby

certify that it is worthy of acceptance.

Signature _____

Typed name: Prof. Rachelle Gaudet

Signature _____

Typed name: Prof. Victoria D'Souza

Signature _____

Typed name: Prof. Vlad Denic

Date: May 21, 2018

OPTIMIZING NANODISCS FOR MEMBRANE PROTEINS AND MAPPING A TRANSLATION INITIATION COMPLEX

A DISSERTATION PRESENTED

BY

DIEGO BAPTISTA

TO

THE DEPARTMENT OF MOLECULAR AND CELLULAR BIOLOGY

FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN THE SUBJECT OF

BIOCHEMISTRY

HARVARD UNIVERSITY

CAMBRIDGE, MASSACHUSETTS

MAY 2018

Acknowledgements

First and foremost, I have to express my gratitude to Gerhard. If it wasn't for the environment you fostered in your laboratory, I would have never been able to perform the work in this thesis. You are an exceptionally inspiring scientist, and you provided me with great support in critical times during my Ph.D. I feel very lucky to have worked for you on exciting projects. The achievements you have made in the world of NMR spectroscopy constantly humbles me, and the legacy you have left in the field of NMR is truly astonishing.

My dissertation advisory committee has been exceptional. Rachelle—I am so happy that you encouraged me to seek out work in Gerhard's lab. Victoria—I would not have ever considered applying to Harvard if it was not for your support and enthusiasm. Vlad—you give me so much hope for the future of science. Guido—I am very thankful for your vast knowledge of biochemical literature, and your enthusiasm for sharing it with me. Other members of the Department of Molecular and Cellular Biology have also had a large impact on my development as a scientist—in particular Andrew Murray, Ben di Bivort, Michael Lawrence, Patty Perez, and Fanuel Muindi.

None of this work would have been possible without the postdocs that have guided me during my Ph.D. Ute Hellmich—at the time in Rachelle's lab—was my first rotation advisor, and she taught me so much biochemistry in the ten weeks we worked together. I can't thank you enough for your time, patience, and support. Members of the Wagner lab were wonderful, especially when I was the only graduate student in the lab. Franz Hagn was my rotation mentor in the Wagner lab, and I learned most of the nanodisc

protocols from him. Josh Ziarek was the best bench-mate that one could dream of—not only was he organized and tidy in the lab, he was always happy to discuss theory, troubleshoot experiments, or teach me NMR. Additionally, Scott Robson, Hari Arthanari, and Jim Sun have been very helpful with giving me advice for how to prepare NMR samples and how to use the spectrometers. Mahmoud and I collaborated on the nanodisc work—it was great to constantly challenge him on the quality of his work. Wanna Chen and I collaborated on the PCS work—she was integral to the experimental design, and helping me collect data. I am excited to see where she takes the project. Other members of the lab that have been wonderful include Melissa Leger-Abraham, Paule Coote, Andras Boeszoermenyi, Simina Grigoriou, Sandeep Chabra, Meng Zhang, Reza Poumerad, and Sven Hyberts. Maura Kilcommons was excellent at navigating the bureaucratic side in the Department of Biological Chemistry and Molecular Pharmacology.

I have to acknowledge the Howard Hughes Medical Institute Gilliam Fellowship for Advanced Study as an excellent means of support during my Ph.D. Not only did it provide funding, but it also provided an exceptional network of diverse scientists, thinkers, and activists. The other fellows in the program are inspirational in their work and demeanor, and I am so happy to have made connections with such an inspiring group of people.

Many of my MCO peers have taught me about great science and served as sources of inspiration. In alphabetical order I have to tell these people how much I appreciate and love them: James Angstman, Fred Chang, Alicia Darnell, Erin Diel, Jason Gehrke, Gonzalo Gonzalez del Pino, Linda Honaker, Sandy Mattei, Tessa Montague, Alyson Ramirez, Matt Reilly, Clarissa Scholes-Rhines, Matt Smith, Georgia Squyres, and

Katie Weiner. You have all had an impact on me in a positive way that I will keep with me for the rest of my life. Special shout out to “bbs near and far”—you made this last year incredible and your support has picked me when my constitution is weak.

To my friends outside of the MCO program—thank you for keeping me sane. In particular, I must thank Jacques Carolan. Your enthusiasm for the world is something that I strive to have. The last few years of my doctorate would have been awful if you hadn’t popped up in my life. You’re the best, mate.

It has been fun to share the toils of graduate school with my family, and I am thankful that they have been open-minded and supportive especially for having no experience of dealing with the struggles and successes of trying to be an academic.

Last, but certainly not least, is a massive thank you to my boo, Gabriel Goulet-Langlois. You have been a great deal of support during my time as a graduate student. Your perspective on the world is incredibly refreshing, and your problem solving abilities are inspiring. You have always asked incredibly thoughtful questions and listened to me rant when things were sour or when things were excellent. I love you, dearly.

Advisor: Gerhard Wagner Diego Baptista

Abstract

Nanodiscs have been an excellent tool for encouraging biochemists and structural biologists to study membrane proteins. They are composed of a discoidal lipid bilayer and two copies of a membrane scaffold protein (MSP). It is possible to reconstitute an integral membrane protein into them—the enhanced stability and ability to provide a lipid environment allows for sophisticated biochemical and biophysical study. In the first part of this thesis, we present a method to improve nanodiscs by covalently circularizing MSP prior to nanodisc assembly. We highlight the enhanced stability of the particles, and show that they assemble more uniformly sized particles. The benefits of this are illustrated in applications for nuclear magnetic resonance (NMR) spectroscopy on voltage dependent anion channel-1 (VDAC-1).

In the second part of this thesis, we explore the complex formed between eukaryotic initiation factor 4A (eIF4A) and the HEAT2 domain of eukaryotic initiation factor 4G (eIF4G-HEAT2). Through chemical modification of the eIF4G-HEAT2 domain, we use paramagnetic lanthanides to monitor psuedocontact shifts (PCSs) in the NMR spectra for both sides of the complex. We present the shifts observed and map them to the complex. In the end, we illustrate that this method could be viable for determining a solution structure for a large, dynamic protein complex.

Contents

1

MEMBRANE PROTEIN STRUCTURAL BIOLOGY

1.1 THE IMPORTANCE OF MEMBRANE PROTEINS

Proteins embedded into plasma membranes are essential to the molecular mechanisms of life. These integral membrane proteins (IMP) respond to external stimuli and communicate those signals to internal machinery, carry electric currents, act as judicious gatekeepers for nutrients, energy sources, and waste, as well as keep close watch on foreign materials within and outside of the cell. They have exceptionally diverse functions and roles to aid in the successful operation of the cell. IMPs account for approximately 23% of the human genome and nearly half of the current pharmacological

targets [1, 2]. Understanding the functional details of IMPs to atomic detail is of utmost importance for understanding many fundamental biological processes, mechanistically identifying the basis of many diseases, and designing and developing new therapies for these diseases. The three main techniques in structural biology have long been challenged with structure determination of IMPs. This obstacle stems from a variety of problems—mainly that biochemical preparations of membrane proteins are challenging, and the special bilayer environment that they live in is difficult to recreate in vitro. The first part of this dissertation develops a method to make it easier to study IMPs with biochemical and structural methods; however, we must first discuss these methods in addition to developments in the membrane protein structure world.

X-ray crystallography has been a powerhouse for determining membrane protein structures since the first structure in 1985 [3]. Over 350 unique structures have been deposited into the Protein Data Bank since then, many more deposited since the development of lipidic cubic phase-based studies in 2012 [4]. More recently, cryoelectron microscopy (cryoEM) has experienced a revolution in a variety of parameters that have enhanced the resolution and sensitivity of data collection [5]. This is due to enhanced technology of detectors [6], the ability to record mini-videos of the samples during collection [7, 8], computation of the data [9], and more niche technology such as phase plates [10]. NMR spectroscopy has positioned itself as a powerful tool for studying the dynamics of protein structure during functional events—such as complex formation, ligand and drug-binding, or chemical environment variations. New developments in biochemical preparations of membrane proteins [11, 12], as well as improvements in

isotopic labeling, pulse programs, and NMR data analysis, have greatly enhanced the technique's ability to ask important questions in the world of membrane protein structure. On their own, each of these techniques can provide a wealth of structural information; however, in combination, these techniques can be exceptionally powerful and provide thorough illustrations of the molecular world. The push to improve biochemical preparations of membrane proteins for each technique has been incredibly beneficial for the field of IMP structure, and cross-pollination between researchers in each structural method has been integral to high-quality preparations of IMPs.

1.2 X-RAY CRYSTALLOGRAPHY

Crystallization of IMPs is a challenging hurdle to overcome and; unfortunately, every novel crystal structure must begin with a crystal. Coaxing a protein to crystallize seems like black magic—obtaining crystals of IMPs adds another level of complexity to this. Not only is crystallization a multiparametric process in which a large variety of biochemical, physical, and chemical parameters need to be explore, but promoting ensembles of delicate proteins and detergent molecules to stabilize during screening is a massive feat. However, this does not stop people from success and crystallography has been a powerhouse of furthering understanding of membrane protein biology. Significant effort has been put towards innovative development to overcome the numerous obstacles associated with X-ray structure determination of IMPs. This includes: overexpression of recombinant IMPs in different expression hosts [13], development of new detergents and lipids for more efficient solubilization and crystallization [14, 15, 16], improvement in

protein stability through mutations, deletions, engineering of fusion partners and monoclonal antibodies to promote diffraction quality of crystals [17], developments in automation, miniaturization, and integration which have contributed to the increasing number of initial crystallization conditions and crystal optimization strategies [18], and synchrotron radiation and beamline developments [19].

1.3 SINGLE PARTICLE CRYO-ELECTRON MICROSCOPY

For quite a long time, cryoEM was referred to as “blobology”—although useful for understanding larger scale resolutions of molecular events, it could not really access atomic scale resolution of biological samples. In the last several years, single particle cryo-EM has come forth as a viable technique for determining atomic resolution structures at better than 4-A resolution [5]. The two major contributors to this “resolution revolution” are advances in detectors [6] and advances in computational image analysis [7, 8].

Modern direct electron detectors have significantly higher detective quantum efficiencies (DQEs) than previous detectors. This means that the images recorded have better signal-to-noise performance over a larger spatial frequency range. They enable recording low-dose cryo-EM images of small particles with much smaller defocus values—and this provides a much better balance between the requirements for both image contrast and high-resolution signal. Additionally, the detectors have an exceptionally fast frame readout rate. This allows for the low total electron dose used in collecting the data to be separated in to several subframes—essentially it is possible to record short movies of the grids [7, 8]. Fractioning the data as a function of time over data

collection allows each of the subframes to be corrected for motion-induced image blurring before averaging them together [6]. This way of handling the data significantly improves data acquisition efficiency, and almost all of the images can be corrected to recover high-resolution information. Some researchers also tinker with the electron dose over time [20], and adding weights to frames that have less information as a result of radiation damage. Cryo-EM is certainly not the “blobology” that it was once considered, and there have been many recent high-resolution structures of IMPs that have been very challenging to crystallize. Although it shows great promise for becoming the technique that will revolution the future of structural biology, there is an exceptionally high barrier to entry into the field—equipment and maintenance requires significant financial investment. Even with automated data acquisition technology and streamlined data processing, image acquisition and processing is still too complex for a novice to learn with minimal training or by studying manuals. While the future of single-particle cryo-EM is bright, it requires strong support from the scientific community as well as massive investment from funding agencies to make single particle cryo-EM as popular as X-ray crystallography.

1.4 NUCLEAR MAGNETIC RESONANCE SPECTROSCOPY

NMR-derived structures can now be found in all major functional groups of membrane proteins: enzymes [21], receptors [22], regulators [23], channels [24] and transporters [25]. Solution NMR has benefited significantly from improvements both in spectroscopic technique with respect to higher magnetic fields, cryoprobe technology, and new pulse sequences. Protein biochemistry development in new isotope labeling

schemes and the creation of effective membrane mimetics have greatly enhanced the questions that can be asked about IMPs with NMR spectroscopy. The most powerful feature of NMR; however, is not the ability to get a structure, but it lies in the power of monitoring a structure during important functional events.

This very attractive feature of solution NMR allows one to probe protein structures in the state of dynamic equilibrium. For example, it is widely recognized that transporters can exist in multiple conformations such as closed, open, inactive, etc. These functional properties can be characterized by transitions between these conformations with NMR spectroscopy. Because NMR allows one to measure conformational dynamics with residue-specific details, it is a valuable tool for understanding IMP function from a dynamics point of view. Comprehending the dynamic nature of IMPs is critical for developing small and large molecule drugs, as their action is likely to depend on conformational specificity. Another important aspect of IMP dynamics is whether active conformations are available for structural interrogation. Despite high functional relevance, active states of many membrane proteins, such as GPCRs, are often weakly populated and therefore challenging to capture by crystallography or electron microscopy. NMR is sensitive enough to detect population levels of just a few percent [26]; however, preparing such challenging proteins for NMR spectroscopy is no small feat. Finally, ligand binding can change the dynamic landscape of a protein, and NMR is an excellent technique to probe subtle effects of ligands which do not necessarily lead to major conformational rearrangements, but can nevertheless affect the equilibrium and modulate protein function.

1.5 THE FUTURE OF IMP STRUCTURE DETERMINATION

The coming years are very exciting for the field of membrane protein structural biology. Technological advances in all of the major techniques used to define structure and understand protein function at the molecular level are laying the groundwork for better workflows, data collection, and data analysis. Researchers are specializing in more than one technique—which is the best way to solve a wide breadth of problems since each technique has advantages and limitations.

The single most important skill that anyone working in membrane structural biology—or really structural biology of any proteins—needs to be skilled at his protein biochemistry. None of the methods described earlier are possible if poor biochemistry is performed to obtain protein molecules. The first part of this thesis deals mainly with making biochemical preparations of membrane proteins better. A fundamental effort such as this will have major implications for all of the techniques used to study membrane proteins, and I am excited to detail these efforts here.

2

NANODISCS

2.1 WHAT IS A NANODISC?

In a perfect world, IMP biochemical and structural investigations should not only give us information about the relevant conformations of the protein, but also their topology with respect to the lipid bilayer and details of the interactions with the lipids—which can vary greatly depending on lipid composition. Some IMPs have exhibited the requirement for certain lipids to be present to be functional [27]. Additionally, a planar bilayer system in which one can control the size of the plane, the number of IMPs that sit in the plane,

and provide access to both sides of the IMP would be ideal. Nanodiscs—developed the Stephen Sligar in 2002 [28]—have proven to be an invaluable tool that can address some of these concerns, and have been useful for deciphering the structure and function of IMPs as well as their interactions with ligands, other proteins, and lipids.

Nanodiscs are discoidal lipid bilayers of tunable diameter on the nanometer scale. They are composed of a cylindrical patch of lipid bilayer that is stabilized by two amphipathic helical membrane scaffold proteins (MSPs)—the hydrophobic side of the helix interacts with the hydrocarbon tails of the bilayer, while the hydrophilic side of the helix interacts with the aqueous solution. Figure 2.1 illustrates a nanodisc that is filled with lipids. Nanodiscs were inspired from the biological process in the small intestine in which lipids and proteins are packaged and transported to the liver [29].

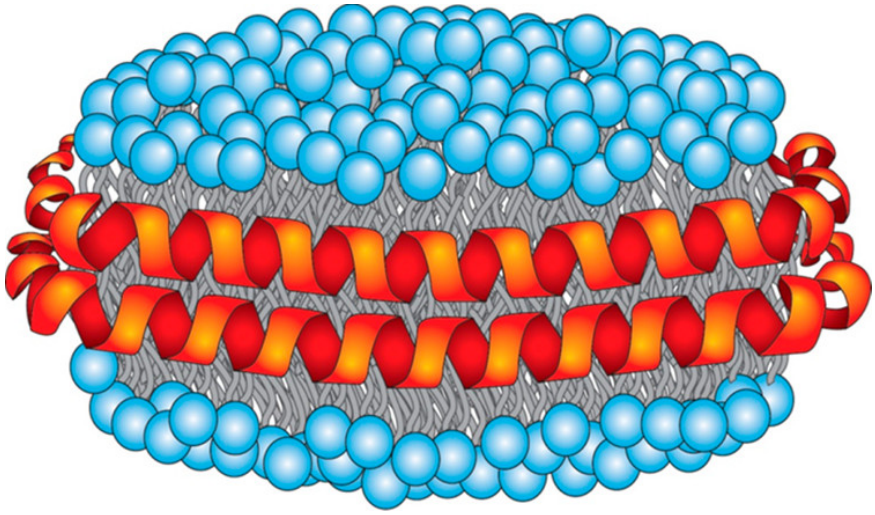


Figure 2.1 An illustration of a nanodisc. MSP proteins are seen in orange—there are two copies per disc. Lipids can be seen in blue with the polar head groups on either side and the aliphatic tails protected by the MSPs.

The scaffold protein used to assemble the discs is an engineered version of the apolipoproteins that performs this function. The MSPs consist of a series of helices that can be treated as modular units, and these units can be removed or repeated to change the length of the protein. Figure 2.2 shows the architecture of MSP1D1, and mutants that have removed single or multiple helices to make smaller particles. Thus, it is possible to tune the diameter of the final, assembled particle for a particular study and protein [30]. Additionally, one can choose the composition of lipids for particle assembly. These qualities make the discs and MSP exceptional tools for reconstituting IMPs into stable particles for molecular investigations [31].

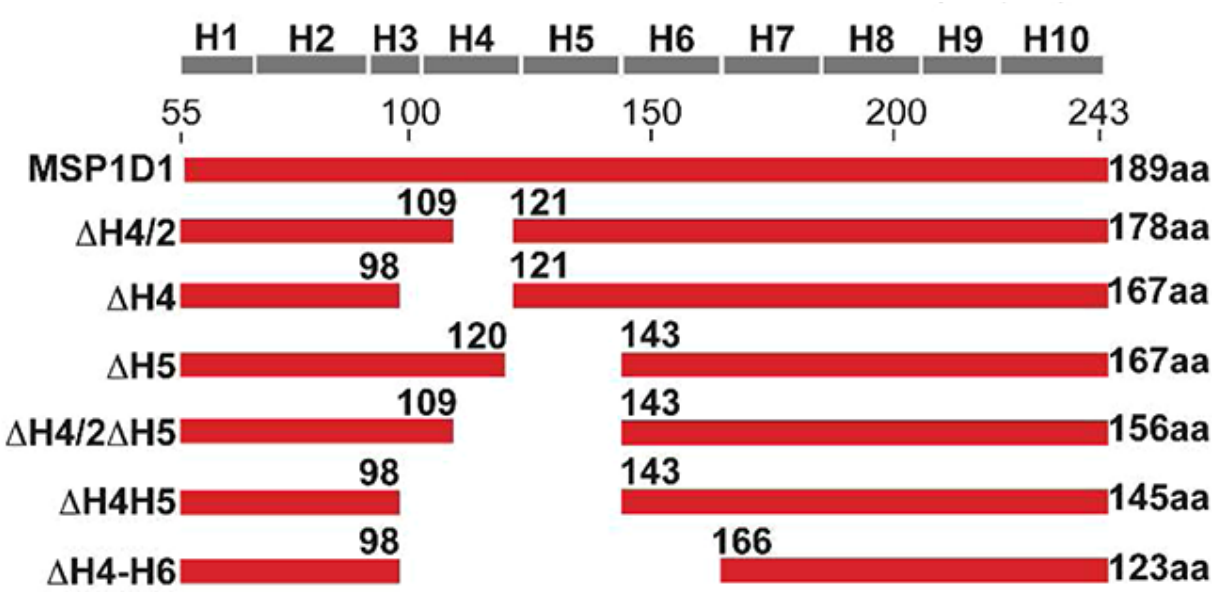


Figure 2.2 An illustration of the modularity of the MSP helices. The protein is composed of 10 helices that can be deleted to make a smaller diameter particle, as shown above. In theory, helices could be repeated to make larger particles. In this thesis, only MSP1D1 and MSPΔH5 were used or modulated [11].

2.2 ASSEMBLING NANODISCS

Since the development of nanodiscs, many types of IMPs with varying structures and topologies have been successfully self-assembled into the particles [32]. Nanodiscs cannot be idealized as a cure-all for IMP investigations, but they can render the protein easier to work. The workflow for IMP reconstitution into nanodiscs can be seen in Figure 2.3. In general, assembly requires detergent-solubilized IMP, lipids, and purified MSP to be mixed into solution. Over time, the lipids will exchange with the detergent in micelles, and with the additional of the anionic surfactant Biobeads, the detergent can be removed from the system. Simultaneously, the MSPs assemble around the lipids, and both particles with and without a reconstituted IMP will form. This requires further downstream purification to remove any of the empty discs that form and can be easily achieved by engineering particular tags on the IMP or size-exclusion chromatography (SEC) if the IMP has sizeable extracellular domains.

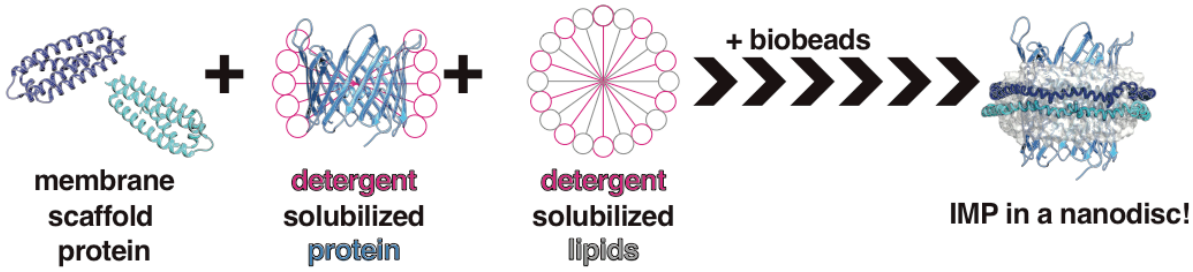


Figure 2.3 The general scheme for assembling nanodiscs and reconstituting IMPs into nanodiscs. MSP, detergent solubilized protein, and detergent solubilized lipids are mixed together and detergent is removed with Biobeads. The nanodiscs self assemble upon detergent removal.

The molar stoichiometric ratio of lipid to MSP to IMP is important for optimizing the final yield of the preparation, and this has to be tailored for each IMP and size of MSP used. Generally, one will calculate the amount of lipids to use based on the length of MSP, or the final particle diameter. If one wants to study an IMP as a monomer, it is best to have the MSPs in significant molar excess to encourage only one IMP to go into each disc. Conversely, if one wants to study an oligomeric IMP, it is best to lower the ratio of MSP to IMP to encourage more IMPs per disc. This approach can provide an control over the homogeneity and composition of the subsequent particles, but it requires that the IMP can be detergent solubilized and stable for at least several hours. This is sometimes not the case, as IMPs are notoriously challenging to work with. In the event the IMP under investigation is difficult to stabilize and purify with recombinant detergent purification techniques, it is possible to directly solubilize cell membranes with detergents of membrane-active polymers [33]. This method, although not used in this thesis, circumvents a separate protein purification step and can potentially preserve the native and functional form of the target IMP.

Successful incorporation of IMPs into nanodiscs requires efficient solubilization in addition to aggregation prevention of the target IMP—these conditions can be achieved by tinkering with the protein purification method, as well as the ratios of IMP to lipids to MSPs. Other factors to consider include the choice of detergent, the speed of detergent removal, and the identities of the lipids.

2.3 NANODISC APPLICATIONS

Once the particles are assembled, they are robust and stable. They can be frozen or lyophilized with or without an incorporated IMP. The MSPs have been extensively engineered to contain a wide range of tags for isolation, imaging, or surface immobilization. The IMP embedded nanodiscs can be treated like soluble particles, which is a massive benefit for performing biochemical and biophysical experimentation. This is evident from the growing body of literature that has been publishing on the mechanistic details of IMPs from studies utilizing nanodisc technology.

Nanodiscs have been used to study many of the the aspects of IMPs. For structural studies, nanodiscs have primarily been used for cryoEM studies and NMR studies. Although the particle tends to make a particular IMP more stable in solution, this does not make it easy to grow crystals with nanodiscs. There are a variety of obvious problems, with the first being that an inserted IMP into the artificial bilayer is azimuthally disordered. The protein and lipids will be mobile if the temperature is above, and this can dampen crystallization efforts. Although there are many challenges to crystallize nanodiscs, there are efforts being made to develop nanodiscs that are amenable to crystallization. This includes using antibodies that attach to the globular domains of the IMPs to provide a larger site for crystal contacts to form, as well as developing nanodiscs that can form a two-dimensional lattice. CryoEM, and to some extent NMR, have taken advantage of nanodiscs to successfully characterize IMP structures.

Paired with the advances in cryo-EM mentioned above, the use of nanodiscs to get high-resolution structures with the technique has been fruitful [34, 35]. For IMPs,

nanodiscs provide a means to avoid aggregation, preserve or enhance target function and structure, as well as provide protection during freezing and mounting. Full three-dimensional reconstructions have yielded high-resolution images of membrane proteins in nanodiscs. Recently, the structure of the TRPV1 ion channel it its unliganded, agonist-bound, and antagonist bound states were imaged to resolutions of around 3 A in nanodiscs [36]. Another study solved the structure of the Tc toxin to an average resolution of 3.46 A and first the first time resolved the trans-membrane fragment [37]. Additionally, the mechanism for direct insertion of the ten helices was illustrated based on comparisons with the prepore complex. There have been many studies utilizing IMPs in nanodiscs to dissect mechanisms with cryo-EM including investigations of lipoxygenase [38], drug efflux pumps [39], a magnesium channel [40], and the ribosome-SecYE complex [41]. Nanodiscs and cryoEM have quickly become a mainstay for the structural determination of IMPs and their interactions to form supramolecular complexes involved in signaling, energy transduction, and transport.

As previously mentioned, NMR has long been used to obtain structural and dynamic information on membrane proteins. However, the inherent tendency for aggregation when membrane proteins are removed from their native bilayer environment can be limiting for NMR spectroscopy. Assembling a membrane protein into nanodiscs; however, has opened a door for interesting, high-resolution studies. In recent years, we have witnessed an explosion in the use solution NMR methods with nanodiscs that have revealed interesting details of IMP structure and function. In the Wagner lab, Franz Hagn solved the first structure of an IMP in a nanodisc [11]. He revealed the subtle

conformational differences in OmpX when in the native bilayer environment, as well as developed nanodiscs that were amenable to NMR spectroscopy.

A recent NMR investigation of the oncogenic protein KRas4b illustrated the conformation was altered when bound to a nanodisc membrane surface [42]. This highlights the need for studying conformational changes of IMPs as they interact with lipid surfaces. The structural details of other membrane proteins involved in signaling at the membrane surface have been reported, such as the neurotrophin receptor in nanodiscs and numerous beta-barrel systems [43, 11, 44]. A direct comparison of the membrane protein BamA in different membrane mimetics and a review of nanodiscs in NMR applications illustrated this approach for structural determination [45]. Structures of the antiapoptotic BCL-2 family protein BCL-XL, a significant anticancer drug target, were compared for the soluble form and for the full-length form with the C-terminal fragment incorporated in nanodiscs [46] [47]. Results show that the C-terminal tail forms an α-helix in the membrane, but retains significant mobility and may be susceptible to proteolysis in soluble form. The mobility of the C-terminal helix may be related to a functional role of this protein in the apoptosis process and in conformational modulation of the globular domain [47].

Short of complete structure determination, nanodiscs can be used to define specific distances and orientations of a membrane protein in a native-like bilayer environment. Examples include conformational changes induced by ligand binding to the maltose transporter [48], critical structural motions of G-protein coupled receptors [49], and the revelation of multiple conformations of SNARE proteins in diverse prefusion

states [50]. Dynamics and transitions between inactive and active conformations of β2-adrenergic receptor in nanodiscs were characterized by NMR of selectively deuterated receptor, and found to be on millisecond scales [51], slower than in the detergent dodecyl maltoside (DDM). Here the population of the active state was higher in nanodiscs than in DDM. Dynamic transitions between closed and multiple open states monitored by NMR signals from methionine residues provided insight into a possible structural mechanism of biased signaling of a μ-opioid receptor [52]. The improved stability of nanodisc incorporated chemokine receptors CCR1 and CCR5, coupled with recently developed methods of NMR spectra analysis, provided necessary resolution for assignment of binding interface between these receptors and their ligand MIP-1α [53]. These results will help in evaluating the effect of two known sites of single nucleotide polymorphism of MIP-1α on the progress of autoimmune disease regulated by these chemokine interactions. The distinct advantages of the nanodisc platform for investigating GPCR structure and function are obvious.

Nanodiscs have proven to be an exceptional tool for the membrane protein biochemist. As they become a more routine tool for labs that are interested in the biochemical and structural mechanisms of IMPs, they will be optimized and tailored for particular studies and needs. Franz presented beautiful work and made custom nanodiscs to benefit the NMR community [11]. I am excited to present the work I have done to optimize nanodiscs for NMR and cryo-EM studies [12].

3

CIRCULARIZED NANODISCS

3.1 PROBLEMS WITH TRADITIONAL NANODISCS

As mentioned in the previous Chapter, nanodiscs have been a great tool for studying the biochemical mechanisms, structural dynamics, and physical insights to challenging membrane proteins. Working in the world of membrane proteins is undoubtedly difficult, and nanodiscs are not a cure-all. When constitution is successful, the particles can be used for a variety of experiments with enhanced stability and ease of workability. There are issues with the particles that need to be discussed to address the

motivation for the following experiments. The biggest problem with traditional nanodiscs is the diameter distribution that is seen in the particle upon reconstitution with a single construct of MSP protein [54].

For NMR experiments, the tumbling rate of the particle has a direct effect on the signals obtained for a particular spin system. Larger particles tumble at much slower rates in solution—this gives the opportunity for excited spins to dissipate their energy to the solution without producing a signal [55]. There are a variety of mechanisms for this relaxation to happen, but generally this is a main reason why larger particles produce poorer NMR spectra. In the case of collecting NMR data on traditional nanodiscs, some signal might be lost due to the difference in size of each particle. In theory, the differently sized particles will be tumbling at varying rates, and this will result in complicated relaxation mechanisms broadening or dissipating the signals. There have been limited structural studies of IMPs in nanodiscs, and this is potentially a major reason for this. There are several reasons that can lead to the non-uniform diameter of the discs in a single prep. Figure 3.1 illustrates some of these reasons. From the structure of a variant of MSP, it is known that the two MSP proteins interact in an anti-parallel orientation [56]. It was shown that helices 4 and 6 interact very strongly, as well as helices 2 and 3 with 7, and helices 8 and 9 with 10. It is possible that the N and C terminus of each MSP can be slightly open or slightly overlap, and still satisfy the necessary interactions to form a stable particle. This means that the disc can accommodate different numbers of lipids, and be different sizes. Not only does this dampen the signals in NMR experiments, but not having exquisite control over the size of the disc can lead to other issues.

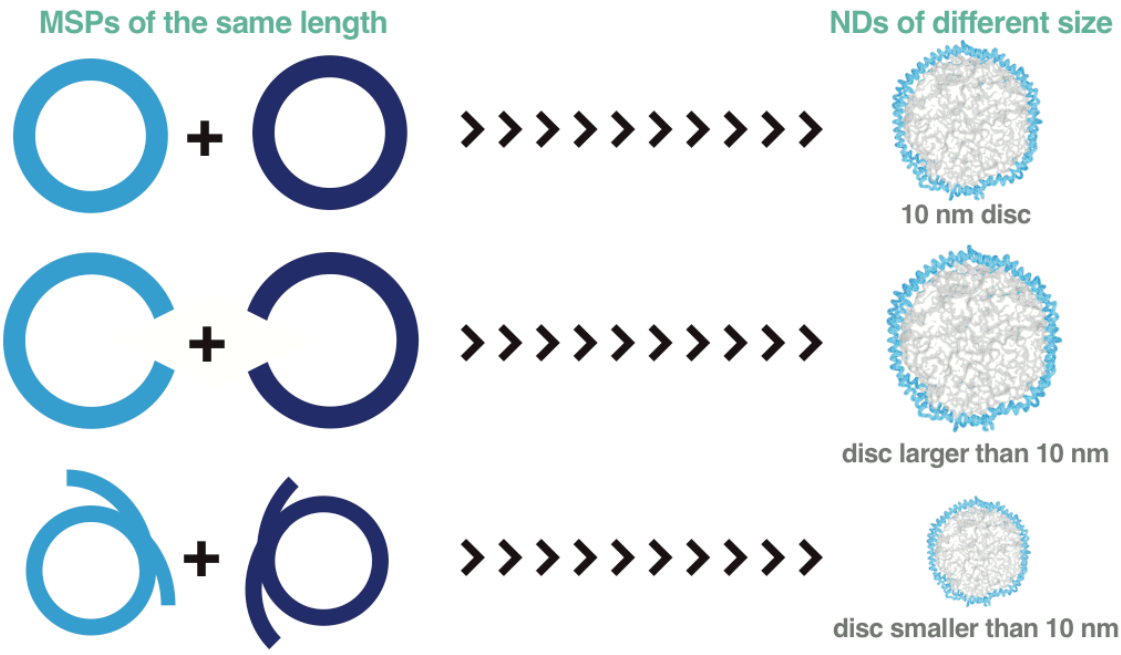


Figure 3.1 Potential reasons for the large diameter distribution in nanodiscs. Above is an example of nanodiscs made with an MSP that codes for a 10 nm particle. The top row indicates what should be seen in theory, but the latter rows indicate potential reasons for diameter variability.

If the discs are not the same size, it is possible that they will be able to accommodate different numbers of IMPs per disc [54]. Depending on the oligomeric nature of the IMP of interest, this can add a layer of complexity making it difficult to do experiments. In the case of VDAC-1, as shown in Figure 3.2, reconstituted in traditional nanodiscs it is possible to end up with three populations of discs with one, two, and three VDACs per particle [54]. If there is any interaction between each of the VDAC molecules—which has been shown for VDAC2 [57]—this can lead to complicated results. Each population of particles will give a different signal in experiments, and for NMR this is particularly bad. For example, any of the residues that interact in a dimer interface will give different signals from those residues in the monomeric form—effectively reducing the signals to be a function of the oligomeric proportions.

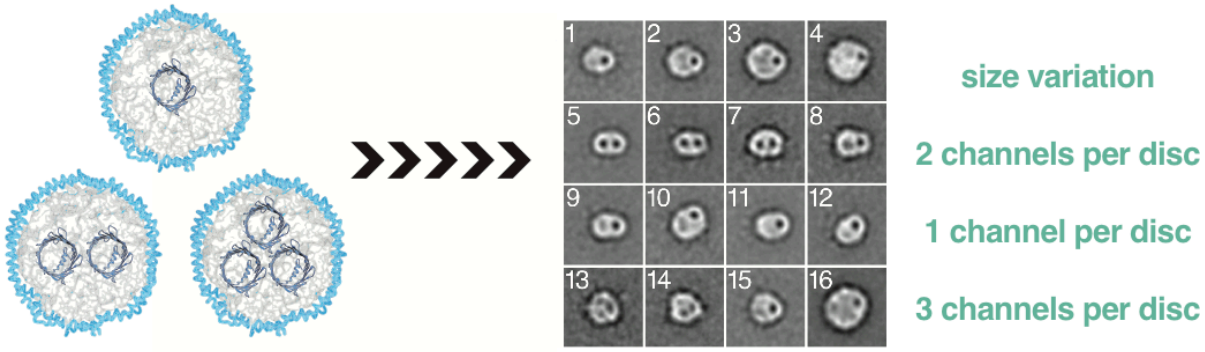


Figure 3.2 Variability in size of the reconstituted nanodiscs can lead to different oligomeric states of an IMP in a single prep. Shown here are EM images from VDAC-1 reconstituted into nanodiscs using MSP1D1. It is clear from the images that there is variation in size across the population of discs in the sample, and that the number of VDAC-1 molecules per disc can vary [54].

Experiments have been done to separate these discs by Analytical Ultracentrifugation (AUC) that take advantage of protein being more dense than lipids [58], but I found these experiments to be difficult. Additionally, working with quantities necessary for good NMR experiments was not feasible to get good resolution between the oligomeric populations with AUC. For years in the people in the Wagner Laboratory had tried to solve the biochemical challenges associate with nanodisc reconstitutions tailored for NMR experiments, and one of the most promising ideas involved using sortase A to covalently link the N- and C-terminus of the MSP protein.

3.2 Engineering MSP for Sortase Linkage

Forming a successful covalent bond between the N- and C-terminus of MSP requires a construct that is amenable to such chemistry. To achieve this, we created constructs of MSP that are shown in Figure 3.3. The MSP constructs previously used in

the lab have a His6-tag on the N-terminus of the protein with a consensus sequence ENLYFQ for Tobacco Entero Virus (TEV) protease cleavage. The constructs used for MSP circularization have the His6-tag at the N-terminus, but have another His6-tag inserted at the C-terminus of the protein. Prior to the C-terminal His6-tag is a consensus sequence LPGTG which is recognized by Sortase A. Additionally, just after the TEV protease consensus sequence is a glycine. This glycine will be the site of covalent linkage to a threonine from the sortase consensus sequence in the C-terminus. Previously, Franz Hagn had engineering a ΔH5 variant of MSP to make nanodiscs that were more amenable to NMR experiments [11]. We engineered this construct to have the same features that of the first MSP construct that will be circularized.

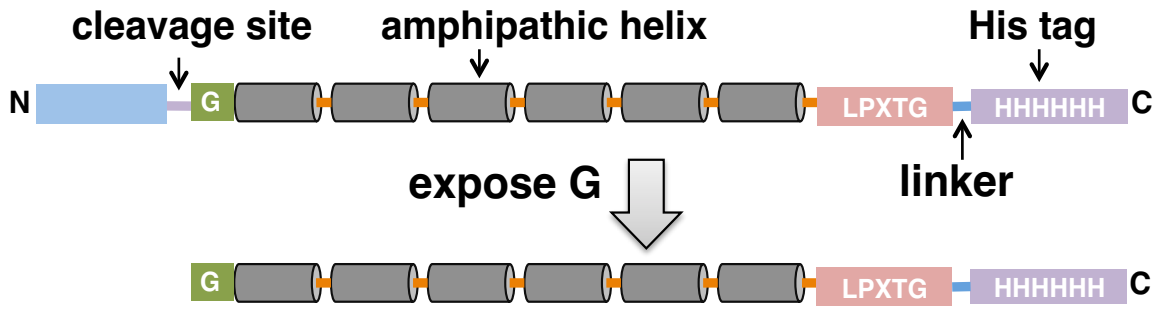


Figure 3.3 The general design for the cMSP constructs. There is an N-terminal tag indicated in blue. The cleavage site is just before an N-terminal glycine illustrated in green. The MSP helices are shown in grey, and can be repeated or deleted from the original construct. The consensus motif LPXTG is shown in pink, with a C-terminal tag that is cleaved upon circulization shown in purple.

3.3 CONSIDERATIONS FOR SORTASE

Sortase A was discovered in bacteria in the early 1990s [59] and produced recombinantly for biochemistry applications in 1999 [60]. Since then, it has become a standard tool in the suite of techniques employed by biochemists. Generally, sortase is used to add a tag of some kind to the protein that is not produced recombinantly with the protein. Thus, sortase is responsible for recognizing two different substrates in each reaction. The general transpeptidation reaction catalyzed by sortase enzymes is two-part. First, a catalytic cysteine acts as a nucleophile and attacks a peptide bond in the first substrate LPXTG recognition motif, between the threonine and glycine. This forms a tetrahedral enzyme-substrate intermediate. This intermediate collapses, releasing the C-terminal domain of the substrate. Second, a free amine from the second substrate—the N-terminal glycine—enters the active site and acts as a nucleophile to attack the enzyme-substrate intermediate. The collapse of this tetrahedral intermediate releases the newly formed peptide bond, regenerating the enzyme [61]. The kinetic mechanism for this reaction is shown in Figure 3.4, and is described as a ping-pong bi-bi hydrolytic shunt kinetic mechanism [62]. There are two pathways that result in regenerated enzyme—one that successfully adds a peptide bond between the two substrates, and one that hydrolyzes the threonine on the N-terminal consensus motif. To apply sortase to a single protein with both substrates sites was challenging, and there were several parameters that needed to be optimized to for successful circularization.

