Exploring Protein Dynamics Through Paradigms of Chaperone Action and Structural Disorder

by

Rishav Mitra

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Molecular, Cellular and Developmental Biology)
in the University of Michigan
2023

Doctoral Committee:

Professor James C.A. Bardwell, Chair
Professor Matthew R. Chapman
Professor Ursula H. Jakob
Professor Lyle A. Simmons
Associate Professor Raymond C. Trievel

Rishav Mitra

mrishav@umich.edu

ORCID iD: 0000-0002-8898-7289

Dedication

This thesis is dedicated to the memory of a beloved mentor Dr. Amitabha Sarkar, who passed away during the COVID-19 pandemic.

Acknowledgments

I would like to thank my advisor Dr. James Bardwell for his support and guidance throughout my journey in graduate school. Jim's immense trust in my abilities as a scientist has always been a driving force in my research. Not only did he give me space to learn and grow, but his unwavering enthusiasm for research has also had a profound influence on my scientific journey. I must also mention his delightful sense of humor and genuine nature, which always brightened our conversations about life beyond the confines of the lab. Thank you, Jim, for everything!

Throughout my academic journey, I received valuable input from my committee members: Ursula Jakob, Lyle Simmons, Matthew Chapman, and Raymond Trievel. Their insightful comments sparked my scientific curiosity and I particularly enjoyed collaborating with Ursula and discussing science with her. Ursula has been an invaluable source of knowledge and essential guidance for enhancing my scientific understanding. During challenging times, she has served as a mentor, helping me grasp the broader context and significance of my work.

As someone with a biology background, my exposure to NMR was limited during my undergraduate and postgraduate studies. I am grateful for the invaluable training I received from Scott Showalter during my two-month laboratory visit at Penn State University, as well as from Debashish Sahu and Minli Xing from the BioNMR core at the University of Michigan. Thanks to their guidance, I was able to extensively utilize

solution state NMR spectroscopy in my current project. Olivia Fraser, Kevin Namitz, and Emery Usher from the Showalter lab have been a huge help. trained me to do NMR experiments and have guided me on several projects. Thank you, Scott, Deb and Minli!

I would also like to thank members of the Jakob lab, especially Kathrin Ulrich, Akash Rai, Bryndon Oleson, and Jian Guan for creating a fun and supportive work environment. I would like to thank my coworkers in the Bardwell lab who have helped me in many ways over the past six years. The current group of undergrads, Matthew Crotteau, Harry Yang, Nathan Clark, and Xiaomeng (Steven) Liu, have been like Luke Skywalkers, and I have felt like Yoda in their scientific journey. Words, truly, fall short when it comes to conveying my gratitude towards my entire cohort, especially Ritvija Agrawal. I am also grateful to Traci Banjanin, Ke Wan, Sylvia Widjaja, and May Tsoi who are integral parts of the fabric of the Bardwell lab and make lab research run smoothly and at full momentum. Traci has always been the source of wisdom and an inspiration for how to be a compassionate human being. Nothing I did would have been possible without all the 99.999% pure proteins that Ke Wan has purified and I am grateful for that.

Expressing my gratitude to my dear friends Mark Dulchavsky, Ben Meinen, Vikramjit Lahiri, Bineet Dash, Janak Bhattarai, Shilpa Padmanaban, Lotte van den Goor, Sania Mancer, Neelanjana Sarkar, Debopriya Chakraborty, and Kevin Wu feels like an immense task. A few lines or even a book wouldn't suffice to convey the depth of my appreciation for each of them. Every moment we shared together, whether in person or through endless hours on the phone, is a memory I'll cherish forever. Mark and Bineet, I owe you more than you would ever know. I am forever grateful for your

unwavering friendship and the countless ways you have impacted my life. You have served as my guiding stars, helping me find my way and navigate through the challenges that life has thrown at me.

I express my heartfelt gratitude to my family, who have been unwavering pillars of support throughout my journey. It feels inadequate to simply thank my parents for "everything" because their influence is intertwined with every aspect of my existence. My mother's unwavering belief in my abilities motivated me to pursue a PhD, while my father's encouragement paved the way for me to embark on this remarkable adventure. I am truly fortunate to have been raised by parents who placed such a high value on education and recognized its profound impact on my life. Additionally, I am grateful to my brother, cousins, uncle, and aunt, who have kept me connected to my roots and evoke a sense of longing for home.

Table of Contents

List of Tables

List of Figures

Abstract

Protein folding and dynamics shape all aspects of biology. Far from being rigid structures, proteins are intrinsically dynamic and undergo hierarchical motions at different time scales. Models help build frameworks to understand complex biological phenomena. In particular, the cross-disciplinary fields of protein folding, and dynamics have benefited immensely by employing model proteins to establish mechanistic paradigms and theoretical models that collectively shape our ways of thinking about the biophysics of proteins. In this thesis, I have used molecular chaperones and intrinsically disordered protein regions as models to investigate how structural dynamics regulate the biological activity of proteins. Molecular chaperones assist in protein folding and prevent protein aggregation. The first part of my thesis explores the mechanism of action of the small ATP-independent chaperone Spy on a topologically complex client, apoflavodoxin. Previous studies have established the dual functions of Spy as a “folding-while-bound” chaperone and as a “holdase” that prevents aggregation. My findings show that the molecular determinant of Spy’s mechanism of action is its affinity for unfolded conformations of the client. Weak affinity interactions with various folding states of the client enables it to fold to its native state while being continuously held by Spy. However, when Spy binds too tightly to (partially) unfolded states, it acts as a kinetic trap that inhibits folding-while-bound. The second part of my thesis involves characterizing the structure and dynamics of the disordered protein SERF in the context of its RNA-binding properties. In my research, I investigate the interplay between the

structural dynamics of SERF and its interaction with RNA, both in dilute solution and within phase-separated RNA-protein condensates. I have developed and characterized an *in vitro* model RNA-protein complex comprising SERF and the model helix-junction-helix motif found in the 5’ end of the human immunodeficiency virus type 1 (HIV-1) transactivating response region (TAR) RNA. The conserved N-terminal segment of SERF is disordered and interacts with RNA. Upon binding TAR, SERF undergoes global compaction, leading to slower structural dynamics. SERF and TAR can form condensates by associative phase separation. In the future, we will use the SERF-TAR system to study molecular features that drive the formation of RNA-protein condensates. Collectively, using two disparate systems offers insights into how the dynamically accessible conformational space of proteins determines their biological functions.

Chapter 1 Introduction

1.1 Structural Dynamics in Mechanisms of Chaperone Function[1]

Protein structures observed through X-ray crystallography primarily appear as rigid macromolecules. However, proteins are flexible and dynamic entities that exist in ensembles of interconverting states[1]. The concept of proteins as statistical ensembles has profoundly inspired biologists to explore the rich conformational diversity of proteins that exists beyond their uniquely defined native state[2]. Molecular chaperones are guardians of the proteome that assist in protein folding and prevent the accumulation of aberrant protein conformations that can lead to proteotoxicity. ATP- independent chaperones do not require ATP to regulate their functional cycle. In recent years, several groups have demonstrated the ability of ATP-independent chaperones to directly affect the folding energy landscape of their client proteins. These chaperones function through mechanisms that involve either modulating their internal dynamics or recognizing client conformations with specific dynamic properties. Here, I will introduce the molecular chaperone Spy, which is the model system chosen for my studies. I will also provide a brief overview of other extensively characterized ATP-independent chaperones that regulate client binding and release through conformational transitions.

[1] Section 1.1 of this chapter was published in, and adapted from *Mitra, R., Wu, K., Lee, C., & Bardwell, J. C. ATP-independent chaperones (2022). Annu Rev Biophys., 51, 409-429*. Annual Reviews permits the reproduction of articles by the author for the purpose of an academic thesis. Kevin Wu and I contributed equally to this article and are co-first authors.

1.1.1 Mechanism of the ATP-independent chaperone Spy

Spy is a 16-kDa ATP-independent chaperone discovered in a genetic selection designed to improve the stability of the protein Im7 in the periplasm of *Escherichia coli*[3]. Spy has a unique fold, a thin, α-helical, cradle-shaped homodimer lacking any globular core (**Fig. 1-1**). The concave surface of Spy, with an average thickness of just 9.2 Å, is thinner than the width of a single α-helix. In solution, Spy is a homodimer held together through an extensive interface that buries ~1,850 Å^2 per monomer and is stabilized by antiparallel coiled- coil interactions[3]. The binding site of Spy contains four hydrophobic patches surrounded by positively charged residues.

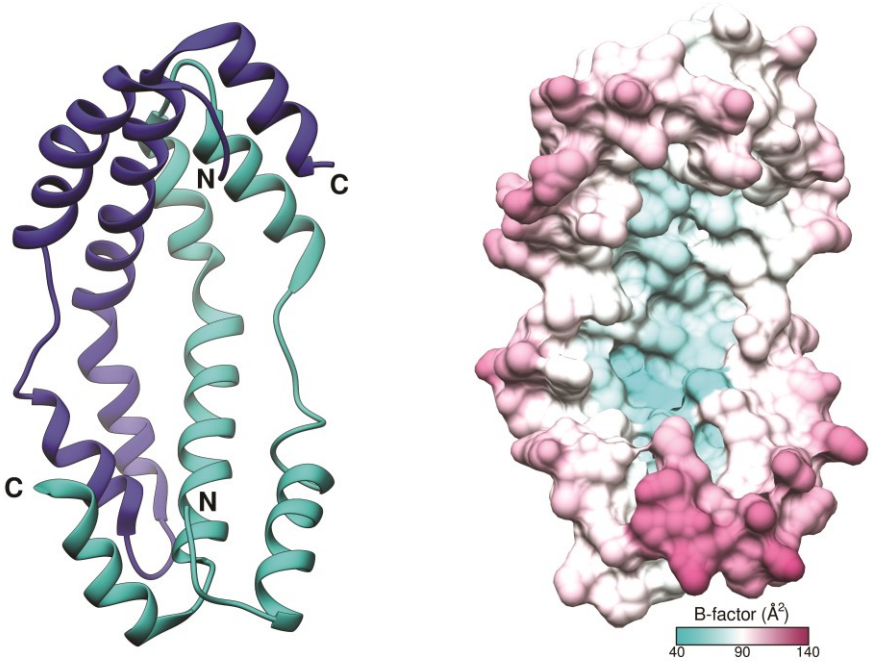


Figure 1-1: Crystal structure of the Spy dimer

(left) Ribbon drawing of Spy dimer (PDB: 3OEO). The N and C termini are labeled. (right) Molecular surface of Spy colored based on average B-factors of each residue. Higher B-factors indicate greater structural flexibility. The disordered N- (residues 1-28) and C-terminal (residues 125-138) regions are missing in the crystal structure.

ATP-independent chaperones have been traditionally thought only to bind partially unfolded clients by hydrophobic interactions and prevent misfolding and

aggregation[4]. Like canonical holdase chaperones, Spy is a potent suppressor of in vitro aggregation of several substrates like malate dehydrogenase (MDH) that is denatured by heating at 43 °C or by treatment with chemical denaturants like urea and ethanol[3]. Spy is a general chaperone with broad substrate specificity, as evidenced by its ability also to suppress the aggregation of other proteins like chemically- denatured glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and aldolase (**Fig. 1-2**)[3]. The well-characterized bacterial immunity protein, Im7, was used to elucidate the thermodynamic, kinetic, and structural features that enable Spy to promote client folding. Im7 populates an on-pathway intermediate in its folding landscape that is stabilized[5,6]. Spy binds folded, intermediate, and unfolded states of Im7[7]. Unlike most chaperones, which are thought to predominantly recognize exposed hydrophobic surfaces, Spy rapidly binds a fully unfolded Im7 variant (L18A L19A L37A H40W) through long-range electrostatic interactions, with association rates of the order of 10^7 M^{-1} s^{-1}[8]. The flexibility and amphiphilic nature of the binding surface allow Spy to accommodate multiple conformations of Im7, a feature that allows the sampling of various folding states while Im7 remains bound[9,10]. Following the initial encounter, the hydrophobic collapse of the client decreases its affinity for Spy and helps favor substrate release in a near-native state (**Fig. 1-3**)[8].

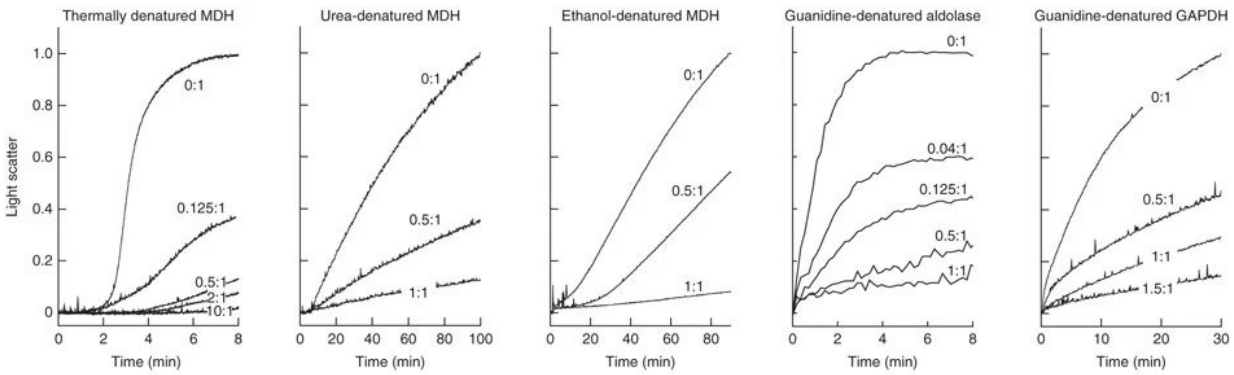


Figure 1-2: Chaperone activity of Spy[3].

Spy prevents substrate protein aggregation as monitored by light scattering. Aggregation of thermally or chemically denatured substrates was measured in the absence or presence of increasing amounts of Spy (ratios given are Spy: substrate). MDH, malate dehydrogenase; GAPDH, glyceraldehyde-3-phosphate dehydrogenase. (Reprinted from Quan, S. et al. Nat. Struct. Mol. Biol. (2011) with permission from Springer Nature)

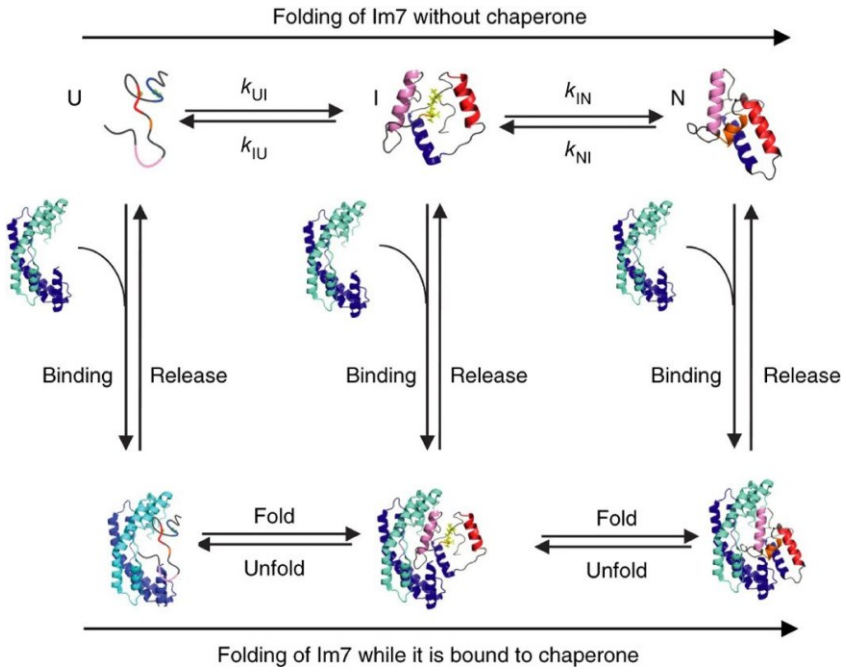


Figure 1-3: Folding of the Im7 in the presence or absence of the Spy[7].

Im7 is shown as a multicolored protein that is helical in both the folding intermediate (I) and in the native folded state (N) and lacks any persistent secondary structure in the unfolded state (U). The Spy homodimer (blue) can bind to all three folding states of Im7, i.e., U, I, and N. k_{UI} and k_{IN} are forward rate constants. k_{IU} and k_{NI} are reverse rate constants. (Reprinted from Stull, F. et al. Nat. Struct. Mol. Biol. (2016) with permission from Springer Nature)

One advantage of this mode of chaperone action is that, by providing a folding-friendly surface that loosely cradles the client protein, Spy allows the client protein to fold while

simultaneously protecting it from aggregation. This is vital because, once a protein is aggregated, it is difficult for it to recover, especially in an ATP-deficient compartment like the periplasm. In general, binding of Spy to either folding intermediates or to the unfolded state will tend to slow folding; however, it enables Spy to inhibit irreversible aggregation of its clients (**Fig. 1-4**).

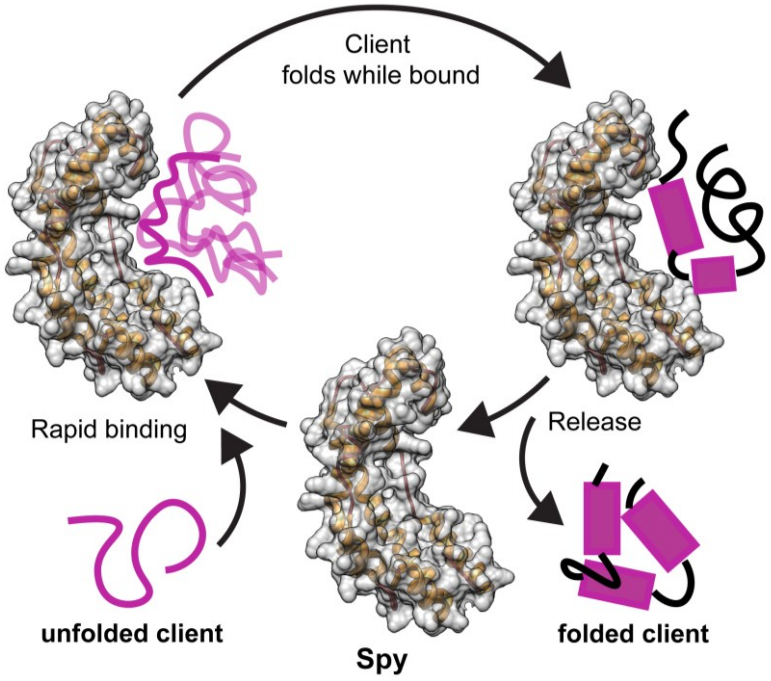


Figure 1-4: The chaperone cycle of Spy[11]***.***

Spy (blue) rapidly associates with Im7 (red) through long-range electrostatic interactions, allowing it to fold while bound. Folding into the native state weakens the interaction with Spy due to hydrophobic burial, triggering Im7 release and completing the chaperone cycle of Spy. (Reprinted from Mitra, R. et al. Annu. Rev. Biophys. (2022) with permission from Annual Reviews)

Given the lack of a regulatory mechanism for substrate binding and release, it is unclear how Spy completes its functional cycle. One possibility is competitive release of the client from Spy, where Spy is displaced by another periplasmic chaperone that has a higher free energy of client binding. In such a scenario, the transfer of clients from Spy could be driven solely by favorable thermodynamics in the absence of ATP-dependent client release mechanisms. Spy can also be replaced by Im7's natural binding partner,

colicin E7. The micromolar affinity of Spy for native Im7 is much weaker than the subnanomolar affinity of Im7 for colicin E7[12]. However, Spy and colicin E7 binds to the same locally flexible regions on fully folded Im7[10]. Therefore, the release of Im7 from Spy can be triggered by folding to the native state due to its energetically favorable interaction with colicin E7[10].

1.1.2 Functionally- relevant dynamics in ATP-independent chaperones

ATP-independent chaperones have evolved diverse strategies to regulate their chaperone activity without nucleotide cofactors and cochaperones. This subsection describes these mechanisms using two E. coli chaperones, SurA and HdeA, as representative examples.

SurA is a 45-kDa periplasmic chaperone essential for cell survival in the stationary growth phase[13]. SurA assists in folding several outer membrane proteins (OMPs) by maintaining them in a state that is competent for insertion or secretion while they transit the periplasm. *E. coli* SurA is a modular protein with a core domain composed of N- and C- terminal regions. SurA has two parvulin- like PPIase domains (P1 and P2). The core domain of SurA contacts the P1 domain to form a globular region with an extended substrate-binding crevice, while the P2 domain is separated from the core by two linkers[14]. The PPIase domains appear to be nonessential under at least some laboratory conditions and P1, though it is parvulin-like in sequence, has no PPIase activity [15]. However, there is some evidence that these parvulin-like domains can regulate or enhance the *in vitro* chaperone activity of SurA. The *in vitro* anti-aggregation activity of SurA for substrates like OmpT is enhanced by its PPIase domain(s)[16].

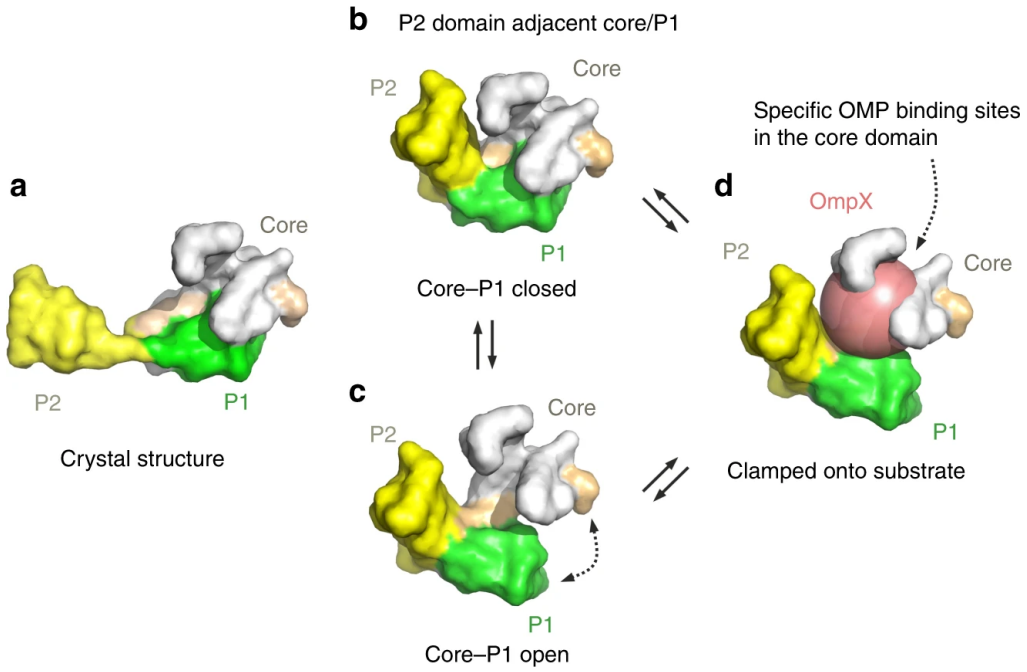


Figure 1-5: Structure and conformational dynamics of SurA[17]

*Inter-domain dynamics in solution allow SurA to populate conformations distinct from the crystal structure (PDB: 1M5Y) shown in (**a**). For example, the P2 domain is close to the core/P1 domains in **(b)** and **(c)**. The binding of unfolded clients results in the conformation **(d)** that is intermediate between the closed **(b)** and open **(c)** conformations adopted by the P1 domain. The SurA-bound OmpX is shown in a collapsed globule state (represented by a sphere). Crosslinking studies have identified multiple OMP binding sites in the core domain.*

Chaperones need to accommodate and, in some cases, facilitate large conformational changes within their substrates, so it is not surprising that they, too, undergo major conformational changes. For SurA, domain-level conformational dynamics appear important, since introducing a disulfide bond that tethers the P1 and core domains together results in impaired OMP activity[18]. A recent study used an integrative approach employing mass spectrometry, single-molecule Förster resonance energy transfer, and molecular dynamics (MD) simulations to show that SurA adopts conformational states in the solution that differ greatly from the extended conformation in its static crystal structure, with the P2 domain lying closer to the core and P1 domains in solution (**Fig. 1-5**) [17]. The results also showed that the P1 domain of SurA adopts

closed and open states at the core–P1 interface, which could be functionally important. However, the relative abundance of the various open and closed conformations is hard to pin down, with another study reporting that the hydrodynamic properties of SurA agree well with the extended arrangement that is seen in the crystal structure where the P2 domain extends away from the core and P1 domains[19]. The free energy of binding of OMPs to SurA is much smaller than that of OMP folding[20]. Therefore, the transfer of unfolded OMPs from SurA to the BAM can be driven solely by favorable thermodynamics in the absence of any ATP-dependent release mechanisms. Regardless, it is still unclear how conformational changes in SurA are coupled to client binding and delivery to the β-barrel assembly machinery (BAM) complex for folding into the outer membrane.

Multiple groups have studied the conformational dynamics of unfolded OMPs bound to SurA[17,19,21,22]. Using single-molecule force spectroscopy, *Thoma et al. (2015)* showed that, upon partial unfolding by mechanical force, that the FhuA receptor populates non-native misfolded conformations when allowed to refold in the absence of chaperones[22]. However, when refolding occurs in the presence of SurA, unfolded FhuA inserts into the lipid membrane and folds into native β-hairpins in a stepwise fashion. SurA alters FhuA folding by decreasing the probability of misfolding and increasing the probability of its folding into native β-hairpins[22]. SurA-bound FhuA exists in a dynamic unfolded ensemble with conformational exchange rates in the milliseconds[22]. Crosslinking mass spectrometry and SANS data on another client, the barrel domain of unfolded OmpA, supports the idea that SurA expands the unfolded client (**Fig. 1-6**)[19]. Two independent studies have also demonstrated that multiple SurA molecules bind to

and expand one molecule of unfolded OmpX[17,21]. Collectively, these findings point to the intriguing possibility that SurA can not only prevent aggregation but can alter the folding pathway of OMPs, steering them away from misfolded states.

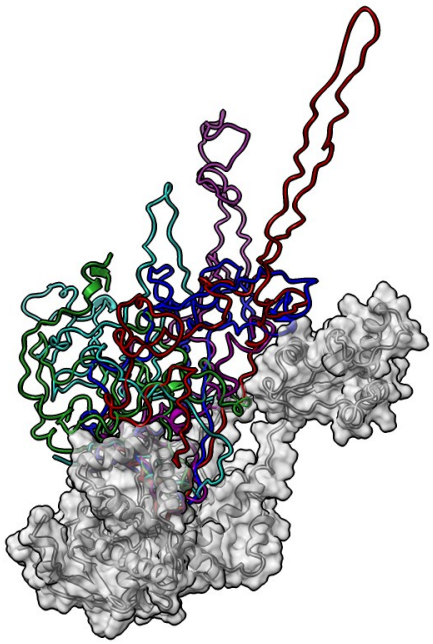


Figure 1-6: Low-resolution model for SurA-unfolded OmpA complex[11]

A sparse ensemble of SurA-unfolded $OmpA_{171}$ complex was constructed by integrating experimental restraints from small-angle neutron scattering and crosslinking-mass spectrometry. (Reprinted from Mitra, R. et al. Annu. Rev. Biophys. (2022) with permission from Annual Reviews)

HdeA is a small, pH-responsive chaperone in the periplasm of enteric bacteria[23,24]. It is highly upregulated under acid stress and protects several periplasmic proteins, including other chaperones like SurA and DegP, from aggregation due to pH-induced denaturation[25]. HdeA is chaperone inactive at neutral pH and exists as a well-folded all-helical homodimer[26]. However, the periplasmic pH rapidly shifts from 7 to 2 when bacteria enter the stomach upon ingestion[24]. Upon the shift to low pH, HdeA rapidly monomerizes, unfolds, and becomes active (**Fig. 1-7**)[27].

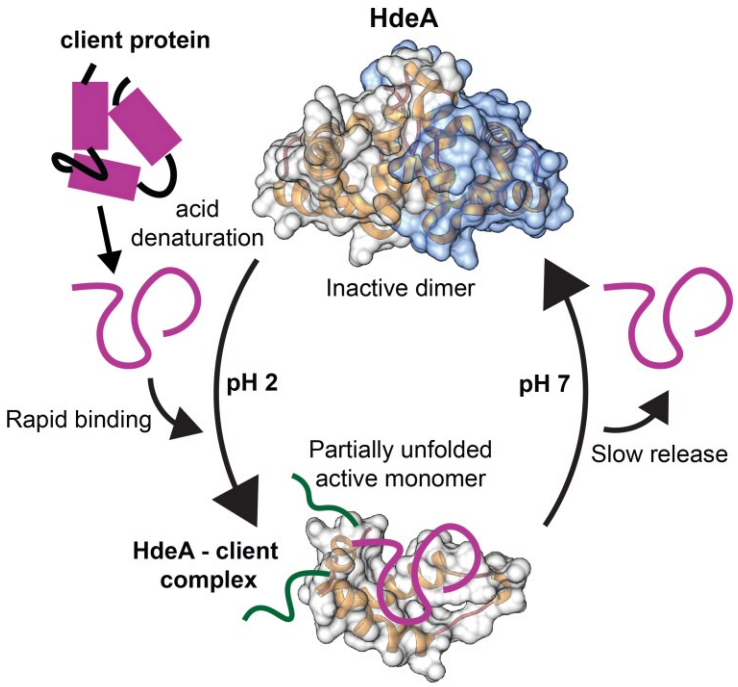


Figure 1-7: Chaperone cycle of HdeA[11]

Client protein is in purple and black, and HdeA in space-filling crystal structure models. At neutral pH, HdeA exists as an inactive dimer. Under acidic conditions (pH 2), HdeA is activated due to monomerization and partial unfolding. Active HdeA binds acid-denatured clients and prevents their aggregation. When pH shifts to neutral (pH 7), HdeA releases its client slowly and forms the dimer again. The monomer structure of HdeA is unknown; for illustration purposes, a single subunit of the dimer is shown, but the actual structure of the monomer is likely to be expanded and dynamic. (Reprinted from Mitra, R. et al. Annu. Rev. Biophys. (2022) with permission from Annual Reviews)

Several groups have studied the activation pathway of HdeA using multiple orthogonal approaches like NMR spectroscopy, constant pH MD simulations, umbrella sampling simulations, and fluorescence-based pH titrations [27–31]. Without ATP- based regulation, order-to-disorder transitions can allow rapid exploration of conformational space. The activation of HdeA is a complex multistep process. Under acidic pH, HdeA populates unfolded states in addition to the folded dimer[27]. Chemical exchange saturation transfer NMR profiles show that the apparent exchange rate constant between folded and unfolded conformations is pH dependent, and the relative population of disordered states increases as pH is lowered from 3.5 to 2.5[29]. *Yu et al.* (2019) identified three pH hotspots for acid-induced conformational transitions[29].

Destabilization of these locks at moderately low pH (3–4) loosens the dimer and exposes client binding sites. At very low pH (1.5–2), HdeA is fully activated and exists as a partially unfolded monomer with highly charged and flexible N- and C- terminal tails and two hydrophobic client binding sites held together by a strictly conserved disulfide bond. Although the authors of this study identified client binding sites on HdeA, obtaining high-resolution NMR structures of the chaperone–client complex of HdeA with the native client MalE is extremely challenging and was not attempted.

In summary, multiple orthogonal approaches combining NMR, FRET, mass spectrometry, and modeling have uncovered complex conformational dynamics of chaperones like SurA and HdeA that can regulate chaperone activity. However, the mechanistic connection between the kinetics of conformational transitions and the kinetics of client binding and release remains elusive. Furthermore, how these chaperones operate in periplasmic proteostasis networks is also largely unclear.

1.2 Dynamic interactions of disordered protein regions with RNA

Intrinsically disordered protein regions (IDPRs) adopt an ensemble of rapidly interconverting conformations in a shallow free-energy landscape. These regions are overrepresented in RNA-binding proteins (RBPs) and often have repetitive motifs and a high prevalence of small polar and charged amino acids, particularly Gly, Ser, Arg, Lys, Gln, Glu, and Asp. They also show a low content of bulky hydrophobic amino acids except for Tyr[32]. Classically, RBPs are recognized as proteins containing canonical RNA-binding domains such as RNA recognition motif, the K homology domain, the DEAD box motif, double-stranded RNA-binding motif, and the zinc-finger domain [32]. However, several recent genome-wide studies have led to the discovery of unconventional RBPs that possess IDPRs capable of directly engaging in or facilitating protein-protein and protein-RNA interactions[33]. RBPs are involved in all aspects of RNA metabolism and associate with RNA in ribonucleoprotein (RNP) complexes like the spliceosome, ribosome, and RNP granules. Here, I will outline some of the distinct roles of disordered regions in RBPs:

- Disordered sequences like RGG/RG motifs in RBPs like G3BP1 and FUS have intrinsic RNA binding activity[34]. RNA can induce a disorder-to-order transition in the RGG/RG motif in disease-related protein FMRP[35].
- Disordered regions can regulate the RBPs of proteins without directly interacting with RNA. The carboxyl terminal domain of RNA polymerase II can regulate the interaction of RNA polymerase II with RBPs involved in posttranscriptional processing of mRNA[36]. The conformational heterogeneity of carboxyl terminal

domain itself is regulated by phosphorylation in serine residues of the YSPTSPS heptads[37].