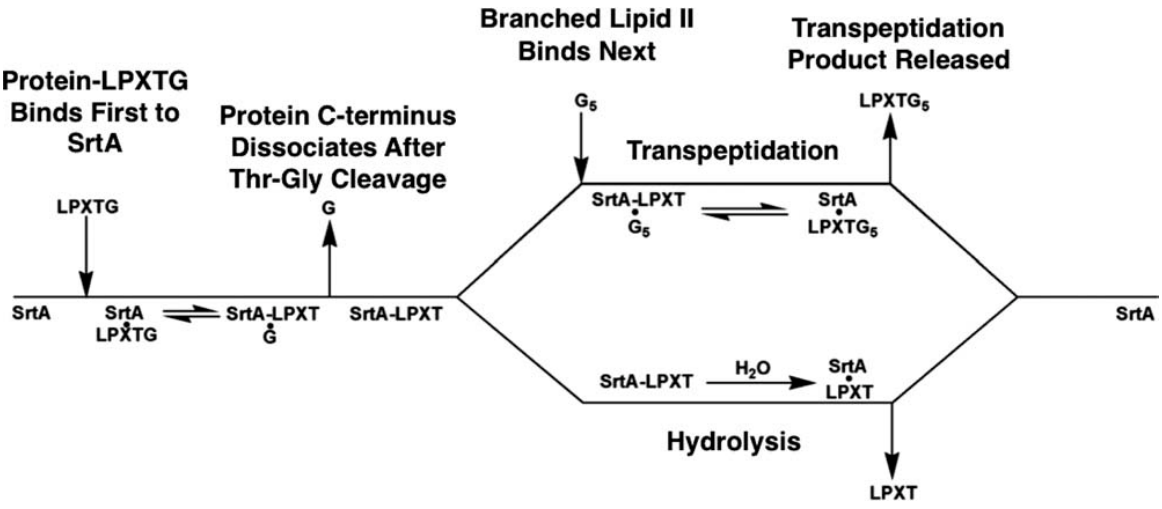


Figure 3.4 The ping-pong bi-bi hydrolytic shunt kinetic mechanism for Sortase A (SrtA). The transpeptidation is the desired pathway for the following experiments, and the hydrolysis pathway is extremely undesireable [62].

Normally, when using sortase to covalently link two molecules, the component containing the C-terminal glycine will be in significantly higher in concentration [61]. This reduces the time that sortase will have wait for nucleophile attack the C-terminal glycine when it has created the enzyme-substrate intermediate. The idea is that once sortase has cleaved between the threonine and glycine in the LPXTG sequence, the free amine from the second substrate is already present for nucleophilic attack. If the initial cleavage has happened, and the second substrate is not present, there is a probability that sortase will release the molecule and hydrolyze the threonine. This is particularly bad for my experiments. The hydrolysis product is virtually indistinguishable from the circularized MSP, and it is very difficult to separate the two. Additionally, if a particular prep is contaminated with the hydrolysis product, the resulting nanodiscs will still have the same properties that were deemed as issues above.

Sortase activity is pH dependent; however, hydrolysis reactions are less likely to happen at lower pHs [61]. If the number of free hydroxides in solution is lower, the probability of a hydroxide acting as a nucleophile in the second step of the sortase mechanism should be lower. To dissuade the pesky hydrolysis product from forming in solution, I decided to do the reactions at a pH of 6.5. Although lower, sortase still has activity at this pH. Additionally, I decided to complete the reactions at cold temperatures (4C). The idea was to lower the entire thermal energy of the reaction pathways in an attempt to completely cut-off the hydrolysis product pathway.

Another issue with using sortase to link two ends of the same protein was the potential for intermolecular bonding between two or more MSPs. I call this “head to tail” linking. As mentioned in the mechanism above, sortase does not care what the free glycine is attached to—it could be the same protein it is currently bound to, or another protein in solution. Depending on how many protein molecules are in solution, the likelihood of sortase linkage happening between the N- and C-terminus of the same molecule versus two different molecules can be skewed. The molecular weights of these products are elegant in that they are a multiple of the base MSP. If the MSP used gives a molecular weight of 11 kDa, the first off-target head-to-tail product will be 22 kDa, and the next will be 33 kDa and so on. These off-target products can be monitored via SDS-PAGE. Contrary to the hydrolysis products, these products of the reaction are relatively easier to separate from the desired target. After nanodisc reconstitution, it is even easier to separate these head-to-tail reactions as the molecular weight of each particle is a function of the diameter. However, it is not very desirable to spend non-trivial effort to

produce and purify MSPs for circularization and then throw away a significant proportion of it to go to unwanted products—especially if the goal is to produce NMR samples which require large amounts of protein. Additionally, even though the difference species might seem to be relatively simple to separate, it is quite a complicated matrix of products that will have an effect on the quality of the final results.

The most effective way to dissuade these head-to-tail products from forming is to keep the concentration of MSP relatively low in solution. In the successful experiments, MSP concentrations did not exceed 10 uM, and were ideally less. The downside of having a relatively dilute concentration of MSP for each reaction was the need for large reaction volumes. This added a significant amount of time to the downstream purification process, which allowed more time for protein to precipitate. However, this was a minor inconvenience to achieving the goal of a clean, precise reaction.

The final major issue for using sortase for this type of chemistry was to inhibit the reverse reaction from taking place, or quench the reaction. The candidate inhibitor that was used to stop the sortase reaction can be seen in Figure 3.5 [63]. AAEK2 covalently binds to the something on the protein, thus blocking it's ability to perform any chemistry. This inhibitor was key in trying to keep sortase from binding to the circularized products and undoing the desired chemistry.

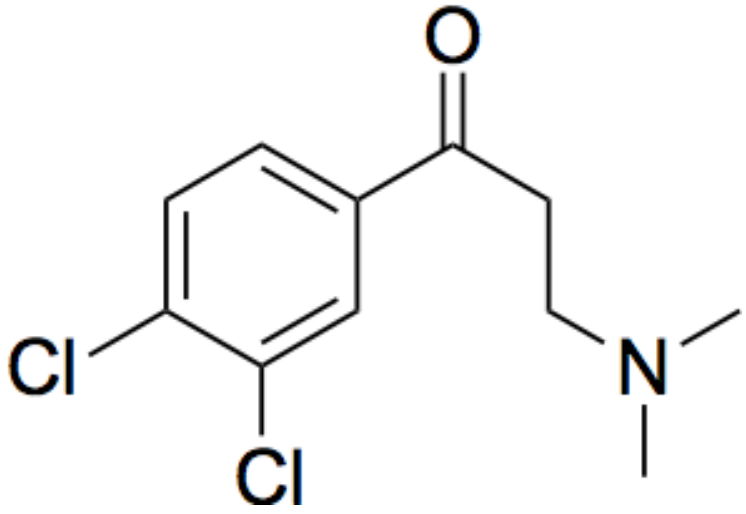


Figure 3.5 The covalent inhibitor AAEK2 binds to sortase A and inhibits function. This is a very useful tool for quenching the circularization reaction and preventing the reverse reaction from undoing the covalent linkages [63].

Once these main issues were tinkered with and optimized, I was able to fully develop a method to circularize MSPs. In the following sections I will discuss how I performed the reactions and made certain that I was making the desired products.

3.4 THE APPROACH TO CIRCULARIZE MSPS

The experimental scheme can be seen in Figure 3.6. The first step is to successfully purify the engineered MSP construct, which involves cleaving the N-terminal His6-tag to reveal the free glycine that will act as a nucleophile for linkage. For the reaction, cMSP and sortase is added to a solution in a 2:1 ratio. The reaction conditions were kept at a pH of 6.5 to dissuade hydrolysis products from forming and the solution was gently shaken overnight at 4°C.

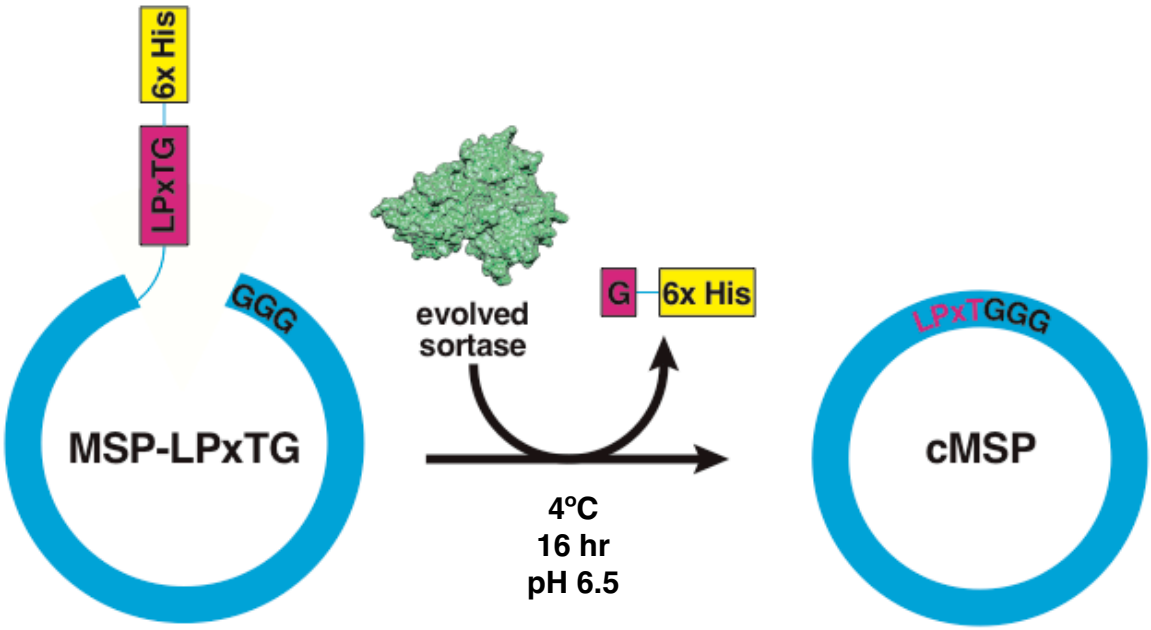


Figure 3.6 The general scheme for MSP circularization is shown. MSP-LPXTG is incubated with sortase at 4°C for 16 hours at pH 6.5. The resulting products are purified by reverse Ni-NTA chromatography and SEC.

Once the reaction had run for a sufficient amount of time, the sortase inhibitor was added and the solution was shaken at room temperature for some time to allow the covalent inhibitor to attach itself to sortase. This solution was concentrated, and run over an Ni-NTA column to remove sortase, as well as any of the unligated MSP. After this purification step, the solution was purified further via SEC to separate any of the head-to-tail products that may have formed. The peak collected from this chromatography experiment can be thought of as the desired product—circularized MSP that can be used to make nanodiscs. However, many tests were necessary to check that the right product was being made before this could be streamlined.

3.5 SORTASE CAN COVALENTLY LINK THE N- AND C-TERMINUS OF MSP

Figure 3.7 shows an SDS-PAGE result for covalently linking the N- and C-terminus of MSP1D1 with sortase. Since 12 amino acids are cleaved from the C-terminus of the protein there should be a shift of roughly 1.35 kDa on a gel. This is certainly possible to detect via SDS-PAGE, but does not tell us whether there was successful cleavage or hydrolysis products. Once we confirmed successful conditions for the sortase reaction, a quick SDS-PAGE result was a very good proxy for how complete the reaction was.

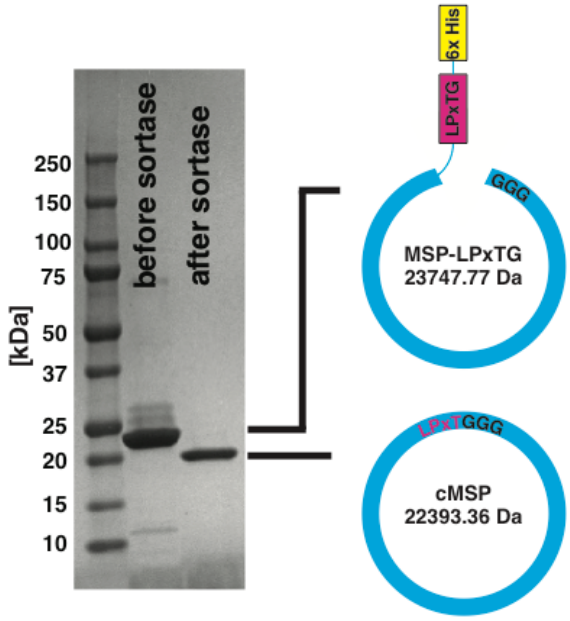
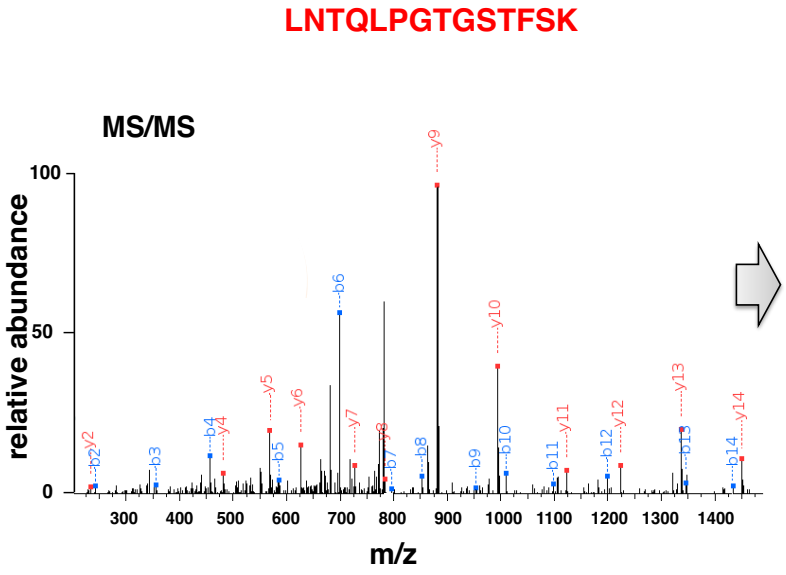


Figure 3.7 An SDS-PAGE result showing the MW of the protein before and after circularization. The MW of MSP-LPXTG is about 23.7 kDa and the MW of cMSP is abouve 22.4 kDa. In the gel it is clear that there is a shift in the product.

It was necessary to confirm the identity of the products seen in this gel, and this was done by mass spectrometry. Shown in Figure 3.8 are the results from tandem MS/MS experiments on the products for cMSP1D1. These experiments use trypsin protease to

digest the protein into fragments, and then reconstruct the protein from the masses of expected peptides. In the case for cMSP1D1, the MS/MS spectrum of a tryptic fragment shows the ligation of the C-terminal motif (**LNTQLPGT**G-His_6) to the N-terminal residues (**GSTFSK**). The intact mass for the peptide sequence **LNTQLPGTGSTFSK** is seen and expected. If the reaction was producing mainly hydrolysis product, the results would show the mass for the peptide sequence **LNTQLPGT**. Expected masses for b and y ions along with the peptide sequence are listed in the table. The b and y ions that were identified in the MS/MS spectrum are highlighted in blue and red. The sequence of peptide measured from cMSP1D1 is shown at the top.



seq	#	b	y	#
K	1	129.102	---	15
L	2	242.186	1450.754	14
N	3	356.229	1337.670	13
T	4	457.277	1223.627	12
Q	5	585.335	1122.579	11
L	6	698.420	994.520	10
P	7	795.472	881.436	9
G	8	852.494	784.384	8
T	9	953.541	727.362	7
G	10	1010.563	626.314	6
S	11	1097.595	569.293	5
T	12	1198.643	482.261	4
F	13	1345.711	381.213	3
S	14	1432.743	234.145	2
K	15	---	147.113	1

Figure 3.8 MS/MS results for the purified cMSP1D1. The m/z of 1450.754 confirms the identity of the peptide LNTQLPGTGSTFSK.

Since these experiments were tailored with the idea to make the reconstituted discs appropriate for NMR experiments, I also circularized the MSPΔH5 proteins. The results for this product are the same as cMSP1D1, so the MS/MS results are exactly the same as those illustrated in Figure 3.8. Table 3.1 gives the results for the calculated and observed masses for MSP1D1, cMSP1D1, MSPΔH5, and cMSPΔH5. I was successful in creating circularized MSP proteins in the face of several technical challenges. The next step was to see if these proteins could indeed form nanodiscs. This was an assumption that could not be assumed to be true, as the dynamics of nanodisc formation are poorly understand and, presumably, removing many degrees of freedom of flexibility by covalently linking the MSP protein to itself could potentially inhibit disc formation.

	Mass, Da			
	Linear (calculated)	Linear (observed)	Circularized (Calculated)	Circularized (Observed)
cMSPΔH5	21191.8	21192.8	19838.4	19838.5
cMSP1D1	23747.7	23752.4	22394.3	22398.1

Table 1. Characterization of intact cMSP proteins by mass spectrometry.

3.6 cNDs ASSEMBLE AND HAVE A SMALLER DIAMETER DISTRIBUTION

After careful purification and chemical linkage of the cMSPs, they can finally be used to assemble nanodiscs. Nanodisc assembly was performed in the same manner as traditional nanodiscs, and it was successful. Figure 3.9 shows a representative EM image of a SEC purified nanodisc assembly using cMSP1D1. As illustrated, the discs have

formed nicely, and analysis can be done to assess the uniform quality of the discs with respect to one another. Early on in these experiments, I worked under the notion that if I made nanodiscs with circularized MSPs the nanodiscs would be much more uniform in their size. Although I did not do any biophysical experiments to assess the quality of particles in solution, for example Dynamic Light Scattering (DLS) experiments, we were able to analyze the EM images for particle diameters. Figure 3.9 also shows the diameter distribution from a SEC purified nanodisc assembly using cMSP1D1. When compared to nanodiscs made with traditional MSP1D1, it is clear that circularization of MSP greatly enhances the uniformity of the resulting nanodiscs. From these results, 89% of the nanodiscs are between 11 and 12 nm for cMSP1D1 in comparison to only 33% of discs between 11 and 12 nm for MSP1D1. This result was both exciting and comforting. Our idea held true, and the potential to use these discs to develop better experiments was high.

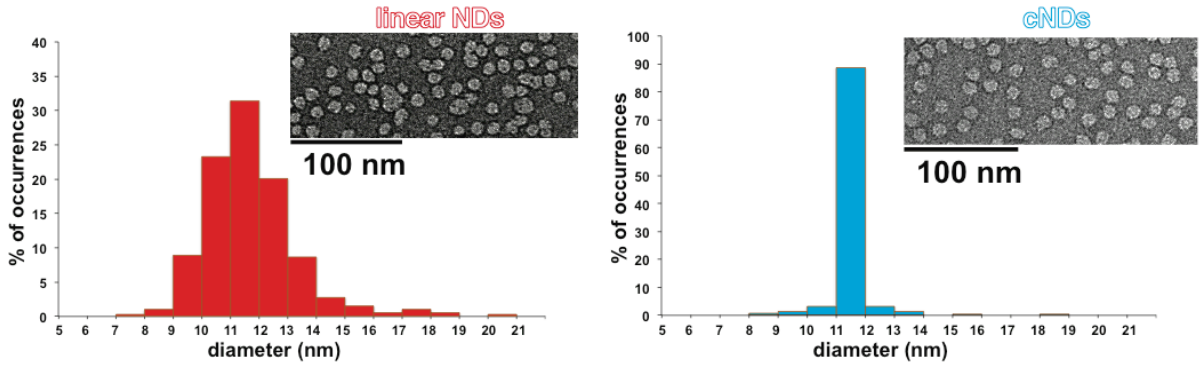


Figure 3.9 Representative EM images of nanodiscs assembled with MSP1D1 and cMSP1D1. Diameter distributions are shown for each, and it is clear that the distribution is tighter for circularized nanodiscs.

Of course, I repeated the assemblies for nanodiscs that were assembled and SEC purified with cMPSΔH5. The results were similar—cMSPΔH5 can assemble nanodiscs, and the quality of the preparation is high. This result was even more promising than the previous preparation since nanodiscs made with MSPΔH5 were shown to be successful for performing sophisticated NMR experiments on the protein residing in the disc. My hope was to move this construct and method forward and see if we can do better NMR experiments with nanodiscs made with cMSPΔH5.

3.7 cNDs ARE MORE STABLE THAN TRADITIONAL NANODISCS

One of the major benefits of using nanodiscs to study an IMP is the increase in stability of the particle as a whole. An increase in stability is generally always desirable for biochemical and biophysical experiments—if the sample does not aggregate or unfold on a quick time-scale, thorough experimentation can be done. It is also great to save time by not having to constantly prepare new samples every time an experiment is to be done. There are several examples where covalently linking a protein to itself increases the protein's stability. I used Circular Dichroism (CD) to test the melting temperatures of MSPs, cMSPs, and nanodiscs formed from each. The results of these melting curves can be seen in Figure 3.10. This result is not trivial. One of the major problems of working with IMPs in the lab is how "delicate" they can be. In general, it is very challenging to collect NMR data on IMPs for multiple days—the high temperatures and long experiment times certainly result in removing an NMR tube that is very cloudy at the end of an experiment.

Having a very stable IMP preparation is extremely beneficial to many other experiments, as well.

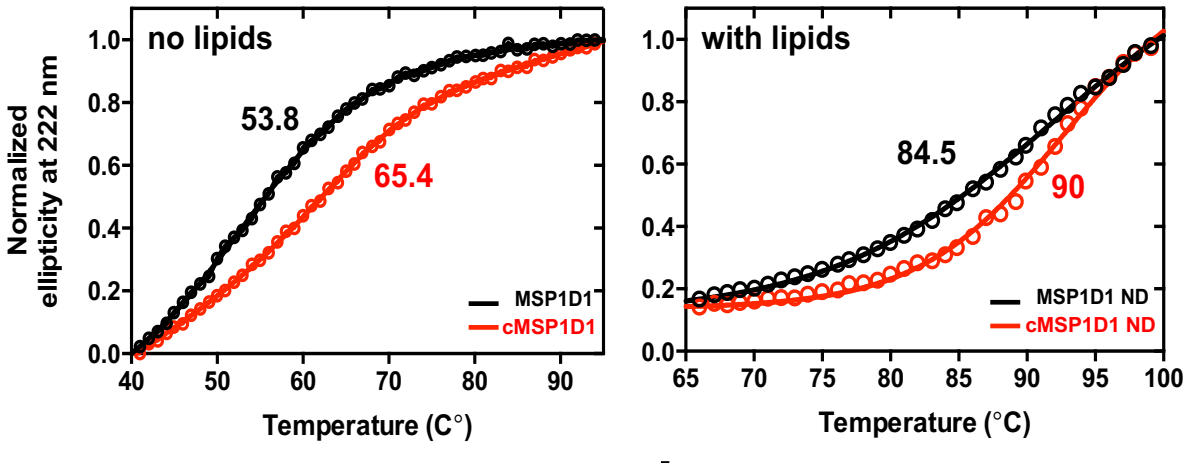


Figure 3.10 Covalent circularization stabilizes MSPs against thermal unfolding without and with lipids, stabilizes embedded VDAC1 and allows control of the number of channels embedded. Thermal unfolding of MSP1D1 (black) and cNW11 (red) without **(a)** and with lipids (**b**) followed by circular dichroism (CD) spectroscopy at 222 nm, the wavelength most characteristic of helical secondary structure. POPC/POPG lipids at a molar ratio of 3:2 were used to generate nanodiscs.

3.8 VDAC-1 IN cNDs IMPROVES STABILITY AND SPECTRAL QUALITY

Obviously, expending a significant amount of effort and being compulsive about trying to make “perfect” nanodiscs is not the most impactful work to be done. The true power in this biochemical force majeure is the potential to collect meaningful data on difficult systems. The main goal of all of this legwork was to see if we can improve the NMR spectra for the protein VDAC-1 in nanodiscs. First, of course, it was necessary to see if we could insert VDAC-1 into nanodiscs.

Figure 3.11 shows the SEC profile and SDS-PAGE result for VDAC-1 incorporation into nanodiscs. Also shown are EM images that indicate VDAC-1 is incorporated into the discs. I was very pleased to see that we could incorporate an IMP that is challenging to work with into the newly developed nanodiscs.

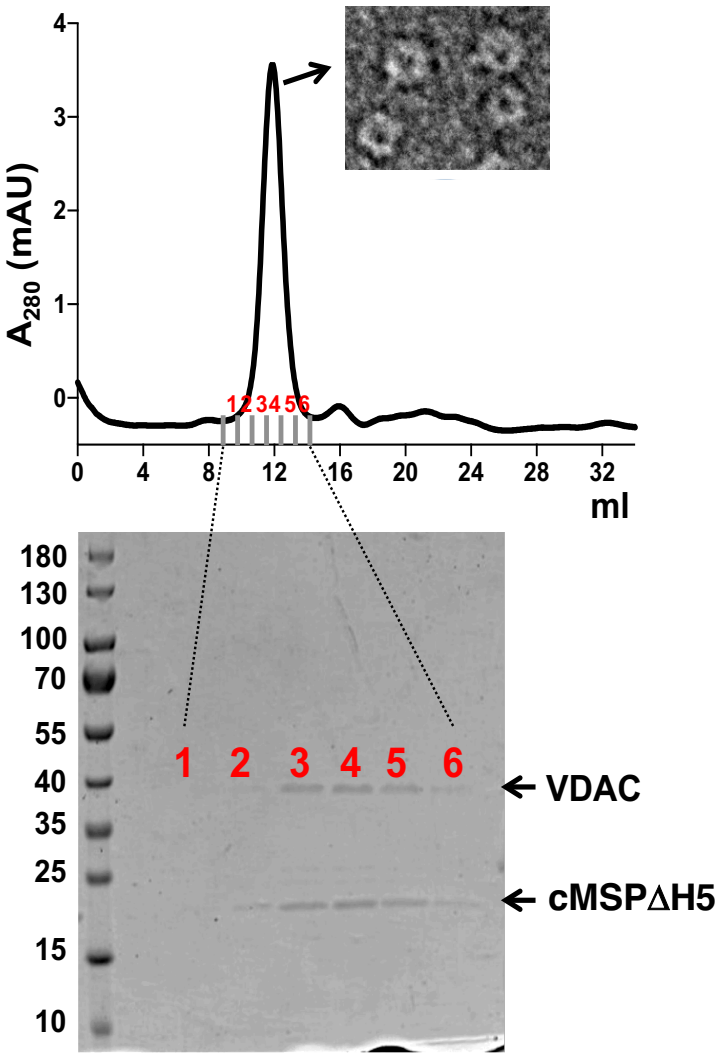


Figure 3.11 SEC profile and SDS-PAGE gel for VDAC-1 incorporation in cMSP1D1. Also shown is an EM image from the combined fractions.

The next step in these experiments was to look at the spectra of VDAC-1 in cNDs versus traditional nanodiscs and compare the quality of the spectra. Figure 3.12 shows HSQCs for VDAC-1 in nanodiscs made from MSPΔH5 and cMSPΔH5. From both spectra, we can see that the overall fingerprint and dispersion is relatively similar. However, if we look at the quality of peaks, we can see that the VDAC-1 in cNDs is of much better quality. This is particularly true for regions that are crowded in the center of the spectrum—that peaks are much better resolved. Additionally, it would seem that there are more peaks in the VDAC-1 in cNDs sample.

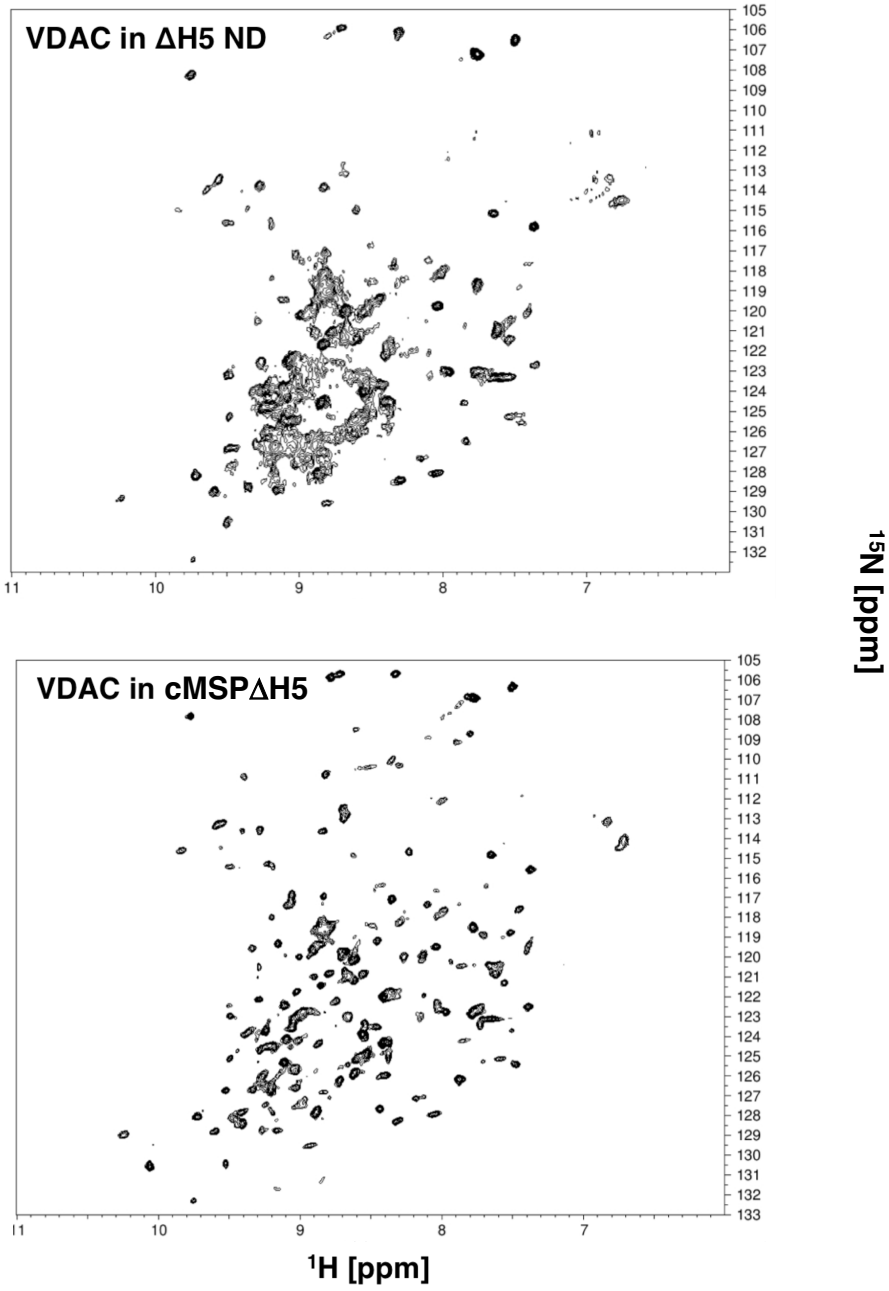


Figure 3.12: Comparison of NMR spectral quality of VDAC-1 in cMSPΔH5 and ΔH5 nanodiscs. (**top**) 15N-TROSY HSQC recorded at 45°C of 100uM ^{15}N-^{2}D-labeled VDAC1 in ΔH5 DMPC:DMPG =3:1 nanodisc acquired overnight on 600MHz spectrometer. (**bottom**) 15N-TROSY HSQC recorded at 45°C of 100uM ^{15}N-^{2}D-labeled VDAC1 in cMSPΔH5 DMPC:DMPG=3:1 nanodisc acquired overnight on 800MHz spectrometer. The spectral quality and sample stability are greatly improved by using cMSPΔH5 ND as compared to ΔH5 nanodiscs.

3.9 MONOMERIC AND DIMERIC RECONSTITUTIONS OF VDAC-1

Another nice application and result from the circularized nanodiscs is the ability to have exquisite control over the number of IMPs that can be reconstituted into each disc. Although ultimate control over this process is dependent on the nature of each IMP; however, reducing the variability of the size of each disc in a single reconstitution can make for more precise control. Based on previous experiments, it is possible to have one, two, or three VDAC-1 proteins per disc. In these experiments, I wanted to see if it was possible to make pure preparations of one or two VDAC-1 molecules per nanodisc. In previous nanodisc reconstitutions with VDAC-1, I have kept the ratio of VDAC-1 to cMSP to be 1:5. Since we hypothesize that VDAC-1 forms a meaningful dimer (like VDAC-2), I decided to vary the ratios of VDAC-1 to cMSP to control the number of molecules per disc. To encourage one VDAC-1 per nanodisc, I did the reconstitutions with a ratio of 1:8:60 VDAC-1:cMSPΔH5:Lipids. The idea was that the VDAC-1 would be dilute enough in solution to discourage dimer formation in the detergent micelles. Additionally, there is quite an excess of cMSPs present in solution with enough lipids to fill them, so plenty of the discs will be empty. If this is the case, it might be possible to make sure one VDAC-1 molecule goes into each disc with plenty of empty discs forming. The downside of this approach is the effort to produce cMSPs that will be used to make empty nanodiscs—although the empty nanodiscs could be salvaged for other experiments. Figure 3.13 shows the resulting SEC, SDS-PAGE, and EM images for VDAC-1 reconstituted into cMSPΔH5 nanodiscs with the attempt to insert one VDAC-1 molecule per disc. From the

EM images, it was clear that generally only one VDAC-1 molecule was inserted into each disc.

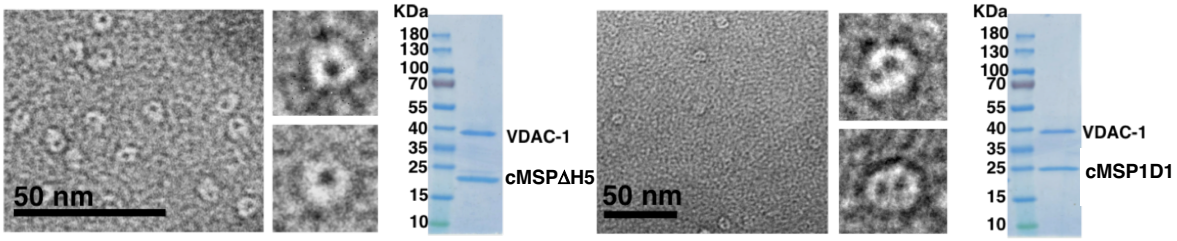


Figure 3.13 VDAC-1 monomeric (left) versus dimeric (right) preparations in cMSPΔH5 and cMSP1D1, respectively. Illustrated are representative EM images for each preparation and gels from the SEC purifications. The VDAC-1 molecule is easily visualized, and it is clear that control over the number of proteins per disc is possible.

To see if I could make nanodisc preparations with two VDAC-1 molecules per disc I used cMSP1D1 for the reconstitutions. Ideally, it would have been quite nice to use the same length of MSP, but I was worried that two VDAC-1 molecules wouldn't physically fit into the bilayer of the nanodisc. To make this preparation, the ratio of VDAC-1:cMSP1D1:Lipids was 1:1:75. The idea here was to keep VDAC-1 in excess so that all of the discs formed would have a pair of molecules in them. Figure 3.13 shows the SDS-PAGE and EM images for these experiments. It is clear from the EM images that there are two VDAC-1 molecules per disc.

The next obvious question to ask was if the NMR spectra for monomer versus dimer VDAC-1 looked any different. Since only about 40% of the residues are assigned for VDAC-1 in MSP1D1 nanodiscs, this experiment was purely to see if the control we have over the nanodisc preparations could be useful for answering interesting questions with NMR spectroscopy. 2H,15N-VDAC-1 was assembled into nanodiscs in the same manner as above, and TROSY-HSQCs were recorded for each sample. Figure 3.14

shows the side by side spectra recorded for the two samples. The regions highlighted indicate areas where the spectra show marked differences. At first, I thought that signals were broadening due to the larger size (and shorter relaxation rates) of nanodiscs made with cMSP1D1 versus cMSPΔH5; however, it is clear that some of the residues are shifted. To confirm that these are meaningful changes in the nature of VDAC-1—one could suggest a dimer interface—the residues will have to be assigned and much more data will need to be collected.

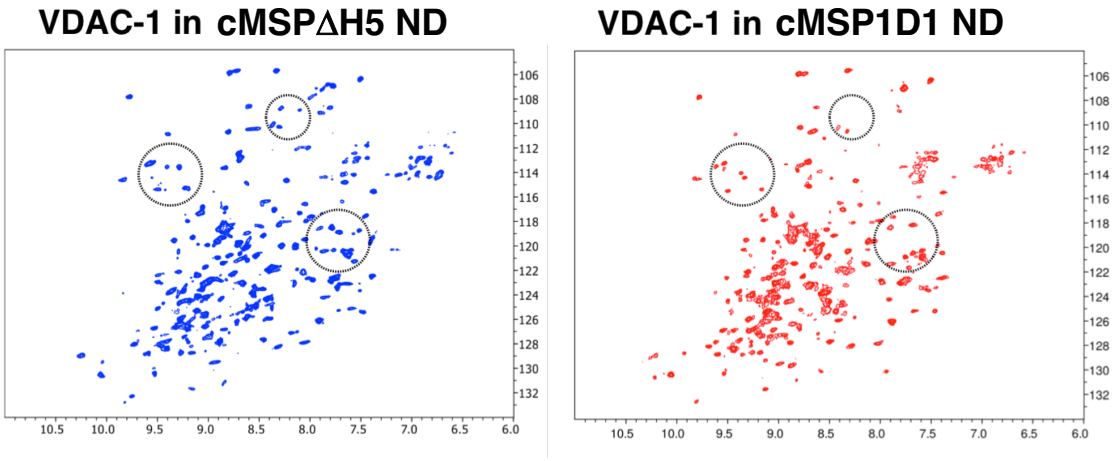


Figure 3.14 A comparison of monomer versus dimer VDAC-1 in cMSPΔH5 and cMSP1D1, respectively. Shown here are TROSY HSQC spectra of [*U*-^{2}H,^{15}N] labeled VDAC-1in either nanodisc.

4

Translation Initiation

4.1 TRANSLATION

The central dogma of biology simplifies an exceptionally complicated system into the blunt transformation of macromolecules. The basis of human genetics, deoxyribonucleic acid (DNA), is transcribed to messenger ribonucleic acids (mRNAs), and these mRNAs are translated to proteins. There are a very large number of players and parameters that dictate this central dogma, and many of the details are still being

discovered—six decades after it was first stated by Francis Crick. Translation refers to the transformation of mRNA to protein, and it is an incredibly important cellular process. In the start of protein synthesis, the small ribosomal subunit assembles with several initiation factors—as well as the initiator transfer RNA (tRNA)—to form the 43S preinitiation complex. This complex binds to the mRNA and scans the nucleotides for the start codon. Once the start site is recognized the Met-charged initiator tRNA is brought to the ribosomal subunit by eukaryotic initiation factor 2 (eIF2), it hydrolyzes GTP, and signals for the dissociation of several factors from the small ribosomal subunit. This leads to the association of the large ribosomal subunit, and the commencement of translation elongation.

4.2 TRANSLATIONAL CONTROL AND INITIATION FACTORS

Translational control is key for regulating cellular protein levels. Having a mechanism to limit the rate of expression at the translation step allows for the cell to respond faster to environmental changes. Although regulation occurs during several steps of the translation process, initiation is typically the rate-limiting step of protein synthesis [64]. It is a complex process with multiple steps that bring together the small and large ribosomal subunits, several initiation factors, and the initiator tRNA at the start codon of an mRNA. The following project attemps to characterize a complex that is involved in the middle steps of initiation—eIF4A and the HEAT2 domain of eIF4G.

Figure 4.1 illustrates the process of recruiting the small ribosomal subunit to the messenger RNA and scanning to the AUG start codon in eukaryotic cap-dependent

translation initiation [65]. mRNAs in eukaryotes utilize a 7-methyl guanosine triphosphate (m^7GTP) cap at the 5' end, and this significantly enchances translation efficiency compared to uncapped mRNA. Initially, eukaryotic Initiation Factor 4F (eIF4F) binds to the 5' end of the mRNA. eIF4F is made of three other eukaryotic initiation factors—eIF4E, eIF4A, and eIF4G. eIF4E is responsible for recognizing the m^7GTP cap, while eIF4G acts as a scaffolding protein, and eIF4A is the RNA helicase.

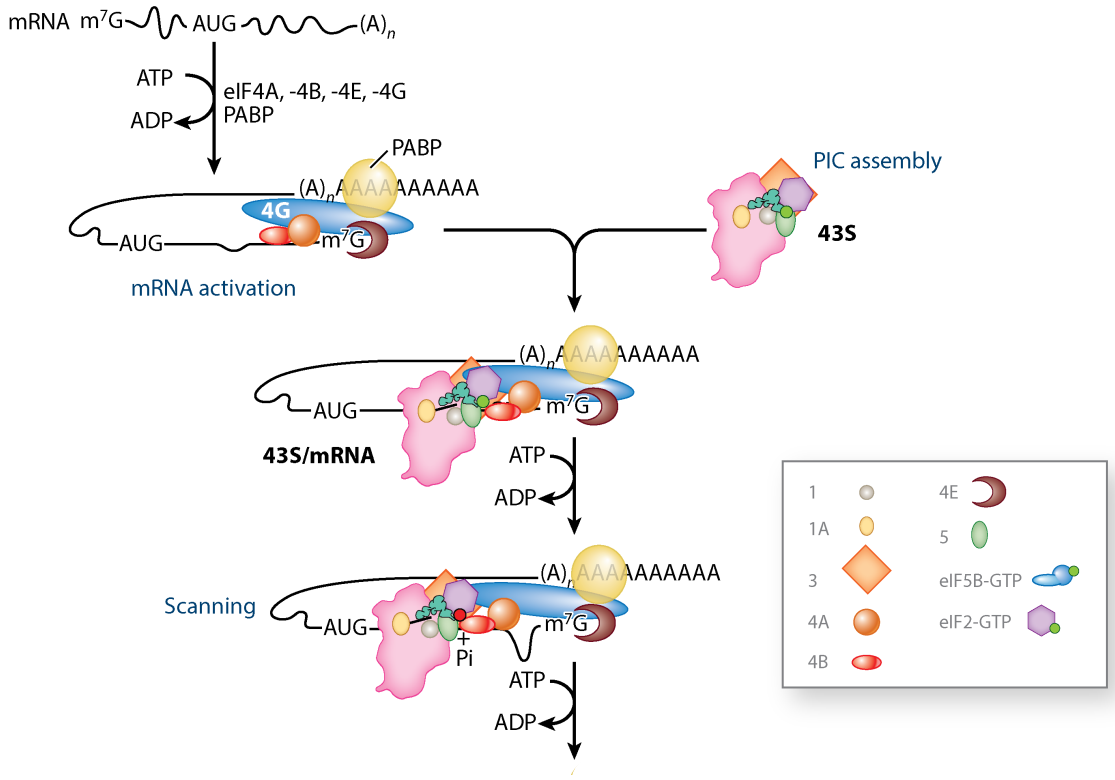


Figure 4.1 Overview of the role of proteins in the eIF family of initiation factors in cap-dependent translation.These proteins are responsible for recruiting the preinitiation complex, including the small ribosomal subunit to the 5' end of the mRNA for scanning. eIF4E recognizes the 5' cap, eIF4A unwinds the mRNA secondary structure, and eIF4G recruits the 43S PIC through its interaction with eIF3, while also serving as a scaffold that brings eIF4G and eIF4A together [65].

Upon assembling onto the mRNA, these factors are responsible for unwinding the secondary structure in the 5'UTR to clear a segment of mRNA for the 43S preinitiation complex to bind. In the subsequent step, the 43S preinitiation complex is recruited to the

mRNA, and this contains the small ribosomal subunit and many other translation initiation factors—including eIF3. This process is mediated by an interaction between eIF4G and eIF3. At this point, the preinitiation complex scans along the mRNA to the start codon with eIF4A unwinding secondary structure in the mRNA. Once the start codon is found, the initiation factors are released, and the 60S large ribosomal subunit binds to form the 80S ribosome and protein synthesis can begin.

Differences in the translation initiation process favor certain features of mRNAs over others. Strongly translated mRNAs have short 5' UTRs with very little secondary structures, and these are easily translated under suboptimal conditions. Weakly translated mRNAs can have long, complex leader sequences with structures than can impede the progress of scanning ribosomes [66]. eIF4A and eIF4G are crucial for unwinding the long, highly structure 5'UTRs commonly found in mRNAs—especially those commonly found encoding growth factors and oncogenes. The work presented in the subsequent sections of this thesis involve these two proteins—eIF4A and the HEAT2 domain of eIF4G.

4.3 EUKARYOTIC INITIATION FACTOR 4A (EIF4A)

eIF4A is the most abundant eukaryotic translation initation factor [67], and it is a prototypical member of the DEAD-box family of ATP-dependent RNA helicases. The DEAD-box family obtains its name from the conserved sequence motif Asp-Glu-Ala-Asp—and this sequence is believed to be responsible for coordinating a magnesium ion which in turn activates a water molecule to hydrolyze the terminal phosphate from ATP.

DEAD-box proteins generally have a core of two recA-like domains, and these contain a set of conserved motifs responsible for binding ATP and RNA, as well as flanking domains with auxiliary functions that may serve to enhance substrate binding or have other roles.

eIF4A consists only of the two core recA-like domains, and it lacks the auxiliary domains that are commonly found in DEAD-box helicases; however, other initiation factors that interact with eIF4A might play some of the same roles as these. There is currently no crystal structure for eIF4A. There are; however, structures of the N-terminal domain [68], the full-length eIF4A from yeast [69], and human eIF4A in complex with the translation repressor protein Pdcd4 [70].

Figure 4.2 illustrates a model of eIF4A. It is 406 amino acids in length and has a molecular weight of 46 kDa. It is an incredibly dynamic protein—with many different binding partners that presumably modulate the helicase function. From the illustration, it is clear that there are two domains in the protein—the N- and C-terminal domains. These domains are very likely to come together to form continuous surfaces for ATP and RNA to perform helicase functions.

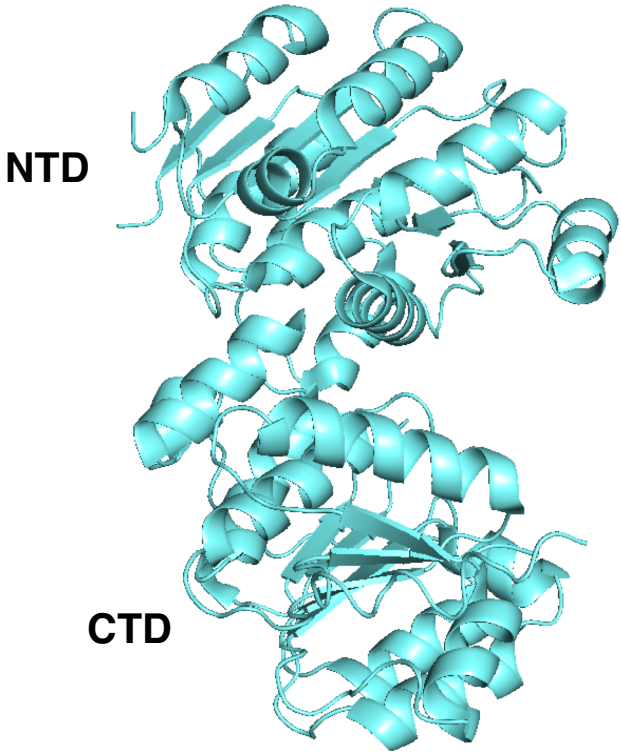


Figure 4.2 Model of human eIF4A. The N- and C-terminal domains are indicated. The protein is 406 amino acids in length and has a molecular weight of 46 kDa. It is the RNA helicase that unwinds mRNA for translation initiation.

eIF4A exhibits very low helicase activity on its own, but its unwinding efficiency is enhanced by interactions with other initiation factors such as eIF4G, eIF4H, and eIF4B [71]. A mechanism that enhances eIF4A activity is to promote the closed, presumably more active form of the helicase at the expense of more open states. The HEAT1 domain has been shown to enhance eIF4A activity through this mechanism [72].

4.4 EUKARYOTIC INITIATION FACTOR 4G AND THE HEAT2 DOMAIN

eIF4G is a large scaffolding protein at the center of the eIF4F complex—and this bridges the 5'-cap binding protein eIF4E and the RNA helicase eIF4A. It has additional

sites for a variety of other proteins involved in translation initiation. Historically, eIF4G has been thought of as three different components, and this was determined by picornaviral protease cleavage products [73]: the N-terminal third contains the eIF4E binding site, as well as a binding site for the Poly-A binding protein that interacts with the 3' end of an mRNA. The middle third of eIF4G interacts with the multisubunit initiation factor eIF3, which binds directly to the ribosome, and the C-terminal third contains a binding site for the kinase MNK1. Deletional and mutational analyses have shown that two distinct eIF4A binding regions are present in the C-terminal two thirds of mammalian eIF4G—one in the middle segment and one in the C-terminal segment [74]. The HEAT2 domain residues in this latter segment.

The role of the interaction between HEAT2 and eIF4A is unclear. It has been proposed to be a regulatory binding partner of eIF4A in conjunction with HEAT1 [75]. A solution NMR structure of the HEAT2 protein has been solved, and a representation of this can be seen in Figure 4.3 [76]. The domain has a molecular weight of 27 kDa, is mainly alpha helical, and can be very well expressed in the laboratory.

Figure 4.3 Structure of the HEAT2 domain of eIF4G.

4.5 THE EIF4A-EIF4G-HEAT2 COMPLEX

From previous experiments, we know that eIF4A and the HEAT2 domain of eIF4G form a stable complex [75]. Scanning or unwinding of mRNAs during translation initiation occurs through dynamic interactions of eIF4A with eIF4G-HEAT1 and HEAT-2 in association with RNA templates. eIF4G and eIF4A have poor affinity for RNA alone, but they efficiently bind RNA in a ternary complex. Prior studies showed that eIF4A binding to eIF4G HEAT1 enables helicase associations with RNA, while binding to HEAT2 counteracts them [75]. It is therefore plausible that increased eIF4A-HEAT2 binding correlates with suppressed RNA binding of the translation initiation helicase complex. A recent study suggested that increased eIF4G HEAT2-eIF4A interactions inhibits RNA-

binding properties of the translation initiation helicase complex. A recent study shows that phosphorylation of Ser1232 in the HEAT2 domain strongly increased its interaction with eIF4A, and this in turn decreased the association of eIF4G and eIF4A with RNA [77].

The details of the this interaction are unknown, and knowing the molecular structure of the complex would be exceptionally advantageous for understanding the basic biological mechanism behind this proposed regulation. Previous experiments in the lab have modelled the structure of the two using Paramagnetic Relaxation Enhancements, and this can be seen in Figure 4.4. HEAT2 is thought to sit slightly between the two lobes of eIF4A—possibly opening it up and encouraging it to become unbound from RNA.

The next section will describe the theory and scheme for using paramagnetic ions and NMR spectroscopy to validate the model of the eIF4A-HEAT2 complex. Then, in Chapter 6, experiments that show this method is feasible for the complex are presented.

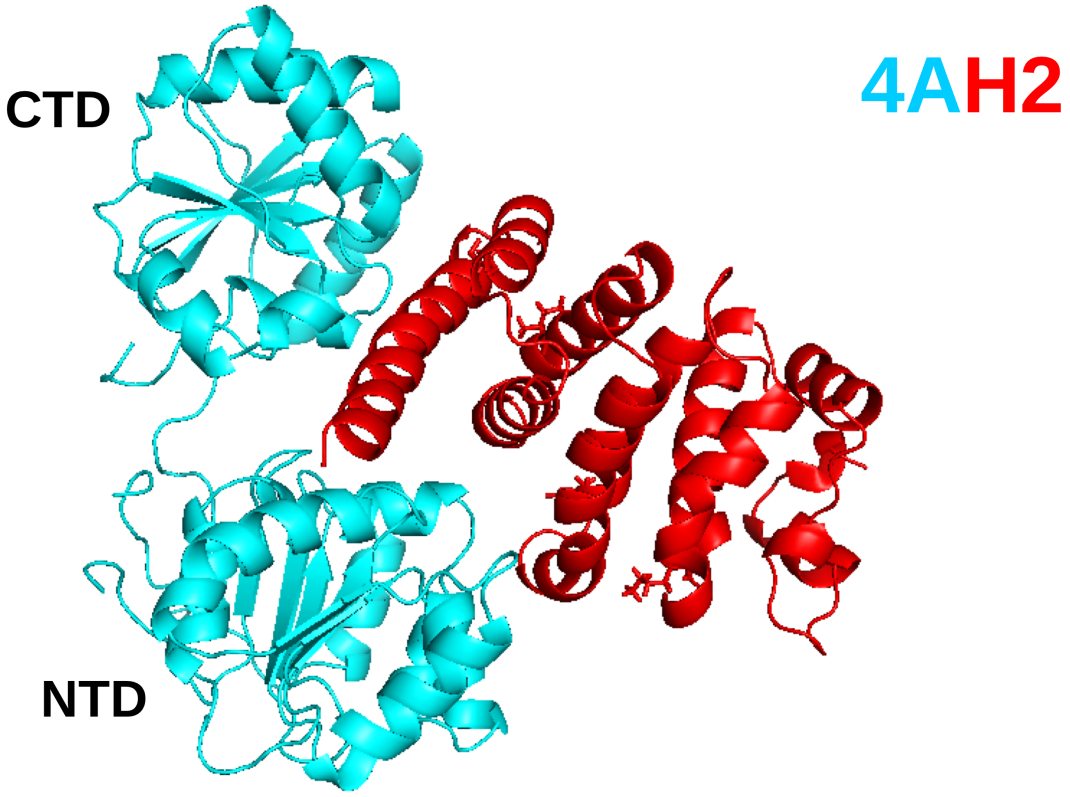


Figure 4.4 A model of the eIF4A-HEAT2 complex from unpublished PRE measurements (C. Wasmer). eIF4A is in cyan, and HEAT2 is in red. It would seem that HEAT2 wedges between the CTD and NTD of eIF4A, potentially reducing the helicase's affinity for RNA.

5

PSUEDOCONTACT SHIFTS

5.1 PARAMAGNETIC METALS IN NMR SPECTROSCOPY

Unpaired electrons can be a powerful tool in NMR spectroscopy—the magnetic moment of a free unpaired electron is about 660 times greater than that of a proton. These single, unpaired electrons can cause large effects in NMR spectra, including significant quenching of the signals surrounding the paramagnetic center in a distance-dependent manner. This paramagnetic relaxation enhancement (PRE) is a result from magnetic dipolar interactions between an NMR-active nuclei and the unpaired electron of the

paramagnetic center that induce faster nuclear relaxation rates. Another effect of an paramagnetic center on an NMR-active nuclei is called the psuedocontact shift (PCS). It is a change in chemical shift due to the presence of a paramagnetic center, and is mediated through-space. The magnitude of the shift is distance-dependent, and—as described below—different paramagnetic centers will induce different shifts on a spin. PCSs can illustrate the details of long-range molecular distances up to 60A—going beyond the limits of the Nuclear Overhauser Effect (NOE) of about 5A. First, I will describe the theory behind the PCS, and then introduce practical applications to utilize the technique in NMR spectroscopy.

5.2 THE ΔX TENSOR

For a point dipole in a rigid molecule—which proteins are not, but this approximation simplifies things—the PCS can be described by an equation (Equation 1, shown below) that includes the location, orientation, and degree of anisotropy of the magnetic susceptibility associated with the paramagnetic center [78].

$$\Delta\delta^{PCS} = \frac{1}{12\pi r^3}[\Delta\chi_{ax}(3cos^2\theta - 1) + 1.5\Delta\chi_{rh}sin^2\theta cos^2\phi]$$

$\Delta\delta$PCS is the PCS (which is a dimensionless number reported in ppm), r is the distance of the paramagnetic center to the nuclear spin, ΔX_{ax} and ΔX_{rh} are the axial and rhombic components of the magnetic susceptibility anisotropy (ΔX) tensor, and theta and rho are the polar angles describing the position of the nuclear spin with respect to the principle axes of the ΔX tensor. The magnetic susceptibility tensor, X, is thought of as an object

with three orthogonal axes centered on the paramagnetic center. The length of each of these axes describes the magnitude of the magnetic moment induced by an external magnetic field if the axis is aligned with the magnetic field. The X tensor is spherical if all three principle axes have the same length; however, PCSs are generated only if the X tensor is anisotropic with the axes having varying lengths. The ΔX_{ax} and ΔX_{rh} components are defined by:

$$\Delta\chi_{ax} = \chi_z \frac{(\chi_x - \chi_y)}{2} \qquad \text{and} \qquad \Delta\chi_{rh} = \chi_x - \chi_y$$

with X_x, X_y, and X_z denoting the susceptibility values along the three principal tensor axes. From this, it is clear why the PCSs are zero if the three principal axes have the same length. The change in magnetic susceptibility tensor, ΔX, can also be represented as isosurfaces that follow the coordinates of PCS values derived from Eq. 1. In this case, the positive and negative PCSs are represented with red and blue, respectively. Figure 5.1 illustrates an example of what the isosurfaces might look like if mapped to a tagged protein. To fully utilize the PCSs there are a few parameters that must be optimized for the protein of interest, the paramagnetic center, and the tag harboring the unpaired electron.

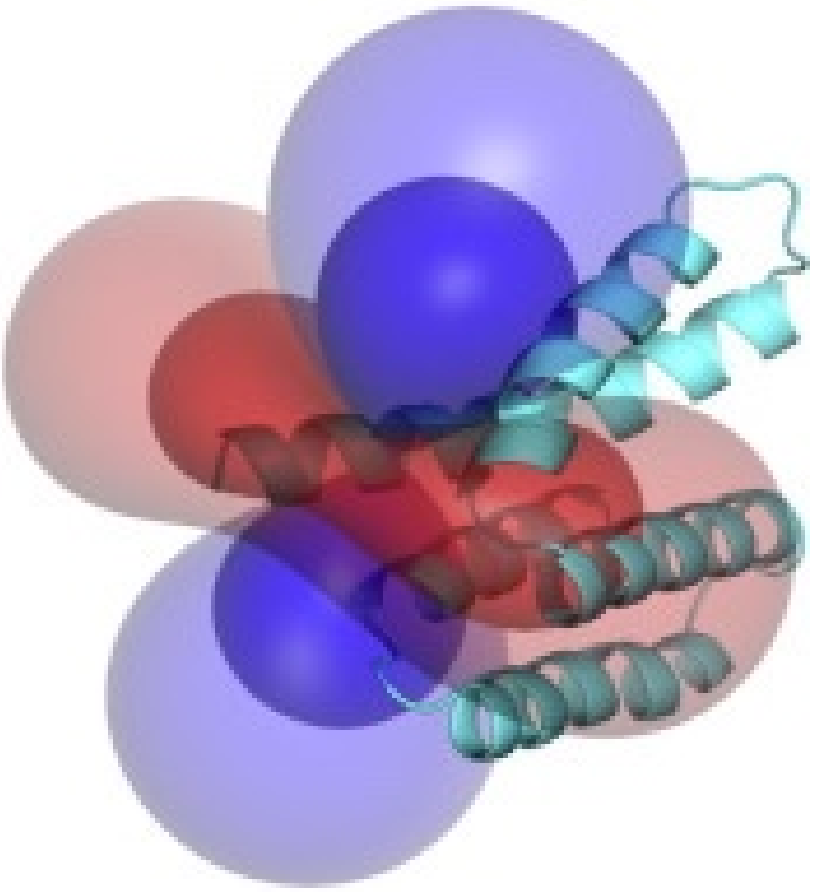


Figure 5.1 An illustrated representation of the isosurfaces that will cause PCSs in a protein. An example protein is shown in cyan, and the isosurfaces are shown in blue and red. Any residues within the blue region would be subject to negative shifts, while any residues within the red region will be subject to positive shifts.

5.3 TAGGING PROTEINS AND TAGS FOR PCS EXPERIMENTS

Most tags that can harbor a paramagnetic ion are thiol-reactive. This is an obvious disadvantage because it requires the protein to have a single, solvent-accessible cysteine to react with the tag. This can normally be achieved through mutation but becomes very challenging if the protein requires cysteines for function or loses stability without it's naturally occurring cysteines. There are methods to circumvent removing cysteines by using genetically encoded amino acids containing other reactive moieties to attach the

tag; however, the expression levels can be low and the non-native amino acid precursors can be cost prohibitive. Other considerations to take into account include the solvent accessibility of the tagging site—the tags themselves can vary a bit in size, but all require the cysteine to be solvent accessible. Tagging reactions are difficult to complete under denaturing conditions, and it is best to pick a site that is not too flexible as this can confound the distance constraints obtained. One must choose the tagging site wisely, and it is likely several experiments to screen for appropriate tagging sites will have to be done. In addition to the physical location of the tagging site, it is also possible to vary the paramagnetic ion used.

Metals are excellent candidates for PCS experiments; however, among the transition metals only high-spin Co^{2+} is a reasonable choice to tag the protein—it combines fairly large PCSs with relatively small PREs. Lanthanides produce equal, if not larger, PCSs to Co^{2+}, and have relatively small PRE effects. There is no known biological function for lanthanides, so specific lanthanide binding sites in proteins are exceptionally rare. The aforementioned point-dipole approximation works exceptionally well for lanthanide ions as their unpaired electrons are found in f-orbitals, and these do not interact as strongly with ligand orbitals like the d-orbitals of transition metals. The electronic properties of lanthanide ions are very similar and their oxidation state (3+) is exceptionally stable—this means that a lanthanide binding site on a synthetic chelating compound can bind any of the lanthanide ions with similar affinity. This feature greatly reduces complexity of combining experiments with different lanthanides—as you can use the same tag for each experiment. In addition to collecting data from experiments that

contain lanthanides displaying large PCSs on the NMR spectra, it is important to run the experiment with a diamagnetic tag for reference. With the diamagnetic reference, it is possible to measure even very small PCSs accurately. The last general parameter to consider is the structure of the tag itself.

Ideally, the perturbation to a protein system would be minimal to preserve as natural of an environment as possible. Synthetic post-translational modifications to proteins can change the structural dynamics and function of a protein, but such tags can also open a whole new realm of questions to be asked about the system. Attaching synthetic tags to a protein post-purification is a challenging problem—proteins contain a diverse range of environment dependent functional groups, and they can be delicate—susceptible to irreversible denaturing or degradation if put in non-ideal conditions. For all of the experiments in this thesis, a single-arm cyclen tag was used to collect data. Double-arm tags can be useful for reducing the metal mobility—thus tightening up the data. Unfortunately, they also carry the potential to form covalent dimers and multimers which can significantly complicate the resulting spectra. The tag used in all of the experiments can be found in Figure 5.2, and has been termed C1 tag. Some tags can generate multiple species due to cis/trans isomerization of the peptide bond with the linker or a slow conformational equilibrium between two different lanthanide coordination geometries. The C1 tag solves the problem of multiple species and produces a single set of paramagnetic peaks through its structure. It also limits the motions of the lanthanide relative to the protein due to its bulkiness, and—depending on the solvent exposure of the tagging site—this allows for the ΔX tensors to be large. Additionally, it's chiral partner

C2 tag could be used to produce a different ΔX tensor at the same protein tagging site. The C2 tag is the enantiometer of C1 of opposite chirality—with all of the aromatic arms at the amides pointing the opposite direction. When all of the aforementioned parameters can be optimized—and are accessible to the protein system of interest—PCSs can be a powerful tool to gain long-range distance information beyond NOEs and have many advantages over PREs.

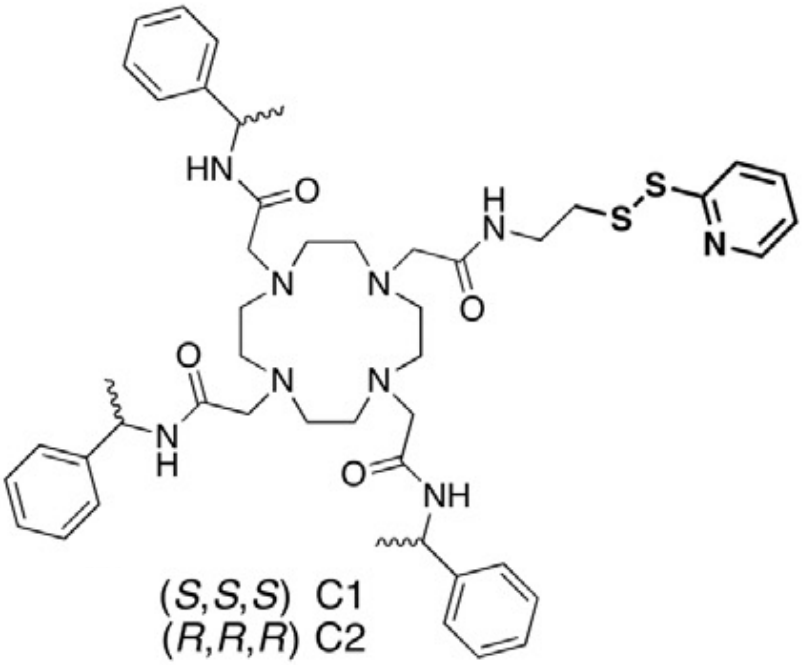


Figure 5.2 The molecular structure for the C1 and C2 tag. The arm on the right portion of the molecule reacts with a free cysteine on a protein to covalently bind to it, while a lanthanide ion can be coordinated in the center heteroatom ring.

If the experiment is prepared thoughtfully, measuring PCSs can be straightforward—it is simply measured as the chemical shift difference between a sample tagged with a paramagnetic metal ion and the diamagnetic reference. This can make experiments quite rapid, since you can generally just use Heteronuclear Single Quantum

Coherence (HSQC) NMR experiments. HSQCs are amenable to lower sample concentrations, and—in high-field magnets with optimized pulse programs—experimental data collection can be much shorter than collecting NOEs, Residual Dipolar Couplings (RDCs), or quantitative PREs. A massive benefit is that intermolecular PCSs easily average to zero. This is due to the fact that the PCSs depend not only on the distance of the nuclear spin from the paramagnetic center, but also on the orientation of the metal complex relative to the coordinate frame of the nuclear spin. If an interaction is non-specific—for example between two of the same tagged proteins—orientational averaging of the paramagnetic center exposes nuclear spins to both positive and negative PCS effects which average to zero. If the PRE effects are quite large and the solution is concentrating, non-specific quenching of signals can confound the spectra significantly.

Incomplete tagging can be a prohibitive problem when measuring quantitative PREs as the tags do not change the chemical shift of the nuclear spins compared with the diamagnetic reference. If the tagging reaction fails to reach 100% completion, the paramagnetic and diamagnetic species will overlap, and this will complicate the analysis of quantitative PRE measurements. Conversely, PCSs can benefit from incomplete tagging. Since the output is a chemical shift of a nuclear spin, if untagged protein is present, both the shifted peak from the paramagnetic tag as well as the unchanged peak from the diamagnetic reference will be present in the spectrum. This can simplify analysis when you overlay the two spectra.

The final benefit of PCS is the superior distance ranges that can be accessed with the technique. As shown in Equation 1, PCSs decrease with increasing distance r from

the paramagnetic center with an r^{-3} dependence. PREs decrease with an r^{-6} dependence. This smaller distance dependence of PCSs is advantageous for structure analysis of large proteins, and especially useful for studying large protein-protein complexes—which is detailed in the following sections.

6

PCS MAPPING ON THE EIF4A-EIF4G-HEAT2 COMPLEX.

6.1 THE MUTATIONAL SITES ON EIF4G-HEAT2

Gathering PCSs rely on a non-native tag that is chemically ligated to a single cysteine mutant. The location of the tagging site is very important for collecting meaningful data, and to get complete data sets there will have to be multiple tagging sites. In the case of the complex formed between eIF4A and HEAT2 it would be ideal to have tagging sites on both eIF4A as well as HEAT2. This would allow for PCSs to be seen on both

sides of the complex without PRE effects dampening the quality of the spectra. However, every attempt to create a single-cysteine mutant of eIF4A was unsuccessful, and trials to undergo the tagging reactions with eIF4A generally left the protein very unhappy. Thus, we decided to only use HEAT2 to harbor the PCS tags—it is generally a very stable protein and can undergo the mutations necessary to make it amenable to PCS experiments.

The dynamics of the site where the tag will attach is very important. The location needs to be somewhere that is solvent accessible so that the tag can come close enough for the attachment chemistry to occur. The site should not be too flexible, though. The tag itself will have unwanted rotational and translational degrees of freedom that can muddle results. If this is paired with an extra flexible region on the protein the distances obtained from the experiments can have large margins of error. An additional concern is to not perturb the binding interface for the two proteins. With all of these considerations in mind, we decided on the tagging sites that are illustrated in Figure 6.1. First, we had to screen these sites to see if protein itself could be tagged at these sites, and then run an additional screen to see if there were visible PCSs on both the eIF4A side of things, as well as the HEAT2 side of things.

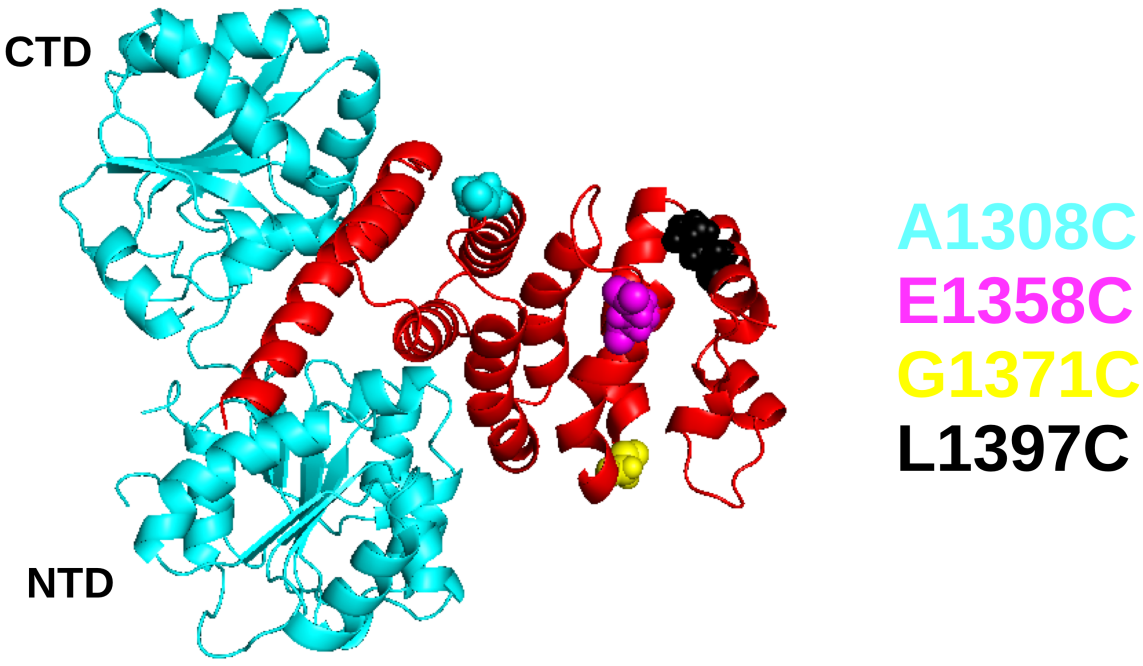


Figure 6.1 The single cysteine mutants of HEAT2 (red) are mapped onto the protein. eIF4A is shown in cyan. The four mutants used in the experiments below are A1308C (cyan), E1358C (magenta), G1371C (yellow) and L1397C (black).

The chosen tagging sites are all on regions of the protein that are not too flexible. There are several tags that are on the distal side of the complex to monitor PCSs on the binding interface between the complex. Additionally, there are tags that are close to the binding interface and should measure PCSs on the N-terminal or C-terminal domain, as well as the distal side of eIF4A. First, I will present results that indicate that HEAT2 was successfully tagged on the sites illustrated, and then show the results from the PCS experiments of the protein in complex with eIF4A.

Table 6.1 illustrates the suite of experiments that were planned for monitoring PCSs on both eIF4A and HEAT2. For each mutant, we have to either isotopically label eIF4A or HEAT2 and put it in complex with a deuterated version of the partner protein. For example, for the HEAT2GC mutant to see PCSs on the HEAT2 side of things we run

an experiment with $^{2}H^{15}N$-HEAT2GC in complex with ^{2}H eIF4A. The experiment has to be done in triplicate—once each for the Y^{3+}, Tm^{3+}, and Tb^{3+} bound form of $^{2}H^{15}N$-HEAT2. To see PCSs on the eIF4A side of things, the same experiments—again, in triplicate—are performed but with eIF4A isotopically labelled, and HEAT2GC deuterated. This experimental scheme resulted in 24 different samples that needed to be prepared.

eIF4A Label	HEAT2 Label	HEAT2 Mutant	Lanthanides
$^{2}H^{15}N$	^{2}H	AC	Tm
			Y
			Tb
^{2}H	$^{2}H^{15}N$	AC	Tm
			Y
			Tb
$^{2}H^{15}N$	^{2}H	EC	Tm
			Y
			Tb
^{2}H	$^{2}H^{15}N$	EC	Tm
			Y
			Tb
$^{2}H^{15}N$	^{2}H	GC	Tm
			Y
			Tb
^{2}H	$^{2}H^{15}N$	GC	Tm
			Y
			Tb
$^{2}H^{15}N$	^{2}H	LC	Tm
			Y
			Tb
^{2}H	$^{2}H^{15}N$	LC	Tm
			Y
			Tb

Table 6.1 This table catalogs all of the combinations of labelling schemes and lanthanides that were used for the experiments to gather PCS data in the eIF4A and HEAT2 complex.

6.2 PCSs on eIF4G-HEAT2 when in complex with eIF4A

In these experiments, we first had to make sure that we were able to successfully tag HEAT2. Not only did we want to tag the protein, but we hoped for the tagging efficiency

to be as high as it is impossible to separate tagged versus untagged protein in a simple manner. The lanthanide ions used were Thulium (Tm^{3+}), Terbium (Tb^{3+}), and Yttrium (Y^{3+}). The Tm^{3+} and Tb^{3+} containing tags act as the paramagnetic samples that will (hopefully) exhibit PRE and PCS effects on the protein, and the Y^{3+} containing tag is the diamagnetic reference. Tm^{3+} and Tb^{3+} are a nice pair to choose for these experiments. If a PCS is seen for a particular residue for Tm^{3+}, there should be a PCS in the opposite direction at a greater magnetic in the Tb^{3+} containing spectrum. When the two paramagnetic and diamagnetic reference spectra are overlaid, there can be diagonal lines drawn through the residues that show PCSs.

Figure 6.2 shows the results from $^{2}H^{15}N$ HEAT2 in complex with ^{2}H eIF4A. In all of the spectra presented in this section and the next, the same color scheme applies to the overlaid spectra: the Y^{3+}-tagged sample is black, the Tm^{3+}-tagged sample is cyan, and the the Tb^{3+}-tagged sample is red. HSQCs for the rest of the $^{2}H^{15}N$ HEAT2 mutants in complex with ^{2}H eIF4A—EC, GC, and LC can be seen in Figures 6.3, 6.4, and 6.5, respectively.

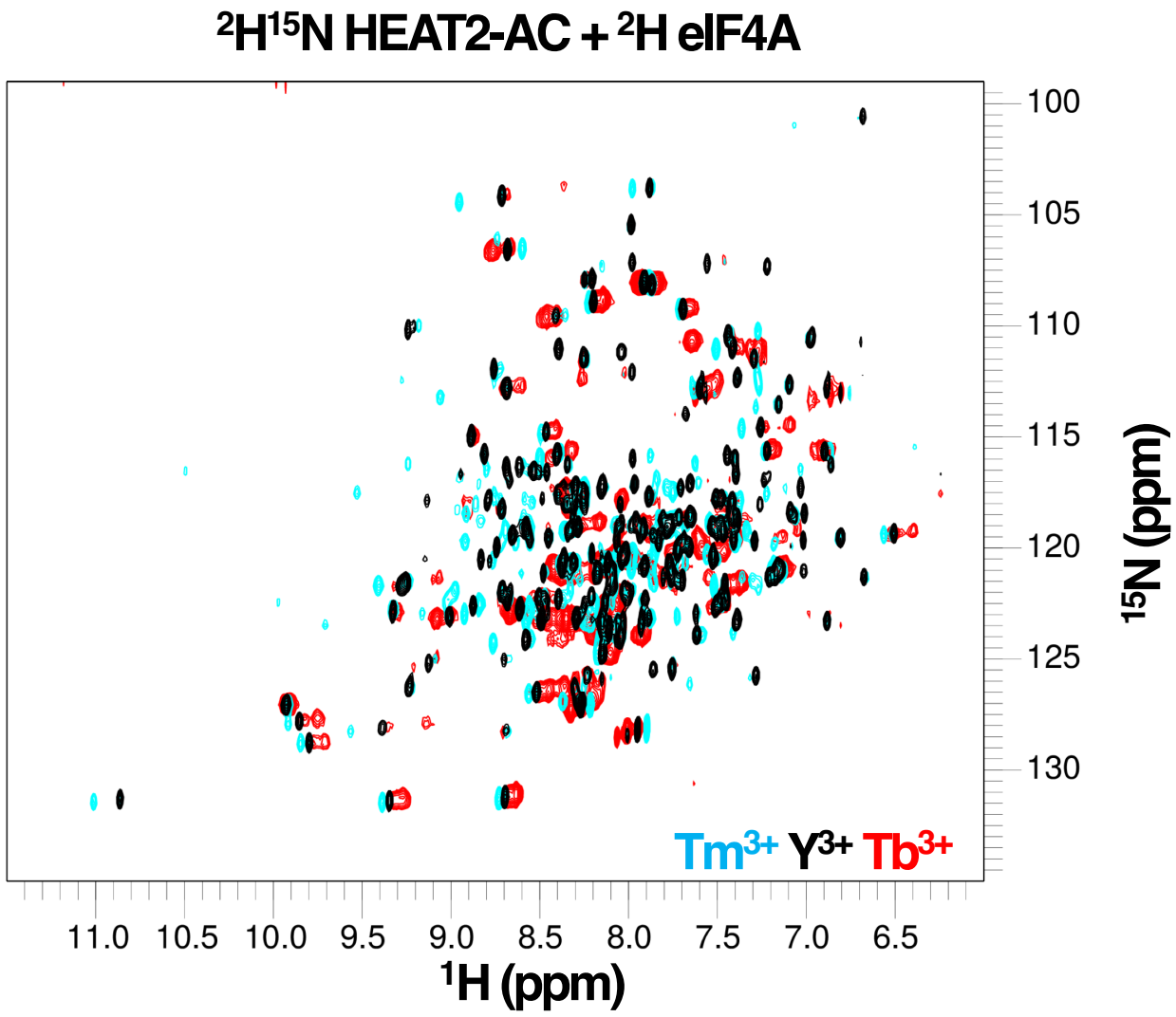


Figure 6.2 HSQC spectra for $^{2}H^{15}N$ HEAT2-AC in complex with 2H eIF4A. Illustrated above is an overlay of three different experiments with Y^{3+} in black, Tm^{3+} in cyan, and Tb^{3+} in red.