- Disordered low-complexity sequences present in RBPs play a regulatory role in the formation of membraneless organelles known as RNP granules. These sequences engage in multivalent interactions, driving the phenomenon of liquid-liquid phase separation[38–40]. In Chapter 3 of my thesis, I investigate the involvement of disordered regions in facilitating RNA-protein interactions within the context of a model phase-separating RNP complex formed by an RNA- binding protein called SERF.

1.2.1 SERF is a small protein with a highly conserved disordered region

MOAG-4/SERF (Modifier of aggregation-4/ Small EDRK-rich factor) is a group of evolutionarily conserved proteins (Pfam: PF04419) that modulate amyloid formation. *Van Ham et al.* identified SERF in a forward genetic screen in *Caenorhabditis elegans* to identify regulators of amyloid aggregation and proteotoxicity[41]. To do this, they mutagenized worms that express a transgene carrying a polyQ stretch of 40 residues fused to a yellow fluorescent protein (Q40::YFP) in the body-wall muscle cells to identify novel *m*odifiers *o*f *ag*gregation (*moag*) that suppress amyloid aggregation (**Fig. 1-8A-C**). One of the mutants (*moag-4*) carried a point mutation, G210A, in the gene Y37E3.4, showing a 75% reduction in the number of aggregates (**Fig. 1-8D**). The *moag-4* mutation is a missense mutation in the MOAG-4 protein, changing the conserved methionine at position 49 into isoleucine. A comprehensive phenotypic characterization of *moag-4* in *C. elegans* models of neurodegenerative diseases led the authors to conclude that the loss/inactivation of MOAG-4 suppresses aggregation and proteotoxicity for various disease-associated amyloidogenic proteins like polyglutamine

protein, amyloid-beta, and alpha-synuclein. In neuroblastoma and HEK 293 cell lines, human orthologs of MOAG-4, SERF1A, and SERF2 increased the aggregation of mutant Huntingtin (Htt-Q74) and increased cell death (**Fig. 1-8E, F**).

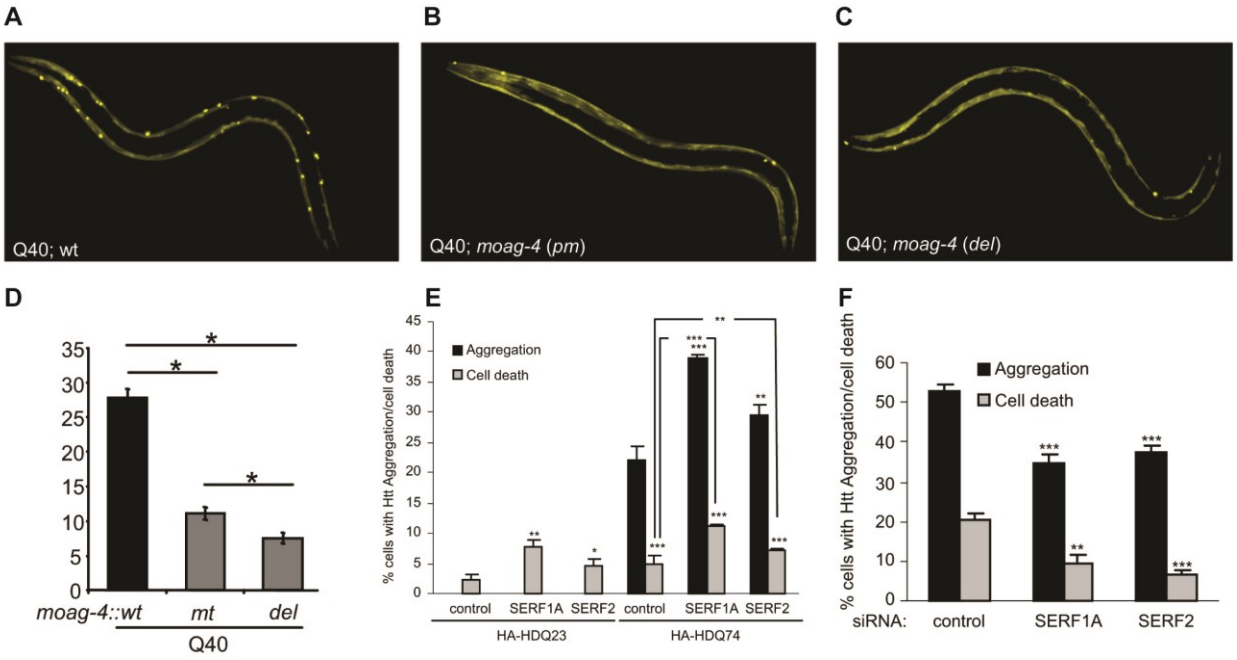


Figure 1-8: Role of MOAG-4/SERF in amyloid aggregation[41]

Confocal images of C. elegans with wild-type (A), G210A point mutant (B), or deleted (C) moag-4 gene, expressing Q40::YFP in the body-wall muscle cells. Soluble and aggregated YFP-tagged polyQ proteins are observed by diffuse and focal fluorescence signals, respectively. (D) Quantification of aggregates in Q40 worms with wild-type, point mutant, or deleted moag-4. (E) The percentage of SK-N-SH neuroblastoma cells showing aggregation and cell death after being transiently transfected with SERF1A or SERF2, or control (dsRed2), and HA-tagged huntingtin exon 1 with either 74 polyQ repeats (HA-HDQ74) or 23 polyQ repeats (HA-HDQ23). Huntingtin exon 1 with 23 polyQ repeats is a control mutant that does not form aggregates. (F) Q74 aggregation and cell death were quantified in HEK 293 cells transfected with a nontargeting siRNA (control) or siRNA targeting SERF1A or SERF2 and HA-tagged Q74 huntingtin. (Reprinted and modified from van Ham T et al. Cell. (2010) with permission from Elsevier)

The kinetic mechanism by which SERF-enhances amyloid formation was studied by *Meinen et al.* using amyloid-β protein (Aβ40) and α-synuclein (αS)[42]. They showed that the SERF homolog *YDL085C-A* from *Saccharomyces cerevisiae* accelerates amyloid formation at substoichiometric concentrations. Although SERF did not alter the basic mechanism of Aβ40 amyloid aggregation, it significantly accelerated the rate constant for primary nucleation. SERF was also able to reduce the half-time of αS

aggregation. Typically, the kinetics of amyloid formation is monitored by measuring the fluorescence of an amyloid-specific dye called thioflavin T (ThT) that binds β-sheet rich structures. The lag time in amyloid formation experiments indicates the energetic barrier for forming ThT-invisible amyloid nuclei from monomers. Since SERF affects primary nucleation, ThT-based kinetic experiments were not particularly informative. Native ion mobility-mass spectrometry (IM-MS) was employed to probe the conformational changes in amyloid proteins in the presence of SERF. This MS approach preserves native-like structures and interactions in the gas phase[43]. Ions generated by nano-electrospray ionization get separated in a drift tube filled with a neutral gas under the influence of a weak magnetic field. Ions migrate based on their orientationally- averaged collision cross-section (CCS), which determines their drift time. Following the ion mobility separation, the ions are injected into a Time-of-Flight analyzer, where they separate based on the mass-to-charge (m/z) ratio.

A broad charge-state distribution was observed for SERF, indicating an intrinsically disordered protein[44]. Furthermore, the lower-charge-state ions detected in the nano-Electrospray ionization process had multiple arrival time distributions in the ion mobility step. The arrival time distributions were fitted to Gaussian functions to obtain CCSs that for SERF ranged from 9 to 22 nm^2. The authors detected 1:1, 1:2, and 2:1 complexes of SERF: Aβ40 and 1:1, 1:2, 2:1, and 2:2 complexes of SERF: αS in equimolar mixtures of SERF and either Aβ40 or αS. The 1:1 complexes of SERF: Aβ40 and SERF: αS had extended conformations and retained the conformational heterogeneity observed for SERF alone. The authors proposed a model wherein weak interactions of SERF and monomeric Aβ40 or αS lead to the formation of “fuzzy

complexes” exhibiting significant conformational and stoichiometric heterogeneity. These extended complexes accelerate the primary nucleation step in amyloid fibril formation by an unknown mechanism.

Recently the human SERF1a protein has been found to localize to nucleoli and can facilitate the nucleolar incorporation of fluorescently labeled RNA[45]. The positively charged N-terminal segment of human SERF1a was previously shown to interact with α-synuclein and single-stranded RNA. While the exact role of SERF in amyloid diseases is still unclear, the cellular localization and physicochemical properties of SERF suggests that RNA binding is its main function in the native cellular context. We were unable to detect any sequence and structure preferences of SERF with the RNA Bind-n-Seq method (*unpublished*). RNA Bind-n-Seq is a high-throughput *in vitro* method for characterizing the binding spectra of RNA-binding proteins based on *in vitro* affinity-based isolation of RNA-protein complexes and deep sequencing of bound RNA[46]. Since, SERF lacks any recognizable folded RNA-binding domain, it is unlikely that it has specific sequence preference. This places SERF in the company of over 95% of unconventional RBPs that lack canonical RNA binding domains, and also do not exhibit sequence specificity for RNA-target[32,47]. Therefore, the RNA-binding properties of SERF can be mechanistically studied with any RNA sequence of choice[48].

Chapter 2 Elucidating Substrate-Specific Mechanisms of the Chaperone Spy[2]

2.1 Background

Topologically complex proteins often populate misfolded intermediates that act as kinetic traps[49]. Such intermediates often expose hydrophobic surfaces that make them prone to aggregation. ATP-dependent molecular chaperones like GroEL- GroES rescue trapped intermediates and facilitate substrate folding[50]. In contrast, ATP-independent chaperones generally bind tightly to non-native substrates and prevent protein aggregation but are not thought to directly facilitate substrate refolding[4]. The simple designations of “foldase” or “holdase” may underemphasize the microscopic structural heterogeneity of chaperone–substrate complexes[51]. Substrates bound to ATP-independent chaperones such as trigger factor, SecB, and the sHsps can adopt a wide range of conformations ranging from near-native to unfolded states[52–54]. We have shown that the ATP-independent chaperone Spy binds protein folding intermediates of substrate proteins such as Im7 and Fyn SH3 and allows for the folding of substrate proteins while they remain chaperone bound[7,55]. Spy loosely binds to different folding states of these model substrate proteins. Although the folding rate constant for these

[2] The contents of this chapter were published in *Mitra, R., Gadkari, V. V., Meinen, B. A., van Mierlo, C. P., Ruotolo, B. T., & Bardwell, J. C. (2021). Mechanism of the small ATP-independent chaperone Spy is substrate specific. Nat. Commun, 12(1), 1-13*. Springer Nature permits the reproduction of articles by the author for the purpose of an academic thesis. I performed all the experiments in this chapter except the analytical ultracentrifugation experiments and native ion mobility-mass spectroscopy. Recombinant proteins were expressed and purified by Ke Wan from the Bardwell lab. All authors contributed to the research design, data analysis, and writing.

two simple substrates decreases with increasing concentrations of Spy, it does not go to zero at saturating Spy concentrations where essentially all the substrate molecules are chaperone bound, evidence that folding of the substrate occurs while it is chaperone-bound[56]. In this mechanism, substrate release is not a prerequisite for substrate folding. We have found that the folding-while-bound mechanism depends on relatively weak chaperone–substrate interactions[55]. Variants of Spy that bind SH3 with stronger affinity than wild-type (WT) Spy significantly slow the folding of bound substrate[55]. The substrates of Spy tested so far have been small, topologically simple, all β-sheet or all α-helical model proteins. We wanted to test if the folding-while-bound model is generalizable to larger substrates with complex topologies such as those found in many proteins.

Here, we chose apoflavodoxin from *Anabaena* PCC 7119 (AnFld) and *Azotobacter vinelandii* (AzoFld), both well-studied folding models that populate kinetically trapped off-pathway intermediates[57–59]. Their α/β topology is ancient and frequently found in common folds such as TIM barrels, Rossman folds, FAD/NAD(P)-binding domains, and P-loop-containing hydrolases. This study aimed to test whether the folding-while-bound mechanism adequately describes the effect of Spy on the folding of proteins with complex topologies using the flavodoxin-like fold as an example. Our results show that Spy's mechanism is surprisingly substrate specific. In the case of AnFld, Spy traps denatured polypeptides in a non-native state, thereby inhibiting AnFld folding. This dual functionality of Spy as both a folding-while-bound chaperone and a holdase rests on opposing thermodynamic requirements. The folding-while-bound paradigm requires Spy to weakly bind the various folding states of the substrate.

In contrast, a holdase binds non-native states tightly, so folding transitions are not feasibly thermodynamically. We show that the mechanism of action of the chaperone Spy is dictated by the difference in binding affinities for partially unfolded states of different substrates to Spy. This work highlights the need to shift from monolithic enzyme-inspired views of molecular chaperones toward more complex models that incorporate the possibility of multiple substrate-specific mechanisms.

2.2 Results

2.2.1 Spy inhibits the folding of ApoFld by kinetic trapping

Spy belongs to an emerging class of chaperones that allow protein folding while the substrate remains bound to the chaperone[56]. The substrates studied in detail so far, namely Im7 and SH3, are small proteins (10 and 7 kDa, respectively) with very simple 3D structures (just α-helices or just β-strands, respectively). We wondered if the folding-while-bound mechanism of Spy applies to more complex substrates that include both α-helices and β-strands and have a more complex folding pathway than the substrate proteins tested so far. To investigate this, we chose the apoflavodoxin AnFld as our substrate, as its native state has an α–β parallel topology and as AnFld exhibits a more complex folding pathway than either Im7 or SH3 in that it folds via a three-state triangular mechanism with an off-pathway intermediate (**Fig. 2-1a**)[59]. In this mechanism, most of the molecules in the intermediate conformation need to unfold prior to refolding into the native conformation. However, a small fraction of molecules can directly fold from the intermediate to the native state. First, we verified that the previously established triangular folding mechanism for flavodoxin folding is functional under our buffer conditions (40 mM HEPES-KOH pH 7.5, 100 mM NaCl), which we considered to be more physiological than the previously used buffer (50 mM MOPS pH 7.0) (**Fig. 2-2**)[7,59].

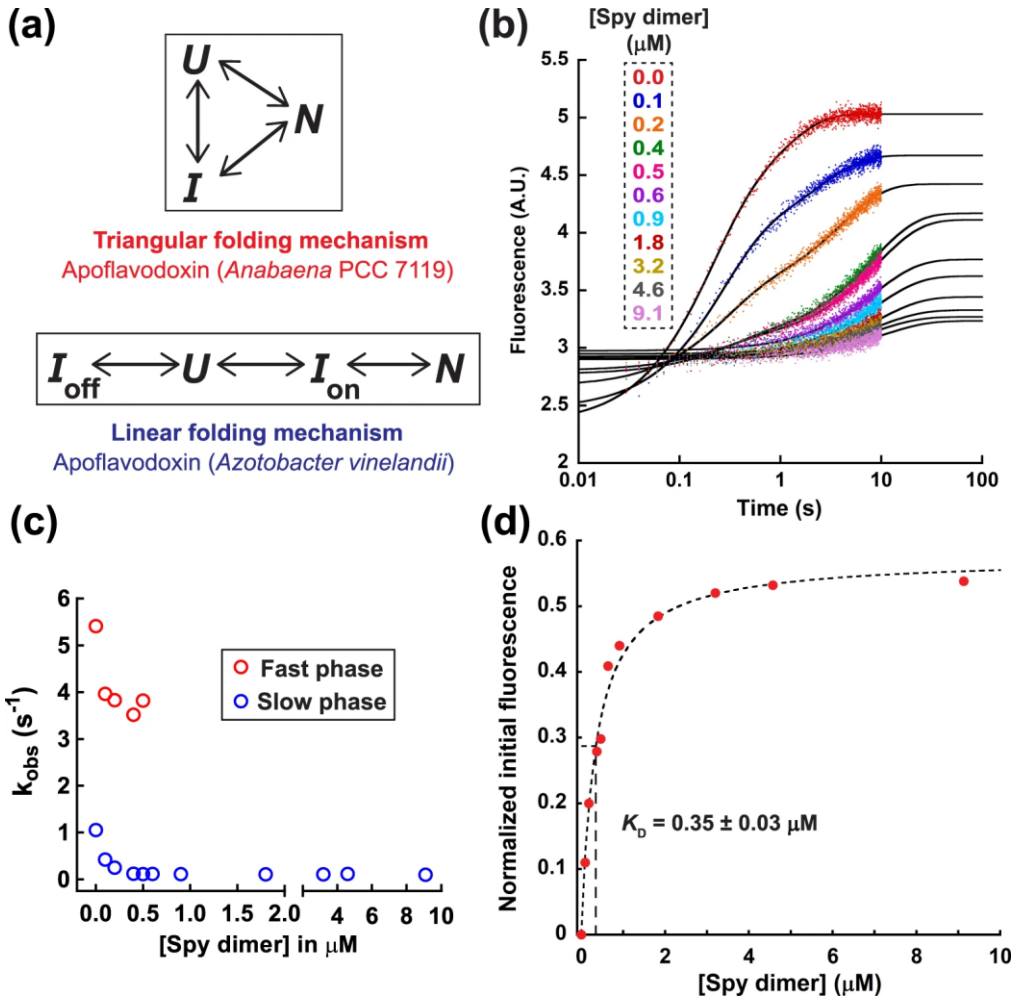


Figure 2-1: Kinetics of apoflavodoxin folding in the presence of Spy.

(a) Folding pathway of apoflavodoxin from Anabaena PCC 7119 (AnFld) and Azotobacter vinelandii (AzoFld). N, U, and I are the native, unfolded, and intermediate states. In the case of AzoFld, $I_{off,}$ and I_{on} are the off-pathway and on-pathway intermediates, respectively. (b) Fluorescence traces for the refolding of 0.09 μM AnFld in the presence of various concentrations of Spy dimer (0–9.1 μM after mixing). The time axis is shown in a logarithmic scale. AU arbitrary units. The traces of AnFld refolding in the presence of 0–0.5 μM Spy dimer could be fit to a sum of two exponentials, and traces for higher Spy concentrations could adequately fit a single exponential function. The absence of the fast phase at higher Spy concentrations most likely indicates that Spy completely blocks the direct folding pathway from unfolded to native AnFld. (c) A plot of the observed rate constants (k_{obs}) for AnFld WT folding as a function of Spy dimer concentration. (d) Initial fluorescence values were calculated from the fit of the kinetic datasets. AU arbitrary units. The binding isotherm was plotted with normalized initial fluorescence intensity during AnFld WT refolding in the y-axis and fitted to a one-site binding model. All experiments were conducted in HN buffer at 25 °C. The value reported is the mean ± s.e.m. of the fit.

We studied the refolding of AnFld in the presence of Spy by monitoring the tryptophan fluorescence in a stopped-flow fluorimeter upon diluting urea-denatured AnFld into refolding buffer containing increasing concentrations of Spy, as described previously[7].

We observe two kinetic phases of AnFld folding; a major kinetic phase, where the amplitude is proportional to the fraction of denatured molecules that directly fold to the native state (N), and a minor phase whose amplitude is proportional to the fraction of molecules that transiently populate the off-pathway intermediate and therefore need to unfold before productive refolding[59]. In the presence of Spy, the observed rate constants (k_{obs}) of both phases decrease substantially until at the highest Spy concentration (~9 μM), a single folding phase remains that has a small k_{obs} value (0.1 s^{-1}) (**Fig. 2-1b, c**). This indicates that Spy significantly slows the folding rate of AnFld, like for Im7 and SH3. However, both Im7 and SH3 can fold to the native state, albeit slowly, even at saturating Spy concentrations[7,55]. In contrast, binding to Spy prevents the folding of AnFld almost completely, as evidenced by the dramatic decrease in the folding amplitude upon increasing Spy concentration (**Fig. 2-1b**). At Spy concentrations higher than ~9 μM, the amplitude of the fluorescence change was extremely small, indicating that most of the denatured AnFld molecules are kinetically trapped in a folding-incompetent bound state. Both the unfolded (U) and intermediate (I) states of AnFld are less fluorescent than the native state[59]. Since AnFld in the presence of Spy is substantially less fluorescent than it is in the native state, we reason that Spy kinetically traps AnFld in a non-native state that is partially or completely unfolded (we term this state U*). It appears that Spy binds very rapidly to AnFld and prevents folding since the fluorescence decreases within the ~25 ms dead time of the stopped-flow instrument. The interaction of Spy and the U* state is tight with a dissociation constant (K_D) of 0.35 μM (**Fig. 2-1d**). The affinity of Spy for $AnFld^{U*}$ in folding experiments is ~13-fold tighter than observed for Spy binding to the intermediate and unfolded states of Im7

(K_D = 4.7 μM for both) and ~8-fold tighter than that between Spy and unfolded SH3 (K_D = 2.9 μM), both of which fold while bound to Spy[7,55]. Due to these high-affinity interactions, the kinetic trapping of AnFld in the Spy-bound U* state apparently necessitates release from Spy before productive folding can occur.

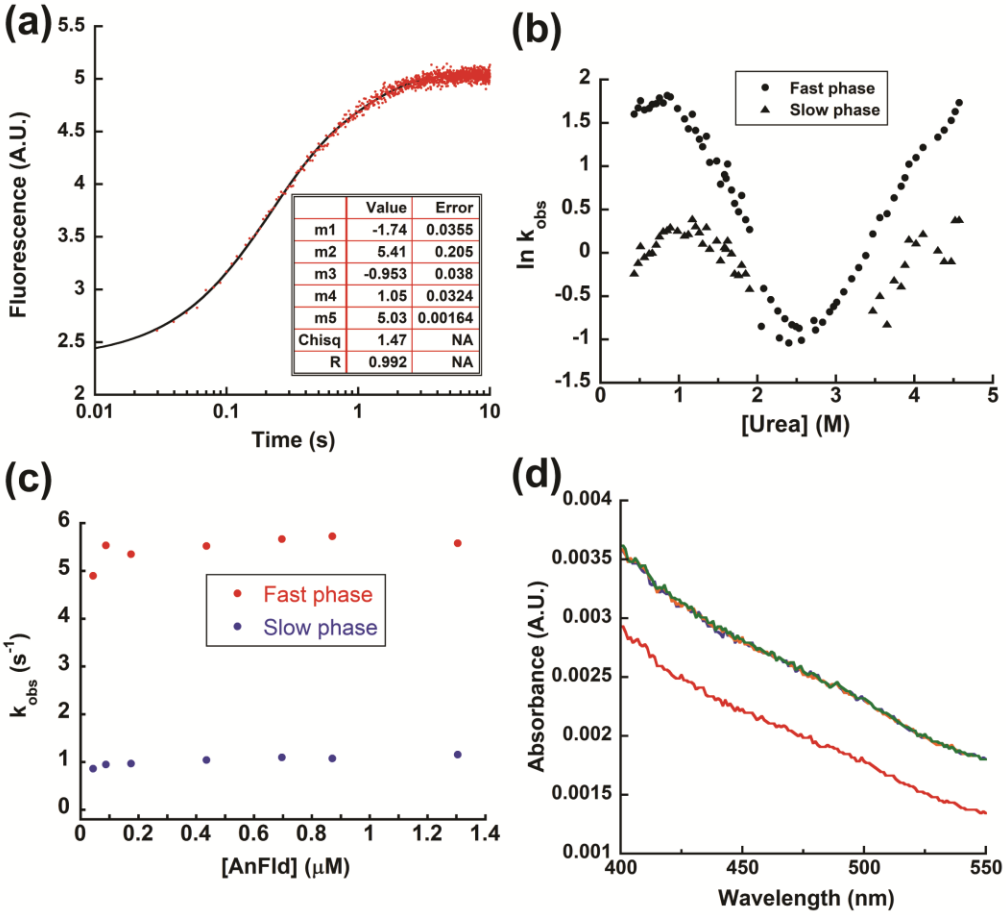


Figure 2-2: Mechanism of apoFld folding; analysis of purified ApoFld.

(a) *The kinetics of AnFld folding has previously been followed by monitoring its intrinsic tryptophan fluorescence[59]. Fluorescence trace for the kinetic refolding of AnFld monitored by 11.5-fold dilution of 1.04 μM AnFld denatured from HN buffer containing 5 M urea into HN buffer. The average of 10 – 12 traces was fitted to a sum of two exponentials given by the equation* $y = m_1 e^{-m_2 t} + m_3 e^{-m_4 t} + m_5$ *where y is the fluorescence at time t,* m_1 *and* m_3 *are the amplitudes,* m_2 *and* m_4 *are the observed rate constants for the major and minor phases, respectively, and* m_5 *is the final fluorescence. A.U., arbitrary units. The black line shows the best fit of the data, and the table shown contains the fit parameters. Note the logarithmic scale in the time axis. AnFld's tryptophan fluorescence increased in two kinetic phases: a major (fast) phase with an amplitude of 64.7 % and a minor (slow) phase with an amplitude of 35.3 %. The rate constant for the major phase is 5.4* s^{-1}, *and that of the minor phase is 1.05* s^{-1}. ***(b) – (c)*** *Before studying the effect that Spy has on AnFld folding, we first wanted to determine if AnFld folds via a similar mechanism in our previously used HN buffer that contains physiological concentrations of NaCl (100 mM), as it does in the 50 mM MOPS buffer used by Fernández-Recio and co-workers that lacks any*

added salt[60]*. To understand AnFld's folding mechanism, we followed AnFld folding and unfolding under conditions of varying urea concentration. The natural logarithms of the folding and unfolding rate constants were plotted as a function of urea concentration to obtain the chevron plot shown in* ***(b)*** *Observation of a logarithmic dependence of both the folding and unfolding rate constants on denaturant concentration would indicate a single transition state barrier with no folding intermediates, i.e., a two-state folding mechanism*[61]*. Our data, however, reveal a major refolding phase and a slower minor refolding phase; the latter shows a marked curvature (nonlinearity) of the folding limb, meaning that the folding rate at low denaturant concentrations increases with increasing denaturant concentration. The observation of such a "rollover" in the folding arm of the chevron plot is taken as strong evidence for the existence of an off-pathway intermediate*[61]*. Fernández-Recio and colleagues found that the slow refolding phase does not represent prolyl cis-trans isomerization in the unfolded state and is independent of protein concentration in the 1 - 18 μM range*[60]*.* ***(c)*** *To verify that AnFld does not aggregate in our kinetic refolding experiments, we varied the protein concentration (0.5 μM – 15 μM) in the folding experiments. Both phases were independent of protein concentration. The unfolding kinetics shows two phases, exhibiting a linear dependence on urea concentration. While a 3-state linear off-pathway mechanism can adequately explain the two folding phases and the fast-unfolding phase, the slow unfolding phase can only be accommodated in a triangular mechanism in the absence of other factors such as aggregation and the purified protein containing a fraction of FMN-bound species (i.e., the holo-protein).* ***(d)*** *Absorbance spectra of HN buffer alone (in red) and 100 μM purified AnFld were recorded thrice (in blue, green, and orange). The mean absorbance value at 450 nm of the protein is 0.0028, similar to the absorbance of the buffer alone. The absence of any signal at 450 nm in the absorbance spectrum of the purified protein indicates the absence of holo-protein in the sample. Also, analytical ultracentrifugation confirmed that the purified native protein is monomeric (Figure S3a). Together, these observations indicate a triangular folding mechanism of AnFld in HN buffer.*

2.2.2 Rapid binding of Spy is coupled to the partial unfolding of ApoFld

Having shown that Spy inhibits the folding of AnFld, we wondered if Spy interacts with AnFld's native state. Holdase chaperones are known to preferentially bind to non-native unfolded or misfolded proteins[62]. On the other hand, folding-while-bound chaperones also bind to the native state. Thus, one way to distinguish between a holdase and a foldase is to determine the relative affinities of the chaperone for native and non-native states. If Spy is functioning as a holdase with AnFld we expect Spy to bind the native state with much weaker affinity than fully and partially unfolded states. We found a binding stoichiometry of 1:1 between AnFld and the Spy dimer using analytical ultracentrifugation (AUC) (**Fig. 2-4a**). There is a very rapid decrease in fluorescence upon mixing Spy and AnFld (**Fig. 2-3a**). The amplitude of this burst phase reaches saturation at high Spy concentrations, consistent with the very rapid binding of Spy to the native AnFld (**Fig. 2-4b**). However, the interaction is weak, with an apparent

dissociation constant (K_{Dapp}) of 40 μM, 114-fold weaker than the K_D for Spy with $AnFld^{U*}$ (**Fig. 2-3b**). The K_{Dapp} for native AnFld is similar to that previously observed for the interaction of Spy with native Im7 (K_D = 20.5 μM) and SH3 (K_D = 50 μM)[7,55]. These comparable binding affinities for diverse substrates indicate that weak (micromolar) affinity for native substrates is a general property of Spy. This makes sense; tight binding of a chaperone to the native state might inhibit substrate function. Whether Spy inhibits folding or allows the substrate to fold while bound may depend mainly on whether Spy has an affinity for non-native states.

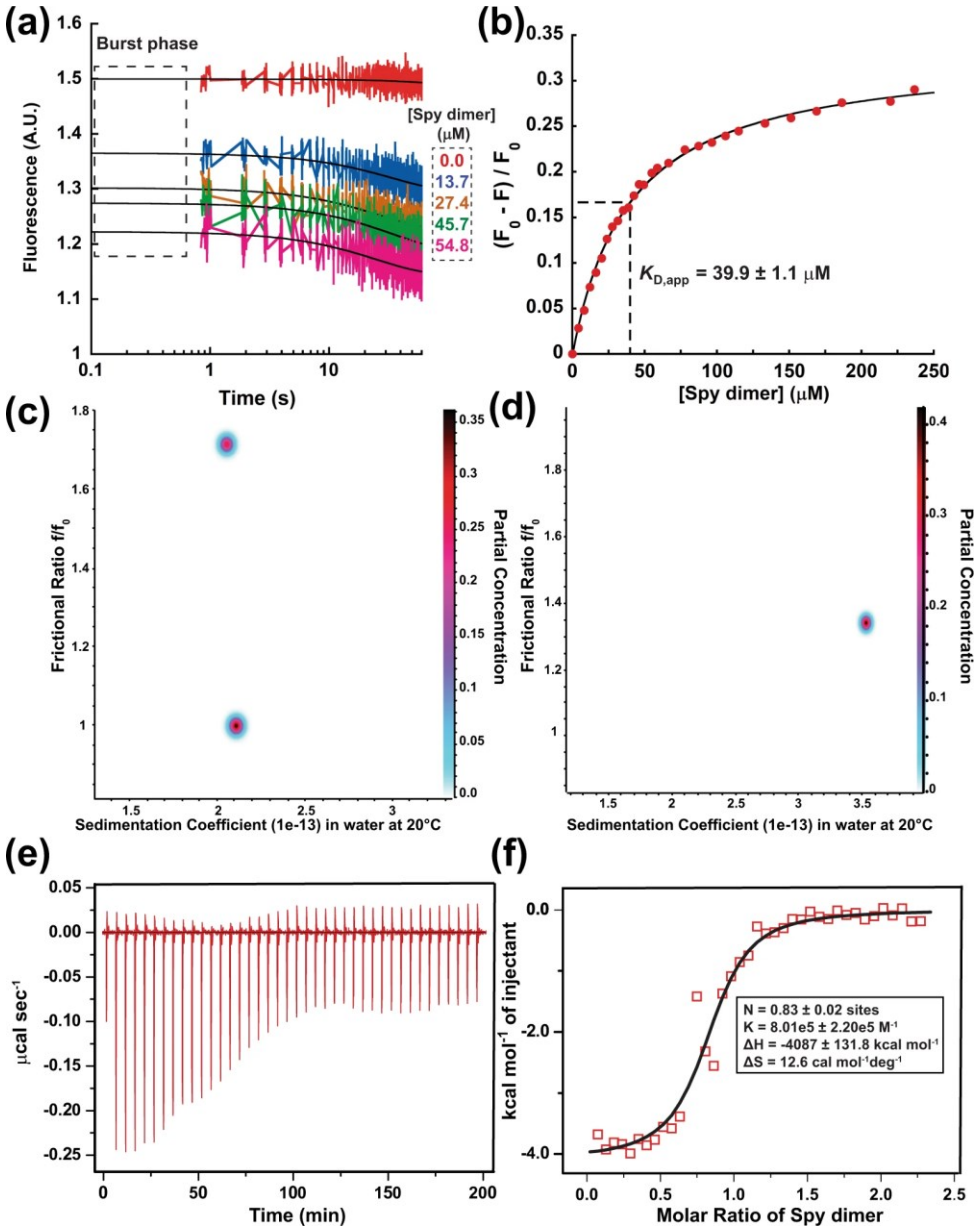


Figure 2-3: Interaction of Spy and native apoflavodoxin.