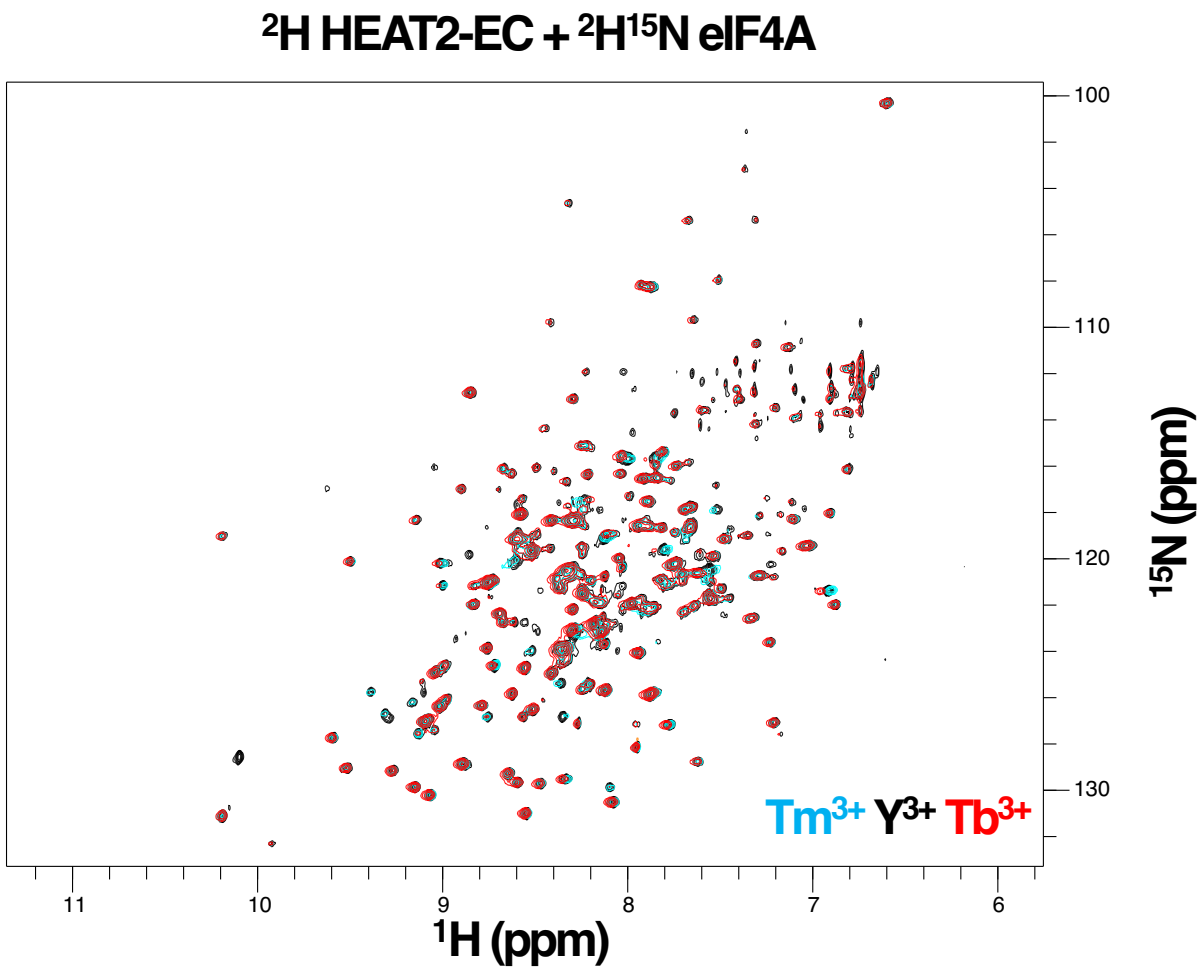


Figure 6.3 HSQC spectra for $^{2}H^{15}N$ HEAT2-EC in complex with 2H eIF4A. Illustrated above is an overlay of three different experiments with Y^{3+} in black, Tm^{3+} in cyan, and Tb^{3+} in red.

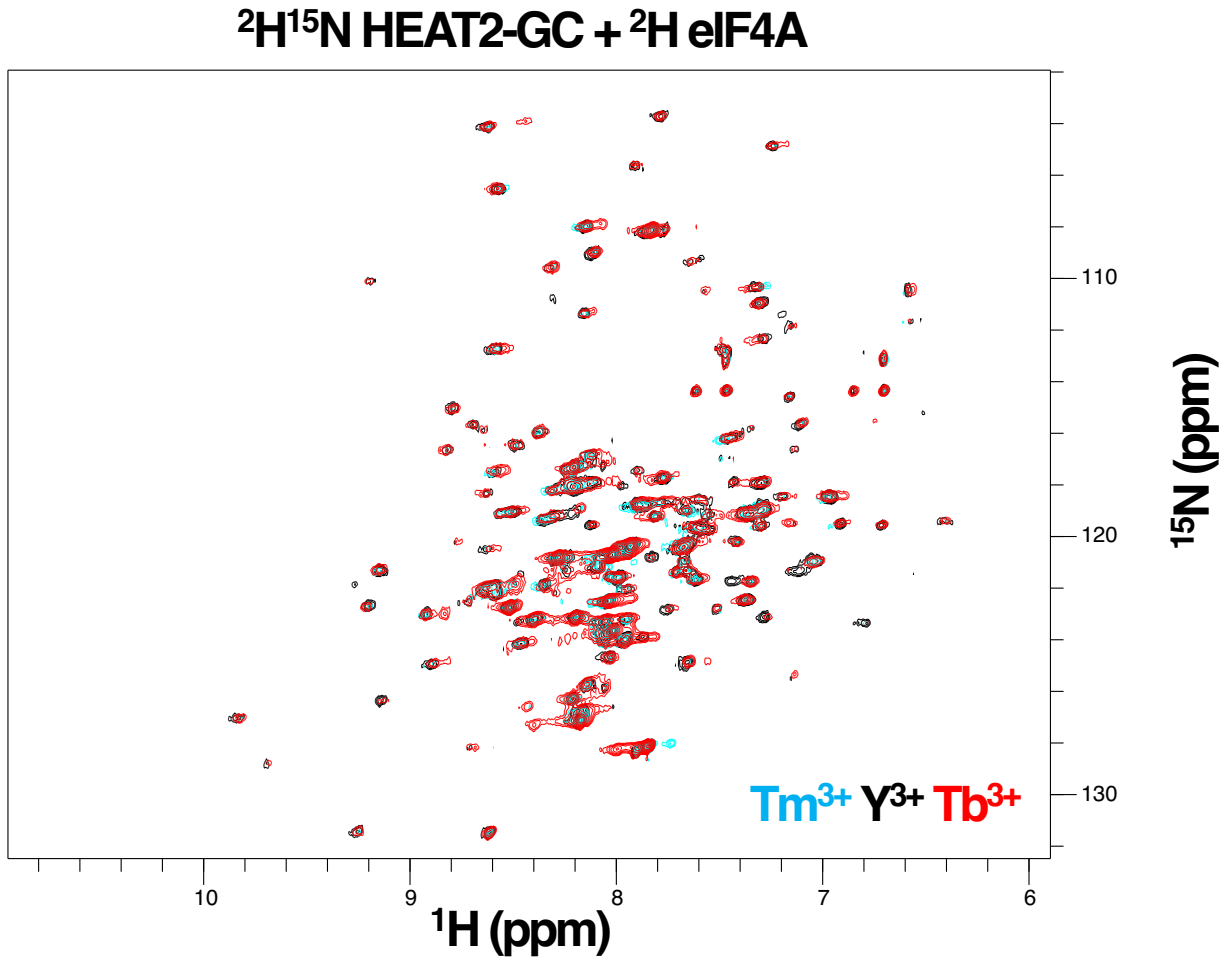


Figure 6.4 HSQC spectra for $^{2}H^{15}N$ HEAT2-GC in complex with 2H eIF4A. Illustrated above is an overlay of three different experiments with Y^{3+} in black, Tm^{3+} in cyan, and Tb^{3+} in red.

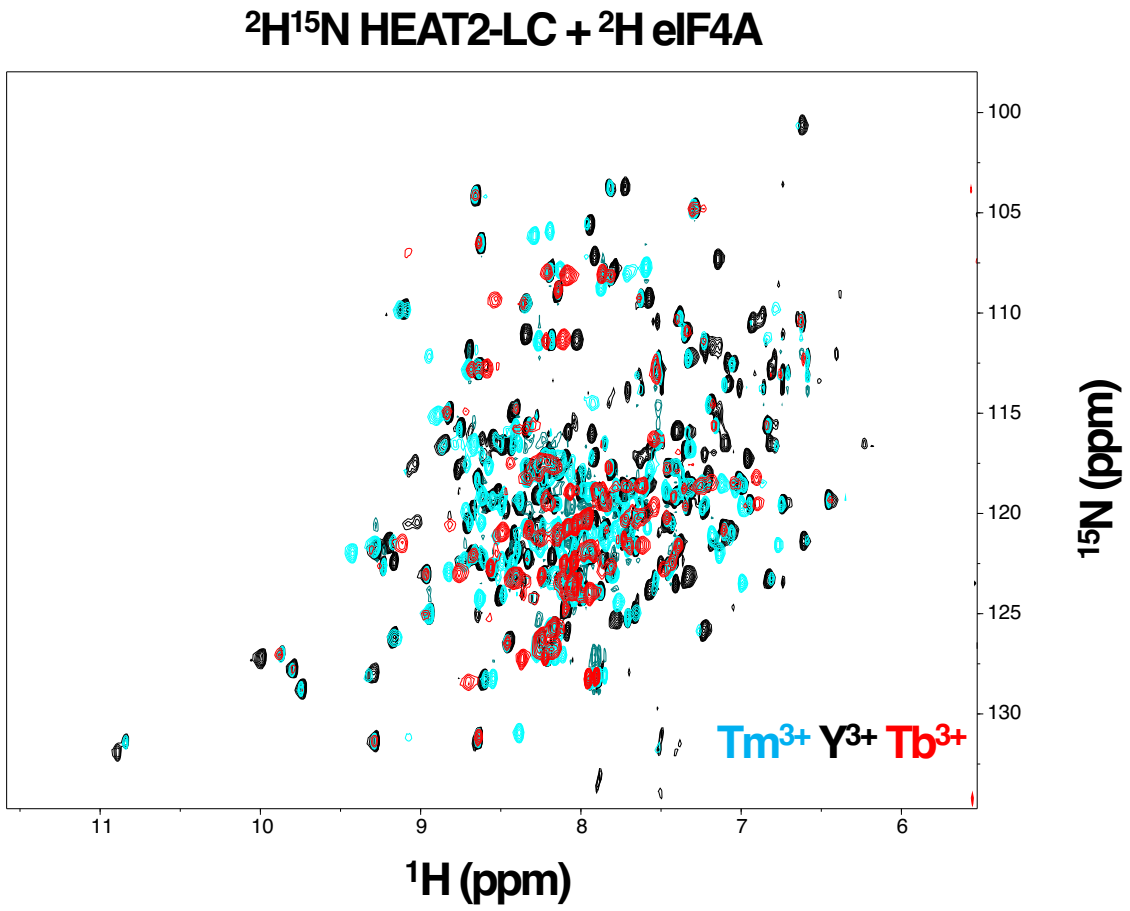


Figure 6.5 HSQC spectra for $^{2}H^{15}N$ HEAT2-LC in complex with 2H eIF4A. Illustrated above is an overlay of three different experiments with Y^{3+} in black, Tm^{3+} in cyan, and Tb^{3+} in red.

From these results, it is immediately obvious that we can monitor PCSs for HEAT2 in complex with eIF4A. In the spectra there are the characteristic diagonal shifts caused by the two paramagnetic ions relative to the diamagnetic reference. Fortunately, the shifts for Tb^{3+} are always greater in magnitude and in the opposite direction relative to the shifts for Tm^{3+}. However, it is clear that Tb^{3+} has a much larger PRE effect than Tm^{3+} from the many times there is a shift in the cyan spectrum, but no opposing shift in the red spectrum.

It is also immediately obvious that some of the tagging locations are better than others. For instance, the AC and LC mutants give much better PCSs than the GC and the EC mutants. In the spectra for the AC and the LC mutants, the shifts seen could likely be

used to calculate a ΔX-tensor. In the GC and the EC mutant spectra, there are relatively few PCSs seen, and the ones that are seen are quite small. There are mostly PRE effects in the spectra with the paramagnetic tags. This indicates that the lanthanide is likely to be too close to the protein, and is quenching the signals that we need to calculate distances.

6.3 PCSs ON eIF4A WHEN IN COMPLEX WITH eIF4G-HEAT2

The complimentary HSQC experiments were performed with isotopically labelled eIF4A in complex with a lanthanide-tagged, deuterated HEAT2 mutant. In general, the data for these experiments was of lesser quality—eIF4A just does not give good spectra like HEAT2. It is important to not forget the size of the complex, and how much this can effect the data obtained. Figures 6.7, 6.8, 6.9, and 6.10 present the HSQCs for $^{2}H^{15}N$ eIF4A in complex with the lanthanide-tagged ^{2}H HEAT2 mutants AC, EC, LC, and GC, respectively.

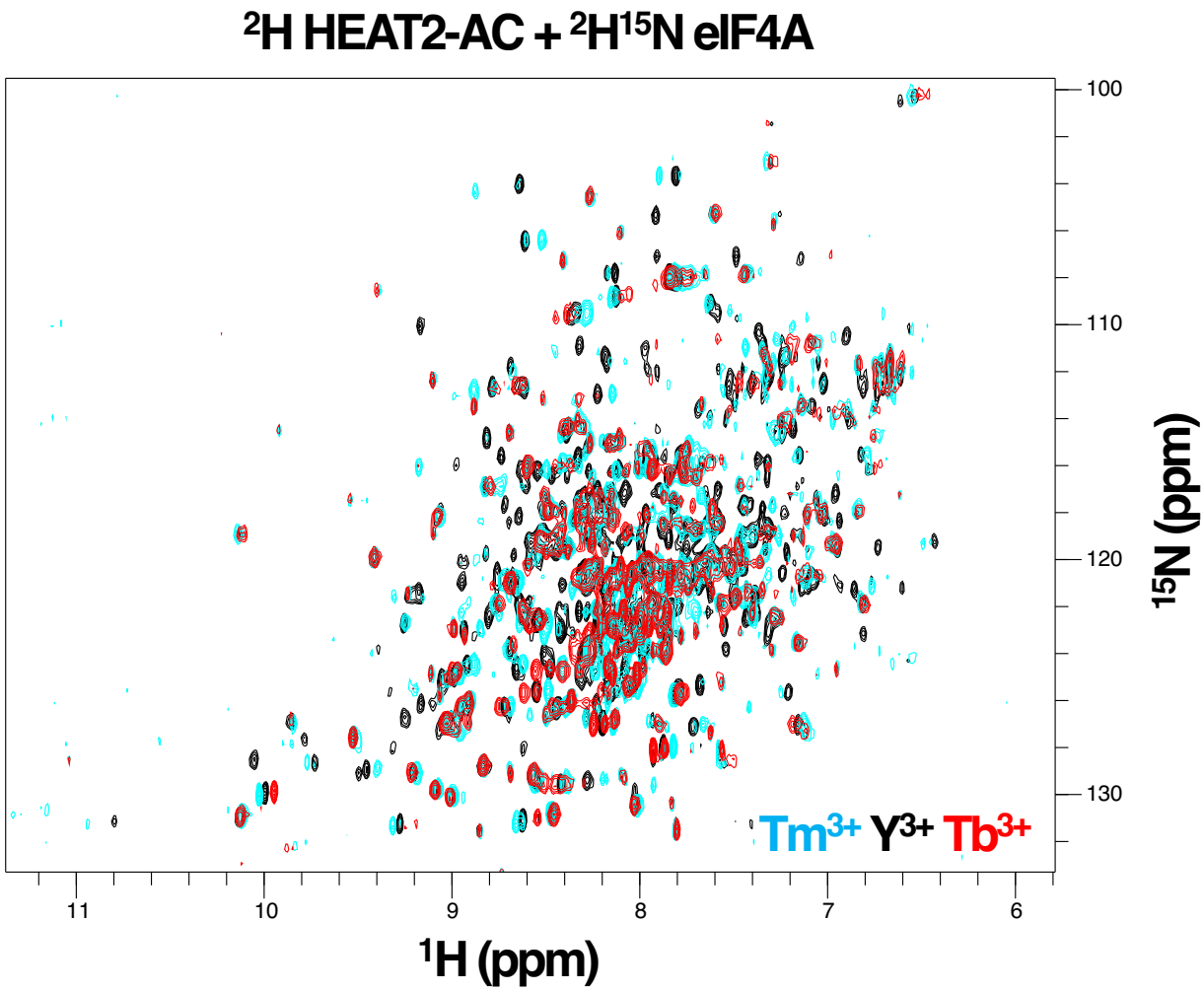


Figure 6.7 HSQC spectra for $^{2}H^{15}N$ HEAT2-AC in complex with ^{2}H eIF4A. Illustrated above is an overlay of three different experiments with Y^{3+} in black, Tm^{3+} in cyan, and Tb^{3+} in red.

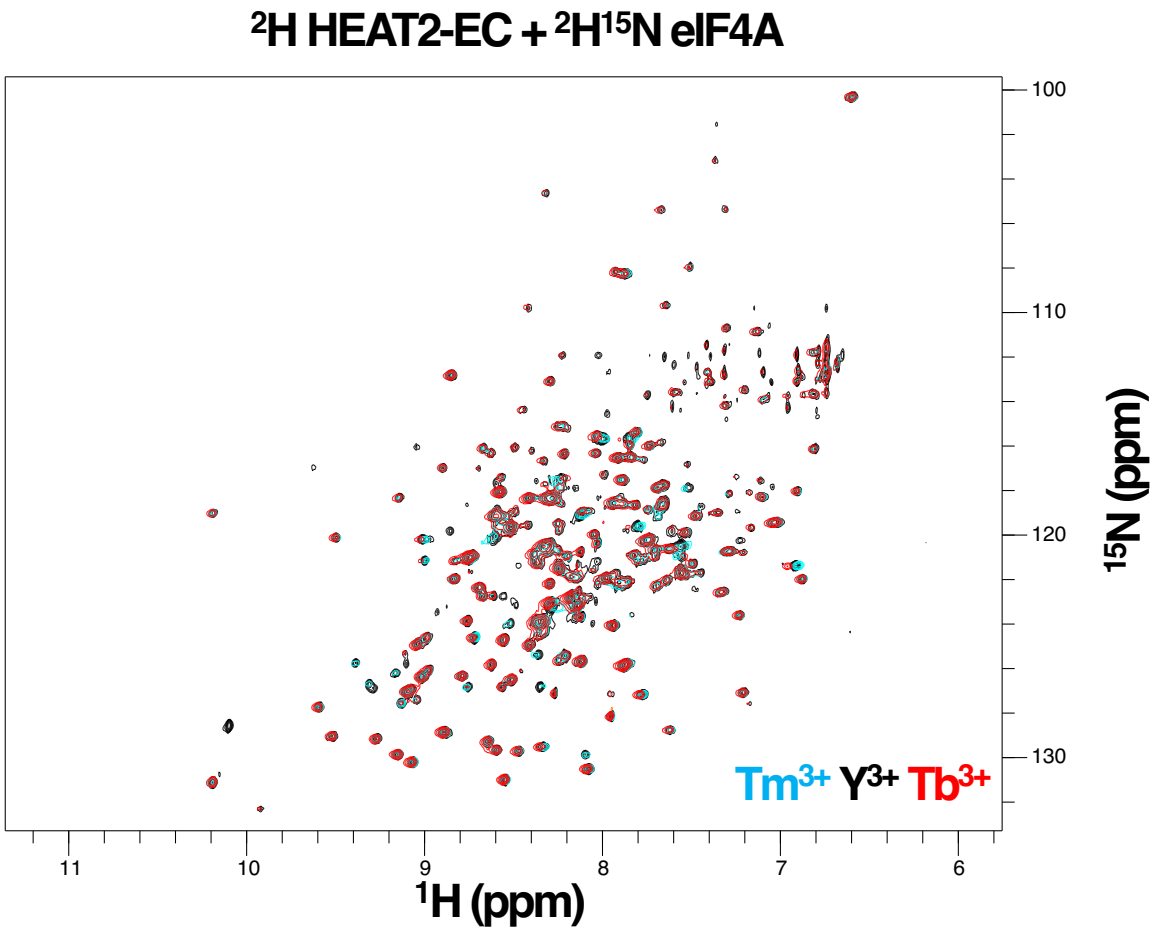


Figure 6.8 HSQC spectra for $^{2}H^{15}N$ HEAT2-EC in complex with 2H eIF4A. Illustrated above is an overlay of three different experiments with Y^{3+} in black, Tm^{3+} in cyan, and Tb^{3+} in red.

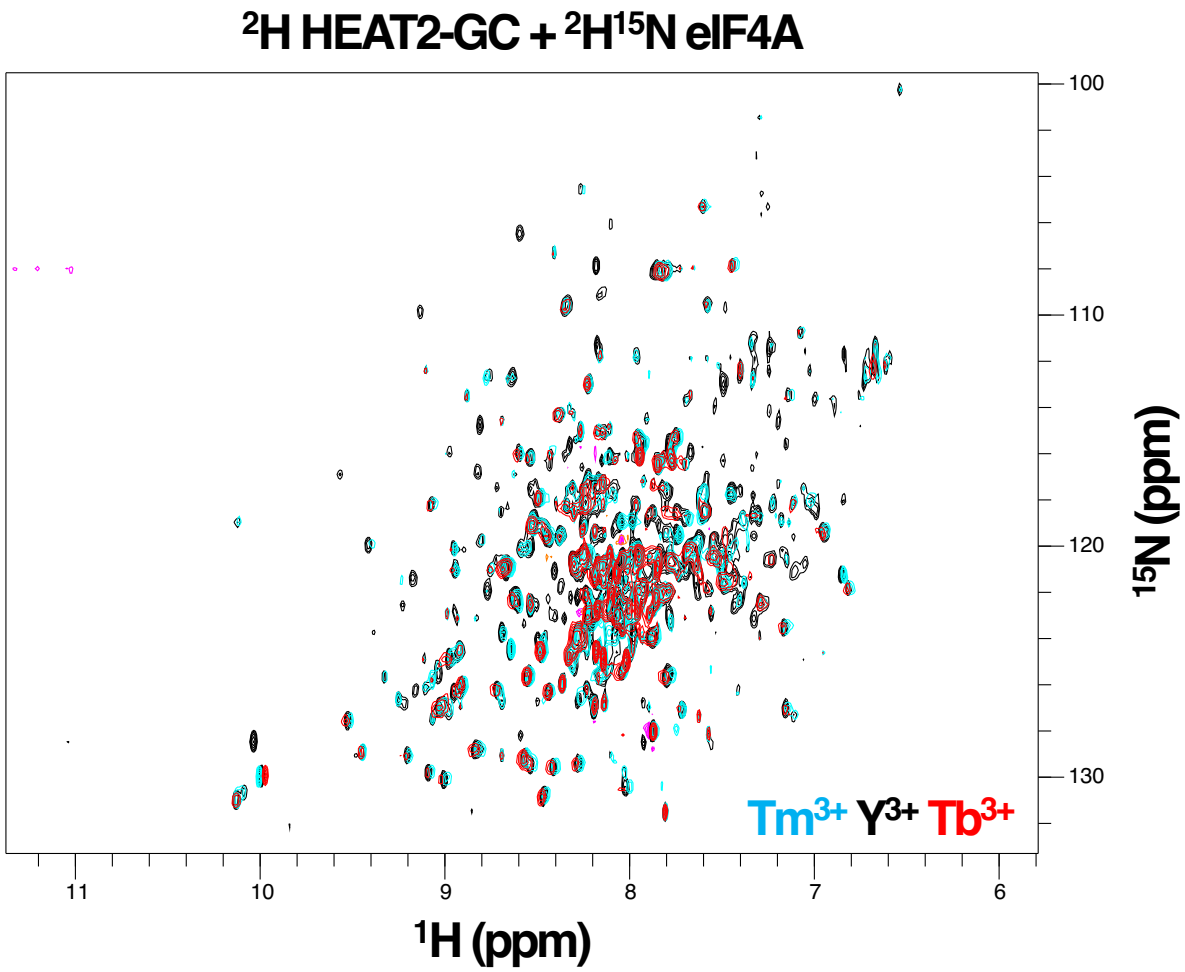


Figure 6.9 HSQC spectra for $^{2}H^{15}N$ HEAT2-GC in complex with 2H eIF4A. Illustrated above is an overlay of three different experiments with Y^{3+} in black, Tm^{3+} in cyan, and Tb^{3+} in red.

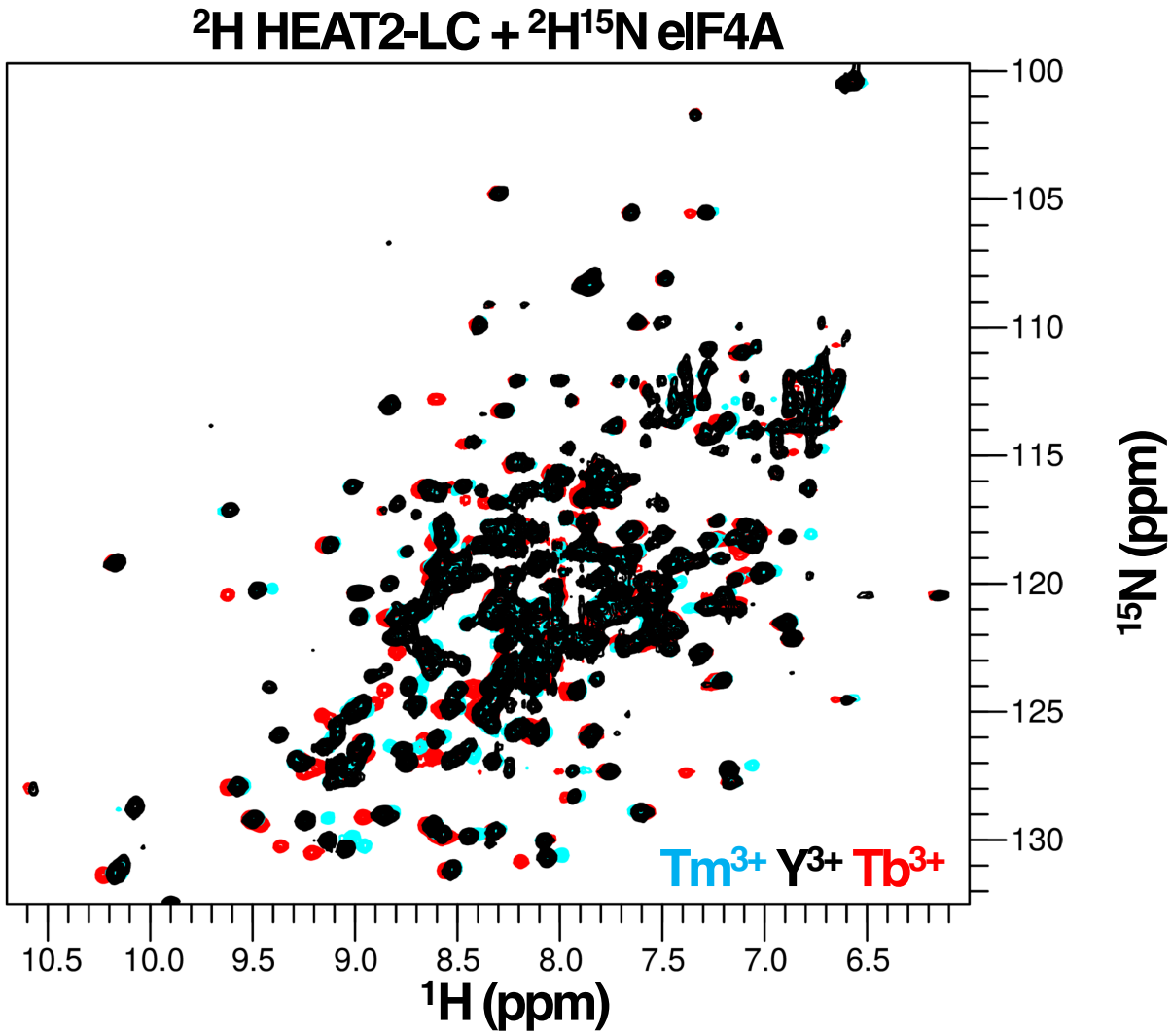


Figure 6.10 HSQC spectra for $^{2}H^{15}N$ HEAT2-LC in complex with 2H eIF4A. Illustrated above is an overlay of three different experiments with Y^{3+} in black, Tm^{3+} in cyan, and Tb^{3+} in red.

In comparison with the results from the HSQCs presented for the HEAT2 side of things, it is very clear that, overall, the magnitude of PSCs are much less. Additionally, PRE effects seem to have a much larger effect on the paramagnetic spectrum for each sample. The triplicate spectra for the EC, GC, and LC mutants give very few PCSs overall. The AC mutant, on the other hand, gives a significant number of PCSs that could potentially be used for structure calculations

7

CONCLUSIONS AND FUTURE DIRECTIONS

7.1 CONCLUDING REMARKS

In this thesis, I have presented a method to make circularized nanodiscs. I go through the method development for the circularization reaction, assembly of circularized nanodiscs, physical benefits of cNDs, and give an example of how they improve NMR experiments through precise biochemical control.

In the second half, I have presented initial experiments using paramagnetic tags to monitor PCSs on the eIF4A-eIF4G-HEAT2 complex. I have shown that it is possible to tag the HEAT2 domain, and it is robust to many perturbations. I presented results showing that it is possible to measure real PCSs on both sides of the complex—even though it is very large. I give examples that indicate the PCSs seen are real, and could potentially be used to construct a better model of the complex.

7.2 CIRCULARIZED NANODISCS

I am obviously very excited about prospect of this technology being picked up by groups that are interested in IMP structural biology. cNDs have already aided in solving the structure of the human epithethial calcium channel TRPV6 [79]. In this study, the authors used cNDs to solve structures for both the open and closed state, which allowed them to present the molecular basis for a novel gating mechanism. Another group used the cNDs to understand the dynamics of SNARE proteins from a biochemical standpoint [80]. As exciting as this technology is, there is still significant work that should be done to improve it.

Firstly, in the experiments we used evolved sortase to perform the circularization [81]. This technology is excellent; however, it was evolved to recognize the consensus sequence for LPETG. The constructs used in these experiments had a consensus sequence for LPGTG. Secondly, the purification of the cMSPs needs to be streamlined somehow. If one is planning to do NMR with the cNDs, the amount of reaction solution you have to concentrate down in the end to obtain a good yield of cMSPs is very large.

This might sound like a moot point, but spending a day concentrating two liters of protein is a bit absurd.

There were many discussions when developing the circularization method regarding ways to squander the hydrolysis product pathway. One idea involved forming nanodiscs with the circularizeable construct, performing the circularization reaction, and then disassembling the nanodiscs and purifying the cMSPs. Although I did not experiment with this, I still firmly believe that making the cMSPs in solution under certain reaction conditions will give much better results. The most obvious point is that the extra steps necessary to make and disassemble nanodisc will be tedious, waste expensive lipids, and likely lead to protein loss due to aggregation. Additionally, given the uncertainty of the dynamics and flexibility of the MSPs, I'm not entirely sure if circularization will be possible in the scenario of a nanodisc being slightly overfilled with lipids.

Another point to consider is trying to engineer better constructs for the cMSPs. The structure of MSPΔH5 was published after I had ceased work on nanodiscs in the lab [56], but it would have been exceptionally helpful for trying to design better nanodiscs for NMR spectroscopy. From Franz's paper, it is quite clear that MSPΔH5 is the only one that has decent stability out of his mutants [11]. Since two particles come together in an antiparallel fashion, there are interactions between helix 4 and helix 6, helices 2 and 3 with helix 7, and helices 8 and 9 with helix 10. Salt bridges and cation-π interaction networks stabilize the interactions between these helices. Potential mutants to try would be Δ2/3/7, Δ4/6, and Δ8/9/10—this might make for stable particles.

A consideration and major limitation of nanodiscs; however, is that if the particle becomes too small will the IMP of interest interact with the helices of the MSP? It's unclear how many lipids are necessary to create a large enough buffer between the MSP and the IMP of interest, and whether or not potential interactions can muddy results. Is there an effect for physical force on the protein? Certain IMPs, such as the mechanosensitive channel MscL, depend on physical force from the lipid bilayer. Is there enough in a nanodisc environment? Is there not enough?

Until it is possible to study the structure of IMPs within the context of the cell, there will be no membrane mimetic that will be perfect. It is up to the researcher to figure out which mimetic is most suitable for the question.

7.3 VDAC-1 IN CIRCULARIZED NANODISCS

The main reason I began work on trying to improve the preparation of nanodiscs was to make it possible to get the full assignments for VDAC-1 in nanodiscs. This obviously did not happen, and is a very safe future direction to discuss for this project. There were interesting differences in the spectra depending on one versus two VDAC-1 molecules per nanodisc, and these differences should certainly be explored further. The first step is to get complete assignments for VDAC-1 in nanodiscs. There are many resonances that are present in the HSQC, but only about 40% of the protein has been assigned. If I had the time, I would start with the smaller circularized nanodisc construct to get the assignments. I once made a 0.9 mM NMR sample of this, and it gave relatively good spectra; however, it did not have the right labeling for assignment experiments. An

additional detail to keep in mind is the use of deuterated lipids. In the lab, a graduate student was working hard on getting the structure of VDAC1 in detergent. A post-doc had the idea to try it with deuterated lipids, and this massively helped increase the relaxation times—allowing for the collection of a complete set of data.

Once the assignments are done in the smaller circularized nanodiscs, it would then be feasible to transfer some of the assignments over to a dimer VDAC-1 preparation in cMSP1D1 nanodiscs. Again, one will likely have to redo the assignment experiments and utilize deuterated lipids. From this, as well as the previous data set, it should be clear if there are any resonances that belong to side chains that are interacting in a dimer interface. This would be an interesting result as there have been several studies indicating that VDAC-1 can form dimers and oligomers, but no molecular resolution evidence for this.

Given the advances in cryo-EM, these samples might be good candidates for testing the detection and resolution limits for “smaller” molecules with the technique. Additional engineering of the cMSP construct to include a site for a gold particle to be attached might be a good idea for cryoEM studies. This way, one will be able to catalog the orientation of the MSPs with respect to the IMP—in this case VDAC-1—inside.

7.4 THE eIF4A-eIF4G-HEAT2 COMPLEX

There is much to be done for this project; however, the data collected shows that PCS can be a viable technique for sorting out the structure of the eIF4A-HEAT2 complex. Some of the mutants—particularly the AC mutant—show large PCSs that can definitely

give enough information to calculate distances from the location of the tag. Although some of the mutants give very little to no PCSs, this is a good starting point for refining the locations of future tags. From this data, we can pare down the regions that do not work for tagging on the protein.

More mutants will have to be made to gather more information since some of the tagging sites do not give very much data. It would seem that mutants on the distal side of HEAT2 give little PCSs for eIF4A, and show large PRE effects on HEAT2. Mutating residues on HEAT2 that are close to the binding interface of the complex should provide a good amount of data for the complex. Since it is very difficult to predict the orientation of the isosurfaces that the tag will have, it might be wise to test several different regions—mutating 3 or 4 residues within a particular hotspot on HEAT2.

Another alternative to collecting more data is to try different tags on the same sites. Since it is the orientation of the paramagnetic ion that governs the shifts in the spectrum, it is possible that a different tag might change the spatial location of the lanthanide relative to the protein.

Obviously, the assignments for each protein needs to be transferred. We were given the assignments of each side of the complex prior to starting the project, but due to poor labelling of the constructs, we have used a different version of HEAT2 that gives a different spectra to the data on hand. This was especially disappointing after collecting such a large quantity of data. However, this is not such a major setback. If assignment experiments were possible with the slightly different construct of HEAT2, they should be possible with the construct of HEAT2 that we used in the PCS experiments.

Another way to move the project forward is to try to tag eIF4A. Wanna has made attempts on this, and based on our discussions it is not a trivial pursuit. An experiment that has not been tried; however, is to look at sequence co-evolution and rather than just mutating the cysteines away, mutate some of the residues that surround the cysteines to residues that have evolved eIF4A proteins in different species.

7

MATERIALS AND METHODS

7.1 PROTEIN EXPRESSION AND PURIFICATION

7.1.1 MSP1D1, MSPΔH5, cMSP1D1, and cMSPΔH5

MSP1D1, MSPΔH5, cMSP1D1, and cMSPΔH5 (all in pET-28b) containing a TEV-cleavable N-terminal His_6 tag and a C-terminal sortase-cleavable His_6 tag (for the latter two) were transformed into BL21-Gold (DE3) competent *E. coli* cells (Agilent). Liters of cell cultures were grown at 37°C with 200 rpm of shaking in Luria broth (LB) medium supplemented with 50 μg/ml Kanamycin. Expression was induced with 1 mM IPTG at an

OD_{600} of 0.6 for 3 hours at 37°C. Cells were harvested by centrifugation (7000x*g*, 15 minutes, 4°C) and cell pellets were stored at -80°C.

Cell pellets expressing MSP1D1 or MSPΔH5 were resuspended in lysis buffer (50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 1% Triton X-100) and lysed by sonication on ice. Lysate was centrifuged (35,000x*g*, 50 minutes, 4°C) and the supernatant was filtered and loaded onto Ni^{2+}-NTA column. The column was washed with lysis buffer then with buffer A (50 mM Tris, pH 8.0, 500 mM NaCl). To recover additional protein from the insoluble fractions, the pellets recovered from lysate centrifugation were dissolved in denaturing buffer (50 mM Tris, pH 8.0, 500 mM NaCl, 6M guanidine hydrochloride), centrifuged (35,000x*g*, 50 minutes, 4°C), and the supernatant was applied to the same Ni^{2+}-NTA column containing bound protein from the soluble fraction. The column was washed with denaturing buffer, and cMSP1D1 and cMSPΔH5 were refolded on-column with 10 column volumes (CV) buffer A. Resin was then washed with 10 CV of the following: buffer A + 1% Triton X-100, buffer A + 50 mM sodium cholate, buffer A, and buffer A + 20 mM imidazole. Proteins were eluted with buffer A + 500 mM imidazole, TEV (His_6-tagged; produced in-house) was added to cleave the N-terminal His_6 tag, and the samples were dialyzed against 50 mM Tris-HCl, pH 8.0, 100 mM NaCl, 1 mM EDTA, 2 mM DTT at 4°C for 16 hours. Protein was concentrated using centricon concentrators (10 kDa MW cutoff, Millipore), and the TEV and uncleaved MSPs were removed with reverse Ni^{2+}-NTA chromatography in 50 mM Tris-HCl pH 8.0, 100 mM NaCl, and 1 mM EDTA.

Cell pellets expressing cMSP1D1 or cMSPΔH5 were resuspended in lysis buffer (50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 1% Triton X-100) and lysed by sonication on ice.

Lysate was centrifuged (35,000x*g*, 50 minutes, 4°C) and the supernatant was filtered and loaded onto Ni^{2+}-NTA column. The column was washed with lysis buffer then with buffer A (50 mM Tris, pH 8.0, 500 mM NaCl). To recover additional protein from the insoluble fractions, the pellets recovered from lysate centrifugation were dissolved in denaturing buffer (50 mM Tris, pH 8.0, 500 mM NaCl, 6M guanidine hydrochloride), centrifuged (35,000x*g*, 50 minutes, 4°C), and the supernatant was applied to the same Ni^{2+}-NTA column containing bound protein from the soluble fraction. The column was washed with denaturing buffer, and NW9 and NW11 were refolded on-column with 10 column volumes (CV) buffer A. Resin was then washed with 10 CV of the following: buffer A + 1% Triton X-100, buffer A + 50 mM sodium cholate, buffer A, and buffer A + 20 mM imidazole. Proteins were eluted with buffer A + 500 mM imidazole, TEV (His_6-tagged; produced in-house) was added to cleave the N-terminal His_6 tag, and the samples were dialyzed against 50 mM Tris-HCl, pH 8.0, 100 mM NaCl, 1 mM EDTA, 2 mM DTT at 4°C for 16 hours. NW9 and NW11 (still containing a C-terminal His_6 tag) were exchanged into nanodisc-assembly buffer (50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 0.02% NaN_3) using centricon concentrators (10 kDa MW cutoff, Millipore).

7.1.2 VDAC-1

VDAC-1 in a pET-21a vector was transformed into BL21-Gold (DE3) competent *E. coli* cells (Agilent). Liters of cell cultures were grown at 37°C with 200 rpm of shaking in Luria broth (LB) medium supplemented with 50 µg/ml Ampicillin. Expression was induced with 1 mM IPTG at an OD_{600} of 0.6 at 37°C and cultures were shaken at 200 rpm for 16-

18 hours. Cells were harvested by centrifugation (7000x*g*, 15 minutes, 4°C) and cell pellets were stored at -80°C.

VDAC was purified from inclusion bodies under denaturing conditions on a Ni-agarose resin (Qiagen) and eluted in Ni-elution buffer (6 M GdHCl, 100 mM NaH_2PO_4, 10 mM Tris-HCl, pH 8.0, and 250 mM imidazole). Cation exchange under denaturing conditions in 8 M urea was necessary as a next step in purification. Cation-exchange chromatography required preparation of the denatured VDAC sample in a low-salt-containing buffer prior to loading on the cation-exchange column to remove any salt or contaminating materials which would inhibit binding of VDAC to the column. Purified VDAC in Ni-elution buffer was precipitated by dialysis against 4 L of 2 mM EDTA in a 12000 MWCO dialysis membrane. Precipitated VDAC was isolated by centrifugation at 18000 rpm in an SS-34 rotor. VDAC was resuspended in cation-exchange start buffer (8 M urea, 25 mM sodium phosphate, pH 6.5, 5 mM DTT) and sonicated to ensure efficient solubilization. VDAC in cation-exchange start buffer was loaded onto a 5 mL HiTrap SP FF cation-exchange column (GE Healthcare) at ~2 mL/min. VDAC was eluted from the cation-exchange column in cation-exchange start buffer plus 100 mM NaCl. Further purification of the denatured VDAC sample was accomplished by gel filtration chromatography under denaturing conditions in denaturing gel filtration buffer (6 M guanidine hydrochloride, 100 mM sodium phosphate, pH 7.0, 100 mM NaCl, 1 mM EDTA, 5 mM DTT) at a flow rate of 0.5 mL/min.

Purified VDAC in denaturing gel filtration buffer was refolded by dropwise dilution at 4 °C with stirring into refold buffer [25 mM sodium phosphate, pH 7.0, 100 mM NaCl, 1

mM EDTA, 5 mM DTT, 1% lauryldimethylamine oxide (LDAO)] to yield a final protein:detergent micelle ratio of ~1:20 (assuming an LDAO aggregation number of 76). After overnight stirring at 4 °C the refolded VDAC sample was then dialyzed against phosphate buffer to remove guanidine hydrochloride. Further purification on a Superdex 200 gel filtration column preequilibrated with LDAO GFC (gel filtration chromatography) buffer (25 mM sodium phosphate, pH 7.0, 1 mM EDTA, 5 mM DTT, 0.1% LDAO) was employed to remove misfolded or aggregated VDAC and trace impurities.

7.1.2 eIF4A, eIF4G-HEAT2, and eIF4G-HEAT2 mutants

eIF4A, eIF4G-HEAT2, and the eIF4G-HEAT2 mutants (all in pET28a) containing a TEV-cleavable N-terminal His_6 tag were transformed into BL21-Gold (DE3) competent *E. coli* cells (Agilent). Liters of cell cultures were grown at 37°C with 200 rpm of shaking in Luria broth (LB) medium supplemented with 50 µg/ml Ampicillin. Expression was induced with 1 mM IPTG at an OD_{600} of 0.6 at 37°C. Cultures were cooled to 18°C, and shaken at 200 rpm for 16-18 hours. Cells were harvested by centrifugation (7000x*g*, 15 minutes, 4°C) and cell pellets were stored at -80°C.

7.2 NANODISC CIRCULARIZATION

A 50 mL reaction was prepared with 10 µM cMSPs and 5 µM (final concentrations) of sortase in 300 mM Tris-HCl, pH 6.5, 150 mM NaCl, and 10 mM $CaCl_2$. The reaction was incubated at 4°C for 16-10 hours with gentle shaking on a rotating platform. A covalent sortase inhibitor AAEK2 was added to a concentration of 500 µM, and the

solution was incubated further for 30 minutes at room temperature with gentle shaking. Proteins that did not undergo circularization were removed by binding to Ni^{2+}-NTA column. Circularized MSPs (cMSPs) were further purified by size-exclusion chromatography (Superdex 75 16/60) equilibrated in buffer containing 20 mM Tris, pH 7.5, 500 mM NaCl plus 50 mM sodium cholate or 1 mM DDM. Purified protein was exchanged into buffer A_{ix} (20 mM Tris, pH 8.2, 1 mM DDM) using centricon concentrators (10 kDa MW cutoff, Millipore) and then was applied to a Resource Q column equilibrated with the same buffer (buffer A_{ix}). A linear salt gradient from 0-60 % buffer B_{ix} (20 mM Tris, pH 8.2, 1 mM DDM, 1M NaCl) was applied. Circularized proteins were eluted around 150-200 mM NaCl.

7.3 RECONSTITUTION OF NANODISCS

cMSPs:lipid ratios of 1:60 and 1:75 were used for cMSPΔH5 and cMSP1D1, respectively. Lipids (DMPC:DMPG 3:1, solubilized in sodium cholate) and cMSPs were incubated on ice for 1 hour. After incubation, sodium cholate was removed by the addition of Bio-beads SM-2 (Bio-Rad) and incubation on ice for 1 hour followed by overnight incubation at 4°C. The nanodisc preparations were filtered through 0.22 µm nitrocellulose-filter tubes to remove the Bio-beads. The nanodisc preparations were further purified by size-exclusion chromatography while monitoring the absorbance at 280 nm on a Superdex 200 10×300 column (for cMSPΔH5 and cMSP1D1 nanodiscs) equilibrated in 20 mM Tris-HCl, pH 7.5, 100 mM NaCl, 0.5 mM EDTA. Fractions

corresponding to the size of each nanodisc were collected and concentrated. The purity of nanodisc preparations was checked using SDS-PAGE.

7.4 RECONSTITUTION OF VDAC-1 IN NANODISCS

To assemble monomeric VDAC-1 into DMPC/DMPG nanodiscs for NMR experiments, 25 μM of VDAC-1, 175 μM of cMSPΔH5, and 9.5 mM lipids (DMPC:DMPG 3:1) were incubated at room temperature for 1 hour. To assemble dimeric VDAC-1 into DMPC/DMPG nanodiscs for NMR experiments, 79 μM of VDAC-1, 75 μM of cMSP1D1, and 3.75 mM lipids (DMPC:DMPG 3:1) were incubated at room temperature for 1 hour. After incubation, detergents were removed by the addition of Bio-beads SM-2 (Bio-Rad) and incubation on ice for 1 hour followed by overnight incubation at room temperature (DMPC/DMPG nanodiscs). The disc preparation was filtered through 0.22 μm nitrocellulose-filter tubes to remove the Bio-beads. To remove the VDAC-free nanodiscs, the sample was mixed with Ni^{2+}-NTA resin for 1 hour at 4°C. The resin bed volume was equal to the assembly mixture. The resin was washed with buffer E (20 mM Tris-HCl, pH 8.0, 0.1 M NaCl, 20 mM imidazole). Nanodiscs containing VDAC were eluted with buffer E containing 0.5 M imidazole. The nanodisc preparation was further purified by size-exclusion chromatography while monitoring the absorbance at 280 nm on a Superdex 200 10 × 300 or a 16/60 Superdex 200 prep grade columns (GE Healthcare) equilibrated in buffer F (20 mM sodium phosphate, 50 mM NaCl, 5 mM DTT, 1 mM EDTA, pH 7.0). Fractions corresponding to the size of the VDAC-nanodisc complex were collected and concentrated. The purity of VDAC-containing nanodiscs was checked using SDS-PAGE.

7.5 TAGGING OF HEAT2

Purified HEAT2 protein was tagged with C1-tag with a 1:3 ratio of tag to protein. Normally, a solution of 3 mL of 50 μM of protein in 20 mM Na-P pH 7.4 and NaCl 100 mM and 150 μM of C1-tag precharged with a lanthanide ion was incubated at room temperature with gentle shaking overnight. The excess tag was removed by using spin-columns with a MWCO of 10 kDa and diluting with the NMR buffer for experiments.

7.6 CIRCULAR DICHROISM

Protein samples were prepared to be between 5 and 10 μM for CD experiments. A Jasco J-1500 Circular Dichroism Spectrophotometer coupled to a peltier device was used for the experiments. Thermal melting curves were produced by measuring at 222 nm and 215 nm every 1°C for a range of 20°C to 100°C. Normally, the reverse temperature ramp was collected, but this data was not presented.

7.7 NEGATIVE-STAIN ELECTRON MICROSCOPY

Samples were prepared by conventional negative staining as described previously[24]. Briefly, 3.5 μl of nanodisc samples were adsorbed to carbon-coated copper grids and stained with 0.75% (w/v) uranyl formate. All EM images (except images for VDAC in cNW9, supplementary figure 5) were collected with a Philips CM10 electron microscope (FEI) equipped with a tungsten filament and operated at an acceleration

voltage of 100 kV. Images were recorded with a Gatan 1 K × 1 K CCD camera (Gatan, Inc., Pleasanton, CA, USA). 2D class averages were computed by EMAN2[25].

7.8 NMR SPECTROSCOPY

[*U*-^{2}H,^{15}N] labeled VDAC-1 in DMPC:DMPG (3:1) lipid nanodiscs (611 μM and 500 μM of VDAC1 in cMSPΔH5 and cMSP1D1 nanodiscs, respectively) were prepared as described above in NMR buffer (20 mM $NaPO_4$, 50 mM NaCl, 5 mM DTT, 1 mM EDTA, pH 7.0, 6% D_2O). ^{15}N-TROSY HSQC data were collected at 45°C on a Bruker 800 spectrometer equipped with TXO cryogenic probe. Data for cMSPΔH5 and cMSP1D1 nanodiscs were acquired with 48 and 96 scans respectively, and 128 complex points in the ^{15}N-indirect dimension.

150 – 250 uM samples of either ^{2}H^{15}N HEAT2 mutant and ^{2}H eIF4A or ^{2}H HEAT2 mutant and ^{2}H^{15}N eIF4A were prepared as described above in NMR buffer (20mM Tris pH 8). ^{15}N-TROSY HSQC data were collected at 25°C on either a Varian 600 or 700 MHz spectrometer equipped with a TXO cryogenic probe.

A

APPENDIX A: "DEFINED AND STABLE COVALENTLY CIRCULARIZED NANODISCS FOR STUDYING MEMBRANE PROTEINS AND VIRAL ENTRY"

A.1 ABSTRACT

We engineered covalently circularized nanodiscs (cNDs) with enhanced stability, defined diameter sizes and tunable shapes. The cNDs enhanced NMR spectral quality for both beta barrel membrane proteins such as VDAC-1, and alpha helical

membrane proteins such as a signaling competent GPCR NTR1 in cND nanodisc. In addition, we utilized cNDs to visualize how simple non-enveloped viruses translocate their genomes across membranes to initiate infection.

A.1 MAIN TEXT

Phospholipid bilayer nanodiscs provide a detergent-free lipid bilayer model, enabling biochemical and biophysical characterization of membrane proteins in a physiologically relevant environment[1]. A traditional nanodisc is composed of a nanometer-sized phospholipid bilayer patch encircled by two α-helical, amphipathic membrane scaffold proteins (MSPs)[2, 3]. MSPs are truncated forms of apolipoprotein A1 (apoA1), which is the major protein component of high-density lipoprotein. To date, however, the utility of this system for structural studies has been limited by the heterogeneity of size and the number of membrane proteins enclosed, and only small nanodiscs could be constructed with the currently available protein scaffolds[4-7]. To resolve these problems we developed three different methods to covalently link the N and C termini of newly engineered variants based on apoA1, and produced nanodiscs with a large range of discrete sizes and defined geometric shapes. The protein constructs we used contain the consensus sequence recognized by sortase A (LPGTG) near the C terminus and a single glycine residue at the N terminus (**Fig. 1a**). These two sites are sufficient to ensure covalent linkage between the N and C termini of a protein [8] while still conserving the function to form nanodiscs.

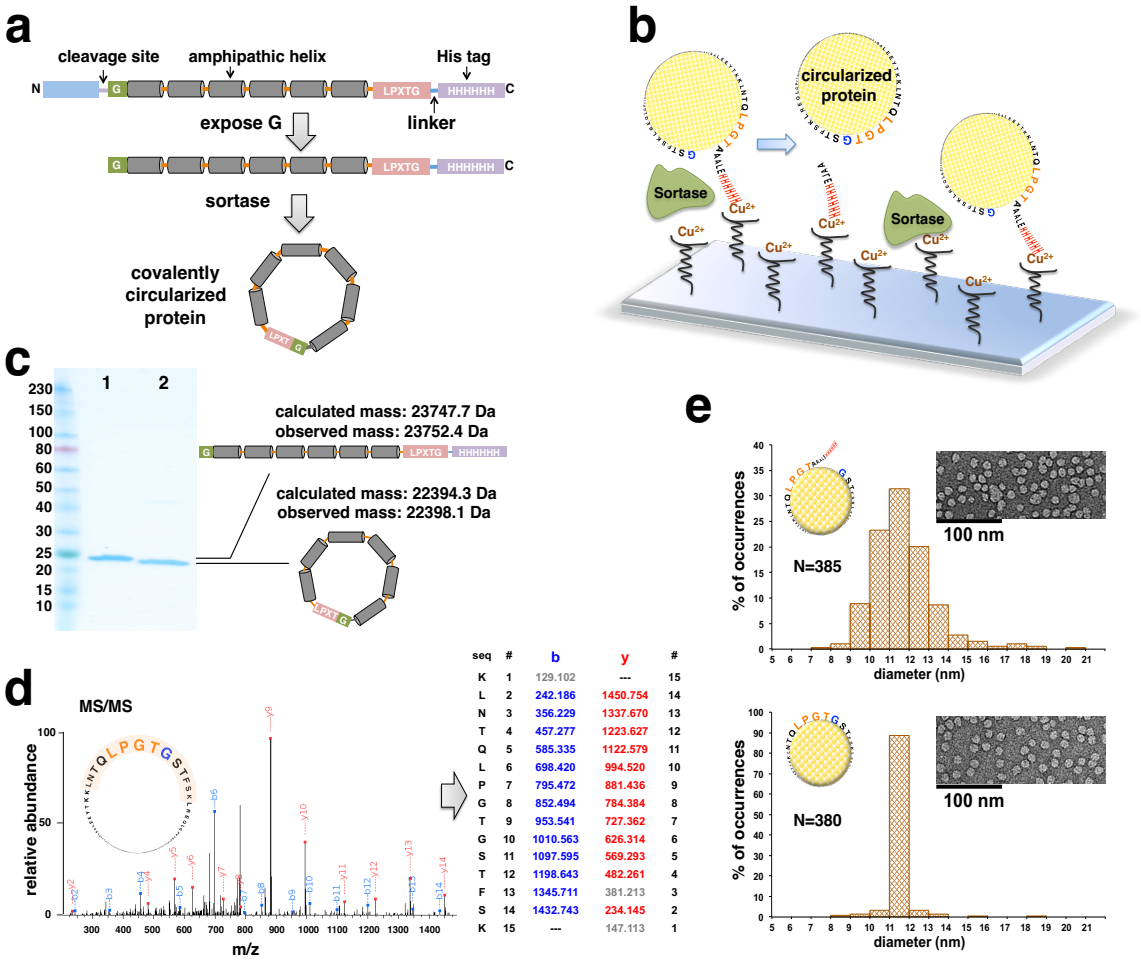


Figure A.1 Producing covalently circularized NW11 and nanodiscs. (**a**) A general outline of the constructs that are used for making covalently circularized nanodiscs. (**b**) Outline of the procedure for creating circularized proteins over Cu^{2+} chip. **(c)** SDS-PAGE analysis of NW11 before (lane 1) and after (lane 2) circularization. **(d)** MS/MS spectrum of a tryptic peptide of cNW11 confirming the ligation of the N-terminal residues (GSTFSK) to the C-terminal LPGTG motif. The b and y ions that were identified in the MS/MS spectrum are highlighted in blue and red. **(e)** Diameter distribution for nanodiscs made using non-circularized NW11 (top) and circularized NW11 (bottom) and representative negative-stain EM images.

First, we used our NW11 construct, which assembles an 11 nm nanodisc, to optimize the circularization over a Cu^{+2} chip (**Fig. 1b**). In this scheme, the Cu^{2+} is saturated with un-circularized NW11 protein prior to evolved sortase[9] addition. Upon successful completion, the circularized NW11 (cNW11) is liberated to the solution and can be further purified via nickel affinity chromatography. Immobilizing NW11 on the Cu^{2+}

chip for circularization reduces the chances for head-to-tail linkage of two neighboring NW11 molecules and also offers a quick reaction time. Reaction completion was confirmed by SDS-PAGE gel shift (**Fig. 1c**), and reaction fidelity was confirmed via tandem mass spectrometry (MS/MS) (**Fig. 1d**). Next, we tested whether the final circularized product was still capable of assembling nanodiscs. Indeed, cNW11 assembled nanodiscs, and the acquired electron microscopy (EM) images revealed uniformly sized nanodiscs. Nearly 89% of the nanodiscs were found to have a diameter between 11-12 nm as opposed to 32% for nanodiscs assembled with the linear counterpart (**Fig. 1e**). Even though circularization over a Cu^{+2} chip usually results in a very clean final product, the approach is limited to small-scale production of circularized protein.

In order to scale up the production of cNW11, we performed the circularization reaction over nickel beads (**Fig. 2a).** SDS-PAGE analysis showed that cNW11 was produced as a mixture with higher molecular weight species that possibly arose from head-to-tail ligation of neighboring NW11 monomers followed by circularization (**Fig. 2a, lane 1**). We noticed that adding lipids to the already immobilized NW11 made the intramolecular circularization the dominant pathway, and higher molecular species were only observed in trace quantities (**Fig. 2a, lane 2**).

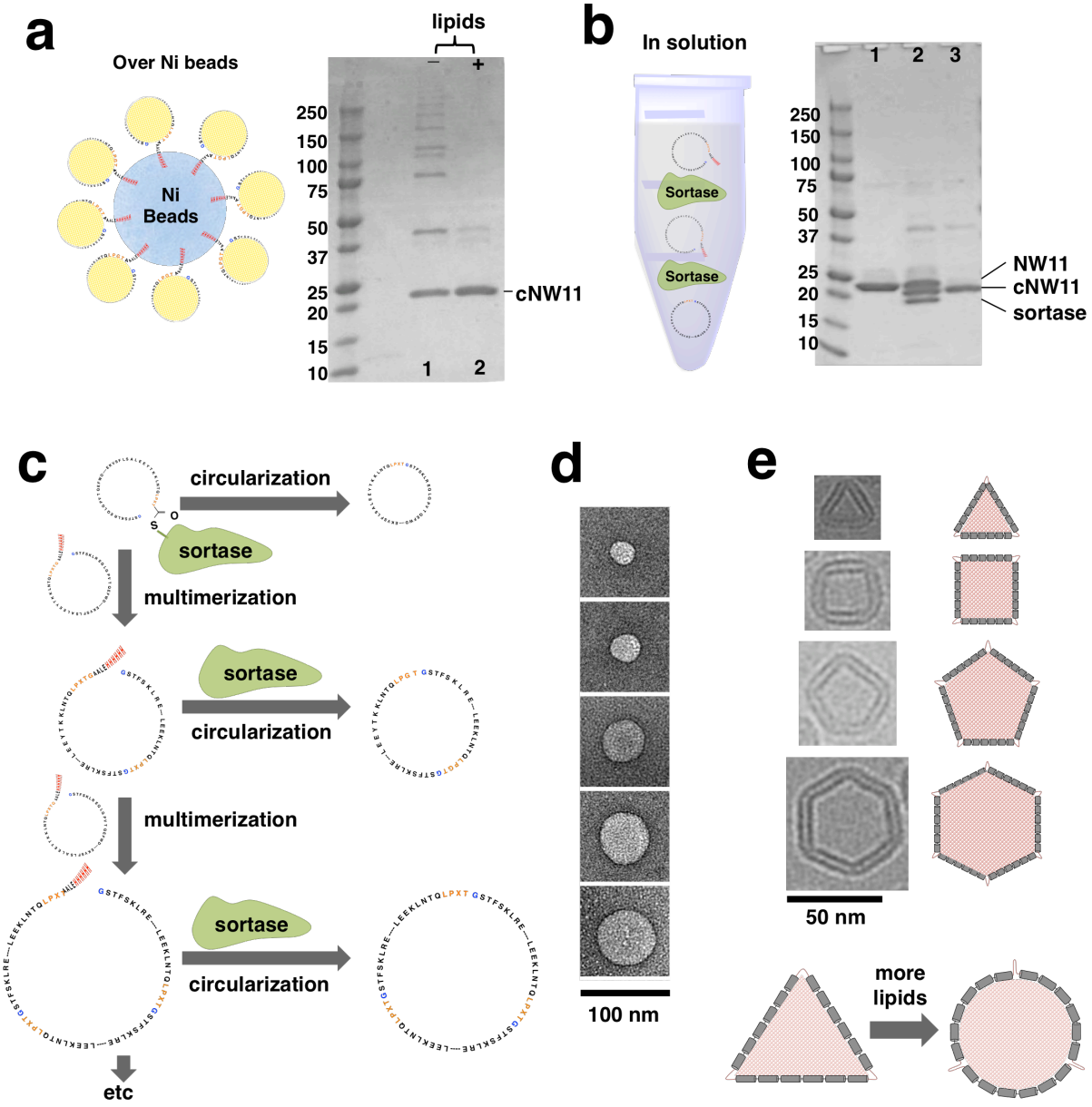

a
Over Ni beads
Ni Beads
lipids
–
+
250
150
100
75
50
37
25
20
15
10
cNW11
1
2
b
In solution
Sortase
Sortase
1
2
3
NW11
cNW11
sortase
c
circularization
sortase
multimerization
sortase
circularization
multimerization
sortase
circularization
etc
d
100 nm
e
50 nm
more lipids
f
NW30
cNW30
50 nm
g
NW50
cNW50
100 nm
cryoEM
50 nm
50 nm

Figure 2. Producing covalently circularized nanodiscs with defined sizes and shapes. (**a**) SDS-PAGE gel showing the final products after circularization over Ni beads with or without lipids. (**b**) Adding evolved sortase to diluted NW11 solution results mainly in NW11 circularization. (**c**) Adding evolved sortase to concentrated NW11 solution leads to multimerization followed by circularization. (**d**) Negative-stain EM images showing large nanodiscs made using oligomeric circularized NW11. (**e**) Cryo-EM images of individual nanodiscs with different shapes. A triangle-shaped nanodisc is made using 3 copies of NW11 covalently linked together and circularized. Adding more lipids can change the shape back to a circular disc. Right: potential molecular arrangements of NW11 molecules around the differently shaped nanodiscs. (**f**) SDS-PAGE analysis of NW30 before and after circularization. The circularized NW30 migrates slower than the linear form. Right: negative-stain EM analysis of the nanodiscs made using circularized NW30 (cNW30) shows the formation of ~15 nm nanodiscs. (**g**) SDS-PAGE analysis of NW50 before and after circularization. Right: negative-stain EM analysis of the nanodiscs made using circularized NW50 (cNW50) shows the formation of ~50 nm nanodiscs. Bottom: Negative stain and cryo-EM images for individual nanodiscs showing top and side views.

To increase the yield of cNW11, we performed the circularization reaction in solution (**Fig. 2b).** Evolved sortase was added to a dilute NW11 solution ([NW11] < 15 uM) to suppress linking two or more copies of NW11. The reaction was quenched with a covalent sortase inhibitor AAEK2[10], and cNW11 was purified via reverse nickel affinity chromatography. Reaction completion was readily confirmed by SDS-PAGE analysis (**Fig. 2b, lane 3**). With this method, we are able to produce mg quantities of cNW11 that is >95% monomeric. Also, we created an array of higher molecular weight circularized species by adding sortase to a concentrated NW11 solution (>100 □M). Indeed we produced larger nanodiscs up to 80 nm in diameter (**Fig. 2c, d and Supplementary Fig. 1**).

We used circular dichroism (CD) spectroscopy to assess the effect of circularization on the thermal stability of lipid-free and lipid-bound cNW11. We found that cNW11 has increased thermostability with an apparent midpoint melting temperature (T_m) of 65.4°C compared to 53.8°C for MSP1D1. Similarly, Lipid-bound cNW11 demonstrated better stability compared to lipid-bound MSP1D1 (T_m of 90°C vs 84.5°C) (**Supplementary Fig. 2a, b**). The stability of Voltage-Dependent Anion Channel 1 (VDAC-1) in LDAO micelles is increased from 63.8 to 73°C by insertion into MSP1D1 nanodiscs, and to 82.2°C when

inserted into cNW11 nanodiscs (**Supplementary Fig. 2c**). Insertion of VDAC1 into the smaller cNW9 nanodiscs (~ 8.5 nm) maintains the high T_m of 82°C (**Supplementary Fig. 2d**) and limits the number of embedded channels to a single one (**Supplementary Fig. 2e, 7**) avoiding the previously observed problem of undefined numbers of embedded channels[5]. Comparisons of proton relaxation rates of lipid resonances in empty cNW9 and MSP1D1ΔH5 nanodiscs[4] indicate more restriction imposed on the lipids inside cNW9 nanodisc by the covalently circularized belt protein. This could partially explain the enhanced thermal stability of circularized nanodiscs as compared to the conventional ones (**Supplementary Fig. 3**). In addition to the improved thermal stability, we show that covalent circularization enhances the proteolytic stability of nanodiscs (**Supplementary Fig. 4**).

During efforts to optimize the lipids/ circularized scaffold proteins ratios, we noticed that at suboptimal lipid proportions the circularized high molecular weight variants could also form nanodiscs of well-defined, non-circular shapes (**Fig. 2e**). The 3X-, 4X-, 5X-, and 6X-cNW11 spontaneously formed triangular, square, pentagonal, and hexagonal shaped nanodiscs. The flexible linkers (LPGTGS) between each copy of NW11 that result from sortase ligation enable these high MW circularized species to assemble nanodiscs with these unusual but well-defined shapes. Each side of these shapes appears to be formed by one copy of NW11. These polygonal nanodiscs, in principle, may prove useful for crystallization efforts by encouraging more efficient crystal packing relative to circular nanodiscs. Efforts to design different sizes of polygonal shapes independent of lipid saturation still in development.

Inspired by the large nanodiscs resulting from ligation of two or more NW11 molecules, we designed DNA constructs NW30 and NW50 to produce sortaggable variants that assemble ~ 30- and 50-nm nanodiscs, respectively. SDS-PAGE analysis (**Fig. 2f, g)** illustrates the purification and circularization of NW30 and NW50. Unlike cNW11, the circularized NW30 and NW50 migrate slower than the linear forms. Surprisingly, the cNW30 forms very homogenous ~15 nm instead of 30 nm nanodiscs as confirmed by negative-stain EM and size exclusion chromatography (**Fig. 2f and Supplementary Fig. 5, 6**). It appears that the protein crosses and folds over itself to form a double belt that can support the bilayer with a single polypeptide. On the other hand, as predicted, circularized NW50 assembled ~ 50 nm nanodiscs (**Fig. 2g**).

With the characterization of the circularized nanodiscs (cNDs) in place, we prepared and recorded 2D ^{15}N TROSY HSQC spectra at 45°C for [*U*-^{2}H,^{15}N] labeled VDAC-1 in cNW9 and cNW11 nanodiscs incorporating one or two copies of the channel respectively. Earlier preparations with open NDs resulted in undefined numbers of embedded VDAC-1 molecules, which yield inconsistent NMR spectra due to sample heterogeneity[5]. Our new data exhibit enhanced signal intensities and spectral resolution than previously reported NMR spectra of VDAC-1 in nanodiscs[5]. The ^{15}N-TROSY HSQC spectra recorded for [*U*-^{2}H,^{15}N] labeled VDAC-1 in cNW9 and cNW11 nanodiscs are significantly different, suggesting conformation differences between monomer and dimer form of VDAC-1 embedded in membrane (**Fig. 3**). These experiments demonstrated the power of the cND approach to control the oligomeric states of dynamically interacting proteins for structural studies. Moreover, the observed stability and spectral qualities of

VDAC-1 in cNDs at 45°C has dramatically improved compared to the open forms (**Supplementary Fig. 8)**, and will greatly facilitate NMR assignments of VDAC-1 in cNDs. We next acquired ^{1}H-^{15}N-TROSY HSQC spectrum of ^{15}N-labeled, signaling-competent variant of rat neurotensin receptor 1 (NTR1) in cNW9 nanodisc. Again, the nanodisc sample was stable at 45°C for longer than 10 days, which enabled to us to further test binding of heterotrimeric G protein composed of $G\alpha_{i1}$, $G\beta_1$, and $G\gamma_1$ to NTR1 in nanodisc and observed numerous spectral changes. The integrity of the covalently circularized nanodiscs was not affected even though small amount of DDM was accompanying G protein titrant at a final concentration of DDM less than 10% of its critical micellar concentration (CMC) values. We believe that cNDs system will greatly facilitate the studying of different dynamic events upon ligand binding in a near- native membrane environment which is critical for understanding GPCR-signal transduction process.

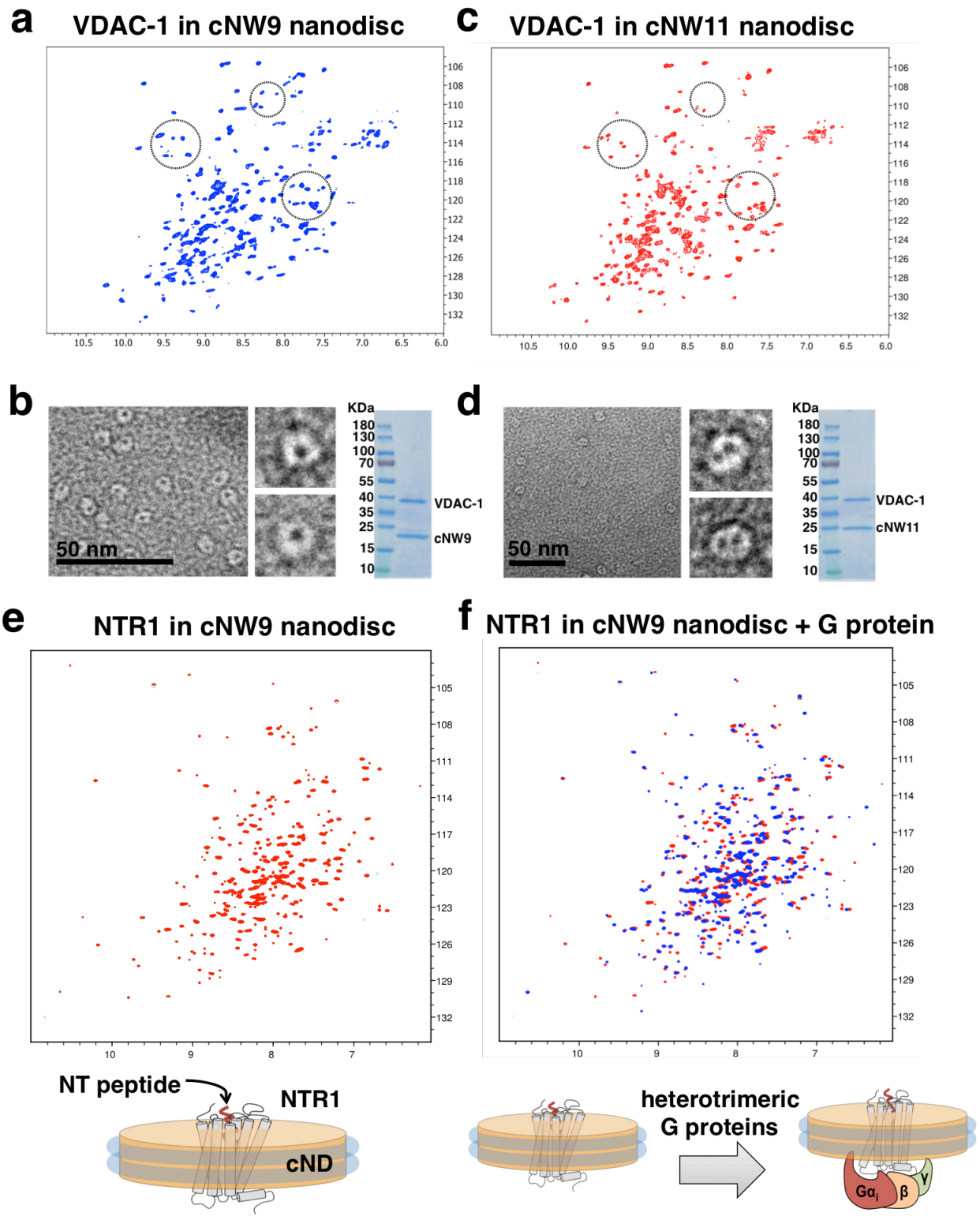

a
VDAC-1 in cNW9 nanodisc
c
VDAC-1 in cNW11 nanodisc
b
50 nm
KDa
180
130
100
70
55
40
35
25
15
10
VDAC-1
cNW9
d
50 nm
KDa
VDAC-1
cNW11
e
NTR1 in cNW9 nanodisc
NT peptide
NTR1
cND
f
NTR1 in cNW9 nanodisc + G protein
heterotrimeric
G proteins
Gαi
β
γ

Figure 3. Analysis of VDAC-1 in different size nanodiscs. (**a**) 2D ^{15}N TROSY HSQC spectrum of monomeric [*U*-^{2}H,^{15}N] labeled VDAC-1 in cNW9 nanodiscs. (**b**) Representative image of negatively stained cNW9 nanodiscs containing single VDAC-1 channel. The stain-filled channels appear as dark spot inside the nanodisc. Right: SDS-PAGE analysis of cNW9 nanodiscs containing VDAC-1. (**c**) 2D ^{15}N TROSY HSQC spectrum of dimeric [*U*-^{2}H,^{15}N] labeled VDAC-1 in cNW11 nanodiscs. (**d**) Representative image of negatively stained cNW11 nanodiscs containing two VDAC-1 channels. Right: SDS-PAGE analysis of cNW11 nanodiscs containing VDAC-1. (**e**) ^{1}H-^{15}N-TROSY HSQC spectrum of 40uM ^{15}N-labeled NTR1 in cNW9 nanodisc acquired at 45ºC on Bruker 800MHz spectrometer. (**f**) Superimposed ^{1}H-^{15}N-TROSY HSQC spectra of 40uM^{15}N-labeled NTR1 in cNW9 before (red) and after (blue) addition of purified heterotrimeric G protein composed of $G\alpha_{i1}$, $G\beta_{1}$, and $G\gamma_{1}$.

Next, we utilized the cNW50 nanodiscs as a model to study the question of how simple non-enveloped viruses transfer their genomes across membranes to initiate infection. Unlike enveloped viruses, non-enveloped viruses lack an external membrane, and the delivery of their genome into cells requires translocation across a membrane to gain access to the inside of the host cell [11, 12]. Although there are now several model systems being used to study this process, the mechanism of genome translocation remains poorly understood[13], and a more detailed structural analysis of the membrane-associated forms of the cell-entry intermediates is required. So far, mechanistic insights have been limited, due in part to technical difficulties involved in direct visualization of viral gene delivery and sample heterogeneity due to size heterogeneity of liposomes. The availability of large nanodiscs with defined size encouraged structural studies that could provide resolutions sufficient to gain insights into the mechanism of RNA translocation. As a proof of principal we have used the cNW50 nanodiscs to visualize the RNA-translocation pore of poliovirus.

Poliovirus (30 nm diameter) is the prototype member of the enterovirus genus of the picornavirus family, which are positive-sense, single-stranded RNA viruses with ~7500b

genomes enclosed by an icosahedral capsid, and lacking an envelope[14]. Viral infection is mediated by a specific receptor, CD155 (also known as the poliovirus receptor, PVR)[15]. Upon raising the temperature from 4°C to 37°C the receptor catalyzes a conformational rearrangement and expansion of the virus particle. The expanded virus is then endocytosed by a non-canonical, actin-independent pathway[16], and the RNA is released across the endosomal membrane, leaving an intact empty particle that is then transported to the perinuclear region. A 50-nm nanodisc is sufficiently large to accommodate multiple CD155 copies and has enough surface area to act as a surrogate membrane for the RNA-translocation complex during viral uncoating (**Fig. 4a**). Similar to studies that used liposomes[17-19], here 50-nm nanodiscs containing lipids derivatized with a NTA nickel-chelating head group were generated and functionalized with the His-tagged CD155 ectodomain. The receptor-decorated nanodiscs were incubated with poliovirus for 5 minutes at 4°C. The complex was then heated to 37°C for 15 minutes to initiate receptor-mediated viral uncoating (**Fig. 4b**). Negative-stain EM confirmed virus binding to the CD155-decorated nanodiscs and subsequent insertion of viral components into and across the membrane (**Fig. 4c**). Additionally, the negative-stain EM images indicated that the virus started to form a pore in the nanodisc. To obtain a view of the molecular interactions involved in the RNA-translocation complex as well as to elucidate in more detail the formation of a pore, we conducted cryo-EM studies using an FEI Polara electron microscope. **Figure 4d and Supplementary Figure 9** show the dark RNA-filled virus next to the slightly larger 50 nm nanodisc. The nanodisc is tilted in **Figure 4e** and tethered to the virus in **Figure 4f**. We were able to visualize the formation of a putative pore inside

the nanodiscs (**Fig. 4g, h**), through which the virus ejects its RNA. The images indicate that the pore is formed by several proteins inside the nanodiscs. We find it unlikely that the putative pore in the nanodisc is a hexametric arrangement of CD155 since we used only the ectodomains of the receptor, which we linked to lipid molecules through His tags. Even still, this is a possibility and the identities of the pore forming proteins are currently being determined. The availability of the circularized 50-nm nanodiscs greatly facilitates imaging as compared to using liposomes because the nanodiscs are more homogenous in size and shape, and allow for the use of a thinner ice layer. Also, RNA can be visualized more easily in the absence of large liposomal membranes. In order to reduce complexity even further, we used cNW30 nanodiscs (~15 nm) decorated with CD155. Surprisingly, the virus also tethers to this smaller nanodisc (**Fig. 4i**), forms a pore and ejects RNA, leaving an empty viral capsid behind (**Fig. 4j and Supplementary Fig. 10**). Three-dimensional reconstruction of these complexes will be a challenging undertaking, and is beyond the scope of this study. However, the quality of the data collected on the virus-nanodisc complexes represents a vast improvement over data previously used to obtain low-resolution structural models of the translocation complex using the receptor-decorated liposome model[20], and we are confident that the nanodisc model will allow determination of structures with greatly improved resolution and quality.