(a) *Change in intrinsic tryptophan fluorescence of 0.09 μM AnFld WT with time in the absence and presence of various concentrations of Spy (0–54.8 μM after mixing). Time axis is shown in logarithmic scale. AU arbitrary units. The black lines shown for each Spy concentration are best fits of each dataset to a single exponential equation. (**b**) AnFld WT-Spy binding as monitored by intrinsic tryptophan fluorescence of AnFld. Spy dimer was titrated into 17.2 μM native AnFld WT at 25 °C. The binding isotherm is plotted with change in fluorescence (F) as normalized by fluorescence intensity relative to native AnFld fluorescence (F_0) in the y-axis. Black line shows the fit to a one-site binding model. **(c)**, **(d)** SV-AUC experiments showing frictional ratios (f/f_0) and sedimentation coefficients for **c.** AnFld WT alone and **(d)** AnFld WT in the presence of 2.5-fold excess of Spy dimer in 40 mM HEPES-KOH (pH 7.5), 25 mM NaCl. Data were analyzed by two-dimensional sedimentation analysis (2DSA) followed by analysis with a genetic algorithm, further validated by a Monte Carlo analysis. **(e),** (**f**) ITC analysis for the interaction of Spy and AnFld F98N at 10 °C. The low temperature minimizes enthalpy changes due to the binding-induced unfolding of AnFld. Five hundred and fifty micromolar Spy dimer in the syringe was*

*titrated into 50 μM AnFld F98N in the cell. The thermogram in **(e)** was integrated and fit to a one-site binding model, as shown in **(f)** to obtain thermodynamic binding parameters. N stoichiometry of binding, K association constant, ΔH enthalpy change, ΔS entropy change. Values reported are the mean ± s.e.m. of the fit.*

We observed a slow exponential decrease in the fluorescence of AnFld in the presence of Spy (k_{obs} = 0.04 s^{-1}) after the burst phase (**Fig. 2-3b**). This decrease is not due to Spy quenching AnFld fluorescence because the final fluorescence decreases hyperbolically with Spy concentration (**Fig. 2-4c**). Instead, we suggest Spy binding leads to a conformational change in AnFld. This conformational change may involve the Spy-mediated unfolding of a minor population of native AnFld to a thermal intermediate-like state (I^T), which was observed to be less fluorescent than the native state[63]. The native and I^T states have largely similar structures, including two hydrophobic cores formed by α-helices packing onto a central β-sheet. Only the loop connecting β4 and α4 and the long loop splitting the strand β5 are disordered in I^T. To further explore the nature of bound AnFld, we used a previously characterized mutant, F98N, that mimics the I^T state[64]. At saturating concentrations of Spy, the fluorescence of the AnFld WT–Spy complex at 340 nm is 0.33 times that of the fluorescence of AnFld WT alone (**Fig. 2-3b**), a value remarkably close to the fluorescence of AnFld F98N (0.39 times that of the WT) (**Fig. 2-4d**). This similarity in fluorescence values supports the idea that the binding of Spy is coupled to the partial unfolding of AnFld to an I^T-like state. Additional support for this proposed partial unfolding comes from studying the interaction of Spy and AnFld in low ionic strength solutions. AUC experiments revealed that in HEPES buffer containing very low salt concentrations (25 mM NaCl), AnFld WT also exists in predominately two conformations that have the same sedimentation coefficient but different frictional ratios (f/f0) (**Fig. 2-4e** and **Fig. 2-3c**). The major conformation (native

state) contributes ~60% of the absorbance signal and has a frictional ratio close to one, typical of globular proteins. The minor conformation contributes ~40% of the signal and resembles the partially unfolded I^T in that it is more expanded with a frictional ratio of ~1.7. In the presence of excess amounts of Spy, only a single AnFld WT species is seen (**Fig. 2-4e** and **Fig. 2-3d**), indicating that the conformational heterogeneity of AnFld seen under non-denaturing conditions is lost upon interaction with Spy. We cannot ascertain whether, in the low ionic strength buffer, only the I^T state of AnFld binds Spy and, in doing so, pulls the equilibrium away from the native state or whether Spy also binds native AnFld and partially unfolds it to a state that resembles I^T. Under physiological salt concentrations, however, most of the AnFld molecules populate the native state (**Fig. 2-4f**). Isothermal titration calorimetry (ITC) experiments showed that Spy binds the WT and F98N mutant with dissociation constants of 28.2 and 1.8 μM, respectively (**Fig. 2-4g, h** and **Fig. 2-3e, f**). The ~16-fold higher affinity of Spy for the I^T mimic than for WT should shift the equilibrium of Spy-bound AnFld towards the more expanded thermal intermediate. Thus, the molecular basis for binding-induced partial unfolding of native substrates appears to be the higher affinity of Spy to more unfolded conformations[55].

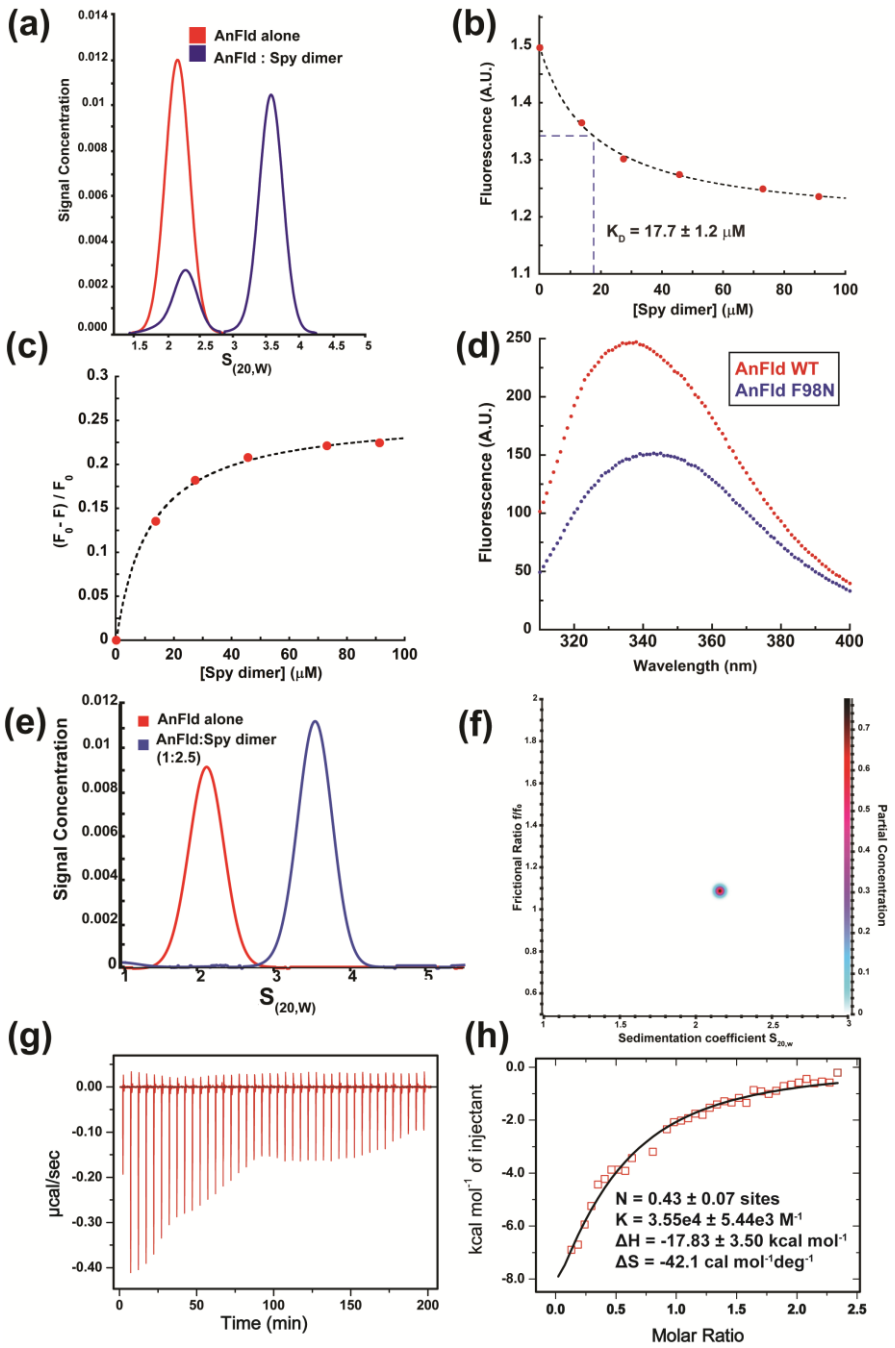


Figure 2-4. Interaction of Spy with AnFld WT and F98N

(a) *Sedimentation profiles of AnFld WT in the absence (red) or presence of Spy dimer (in 1:1 mole ratio) (blue) from analytical ultracentrifugation experiments conducted at 48,000 rpm and 20 °C. The protein samples were prepared in 10 mM potassium phosphate pH 7.5 buffer, and the concentration of AnFld was 8.8 μM.* ***(b)*** *initial and* ***(c)*** *final fluorescence was obtained from the fit of the kinetic data sets for the interaction of Spy and 0.09 μM AnFld WT. The plot shows the relative fluorescence signal normalized to fluorescence of AnFld WT alone (F_0), i.e., $\Delta F/F_0$ where $\Delta F = F_0 - F$, as a function of Spy concentration. The black line shows the best fit for a one-site binding model.* ***(d)*** *Fluorescence emission spectra of 5 μM each of AnFld WT (red) and F98N (blue) in HN buffer at 25 °C. The excitation wavelength used was 295 nm and emission were monitored from 310 nm to 400 nm. The slit widths were 5 nm each. Note the*

fluorescence intensity at 340 nm. ***(e)*** *Sedimentation of AnFld WT in the absence (red) or presence of Spy dimer (in 1:2.5 mole ratio) (blue) was conducted at 48,000 rpm and 20 °C. The protein samples were prepared in 40 mM HEPES-KOH pH 7.5, 25 mM NaCl buffer, and the concentration of AnFld was 8.7 μM.* ***(f)*** *SV-AUC experiment showing frictional ratios (f/f_0) and sedimentation coefficient for 17.6 μM AnFld WT in 40 mM HEPES-KOH (pH 7.5), 100 mM NaCl at 50,000 rpm, and 22 °C. The plot shows the two-dimensional sedimentation analysis (2DSA) followed by analysis with a genetic algorithm, further validated by a Monte Carlo analysis.* ***(g)*** *–* ***(h)*** *ITC data showing the binding of Spy dimer to AnFld WT at 10 °C. The titration syringe was filled with 550 μM Spy dimer and 50 μM AnFld WT was filled in the cell. The thermogram in* ***(g)*** *was integrated and fit to a one-site binding model, as shown in* ***(h)*** *to obtain thermodynamic binding parameters. N, the stoichiometry of binding; K, association constant; ΔH, enthalpy change; ΔS, entropy change. Values reported are the mean ± s.e.m. of the fit. A.U., arbitrary units.*

2.2.3 Spy binds tightly to a partially unfolded mutant of ApoFld

If the binding of Spy tightly to partially or fully unfolded conformations of AnFld inhibits AnFld refolding, then Spy should have a higher affinity to destabilized mutants of AnFld than it does to the WT protein. We combined two destabilizing mutations in AnFld (L105A and I109A) to generate AnFld 2A[65]. Sedimentation equilibrium experiments showed that this protein is monomeric and soluble (**Fig. 2-6a**). Multiple lines of evidence indicate AnFld 2A is partially unfolded. CD spectra indicate the A2 mutant has lost about 30% of its alpha-helical content (**Fig. 2-5a** and **Table 1**), though its β-sheet content is unchanged (**Table 1**). Other evidence that the 2A mutant is partially unfolded includes a decrease in intensity and a redshift in tryptophan fluorescence (**Fig. 2-5b**) and an increase in ANS (1-anilino-8-napthalene sulfonic acid) binding (**Fig. 2-5c**). Sedimentation profiles show that the mutant is expanded relative to the WT protein (**Fig. 2-6a**) and urea denaturation profiles indicate that the mutant is highly destabilized, (**Fig. 2-5d**) and it unfolds more rapidly than WT (**Fig. 2-6b**). Limited dispersion in the ^{1}H dimension of the 2D [$^{1}H-^{15}N$] HSQC–TROSY NMR spectra of the mutant between 7.5 and 8.5 ppm indicates that it is mostly disordered, though 48 residues maintain native-like backbone ^{15}N shifts (**Fig. 2-5e** and **Fig. 2-6c, d**)[66]. Together, our experiments indicate that the AnFld 2A mutant is partially unfolded. Collectively these observations

support our model that Spy kinetically traps AnFld in a partially unfolded state by rapid and tight binding, thus functioning as a holdase.

Table 1: Secondary structure determination from far UV CD spectra

Protein	Secondary structural element	Percentage
AnFld WT	Helix	35.4 %
	Sheet	16.6 %
	Turn	5.3 %
	Others	42.6 %
AnFld 2A	Helix	25.2 %
	Sheet	16.2 %
	Turn	12.3 %
	Others	46.3 %

Fluorescence experiments indicate that Spy rapidly binds the 2A mutant with a $K_{D\ app}$ of 0.1 μM (**Fig. 2-5f**). ITC and kinetic unfolding experiments indicate that AnFld 2A exists in two binding-competent conformations interacting with Spy with comparable affinities of 0.38 and 0.23 μM (**Fig. 2-6e, f**). These affinities measured for AnFld 2A are similar to the 0.4 μM K_D found for Spy's interaction with the U* state of AnFld WT.

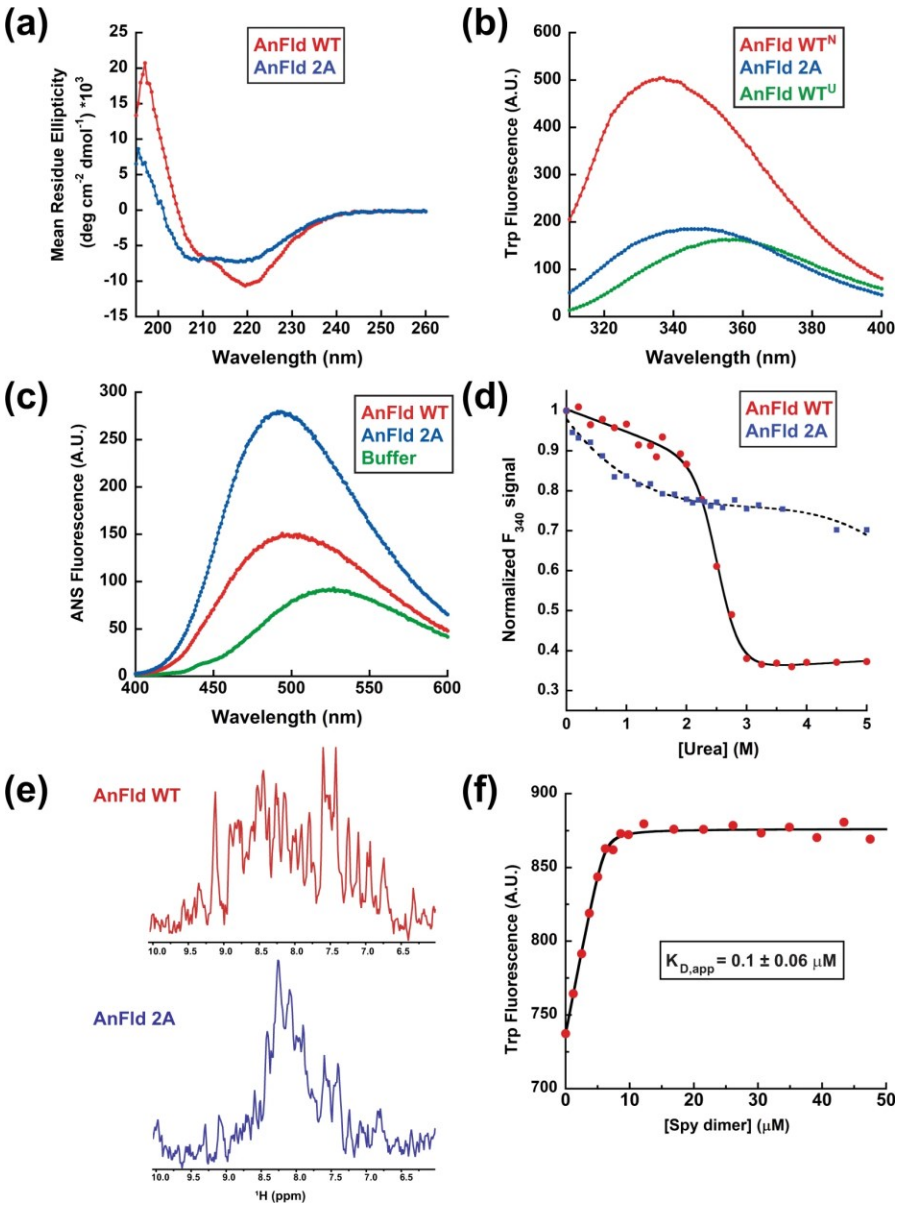


Figure 2-5: Characterization of the AnFld2A mutant and its interaction with Spy.

(a) *Far-UV CD spectra of AnFld WT (red) and AnFld2A mutant (green) in 0.1 M potassium phosphate buffer (pH 7.5) at 25 °C.* ***(b)*** *Intrinsic tryptophan fluorescence emission spectra of the native states of AnFld WT (green), AnFld 2A mutant (blue), and denatured AnFld WT in 5 M urea (red) in HN buffer at 25 °C.* ***(c)*** *Fluorescence emission spectra of ANS alone (red) in HN buffer and in the presence of AnFld WT (green) or AnFld 2A mutant (blue) in HN buffer at 25 °C. The fluorescence intensity of ANS increases 2.8-fold upon binding the mutant compared to the WT.* ***(d)*** *Urea-induced denaturation of AnFld WT (red) and AnFld 2A mutant (blue) monitored by tryptophan fluorescence emission at 340 nm in HN buffer at 25 °C. The data were normalized to each protein's fluorescence signal (F_{340}) in the absence of urea. The denaturation curve of the WT protein was fitted to a two-state unfolding model (solid black line) and that of AnFld 2A was fitted to a third-degree polynomial (broken black line).* ***(e)*** *^{1}H dimension of the 2D [^{1}H-^{15}N] HSQC–TROSY spectra of 0.2 mM AnFld WT (red) and 2 A (blue).* ***(f)*** *Titration of Spy dimer to 10 μM AnFld 2A as monitored by the intrinsic tryptophan fluorescence. The black line shows the fit of the fluorescence intensity as a function of Spy dimer concentration to a tight binding model. A.U. arbitrary units. The value reported is the mean ± s.e.m. of the fit.*

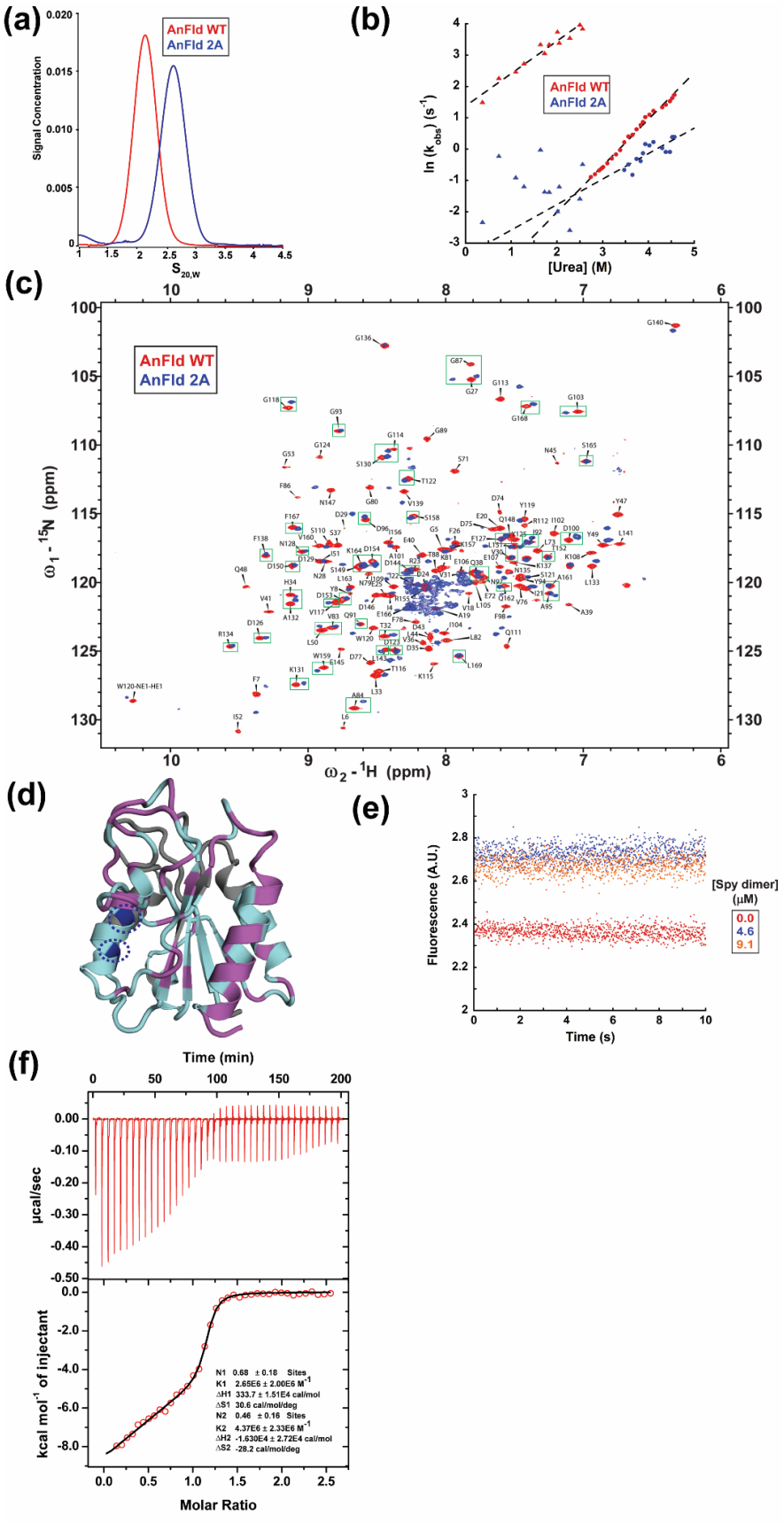

(a)
AnFld WT
AnFld 2A
Signal Concentration
S20,w
(b)
AnFld WT
AnFld 2A
ln (kobs) (s-1)
[Urea] (M)
(c)
AnFld WT
AnFld 2A
ω1 - 15N (ppm)
ω2 - 1H (ppm)
(d)
(e)
Fluorescence (A.U.)
Time (s)
[Spy dimer] (μM)
0.0
4.6
9.1
(f)
Time (min)
μcal/sec
kcal mol-1 of injectant
Molar Ratio
N1 0.68 ± 0.18 Sites
K1 2.65E6 ± 2.00E6 M-1
ΔH1 333.7 ± 1.51E4 cal/mol
ΔS1 30.6 cal/mol/deg
N2 0.46 ± 0.16 Sites
K2 4.37E6 ± 2.33E6 M-1
ΔH2 -1.630E4 ± 2.72E4 cal/mol
ΔS2 -28.2 cal/mol/deg

Figure 2-6: Further characterization of AnFld2A and Spy-AnFld2A interactions

(a) *Sedimentation of (17.6 μM each) AnFld WT (red) and 2A mutant (blue) at 50,000 rpm and 22 °C. The protein samples were prepared in HN buffer.* ***(b)*** *Major and minor unfolding phases of AnFld WT (red and blue circles, respectively) and those of the 2A mutant (red and blue triangles, respectively) were obtained by fitting the unfolding kinetics in various urea concentrations to a sum of two exponentials. Black lines show the linear fit of the natural logarithm of k_{obs} for the major unfolding phase as a function of urea concentration according to the equation $ln\, k_{uf} = ln\, k_{uf}^{H_2O} + \frac{m_{uf}}{RT}[Urea]$, where $k_{uf}^{H_2O}$ is the unfolding rate constant in 0 M urea, m_{uf} is the urea-dependence of the unfolding rate constant, R is the gas constant, and T is the temperature in Kelvin. The $k_{uf}^{H_2O}$ values for the major unfolding phase of AnFld WT and 2A mutant are 0.007 s^{-1} and 3.96 s^{-1}, respectively. The m_{uf} values for the major unfolding phase of the AnFld WT and 2A mutant are 0.88 kcal mol^{-1} K^{-1} and 0.61 kcal mol^{-1} K^{-1}, respectively. The $k_{uf}^{H_2O}$ and m_{uf} values for the minor phase in AnFld WT unfolding are 0.033 s^{-1} and 0.48 kcal mol^{-1} K^{-1}, respectively. Observing two unfolding phases for the A2 mutant is consistent with two populations at equilibrium without aggregation.* ***(c)*** *2D [^{1}H-^{15}N] HSQC- TROSY spectra of 0.2 mM AnFld WT (red) and 2A (blue). The cross peaks in the spectra of the 2A mutant that could be assigned unambiguously are shown in green boxes along with the corresponding peak in the WT protein.* ***(d)*** *Crystal structure of AnFld WT showing the residues that could be assigned in the 2A mutant (magenta), residues L105 and I109 that are mutated (blue and encircled in blue dotted circles), unassigned residues in the spectra of the mutant (cyan) and residues also unassigned in the WT protein (gray).* ***(e)*** *The interaction of AnFld2A mutant and Spy was studied in a stopped-flow fluorometer by monitoring tryptophan fluorescence of AnFld following 11.5-fold dilution in HN buffer containing 0 μM (red), 4.6 μM (blue) and 9.1 μM (orange) Spy dimer.* ***(f)*** *ITC data showing the binding of Spy dimer to AnFld 2A mutant at 5 °C. The titration syringe was filled with 1.2 mM Spy dimer and 100 μM AnFld2A was filled in the cell. The thermogram was fitted to a two-site binding model, which represents two conformations of AnFld 2A in equilibrium that can both bind Spy. The thermodynamic parameters for the two binding species are shown. N, the stoichiometry of binding; K, association constant; ΔH, enthalpy change; ΔS, entropy change.*

2.2.4 Spy forms a compact chaperone–client complex with ApoFld

To probe the interaction between Spy and AnFld, we employed native mass spectrometry (MS), which detects non-covalent protein–protein interactions in the gas phase. Native MS is particularly well-suited for analyzing mixtures such as co-incubated proteins and their complexes, as each population is separated by mass[67]. When coupled to ion mobility (IM), native IM-MS enables gas phase structural measurements of the mass-resolved proteins and complexes. The IM arrival time distributions (ATD) of individual ions can be converted to rotationally averaged collision cross-section distributions (CCSD). The centroid of these distributions can be reported as the rotationally averaged collision cross section ($^{DT}CCS_{N2}$, CCS). CCS is a charge-independent structural parameter that correlates with the surface area of the ion[68]. It is

sensitive to domain-level rearrangements that occur in solution prior to ionization. For instance, a rigid protein with a single solution conformation exhibits a Gaussian ATD/CCSD, whereas a dynamic protein exhibits a multimodal or extended ATD/CCSD[69].

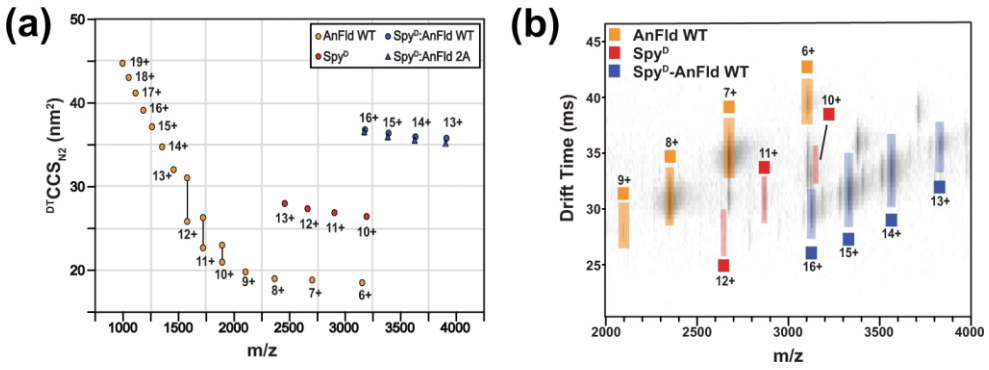


Figure 2-7: Native IM-MS to study AnFld and its complex with Spy.

(a) *Collision cross-sections of AnFld WT (orange circles), Spy (red circles), Spy: AnFld WT (blue circles) versus m/z. The complex of Spy: AnFld 2A (blue triangles) is also plotted to compare the complexes. (**b**) Ion mobility mass spectrum of AnFld WT (orange) co-incubated with Spy^D (red). Complexes of Spy^D: AnFld WT are highlighted in blue.*

Using native IM-MS, we first compared WT AnFld and the destabilized 2A mutant. The CCSs of AnFld 2A are similar to those of AnFld WT, indicating that both samples have a similar structural ensemble (**Fig. 2-8b** and **Table 2**). The small difference observed in CCS can be attributed to the small shift in mass of 2A vs. WT. The broad CCSD of both AnFld's (**Fig. 2-7** and **Fig. 2-8b**), ranging from 18 to 20 nm^2 for the compact states and 20 to 45 nm^2 for the extended unfolded states, indicates that the proteins are structurally heterogeneous in solution. The highest charge states exhibit collision cross-sections equal to those observed for proteins nearly two to three times the size of apoflavodoxin[70,71]. Such conformational heterogeneity is expected as AnFld is destabilized in low ionic strength buffers like the 20 mM ammonium acetate solution used in these experiments[63,72]. Using this buffer is fortuitous as it allows us to detect the

native state and otherwise sparsely populated partially folded states of AnFld. The degree of protein folding can also be assessed by inspection of ions' charge state distribution (CSD). Natively folded compact proteins with rigid structures have a CSD spanning 2–4 charge states. In contrast, both AnFld proteins analyzed in this work have a broad multimodal CSD spanning 14 charge states, including “completely unfolded” higher charge states (12+ to 19+), some “partially unfolded” charge states (9+ to 11+), and some “folded”, compact charge states (6+ to 8+) (**Fig. 2-8**) confirming that these proteins populate various conformations in solution. The unfolded 9+ to 19+ charge states account for 15 ± 2% of total AnFld 2A signal intensity while only accounting for 1.5 ± 0.6% of total AnFld WT signal intensity (**Fig. 2-8a**). Although AnFld WT and AnFld 2A both apparently sample the same structural ensemble, the equilibrium of AnFld 2A is clearly more shifted towards the unfolded states than AnFld WT. Next, we analyzed complexes formed between Spy, and both AnFld variants. Three CSDs were observed between 2000 and 4000 m/z corresponding to AnFld, Spy and Spy in complex with AnFld (**Fig. 2-7** and **Table 2**). The Spy: AnFld complexes, containing either AnFld WT or AnFld 2A, consistently exhibit lower CCSs than the larger more unfolded forms of AnFld (z = 15+ to 19+).

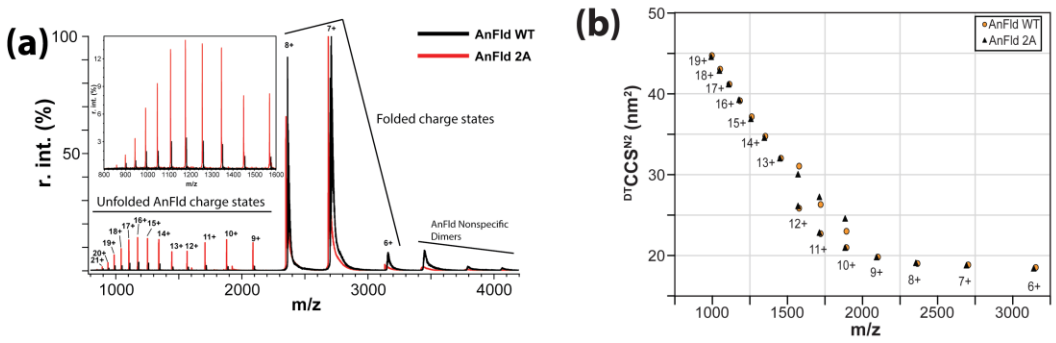


Figure 2-8: IM-MS spectra and CCSDs of AnFld WT and 2A mutant

(a) *The mass spectra of AnFld WT and AnFld 2A were normalized by intensity and overlaid to compare the charge state distribution and their relative intensities. The AnFld WT mass spectrum is shown as a black trace, and the AnFld 2A mass spectrum is shown as a red trace. The inset shows the mass range between 800 and 1600 m/z zoomed in. The intensities of these low m/z "unfolded" forms of AnFld 2A are greater than those of AnFld WT, indicating that more of AnFld 2A is unfolded in solution.* ***(b)*** *The* $^{DT}CCS_{N2}$ *of AnFld WT and AnFld 2A plotted together to show the similarity in CCS between the two forms of AnFld.*

Table 2: CCSDs of AnFld (WT and 2A), Spy, and Spy-AnFld complex

ID	**z**	**AnFld WT** (18920.4 ± 0.4 Da)			**AnFld 2A** (18,837.0 ± 0.3 Da)		
		CCS (nm^2)	**STD**	**R.S.D. (%)**	**CCS (nm^2)**	**STD**	**R.S.D. (%)**
AnFld	19	44.73	0.27	0.60	44.56	0.07	0.16
	18	43.03	0.06	0.15	42.86	0.08	0.19
	17	41.17	0.20	0.49	41.17	0.17	0.42
	16	39.13	0.16	0.41	39.26	0.16	0.40
	15	37.16	0.17	0.47	36.90	0.12	0.32
	14	34.75	0.20	0.58	34.57	0.09	0.25
	13	32.03	0.15	0.48	32.02	0.37	1.15
	12	25.85	0.16	0.63	26.14	0.13	0.50
	12	31.06	0.26	0.83	30.05	0.10	0.33
	11	22.71	0.18	0.78	22.84	0.07	0.30
	11	26.31	0.48	1.81	27.30	0.31	1.13
	10	21.00	0.12	0.58	20.99	0.06	0.29
	10	23.01	0.51	2.20	24.60	0.03	0.14
	9	19.83	0.04	0.19	19.81	0.01	0.06
	8	19.02	0.02	0.12	19.11	0.06	0.30
	7	18.86	0.03	0.14	18.81	0.05	0.27
	6	18.53	0.03	0.14	18.43	0.05	0.27
ID	**z**	**Spy-AnFld WT** (50,849 ± 1 Da)			**Spy-AnFld 2A** (50,762 ± 1 Da)		
		CCS (nm^2)	**S.D.**	**R.S.D. (%)**	**CCS (nm^2)**	**S.D.**	**R.S.D. (%)**
Spy-AnFld	13	36.79	0.02	0.05	36.51	0.06	0.16
	12	36.39	0.07	0.19	35.91	0.06	0.16
	11	35.94	0.06	0.16	35.50	0.08	0.22
	10	35.79	0.19	0.54	35.26	0.05	0.14
ID	**z**	**Spy dimer** (31,931 ± 2 Da)					
		CCS (nm^2)	**S.D.**	**R.S.D. (%)**			
Spy dimer	11	26.29	0.08	0.30			
	10	25.65	0.15	0.59			
	9	24.87	0.21	0.83			
	8	23.81	0.21	0.89			

Since Spy does not undergo major changes in structure and dynamics upon substrate binding, we attribute the changes in CCS upon complex formation to changes in the conformation of AnFld upon Spy binding[9,10]. Measurements of intermolecular paramagnetic relaxation enhancement effects in the Spy–Im7 complex by NMR spectroscopy showed that the binding of Spy induces compaction of a dynamic ensemble of unfolded Im7 L18A L19A L37A[10]. The compaction of an unfolded client in complex with Spy is proposed to facilitate intramolecular contacts necessary to fold while bound[8,10]. Collectively, our results from spectroscopy-based kinetic experiments and IM-MS reveal that such client compaction may also stabilize kinetically trapped states that cannot fold in a chaperone-bound form.