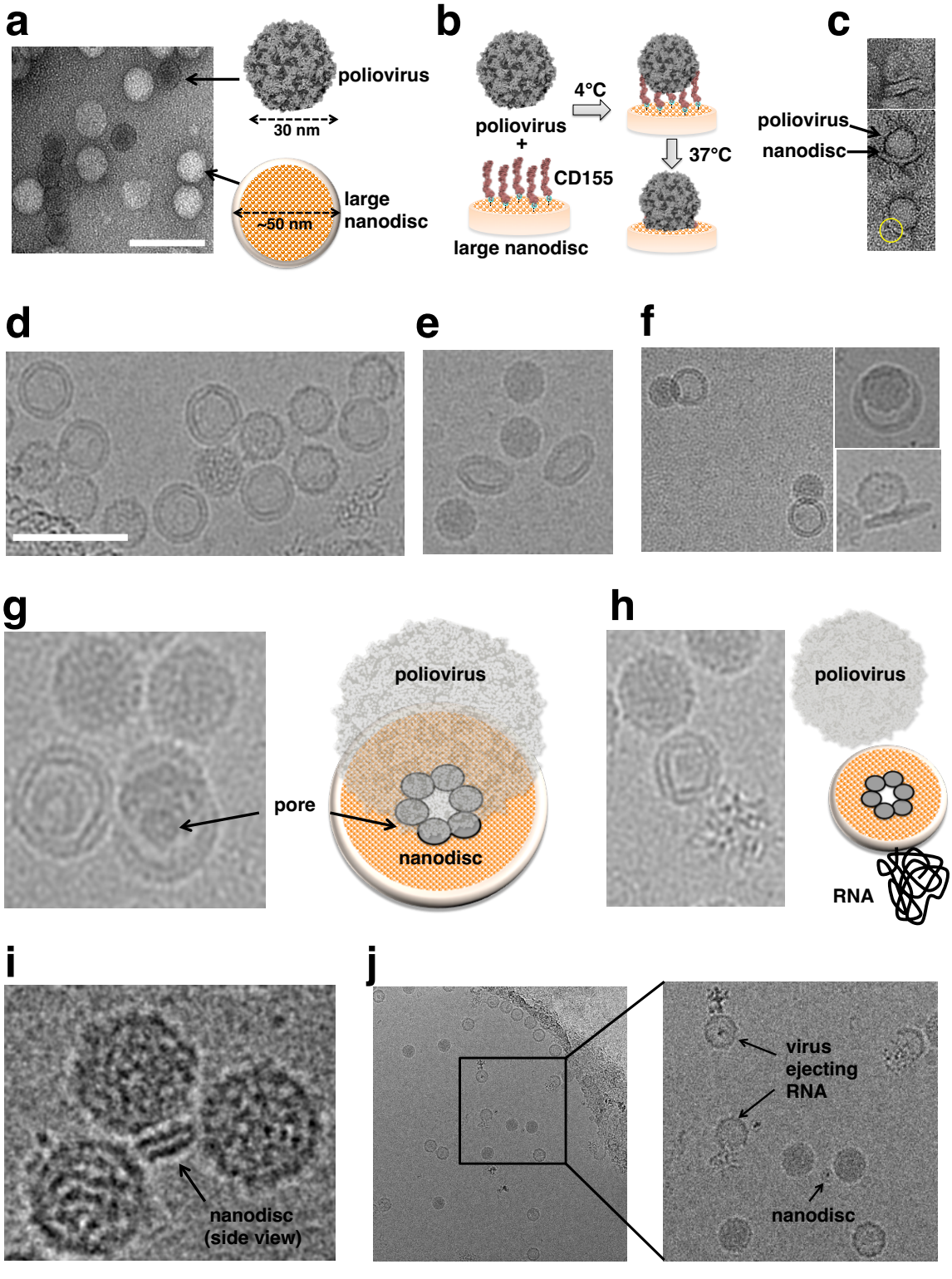

a
poliovirus
30 nm
large
nanodisc
~50 nm
b
poliovirus
+
CD155
large nanodisc
4°C
37°C
c
poliovirus
nanodisc
d
e
f
g
pore
poliovirus
nanodisc
h
poliovirus
RNA
i
nanodisc
(side view)
j
virus
ejecting
RNA
nanodisc

Figure 4. Poliovirus caught in the act. (**a**) Negative-stain EM of 50-nm circularized nanodiscs plus poliovirus. A control (nanodiscs without CD155) is shown to illustrate the relative dimensions of the 30-nm poliovirus and the 50-nm nanodisc. (**b**) Outline of the procedure used to initiate poliovirus bridging and fusion with nanodiscs decorated with CD155. (**c**) Negative-stain EM images showing individual viruses tethered to nanodiscs. (**d**) Cryo-EM image of 50-nm nanodiscs plus poliovirus. (**e**) Cryo-EM image showing a tilted view of 50-nm nanodisc. (**f**) Cryo-EM images showing individual viruses tethered to nanodiscs. (**g, h**) Cryo-EM images showing the creation of a putative pore in the nanodisc by the poliovirus. (**i**) Cryo-EM image showing three viral particles around a 15-nm nanodisc. (**j**) Cryo-EM image showing individual viruses ejecting RNA after incubation with CD155-decorated 15-nm nanodiscs.

In conclusion, we have demonstrated the construction of covalently circularized nanodiscs with a wide range of geometric shapes and sizes. The ability to make stable cNDs at multiple defined sizes up to 80 nm diameter provides a tool to tightly embed much larger membrane proteins or their intra- and extra-membrane complexes than previous nanodiscs systems have allowed. Moreover, we have shown that the newly engineered covalently circularized nanodiscs produce nanodiscs with high homogeneity in size and shape and with significantly improved stability compared to non-circularized forms, both of which would greatly facilitate their use for NMR and cryo-EM. We have demonstrated the utility of this model system to probe an outstanding question in the field of virology, and believe that the system will be similarly enabling for the structural and functional study of other large protein/membrane complexes.

A.3 SUPPLEMENTARY INFORMATION

A.3.1 Expression of NW9, NW11, NW30 and NW50.

NW9, NW11, NW30 and NW50 (all in pET-28a) containing a TEV-cleavable N-terminal His_6 tag and a C-terminal sortase-cleavable His_6 tag were transformed into BL21-Gold (DE3) competent *E. coli* cells (Agilent). 3L cell cultures were grown at 37°C, 200

rpm in Luria broth (LB) medium supplemented with 50 µg/ml Kanamycin, and expression was induced with 1 mM IPTG at an OD_{600} of 0.6 for 3 hours at 37°C (NW9 and NW11) or 16 hours at 18°C (NW30 and NW50). Cells were harvested by centrifugation (7000x*g*, 15 minutes, 4°C) and cell pellets were stored at -80°C.

A.3.2 Purification of NW9 and NW11.

Cell pellets expressing NW9 or NW11 were resuspended in lysis buffer (50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 1% Triton X-100) and lysed by sonication on ice. Lysate was centrifuged (35,000x*g*, 50 minutes, 4°C) and the supernatant was filtered and loaded onto Ni^{2+}-NTA column. The column was washed with lysis buffer then with buffer A (50 mM Tris, pH 8.0, 500 mM NaCl). To recover additional protein from the insoluble fractions, the pellets recovered from lysate centrifugation were dissolved in denaturing buffer (50 mM Tris, pH 8.0, 500 mM NaCl, 6M guanidine hydrochloride), centrifuged (35,000x*g*, 50 minutes, 4°C), and the supernatant was applied to the same Ni^{2+}-NTA column containing bound protein from the soluble fraction. The column was washed with denaturing buffer, and NW9 and NW11 were refolded on-column with 10 column volumes (CV) buffer A. Resin was then washed with 10 CV of the following: buffer A + 1% Triton X-100, buffer A + 50 mM sodium cholate, buffer A, and buffer A + 20 mM imidazole. Proteins were eluted with buffer A + 500 mM imidazole, TEV (His_6-tagged; produced in-house) was added to cleave the N-terminal His_6 tag, and the samples were dialyzed against 50 mM Tris-HCl, pH 8.0, 100 mM NaCl, 1 mM EDTA, 2 mM DTT at 4°C for 16 hours. NW9 and NW11 (still containing a C-terminal His_6 tag) were exchanged into nanodisc-assembly buffer (50 mM

Tris-HCl, pH 8.0, 500 mM NaCl, 0.02% NaN_3) using centricon concentrators (10 kDa MW cutoff, Millipore).

A.3.3. Purification of NW30 and NW50.

NW30 and NW50 were purified under denaturing conditions and refolded as follows. Pellets of cells expressing NW30 or NW50 were resuspended in denaturing lysis buffer (50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 6M guanidine hydrochloride) and lysed by sonication on ice. Lysate was centrifuged (35000x*g*, 50 minutes, 4°C) and the supernatant was filtered and loaded onto Ni^{2+}-NTA column. Resin was washed with 10 column volumes (CV) of denaturing lysis buffer to remove unbound proteins, and NW30 and NW50 were refolded on-column with 10 CV buffer A (50 mM Tris HCl, pH 8.0, 500 mM NaCl). Resin was washed with 10 CV of the following: buffer A + 1% Triton X-100, buffer A + 50 mM sodium cholate, buffer A, and buffer A + 20 mM imidazole. Proteins were eluted with buffer A + 500 mM imidazole, TEV was added to cleave the N-terminal His_6 tag, and samples were dialyzed against 50 mM Tris, pH 8.0, 100 mM NaCl, 1 mM EDTA, 2 mM DTT at 4°C for 16 hours. NW30 and NW50 (still containing a C-terminal His_6 tag) were further purified by size exclusion chromatography (SEC; Superdex 200 16/60 [GE Healthcare] equilibrated in 20 mM Tris-HCl, pH 7.5, 500 mM NaCl, 50 mM sodium cholate, 0.5 mM EDTA). SEC fractions containing NW30 and NW50 were further purified over Ni^{2+}-NTA resin to remove truncation products (which lack a C-terminal His_6 tag). Purified proteins were exchanged into nanodisc assembly buffer (50 mM Tris HCl, pH

8.0, 500 mM NaCl, 0.02% NaN_3) using centricon concentrators (30 kDa MW cutoff, Millipore).

A.3.4. MSP circularization.

A 50 mL reaction was prepared with 10 µM NWs and 5 µM (final concentrations) freshly made evolved sortase in 300 mM Tris-HCl, pH 7.5, 150 mM NaCl, and 10 mM $CaCl_2$. The reaction was incubated at 37°C for 3-4 hours or at 4°C overnight with gentle shaking on a rotating platform. A covalent sortase inhibitor AAEK2 was added to a concentration of 500 µM, and the solution was incubated further for 30 minutes at room temperature with gentle shaking. Proteins that did not undergo circularization were removed by binding to Ni^{2+}-NTA column. Circularized NWs (cNWs) were further purified by size-exclusion chromatography (Superdex 75 16/60) equilibrated in buffer containing 20 mM Tris, pH 7.5, 500 mM NaCl plus 50 mM sodium cholate or 1 mM DDM. Purified protein was exchanged into buffer A_{ix} (20 mM Tris, pH 8.2, 1 mM DDM) using centricon concentrators (10 kDa MW cutoff, Millipore) and then was applied to a Resource Q column equilibrated with the same buffer (buffer A_{ix}). A linear salt gradient from 0-60 % buffer B_{ix} (20 mM Tris, pH 8.2, 1 mM DDM, 1M NaCl) was applied. Circularized proteins were eluted around 150-200 mM NaCl.

A.3.5 Reconstitution of cNW11, cNW30 and cNW50 nanodiscs.

cNWs:lipid ratios of 1:60, 1:75, 1:1000 and 1:4000 were used to assemble cNW9**,** cNW11, cNW30, cNW50 nanodiscs respectively. Lipids (POPC:POPG 3:2, solubilized in

sodium cholate) and cNWs were incubated on ice for 1 hour. After incubation, sodium cholate was removed by the addition of Bio-beads SM-2 (Bio-Rad) and incubation on ice for 1 hour followed by overnight incubation at 4°C. The nanodisc preparations were filtered through 0.22 μm nitrocellulose-filter tubes to remove the Bio-beads. The nanodisc preparations were further purified by size-exclusion chromatography while monitoring the absorbance at 280 nm on a Superdex 200 10×300 column (for cNW9 and cNW11 nanodiscs) or Superose 6 10/300 column (for cNW30 and cNW50 nanodiscs) equilibrated in 20 mM Tris-HCl, pH 7.5, 100 mM NaCl, 0.5 mM EDTA. Fractions corresponding to the size of each nanodisc were collected and concentrated. The purity of nanodisc preparations was checked using SDS-PAGE.

A.3.6 In vitro reconstitution of VDAC-1 into POPC/POPG or DMPC/DMPG nanodiscs.

In order to assemble VDAC-1 into POPC/POPG nanodiscs, 15 μM of VDAC-1, 150 μM of cNW11 and 8.5 mM lipids (POPC:POPG 3:2, solubilized in sodium cholate) were incubated over ice for 1 hour. To assemble monomeric VDAC-1 into DMPC/DMPG nanodiscs for NMR experiments, 25 μM of VDAC-1, 175 μM of cNW9, and 9.5 mM lipids (DMPC:DMPG 3:1) were incubated at room temperature for 1 hour. To assemble dimeric VDAC-1 into DMPC/DMPG nanodiscs for NMR experiments, 79 μM of VDAC-1, 75 μM of cNW11, and 3.75 mM lipids (DMPC:DMPG 3:1) were incubated at room temperature for 1 hour. After incubation, detergents were removed by the addition of Bio-beads SM-2 (Bio-Rad) and incubation on ice for 1 hour followed by overnight incubation at 4°C (POPC/POPG nanodiscs) or overnight incubation at room temperature (DMPC/DMPG

nanodiscs). The disc preparation was filtered through 0.22 µm nitrocellulose-filter tubes to remove the Bio-beads. To remove the VDAC-free nanodiscs, the sample was mixed with Ni^{2+}-NTA resin for 1 hour at 4°C. The resin bed volume was equal to the assembly mixture. The resin was washed with buffer E (20 mM Tris-HCl, pH 8.0, 0.1 M NaCl, 20 mM imidazole). Nanodiscs containing VDAC were eluted with buffer E containing 0.5 M imidazole. The nanodisc preparation was further purified by size-exclusion chromatography while monitoring the absorbance at 280 nm on a Superdex 200 10 × 300 or a 16/60 Superdex 200 prep grade columns (GE Healthcare) equilibrated in buffer F (20 mM sodium phosphate, 50 mM NaCl, 5 mM DTT, 1 mM EDTA, pH 7.0). Fractions corresponding to the size of the VDAC-nanodisc complex were collected and concentrated. The purity of VDAC-containing nanodiscs was checked using SDS-PAGE.

A.3.7. Production and reconstitution of NTR1 into DMPC/DMPG nanodiscs.

The expression and the purification of a signaling-competent, thermo stabilized variant of rat neurotensin receptor 1 (termed HTGH4) were performed with some modifications from previously described protocols[21,22]. Briefly, the full length fusion protein consisting of the maltose-binding protein (MBP), followed by a His tag, a 3C protease recognition site, the NTR1, a second 3C protease recognition site, thioredoxin (TrxA), and a His tag at the C terminus was purified by Ni^{2+} affinity chromatography. The purified full length fusion protein was immediately incorporated into DMPC:DMPG (3:1) cNW9 nanodiscs. The assembled nanodiscs were subjected to Ni^{2+} affinity chromatography to remove empty nanodiscs from the receptor containing nanodiscs utilizing the His tag of

the fusion protein. The receptor containing nanodiscs were further purified by size exclusion chromatography followed by incubation with 3C protease cleavable neurotensin peptide bound resin (pD-NT) column to allow the enrichment of nanodiscs containing correctly folded receptors. Next, His-tagged 3C protease (produced in house) was added to the pD-NT column followed by incubation for 2 h. The eluted proteins (neurotensin bound-NTR1 in nanodisc, MBP, TrxA and 3C protease) were subjected to another Ni affinity chromatography followed by size exclusion using a Superdex-200 10/300 column that was pre-equilibrated with running buffer containing 20 mM sodium phosphate, pH 6.9, 50 mM NaCl, 5 mM DTT, 1 mM EDTA.

The heterotrimeric G protein ($\alpha_{i1}\beta_1\gamma_1$) was expressed in Sf9 cells using a single baculovirus encoding all three subunits as described by Egloff et al.[22] and purified following the procedure described by Rasmussen et al.[23]

A.3.8 Negative-stain electron microscopy

Samples were prepared by conventional negative staining as described previously[24]. Briefly, 3.5 µl of nanodisc samples were adsorbed to carbon-coated copper grids and stained with 0.75% (w/v) uranyl formate. All EM images (except images for VDAC in cNW9, supplementary figure 5) were collected with a Philips CM10 electron microscope (FEI) equipped with a tungsten filament and operated at an acceleration voltage of 100 kV. Images were recorded with a Gatan 1 K × 1 K CCD camera (Gatan, Inc., Pleasanton, CA, USA). 2D class averages were computed by EMAN2[25].

A.3.9 NMR spectroscopy

[*U*-^{2}H,^{15}N] labeled VDAC-1 in DMPC:DMPG (3:1) lipid nanodiscs (611 μM and 500 μM of VDAC1 in cNW9 and cNW11 nanodiscs, respectively) were prepared as described above in NMR buffer (20 mM $NaPO_4$, 50 mM NaCl, 5 mM DTT, 1 mM EDTA, pH 7.0, 6% D_2O). ^{15}N-TROSY HSQC data were collected at 45°C on a Bruker 800 spectrometer equipped with TXO cryogenic probe. Data for cNW9 and cNW11 nanodiscs were acquired with 48 and 96 scans respectively, and 128 complex points in the ^{15}N-indrect dimension.

^{1}H-^{15}N-TROSY HSQC spectrum of 40uM ^{15}N-labeled NTR1 in cNW9 nanodiscs was acquired at 45ºC on Bruker 800MHz spectrometer. Data were collected with 2048 scans per FID and 36 non-uniformly sampled ^{15}N-dimension complex points (maximum point of 64). The ^{15}N-dimension time domain data was predicted with hmsIST[26] and the spectrum processed by NMRPipe software programs[27].

A.3.10 Cryo-electron microscopy

A 3.5 ul droplet of the poliovirus-nanodisc complex was loaded onto a glow-discharged holey grid (Protochips, Morrisville, NC). The excess liquid was removed from the grid surface before it was rapidly plunged into liquid ethane. Grids were transferred to an FEI Polara electron microscope operating at an acceleration voltage of 300 keV. Micrographs were acquired on a K2 Summit camera (Gatan, Pleasanton, CA) in super-resolution mode using SerialEM[28], whereby 25 frames were collected to a total dose of 30 electrons per square angstrom. These frames were aligned and averaged using

motioncorr[29].

A.3.11 Protein sequences

NW9 (assemble ~8.5 nm nanodisc)

MGSSHHHHHHENLYFQGSTFSKLREQLGPVTQEFWDNLEKETEGLRQEMSKDLEEV

KAKVQPYLDDFQKKWQEEMELYRQKVEPLGEEMRDRARAHVDALRTHLAPYSDELR

QRLAARLEALKENGGARLAEYHAKATEHLSTLSEKAKPALEDLRQGLLPVLESFKVSF

LSALEEYTKKLNTQLPGTGAAALEHHHHHH

NW11 (assembles ~ 11 nm nanodiscs)

MGSSHHHHHHENLYFQGSTFSKLREQLGPVTQEFWDNLEKETEGLRQEMSKDLEEV

KAKVQPYLDDFQKKWQEEMELYRQKVEPLRAELQEGARQKLHELQEKLSPLGEEMR

DRARAHVDALRTHLAPYSDELRQRLAARLEALKENGGARLAEYHAKATEHLSTLSEK

AKPALEDLRQGLLPVLESFKVSFLSALEEYTKKLNTQLPGTGAAALEHHHHHH

NW30 (assemble ~15 nm nanodisc)

MGSSHHHHHHENLYFQGSTFSKLREQLGPVTQEFWDNLEKETEGLRQEMSKDLEEV

KAKVQPYLDDFQKKWQEEMELYRQKVEPLRAELQEGARQKLHELQEKLSPLGEEMR

DRARAHVDALRTHLAPYSDELRQRLAARLEALKENGGARLAEYHAKATEHLSTLSEK

AKPALEDLRQGLLPVLESFKVSFLSALEEYTKKLNTQGTPVTQEFWDNLEKETEGLRQ

EMSKDLEEVKAKVQPYLDDFQKKWQEEMELYRQKVEPLRAELQEGARQKLHELQEK

LSPLGEEMRDRARAHVDALRTHLAPYSDELRQRLAARLEALKENGGARLAEYHAKAT

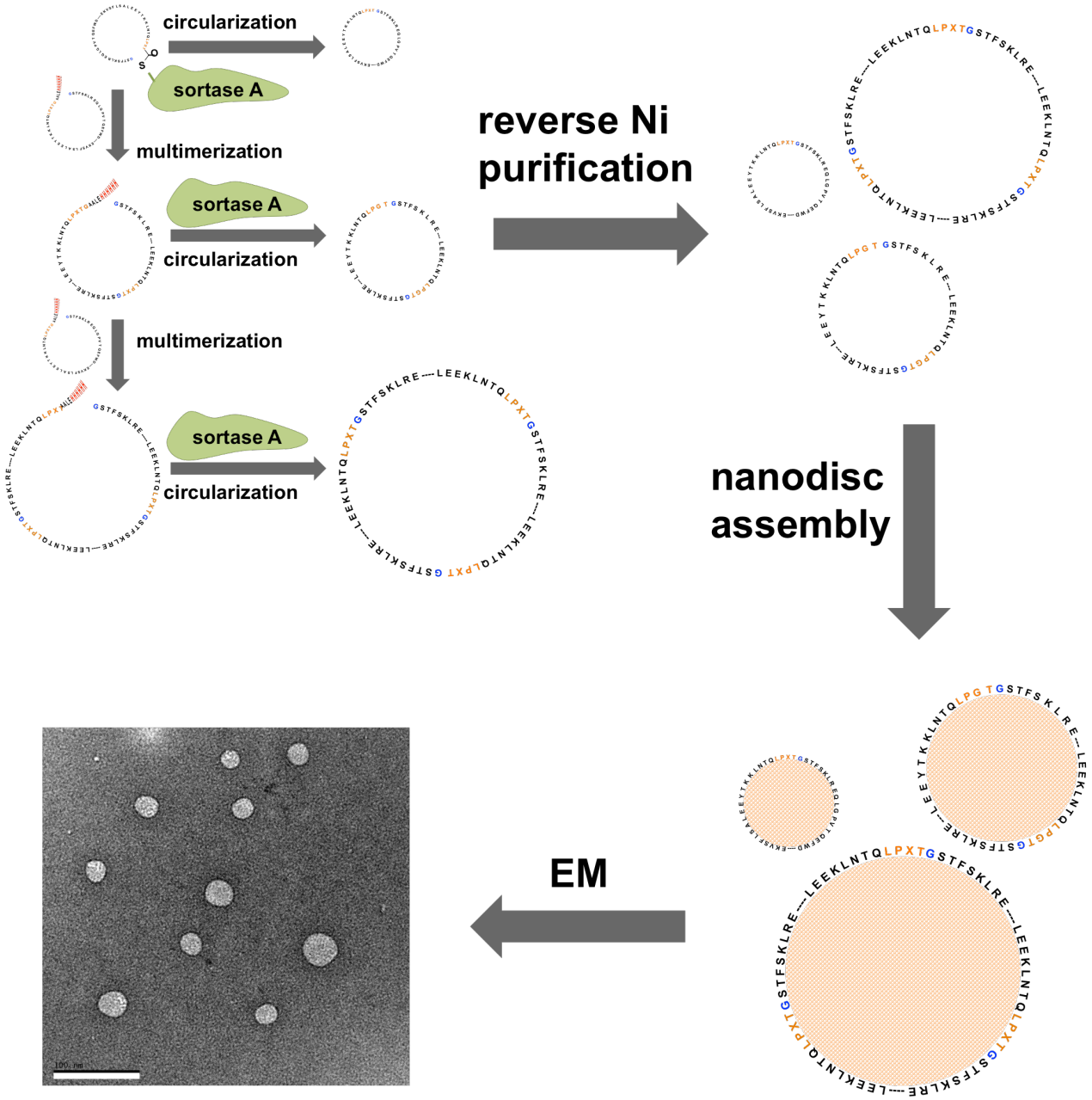
EHLSTLSEKAKPALEDLRQGLLPVLESFKVSFLSALEEYTKKLNTQGTPVTQEFWDNL

EKETEGLRQEMSKDLEEVKAKVQPYLDDFQKKWQEEMELYRQKVEPLRAELQEGAR
QKLHELQEKLSPLGEEMRDRARAHVDALRTHLAPYSDELRQRLAARLEALKENGGAR
LAEYHAKATEHLSTLSEKAKPALEDLRQGLLPVLESFKVSFLSALEEYTKKLNTQLPGT
GAAALEHHHHHH

NW50 (assemble ~ 50 nm nanodisc)

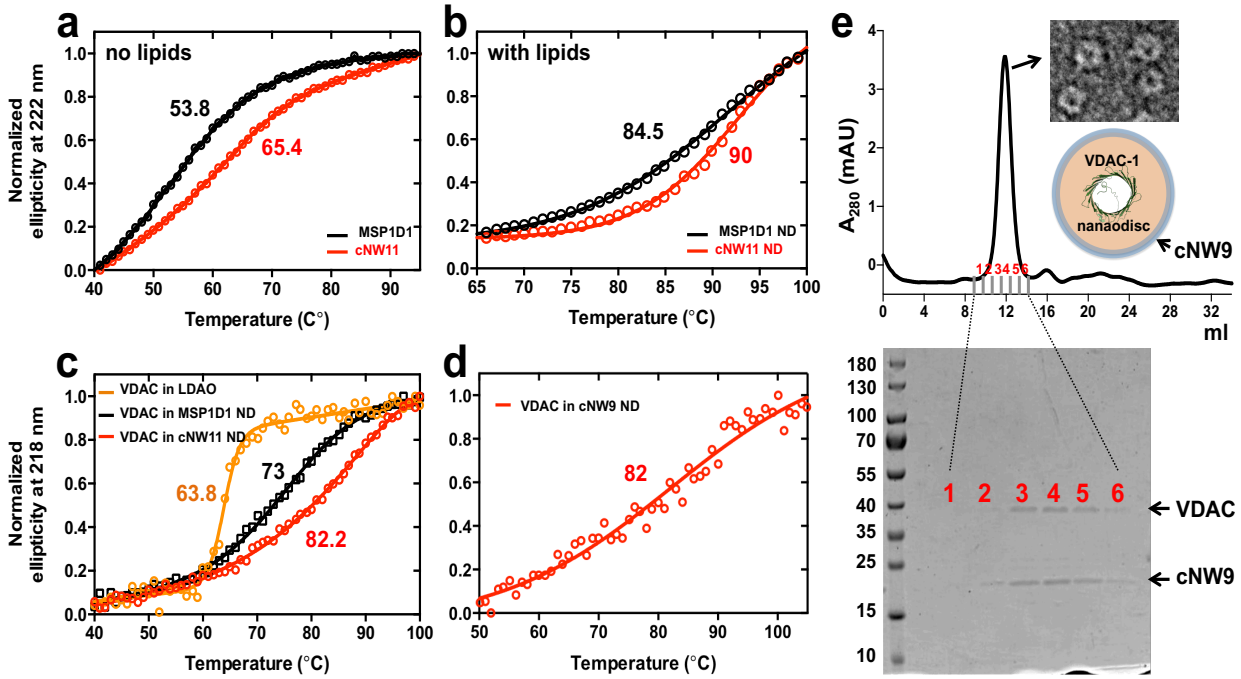
MGSSHHHHHHENLYFQGSTFSKLREQLGPVTQEFWDNLEKETEGLRQEMSKDLEEV
KAKVQPYLDDFQKKWQEEMELYRQKVEPLRAELQEGARQKLHELQEKLSPLGEEMR
DRARAHVDALRTHLAPYSDELRQRLAARLEALKENGGARLAEYHAKATEHLSTLSEK
AKPALEDLRQGLLPVLESFKVSFLSALEEYTKKLNTQGTPVTQEFWDNLEKETEGLRQ
EMSKDLEEVKAKVQPYLDDFQKKWQEEMELYRQKVEPLRAELQEGARQKLHELQEK
LSPLGEEMRDRARAHVDALRTHLAPYSDELRQRLAARLEALKENGGARLAEYHAKAT
EHLSTLSEKAKPALEDLRQGLLPVLESFKVSFLSALEEYTKKLNTQGTPVTQEFWDNL
EKETEGLRQEMSKDLEEVKAKVQPYLDDFQKKWQEEMELYRQKVEPLRAELQEGAR
QKLHELQEKLSPLGEEMRDRARAHVDALRTHLAPYSDELRQRLAARLEALKENGGAR
LAEYHAKATEHLSTLSEKAKPALEDLRQGLLPVLESFKVSFLSALEEYTKKLNTQGTPV
TQEFWDNLEKETEGLRQEMSKDLEEVKAKVQPYLDDFQKKWQEEMELYRQKVEPLR
AELQEGARQKLHELQEKLSPLGEEMRDRARAHVDALRTHLAPYSDELRQRLAARLEA
LKENGGARLAEYHAKATEHLSTLSEKAKPALEDLRQGLLPVLESFKVSFLSALEEYTK
KLNTQGTPVTQEFWDNLEKETEGLRQEMSKDLEEVKAKVQPYLDDFQKKWQEEMEL
YRQKVEPLRAELQEGARQKLHELQEKLSPLGEEMRDRARAHVDALRTHLAPYSDELR
QRLAARLEALKENGGARLAEYHAKATEHLSTLSEKAKPALEDLRQGLLPVLESFKVSF

LSALEEYTKKLNTQLPGTGAAALEHHHHHH

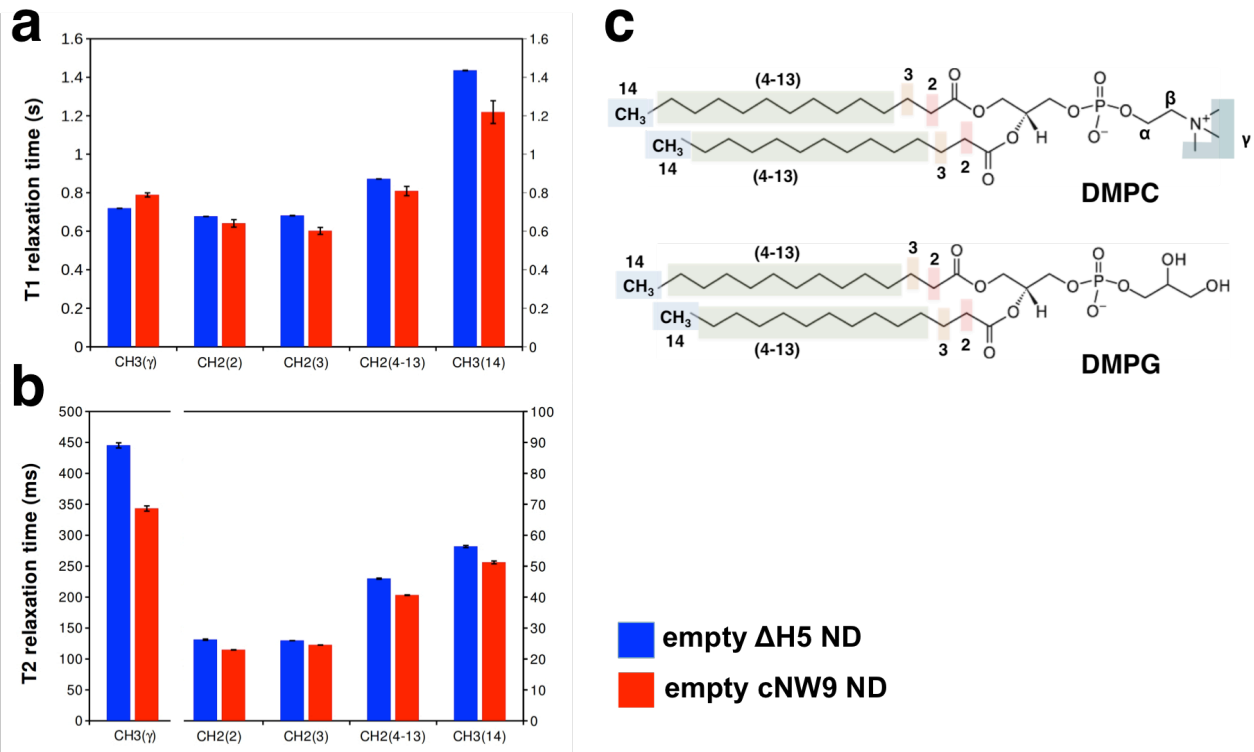


Supplementary Figure 1: Adding evolved sortase to concentrated NW11 solution leads to multimerization of NW11 followed by circularization. The oligomeric, circularized species containing variable numbers of NW11 assemble into nanodiscs of

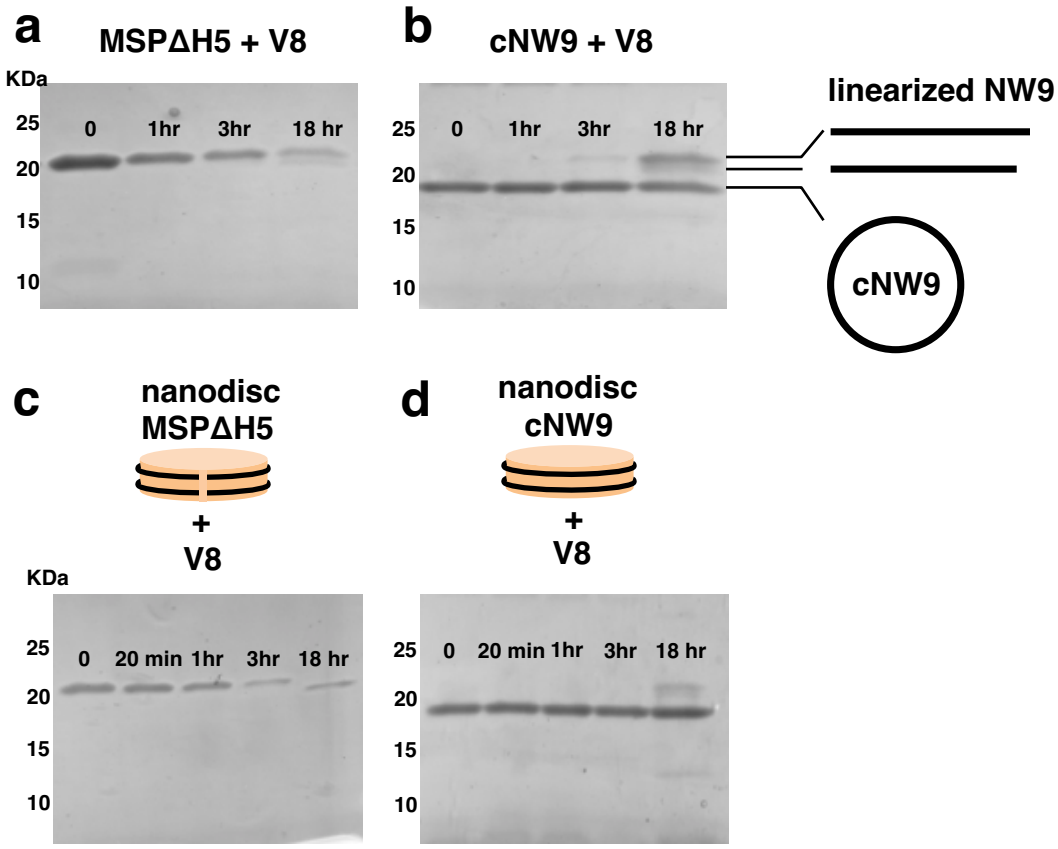
various sizes. **Bottom**: Negative-stain EM image showing large nanodiscs made using oligomeric, circularized NW11. Scale bar represents 100 nm.



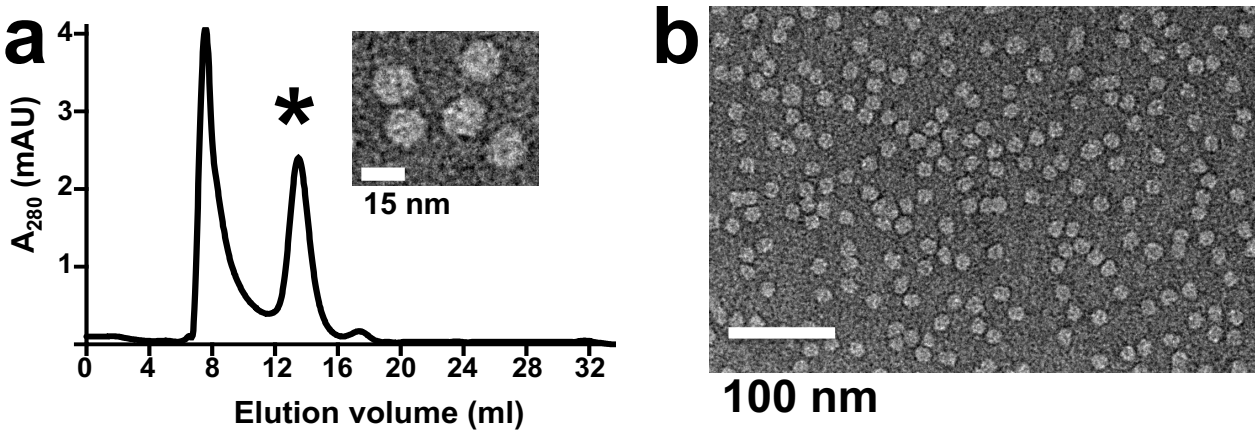
Supplementary Figure 2. Covalent circularization stabilizes MSPs against thermal unfolding without and with lipids, stabilizes embedded VDAC1 and allows control of the number of channels embedded. Thermal unfolding of MSP1D1 (black) and cNW11 (red) without **(a)** and with lipids (**b**) followed by circular dichroism (CD) spectroscopy at 222 nm, the wavelength most characteristic of helical secondary structure. POPC/POPG lipids at a molar ratio of 3:2 were used to generate nanodiscs. (**c**) Placement in MSP1D1 nanodiscs increases the melting temperature of VDAC1 by 9.2 degrees over that of VDAC1 in an LDAO micelle environment, and covalent circularization of the scaffold protein (cNW11) raises T_m by additional 9.2 degrees. Thermal unfolding of human VDAC-1 was followed by CD spectroscopy at 218 nm, the wavelength most characteristic of β-sheet secondary structure. Orange: VDAC1 in 0.1% LDAO. black: VDAC1 reconstituted into conventional nanodiscs (assembled using MSP1D1), and red: VDAC1 reconstituted into circularized nanodiscs (assembled using cNW11). Nanodiscs were made with POPC/POPG 3:2 lipids. (**d**) Thermal unfolding of VDAC1 reconstituted into circularized nanodiscs (assembled using cNW9) followed by CD spectroscopy at 218 nm. Nanodiscs were made with POPC/POPG lipids at a molar ratio of 3:2. All samples were in 20 mM Tris-HCl, pH 7.5, 100 mM NaCl. **(e)** Analysis of the VDAC1 nanodisc assembly reaction. Top: size-exclusion chromatography and negative-stain EM of VDAC1 in cNW9 nanodiscs. Negative-stain image shows nanodiscs containing a single channel. The stain-filled channels appear as dark spots inside nanodiscs. Bottom: SDS-PAGE analysis of the nanodisc assembly. Fractions 1-6 were collected and analyzed.



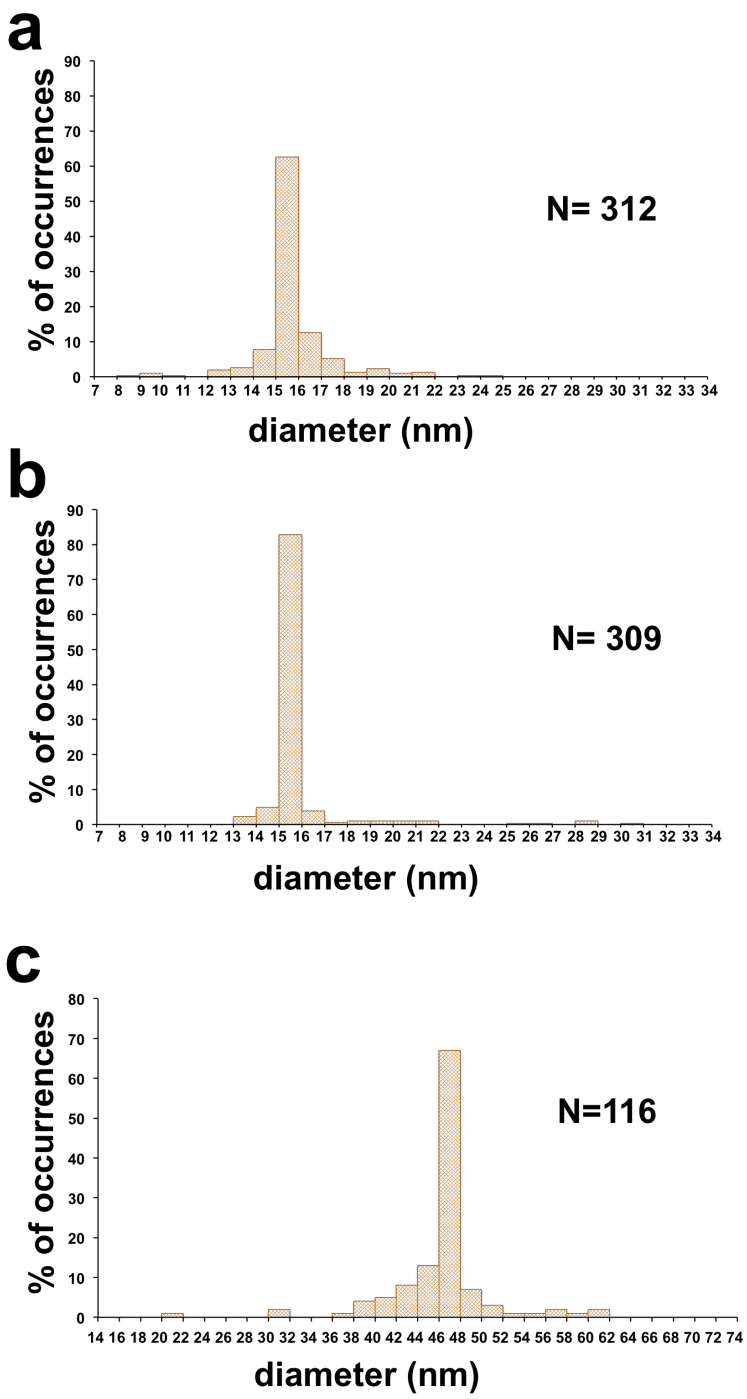
Supplementary Figure 3. Relaxation analysis of DMPC/DMPG (3:1) phospholipids in cNW and ΔH5 nanodiscs. (**a**) ^{1}H-NMR T1 relaxation times of DMPC/DMPG lipid signals in empty ΔH5 (blue) and cNW9 (red) nanodiscs. (**b**) ^{1}H-NMR T2 relaxation times of DMPC/DMPG lipids signals in empty ΔH5 (blue) and cNW9 (red) nanodiscs (**c**) Chemical structure of DMPC and DMPG.
Except the DMPC γ methyl group located outside the bilayer region, the lipids in cNW9 show shorter T1 and T2 times consistent with smaller nanodisc size and dynamics indicative of the more restrictive circularized cNW9 belt. Measurements were acquired at 45ºC on Bruker 500MHz spectrometer using inversion recovery for T1 and CPMG refocusing train for T2.


a
MSPΔH5 + V8
KDa
25
20
15
10
0
1hr
3hr
18 hr
b
cNW9 + V8
linearized NW9
cNW9
c
nanodisc
MSPΔH5
+
V8
KDa
0
20 min
1hr
3hr
18 hr
d
nanodisc
cNW9
+
V8

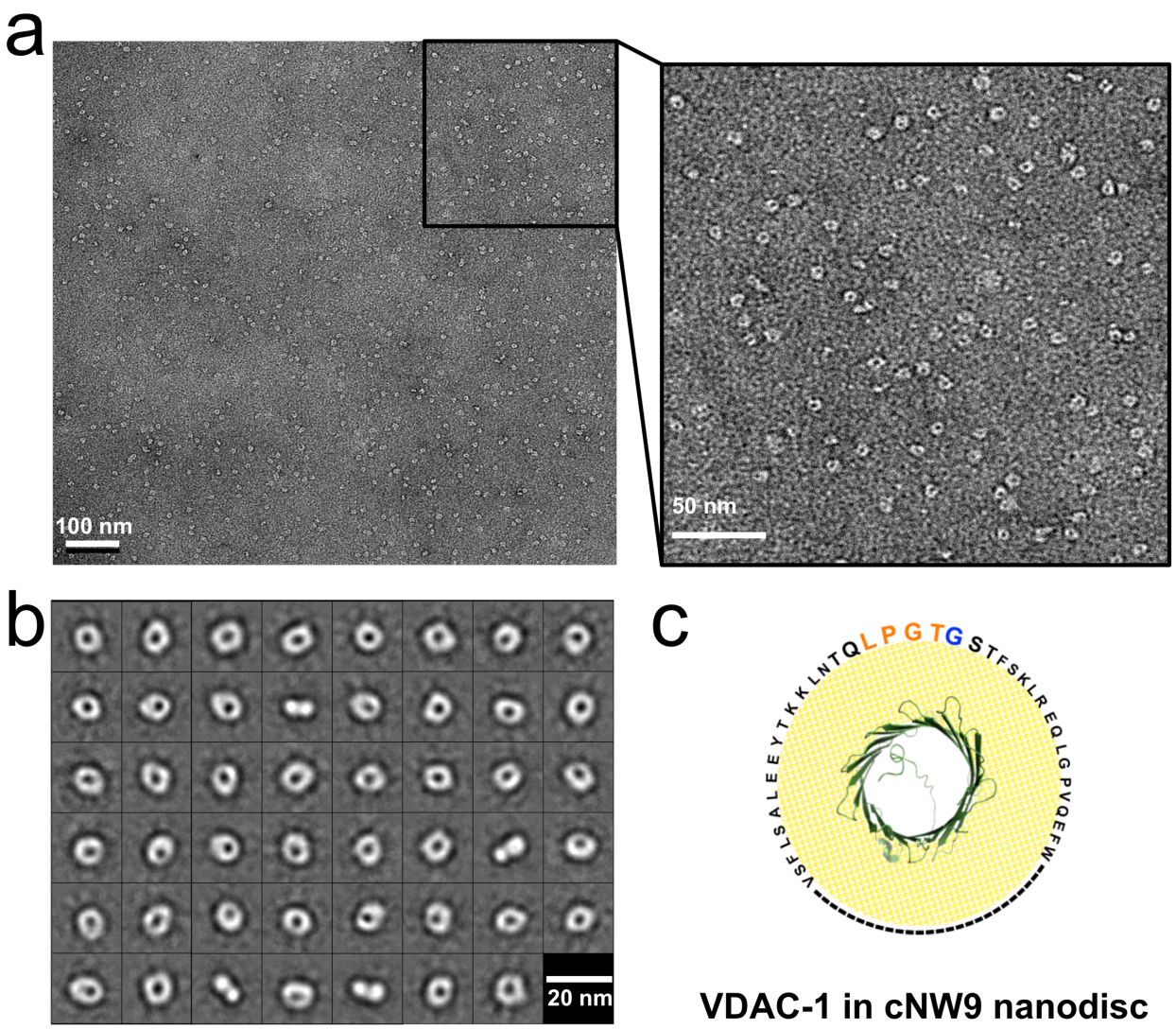
Supplementary Figure 4: Covalent circularization stabilizes MSPs against digestion by V8 protease without and with lipids. (**a, b**) SDS-PAGE analysis of the proteolysis of lipid-free MSPΔH5 and cNW9. Samples were treated with V8 protease for 0 min (before addition of V8), and for 1, 3 and 18 hours. Lanes are labeled with times of protease treatment. Proteolysis was performed in 20 mM Tris-HCl, pH 7.5, 100 mM NaCl at 37°C and using a protein:protease ratio (w/w) of 1000:1. Treatment of MSPΔH5 with V8 resulted in the appearance of a large peptide, which is close in size to MSPΔH5 and is not discernible until about 18 hours after addition of V8; this peptide may be generated earlier but is not visible due to the low amount or overlap with uncleaved MSPΔH5. The intensity of the MSPΔH5 band decreased by 75% after 3 hours and by 93% after 18 hours. On the other hand, the intensity of the cNW9 band decreased by only 20% after 18 hours. A band that corresponds to linearized NW9 was observed after 3 hours and increased in intensity after 18 hours. (**c, d**) SDS-PAGE analysis of the proteolysis of nanodiscs assembled with MSPΔH5 and cNW9. Samples were treated with V8 protease for 0 min (before addition of V8), and for 20 min, 1, 3 and 18 hours. Proteolysis was performed at 37°C at pH 7.5 in 20 mM Tris-HCl, 100 mM NaCl using a 100:1 protein:protease (w/w). The band intensity of MSPΔH5 decreased by 81% after 3 hours. There were no decreases in cNW9 band intensity up to 3 hours. ImagJ software was for analyzing band intensities.



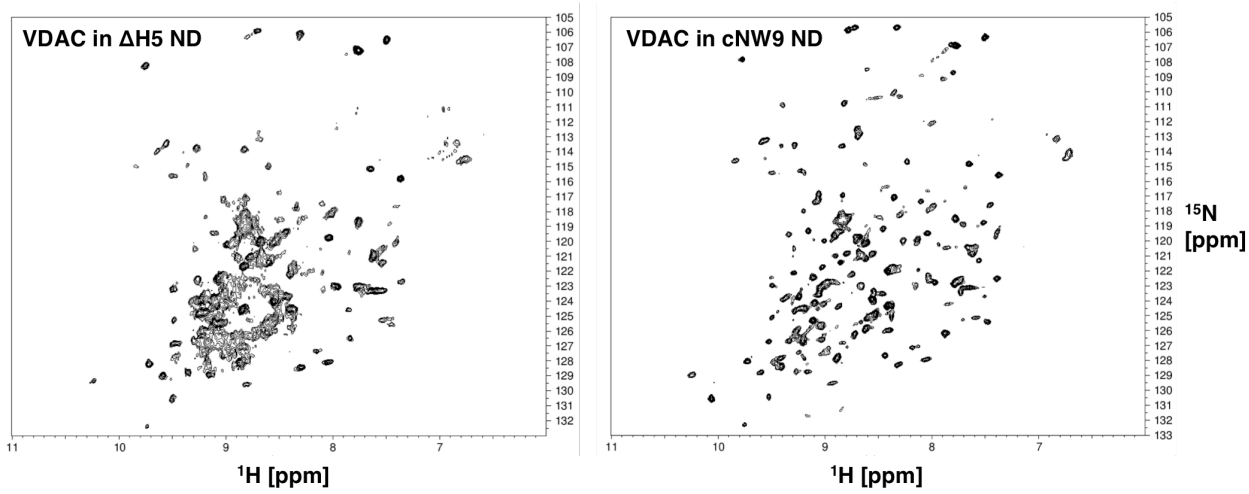
Supplementary Figure 5. Analysis of nanodiscs assembled using a cNW30:lipid ratio of 1:1000. (**a**) Size-exclusion chromatography (SEC) analysis of the assembled nanodiscs. The SEC column (Superose 6 10/300) was equilibrated in 20 mM Tris-HCl, pH 7.5, 100 mM NaCl, 0.5 mM EDTA. The peak (labeled with *) was collected and analyzed by negative-stain EM (**b**).



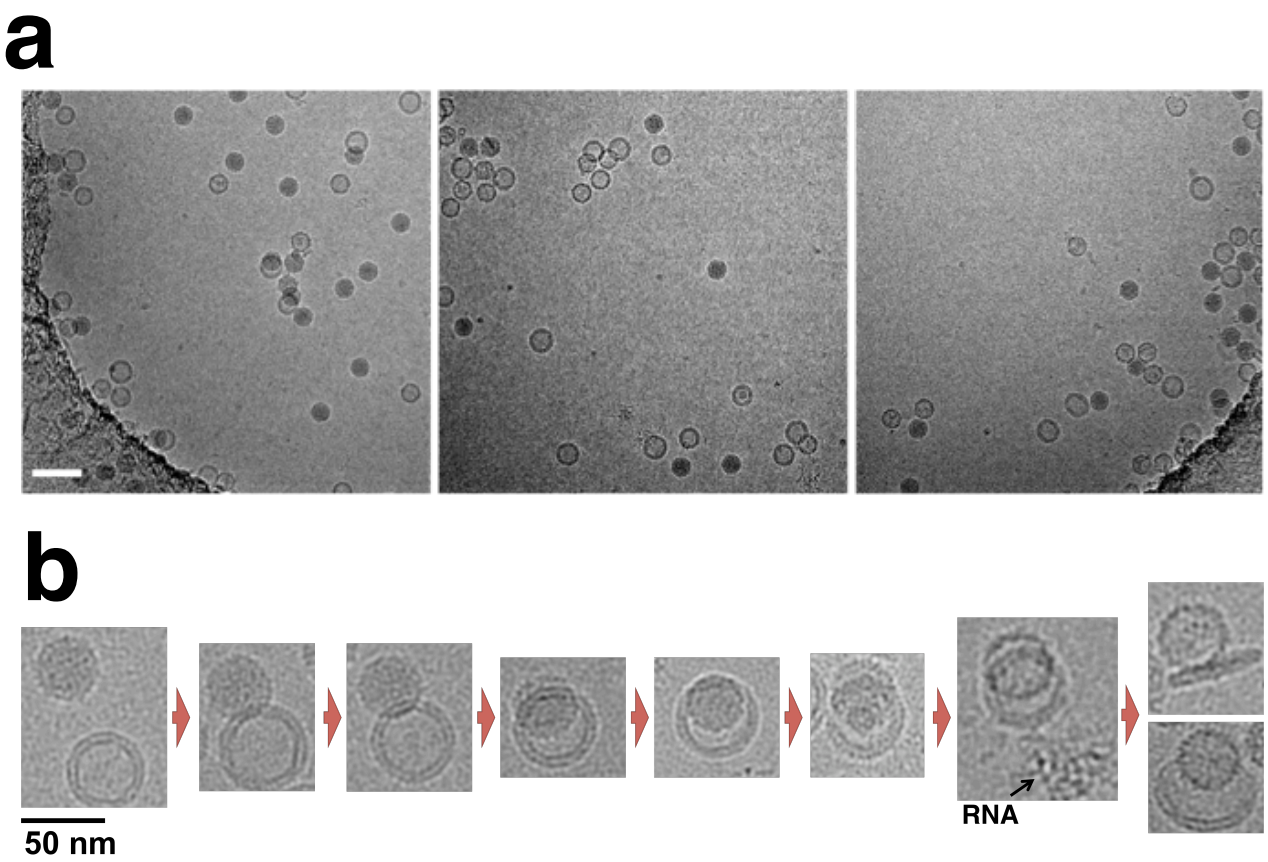
Supplementary Figure 6. Diameter distribution for nanodiscs assembled using non-circularized (**a**) and circularized (**b**) NW30 proteins. (**c**) Diameter distribution for nanodiscs made using circularized NW50. Analyses were performed with the ImageJ software[30].



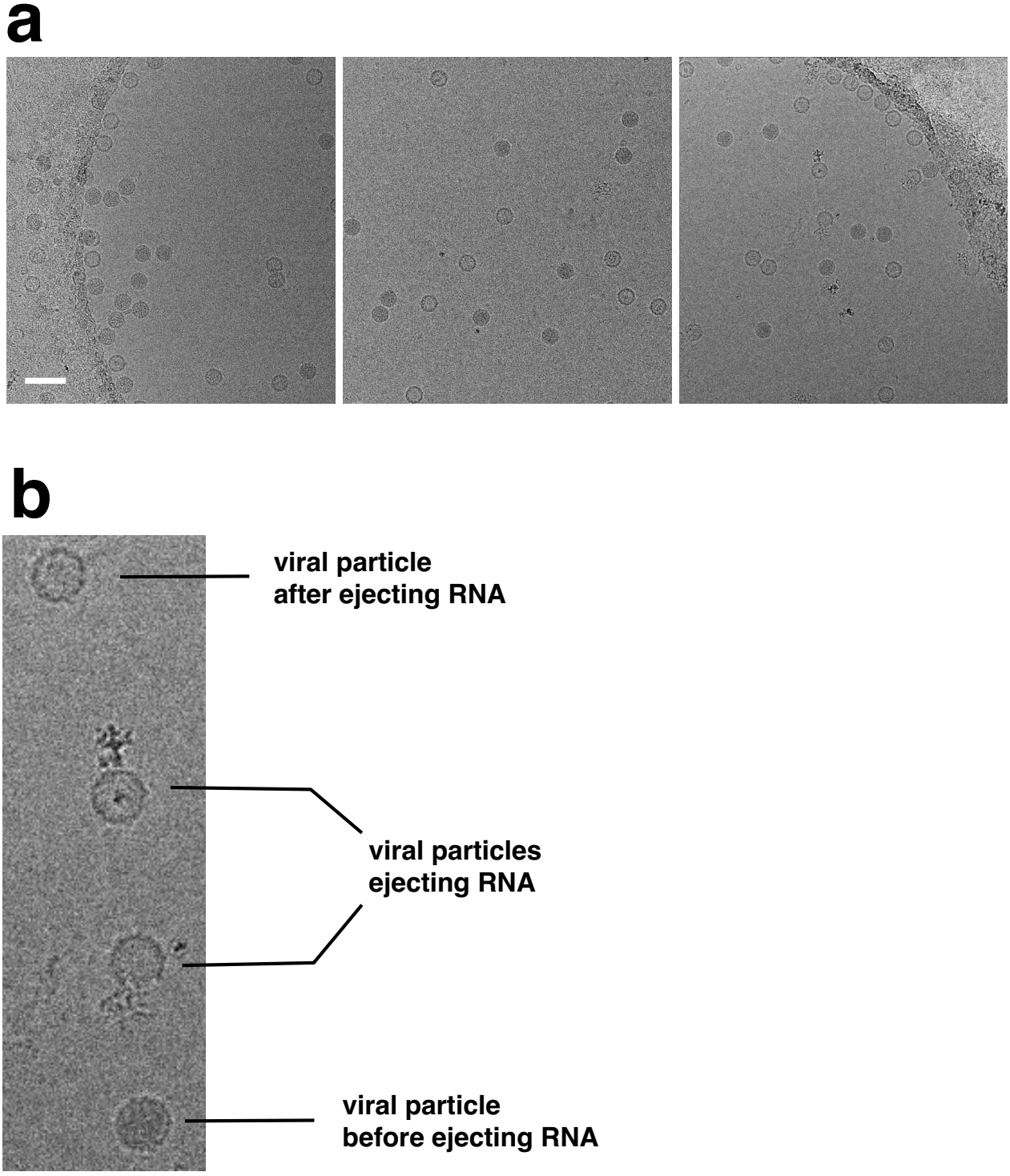
Supplementary Figure 7. Single-particle EM of negative stained VDAC-1 in cNW9 nanodisc. (**a**) Representative field micrograph of VDAC-1 in cNW9. Right. Magnified image of the rectangular area showing one VDAC channel per nanodisc. Images were collected on a Tecnai T12 electron microscope (FEI) operated at 120 kV. (**b**) Representative two-dimensional (2D) class averages of VDAC-1 in cNW9 nanodisc showing distinct views (top, tilted and side views). (**c**) Cartoon representation of VDAC-1 in cNW9 nanodisc.



Supplementary Figure 8: Comparison of NMR spectral quality of VDAC-1 in cNW9 and ΔH5 nanodiscs. (**a**) 15N-TROSY HSQC recorded at 45°C of 100uM ^{15}N-^{2}D-labeled VDAC1 in ΔH5 DMPC:DMPG =3:1 nanodisc acquired overnight on 600MHz spectrometer. (B) 15N-TROSY HSQC recorded at 45°C of 100uM ^{15}N-^{2}D-labeled VDAC1 in cNW9 DMPC:DMPG=3:1 nanodisc acquired overnight on 800MHz spectrometer. The spectral quality and sample stability are greatly improved by using cNW ND as compared to ΔH5 nanodiscs.



Supplementary Figure 9: (**a**) Representative cryo-EM images of poliovirus plus CD155-decorated 50-nm nanodiscs in vitreous ice. Scale bar represents 100 nm. (**b**) Representative poliovirus-nanodisc complexes showing different intermediate states observed in vitreous ice.



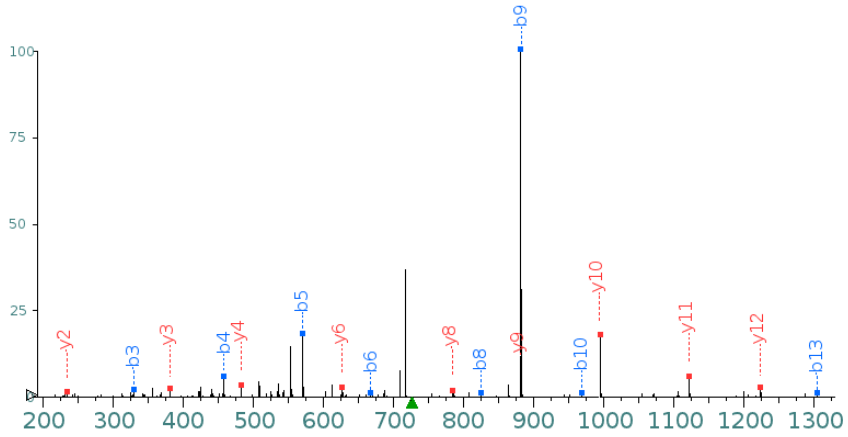
Supplementary Figure 10: (a) Representative cryo-EM images of poliovirus plus CD155-decorated 15-nm nanodiscs in vitreous ice. Scale bar represents 100 nm. **(b)** Cryo-EM image showing individual viruses ejecting RNA after incubation with CD155-decorated 15-nm nanodiscs.

MS/MS confirmation of the circularization reactions for NW9, NW30 and NW50

NW9 (amino acid sequence after TEV cleavage)
GSTFSKLREQLGPVTQEFWDNLEKETEGLRQEMSKDLEEVKAKVQPYLDDFQKKWQ
EEMELYRQKVEPLGEEMRDRARAHVDALRTHLAPYSDELRQRLAARLEALKENGGAR
LAEYHAKATEHLSTLSEKAKPALEDLRQGLLPVLESFKVSFLSALEEYTKK**LNTQLPGT**
GAAALEHHHHHH

Peptide sequence

LNTQLPGTGSTFSK



seq	#	b	y	#
L	1	114.091	---	14
N	2	228.134	1337.670	13
T	3	329.182	1223.627	12
Q	4	457.241	1122.579	11
L	5	570.325	994.520	10
P	6	667.377	881.436	9
G	7	724.399	784.384	8
T	8	825.446	727.362	7
G	9	882.468	626.314	6
S	10	969.500	569.293	5
T	11	1070.548	482.261	4
F	12	1217.616	381.213	3
S	13	1304.648	234.145	2
K	14	----	147.113	1

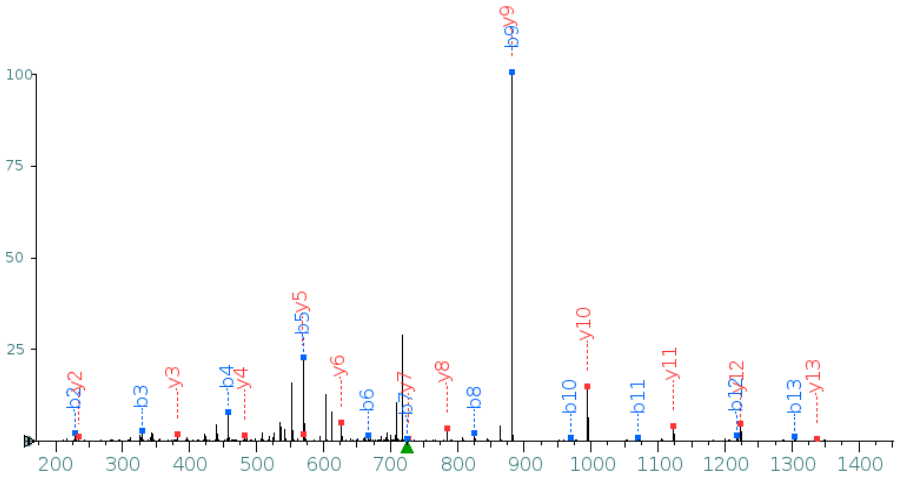
Supplementary Figure 11: Characterization of cNW9 by MS/MS confirms the ligation of the C terminus to the N terminus. MS/MS spectrum of a tryptic fragment of cNW9 showing the ligation of the C-terminal motif (**LNTQLPGT**G-His_6) to the N-terminal residues (**GSTFSK**). Expected masses for b and y ions along with the peptide sequence are listed in the table. The b and y ions that were identified in the MS/MS spectrum are highlighted in blue and red. The full amino acid sequence of linear NW9 is shown at the top.

NW30 (amino acid sequence after TEV cleavage)

GSTFSKLREQLGPVTQEFWDNLEKETEGLRQEMSKDLEEVKAKVQPYLDDFQKKWQE
EMELYRQKVEPLRAELQEGARQKLHELQEKLSPLGEEMRDRARAHVDALRTHLAPYSD
ELRQRLAARLEALKENGGARLAEYHAKATEHLSTLSEKAKPALEDLRQGLLPVLESFKV
SFLSALEEYTKKLNTQGTPVTQEFWDNLEKETEGLRQEMSKDLEEVKAKVQPYLDDFQ
KKWQEEMELYRQKVEPLRAELQEGARQKLHELQEKLSPLGEEMRDRARAHVDALRTH
LAPYSDELRQRLAARLEALKENGGARLAEYHAKATEHLSTLSEKAKPALEDLRQGLLPV
LESFKVSFLSALEEYTKKLNTQGTPVTQEFWDNLEKETEGLRQEMSKDLEEVKAKVQP
YLDDFQKKWQEEMELYRQKVEPLRAELQEGARQKLHELQEKLSPLGEEMRDRARAHV
DALRTHLAPYSDELRQRLAARLEALKENGGARLAEYHAKATEHLSTLSEKAKPALEDLR
QGLLPVLESFKVSFLSALEEYTKK**LNTQLPGT**GAAALEHHHHHH

Peptide sequence

LNTQLPGTGSTFSK



seq	#	b	y	#
L	1	114.091	---	14
N	2	228.134	1337.670	13
T	3	329.182	1223.627	12
Q	4	457.241	1122.579	11
L	5	570.325	994.520	10
P	6	667.377	881.436	9
G	7	724.399	784.384	8
T	8	825.446	727.362	7
G	9	882.468	626.314	6
S	10	969.500	569.293	5
T	11	1070.548	482.261	4
F	12	1217.616	381.213	3
S	13	1304.648	234.145	2
K	14	----	147.113	1

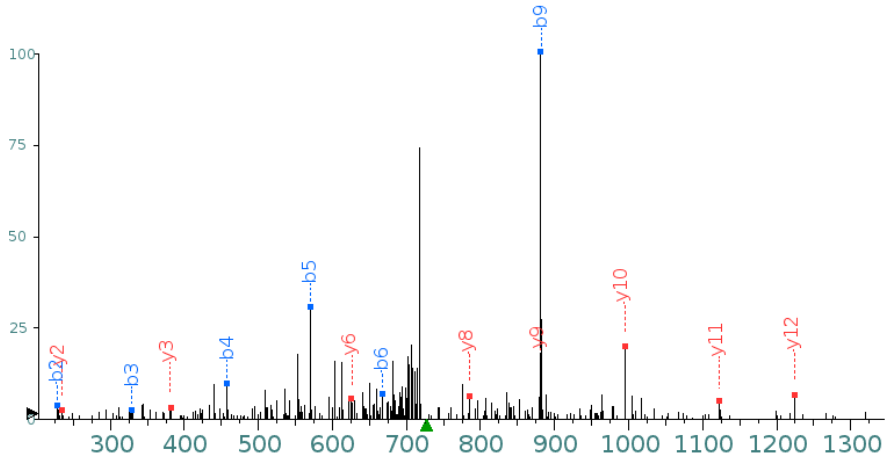
Supplementary Figure 12: Characterization of cNW30 by MS/MS confirms the ligation of the C terminus to the N terminus. MS/MS spectrum of a tryptic fragment of cNW30 showing the ligation of the C-terminal motif (**LNTQLPGT**G-His_6) to the N-terminal residues (**GSTFSK**). Expected masses for b and y ions along with the peptide sequence are listed in the table. The b and y ions that were identified in the MS/MS spectrum are highlighted in blue and red. The full amino acid sequence of linear NW30 is shown at the top.

NW50 (amino acid sequence after TEV cleavage)

GSTFSKLREQLGPVTQEFWDNLEKETEGLRQEMSKDLEEVKAKVQPYLDDFQKKWQE
EMELYRQKVEPLRAELQEGARQKLHELQEKLSPLGEEMRDRARAHVDALRTHLAPYSD
ELRQRLAARLEALKENGGARLAEYHAKATEHLSTLSEKAKPALEDLRQGLLPVLESFKV
SFLSALEEYTKKLNTQGTPVTQEFWDNLEKETEGLRQEMSKDLEEVKAKVQPYLDDFQ
KKWQEEMELYRQKVEPLRAELQEGARQKLHELQEKLSPLGEEMRDRARAHVDALRTH
LAPYSDELRQRLAARLEALKENGGARLAEYHAKATEHLSTLSEKAKPALEDLRQGLLPV
LESFKVSFLSALEEYTKKLNTQGTPVTQEFWDNLEKETEGLRQEMSKDLEEVKAKVQP
YLDDFQKKWQEEMELYRQKVEPLRAELQEGARQKLHELQEKLSPLGEEMRDRARAHV
DALRTHLAPYSDELRQRLAARLEALKENGGARLAEYHAKATEHLSTLSEKAKPALEDLR
QGLLPVLESFKVSFLSALEEYTKKLNTQGTPVTQEFWDNLEKETEGLRQEMSKDLEEV
KAKVQPYLDDFQKKWQEEMELYRQKVEPLRAELQEGARQKLHELQEKLSPLGEEMRD
RARAHVDALRTHLAPYSDELRQRLAARLEALKENGGARLAEYHAKATEHLSTLSEKAKP
ALEDLRQGLLPVLESFKVSFLSALEEYTKKLNTQGTPVTQEFWDNLEKETEGLRQEMSK
DLEEVKAKVQPYLDDFQKKWQEEMELYRQKVEPLRAELQEGARQKLHELQEKLSPLG
EEMRDRARAHVDALRTHLAPYSDELRQRLAARLEALKENGGARLAEYHAKATEHLSTL
SEKAKPALEDLRQGLLPVLESFKVSFLSALEEYTKK**LNTQLPGT**GAAALEHHHHHH

Peptide sequence

LNTQLPGTGSTFSK



seq	#	b	y	#
L	1	114.091	---	14
N	2	228.134	1337.670	13
T	3	329.182	1223.627	12
Q	4	457.241	1122.579	11
L	5	570.325	994.520	10
P	6	667.377	881.436	9
G	7	724.399	784.384	8
T	8	825.446	727.362	7
G	9	882.468	626.314	6
S	10	969.500	569.293	5
T	11	1070.548	482.261	4
F	12	1217.616	381.213	3
S	13	1304.648	234.145	2
K	14	----	147.113	1

Supplementary Figure 13: Characterization of cNW50 by MS/MS. Spectrum of a tryptic fragment of cNW50 showing the ligation of the C-terminal motif (**LNTQLPGT**G-His_6) to the N-terminal residues (**GSTFSK**). Expected masses for b and y ions along with the peptide sequence are listed in the table. The b and y ions that were positively identified in the MS/MS spectrum are highlighted in blue and red. The amino acid sequence of linear NW50 (after TEV cleavage) is shown at the top.

	Mass, Da			
	Linear (calculated)	Linear (observed)	Circularized (Calculated)	Circularized (Observed)
NW9	21191.8	21192.8	19838.4	19838.5
NW11	23747.7	23752.4	22394.3	22398.1
NW30	65506.9	65520.6	64153.5	64162.4
NW50	107266.1	107304.7	105912.7	105940.7

Supplementary Table 1. Characterization of intact NW proteins by mass spectrometry.

B

Appendix B: "Recent developments in solution nuclear magnetic resonance (NMR)-based molecular biology"

B.1 Clinical significance – The awesome power of NMR

Visualizing post-translational modifications, conformations, and interaction surfaces of protein structures at atomic resolution underpins the development of novel therapeutics to combat disease. As computational resources expand, in silico calculations

coupled with experimentally-derived structures and functional assays has led to an explosion in structure based drug design (SBDD) with several compounds in clinical trials. It is increasingly clear that 'hidden' transition-state structures along activation trajectories can be harnessed to develop novel classes of allosteric inhibitors. The goal of this mini-review is to empower the clinical researcher with a general knowledge of the strengths and weaknesses of nuclear magnetic resonance (NMR) spectroscopy in molecular medicine. Although NMR can determine protein structures at atomic resolution, its unrivaled strength lies in sensing subtle changes in a nuclei's chemical environment as a result of intrinsic conformational dynamics, solution conditions, and binding interactions. These can be recorded at atomic resolution, without explicit structure determination, and then incorporated with static structures or molecular dynamics simulations to produce a complete biological picture.

B.2 THE NICHE OF NMR IN MODERN STRUCTURAL BIOLOGY

Static structural representations subconciously train scientists to envision a rigid protein architecture. However, macromolecues are exceptionally acrobatic possessing fast, local fluctuations and slow, concerted structural motions (Fig. 1). Therefore, a more accurate depiciton is a series of stuctures where the kinetics, thermodynamics, and function of each conformation varies significantly within a statistical ensemble. Understanding how these dynamic molecules behave in solution with atomic resolution is essential for deciphering their roles in biological processes such as ligand binding and protein-protein interactions. A classic example is how oxygen accesses the cavity of myoglobin – whose

crystal structures reveal no channels for the ligand to enter the heme-binding site [1]. NMR spectroscopy is the only technique available to probe these dynamics, at atomic resolution, on timescales ranging over 15 orders of magnitude. As illustrated in Figure 1, there are NMR experiments to look at backbone vibrations, side-chain motions, secondary structure movements, domain rearrangements, and even folding all under physiological solution-state conditions [2-5]. This unique perspective from NMR spectroscopy lends major contributions to understanding the biological impact of inherently flexible systems such as intrinsically disordered proteins (IDPs) and integral membrane proteins (IMPs). It should be emphasized that this mini-review is written for the NMR novice. We aim to give an introduction for modern NMR applications to clinically relevant systems such as structure-based drug design, IDPs and IMPs. Recommendations to indepth reviews are found throughout the text for the interested researcher.

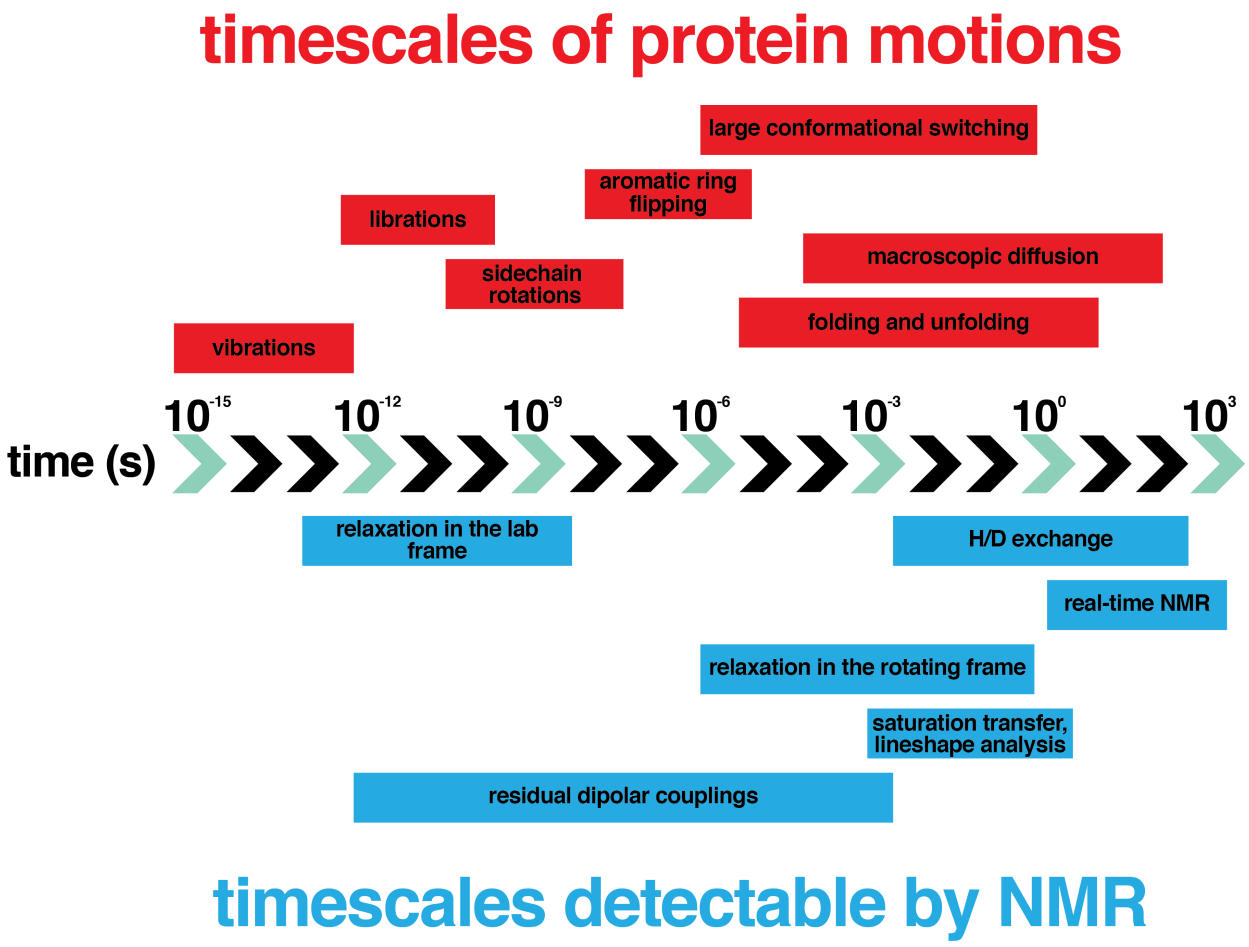


Fig. 1. NMR is capable of monitoring molecular motions over 15 orders of magnitude at atomic resolution by probing a multitude of properties with a variety of experiments.

B.3 THE BUSINESS END OF NMR: CHEMICAL SHIFTS

Spectroscopy concerns the interaction between light and matter. NMR spectroscopy probes nuclei that possess non-zero quantum spin angular momentum, or simply *spin*. Biomolecular NMR typically focuses on spin ½ nuclei which produce easily interpretable spectra. The primary atomic constituents of biomolecules: carbon, nitrogen, and hydrogen all possess a stable spin ½ isotope (e.g. – ^{13}C, ^{15}N, and ^{1}H). When placed in an external magnetic field, electromagnetically irradiated nuclei resonate at a frequency dictated by

their chemical environment – what atoms is it bonded to or spatially nearby. In other words, the resonance frequency of each nuclei is modulated by its primary, secondary, tertiary and quaternary structure. This unique chemical dependence enables structure determination as well as the ability to monitor changes in motion, and even map atom-specific changes upon ligand binding. For practical purposes, spectroscopists express resonance frequency in terms of *chemical shift* (units of parts per million; ppm) – which is the resonance frequency scaled by magnetic field strength. Chemical shifts are the most easily accessible NMR parameter that can be measured with high accuracy. As illustrated in Figure 2a, the unique chemistry of each functional group causes them to occupy a particular region of the spectrum. Subsequently, each amino acid gives rise to distinct chemical shift ranges assuming the residue exists in isolation, or an unstructured conformation, known as a random coil chemical shift (Fig. 2b). These chemical shifts are further altered by the nuclei's oxidation state, orientation of covalent bonds, isotope of neighboring atoms, hydrogen bonding, and proximity to groups with high magnetic susceptibility such as carbonyl, aromatic or paramagnetic ions (metals, lanthanides, etc.). In fact, a common technique involves intentional incorporation of paramagnetic ions to modulate the chemical shift for unique structure and dynamic information (as reviewed in [6]).