2.2.5 The chaperone-binding surface of ApoFld is dynamic

Using NMR spectroscopy, we set out to characterize the binding surface of AnFld at the residue level. Upon the addition of sub-stoichiometric amounts of unlabeled Spy, we observed a loss of signal intensity, as quantified by the peak heights in the (2D) [^{15}N, ^{1}H]–TROSY NMR spectra of [U-^{2}H, ^{15}N, ^{13}C]-labeled AnFld (**Fig. 2-10a**). Considering both the increase in size upon complex formation (size of apo-AnFld is ~19 kDa and that of the AnFld-Spy complex is ~50 kDa) and the micromolar affinity of the interaction, we reasoned that line broadening could be caused by increased tumbling time and/ or increased chemical exchange in the µs-ms time scale. Therefore, we attempted to map the binding surface of AnFld for Spy by plotting the ratio of peak intensities of AnFld residues in the absence and presence of 0.75× unlabeled Spy. Residues were considered to be part of the interaction site if they underwent greater than 2/3 of the average loss in peak intensity upon Spy binding (**Fig. 2-9a**). These residues largely

colocalize on one surface of native AnFld (**Fig. 2-9b, c** and **Fig. 2-10b**). However, we cannot eliminate the possibility of additional binding sites on AnFld as chemical shift assignments were unavailable for stretches of contiguous residues G9–S17 at the phosphate-binding site and C54–Y70 at the FMN-binding site[73].

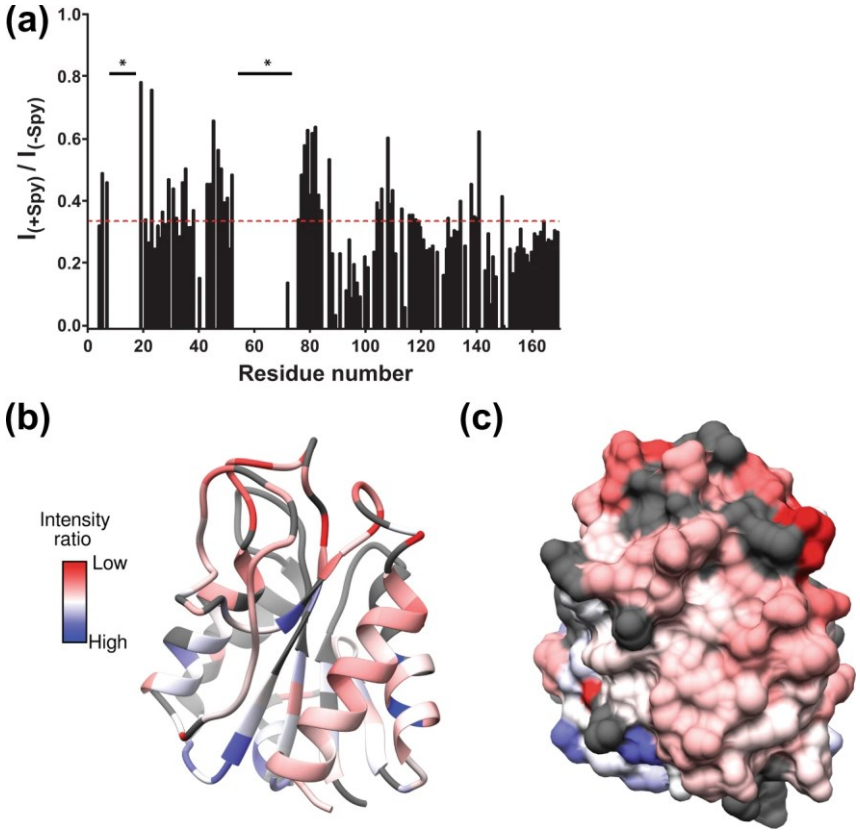


Figure 2-9: NMR spectroscopy to map the Spy binding surface of AnFld.

(a) *A plot of the ratios of intensities of each assigned cross peak in the [^{1}H-^{15}N]–TROSY NMR spectra of AnFld WT in the presence of a sub-stoichiometric amount of Spy (0.75×) and in the absence of Spy [$I_{(+Spy)}$ and $I_{(-Spy)}$, respectively]. The asterisks denote unassigned residues in the spectra, and the dashed red line shows the cutoff of average peak broadening for all mapped residues.* ***(b)****,* ***(c)*** *A red-to-blue color scale has been used to map the NMR peak intensity ratios on the crystal structure of AnFld (Protein Data Bank ID code 1FTG) in* ***(b)*** *ribbon and* ***(c)*** *surface representations. Red and blue represent the highest and lowest intensity ratios in the dataset at 0.784 and −0.002, respectively. A lower intensity ratio, i.e., [$I_{(+Spy)}/I_{(-Spy)}$] is indicative of a higher degree of backbone perturbation in the presence of Spy. The unassigned residues in AnFld are shown in gray.*

We note that several residues in the two loops, 90–100 and 120–139, that were found to interact with Spy based on their intensity ratios have previously been found to be

unstructured in the otherwise natively folded thermal intermediate of AnFld (**Fig. 2-10c**)[74]. The binding of Spy to these dynamic regions on AnFld may help stabilize the partially disordered AnFld in a high-affinity complex. This is consistent with our native IM-MS experiments, which indicate that the AnFld-Spy complex is more compact than unfolded conformations of unbound AnFld in the solution. Interestingly, Monte Carlo simulations using a nucleation-growth model showed that the folding nucleus of AnFld comprises the loop 90–100 and the segment 120–160 that includes α4, β5, most of the C-terminal helix α5, and connecting loops; in other words, almost completely overlaps with the binding surface for Spy[73]. While the implications of this finding are unclear, we speculate that Spy may kinetically trap refolding AnFld in the non-native U* state by tightly binding to a folding nucleus formed early during folding.

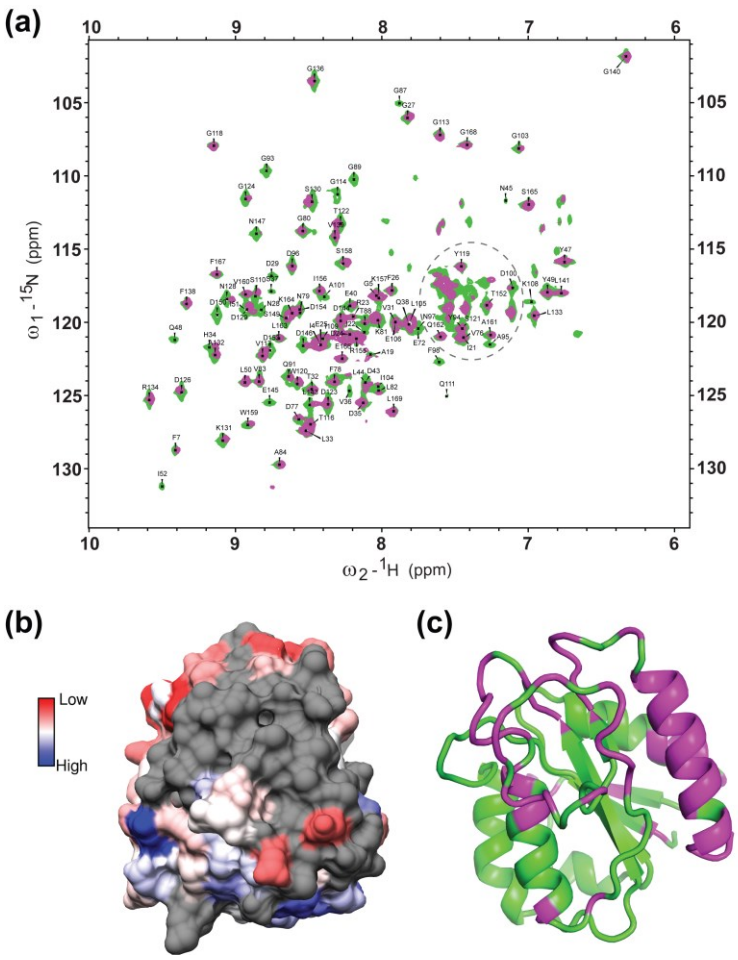


Figure 2-10: Interaction of Spy with AnFld

(a) *Two-dimensional [^{1}H-^{15}N] HSQC- TROSY NMR spectra of [U- ^{2}H, ^{15}N, ^{13}C] -labeled AnFld in the absence (green) and presence of 0.75x Spy (magenta) in 90% (v/v) 50 mM potassium phosphate buffer (pH 7.5) (in $^{1}H_2O$) and 10% (v/v) $^{2}H_2O$. Gray dotted circle shows a region with several overlapping peaks that could not be assigned. 70.5% of the published amide cross peaks of AnFld (BMRB accession number 5011) could be assigned unambiguously in our experiments.* ***(b)*** *Structural representation of the backside of the Spy-binding surface in AnFld mapped by NMR peak intensity ratios using a red-to-blue color scale. The unassigned residues are shown in gray.* ***(c)*** *Crystal structure of AnFld showing the residues that interact with Spy in magenta. Amide cross peaks of these residues have intensity ratios lower than the average value of 0.334. The rest of the protein is colored green, including unassigned residues in our experiments.*

2.2.6 Spy kinetically traps the misfolded molten globule of ApoFld

We have shown that Spy's ability to inhibit the folding of Anabaena flavodoxin apparently results from its tightly binding to partially unfolded states. Given the striking differences between Spy's mechanism of action for AnFld and simple substrates like Im7 and Fyn SH3, we wished to test the holdase activity of Spy with apoflavodoxin from

another species, *Azotobacter vinelandii*, that populates a stable molten globule intermediate[57]. Molten globules are a class of compact folding intermediates that lack persistent native tertiary structure but possess a significant amount of native-like secondary structure[75]. The flavodoxin from *Azotobacter vinelandii*, which we term AzoFld, shares 47% sequence identity and is structurally very similar (RMSD of 0.6 Å) to the flavodoxin from *Anabaena* with which we did all of the proceeding experiments (**Fig. 2-11a**). AzoFld folds via a complex four-state mechanism that involves both the formation of a stable off-pathway molten globule intermediate (MG_{off}) that acts as a kinetic trap and an obligate high-energy on-pathway intermediate (I_{on}) (**Fig. 2-1a**)[57]. The major kinetic phase during refolding reports on the rate-limiting step of unfolding the molten globule intermediate formed rapidly after diluting from denaturant [57]. The sedimentation profiles of AzoFld alone and in the presence of Spy showed that the Spy dimer forms a 1:1 complex with AzoFld (**Fig. 2-11b**). By monitoring the refolding of urea- denatured AzoFld in buffer containing increasing concentrations of Spy, we observed that Spy decreased both the folding rate and yield of AzoFld (**Fig. 2-12a**). Like our observations with AnFld, the decreasing fluorescence amplitude with increasing Spy concentrations indicates that AzoFld cannot undergo efficient folding when bound to Spy. Kinetic refolding of AzoFld occurs in two phases in the absence of Spy. The minor folding phase disappears with even low amounts of Spy (> 4.6 μM). The k_{obs} for the major folding phase decreased hyperbolically with increasing concentrations of Spy and asymptotically reached the very low value of 0.09 s^{-1} at saturating Spy concentrations (**Fig. 2-12b**). The simplest explanation for this observation is that Spy binds more tightly to both MG_{off} and U states than it does to native AzoFld, thereby allowing Spy to inhibit

the folding of AzoFld. The small non-zero k_{obs} value suggests that the unfolding of MG_{off} can still occur while bound to Spy, albeit at a much slower rate. Notably, the U and MG_{off} states are indistinguishable species without denaturant due to a gradual second-order-like collapse transition of the unfolded protein to the MG_{off} state under equilibrium conditions[76]. Given that the MG_{off} state is reported to be aggregation-prone in the presence of cytosol-mimicking crowding agents, binding to such misfolded intermediates is additional strong evidence for Spy having a holdase activity for flavodoxin-fold substrates[77,78].

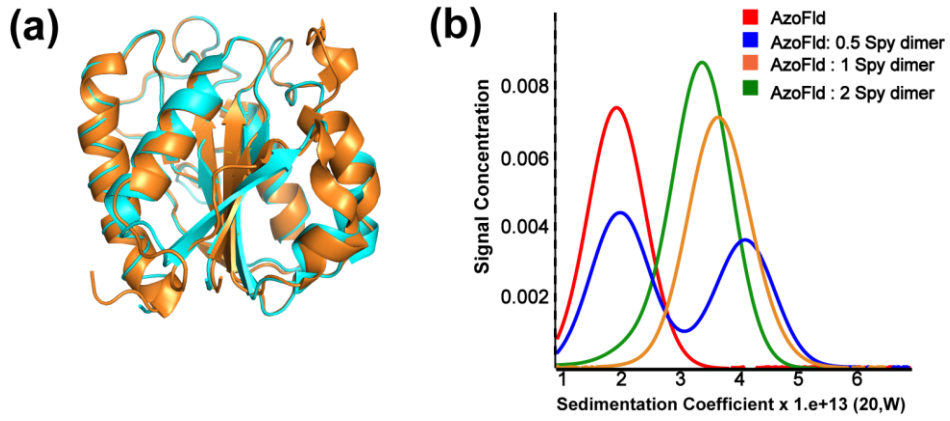


Figure 2-11: Structure of AzoFld and its interaction with Spy

(a) *Structural alignment of AnFld (PDB ID 1FTG) (cyan) and AzoFld (PDB ID 1YOB) (orange).* ***(b)*** *Samples of AzoFld (13.8 µM) in the absence or presence of 0.5x, 1x and 2x Spy dimer in KP buffer were analyzed by analytical ultracentrifugation at 50,000 rpm and 25°C.*

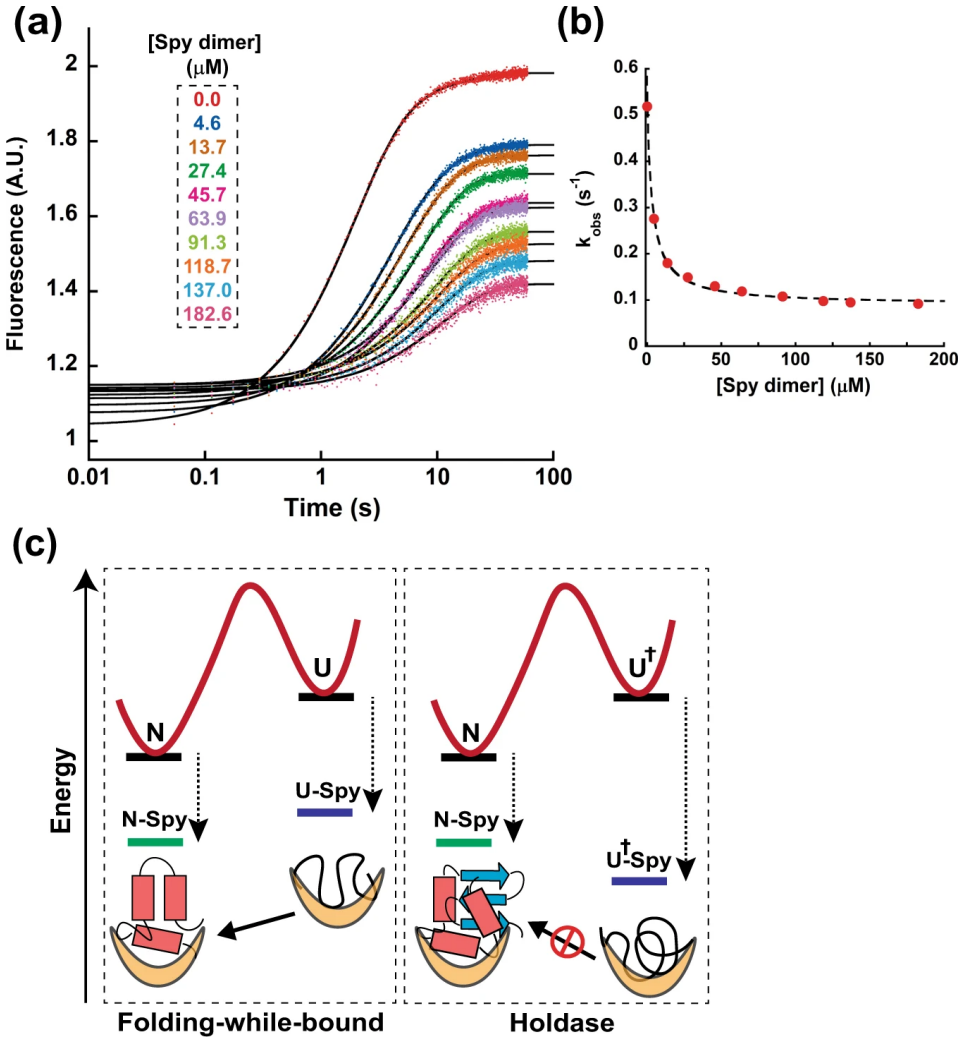


Figure 2-12: Holdase activity of Spy inhibits the folding of apoflavodoxin.

(a) *Fluorescence traces of the refolding of 0.1 μM AzoFld C69A in the presence of various concentrations of Spy dimer (0–182.6 μM after mixing) in KP buffer. The time axis is shown in a logarithmic scale. (**b**) A plot of the observed major rate constant for AzoFld refolding (k_{obs}) as a function of Spy dimer concentration. **(c)** Mechanism of substrate-specific action of Spy. Simple substrates can fold while continuously chaperone-bound because Spy binds the native (N) and unfolded (U) states with weak affinities. Even at saturating concentrations of Spy, folding occurs because the N–Spy complex is the most energetically stable form of the substrate. However, large and topologically complex substrates cannot fold while bound because of Spy's strong affinity for (partially) unfolded states (denoted generically by $U^{\dagger}$). Spy acts as a holdase for such substrates by sequestering aggregation-prone $U^{\dagger}$ states. Kinetically, such holdase-like action makes Spy a competitive inhibitor of folding because the $U^{\dagger}$–Spy complex is the most energetically stable form of the substrate.*

2.3 Discussion

Misfolded and unfolded states of cellular proteins are recognized by ATP-independent holdase chaperones that act by sequestering them to prevent aggregation[4]. Our results with the small bacterial ATP-independent chaperone Spy showed that simple model proteins like Im7 and Fyn SH3 could fold while continuously bound to Spy, provided the different folding states bind the chaperone only weakly[55]. The relative affinity of Spy to the different folding states of the substrate appears to determine whether substrate folding occurs while bound to Spy. We previously observed that the Spy mutants Q100L and H96L inhibit SH3 refolding[55]. The mechanistic basis for the unfolding activity of these Spy mutants was an increase in the binding affinity of Spy towards unfolded SH3. Furthermore, the affinity of Spy to unfolded SH3 correlates with the folding rate constant in the presence of Spy. These Spy mutants are also potent holdase chaperones that suppress the *in vitro* aggregation of denatured substrates like aldolase and α-lactalbumin[79]. We postulate that only by binding loosely and with comparable affinity to unfolded, intermediate, and native states can a chaperone allow folding while bound without interfering with the substrate protein's function. In contrast, binding and sequestration of aggregation-prone denatured states is a hallmark of ATP-independent holdase chaperones[56]. Nanomolar affinities for unfolded polypeptides and folding intermediates have been observed for the interaction of substrates with ATP-independent chaperones such as trigger factor, Skp, SecB, and sHsps[80–84]. For instance, the periplasmic holdase chaperone Skp binds the unfolded outer membrane protein OmpA with a K_D of 22 nM. The chaperone activity of the human small heat-shock protein Hsp27 is enhanced upon stress-induced

phosphorylation. The phosphorylation mimic of Hsp27, S15D/S78D/S82D binds a destabilized T4 lysozyme variant with an apparent affinity of 4 nM[85]. Although the exact mechanism varies, these chaperones generally maintain substrates in an unfolded state, thereby preventing aggregation. That these chaperones bind tightly to non-native states with different degrees of unfolding intuitively supports their holdase activity. In the case of Im7 and Spy, the various Im7 folding states bound to Spy can interconvert while chaperone bound, apparently at least partly due to the weaker nature of the interaction[9,10]. In the case of AnFld, the interaction of partially unfolded states with Spy is substantially tighter, and this tighter binding apparently prevents the folding of the chaperone-bound substrate. These observations make intuitive sense; tight binding of the chaperone will hinder conformational transitions in the bound substrate and thus prevent folding while bound. The inability of sHsps to work as folding-while-bound chaperones may be partly due to their reliance on short-range hydrophobic interactions to recognize non-native clients[84]. This mode of client recognition also prevents them from binding native proteins, which normally do not expose hydrophobic surfaces. Tight binding to unfolded and intermediate states render sHsps incapable of spontaneous release and refolding of clients. The chaperone function of sHsps is therefore limited to sequestering aggregation-prone unfolded proteins [83]. On the other hand, Spy utilizes mainly long-range electrostatic interactions to recognize and bind unfolded substrates. Although the Spy-unfolded substrate complex is also stabilized by hydrophobic interactions, for substrates like Im7 and SH3, binding is weak and allows the substrate to fold to its native state[7,55]. Our work with apoflavodoxin shows that the affinity of Spy

to partially unfolded substrates is crucial in determining whether Spy's chaperone function mirrors that of sHsps (**Fig. 2-12c**).

Taken together, this work on Spy and apoflavodoxin and our past work on the effect of Spy on the folding of Im7 and Fyn SH3 establish the central role of the folding landscape of the substrate in determining chaperone mechanisms. Spy can allow simple substrates like Im7 and SH3 to fold while bound because unfolded conformations can transition to the native state on the cradle of Spy. However, relatively large and topologically complex substrates like apoflavodoxin will be energetically stabilized in a chaperone-bound unfolded conformation. Our results, therefore, highlight the importance of substrate-specific mechanisms rather than the paradigmatic roles of ATP-independent chaperones in protein quality control. Intuitively, multi-functionality may be advantageous for cellular and organismal fitness as it enables chaperones to bind to different conformational states of substrates.

2.4 Materials and Methods

2.4.1 Protein expression and purification

The genes for AnFld and AzoFld C69A were codon-optimized for expression in *E. coli* and synthesized by GenScript. The flavodoxin genes were inserted into a pET28b-based vector with an N-terminal His_6-SUMO tag by standard restriction digestion and ligation-based cloning using the enzymes BamHI and XhoI. The F98N, L105A and I109A mutations were inserted into AnFld simultaneously using site-directed mutagenesis with a QuikChange kit (Agilent). The primers used for mutagenesis are listed in **Table 3**. Plasmid with the Spy gene was obtained from lab collection[3]. Spy was purified as described previously[55]. The modified pET28b plasmid containing the flavodoxin gene was transformed into *E. coli* BL21 (DE3) cells for protein expression. Cells were grown at 37 °C to early log phase in PEM media containing 100 µg/ml kanamycin and then shifted to 20 °C for induction by addition of 0.1 mM IPTG. After 16h of protein expression, cells were pelleted by centrifugation for 20 min at 5000 x g at 4°C. The pellet was resuspended in lysis buffer containing 50 mM sodium phosphate, 400 mM NaCl, 10% glycerol, pH 8.0 and 0.05 µg/ml DNase and protease inhibitor cocktail (complete mini EDTA-free, Roche). Following sonication for 8 min on ice, the cell lysate was centrifuged twice at 36,000 x g for 30 min at 4 °C to pellet the cell debris. The supernatant was loaded with 20 mM imidazole onto a Ni-His Trap column equilibrated with water and lysis buffer. The column was washed with lysis buffer containing 30 mM imidazole. His-tagged flavodoxin was eluted by adding lysis buffer containing 500 mM imidazole to the column. 500 µg of ULP1 protease was added to the elution to cleave the His-SUMO tag from the protein. The elution was then dialyzed

against buffer containing 40 mM Tris, 400 mM NaCl, pH 8.0 at 4 °C overnight. The cleaved His tag was removed following ULP1 digestion and dialysis by passing through a Ni-His Trap column that had been equilibrated with lysis buffer containing 30 mM imidazole. The flow through was collected and 25 mM Tris pH 8.0 buffer was added to dilute the sample 5-fold. The sample was then loaded onto a HiTrap Q column and the protein was eluted with a NaCl gradient in 25 mM Tris pH 8.0. The fractions containing holoflavodoxin were yellow in color and were discarded. The colorless fractions contain purified apoflavodoxin and hence were collected and concentrated. The protein was then passed through a HiLoad Superdex 75 column equilibrated with 40 mM HEPES-KOH (pH 7.5), 100 mM NaCl (HN buffer) for desalting and buffer exchange. The purified protein was then aliquoted, flash frozen in liquid nitrogen, and stored at −80 °C. Expression and purification of AnFld F98N, AnFld2A, and AzoFld were done with the same protocol. ^{15}N-labeled AnFld WT and 2A mutant were expressed in M9 minimal media containing $^{15}NH_4Cl$ as the sole nitrogen source. The protocol of Marley and coworkers was used to express [*U*- ^{2}H, ^{15}N, ^{13}C] –labeled AnFld WT in M9 minimal media prepared in $^{2}H_2O$ and supplemented with $^{15}NH_4Cl$, ^{13}C-glucose and ISOGRO growth supplement (Sigma)[86]. The isotopically labeled proteins were purified with the same protocol as that used for unlabeled proteins except that the holo-protein fractions were also collected. Apoflavodoxin was obtained by trichloroacetic acid precipitation to remove the bound FMN[57]. Protein concentrations were determined using the molar extinction coefficients of 34,100 $M^{-1}cm^{-1}$ for AnFld (WT, F98N and 2A mutants) and 29,000 $M^{-1}cm^{-1}$ for AzoFld[87,88].

Table 3: Primers used in this study

Primer	Mutation	DNA Sequence (5' to 3')
RM01	L105A, I109A (forward primer)	CACGCTGGCTAGCCTTTTCCTCCGCAATGCCGATCGCGTCTTGGA
RM02	L105A, I109A (reverse primer)	TCCAAGACGCGATCGGCATTGCGGAGGAAAAGGCTAGCCAGCGTG
RM03	F98N (forward primer)	CGGTTATGCGGATAACAACCAAGACGCGATCGGC
RM04	F98N (reverse primer)	GCCGATCGCGTCTTGGTTGTTATCCGCATAACCG

2.4.2 Analytical ultracentrifugation

Analytical ultracentrifugation (AUC), sedimentation was monitored by absorbance at 280 nm at 50 krpm for AnFld WT and AnFld 2A mutant. For the binding experiments containing Spy and AnFld or AzoFld, sedimentation was conducted at 48 krpm and monitored by absorbance at 280 nm. Sedimentation velocity AUC (SV-AUC) was carried out using 450 µl loaded into two-sector epon centerpieces with 1.2 cm path-length in an An60Ti rotor in a Beckman Optima Xl-I analytical ultracentrifuge. Measurements were completed in intensity mode. All SV-AUC data were analyzed using UltraScan 4 software, version 4.0 and fitting procedures were completed on XSEDE clusters at the Texas Advanced Computing Center (Lonestar, Stampede) through the UltraScan Science Gateway (https://www.xsede.org/web/guest/gateways-listing)[89]. The partial specific volume (vbar) of the protein samples (Spy, AnFld, AnFld^{2A} mutant, AzoFld) was estimated within UltraScan III based on the protein sequence[49]. Raw intensity data were converted to pseudo-absorbance by using the intensity of the air above the meniscus as a reference, then edited. Next, 2-dimensional sedimentation spectrum analysis (2DSA) was conducted to subtract time-invariant noise, and the meniscus was fit using ten points in a 0.05 cm range[90]. The arrays were fit using an S range of 1–8, an f/f_0 range of 1–4 with 64 grid points for each, ten uniform grid

repetitions, and 400 simulation points. 2DSA was then repeated at the determined meniscus to fit radially invariant and time-invariant noise together using ten iterations. The 2DSA analysis was refined by a genetic algorithm, which helps to define the solutes and to eliminate any false positive solutions. The distribution between AnFld native state and AnFld intermediate I^T was determined by integrating the peak intensities after performing the genetic algorithm analysis. The results from the genetic algorithm were evaluated using a Monte Carlo algorithm[91].

2.4.3 Isothermal titration calorimetry

Thermodynamic analysis of Spy binding to AnFld WT and AnFld F98N was conducted by ITC experiments at 10 °C using a MicroCal iTC200 (Malvern Instruments). The proteins were buffer exchanged into assay buffer containing 40 mM HEPES-KOH (pH 7.5), 100 mM NaCl, using a PD10 desalting column. For the WT and F98N mutants, 550 μM Spy dimer was added in the titration syringe and 50 μM AnFld in the cell. For the 2A mutant, 1.2 mM Spy dimer was added in the titration syringe and 100 μM AnFld in the cell. The fitting of thermograms to one-site or two-site models was done using Origin software (OriginLab) that was provided with the instrument.

2.4.4 Stopped-flow fluorescence

All stopped-flow experiments were conducted in a KinTek SF-300X stopped-flow instrument at 25 °C. The tryptophan residues of both the flavodoxins were excited at 295 nm and the emitted fluorescence signal was collected using a 320 nm long-pass filter provided with the instrument. Urea dependence of AnFld refolding was monitored by 11.5-fold dilution of 1.04 μM AnFld that was denatured in HN buffer containing 5 M

urea into HN buffer containing various urea concentrations in a 1:10.5 mix. AnFld unfolding was monitored in a similar way, except that native AnFld in HN buffer was diluted 11.5-fold in HN buffer containing increasing concentrations of urea between 0 and 5 M. The final concentration of AnFld in all cases was 0.09 μM. 10–12 traces of 10 s were acquired for each data point and averaged. For refolding experiments at 0.43–1.9 M final urea concentration, two kinetic phases were observed. However, one exponential was sufficient to fit the traces from experiments at higher urea concentrations. Average traces from unfolding experiments at 2.05–3.38 M final urea concentration were fit to a single kinetic phase, whereas data from experiments at higher final urea concentrations were fit to two phases. AnFld refolding in the presence of Spy was monitored by mixing in a 1:10.5 mix, 1.04 μM AnFld in HN buffer containing 5 M urea with HN buffer containing various Spy concentrations (0–27.39 μM Spy dimer after mixing). 10–12 traces of 10 s duration were acquired for each data point and averaged. AzoFld refolding in the presence of Spy was monitored using tryptophan fluorescence intensity as signal in a similar way as AnFld, i.e., by 11.5-fold dilution of 1.4 μM AzoFld in 10 mM potassium phosphate pH 6.0 (KP) buffer containing 6 M urea into KP buffer containing various concentrations of Spy (0–228.3 μM Spy dimer after mixing). Each kinetic trace is an average of 7–10 traces that were collected for 50 s. For stopped-flow experiments to monitor the interaction of native AnFld and Spy, 1.04 μM AnFld in HN buffer was diluted 11.5-fold into HN buffer containing various concentrations of Spy. Each kinetic trace was an average of 5–6 traces of 50 s duration that were acquired with the auto-shutter on to minimize photobleaching during the experiment. The average trace for each data point was fit to a single exponential. In all

the experiments with Spy, the background signal of Spy due to its tyrosine residues was removed. This was done by collecting shots of each concentration of Spy diluted into buffer and subtracting the average signal of Spy from the kinetic trace of flavodoxin folding at that Spy concentration. Individual traces were fitted to sums of exponentials in KaleidaGraph (Synergy Software) to obtain observed rate constants.