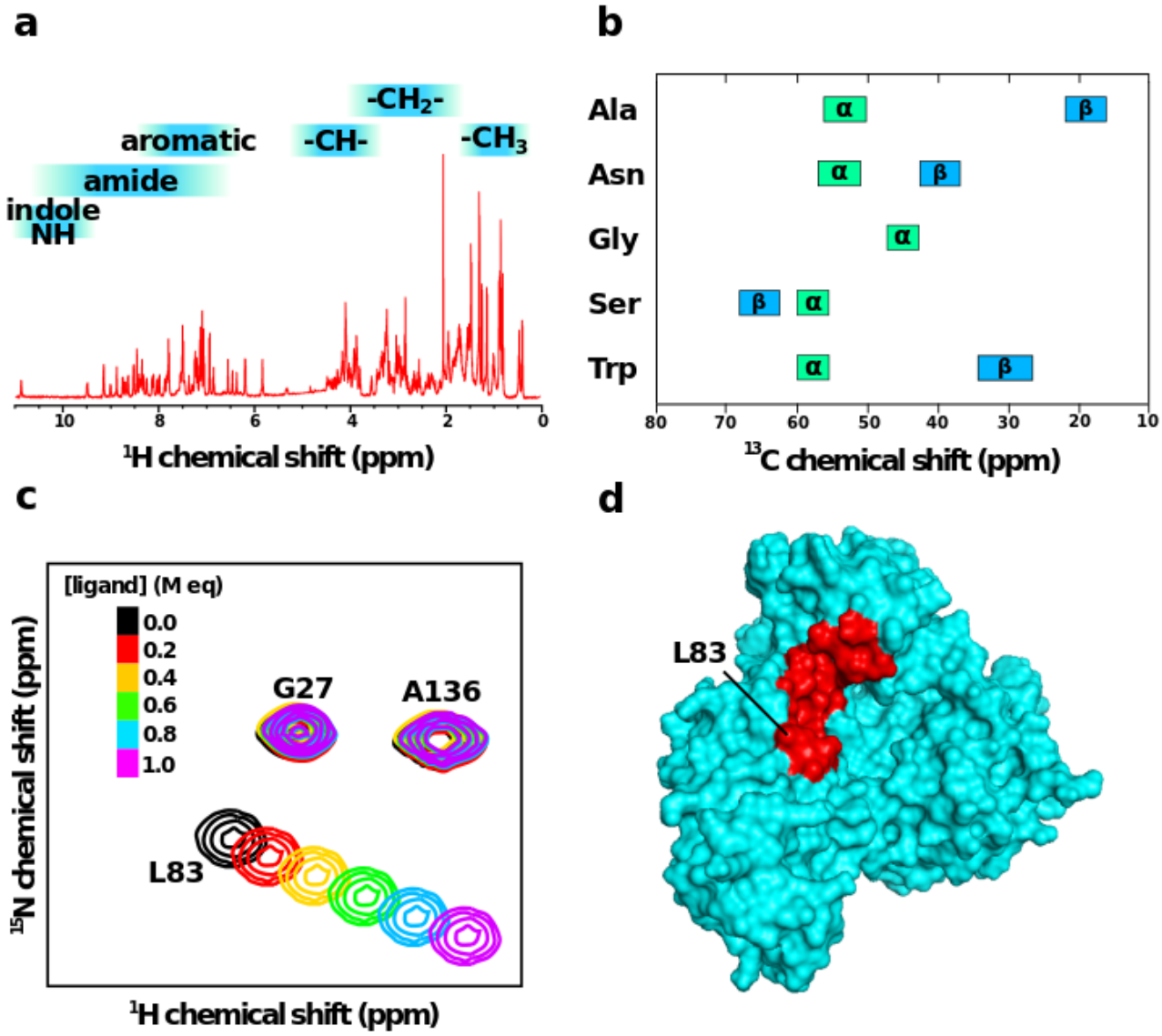


Fig. 2. NMR is highly sensitive to chemical environment based on primary, secondary, tertiary and quaternary protein structure. a) Each atom within a functional group possesses a distinct chemical shift range as illustrated for ^{1}H nuclei. b) This translates to unique values for each atoms on a per residue basis as illustrated for ^{13}C groups of select amino acids. A complete list of predicted chemical shift values is available online at the Biological Magnetic Resonance Data Bank [53]. c) A series of 2D spectra collected with increasing ligand concentrations (colored from black to magenta); only L83 is specifically located at the binding interface as demonstrated by its large chemical shift perturbations. The complex affinity can be quantified by fitting changes in chemical shift, and/or peak intensity, as a function of ligand concentration to an appropriate binding equation [28]. d) Finally, the residues that experience chemical shift changes can be highlighted on the surface of structure or homology model to map the binding interface.

It was not until the early 1990s that the NMR structural database reached a sufficient sample size for 'chemical shift homology' searches to predict secondary structures for sequence alignment, fold classification, protein visualization and homology modeling. Qualitative secondary structure predictions first measured the difference between a residue's chemical shift from the random coil value, known as the secondary chemical shift, and plotted it as a function of primary sequence. Initial estimates focused on how $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ secondary chemical shifts could discriminate the electronic distribution of beta-strand and alpha-helical regions [7] that formed the Chemical Shift Index (CSI) method [8]. CSI then paved the way for more robust methods that utilize ^{15}N, $^{1}H^{\alpha}$, $^{13}C'$, $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts in numerous web-based or downloadable programs to estimate secondary structure based on chemical shifts, with DANGLE and PsiCSI reporting the highest fidelity [9, 10]. The Random Coil Index (RCI), developed by the Wishart group, goes further by using chemical shifts alone to accurately quantify protein flexibility [11]. Without explicit knowledge of the 3D structure or additional NMR experiments, the chemical shift values of nuclei that sample multiple conformations on a sub-millisecond scale are population averaged. Comparison of these data with molecular dynamics simulations or crystallographic B-factors can quickly identify whether the B-factor is a result of static conformational heterogeneity or true conformational dynamics. Backbone chemical shift 'homology searches' are also exploited to predict backbone ϕ and ψ dihedral angles within $\leq 30°$ for at least 85% of sites; the most popular programs are PREDITOR, TALOS+, and DANGLE [10, 12, 13]. Stereospecific methyl labeling has lead to the development of a simple formula for sidechain rotamers of Val ($\chi 1$), Leu ($\chi 2$),

Ile ($\chi 2$) and calculations of further polar and aromatic residues should be possible with larger structural datasets [14, 15].

Current structure determination procedures rely primarily on cataloging short-range (typically ≤5 Å) ^{1}H-^{1}H distance constraints to define the polypeptide architecture. The holy grail of NMR would be tertiary and quaternary structure calculations from chemical shifts alone. Since the year 2000, international structural genomics initiatives rapidly added datasets to the point that high-resolution structures can now be calculated solely from chemical shifts. The dominant methods, CHESSIRE, CS-ROSETTA, CS23D, rely on backbone chemical shift assignments alone but can faithfully calculate model structures with backbone RMSD < 1.5 Å from reference structures [16-18]. These model structures are currently limited to ~110 amino acids but inclusion of side chain chemical shifts, residual dipolar couplings (RDCs), and unassigned nuclear Overhauser effect (NOE) restraints has been successfully applied to proteins up to 200 amino acid proteins [19]. More recent iterations model oligomeric structures from chemical shift perturbations upon complex formation. Programs such as HADDOCK, CamDock and CS-ROSETTA sequentially, or simultaneously, combine docking and structure determination algorithms to produce tertiary structures from inputs of the free state structures [20-22].

B.4 APPLICATIONS

B.4.1 Solution NMR in drug discovery

In the last two decades, drug discovery's focus has shifted to screening fragment molecules (typically containing less than 20 atoms) for development into high-affinity lead

compounds. Initial fragments with optimal binding efficiencies (affinity v. number of atoms) can be further enhanced into lead-like molecules through the addition of chemical functionalities, and/or the linking of ligands that target disparate sites. However, fragments are inherently low-affinity (micromolar affinity) molecules that do not easily crystalize and are difficult to identify in functional assays. In contrast, NMR is not affinity limited and can directly determine binding constants during compound screening to rank them for optimization and lead selection. Binding kinetics directly determine accurate quantification of drug affinity by NMR (for detailed texts on NMR kinetics, see [23-25]). The particular exchange regime of a binding reaction can be readily identified by observing each peak throughout a titration; it is important to note that each atom can exhibit its own rate of exchange for a particular binding reaction.

First spearheaded by Fesik and colleagues, structure-activity relationships (SAR) by NMR [26] has matured into a moderate-throughput screening technology. Titrations can be performed in two different formats: 1) ligand-based or 2) target-based. Ligand-based screening relies on excess concentrations of molecular fragments and monitors chemical shift changes through one-dimensional spectra collected in minutes. A positive binding interaction alters the chemical shift and/or the line shape of a given resonance. Other experimental varieties monitor changes in the ligands' diffusion coefficient or only allow visualization of atoms within a particular distance of the target protein (such as saturation transfer difference spectroscopy). The high-resolution of small molecules allows batch screening of 5-10 molecules to improve throughput (for details we recommend this excellent review [27]). Ligand-based screening is further advantageous

as the target protein does not need to be isotopically-labeled and usually only studied in small concentrations. Conversely, target-based screening follows changes to on the protein side. Once the chemical shift assignments of the target protein are known, isotopically-enriched protein can be titrated with a putative ligand or binding partner and the resulting chemical shift changes in monitored by 2D correlation spectroscopy (Fig. 2c) to determine binding constants and stoichiometry [28]. Highlighting perturbed residues on the surface of a known protein structure, called *chemical shift mapping*, can be used to identify the binding pocket (Fig. 2d). 2D ^{1}H-^{15}N heteronuclear single-quantum coherence (HSQC) spectroscopy is the most common method for mapping backbone chemical shifts because every residue type (except proline) contains a highly sensitive probe to changes in local magnetic environment. HMQC [29] and TROSY [30] are alternative pulse schemes that supply equivalent spectral information and provide advantages in certain situations such as large molecular weight proteins. As most protein-ligand interactions are sidechain mediated, 2D ^{1}H-^{13}C HSQCs can provide additional information, but sidechain signals typically suffer from overlap in large target proteins. Thus methyl-selective ^{1}H,^{13}C-labeling of Ile, Leu, Ala and Val residues can significantly improve resolution without compromising sensitivity [31].

B.4.2 Intrinsically-disordered proteins (IDPs)

Since the initial discovery of functional proteins without a stable architecture it's been estimated that >30% of the human proteome constitute partially, or completely, unstructure polypeptides [32]. IDPs exist in an ensemble of high-energy states that may

kinetically sample globular structures and often only in the presence of specific binding partners. They evolved to maintain a level of solubility required for optimal function and have significant biological roles in signaling and regulation – in part due to their promiscuous binding activity. A large proportion of the peptides and proteins responsible for abherrant misfolding diseases are intrinsically disordered in their free soluble forms – amyloid-β in Alzheimer's [33], α-synuclein in parkinsons [34], and amylin in type II diabetes [35]. Whereas X-ray crystallography may obtain structures when IDPs adopt well-ordered bound states, NMR spectroscopy can report on both the bound-state and the apo, high-energy ensemble. Besides NMR, small-angle X-ray scattering (SAXS) is now increasingly used to characterize IDPs. Their high flexibility results in spectroscopic properties that are characteristic of small molecules, allowing for backbone chemical shift assignment even for high molecular weight proteins. Once assigned, one can monitor changes upon interaction with a particular ligand using chemical shift mapping as described in the SBDD section. Although NMR has monopolized their structural characterization, IDPs do present unique technical challenges that are beginning to be addressed by low-gamma nuclei (^{13}C and ^{15}N) detection experiments. In brief, using low-gamma nuclei instead of tranditional ^{1}H-detected experiments potentially results in larger signal dispersion, sensitivity and resolution. Interested readers are directed towards an excellent review by Takeuchi and colleagues [36].

B.4.3 Integral membrane proteins (IMPs) and nanodisc technology

Proteins embedded into plasma membranes are critical to the molecular mechanisms of life, responding to extracellular stimuli and communicating changes in the external environment to intracellular machinery. Some estimates suggest that IMPs comprise one third of the genome and nearly half of the current pharmacology targets [37, 38]. Despite being ubiquitous in healthy and pathological systems, structural studies of IMPs have been limited. Until recently, X-ray crystallography was unable to handle the fluid-like environment of lipid membranes. In the past decade, lipidic cubic phase-based studies have revolutionized crystallography of membrane proteins with more than 200 structures, nearly half of which were determined in the last five years [39]. In contrast, NMR spectroscopy has steadily contributed to IMP structural information - being limited primarily by isotope labeling, molecular weight and detergent stability. However, the size limitations have decreased significantly following the invention of TROSY methods and improved isotopic labeling schemes. At this point, a major limitation is the loss of structure and function that accompanies purification in detergent and removal of the native lipid environment. If a suitable detergent environment can be found, an additional challenge arises in characterizing protein-protein interactions which are frequently compromised by detergents. A promising strategy is to incorporate IMPs into discrete phospholipid-bilayer mimetics called nanodiscs.

Nanodiscs are an elegant solution created by Sligar and colleagues [40]. Modeled after high-density lipoprotein (HDL) particles, they comprise of a discoidal lipid bilayer stabilized by a polypeptide belt, termed a membrane scaffold protein (MSP). In 2013, Hagn et al. modified the MSP primary sequence to reduce the disc diameter and

subsequently rotational correlation times - critical to making nanodiscs accessible to the NMR community [41]. The nanodiscs range from 4-10 nm (approximately equivalent to 52-124 kDa) and dominate the IMP's relaxation properties assuming there are minimal extramembrane portions. In 2017, an empty nanodisc structure was determined by NMR spectroscopy [42]. As suggested by previous bioinformatics and biochemical experiments, MSP dimerizes and adopts a series of alpha-helical repeats that wrap around the lipids in an anti-parallel alignment. One suggested limitation of nanodisc technology is inherent heterogeneity of the disc diameter and difficulty in controlling the stoichiometry of embedded IMPs. Recently, Wagner and colleagues have modified the MSP construct to be covalently-circularized using sortase technology to control 95% of discs to within a 1 nm size distribution [43]. MSP-based nanodiscs rely on detergent molecules to solubilize protein:lipid mixtures prior to disc formation. An alternative styrene maleic acid polymer is capable of directly incorporating IMPs without detergent solubilization and may gain traction in the future once disc diameters can be faithfully formed below 15 nm [44, 45].

B.5 TECHNICAL CONSIDERATIONS IN BRIEF

Sensitivity. The magnitude of a NMR signal is dependent on the population difference between high and low energy states. The population difference is directly proportional to the static magnetic field strength – which is the primary reason for building higher-field spectrometers. For a given field strength, signals scale linearly with sample concentration and signal-to-noise (S/N) ratio scales as a $\sqrt{2}$. It is important to recognize that signal

strength reflects the total mole fraction of NMR-active spins rather than the molar concentration, which means the spectra of a 200 µl sample at 100 µM is effectively equivalent to a 400 µl sample at 50 µM (in the absence aggregation, oligomerization ,etc.). Samples volumes are typically around 300-500 µl. However, smaller tubes are available that reduce sample volume to 175-200 µl. Modern spectrometer hardware has reduced the minimum protein concentration to ≤15 µM (in 500 µl) for collecting a 2D spectrum within a couple hours, although higher concentrations will directly result in shorter experimental times. Samples are tolerant to a wide range of solution conditions and temperatures – limited mainly by high salt concentrations (special tubes should be used at > 200 mM salt). Temperatures are only limited by the solvent's freezing point and the spectrometer's hardware boundary; modern cold probes operate at temperatures between -40 °C and 80 °C.

Molecular weight. The details of spin physics are beyond the scope of this review; however, it is important for any scientists to have an idea of the limitations of each technique. For an accessible, yet thorough, text we recommend Understanding NMR Spectroscopy by James Keeler [46]. The maximum theoretical signal is proportional to the number of NMR-active spins within a sample. This signal is then modulated by exponential decay functions that reflect *relaxation*, or the return of a spin system to thermodynamic equilibrium. The relaxation source divides signal decay into two types: T1 (aka longitudinal or spin-lattice relaxation) and T2 (aka transverse or spin-spin relaxation). T2 relaxation is particularly important in biomolecular NMR spectroscopy – with the decay rate increasing as a function of molecular weight. Many researchers are quick to note that

T2 relaxation limits the application of NMR to low molecular weight systems. However, as we have detailed in this review – although structure determination becomes challenging at higher molecular weights, resonances are particular sensitive to changes in protein structure and dynamics without complete structure determination.

Isotope labeling. The major limitation of NMR spectroscopy is the requirement of NMR-active atoms, specifically spin-1/2 nuclei. Luckily for biological systems, hydrogen's most common isotope is spin-1/2 (^{1}H; 99.9% natural abundance); however, samples must be enriched with ^{15}N and ^{13}C-atoms as their natural abundance is ≤1%. The details of isotope enrichment depend on the specific recombinant protein expression system. In brief, E. coli cultures can be cheaply grown with simple isotopically labeled carbon and nitrogen sources (typically, glucose and ammonium chloride, respectively) whereas eukaryotic systems require supplementation with more costly labeled amino acids. Depending on yields, isotopic labeling can be prohibitively expensive. We recommend these articles for E. coli [47], yeast [47], insect cell [48], or mammalian expression systems [48]. As molecular weights increase (typically ≥ 30 kDa), non-exchangable hydrogens must be partially or fully deuterated (^{2}H) to improve T2 relaxation properties (as reviewed in [47]).

B.6 CONCLUSION

NMR has always been chemists' default technique for robust small-molecule structure determination, but its role in the biomolecular arena has been slightly more contested. X-ray crystallography remains the workhorse of structural biology – providing 90% of the Protein Data Bank's (PDB) entries. Although crystallization has contributed priceless

amounts of architectural information, it inherently stabilizes the lowest-energy conformation at the expense of eliminating protein conformational dynamics. It requires a variety of schemes to trap the reaction substates or coax all molecules within the crystal to adopt the same substate conformation. Even then, the crystallization procedure itself has the potential to modify the structure, or introduce artifacts in the form of oligomerization, non-specific protein-ligand interactions, or occlusion of binding interfaces. In contrast, solution-state NMR spectroscopy can be performed at physiological temperatures and buffer conditions. Experiments exist to quantitatively probe motions over all timescales from picoseconds to days, and most of these experiments provide atomic resolution information about conformations without the need to determine the complete structure. One glaring absence in this review is the description of NMR experiments probing conformational motions. Even cursory explanation of NMR-based dynamics requires in-depth discussion of spin physics for numerous experiments (Fig. 1). Interested readers are directed to the texts Spin Dynamics [25], Structural Biology: Practical NMR Applications [49] and Protein NMR Spectroscopy [50]. Protein structure determination methods have improved considerably over the past two decades as a result of improved hardware, software and isotopic labeling strategies. Hardware advancements such as cryogenic probes increase the sensitivity allowing for low concentration studies. Data processing advancements such as non-uniform sampling (NUS) allow faster data collection or the allocation of a given time to collect more data for higher resolution or signal:noise ratio. Novel pulse sequences, especially transverse relaxation-optimized spectroscopy (TROSY) continue to push the molecular weight limits

to nearly 1 megadalton complexes [51, 52]. NMR spectroscopy will continue to be a valuable tool for not only exploring fundamental biological principles, but also understanding the molecular details of a variety of diseases and provide critical aid for designing robust medical therapies.

References

[1] E. Wallin and G. von Heijne, "Genome-wide analysis of integral membrane proteins from eubacterial, archaen, and eukaryotic organisms," *Protein Science,* pp. 1029-1038, 1998.

[2] J. Overington, B. Al-Lazikani and A. Hopkins, "How many drug targets are there?," *Nature Reviews of Drug Discovery,* pp. 993-996, 2006.

[3] J. Deisenhofer, O. Epp, K. Miki, R. Huber and H. Michel, "Structure of the protein subunits in the photosynthetic reaction centre of Rhosopsuedomonas viridis at 3A resolution," *Nature,* pp. 618-624, 1985.

[4] I. Moraes, G. Evans, J. Sanchez-Weatherby, S. Newstead and P. Shaw Steward, "Membrane protein structure determination -- The next generation," *Biochimica et Biophysica Acta,* pp. 78-87, 2014.

[5] Y. Cheng, "Single-particle cryo-EM at crystallographic resolution," *Cell,* pp. 450-457, 2015.

[6] X. Li, P. Mooney, S. Zheng, C. Booth, M. Braunfield, S. Gubbens, D. Agard and Y. Cheng, "Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM," *Nature Methods,* pp. 584-590, 2013.

[7] A. Brilot, J. Chen, A. Cheng, J. Pan, S. Harrison, C. Potter, B. Carragher, R. Henderson and N. Grigorierff, "Beam-induced motion of vitrified specimen on holey carbon film," *Journal of Structural Biology,* pp. 630-637, 2012.

[8] M. Campbell, A. Cheng, A. Brilot, A. Moeller, D. Lyumkis, D. Veesler, J. Pan, S. Harrison, C. Potter, B. Carragher and N. Grigorieff, "Movies of ice-embedded particles enhance resolution in electron cryo-microscopy," *Structure,* pp. 1823-1828, 2012.

[9] D. Lyumkis, A. Brilot, D. Theobald and N. Grigorieff, "Likelihood-based classification of cryo-EM images using FREALIGN," *Journal of Structural Biology,* pp. 377-388, 2013.

[10] R. Danev and W. Baumeister, "Cryo-EM single particle analysis with the Volta phase plate," *eLife,* 2016.

[11] F. Hagn, M. Etzkorn, T. Raschle and G. Wagner, "Optimized phospholipid bilayer nanodiscs facilitate high-resolution structure determination of membrane proteins," *Journal of American Chemical Society,* pp. 1919-1925, 2013.

[12] M. Nasr, D. Baptista, M. Strauss, J. Zhen-Yu, S. Grigoriu, S. Huser, A. Pluckthun, F. Hagn, T. Walz, J. Hogle and G. Wagner, "Covalently circularized nanodiscs for studying membrane proteins and viral entry," *Nature Methods,* pp. 49-52, 2017.

[13] S. Wagner, M. Klepsch, S. Schlegel, A. Appel, R. Draheim, M. Tarry, M. Hogbom, K. van Wijk, D. Slotbloom, J. Persson and J. de Gier, "Tuning Escherichia coli for membrane protein overexpression," *Proceedings of the National Academy of Science,* pp. 14371-14376, 2008.

[14] G. Prive, "Detergents for the stabilization and crystallization of membrane proteins," *Methods,* pp. 388-397, 2007.

[15] P. Chae, S. Rasmussen, R. Rana, K. Gotfryd, R. Chandra, M. Goren, A. Kruse, S. Nurva, C. Loland, Y. Pierre, D. Drew, J. Popot, D. Picot, B. Fox, L. Guan, U. Gether, B. Byrne, B. Kobilka and S. Gellman, "Maltose-neopentyl glycol (MNG) amphiphiles for solubilization, stabilization, and crystallization of membrane proteins," *Nature Methods,* pp. 1003-1008, 2010.

[16] E. Serebryana, G. Zhu and E. Yan, "Artificial membrane-like environments for in vitro studies of purified G-protein coupled receptors," *Biochimica et Biophysica Acta,* pp. 225-233, 2012.

[17] E. Chun, A. Thompson, W. Liu, C. Roth, M. Griffith, V. Katritch, J. Kunken, F. Xu, V. Cherezov, M. Hanson and R. Stevens, "Fusion partner toolchest for stabilization and crystallization of G protein-coupled receptors," *Structure,* pp. 967-976, 2012.

[18] R. Stevens, S. Yokoyama and I. Wilson, "Global efforts in structural genomics," *Science Signaling,* p. 89, 2001.

[19] E. Duke and L. Johson, "Macromolecular crystallography at synchotron radiation sources: current status and future developments," *Proceedings of the Royal Society,* pp. 3421-3452, 2010.

[20] L. Baker and J. Rubenstein, "Radiation damage in electron cryomicroscopy," *Methods of Enzymology,* 2010.

[21] W. Van Horn, H. Kim, C. Ellis, A. Hadziselimovic, E. Sulistijo, M. Karra, C. Tian, F. Sonnichsen and C. Sanders, "Solution nuclear magnetic resonance structure of membrane-integral diacylglycerol kinase," *Science,* pp. 1726-1729.

[22] A. Gautier, H. Mott, M. Bostock, J. Kirkpatrick and D. Nietlispach, "Structure determination of the seven-helix transmembrane receptor sensory rhodopsin II by solution NMR spectrsocopy," *Nature Structural and Molecular Biology,* pp. 768-774, 2010.

[23] K. Oxenoid and J. Chou, "The structure of phospholamban pentamer reveals a channel-like architecture in membranes," *Proceedings of the National Academy of Science,* pp. 10870-10875, 2005.

[24] S. Hiller, R. Garces, T. Malia, V. Orekhov, M. Colombini and G. Wagner, "Solution structure of the integral human membrane protein VDAC-1 in detergent micelles," *Science,* pp. 1206-1210.

[25] J. Williamson, S. Cho, J. Ye, J. Collet, J. Beckwith and J. Chou, "Structure and function of the transmembrane electron transporter CcdA," *Nature Structural and Molecular Biology,* pp. 809-814, 2015.

[26] A. Palmer, C. Kroenke and J. Loria, "Nuclear magnetic resonance methods for quantifying microsecond-to-millisecond motions in biological macromolecules," *Methods in Enzymology,* pp. 204-238, 2001.

[27] T. Cross, D. Murray and A. Watts, "Helican membrane protein conformations and their environment," *European Journal of Biophysics,* pp. 731-755, 2013.

[28] T. Bayburt, Y. Grinkova and S. Sligar, "Self-assembly of discoidal phospholipid bilayer nanoparticles with membrane scaffold proteins," *NanoLetters,* pp. 853-856.

[29] C. Brouillette, G. Anantharamaiah, J. Engler and D. Borhani, "Structural models of human apolipoprotein A-I: a critical analysis and review," *Biochimica et Biophysica Acta,* pp. 1388-1981, 2001.

[30] I. Denisov, Y. Grinkova, A. Lazarides and S. Sligar, "Directed self-assembly of monodisperse phospholipid bilayer nanodiscs with controlled size," *Journal of the American Chemical Society,* pp. 3477-3487, 2004.

[31] T. Bayburt and S. Sligar, "Membrane protein assembly into nanodisks," *FEBS Letters,* pp. 1721-1727, 2010.

[32] I. Denisov and S. Sliger, "Nanodiscs in membrane biochemistry and biophysics," *Chemical Reviews,* pp. 4669-4713, 2017.

[33] N. Shirzad-Wasei, J. van Oostrum, P. Bovee-Geurts, L. Kusters, G. Bosman and W. Degrip, "Rapid transfer of overexpressed integral membrane protein from the host membrane into soluble lipid nanodiscs without previous purification," *Biological Chemistry,* pp. 903-916, 2015.

[34] R. Efremov, A. Leitner, R. Aebersold and S. Raunser, "Architecture and conformational switch mechanism of the ryanodine receptor," *Nature,* pp. 39-43, 2014.

[35] R. Efremov, C. Gatsogiannis and S. Raunser, "Lipid nanodiscs as a tool for high-resolution structure determination of membrane proteins by single-particle cryo-EM," *Methods in Enzymology,* pp. 1-30, 2017.

[36] Y. Gao, E. Cao, D. Julius and Y. Cheng, "TRPV1 structures in nanodiscs reveal mechanisms of ligand and lipid action," *Nature,* pp. 347-351, 2016.

[37] C. Gatsogiannis, F. Merina, D. Prumbaum, D. Roderer, F. Leidreiter, D. Meunsch and S. Raunser, "Membrane insertion of a Tc toxin in near-atomic detail," *Natural Structural and Molecular Biology,* pp. 884-890, 2016.

[38] R. Kumar, L. Zhu, H. Idbord, O. Raadmark, P. Jakobsson, A. Rinaldo-Mathis, H. Herbert and Jegerschoeld, C, "Structural and functional analysis of calcium ion mediated binding of 5-lipoxygenase to nanodiscs," *PLoS One,* 2016.

[39] L. Daury, F. Orange, J. Taveau, A. Verchere, L. Monlezun, C. Gounou, R. Marreddy, M. Picard, I. Broutin, K. Pos and L. Olivier, "Tripartite assembly of RND multidrug efflux pumps," *Nature Communications,* 2016.

[40] D. Matthies, O. Dalmas, M. Borgnia, P. Dominik, A. Merk, P. Rao, B. Reddy, S. Islam, A. Bartesaghi, E. Perozo and S. Subramaniam, "Cryo-EM structures of the magnesium channel CorA reveal symmetry break upon gating," *Cell,* pp. 747-756, 2016.

[41] J. Frauenfeld, J. Gumbart, E. van der Sluis, S. Funes, M. Gartmann, B. Beatrix, T. Mielke, O. Berninghausen, T. Becker, K. Schulten and R. Beckmann, "Cryo-EM structure of the ribosome-SecYE complex in the membrane environment," *Nature Structural and Molecular Biology,* pp. 614-621, 2011.

[42] M. Mazhab-Jafari, C. Marshall, M. Smith, G. Gasmi-Seabrook, P. Stathopulos, F. Inagaki, L. Kay, B. Neel and M. Ikura, "Oncogenic and RASopathy-associated K-RAS mutations relieve membrane-dependent occlusion of the effector-binding site," *Proceedings of the National Academy of Sciences,* pp. 6625-6630, 2015.

[43] K. Mineev, S. Goncharuk, P. Kuzmichev, M. Vilar and A. Arseniev, "NMR dynamics of transmembrane and intracellular domains of p75NTR in lipid-protein nanodiscs," *Biophysical Journal,* pp. 772-782, 2015.

[44] I. Kucharska, T. Edrington, B. Liang and L. Tamm, "Optimizing nanodiscs and bicelles for solution NMR studies of two beta-barrel membrane proteins," *Journal of Biomolecular NMR,* pp. 261-274, 2015.

[45] L. Morgado, K. Zeth, B. Burmann, T. Maier and S. Hiller, "Characterization of the insertase BamA in three different membrane mimetics by solution NMR spectroscopy," *Journal of Biomolecular NMR,* pp. 333-345, 2015.

[46] Y. Yao, D. Nisan, A. Antignani, A. Barnes, N. Tjandra, R. Youle, F. Marassi and L. Fujimoto, "Characterization of the membrane-inserted C-terminus of cytoprotective BCL-XL," *Protein Expression and Purification,* pp. 56-63, 2016.

[47] Y. Yao, L. Fujimoto, N. Hirshman, A. Bobkov, A. Antignani, R. Youle and F. Marassi, "Conformation of BCL-XL upon membrane integration," *Journal of Molecular Biology,* Vols. 2262-2270, 2015.

[48] Alvarez, FJD, Orelle, C, Y. Huang, R. Bajaj, R. Everly, C. Klug and A. Davidson, "Full engagement of liganded maltose-binding protein stabilizes a semi-open ATP-binding cassette dimer in the maltose transporter," *Molecular Microbiology,* pp. 878-894, 2015.

[49] Y. Kang, X. Zhou, X. Gao, Y. He, W. Liu, A. Ischenko, A. Barty, T. White, O. Yefanov, G. Han, Q. Xu, P. de Waal, J. Ke, M. Tan, C. Zhang and A. Moeller, "Crystal structure of rhodopsin bound to arrestin by femtosecond X-ray laser".

[50] J. Shin, X. Lou, D. Kweon and Y. Shin, "Multiple conformations of a single SNAREpin between two nanodiscs membranes reveal diverse pre-fusion states," *Biochemical Journal,* pp. 95-102, 2014.

[51] Y. Kofuku, U. Takumi, J. Okude, Y. Shiraishi, K. Kondo, T. Mizumura, S. Suzuki and I. Shimada, "Functional dynamics of deuterated B2-adrenergic receptor in lipid bilayers revealed by NMR spectroscopy," *Angewandte Chemie,* pp. 13376-13379, 2014.

[52] J. Okude, U. Takumi, Y. Kofuku, M. Sato, N. Nobuyama, K. Kondo, Y. Shiraishi, T. Mizumura, K. Onishi, M. Natsume, M. Maeda, H. Tsujishita, T. Kuranaga, M. Inoue and I. Shimada, "Identification of a conformational equilibrium that determines the efficacy and functional selectivity of the u-opioid receptor," *Angewandte Chemie,* pp. 15771-15776, 2015.

[53] C. Yoshiura, T. Ueda, Y. Kofuku, M. Matsumoto, J. Okude, K. Kondo, Y. Shiraishi and I. Shimada, "Journal of Biomolecular NMR," *2015,* pp. 333-340.

[54] T. Raschle, S. Hiller, T. Yu, A. Rice, T. Walz and G. Wagner, "Structural and functional characterization of the integral membrane protein VDAC-1 in lipid bilayer nanodiscs," *Journal of the American Chemical Society,* pp. 10Th 7777-17779, 2009.

[55] G. Wider and K. Wuthrich, "NMR spectroscopy of large molcules and multimolecular assemblies in solution," *Current Opinion in Structural Biology,* pp. 594-601, 1999.

[56] S. Bibow, Y. Polyhach, C. Eichmann, C. Chi, J. Kowal, S. Albiez, R. McLeod, H. Stahlberg, G. Jeschke, P. Guntert and R. Riek, "Solution structure of discoidal high-density lipoprotein particles with shortened apolipoprotein A-I," *Nature Structural and Molecular Biology,* pp. 187-196, 2017.

[57] J. Schredelseker, A. Paz, C. Lopez, C. Altenbach, C. Leung, M. Drexler, J. Chen, W. Hubbell and J. Abramson, "High resolution structure and double electron-

electron resonance of the zebrafish voltage-dependent anion channel reveal an oligomeric population," *Journal of Biological Chemistry,* pp. 12566-12577, 2014.

[58] T. Bayburt, A. Leitz, G. Xie, D. Oprian and S. Sligar, "Transducin activation by nanoscale lipid bilayers containing one and two rhodopsins," *Journal of Biological Chemistry,* pp. 14875-14881, 2007.

[59] O. Schneewind, P. Model and V. Fischetti, "Sorting of protein A to the staphylococcal cell wall," *Cell,* pp. 267-281, 1992.

[60] H. Ton-That, G. Liu, S. Mazmanian, K. Faull and O. Schneewind, "Purification and characterization of sortase, the transpeptidase that cleaves surface proteins of staphylococcus aureaus at the LPXTG motif," *Proceedings of the National Academy of Sciences,* pp. 12424-12429, 1999.

[61] D. McCafferty and J. Melvin, "Sortases," *Handbook of Proteolytic Enzymes,* pp. 2459-2465, 2013.

[62] K. Clancy, J. Melvin and D. McCafferty, "Sortase transpeptidases: Insights into mechanism, substrate, specificity, and inhibition," *Peptide Science,* pp. 385-396, 2010.

[63] A. Maresso, R. Wu, J. Kern, R. Zhang, D. Janik, D. Missiakas, M. Duban, A. Joachimiak and O. Schneewind, "Activation of inhibitors by sortase triggers irreversible modification of the active site," *Journal of Biological Chemistry,* pp. 23129-23139, 2007.

[64] R. Palmiter, "Regulation of protein synthesis in chick ovidecut," *Journal of Biological Chemistry,* 1972.

[65] A. Hinnebusch, "The scanning mechanism of eukaryotic translation initiation," *Annual Review of Biochemistry,* pp. 779-812, 2014.

[66] Y. Svitkin, A. Pause, A. Haghighat, S. Pyronnet, G. Witherall, G. Belsham and N. Sonenberg, "The requirement for eukaryotic initiation factor 4A (eIF4A) in translation is in direct proportion to the degree of mRNA 5' secondary structure," *RNA,* pp. 382-394, 2001.

[67] R. Duncan, S. Milburn and J. Hershey, "Regulated phosphorylation and low abundance of HeLa cell initiation factor eIF-4F suggest a role in translational control," *Journal of Biological Chemistry,* pp. 380-388, 1987.

[68] P. Shutz, T. Kalbert, S. van den Berg, R. Collins, L. Lehtio, M. Hogbom, L. Holmberg-Schiavone, W. Tempel, H. Park, M. Hammarstrom, M. Moche, A. Thorsell and H. Shuler, "Comparitive structural analysis of human DEAD-box RNA helicases," *PLoS ONE,* 2010.

[69] J. Caruthers, E. Johnson and D. McKay, "Crystal structure of yeast initiation factor 4A, a DEAD-box RNA helicase," *Proceedings of the National Academy of Sciences,* pp. 13080-13085, 2000.

[70] J. Chang, Y. Cho, S. Sohn, J. Choi, A. Kim, Y. Kim, S. Jang and Y. Cho, "Crystal structure of the eIF4A-PDCD4 complex," *Proceedings of the National Academy of Sciences,* pp. 3148-3153, 2009.

[71] G. Rogers, N. Richter, W. Lima and W. Merrick, "Modulation of the helicase activity of eIF4A by eIF4B, eIF4H, and eIF4F," *Journal of Biological Chemistry,* 2001.

[72] M. Oberer, A. Marintchev and G. Wagner, "Structural basis for the enhancement of eIF4A helicase activity," *Genes and Development,* 2005.

[73] B. Lamphear, R. Kirchweger, T. Skern and R. Rhoads, "Mapping of functional domains in eukaryotic protein synthesis initiation factor 4G (eIF4G) with picornaviral proteases," *Journal of Biological Chemistry,* 1995.

[74] H. Imataka and N. Sonenberg, "Human eukaryotic translation initiation factor 4G (eIF4G) possesses two separate and independent binding sites for eIF4A," *Molecular and Cellular Biology,* 1997.

[75] A. Marintchev, K. Edmonds, B. Marintcheva, E. Hendrickson, M. Oberer, C. Suzuki, B. Herdy, N. Sonenberg and G. Wagner, "Topology and regulation of the Human eIF4A/4G/4H helicase complex in translation initiation," *Cell,* pp. 447-460, 2009.

[76] K. Edmonds and G. Wagner, "1H, 13C, and 15N backbone and sidechain chemical shift assignments for the HEAT2 domain of human eIF4GI," *Biomolecular NMR Assignments,* 2015.

[77] M. Dobrikov, M. Shveygert, M. Brown and M. Gromeier, "Mitotic phosphorylation of eukaryotic initiation factor 4G (eIF4G) at Ser1232 by Cdk:cyclic B inhibits eIF4A helicase complex binding with RNA," *Molecular and Cellular Biology,* 2014.

[78] C. Nitsche and G. Otting, "Psuedocontact shifts in biomolecular NMR using paramagnetic metal tags," *Progress in Nuclear Magnetic Resonance Spectroscopy,* pp. 20-49, 2017.

[79] L. McGoldrick, A. Singh, K. Saotome, M. Yelshanskaya, E. Twomey, R. Grassucci and A. Sobolevsky, "Opening of the human epithelial calcium channel TRPV6," *Nature,* 2018.

[80] H. Bao, D. Das, N. Courtney, Y. Jiang, J. Briguglio, X. Lou, D. Roston, Q. Cui, B. Chanda and E. Chapman, "Dynamics and number of trans-SNARE complexes determine nascent peptide fusion pore properties," *Nature,* 2018.

[81] I. Chen, B. Dorr and D. Liu, "A general strategy for the evolution of bond-forming enzymes using yeast display," *Proceedings of the National Academy of Sciences,* 2011.