2.4.5 Fluorescence spectroscopy

All fluorescence spectra were acquired on a Cary Eclipse Fluorescence Spectrophotometer. In tryptophan fluorescence-based titrations, binding was monitored by tryptophan fluorescence of apoflavodoxin; 17.2 μM of AnFld WT or 10 μM AnFld 2A in 1 ml of HN buffer was titrated with increasing concentrations of Spy in a 10 mm quartz cuvette containing a magnetic stirrer. After each addition, the contents of the cuvette were manually mixed with a pipette and allowed to equilibrate inside the spectrophotometer for 2–3 min. Spectra were collected at 25 °C with an excitation wavelength of 295 nm and emission wavelengths scanning from 310–400 nm. Both excitation and emission slit widths were set to 5 nm. For each addition of Spy, spectra were collected in triplicate over a period of 0.5 min and then averaged. Average spectra for each titration point were corrected for dilution from the titrant additions. For AnFld WT, the fluorescence emission at 340 nm for every titration point was normalized with respect to the emission of flavodoxin alone, plotted as a function of Spy dimer concentration, and the binding isotherm fitted to a square hyperbola equation, Eq. (1). For AnFld 2A, the fluorescence emission at 340 nm was plotted as a function of Spy dimer concentration and fitted to a quadratic equation for tight binding, Eq. (2).

$$(F - F_i) / F_i = (B_{max} * [L]) / (K_D + [L]) \quad (1)$$

$$F = 0.5 * (F_0 (C + [L] + K_D - \sqrt{(C + [L] + K_D)^2 - 4 * C * [L]}) + I \quad (2)$$

where, F is the fluorescence signal, F_i is the fluorescence of AnFld WT in the absence of Spy, [L] is the Spy dimer concentration, B_{max} is the maximum fluorescence change, K_D is the dissociation constant, F_0 is a fluorescence correction factor, C is the concentration of AnFld 2A, and I is the y-intercept. The fluorescence emission spectra of native AnFld WT and the AnFld 2A mutant were recorded by dissolving the protein in HN buffer, and the spectra of denatured AnFld WT was recorded in HN buffer containing 5 M urea. All the proteins were used at a concentration of 10 μM. The spectra were collected at 25 °C with an excitation wavelength of 295 nm and emission wavelengths from 310–400 nm. Both the slit widths were 5 nm. Each spectrum was acquired three times and averaged.

ANS binding assays were conducted by monitoring the fluorescence emission spectra of 1-anilino-8-napthalene sulphonate (ANS) in HN buffer in the presence and absence of 5 μM AnFld WT or 5 μM 2A mutant. The final concentration of ANS was 250 μM. ANS was excited with a wavelength of 385 nm and emission was recorded from 400 to 700 nm. The excitation and slit widths were 5 nm. The spectra shown are each an average of five scans.

For chemical denaturation by urea, 10 μM AnFldWT or 5 μM AnFld2A was added to 1 ml HN buffer containing various concentrations of urea ranging from 0–5 M and incubated for ~30 min at room temperature. Fluorescence spectra were recorded at 25 °C with excitation wavelength at 280 nm and emission monitored at 340 nm.

2.4.6 Circular dichroism spectroscopy

Far-UV CD spectra were acquired on a Jasco J-1500 CD spectrometer. The spectra were recorded in a quartz cell with a path length of 0.1 cm using ~0.2 mg/ml of protein concentration. The spectra were acquired from 260 to 190 nm with a 0.5 nm data interval, 50 nm/min scan speed, and at 25 °C unless otherwise specified. Five scans were collected and averaged. The mean residue ellipticity (*MRE*) was calculated using Eq. (3)

$$MRE = (MRW * \Theta) / (10 * d * c) \quad (3)$$

Where, *MRW* is the mean residue weight, *Θ* is the observed ellipticity (degrees), *d* is the path length (0.1 cm), and *c* is the protein concentration in mg/ml. The average CD spectra were deconvoluted with the BestSel algorithm to determine the secondary structure of the proteins in terms of the percentages of α-helix, β-sheet, turn, and unordered components[92].

2.4.7 Nuclear magnetic resonance spectroscopy

The [^{1}H-^{15}N] – TROSY HSQC NMR spectra of AnFld WT and the 2A mutant were recorded using Watergate solvent suppression at 25 °C on a Bruker 600 MHz spectrometer equipped with the (^{1}H/^{19}F)-X broadband CryoProbe Prodigy. [*U*- ^{15}N] – labeled proteins were exchanged into NMR sample buffer (50 mM potassium phosphate buffer (pH 7.5) containing 0.1 M NaCl) using a PD10 desalting column (GE Healthcare). The samples were 0.2 mM of protein dissolved in a mixture of 90% (v/v) NMR sample buffer (in $^{1}H_2O$) and 10% (v/v) $^{2}H_2O$. The spectra were acquired using a total of 2048 complex points in t2 and 512 increments in t1 with 8 scans per increment over a spectral width of 9.6 kHz and 2.1 kHz in the ^{1}H and ^{15}N dimensions, respectively. To

study the interaction of AnFld and Spy, [*U*- ^{2}H, ^{15}N, ^{13}C] –labeled AnFld and Spy were exchanged into 50 mM potassium phosphate buffer (pH 7.5) without any salt using a PD10 desalting column (GE Healthcare). ^{15}N-TROSY spectra were collected at 25 °C with 0.2 mM AnFld dissolved in 90% (v/v) buffer (in $^{1}H_2O$) and 10% (v/v) $^{2}H_2O$ and then titrations were done by addition of appropriate volume of Spy to the NMR tube. A total of 2048 data points in t2 dimension and 256 increments in t1 with 16 scans per increment were acquired. The spectra were processed using NMRPipe suite, analyzed, and plotted in NMRFAM Sparky[93,94].

2.4.8 Native ion mobility-mass spectrometry

Prior to analysis, Spy and AnFld were exchanged from storage buffer into 20 mM ammonium acetate (pH 7.4) by Micro Bio-spin size exclusion spin columns (Bio-Rad). Spy-AnFld complexes were formed by mixing Spy dimer and AnFld at 1 μM final concentration in 20 mM ammonium acetate solution (pH 7.4) and co-incubating on ice for ~5 min. Aqueous samples were transferred into the gas phase via nano electrospray ionization (nESI) using a capillary voltage of 900–1,000 V. The ions were analyzed on a modified Agilent 6560 drift tube ion mobility quadrupole time-of-flight mass spectrometer (DTIM-Q-TOF) (Agilent)[71,95]. The drying gas (N_2) temperature was set to 25 °C, and the flow was reduced to 1.5 l/min. The instrument was operated with 99.9999% nitrogen gas, front funnel pressure of 4.52 torr, trap funnel pressure of 3.80 torr, drift tube pressure of 3.947 torr, quadrupole pressure of 2.48 $*10^{-5}$ torr, and TOF flight tube pressure of 1.699 $*10^{-7}$ torr. The IM separation was carried out under low field conditions (~18.5 V/cm), and the ion arrival time distributions were fit to Gaussian functions using using CIUSuite2[96]. The centroids of the fit Guassian functions were

converted to rotationally averaged collision cross section ($^{DT}CCS_{N2}$) by a single-field calibration using Agilent Tune Mix ions as previously described[97]. The $^{DT}CCS_{N2}$ measurements reported are the average of three technical replicate measurements, and the error reported is the standard deviation. Raw data were analyzed using Agilent IM-MS Browser 10.0 and mMass[98,99].

Chapter 3 Characterizing the Structural Dynamics of SERF-RNA Interactions[3]

3.1 Background

Ribonucleoprotein (RNP) granules are founding members of a growing class of biomolecular condensates formed through multivalent interactions of RNA and RNA-binding proteins (RBPs)[100–102]. Numerous examples have established the role of RNA-RNA and RNA-protein interactions in regulating the composition, material properties, and dynamics of RNP granules[103–108]. However, as pointed out recently, the structural and dynamic properties of proteins and nucleic acids that give rise to these condensates are largely unclear[109]. In the early stages of phase separation, critical nuclei are formed by spontaneous assembly of mesoscopic RNA-RBP clusters[110,111]. Unfortunately, the sub-diffractive assemblies in these early stages are invisible to diffraction-limited microscopy[112]. We think the lack of suitable model systems has hindered the comprehensive structural analysis of early RNA-RBP clusters. Most extensively studied RBPs, such as FUS, hnRNPA1, and G3BP1, are large and have multi-domain architecture[104,113–117]. Additionally, lncRNAs and mRNAs that have been identified in RNP condensates are kilobases in length and adopt complex structures[118–121]. Although these native target RNAs are ideal for *in vitro* reconstitution, synthesizing

[3] The contents of this chapter reflect a manuscript in preparation titled "*Mechanisms underlying RNA-protein condensation"*. Kevin Namitz from the Penn State Huck Institutes of the Life Sciences performed AUC experiments and analyzed the data. Microscopic imaging was conducted by Matthew Crotteau and myself from the lab of James Bardwell, in collaboration with Jian Guan from the lab of Ursula Jakob. Recombinant proteins were expressed and purified by Ke Wan from the lab of James Bardwell.

them in bulk for phase separation assays is difficult or prohibitively expensive. These considerations make it challenging, if not impossible, to probe the structural properties of RBPs and RNAs in the context of phase separation (**Fig. 3-1**). Hence, there is a need for tractable RBP-RNA pairs that can serve as model systems to characterize biomolecular structure in condensed phases quantitatively. To maximize our potential for obtaining structural insights into phase- separating RNP complexes, we have limited our selection to the remarkably small number of well-characterized RNAs that also contain ubiquitous secondary structure elements such as stems, loops, and bulges[122]. A model RNA-RBP system allows for elucidating the structure and dynamics of mesoscopic clusters in sub-saturated solutions that give rise to a macroscopic condensed phase[111,123].

The small EDRK-rich factor (SERF) is a small eukaryotic protein with a highly conserved N-terminal disordered region. SERF acts as a modifier of amyloid formation[41,124]. We have previously reported that the SERF accelerates the primary nucleation of amyloid proteins by forming fuzzy complexes[125]. The highly conserved N-terminal ~20 residues of SERF (N-SERF domain) form an unstructured module tethered to a folded region[126]. Recent work has shown that the human SERF1a protein localizes to nucleoli and can facilitate the nucleolar incorporation of fluorescently labeled RNA[45]. While the exact role of SERF in amyloid diseases is still unclear, the cellular localization and physicochemical properties of SERF suggest that RNA binding is its main function in the native cellular context. Over 95% of RBPs that lack canonical RNA binding domains do not exhibit sequence specificity for their RNA targets[32,47]. Since RBPs often recognize bulges and internal loops in RNA, we selected a model helix-junction-helix

motif to build our synthetic RNA-RBP condensate[127]. Small secondary structural modules like helices and junctions are often used to study RNA folding[128–133]. The apical region (residues 17 - 45) of the trans-activation response element (TAR) RNA from human immunodeficiency virus type-1 (HIV-1) is the best-studied helix-junction-helix motif[134,135]. This 29-nucleotide RNA (depicted in **Fig. 3-4a**) contains two A-form helical stems (Domain I and Domain II), a trinucleotide bulge, and a hexanucleotide apical loop. TAR undergoes dynamic motions over a broad range of time scales[136]. Here, we have established SERF and TAR as a tractable *in vitro* RBP-RNA complex and comprehensively characterized their interaction using various complementary methods. Using solution NMR, we show that the conserved N-SERF domain interacts with TAR and perturbs its structure. Although binding to TAR slows down the backbone dynamics of SERF, it does not result in the adoption of a folded structure of SERF in the bound state. SERF:TAR complexes are compact and not as conformationally heterogenous as SERF: amyloid protein complexes. Future work will investigate how multivalent interactions drive the assembly of binary SERF-TAR condensates.

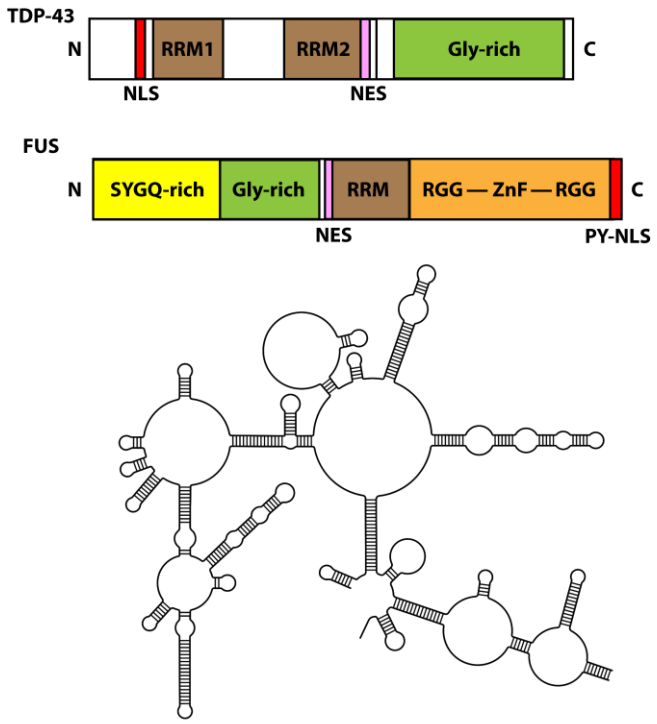


Figure 3-1: Modular architecture of RBPs and RNAs in RNP condensates

The two RNA-binding proteins shown here, TDP-43 and FUS are large and contain multiple domains[137]*. The RNA molecule depicted here reflects the secondary structural complexity of mRNAs and lncRNAs often found in RNP granules. RRM. RNA-recognition motif; ZnF. Zinc-finger; NLS. Nuclear localization signal; NES. Nuclear export signal; PY-NLS Proline-tyrosine-rich nuclear localization signal. (The cartoon RNA was created with BioRender.com)*

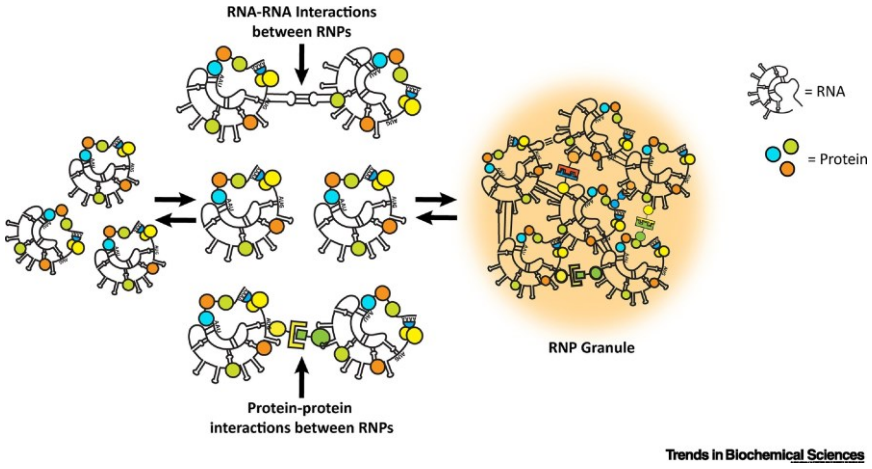


Figure 3-2: Homotypic and heterotypic interactions that driving RNP granule formation[138]

Ribonucleoprotein (RNP) granules are assembled through multivalent protein–protein, RNA–RNA, and protein–RNA interactions. The molecular mechanisms that explain the connection between these multivalent interactions and the stability, composition, and material properties of RNP granules are not well understood. (Reprinted from Tauber, D. et al. Trends. Biochem. Sci. (2020) with permission from Springer)

3.2 Results

3.2.1 NMR illuminates a dynamic, disordered N-SERF domain

As the first step towards comprehensively characterizing the interactions of SERF and TAR in our model RNP complex, we utilized solution NMR methods to obtain high-resolution information about the conformational dynamics of SERF. The $^{1}H,^{15}N$-HSQC spectrum of SERF, like most IDPs, shows narrow chemical shift dispersion in the direct (^{1}H) dimension at room temperature (**Fig. 3-7a**). To circumvent the limitation of poor signal dispersion, we turned to ^{13}C Direct-Detect NMR spectroscopy[139–143]. The $^{15}N,^{13}C$-CON spectrum shows excellent dispersion, allowing us to obtain unambiguous backbone and sidechain resonance assignments for every residue in SERF (**Fig. 3-3**). $^{13}C^{\alpha}$, $^{13}C^{\beta}$, $^{13}C'$, and ^{15}N chemical shifts are sensitive to backbone dihedral angles and report on site-specific conformational (dis)order[144]. We measured per-residue secondary structure propensities as the deviations of the measured $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts from the sequence-specific random coil chemical shifts for SERF calculated by the *ncIDP* predictor [145]. The C-terminal region (CTR) of SERF spanning E36-K65 shows a high propensity to adopt an α-helix. The human SERF1A protein also possess a structured C-terminal region with helical propensity[124]. On the other hand, the conserved N-SERF domain displays conformational disorder but also exhibits small but measurable α-helical content. This is evident from the slight positive deviations in secondary chemical shifts from typical random coil values observed in IDPs (**Fig. 3-4a**). Further evidence for the disordered nature of the N-SERF segment comes from the backbone spin relaxation measurements[146,147]. Both T_1 and T_2 rates are sensitive to global tumbling time and local backbone motions. T_2 rates are also sensitive to

conformational exchange in the µs-ms time scales. ^{15}N-T_1 profile of SERF is largely featureless, and the high baseline value is consistent with an overall disordered protein. The elevated T_1 rates observed in the vicinity of P31 compared to the baseline may be indicative of enhanced flexibility, potentially arising from hinge-like motions around the proline residue (**Fig. 3-4b**). Comparing the residue-specific ^{15}N-T_2 relaxation profiles of the N-SERF domain and the CTR reveals backbone motions occurring at two different timescales, indicating the presence of long µs to short ms dynamics within the N-SERF domain (**Fig. 3-4c**). Collectively, spin relaxation parameters reveal fast backbone motions and slower conformational exchange in the N-SERF domain in the free form. Despite the lack of a persistent structure, transient tertiary contacts and sparsely sampled compact conformations are often detected in the heterogenous ensembles of IDPs like α-synuclein[148,149]. Paramagnetic relaxation enhancement methods (PREs) are highly sensitive to sparsely populated states with long-range structural contacts[150]. We conducted paramagnetic relaxation enhancement (PRE) experiments to probe potential transient tertiary contacts between N- and C-terminal regions of SERF by monitoring resonance broadening in the ^{15}N,^{13}C-CON spectrum. Since wildtype SERF lacks cysteines, we covalently attached a cysteine-specific spin-label to a SERF variant with Ala63 mutated to Cys. Any residue within ~20 Å of the labeled residue will decrease in intensity due to enhanced relaxation[151]. We noted peak broadening for residues in the vicinity of the C63 site, consistent with the presence of local backbone contacts. However, no PRE effects were detected for residues distant in sequence from C63 (**Fig. 3-4d**). Since PREs are highly sensitive to sparsely sampled compact conformations,

the absence of PREs strongly indicates that no long-range interactions exist between the N- and C-terminal regions of SERF, possibly due to strong intrachain repulsion.

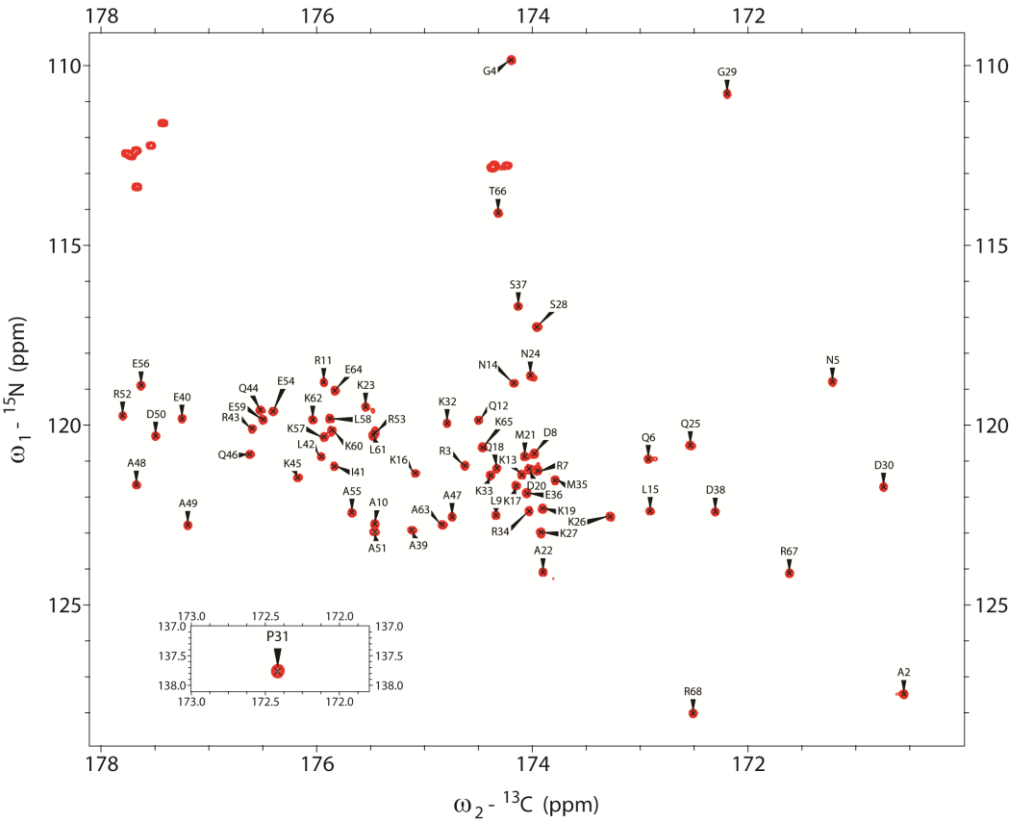


Figure 3-3: $^{15}N,^{13}C$-CON spectrum of SERF

Cross-resonances for every residue of SERF were assigned using ^{13}C direct detection NMR.

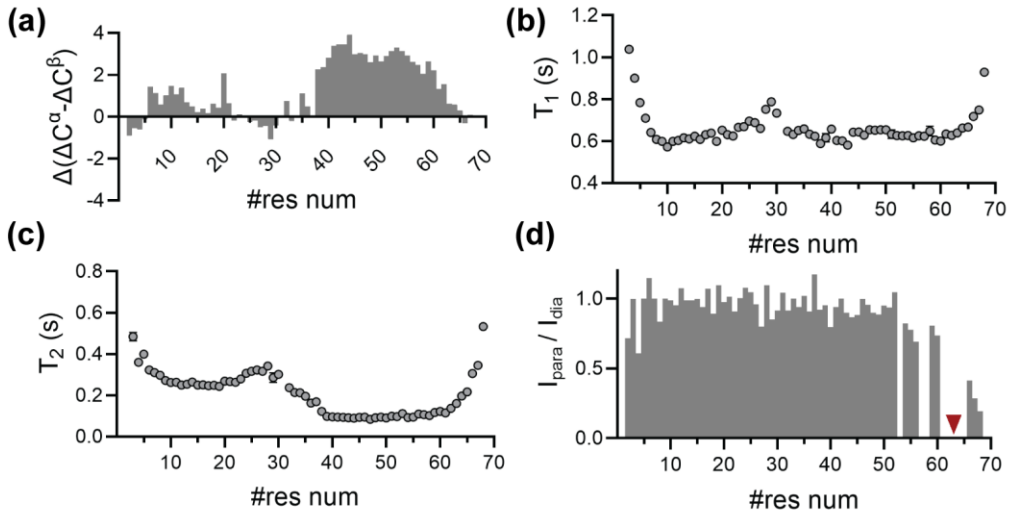


Figure 3-4: NMR characterization of SERF

(a) *Secondary structure propensities reported as residue-specific secondary chemical shifts ($\Delta(C^{\alpha}_{calc}-C^{\alpha}_{RC}) - (C^{\beta}_{calc}-C^{\beta}_{RC})$). Backbone relaxation profiles* ***(b)*** *^{15}N-T_1,* ***(c)*** *^{15}N-T_2. Error bars represent uncertainty in fitting.* ***(d)*** *PRE profiles for MTSSL conjugation to C63 (red arrow), in the A63C mutant of SERF. The profile in* ***(d)*** *reports data collected using the carbon-start ^{13}C,^{15}N-CON. Residues adjacent to C63 could not be assigned.*

3.2.2 A homogenous SERF-TAR complex

Polycationic proteins and nucleic acids can form polyelectrolyte complexes like those observed when oppositely charged synthetic polymers are mixed in an aqueous solution[152]. We first tested whether SERF binds with a well-characterized RNA, the HIV-1 TAR stem-loop, using 3'-FAM-labeled TAR and fluorescence anisotropy. By monitoring the increase in fluorescence anisotropy, we found that SERF binds TAR with an apparent dissociation constant (K_D^{app}) of 0.74 μM (**Fig. 3-5b**). Sedimentation profiles from analytical ultracentrifugation experiments showed that SERF and TAR forms a homogenous complex with a 1:1 stoichiometry (**Fig. 3-5c**).

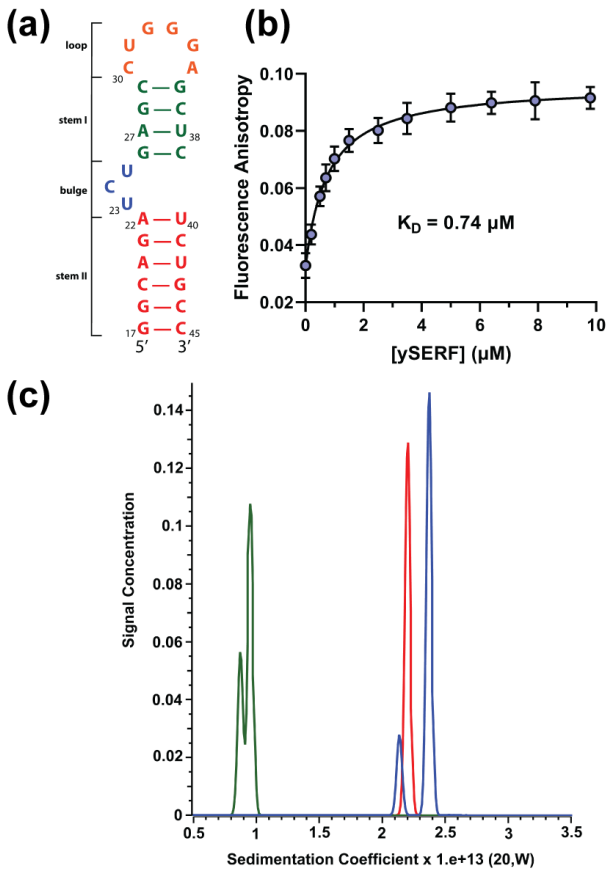


Figure 3-5: SERF binds TAR RNA

(a) *Secondary structure of TAR RNA depicting the two stems, apical loop and trinucleotide bulge.* ***(b)*** *Binding curve showing the change in fluorescence anisotropy of FAM-labeled TAR as a function of SERF concentration. Black line shows the fit to a one-site binding model.* ***(c)*** *Sedimentation coefficient distributions from analytical ultracentrifugation experiments on samples containing SERF alone, monitored at 230 nm (green), TAR alone, monitored at 260 nm (red), and 2.5:1 ratio of SERF and TAR monitored at 260 nm (blue).*

3.2.3 Mapping the interaction sites on SERF and TAR

The CD spectrum of TAR exhibits a positive band at 260 nm (CD_{260}), characteristic of the trinucleotide bulge region. CD_{260} reports on base stacking and may be sensitive to the binding of ligands. The binding of SERF leads to a saturable decrease in CD_{260} signal, likely due to the unstacking of bulge nucleotides (**Fig. 3-6a**). A similar decrease in CD intensity is also observed when the HIV Tat protein binds to the bulge region[153]. We titrated SERF into TAR and monitored imino proton resonances in

the ^{1}H spectrum, and pyrimidines H5/H6 cross-peaks in the ^{1}H-^{1}H TOCSY spectrum of TAR. The positions and intensities of several ^{1}H resonances in the imino region (10 - 15 ppm) of TAR changed on addition of SERF (**Fig. 3-6b**). Chemical shift changes at a particular site can result from changes in the local chemical environment due to ligand interactions or conformational changes due to binding at another site. To probe the conformational transitions in TAR that are coupled to SERF binding, we employed 2D NMR experiments that correlate nucleobase spin systems. The chemical shifts of pyrimidine H5-H6 cross-resonances in the ^{1}H-^{1}H TOCSY spectrum of TAR were analyzed and compared in the presence and absence of SERF. We observed notable chemical shift changes and line broadening in the bulge region (U23, C24, U25) as well as in several nucleobases adjacent to the bulge, such as C37, C41, and C39, in the stems (**Fig. 3-6c**). However, minor or no chemical shift changes were observed for nucleotides farther away from the bulge, like C19, C45, C44, and U31. Interactions with the bulge region of TAR is expected to arrest interhelical motions in TAR by inducing changes in the inter-helical angle between the two stems[154,155]. We are currently probing these conformational differences between the free- and bound- states of TAR using NMR residual dipolar coupling measurements that are commonly used to study long-range constraints in TAR[154].

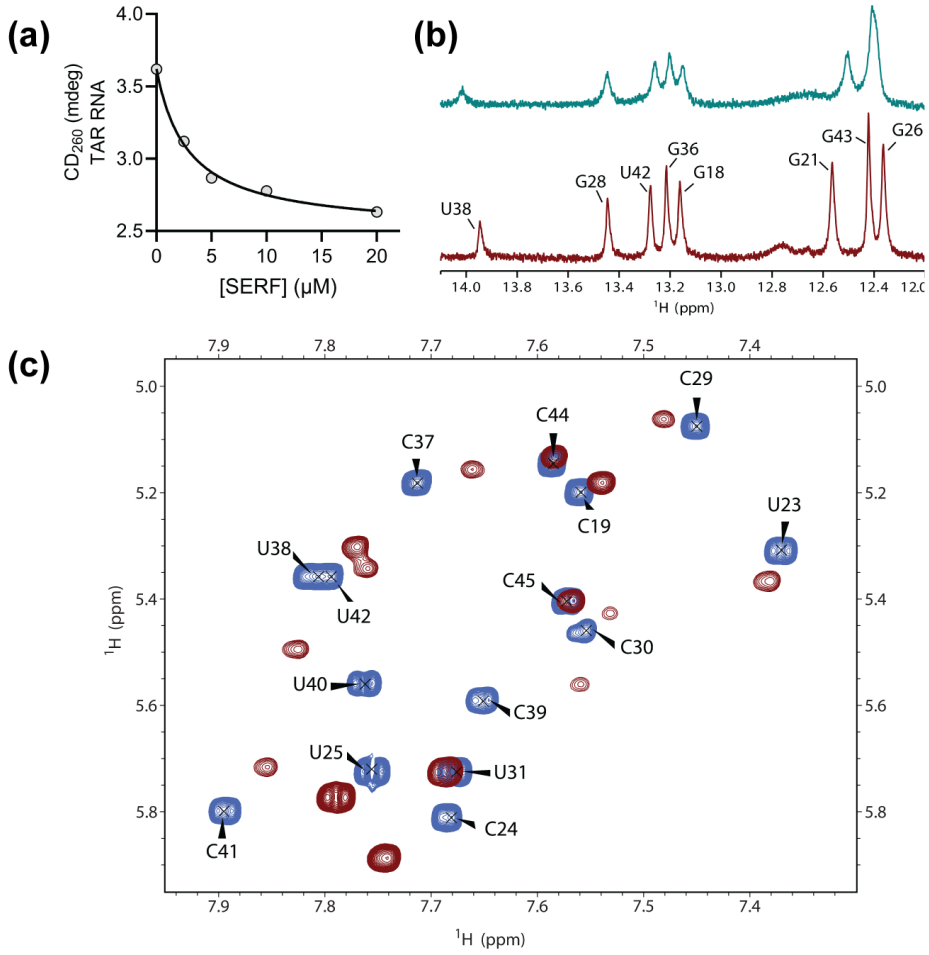


Figure 3-6: SERF perturbs the structure of TAR

(a) *Ellipticity at 260 nm of TAR from circular dichroism spectra. NMR samples for titrations were prepared with 250 μM TAR* ***(b)*** *^{1}H spectra of TAR in the absence (red) and presence of 200 μM SERF (green).* ***(c)*** *^{1}H, ^{1}H- TOCSY spectra of TAR in the absence (blue) and presence of 200 μM SERF (red).*

We observed significant signal loss due to resonance broadening in the ^{15}N,^{13}C-CON spectrum of SERF upon adding TAR. Therefore, we used ^{1}H-detection-based NMR experiments to monitor TAR-induced chemical shift differences in SERF. We could unambiguously assign amide crosspeaks for 62 out of the 67 non-proline residues of SERF using the 2D ^{15}N,^{13}C-CON fingerprint spectrum and 3D HNCO/HN(CA)CO spectra collected at 25 °C The interaction sites for RNA are not broadly distributed in the sequence of SERF. Significant chemical perturbations were observed for residues located in the conserved N-SERF domain. In contrast, residues in the C-terminal helix

show little chemical shift change on adding TAR (**Fig. 3-7b**). Although most IDPs adopt folded structures to varying degrees upon binding to their targets, some retain conformation disorder in the bound state[156,157]. TAR did not induce stable secondary structure in the disordered N-SERF domain. The negative secondary chemical shifts of some of the residue in this region indicates some preference for extended conformations in the bound-state, perhaps to maximize intermolecular contacts (**Fig. 3-7c**).^{15}N-T_2 relaxation measurements show reduced backbone motions on binding TAR (**Fig. 3-7d, e**). The decrease in T_2 rates observed in the TAR-bound state may be attributed to slower microsecond-millisecond conformational exchange processes. The small values of ^{15}N-^{1}H NOEs in the N-SERF domain of free SERF indicates a high degree of backbone fluctuations in fast picosecond-nanosecond timescales. By contrast, the region spanning residue 40 – 60 is more ordered with ^{15}N-^{1}H NOEs close to 0.5, indicating alpha-helical conformation (**Fig. 3-7e**). Although the TAR-bound state is less dynamic than free SERF, the ^{15}N-^{1}H NOEs are below 0.5, suggesting that the N-SERF domain does not adopt a stable helical conformation when bound to TAR. Collectively, our NMR relaxation analysis suggest that the interactions of SERF with amyloid proteins are structurally distinct from those with RNA. SERF forms conformationally heterogenous fuzzy complexes with Aβ40 and α-synuclein[42]. The conformations adopted by SERF in the bound state presumably depends on the nature and lifetime of the interactions that drive complex formation. SERF does not adopt a defined structure in the presence of TAR but loses some of its internal dynamics. The precise modulation of structure and dynamics within the ligand-bound ensemble of a disordered domain could play a crucial role in determining the functional outcome of

RNA binding in a physiological context[158]. For instance, the composition, localization, and material properties of RNP granules can be altered by RNA[159–161]. Next, we directed our focus towards investigating the global hydrodynamic properties of the SERF-TAR complexes in order to assess the long-range organization of these assemblies.

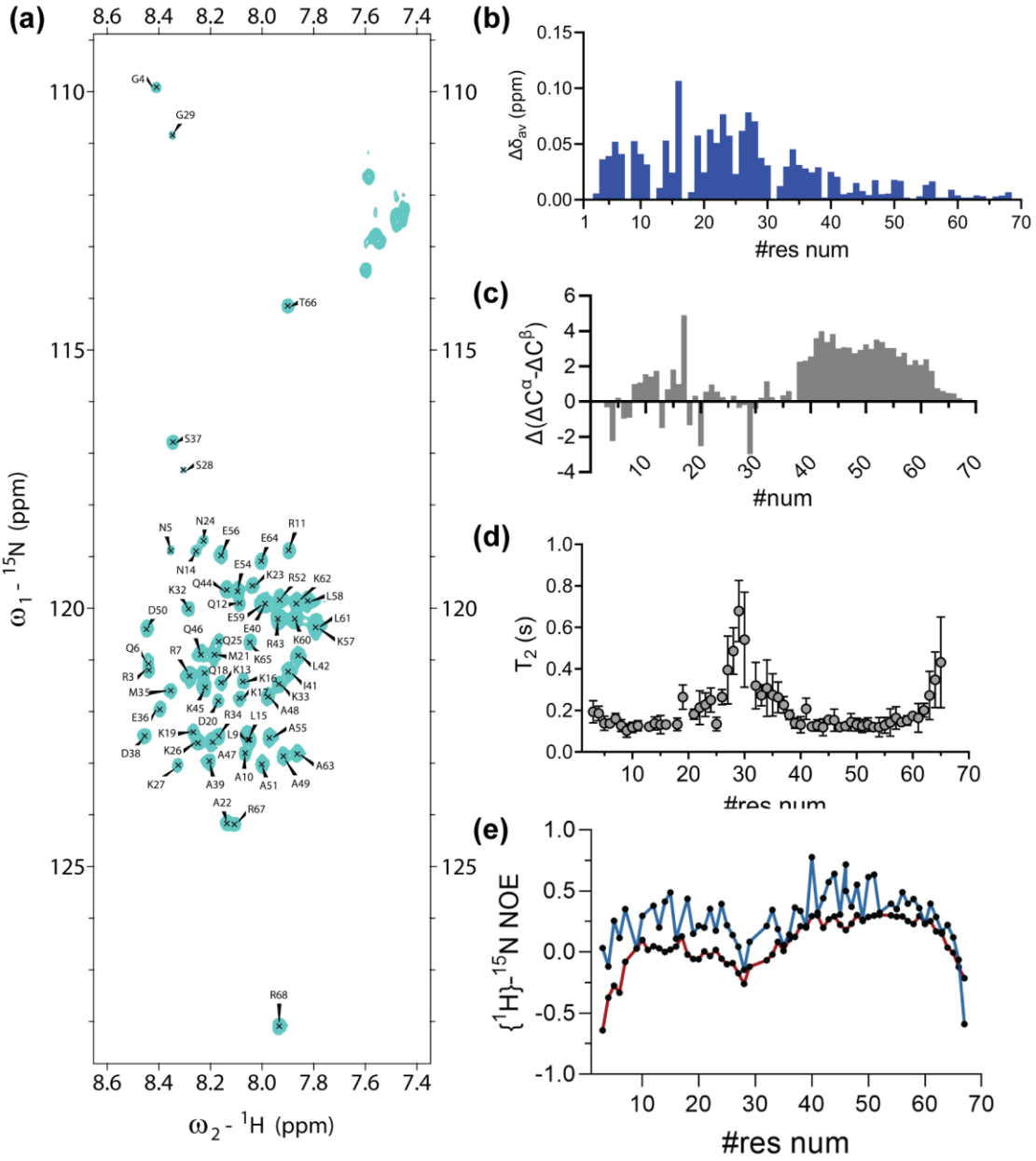


Figure 3-7: Interaction with TAR constrains SERF dynamics

(a) *^{1}H, ^{15}N- HSQC NMR spectrum showing cross-peak assignments for 62 out of the 67 non-proline residues of SERF.* ***(b)*** *chemical shift differences in $^{15}N,^{1}H$-HSQC NMR spectrum of $^{15}N,13C$-labeled SERF in the presence of TAR. Average chemical shifts differences in amide cross-peaks were calculated as $\Delta\delta_{N-H} = [(\Delta\delta_{HN})^2+(\Delta\delta_N/5)^2]^{1/2}$ where $\Delta\delta_{HN}$, and $\Delta\delta_N$ are the chemical shifts (ppm) in $^{1}H^{N}$ and ^{15}N dimensions respectively. Residue-specific* ***(c)*** *secondary chemical shifts ($\Delta(C^{\alpha}_{calc}-C^{\alpha}_{RC}) - (C^{\beta}_{calc}-C^{\beta}_{RC})$).* ***(d)*** *$^{15}N$-$T_2$ relaxation* ***(e)*** *$\{^{1}H\}$-^{15}N heteronuclear NOE of SERF alone (red), and in the presence of TAR (blue). Error bars represent uncertainty in fitting.*

3.2.4 The SERF-TAR complex is globally compact

Our findings indicate that the SERF-TAR complex exhibits distinct dynamic properties compared to the unbound components. Both TAR and SERF sediment as monomers in analytical ultracentrifugation experiments. The frictional ratio (f/f_0) reports on particle asymmetry and is independent of molecular mass[90]. Frictional ratios of 1.1 to 1.35 have been observed for globular compact proteins[162]. IDPs have extended structures with significantly larger values of f/f_0[163,164]. The pseudo-3D plot shows that SERF exists in a rod-shaped (extended β strand-like) dimer and a compact monomer, characterized by frictional ratios of ~1.9 and ~1.45, respectively (**Fig. 3-8a**). A frictional ratio of ~1.4 for TAR indicates that it exists in a globally compact conformation in solution (**Fig. 3-8b**). Strikingly, the 1:1 complex of SERF and TAR is compact and appear to be similar in global dimension to TAR itself (**Fig. 3-8c**). The van Holde-Weischet (vHW) analysis (**Fig. 3-8d**) involves utilizing the sedimentation coefficients obtained from SV-AUC data and plotting the boundary fraction of each species[165]. The sedimentation coefficient distribution of SERF alone and TAR alone shows a peak at ~0.9 and ~2.2, respectively. In a sample containing TAR and 2.5-fold molar excess of SERF, unbound SERF is detected at a sedimentation coefficient of 1.0 (60% of boundary fraction) and TAR-bound SERF between 1.25 and 2.2 (40% of boundary fraction). 100% of TAR is bound and sediments with a uniform boundary of ~2.4. Although the binding sites for amyloid proteins like α-synuclein also reside in the N-

SERF domain, the binding affinity for monomeric α-synuclein is 9 μM, a value that is over 10-fold weaker than the affinity of 0.7 μM for TAR (**Fig. 3-5b**)[42]. The tighter binding to RNA may explain the loss of conformational heterogeneity in SERF when it is bound to RNA. In contrast, when interacting with α-synuclein, SERF forms heterogeneous complexes with broad cross-sections [42]. The residue-specific interactions in IDRs that lead to single-chain compaction in dilute solutions, also significantly contribute to driving LLPS[166,167].

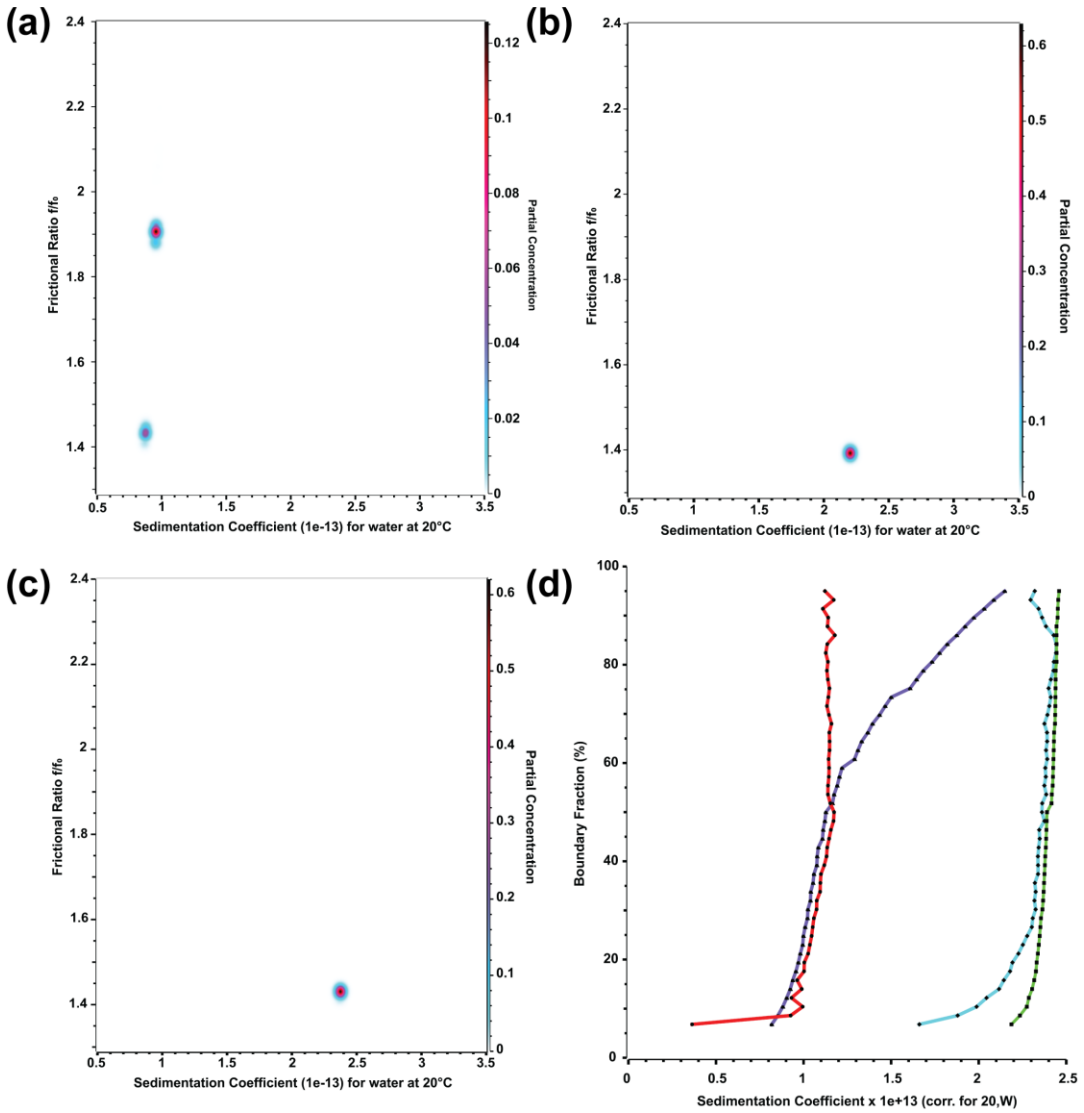


Figure 3-8: Sedimentation behavior of SERF and TAR in free and bound states

*Sedimentation velocity- analytical ultracentrifugation (SV-AUC) experiments showing Pseudo-3D plots of frictional ratios (f/f_0) and sedimentation coefficients of **(a)** SERF monitored at 230 nm, **(b)** TAR monitored at 260 nm, and (c) SERF-TAR complexes monitored at 260 nm in solution containing 2.5:1 ratio of SERF and TAR. Only TAR-containing species will be detected in **(c)**. Data were analyzed by two-dimensional sedimentation analysis (2DSA) followed by analysis with a genetic algorithm, which was further validated by a Monte Carlo analysis. **(d)** vHW analysis of multi-wavelength AUC experiments with SERF and TAR mixed in a 2.5:1 ratio (extinction coefficient of TAR (green), extinction coefficient of SERF (violet)), TAR alone (cyan), and SERF alone (red).*

3.2.5 SERF undergoes associative phase separation with TAR RNA

Beyond saturation concentration, IDPs can spontaneously condense into phase-separated droplets in the presence of oppositely charged multivalent partners. To test if SERF and RNA can form charge-driven condensates, we added TAR to an aqueous solution of SERF with PEG as a molecular crowder. This resulted in the formation of spherical micrometer-sized spherical droplets that fused with the conservation of volume. Confocal fluorescence images showed the co-enrichment of SERF and TAR in droplets doped with fluorescently labeled components (**Fig. 3-9a**). No droplets were observed under the same conditions without RNA. The phase separation of the SERF-TAR system can be best described by complex coacervation, a type of associative phase separation that occurs primarily due to electrostatic attractions between oppositely charged polymers[168]. We observed an increase in the number of spherical droplets when the concentration of SERF was increased from 50 μM to 100 μM, while keeping the TAR concentration constant at 20 μM. The number of droplets did not change appreciably between 100 μM and 500 μM of SERF (**Fig. 3-9c**). The fluorescence of cy5-SERF recovered after photobleaching within a window of 3 min (**Fig. 3-9b**). There was little to no fluorescence recovery of cy3-TAR over ~ 3 min, indicating that TAR in these droplets is immobile. This suggests that SERF is liquid-like in the phase-separated droplets, but RNA displays solid-like properties. Differences in the lifetimes of interactions among heterotypic SERF-TAR pairs and homotypic SERF-

SERF and TAR-TAR pairs can result in differences in material properties inside condensates[169]. G-rich sequences like poly-rG can undergo phase separation into gel-like condensates owing to the formation of G-quadruplex structures[169,170]. RNA can also drive liquid-to-solid transition in droplets containing RBPs like FUS and hnRNPA1 [106]. Hence, we wondered if the hairpin structure of TAR is responsible for the gel-like dynamics. We mixed SERF with 30-nucleotide rA RNA (rA_{30}) and observed spherical coacervates similar to those with TAR. Like TAR, we observed slow FRAP recovery of FAM-labeled rA_{30} doped into the SERF-rA_{30} mixture phase-separation conditions. (**Fig. 3-9d**). A complex coacervation mechanism has been proposed for the formation of nuclear bodies by the disordered domain of Nephrin in mammalian cells[171]. Simple model peptide/ RNA systems like cationic peptide-poly(U) RNA complexes have been demonstrated to form liquid-like condensates by complex coacervation[172]. The observed differences in FRAP recovery between SERF and RNA within the complex coacervates formed under our experimental conditions indicate that binary RBPs-RNA condensates can exhibit heterogeneous internal dynamics, which is not replicated in coacervates formed by cationic peptides and homopolymeric RNAs.

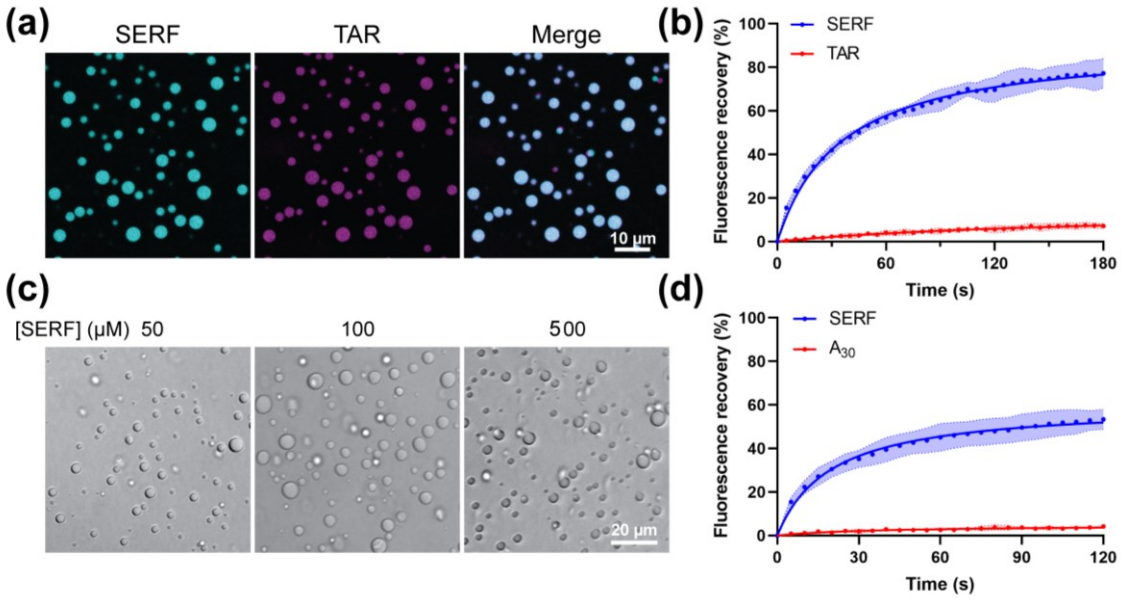


Figure 3-9: Co-phase separation of SERF and RNA

(a) *Confocal images showing colocalization of SERF (cyan) and TAR (pink).* ***(b)*** *FRAP recovery of SERF and TAR fluorescence.* ***(c)*** *Brightfield images of SERF-TAR droplets with different concentrations of SERF at a fixed concentration of 20 µM TAR.* ***(d)*** *FRAP recovery of SERF and rA30 RNA fluorescence.*

3.3 Discussion

3.3.1 SERF and TAR form a tractable RNP complex

Most *in vitro* studies on RNP granules have focused on the phase separation of RBPs like TIA1, hnRNPA1, and FUS under conditions that promote the formation of protein-only condensates[173]. However, the behavior of purified RBPs in a buffer may substantially differ from that in cellular RNP granules, where they engage in multivalent interactions with other proteins and RNAs. It is interesting to note that the specific and non-specific RNA-RBP interactions that drive RNP granule formation, remain largely uncharacterized[174]. Therefore, we aimed to identify a model RNA-RBP pair that can allow us to investigate the role of these interactions in condensate assembly with residue-specific biophysical approaches. The SERF-TAR complex exhibits key characteristics observed in RNP complexes that drive liquid-liquid phase separation through complex coacervation. SERF and TAR form weak, non-covalent interactions that are driven by electrostatic interactions[175]. The conformational dynamics of SERF changes dramatically on binding RNA. Although SERF does not undergo a binding-induced folding transition, possibly due to its charge-rich sequence, hydrodynamic measurements indicate that the SERF-TAR complex exhibits global compaction. The entropic cost associated with condensation into RNP granules can be offset by the compaction that takes place in the dilution and subsaturated regime[176]. Additionally, locally extended conformations adopted by the RNA-binding region of SERF allows more enthalpically-favorable interprotein contacts that promote condensation[177]. Collectively, The comprehensive analysis of the molecular properties of the SERF-TAR complex offers valuable insights into understanding the structure and dynamics of

phase-separating RNP complexes. This study provides a potential roadmap for further investigations in this field.

3.3.2 Evolutionary conservation of the N-SERF domain

Alignment-based methods, commonly used to track the evolutionary divergence of proteins with uniquely folded structures, are not as informative for IDPRs. The lack of selective constraints to preserve a well-defined 3D structure explains the high evolutionary rates of IDPRs at the sequence level[178]. Several studies have detected widespread evolutionary pressure to preserve molecular features that impart “disorder” to proteins. Here, I will discuss potential evolutionary scenarios that could explain the high degree of sequence conservation in the disordered N-SERF domain (**Fig. 3-10**). One possible explanation is that the N-SERF domain is under selective pressure to maintain interactions modules, such as short linear motifs (SLiMs). SLiMs are evolutionary conserved mono-partite sequences that are enriched in disordered regions. SLiMs often mediate multivalent protein–protein interactions or serve as sites of posttranslational modifications such as phosphorylation[179]. While we cannot rule out the possibility that the N-SERF domain contains SLiMs, we did not find any such motif patterns in the annotated Eukaryotic Linear Motif (ELM) database[180]. Sequence conservation in IDPRs cannot be a consequence of the dangers of native state destabilization. Nonetheless, evolutionary conservation can arise from structural constraints to preserve a limited number of specific conformations. The transcriptional activators of the ETV/PEA3 family exemplify this conformational sensitivity where mutations in the four-residue disordered N-terminus lead to conformational rearrangement in the complex of the transcriptional activation domain with the Med25

coactivator[181]. “Constrained disorder” is a term that has been used to describe residues where both the amino acid sequence and the structural property of disorder are conserved[182]. This form of conservation has been documented for the C-terminal domain of the 60S ribosomal protein Rpl5 and the VHL tumor suppressor protein[182]. The similar *in vitro* binding affinities of the N-SERF domain for various amyloid proteins and nucleic acids raises another alternative possibility: evolutionary pressure to prevent nonspecific interactions that may be detrimental to the organism[183]. Interestingly, SERF proteins are not enriched in residues commonly found in low-complexity domains of RBPs such as F, Y, S, G, Q, and N. Perhaps the unusual sequence features and evolutionary history of SERF can be best explained by the presence of “fractional order” in its conformational ensemble, i.e., the presence of partially- ordered conformations that are poised to interact with a binding partner.

```
Homo sapiens (SERF2) MTRGNQRELA RQKNMKKQSD SVKGKRRDD- ----GLSAAA RKQRDSEIMQ QKQKKANEKK EEPK
Pongo abelii         MTRGNQRELA RQKNMKKQSD SVKGKRRDD- ----GLSAAA RKQRDSEIMQ QKQKKANEKK EEPK
Mus musculus         MTRGNQRELA RQKNMKKQSD SVKGKRRDD- ----GLSAAA RKQRDSEIMQ QKQKKANEKK EEPK
Xenopus laevis       MTRGNQRELA RQKNLKKNQD KSK-KEKDD- ----GLSAAA RKQRDAQIMQ EKQKKATEKK DDPK
Homo sapiens (SERF1) MARGNQRELA RQKNMKKTQE ISKGKRKED- ----SLTASQ RKQRDSEIMQ EKQKAANEKK SMQTREK
Bos taurus           MARGNQRELA RQKNMKKSQE ISKGKRKED- ----SLTTSQ RKQRDSEIMQ QKQKAANERK SMQTREK
Danio rerio          MTRGNQRELA RQKNAKKQTD SSKGKRNED- ----GLSAAA RKQRDAEIMQ QKQKKKKK
Sus scrofa           MARGNQRELA RQKNMKKSQE ITKGKRKED- ----SLTTSQ RKQSSRSQKL KIEVQPGTYF L
Schistosoma mansoni  MTRGNQRDLA REKNLKKQSE QRKSKASSQK DGNKGLTLEE RRLRDAEALR AKQQAKSQAS VPKA
S. pombe             MSRGNQRDVD RARNLKKSQ- -ASKKKQAG- ------DPTK RLEAQAEIMR AKQRAADERK AAEANGGSKG KK
S. cerevisiae        MARGNQRDLA RQKNLKKQKD MAKNQKKSG- ------DPKK RMESDAEILR QKQAAADARR EAEKLEKLKA EKTRR
Canis familiaris     MTRGNQRELA RQKNMKKQSD SVKGKRRDD- ----GLSAAA RKQSF
Aedes aegypti        MTRGNQRELA RERNMKKQQA AQKKNKED-- ----GLTHEQ RKQRFFPP
```

Figure 3-10: Multiple sequence alignment of SERF proteins

The software MultAlin was used to align the SERF sequences[184]. Red denotes > 90 % identity and blue denotes > 50 % identity.

3.4 Materials and Methods

3.4.1 Oligonucleotides

The HIV-1 TAR RNA fragment used for all our experiments has the sequence 5′-GGCAGAUCUGAGCCUGGGAGCUCUCUGCC-3′. Unlabeled RNA from Horizon Discovery and 6-carboxyfluorescein (6-FAM) 3′-labeled TAR RNA from Integrated DNA Technologies were purchased as HPLC-purified lyophilized solids and were used without further purification. All solutions were prepared in nuclease-free water (Ambion) and larger volumes of buffer were prepared in DEPC-treated double-distilled water to remove RNases. Typically, labeled RNA was dissolved in nuclease-free water to a stock concentration of 100 μM. The RNA was annealed by heating at 95 °C for 5 min followed by rapid cooling on ice.

3.4.2 Protein expression and purification

The SERF protein was expressed and purified as previously described[42]. Point mutations were inserted into the codon-optimized *S. cerevisiae* SERF gene (amino acid: MARGNQRDLARQKNLKKQKDMAKNQKKSGDPKKRMESDAEILRQKQAAADARREAE KLEKLKAEKTRR) by site-directed mutagenesis with a QuikChange kit (Agilent) and were verified by DNA sequencing. The wildtype and mutant SERF genes were cloned into a pET28 vector that encodes an N-terminal His_6-SUMO fusion tag for purification and scarless cleavage by ULP1 protease. The plasmid was transformed into *E. coli* BL21 (DE3) cells for protein expression. Cells were grown to early logarithmic phase at 37 °C in Protein Expression Medium (Gibco) or M9 minimal medium supplemented with $^{15}NH_4Cl$ (for expressing uniformly ^{15}N labeled protein), both containing 100 μg/mL

kanamycin. Prior to induction, cells were shifted to 20 °C and then expression was induced by adding 0.1 mM isopropyl b-D-1-thiogalactopyranoside (IPTG). After 16 hours of protein expression, cells were pelleted by centrifugation and resuspended in lysis buffer containing 40 mM Tris, 10 mM sodium phosphate (pH 8.0), 10% glycerol, three tablets of protease inhibitor cocktail (cOmplete mini EDTA-free, Roche), 0.375 mM $MgCl_2$, 0.05 µg/ml DNase I and RNase A. The cells were harvested by centrifugation and lysed by sonication on ice for 8 min and the lysate was centrifuged twice at 37,500 × g for 30 min at 4 °C. The supernatant was loaded on a 5 mL HisTrap pre-packed column (Cytiva) that had been equilibrated with water and lysis buffer. The column was washed with lysis buffer and the His_6-SUMO tagged protein was eluted by adding lysis buffer containing 500 mM imidazole. To remove any undegraded nucleic acids, 2ul Pierce™ Universal Nuclease (Cat# 88701) and 20 µl $MgCl_2$ (1 M) were added to the eluate. ULP1 and 10 µL of β-mercaptoethanol were also added to the eluted material before dialyzing against a buffer containing 40 mM Tris (pH 8.0), 300 mM NaCl overnight at 4 °C. The cleaved affinity tag was separated from the SERF by passing the protein solution over a HisTrap column that had been equilibrated with water and lysis buffer. The flow-through from the column containing the tag-free protein and diluted by adding anion exchange buffer (50 mM sodium phosphate (pH 6.0), 100 mM NaCl). Following centrifugation, the protein was passed over HiTrap SP anion exchange column (Cytiva) that had been equilibrated with the anion exchange buffer. The protein was eluted with a linear gradient using buffer containing 1 M NaCl. The eluted fractions were concentrated and loaded on a HiLoad Superdex S75 gel filtration column (Cytiva) equilibrated with 40 mM HEPES (pH 7.5), 100 mM NaCl. Purified SERF was flash-

frozen in liquid nitrogen, and either aliquoted and stored at -80 °C, or subjected to overnight dialysis in 50 mM ammonium bicarbonate pH 8.0 at 4 °C followed by freeze-drying and storage at -80 °C. Protein concentration was determined with the Qubit protein assay (ThermoFisher Scientific) following manufacturer's protocol. Uniformly ^{15}N and ^{13}C isotope- enriched proteins were expressed following the protocol of Marley et al. in M9 minimal medium supplemented with $^{15}NH_4Cl$, U-^{13}C-glucose, and ISOGRO growth supplement (Sigma)[86].

3.4.3 Fluorescence anisotropy measurements

The apparent dissociation constant for the SERF-TAR complex was determined by titrating SERF (0.5 mM stock) into a solution containing 200 nM 3'6-FAM-labeled TAR RNA (100 µM stock) in 1 mL RNA buffer (15 mM sodium phosphate (pH 6.4), 50 mM NaCl, 0.1 mM EDTA) that was heated to 95 °C for 5 min and then cooled rapidly in an ice bath for 10 min or until use. Fluorescence signal was recorded at 25 °C with a Cary Eclipse Spectrofluorometer (Agilent) using excitation and emission wavelengths of 493 nm and 520 nm (5 nm bandpass for both) respectively, and anisotropy values were calculated using the following equations:

$$G = \frac{I_{hv}}{I_{hh}} ; r = \frac{I_{vv} - G.I_{vh}}{I_{vv} + 2G.I_{vh}}$$

where *G* is the instrument correction factor, *I* is the measured fluorescence intensity with polarizers oriented in directions that are indicated in subscripts (*v* is vertical and *h* is horizontal), and *r* is the anisotropy. The titration data points were fitted to a single-site binding equation:

$$r = r_0 + \frac{B_{max} * [SERF]}{K_{d,app} + [SERF]}$$

Where r is the anisotropy signal, B_{max} is the total signal change, r_0 is the initial fluorescence signal of 3'6-FAM-labeled TAR without SERF addition, and $K_{d,app}$ is the apparent dissociation constant.

3.4.4 Spin-Labeling Cysteine Mutants

MTSSL (1-Oxyl-2,2,5,5-tetramethylpyrroline-3-methyl) methanethiosulfonate) stock was prepared in acetone at a concentration of 100 mg/mL. ^{15}N,^{13}C- enriched SERF A10C sample was diluted to 1 mL in Tris buffer, and freshly prepared DTT was added to a final concentration of 10 mM. The SERF samples were incubated for at least 1 h at 4 °C in the dark to reduce cysteine. The samples were then buffer-exchanged into 50 mM potassium phosphate (pH 6.5), 50 mM KCl, 1 mM $MgCl_2$, 0.01% NaN_3 using a PD-10 buffer exchange resin (Cytiva) to remove the excess DTT. MTSSL was added to a final concentration of 1 mg/mL, and the sample was incubated for 1 hr at room temperature in the dark. The unconjugated MTSSL was removed from MTSSL-labeled SERF samples using a PD-10 resin equilibrated with 50 mM potassium phosphate (pH 6.5), 50 mM KCl, 1 mM $MgCl_2$, 0.01% NaN_3. Labeling was confirmed by mass spectrometry.

3.4.5 Solution-state NMR spectroscopy

Chemical Shift Assignment and Secondary Structure Analysis

^{15}N,^{13}C- labeled SERF was buffer-exchanged into 50 mM potassium phosphate (pH 6.5), 50 mM KCl, 1 mM $MgCl_2$, 0.01% NaN_3 using a Sephadex G-10 desalting column and concentrated to 0.8 – 1 mM. Samples for NMR experiments were made to a volume of 500 µL with 5 % D_2O for the deuterium spin- lock. All NMR spectra were collected at 27 °C on a Bruker Avance NEO 600 MHz spectrometer with a 5 mm TCI triple-

resonance cryoprobe. On-instrument processing of all spectra was performed using the Bruker Topspin software. Peak picking and manual assignments were carried out in NMRFAM-Sparky [93]. Backbone and side-chain chemical shift assignments were acquired using the ^{13}C direct-detect experiments developed by Scott Showalter's group[141]. Nearest-neighbor correlations to adjacent residues were established using 3D (HACA)N(CA)CON and (HACA)N(CA)NCO experiments and 2D amino acid-filtered experiments collected as CACON/CANCO spectra for Asp, Ala, Glu, and CACON spectrum for Leu/Ala resonances. Side-chain chemical shifts were assigned by collecting 3D CCCON spectra for aliphatic carbon and 3D H(CC)CON aliphatic proton resonances. The assignments were mapped onto the ^{1}H,^{15}N-HSQC spectrum using standard 3D HNCACB, CBCA(CO)NH, HNCO, and HN(CA)CO spectra. Secondary structure propensity was determined by calculating residue-specific secondary chemical shifts, i.e., the difference in the measured $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts from the sequence-specific random coil chemical shifts calculated by the *ncIDP* predictor for SERF[185].

Spin Relaxation Measurements

^{15}N T_1 and T_2 spin relaxation data were collected as pseudo-3D spectra using $^{1}H_N$- start CON experiments developed previously for ^{13}C direct-detect, ^{15}N spin relaxation measurements[186]. For T_1 measurements, 13 fully interleaved spectra were collected with relaxation delays of 0.02, 0.05, 0.08, 0.1, 0.15, 0.18, 0.2, 0.3, 0.4, 0.6, 0.75, 0.8, 1, and 2 s. For T_2 measurements, 16 fully interleaved spectra were collected with relaxation delays of 15.68, 31.36, 62.72, 78.4, 94.08, 109.76, 125.44, 156.8*, 172.48, 188.16, 219.52, 235.2, 250.88, 282.24, and 313.6 ms (the asterisks indicate duplicate

measurements). The decay curves were fitted to a single-exponential function and reported error bars represent uncertainty in fitting.

Paramagnetic Relaxation Enhancement (PRE) Measurements

$^{13}C,^{15}N$-CON spectrum was acquired on the freshly prepared paramagnetic sample containing spin-labeled SERF A10C in 50 mM potassium phosphate (pH 6.5), 50 mM KCl, 1 mM $MgCl_2$, 0.01% NaN_3 supplemented with 5 % D_2O. Following data acquisition, the MTSSL radical was quenched by mixing 1 µL of 0.5 M sodium ascorbate directly into the NMR tube containing 500 µL of the sample. Following incubation at room temperature for 10 min, identical NMR spectra were acquired for the diamagnetic sample. PREs were reported as the ratio of cross-peak intensities under paramagnetic (I_{para}) and diamagnetic conditions (I_{dia}).

RNA titration experiments

The NMR samples contained 100 µM $^{13}C,^{15}N$ isotope labeled SERF and 100 µM TAR in 50 mM potassium phosphate (pH 6.5), 50 mM KCl, 1 mM $MgCl_2$, 0.01% NaN_3 supplemented with 8 % D_2O. Resonance assignments of SERF in the presence of TAR was generated based on a ^{1}H, ^{15}N HSQC experiment. The 3D experiments that were collected for assignments include HNCO, HN(CA)CO, CBCA(CO)NH, and HNCACB. Spectral comparison was carried out using the assigned ^{1}H, ^{15}N HSQC spectrum of SERF alone as a reference. The T_2 measurement was carried out with the same relaxation delays as for SERF alone. The $\{^{1}H\}-^{15}N$ NOE was measured with proton saturation for 2.5 s and determined from the ratio of peak intensity measured in the saturated and unsaturated spectra.

1D and 2D NMR for TAR

NMR samples contained (unlabeled) TAR RNA dissolved in buffer containing 15 mM NaP (pH 6.4), 50 mM NaCl, 0.1 mM EDTA supplemented with 10 % D_2O. NMR spectra were collected at 25 °C on a Bruker 800 MHz spectrometer equipped with a 5 mm TCI triple-resonance cryoprobe. The spectra were initially processed and analyzed in the Topspin (Bruker) and final images were prepared in MNova (Mestrelab) and NMRFAM-Sparky[93]. 1D ^1H spectra were acquired using a jump-return echo sequence. The imino proton resonances of the guanines and uracils are found between 10 - 15 ppm. 2D ^1H-^1H TOCSY spectrum was collected with a mixing time of 80 ms.

3.4.6 Circular dichroism (CD) spectroscopy

CD spectra of TAR with different concentrations of SERF in 50 mM potassium phosphate (pH 6.5), 50 mM KCl, 1 mM $MgCl_2$, 0.01% NaN_3 were acquired on a Jasco J-1500 CD spectrometer purged with N_2. Measurements were taken in a 0.1 mm quartz cuvette from 200 nm to 325 nm wavelength at 25 °C. Five separate scans were collected for averaging in each experiment. The average CD_{260} signal was plotted as a function of the SERF concentration and fitted to a one-site binding model.

3.4.7 Analytical ultracentrifugation (AUC)

All sedimentation velocity analytical ultracentrifugation (SV-AUC) was performed on a Beckman-Coulter Optima AUC equipped with absorbance and interference optics. Data were collected solely in absorbance mode, at wavelengths of either 230 nm (for SERF peptide backbone absorbance), 260 nm (for TAR RNA absorbance), or every other wavelength from 220 – 290 nm (36 wavelengths total) for the multiwavelength (MW-AUC) experiment. SERF, TAR and SERF:TAR complexes were prepared in 50 mM

potassium phosphate (pH 6.5), 50 mM KCl, 1 mM $MgCl_2$ buffer without NaN_3, which has a strong absorbance at and below 230 nm. The cells used for these experiments contained 12 mm titanium centerpieces and sapphire windows (Nanolytics) and were disassembled and washed with RNase Away™ to avoid RNA cleavage before re-assembly. Each sector was filled with ~400 µL of sample and placed in an An60 titanium rotor. Using Ultrascan III software, all four rotor positions could be used for cells with samples (for the MW-AUC run, only one cell was used because of the time constraint of collecting 36 wavelengths/cell)[89]. The rotor was then placed in the chamber, and a vacuum was established, followed by a 2 hr equilibration time at the selected run temperature of 22°C. A rotor speed of 55,000 RPM was used for each run. Acceleration of the rotor was set to 400 RPM/sec, and the time between scans was set to 3 min for the 230/260 nm runs and was 10 min, 48 sec by default for the MW-AUC experiment. Cells were scanned for a total of 17.5 hr.

Single-wavelength data analysis

Data were imported into the LIMS database directly from the Optima instrument and stored locally on the Penn State LIMS server. Data were edited to convert from radial intensity to pseudo-absorbance, using the intensity of the air above the sample in each sector as a reference. Initial meniscus estimates and the number of scans included in the analysis were made manually. First, a 2-Dimensional Spectrum Analysis (2DSA) (0.5 – 10 S with 150 resolution; 1 – 4 f/f_0 with 64 resolution) with time-invariant noise fitting was performed, followed by the same analysis with additional radially invariant noise fitting and meniscus fitting (2DSA-FM). A final time- and radially invariant noise fitting with iterative boundary fitting (2DSA-IT) was run, and then the selected data were

fit using the Genetic Algorithm with Monte-Carlo simulations (GA-MC) implemented in the Ultrascan software. The data were automatically binned for the GA analysis to avoid bias. For the MC analysis, 32 simulations were used, with 16 parallel threads selected for processing. Pseudo-3D plots of the GA-MC data were exported for presentation, and 2D distributions of the $S_{20,w}$, molecular weight, and frictional ratio plots.

Multiwavelength data analysis

Data were again imported into the LIMS database from the Optima and edited to pseudo-absorbance. The first meniscus estimate was made and the scans to exclude were determined and then propagated to all 36 wavelengths for the sector. All 36 wavelengths' data were fit with 2DSA (same dimensions as above). The 260 nm data was selected for the meniscus fitting refinement and then that fit was propagated to all the wavelengths for the subsequent 2DSA-IT fit. Once all the wavelengths were at the iterative fitting stage, five two-fold serial dilutions of SERF alone (200 μM – 12.5 μM) and TAR alone (5 μM – 1 μM) were put into a UV/visible spectrophotometer and the signal from 200 – 300 nm was recorded for each and exported into the Ultrascan "spectrum fitter" program. All five dilutions were globally fit and the extinction coefficient of SERF at 230 nm was estimated (20,100 M^{-1}, cm^{-1}) and the extinction coefficient of TAR RNA was estimated using an online estimator (https://www.biosearchtech.com/oligospec-calculator-6628#:~:text=Extinction%20Coefficient%20Calculation%20%2D%20The%20extinction,0.9%2C%20to%20adjust%20for%20hyperchromicity.; 265,275 M^{-1}, cm^{-1}). The data were then run through the Ultrascan "Optima MWL fit simulation" program to synchronize the time-grid at all wavelengths, using the original speed and time of the

experiment. These ISSF files were then imported and edited like standard AUC data, then put through the “MWL species fit” program, with the species fits from earlier incorporated. The data were saved as SSF-ISSF files and then imported and edited. The two files were then analyzed using enhanced van Holde- Weischet (vHW) and the contributions of each species could then be observed in a vHW plot and compared to SERF and TAR individually.

3.4.8 Microscopy

All phase separation assays were performed in 15 mM NaP (pH 6.4), 50 mM NaCl, 0.1 mM EDTA containing 10 % PEG 8000 (w/v) at room temperature. Samples were loaded into glass bottom 96-well plates (Corning) pretreated overnight with 5% (w/v) Pluronic acid (Thermo Fisher Scientific) and rinsed thoroughly with ddH_2O. All fluorescent imaging and FRAP measurements were performed on a Leica SP8 confocal microscope equipped with LAS X Life Science Microscope Software and 100x objective lens. FAM-labeled RNA was imaged using 495 nm excitation and 505-550 nm emission. Cy5 labeled SERF A63C was imaged using 650 nm excitation and 665-740 nm emission. Photobleaching was performed using Zoom-In mode, with 405 nm and 495 nm (FAM) or 650 nm (Cy5) lasers at 100% power applied to selected condensates for a duration of 200 msec. The time-lapse was set up to capture three pre-bleach images, followed by post-bleach images with 5 sec intervals. Image analysis was performed on LAS X software, and the intensity data of 8 droplets per condition were fitted in GraphPad Prism.

Chapter 4 Conclusion and Future Directions[4]

The broad aim of this thesis has been to study the role of conformational dynamics in proteins. I have focused on proteins whose dynamics play a central role in regulating function. The first part of my thesis has explored the mechanism of action of the small ATP-independent chaperone Spy on a topologically complex client. My research uncovered substrate-specificity in the mode of action of Spy, prompting us to propose a model wherein the relative affinity for partially- unfolded states of the client determines the mode of chaperone action. In the second part of my thesis, I investigated the structure and dynamics of SERF, a small eukaryotic protein with a highly conserved disordered N-terminal domain. Furthermore, I have characterized the intermolecular interactions between SERF and the extensively studied HIV-1 TAR RNA. I have shown how the SERF-TAR complex can be used as a model system to study the *in vitro* phase separation of RNA-protein condensates. I am currently exploring how molecular-scale conformational properties of SERF-RNA complexes influence the assembly of phase-separated condensates. Here, I summarize the major conclusions drawn from my research and discuss potential future directions.

---

[4] Section 4.1 of this chapter was published in and adapted from: he contents of this chapter were published in *Mitra, R., Gadkari, V. V., Meinen, B. A., van Mierlo, C. P., Ruotolo, B. T., & Bardwell, J. C. (2021). Mechanism of the small ATP-independent chaperone Spy is substrate specific. Nat. Commun, 12(1), 1-13*. Springer Nature permits the reproduction of articles by the author for the purpose of an academic thesis.
Section 4.2 of this chapter was published in and adapted from: *Mitra, R., Wu, K., Lee, C., & Bardwell, J. C. ATP-independent chaperones (2022). Annu. Rev. Biophys., 51, 409-429*. Annual Reviews permits the reproduction of articles by the author for the purpose of an academic thesis. Kevin Wu and I contributed equally to this article and are co-first authors.

4.1 Biological implications of substrate-specific chaperone action

Spy's substrate-specific mode of action may enable it to be very effective as an ATP-independent chaperone. Spy has a dual function; it can facilitate protein folding by allowing clients to fold while being chaperone-bound or preventing aggregation by sequestering unfolded substrates or folding intermediates. Foldase chaperones such as Hsp70, Hsp90, and GroEL exhibit much more complex chaperone cycles that involve co-chaperone binding and ATP hydrolysis[4]. Hsp70 chaperones, for instance, depend on J-domain proteins and nucleotide exchange factors for substrate binding and release. These cochaperones modulate Hsp70 function by regulating its ATPase cycle[187]. Unlike these complex and highly evolved foldase chaperones, Spy lacks any known cochaperones or cofactors that can act to modulate substrate interaction. Instead, Spy's interaction with its substrates is, at least in the cases of Im7 and SH3, apparently finely tuned by evolution to allow for loose binding and, thus, folding while bound. In other cases, such as the one studied here with apoflavodoxin, Spy binds tightly enough to inhibit folding while bound, thereby mimicking the action of well-studied holdase chaperones. For topologically complex model substrates like apoflavodoxin, binding tighter to aggregation-prone misfolded states may be the only thermodynamic outcome in the absence of ATPase activity. Our results reveal a substrate-specific mechanism for Spy where this chaperone exists with a foot in both the "foldase" and "holdase" worlds and provides interesting insights into both. Since Spy can bind to native proteins with low affinity and non-native proteins with higher affinity, the question arises as to why Spy does not interfere with protein function in the cell and why Spy does not become clogged by high-affinity interactions with folding intermediates. We have previously

kinetically and thermodynamically characterized the mechanism whereby Spy binds to, folds, and releases its best-characterized substrate Im7. Periplasmic proteins are highly stable under normal conditions, which strongly decreases the abundance of periplasmic folding intermediates which may preclude the need for high levels of Spy under normal conditions[188]. However, Spy is overproduced up to 500-fold in response to treatment by protein unfolding agents that lead to the accumulation of (un)folding intermediates in the periplasm. Spy can make up to 25% of the periplasmic protein content[3]. It is unclear how clients with no obvious binding partners can be released from Spy during stress or early in stress recovery. One possibility is that Spy is simply diluted out by cellular growth following the cessation of stress. A recent study employed comparative proteomics approaches to identify differentially abundant envelope proteins in spy null mutant vs. spy-containing *Escherichia coli* K-12 MG1655 strains under butanol stress. Four outer membrane proteins (OMPs), OmpC, OmpX, OmpT, and LamB, were significantly less abundant in *spy* null mutant cells[189]. Periplasmic chaperones, like SurA, Skp, FkpA, and DegP, perform several redundant functions in OMP assembly pathways[190]. This raises the possibility that Spy is part of a parallel quality control pathway that manages a specific subset of unfolded OMPs during transit across the periplasm. Evidence in support of this hypothesis comes from another study that showed that overexpression of Spy can compensate for defects in the assembly of OmpX, OmpC, OmpA, and LptD in a *Δskp ΔfkpA* double mutant[189,191]. The holdase-like mechanism of Spy might be important in preventing the aggregation of unfolded OMPs under conditions that inhibit the chaperone activity of other periplasmic chaperones.

4.2 ATP-independent chaperones affect client folding pathways

The folding process is usually an equilibrium reaction, but aggregation and proteolysis are generally irreversible and thus need to be avoided by the cell. The action of ATP-independent chaperones can in large part be explained by models that involve kinetic partitioning among the association rate of chaperone binding to an unfolded or misfolded client, the rate of protein aggregation, and the rate of client protein refolding. ATP-independent chaperones will tend to function as 'holdase' chaperones when the client association rate exceeds the client aggregation and refolding rates (**Fig. 4-1a**). Under proteotoxic stress, the cellular concentrations of active chaperones are upregulated at either the transcriptional or posttranslational levels. Interactome and gene expression analysis of aging human brains revealed induction of ATP-independent chaperones[192]. The association rate between a chaperone and its client protein increases at higher chaperone levels, making binding more likely than refolding. Thus, holdase chaperones tend to become kinetic traps that prevent protein aggregation and inhibit protein folding (**Fig. 4-1a**). Tight binding with nanomolar affinity has been reported for many ATP-independent chaperones such as SecB, Skp, and the sHsp family member Hsp27[22,53,82,85,193,194]. However, this is not a universal feature for ATP-independent chaperones. Many of these chaperones, such as trigger factor, Spy, and SurA, have been shown to bind to their client proteins with low affinity, with apparent dissociation constants in the micromolar range[7,17,22,81,195]. These weak associations apparently allow these chaperones to suppress protein aggregation effectively. While these chaperones may not form a stable complex with unfolded proteins, they employ other mechanisms to protect proteins. One mechanism by which weak-affinity

chaperones may facilitate folding is by associating with productive partially folded intermediates. This mechanism has been proposed for the action of trigger factor, which suppresses protein aggregation and promotes correct folding by stabilizing an aggregation-prone intermediate on the folding pathway (**Fig. 4-1b**)[81,195]. Association with partially folded proteins has also been observed for two classical holdase chaperones, the small heat shock proteins (sHsps) Hsp26 and Hsp42[54]. Another novel mechanism for the action of ATP-independent chaperones is to allow the protein to fold while bound (**Fig. 4-1c**)[7]. This mechanism has been intensively studied by me (**Chapter 2**) and others (reviewed in **Chapter 1**) using Spy[7,10]. In summary, the diverse binding affinities of ATP-independent chaperones (listed in **Table 5**) enable them to do much more than prevent protein aggregation.

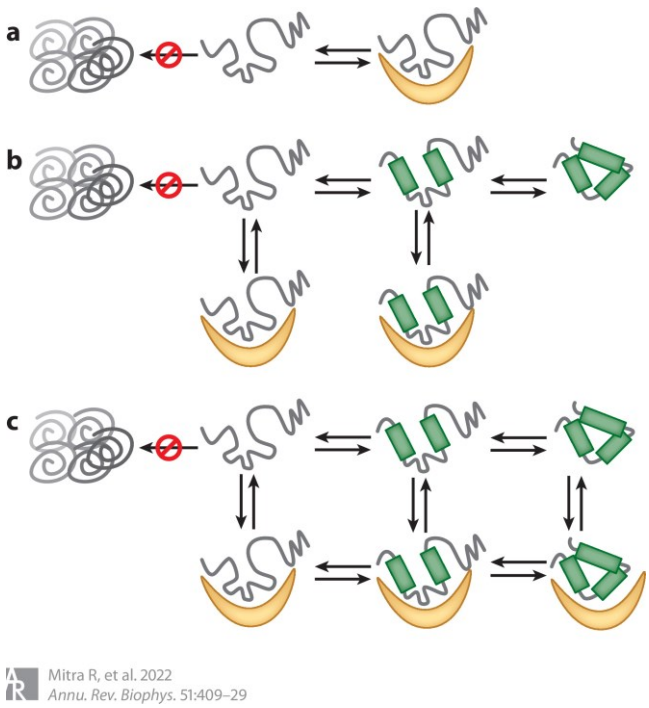


Figure 4-1: Diverse mechanisms of ATP-independent chaperones[11]

Chaperones are shown as orange crescent-shaped structures, and client proteins are shown in green and gray. (a) Holdase chaperones. (b) Chaperones that promote client folding by stabilizing productive intermediates. (c) Chaperones that can allow client folding- while- bound. (Reprinted from Mitra, R. et al. Annu. Rev. Biophys. (2022) with permission from Annual Reviews)

Table 4: Client affinities of ATP- independent chaperones[11]

ATP-independent chaperones	Unfolded Client	Binding affinity
Trigger factor	mature PhoA [195]	2 μM
	mature MBP [53]	0.5 μM
	denatured GAPDH [81]	0.32 μM
	denatured luciferase [81]	4.5 μM
SurA	OmpA [16]	1.8 μM
	OmpT [16]	9.3 μM
	OmpX [17]	0.8 μM
Spy	Im7[7]	4.7 μM
	Fyn SH3[55]	2.9 μM
	Apoflavodoxin *(this study)*	0.35 μM
Skp	OmpLA/ OmpW/ PagP [196]	10 nM
	OmpA/ OmpG [82]	20 nM
SecB	Barnase [197]	2 nM
	pre- mature MBP [193]	0.8 nM
	mature MBP [193]	80 nM
	mature PhoA [53]	0.5 μM
Hsp27	T4 lysozyme [198]	40-200 nM
ttHsp40	mature PhoA[199]	10 μM

4.3 Structural properties of IDPRs in condensed phase

Solution-state NMR spectroscopy is the most common technique to study conformational properties of IDPRs at atomic resolution[200]. NMR is also well-suited for probing the interactions, motions, and dynamics of IDPRs like Ddx4, FUS, TDP-43, NPM1, CAPRIN1, and hnRNPA2 that form phase-separated condensates[201]. The type of information obtained from NMR experiments on LLPS depends on the phases present in the NMR sample. *Murthy and Fawzi (2020)* have categorized NMR samples into three groups: dispersed-phase, biphasic, and condensed-phase samples[201]. NMR studies with dispersed solutions use protein concentrations below the saturation concentration for LLPS. The enhanced relaxation rates of the protein fraction in the condensed phase led to severe signal broadening for the condensed fraction, making it possible to monitor the onset of LLPS by observing loss in signal intensity from the remaining molecules in the dispersed phase. Any changes in the chemical environment of residues involved in homotypic and heterotypic interactions can be monitored by observing chemical shift perturbations as a function of protein concentration, as was done to identify residues that are important for LLPS of the C-terminal region of TDP-43[202]. This approach does not always identify specific LLPS-triggering residues, as seen in the low-complexity domain of hnRNPA2, where minor chemical shift changes were observed across the entire protein sequence[115]. Biphasic samples contain a suspension of phase-separated droplets. The protein molecules in the dilute and dense phases generally have different translational diffusion rates and rotational tumbling times. This enables the selective isolation of signals from either of the two phases by diffusion editing or relaxation editing schemes, as shown in a study on the LLPS of an elastin-like

peptide[203]. In diffusion-editing, pulsed field gradients are used to separate protein populations based on the difference in molecular diffusion in the dilute and condensed phases. Relaxation editing exploits the difference in transverse relaxation rates between protein populations in the two phases. Significantly higher protein concentrations than the saturation concentration for LLPS is required to overcome the low signal intensity of the condensed phase caused by the higher viscosity and enhanced R_2 relaxation rates[201]. The most direct method to probe molecules in the condensed phase is to prepare a macroscopic condensate that completely occupies the observation volume of the NMR coil. This approach has been successfully used to study phase-separated RNA-binding proteins like FUS, CAPRIN1, and Ddx4[117,204,205] . However, probing site-specific interactions in the condensed phase is generally challenging. One successful approach involves preparing mixed-state samples that contain proteins purified under specific isotope labeling schemes. In the case of CAPRIN1, intermolecular NOEs between backbone amide protons of one chain and aliphatic or aromatic protons of adjacent chains in the condensed phase have been recorded by using specific isotope labeling schemes to filter-out NOE signals from intramolecular contacts. For instance, $^{12}C/^{15}N/^{2}H$- labeled and $^{13}C/^{14}N/^{1}H$- labeled CAPRIN1 were mixed to obtain residue-specific contact maps for the condensed phase[206]. Intermolecular NOEs revealed contact sites distributed throughout the sequence of CAPRIN1. Similar intermolecular NOE-based experiments with 1:1 mixture of either $^{15}N/^{12}C$, $^{14}N/^{13}C$ or $^{15}N/^{12}C$, $^{14}N/^{12}C$ samples for the low-complexity domain of FUS showed that the condensed phase of FUS, like CAPRIN1, is stabilized by hydrogen bonding, π/sp^2, and hydrophobic interactions involving residues dispersed throughout the sequence[117].

4.4 Proposed model for the assembly of SERF-TAR condensates

A key question we wish to answer is whether RNA-protein interactions in the dilute phase are maintained in the condensed phase. The choice of a tractable RNA-RBP pair allows us to probe the role of heterotypic and homotypic multivalent interactions in stabilizing protein-RNA condensates. In our model of TAR-dependent assembly, SERF monomers and dimers cannot condense into LLPS droplets due to interchain and intrachain repulsion. Complexation with TAR neutralizes the charges on SERF allowing new inter-chain contacts to be established under conditions that promote phase separation. SERF oligomers can reduce the entropic cost of demixing by giving rise to “emergent stickers” that are capable of driving phase separation[176]. Therefore, TAR acts as a nucleator of phase separation but does not provide additional sources of multivalency in our system. The emergent multivalent network exhibits a beads-on-a-string-like structure, where RNA acts as an immobile scaffold with adhesive points that bind SERF molecules. This enables the formation of SERF-SERF crosslinks, leading to the formation of a macroscopic condensed phase. A few examples demonstrating a similar mechanism of condensate assembly have been reported so far. Photoinduced oligomerization drives phase separation by increasing the valence of IDRs in a synthetic Corelet system[207]. RNA binding can nucleate higher-order fibrous assemblies of the FUS protein[208]. An auxin response transcription factor, ARF19 in *Arabidopsis* can undergo spontaneous polymerization through electrostatic interactions driven by a folded PB1 domain[209]. Molecular simulations are needed to test our model of cluster formation in the dispersed phase by RNA-mediated higher-order assembly of SERF[176].

Appendix

Characterizing an oxygen tunnel in the flavoenzyme oxidase NicA2[5]

This study focuses on nicotine oxidoreductase (NicA2), a therapeutically relevant Flavin-dependent enzyme that catalyzes the degradation of nicotine into a non-psychoactive metabolite, N-methylmyosmine (NMM)[210]. As a Flavin-dependent amine oxidase family member, NicA2 catalyzes reduction and oxidation reactions by first accepting electrons from nicotine. This leads to the conversion of its bound Flavin adenine dinucleotide (FAD) to the reduced hydroquinone form ($FADH_2$), which must be oxidized after this reductive step to complete the catalytic cycle. O_2 is the canonical electron acceptor for Flavin-dependent amine oxidases[211]. This enzyme has recently been shown to reduce nicotine levels in the blood and brain of rats following pretreatment in a dose-dependent manner[212–215]. Rats pretreated with NicA2 also show reduced addiction to nicotine in a self-administration model. These findings have established NicA2 as a promising therapeutic to help in smoking cessation. However, the WT NicA2 enzyme has extremely slow catalytic turnover with a k_{cat} of ~0.007 s^{-1}, making it a poor therapeutic[211,216]. In this work, we used a genetic selection approach to isolate variants with up to a 150-fold improvement in their rate constant for oxidation by O_2 and a 10-fold increase in k_{cat}/K_M. Kinetics studies revealed that the mutations found

[5] The following materials are adapted from results published in *Dulchavsky, M., **Mitra, R.**, Wu, K., …, George, O., Stull, F., Bardwell, J.C. Directed evolution unlocks oxygen reactivity for a nicotine degrading flavoenzyme. (In revision at Nat. Chem. Biol.)* All proteins were expressed and purified by Mark Dulchavsky. I designed the NMR experiments and analyzed the results. HDX-MS experiments were conducted by Antonio Calabrese and the crystal structures were determined by Kevin Wu. Mark Dulchavsky and I performed the NMR experiments and wrote the relevant sections.

in these variants improve catalysis by increasing the rate at which O_2 oxidizes NMM-bound enzyme.

I have characterized the structural features of one such variant of NicA2 *v321*, which contains seven mutations, F104L, A107T, S146I, G317D, H368R, L449V, and N462S, using NMR spectroscopy. The crystal structures of NMM-bound NicA2 v321 and WT are very similar, with an overall RMSD of < 0.5Å. No obvious difference in these structures would explain the striking difference in their activities. This finding, which indicates that there may be catalytically relevant dynamic states of NicA2 that are not captured in "frozen" crystallographic snapshots, prompted us to employ solution NMR spectroscopy to investigate the structural basis for the increased oxidation rates of NicA2 v321. Flavoenzyme oxidases often contain tunnels within the protein structure for the guided diffusion of dioxygen into the protein core that houses the Flavin cofactor[217]. It has been shown that obstruction or modification of these tunnels can decrease oxidase activity, which supports the direct role of channels in bringing O_2 to the active site [39,40,43]. One-way flavoenzymes may modulate the oxidation of $FADH_2$ is by changing the accessibility of these O_2 access tunnels. The observed increase in the oxidation rate of NMM-bound NicA2 in variant v321 raised an intriguing possibility: the collection of mutations in this variant alters the transport pathway of O_2. Upon inspection of the solvent-accessible surface in the crystal structure of NicA2, we noticed a tunnel in the region rich in the activity-enhancing mutations in v321. This tunnel extends from the active site to the solvent-exposed surface of the protein.

To further elucidate the chemical environment of this tunnel, we employed an NMR site-specific probe. Although hitherto under-utilized for studying flavoenzymes,

NMR spectroscopy can provide structural and kinetic information on transient events in enzyme dynamics[218]. Given the large size of the NicA2 homodimer (110 kD) and the subtle conformational changes we wish to probe, we decided to use 1D ^{19}F NMR. This approach has two advantages: fluorine serves as an exquisitely sensitive structural probe for side-chain dynamics, and observation of anything other than a single sharp peak in the ^{19}F NMR spectrum indicates conformational heterogeneity in the local environment of the fluorine nucleus[219,220]. In addition, ^{19}F NMR often works better for large proteins than more traditional two-dimensional ^{1}H, ^{15}N-based NMR[221]. Therefore, we used an amber codon suppression strategy to site-specifically incorporate the unnatural amino acid, 4-trifluoromethyl-L-phenylalanine (tfmF) in the place of Y342, a residue that is located adjacent to the tunnel observed in the WT protein (**Appendix Fig. 1a**)[222].

We observed a single, broad signal in the ^{19}F spectrum of WT NicA2 at -59.8 ppm that can be deconvoluted into two highly overlapping peaks indicative of conformational motion in intermediate-to-slow exchange in the interior of the tunnel (**Appendix Fig. 1b**). On the other hand, the ^{19}F spectrum of NicA2 v321 is strikingly different, composed of a broad up-field shifted major peak at -60.2 ppm and three highly overlapping down-field minor peaks (**Appendix Fig. 1c**). In contrast to the ^{19}F NMR spectrum for the fluorine incorporated at the tunnel- adjacent Y342 site, a control substitution at Y313, a position that lies far from the tunnel region, we observed single narrow peaks in both WT NicA2 and v321 (**Appendix Fig. 2a**). This suggests that the multiple peaks we detect with a single tfmF incorporated at Y342 cannot be explained solely by global structural perturbation. Furthermore, the greater conformational

heterogeneity in the vicinity of Y342 in NicA2 v321 cannot be explained by the lower stability of the mutant variant since the WT and mutant enzymes exhibit similar thermal denaturation curves and activities (**Appendix Fig. 3**). WT and v321 also demonstrate a nearly identical elution profile in size exclusion chromatography, excluding the possibility that the chemical shift differences arise from a difference in size distribution (**Appendix Fig. 4**).

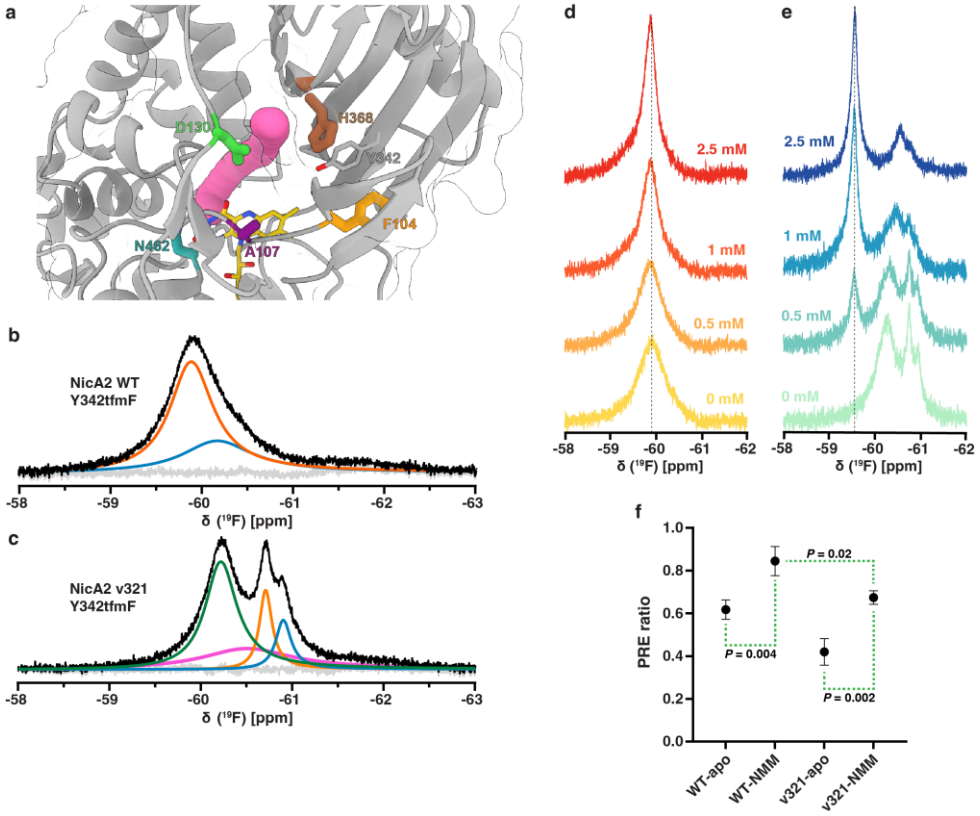
In summary, the ^{19}F NMR spectra of the two ligand-free NicA2s demonstrate higher conformational heterogeneity in the proposed oxygen tunnel in v321 than in the WT protein. To explore the structural basis of the increased rate of oxidation of v321 over the WT enzyme only when bound by NMM, we decided to compare the ^{19}F NMR spectra NMM-bound state of wild type and NicA2 v321. NMM addition does not cause any significant chemical shift perturbation in the WT NicA2, although the peak does narrow, which suggests that ligand binding may restrict conformational fluctuations in the tunnel (**Appendix Fig. 1d**). In contrast, upon the addition of NMM to NicA2 v321, we observed a substantial downfield chemical shift in the major peak of the enzyme, possibly due to conformational rearrangement in the tunnel region due to product binding (**Appendix Fig. 1e**). Furthermore, at increasing molar ratios, NMM yields a narrow major peak whose intensity increases with a concomitant decrease in the intensity of the apo-state peaks. Thus, NMM binding “pulls” the distinct apo-state sub-ensembles in slow exchange into a single major conformation, presumably resembling the crystallized conformation of the NMM-bound NicA2 v321 complex. We confirmed that these conformational changes are relevant to catalysis by adding excess nicotine to the enzyme to achieve the NMM-bound, reduced enzyme state that O_2 rapidly oxidizes

in our kinetic experiments. The resulting spectra after nicotine addition closely overlap with those representing NMM-bound oxidized enzyme, indicating that NMM-bound oxidized and reduced states are conformationally indistinguishable (**Appendix Fig. 2b**). Therefore, observations made using the NMM-bound oxidized enzyme report on the enzyme conformation relevant to the oxidation step observed in stopped-flow experiments.

Paramagnetic relaxation enhancement (PRE) is widely used to obtain long-range distance restraints in NMR spectroscopy[223]. When a soluble paramagnetic reagent like 4-hydroxy-2,2,6,6-tetramethyl-piperidine-1-oxyl (TEMPOL) is added to the solvent, the extent of paramagnetic broadening depends on the solvent accessibility of fluorine. Therefore, we decided to use solvent accessibility as a proxy to ascertain the access of O_2 to the proposed tunnel. In the presence of TEMPOL, we observed line broadening in the major peaks in both the unbound WT and mutant NicA2 (59.8 and 60.2 ppm, respectively), suggesting that these peaks represent accessible conformations (**Appendix Fig. 5**). The major populations of both wildtype and v321 NicA2 appear less accessible to solvent when bound by NMM than when ligand-free (**Appendix Fig. 1f**). However, in both the ligand-free and NMM-bound states, the predominant conformations of NicA2 v321 have a more solvent-accessible tunnel than the respective states of the wildtype protein. This may indicate that mutations encoded in v321 allow greater diffusion of O_2 to flavin in this catalytically competent state. This could, at least partially, explain the increased oxidation rate of NicA2 v321.

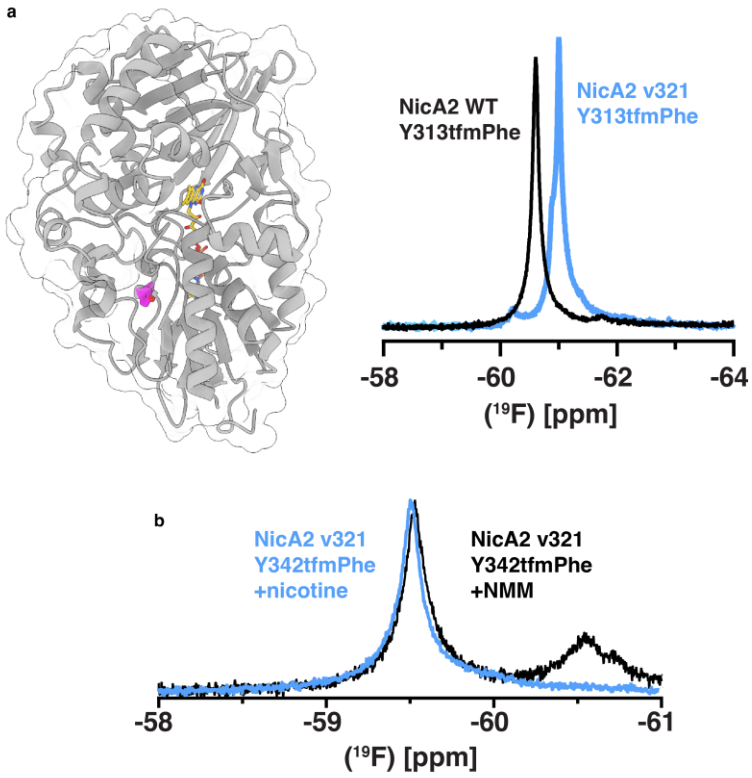
It remains unclear how exactly the mutations in v321 affect the steering and orienting of O_2 to a catalytically productive location near Flavin. Nonetheless, our results

pave the way for future studies using approaches like molecular dynamics simulations with enhanced sampling to probe the connection between allosteric networks and dynamics of O_2 permeation in NicA2 in greater depth.

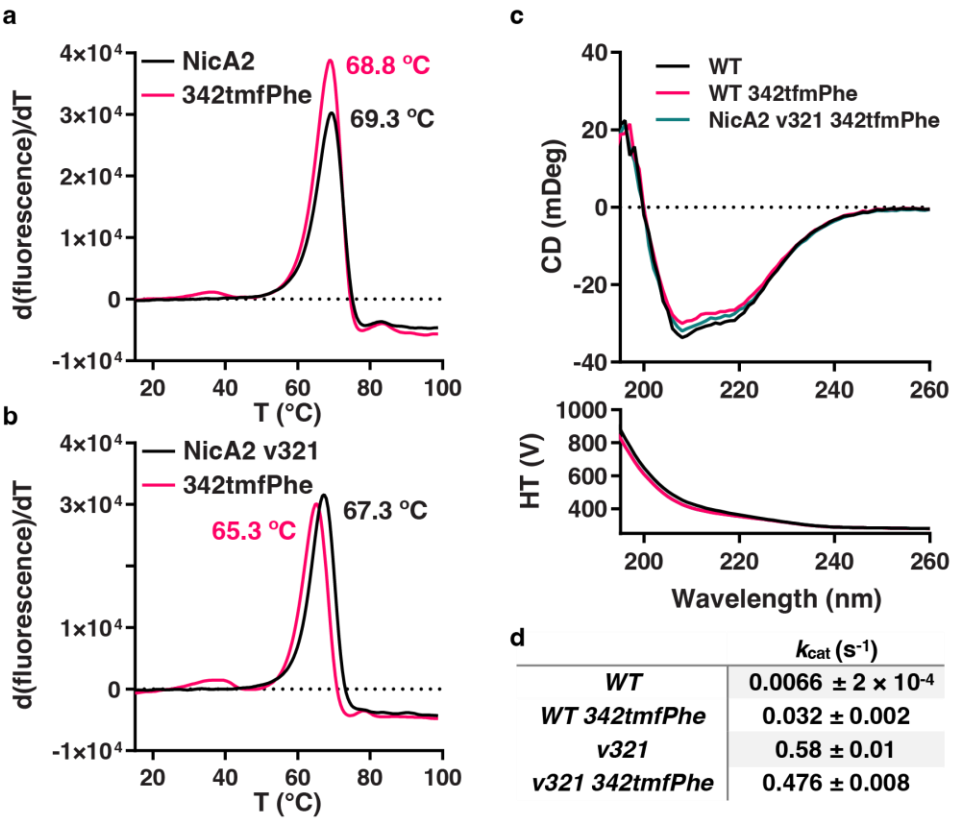


Appendix Figure 1: Conformational landscapes of wildtype and v321 NicA2.

a. *NicA2 wildtype's crystal structure (PDB ID: 5TTJ) is displayed with an ~ 1.4 Å diameter tunnel identified by CAVER simulation rendered in magenta*[224]*. FAD is rendered in yellow, Y342 is rendered in grey and labeled, and frequently mutated residues are rendered F104 in orange, A107 in purple, D130 in green, H368 in brown, and N462 in turquoise.* ***b.*** *^{19}F NMR spectrum of wildtype NicA2 with the Y342tmfF substitution. The black trace represents the raw data, the colored curves represent fits deconvoluted using decon1d, and the grey trace represents residuals from the fit.* ***c.*** *^{19}F NMR spectrum of NicA2 v321 with the Y342tfmF substitution.* ***d.*** *The signal for wildtype NicA2 Y342tfmF titrated with 0, 0.1, 0.5, and 2.5 mM NMM narrows with increasing NMM, indicating a more restricted sampling of local conformations.* ***e.*** *NicA2 v321 Y342tfmF titrated with NMM of the same concentrations as labeled, showing collapse into a single population.* ***f.*** *The PRE ratio for wildtype and v321 in both apo- and NMM-bound states are plotted. Error bars represent the standard deviation of 3 replicates, and data were analyzed by 1-way ANOVA with Tukey's multiple comparisons post hoc. All NMR spectra were acquired at 298 K on a Bruker 600 MHz instrument equipped with Prodigy (^{1}H/^{19}F)-X broadband cryoprobe and operating at a basic transmitter frequency of 563.6701729 MHz for the ^{19}F nucleus. One-dimensional un-decoupled ^{19}F spectra were recorded with 16K data points, 30 kHz sweep width, an acquisition time of 0.288 s, relaxation delay of 2 s, and 1000 - 5000 transients accumulated per experiment. The ^{19}F chemical shifts were referenced to either TFA (set at -75.39 ppm) or tfmF (set at -62.0832 ppm). The spectra were processed in TopSpin 4.1.4 and deconvoluted using the Python-based decon1d fitting program that assumes Lorentzian peak shapes*[225]*.*



Appendix Figure 2: ^{19}F NMR spectra of Y313tfmF and Y342tfmF substituted NicA2 enzymes
a. *Y313 (magenta) was mutated to the amber stop codon, and tfmF incorporated into this position to serve as a probe for ^{19}F NMR. This position was chosen because it is conservative substitution distant from the tunnel region and buried in the core of the flavin-binding domain. It should therefore be an independent readout about global protein dynamics that is not influenced by local tunnel dynamics. Also displayed are the ^{19}F NMR spectra of the Y313tfmF for wildtype and variant NicA2s, which are similar.* ***b.*** *Nicotine was titrated into NicA2 v321 Y342tfmF to achieve the reduced, NMM-bound state. NMM was also titrated into the oxidized enzyme. These major species of these spectra overlay closely, indicating that NMM-bound oxidized and reduced states are conformationally indistinguishable.*

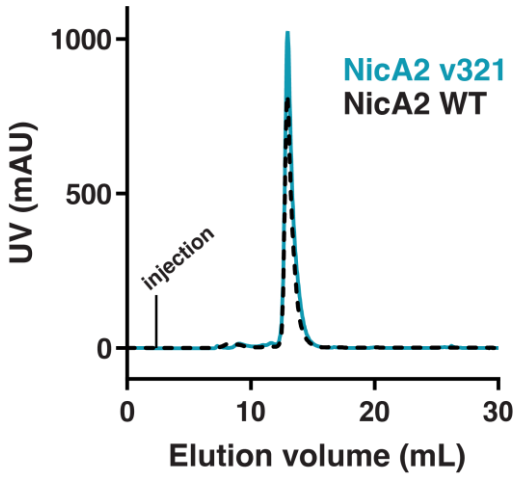


d

	k_{cat} (s^{-1})
WT	0.0066 ± 2 × 10^{-4}
WT 342tmfPhe	0.032 ± 0.002
v321	0.58 ± 0.01
v321 342tmfPhe	0.476 ± 0.008

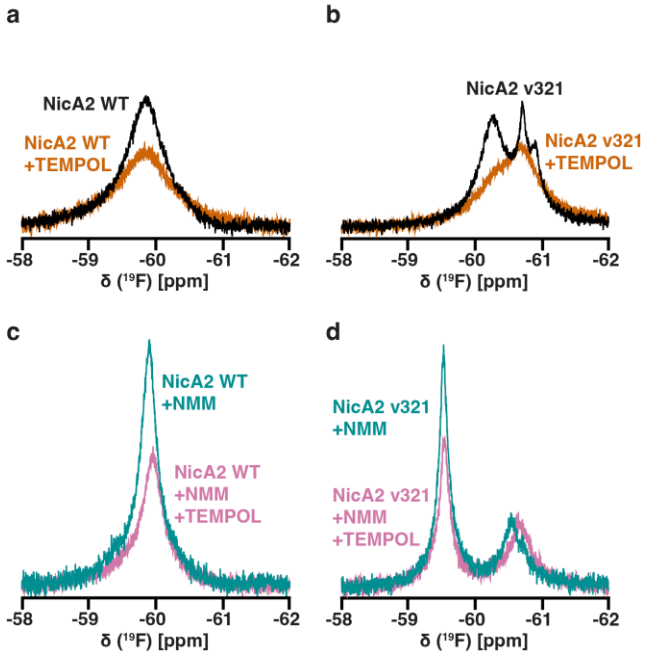
Appendix Figure 3: Qualities of NicA2 19F containing variants.

a. *NicA2 wildtype and NicA2 wildtype with tfmF substituted at position 342 have a similar melting temperature. The intrinsic fluorescence of FAD is quenched upon being bound by protein. As the temperature of the protein is raised and it begins to unfold, the released FAD is once again fluorescent. This unfolding signal is monitored in a qPCR machine. A volume of 20 µL of 10 µM NicA2 enzymes were pipetted to the bottom of a qPCR tray. This was monitored for fluorescence as the temperature was ramped at 0.5 °C per second with a 30 second hold time per step in a QuantStudio 3 Real-Time PCR System. Traces represent an average of 3 replicates.* ***b.*** *NicA2 v321 and NicA2 v321 with tfmF substituted at position 342 have a slightly different, but still similar melting temperature. Traces represent an average of 3 replicates.* ***c.*** *Circular dichroism spectra of wildtype and the two ^{19}F containing constructs maintain secondary structure. Samples of 10 µM NicA2 proteins (flavin concentration) were prepared in 10 mM potassium phosphate pH 7.4, 2.5% (w/v) glycerol for circular dichroism experiments. Solutions were added to 1 mm quartz cuvette. Five room temperature scans were averaged from 260-195 nm using a Jasco J-1500 circular dichroism spectrophotometer.* ***d.*** *Steady-state kinetic data for the reaction of NMR variants with nicotine. Values are reported with 95% confidence intervals.*



Appendix Figure 4: Elution profiles of NicA2 wildtype and v321.

NicA2 enzymes were prepared at NMR concentrations (100 µM). 200 µL of each enzyme were injected onto a Superdex 200 10/300 GL column equilibrated in 40 mM HEPES pH 7.4, 100 mM NaCl, 10% (w/v) glycerol. An Äkta-FPLC system was used to maintain a flowrate of 0.2 mL min^{-1}, and absorbance of the sample monitored at 280 nm. The elution profile of NicA2 wildtype is displayed as a dashed black line, v321 is a solid cyan line



Appendix Figure 5: NicA2 enzymes upon addition of TEMPOL

a. *Scans were taken before and after TEMPOL was added to 8 mM end concentration in samples of NicA2 Y342tfmF wildtype or* ***b.*** *v321.* ***c.*** *The same experiment was repeated with 5 mM NMM included in samples of NicA2 wildtype and* ***d****. v321*

Bibliography

1. Medina, E., R. Latham, D. & Sanabria, H. Unraveling protein's structural dynamics: from configurational dynamics to ensemble switching guides functional mesoscale assemblies. *Curr Opin Struct Biol* **66**, 129–138 (2021).
2. Mirsky, A. E. & Pauling, L. On the Structure of Native, Denatured, and Coagulated Proteins. *Proceedings of the National Academy of Sciences* **22**, 439–447 (1936).
3. Quan, S. *et al.* Genetic selection designed to stabilize proteins uncovers a chaperone called Spy. *Nat Struct Mol Biol* **18**, 262–269 (2011).
4. Hartl, F. U., Bracher, A. & Hayer-Hartl, M. Molecular chaperones in protein folding and proteostasis. *Nature* **475**, 324–332 (2011).
5. Spence, G. R., Capaldi, A. P. & Radford, S. E. Trapping the on-pathway folding intermediate of Im7 at equilibrium. *J Mol Biol* **341**, 215–226 (2004).
6. Capaldi, A. P., Kleanthous, C. & Radford, S. E. Im7 folding mechanism: Misfolding on a path to the native state. *Nat Struct Biol* **9**, 209–216 (2002).
7. Stull, F., Koldewey, P., Humes, J. R., Radford, S. E. & Bardwell, J. C. A. Substrate protein folds while it is bound to the ATP-independent chaperone Spy. *Nat Struct Mol Biol* **23**, 53–58 (2016).
8. Koldewey, P., Stull, F., Horowitz, S., Martin, R. & Bardwell, J. C. A. Forces Driving Chaperone Action. *Cell* **166**, 369–379 (2016).
9. Horowitz, S. *et al.* Visualizing chaperone-assisted protein folding. *Nat Struct Mol Biol* **23**, 691–697 (2016).
10. He, L., Sharpe, T., Mazur, A. & Hiller, S. A molecular mechanism of chaperone-client recognition. *Sci Adv* **2**, e1601625–e1601625 (2016).
11. Mitra, R., Wu, K., Lee, C. & Bardwell, J. C. A. ATP-Independent Chaperones. *Annu Rev Biophys* **51**, 409–429 (2022).

12. Knowling, S. E., Figueiredo, A. M., Whittaker, S. B. M., Moore, G. R. & Radford, S. E. Amino Acid Insertion Reveals a Necessary Three-Helical Intermediate in the Folding Pathway of the Colicin E7 Immunity Protein Im7. *J Mol Biol* **392**, 1074–1086 (2009).

13. Stull, F., Betton, J.-M. & Bardwell, J. C. A. Periplasmic Chaperones and Prolyl Isomerases. *EcoSal Plus* **8**, (2018).

14. Bitto, E. & McKay, D. B. Crystallographic structure of SurA, a molecular chaperone that facilitates folding of outer membrane porins. *Structure* **10**, 1489–1498 (2002).

15. Behrens, S., Maier, R., De Cock, H., Schmid, F. X. & Gross, C. A. The SurA periplasmic PPIase lacking its parvulin domains functions in vivo and has chaperone activity. *EMBO Journal* **20**, 285–294 (2001).

16. Humes, J. R. *et al.* The Role of SurA PPIase Domains in Preventing Aggregation of the Outer-Membrane Proteins tOmpA and OmpT. *J Mol Biol* **431**, 1267–1283 (2019).

17. Calabrese, A. N. *et al.* Inter-domain dynamics in the chaperone SurA and multi-site binding to its outer membrane protein clients. *Nat Commun* **11**, (2020).

18. Soltes, G. R., Schwalm, J., Ricci, D. P. & Silhavy, T. J. The activity of Escherichia coli chaperone SurA is regulated by conformational changes involving a parvulin domain. *J Bacteriol* **198**, 921–929 (2016).

19. Marx, D. C. *et al.* SurA is a cryptically grooved chaperone that expands unfolded outer membrane proteins. *Proc Natl Acad Sci U S A* **117**, 28026–28035 (2020).

20. Fleming, K. G. Energetics of membrane protein folding. *Annu Rev Biophys* **43**, 233–255 (2014).

21. Chamachi, N. *et al.* Chaperones Skp and SurA dynamically expand unfolded OmpX and synergistically disassemble oligomeric aggregates. *Proceedings of the National Academy of Sciences* **119**, (2022).

22. Thoma, J., Burmann, B. M., Hiller, S. & Müller, D. J. Impact of holdase chaperones Skp and SurA on the folding of β-barrel outer-membrane proteins. *Nature Structural and Molecular Biology* vol. 22 795–802 Preprint at https://doi.org/10.1038/nsmb.3087 (2015).

23. Gajiwala, K. S. & Burley, S. K. HDEA, a periplasmic protein that supports acid resistance in pathogenic enteric bacteria. *J Mol Biol* **295**, 605–612 (2000).

24. Hong, W., Wu, Y. E., Fu, X. & Chang, Z. Chaperone-dependent mechanisms for acid resistance in enteric bacteria. *Trends Microbiol* **20**, 328–335 (2012).

25. Zhang, M. *et al.* A genetically incorporated crosslinker reveals chaperone cooperation in acid resistance. *Nat Chem Biol* **7**, 671–677 (2011).

26. Yang, F., Gustafson, K. R., Boyd, M. R. & Wlodawer, A. Crystal structure of Escherichia coli HdeA. *Nat Struct Biol* **5**, 763–764 (1998).

27. Foit, L., George, J. S., Zhang, B. W., Brooks, C. L. & Bardwell, J. C. A. Chaperone activation by unfolding. *Proceedings of the National Academy of Sciences* **110**, (2013).

28. Zhai, Z. *et al.* Roles of structural plasticity in chaperone HdeA activity are revealed by 19F NMR. *Chem Sci* **7**, 2222–2228 (2016).

29. Yu, X.-C., Hu, Y., Ding, J., Li, H. & Jin, C. Structural basis and mechanism of the unfolding-induced activation of HdeA, a bacterial acid response chaperone. *Journal of Biological Chemistry* **294**, 3192–3206 (2019).

30. Salmon, L. *et al.* The Mechanism of HdeA Unfolding and Chaperone Activation. *J Mol Biol* **430**, 33–40 (2018).

31. Dickson, A., Ahlstrom, L. S. & Brooks, C. L. Coupled folding and binding with 2D Window-Exchange Umbrella Sampling. *J Comput Chem* **37**, 587–594 (2016).

32. Van Nostrand, E. L. *et al.* A large-scale binding and functional map of human RNA-binding proteins. *Nature* **583**, 711–719 (2020).

33. Calabretta, S. & Richard, S. Emerging Roles of Disordered Sequences in RNA-Binding Proteins. *Trends in Biochemical Sciences* vol. 40 662–672 Preprint at https://doi.org/10.1016/j.tibs.2015.08.012 (2015).

34. Thandapani, P., O'Connor, T. R., Bailey, T. L. & Richard, S. Defining the RGG/RG Motif. *Molecular Cell* vol. 50 613–623 Preprint at https://doi.org/10.1016/j.molcel.2013.05.021 (2013).

35. Phan, A. T. *et al.* Structure-function studies of FMRP RGG peptide recognition of an RNA duplex-quadruplex junction. *Nat Struct Mol Biol* **18**, 796–804 (2011).

36. Portz, B. *et al.* Structural heterogeneity in the intrinsically disordered RNA polymerase II C-terminal domain. *Nat Commun* **8**, (2017).

37. Gibbs, E. B. *et al.* Phosphorylation induces sequence-specific conformational switches in the RNA polymerase II C-terminal domain. *Nat Commun* **8**, (2017).

38. Abyzov, A., Blackledge, M. & Zweckstetter, M. Conformational Dynamics of Intrinsically Disordered Proteins Regulate Biomolecular Condensate Chemistry. *Chemical Reviews* vol. 122 6719–6748 Preprint at https://doi.org/10.1021/acs.chemrev.1c00774 (2022).

39. Jo, Y., Jang, J., Song, D., Park, H. & Jung, Y. Determinants for intrinsically disordered protein recruitment into phase-separated protein condensates. *Chem Sci* **13**, 522–530 (2022).

40. Ahmed, R. & Forman-Kay, J. D. NMR insights into dynamic, multivalent interactions of intrinsically disordered regions: From discrete complexes to condensates. *Essays Biochem* **66**, 863–873 (2022).

41. van Ham, T. J. *et al.* Identification of MOAG-4/SERF as a regulator of age-related proteotoxicity. *Cell* **142**, 601–612 (2010).

42. Meinen, B. A., Gadkari, V. V., Stull, F., Ruotolo, B. T. & Bardwell, J. C. A. SERF engages in a fuzzy complex that accelerates primary nucleation of amyloid proteins. *Proceedings of the National Academy of Sciences* **116**, 23040–23049 (2019).

43. Eschweiler, J. D., Kerr, R., Rabuck-Gibbons, J. & Ruotolo, B. T. Sizing up protein-ligand complexes: The rise of structural mass spectrometry approaches in the pharmaceutical sciences. *Annual Review of Analytical Chemistry* **10**, 25–44 (2017).

44. Beveridge, R. *et al.* A mass-spectrometry-based framework to define the extent of disorder in proteins. *Anal Chem* **86**, 10979–10991 (2014).

45. Meyer, N. H. *et al.* Structural Fuzziness of the RNA-Organizing Protein SERF Determines a Toxic Gain-of-interaction. *J Mol Biol* **432**, 930–951 (2020).

46. Lambert, N. *et al.* RNA Bind-n-Seq: Quantitative Assessment of the Sequence and Structural Binding Specificity of RNA Binding Proteins. *Mol Cell* **54**, 887–900 (2014).

47. Ray, D. *et al.* RNA-binding proteins that lack canonical RNA-binding domains are rarely sequence-specific. *Sci Rep* **13**, 5238 (2023).

48. Hafner, M. *et al.* PAR-CliP - A Method to Identify Transcriptome-wide the Binding Sites of RNA Binding Proteins. *Journal of Visualized Experiments* (2010) doi:10.3791/2034.

49. Bollen, Y. J. M. & Van Mierlo, C. P. M. Protein topology affects the appearance of intermediates during the folding of proteins with a flavodoxin-like fold. *Biophys Chem* **114**, 181–189 (2005).

50. Chakraborty, K. *et al.* Chaperonin-catalyzed rescue of kinetically trapped states in protein folding. *Cell* **142**, 112–122 (2010).

51. Koldewey, P., Horowitz, S. & Bardwell, J. C. A. Chaperone-client interactions: Non-specificity engenders multifunctionality. *Journal of Biological Chemistry* **292**, 12010–12017 (2017).

52. Mashaghi, A. *et al.* Reshaping of the conformational search of a protein by the chaperone trigger factor. *Nature* **500**, 98–101 (2013).

53. Huang, C., Rossi, P., Saio, T. & Kalodimos, C. G. Structural basis for the antifolding activity of a molecular chaperone. *Nature* **537**, 202–206 (2016).

54. Ungelenk, S. *et al.* Small heat shock proteins sequester misfolding proteins in near-native conformation for cellular protection and efficient refolding. *Nat Commun* **7**, 13673 (2016).

55. Wu, K., Stull, F., Lee, C. & Bardwell, J. C. A. Protein folding while chaperone bound is dependent on weak interactions. *Nat Commun* **10**, 4833 (2019).

56. Horowitz, S., Koldewey, P., Stull, F. & Bardwell, J. C. Folding while bound to chaperones. *Curr Opin Struct Biol* **48**, 1–5 (2018).

57. Bollen, Y. J. M., Sánchez, I. E. & Van Mierlo, C. P. M. Formation of on- and off-pathway intermediates in the folding kinetics of Azotobacter vinelandii apoflavodoxin. *Biochemistry* **43**, 10475–10489 (2004).

58. Houwman, J. A. & van Mierlo, C. P. M. Folding of proteins with a flavodoxin-like architecture. *FEBS J* **284**, 3145–3167 (2017).

59. J. Fernández-Recio, ‡, C. G. Genzor, § and & Sancho*, J. Apoflavodoxin Folding Mechanism: An α/β Protein with an Essentially Off-Pathway Intermediate†. (2001) doi:10.1021/BI010216T.

60. Fernández-Recio, J., Genzor, C. G. & Sancho, J. Apoflavodoxin folding mechanism: An α/β protein with an essentially off-pathway intermediate. *Biochemistry* **40**, 15234–15245 (2001).

61. Brockwell, D. J. & Radford, S. E. Intermediates: ubiquitous species on folding energy landscapes? *Curr Opin Struct Biol* **17**, 30–37 (2007).

62. Mattoo, R. U. H. & Goloubinoff, P. Molecular chaperones are nanomachines that catalytically unfold misfolded and alternatively folded proteins. *Cellular and Molecular Life Sciences* **71**, 3311–3325 (2014).

63. Irún, M. P., Garcia-Mira, M. M., Sanchez-Ruiz, J. M. & Sancho, J. Native hydrogen bonds in a molten globule: The apoflavodoxin thermal intermediate. *J Mol Biol* **306**, 877–888 (2001).

64. Ayuso-Tejedor, S. *et al.* Design and structure of an equilibrium protein folding intermediate: A hint into dynamical regions of proteins. *J Mol Biol* **400**, 922–934 (2010).

65. Bueno, M., Ayuso-Tejedor, S. & Sancho, J. Do proteins with similar folds have similar transition state structures? A diffuse transition state of the 169 residue apoflavodoxin. *J Mol Biol* **359**, 813–824 (2006).

66. Konrat, R. NMR contributions to structural dynamics studies of intrinsically disordered proteins. *Journal of Magnetic Resonance* **241**, 74–85 (2014).

67. Hyung, S. J. & Ruotolo, B. T. Integrating mass spectrometry of intact protein complexes into structural proteomics. *Proteomics* **12**, 1547–1564 (2012).

68. Mason, E. A. & McDaniel, E. W. *Transport Properties of Ions in Gases*. *Transport Properties of Ions in Gases* (Wiley, 1988). doi:10.1002/3527602852.

69. Gadkari, V. V. *et al.* Investigation of sliding DNA clamp dynamics by single-molecule fluorescence, mass spectrometry and structure-based modeling. *Nucleic Acids Res* **46**, 3103–3118 (2018).

70. Bush, M. F. *et al.* Collision cross sections of proteins and their complexes: A calibration framework and database for gas-phase structural biology. *Anal Chem* **82**, 9557–9565 (2010).

71. Gadkari, V. V. *et al.* Enhanced Collision Induced Unfolding and Electron Capture Dissociation of Native-like Protein Ions. *Anal Chem* **92**, 15489–15496 (2020).

72. Maldonado, S. *et al.* Salt-induced stabilization of apoflavodoxin at neutral pH is mediated through cation-specific effects. *Protein Science* **11**, 1260–1273 (2002).

73. Nelson, E. D. & Grishin, N. V. Alternate Pathways for Folding in the Flavodoxin Fold Family Revealed by a Nucleation-growth Model. *J Mol Biol* **358**, 646–653 (2006).

74. Campos, L. A., Bueno, M., Lopez-Llano, J., Jiménez, M. Á. & Sancho, J. Structure of stable protein folding intermediates by equilibrium φ-analysis: The apoflavodoxin thermal intermediate. *J Mol Biol* **344**, 239–255 (2004).

75. Ptitsyn, O. B. Molten globule and protein folding. in *Advances in Protein Chemistry* vol. 47 83–229 (1995).

76. Lindhoud, S., Pirchi, M., Westphal, A. H., Haran, G. & Van Mierlo, C. P. M. Gradual Folding of an Off-Pathway Molten Globule Detected at the Single-Molecule Level. *J Mol Biol* **427**, 3148–3157 (2015).

77. Houwman, J. A., André, E., Westphal, A. H., Van Berkel, W. J. H. & Van Mierlo, C. P. M. The ribosome restrains molten globule formation in stalled nascent flavodoxin. *Journal of Biological Chemistry* **291**, 25911–25920 (2016).

78. Engel, R. *et al.* Macromolecular crowding compacts unfolded apoflavodoxin and causes severe aggregation of the off-pathway intermediate during apoflavodoxin folding. *Journal of Biological Chemistry* **283**, 27383–27394 (2008).

79. Quan, S. *et al.* Super Spy variants implicate flexibility in chaperone action. *Elife* **2014**, 1–22 (2014).

80. Bornemann, T., Holtkamp, W. & Wintermeyer, W. Interplay between trigger factor and other protein biogenesis factors on the ribosome. *Nat Commun* **5**, (2014).

81. Wu, K., Minshull, T. C., Radford, S. E., Calabrese, A. N. & Bardwell, J. C. A. Trigger factor both holds and folds its client proteins. *Nat Commun* **13**, 1–15 (2022).

82. Qu, J., Mayer, C., Behrens, S., Holst, O. & Kleinschmidt, J. H. The trimeric periplasmic chaperone Skp of Escherichia coli forms 1:1 complexes with outer membrane proteins via hydrophobic and electrostatic interactions. *J Mol Biol* **374**, 91–105 (2007).

83. Haslbeck, M., Weinkauf, S. & Buchner, J. Small heat shock proteins: Simplicity meets complexity. *Journal of Biological Chemistry* **294**, 2121–2132 (2019).

84. Basha, E., O'Neill, H. & Vierling, E. Small heat shock proteins and α-crystallins: Dynamic proteins with flexible functions. *Trends Biochem Sci* **37**, 106–117 (2012).

85. Jovcevski, B. *et al.* Phosphomimics destabilize Hsp27 oligomeric assemblies and enhance chaperone activity. *Chem Biol* **22**, 186–195 (2015).

86. Marley, J., Lu, M. & Bracken, C. A method for efficient isotopic labeling of recombinant proteins. *J Biomol NMR* **20**, 71–75 (2001).

87. Genzor, C. G. *et al.* Conformational stability of apoflavodoxin. *Protein Science* **5**, 1376–1388 (1996).

88. Van Mierlo, C. P. M., Van Dongen, W. M. A. M., Vergeldt, F., Van Berkel, W. J. H. & Steensma, E. The equilibrium unfolding of Azotobacter vinelandii apoflavodoxin II occurs via a relatively stable folding intermediate. *Protein Science* **7**, 2331–2344 (1998).

89. Demeler, B. & Gorbet, G. E. Analytical ultracentrifugation data analysis with ultrascan-III. in *Analytical Ultracentrifugation: Instrumentation, Software, and Applications* 119–143 (Springer Japan, 2016). doi:10.1007/978-4-431-55985-6_8.

90. Brookes, E., Cao, W. & Demeler, B. A two-dimensional spectrum analysis for sedimentation velocity experiments of mixtures with heterogeneity in molecular weight and shape. *European Biophysics Journal* **39**, 405–414 (2010).

91. Brookes, E. & Demeler, B. Genetic algorithm optimization for obtaining accurate molecular weight distributions from sedimentation velocity experiments. *Prog Colloid Polym Sci* **131**, 33–40 (2006).

92. Micsonai, A. *et al.* BeStSel: A web server for accurate protein secondary structure prediction and fold recognition from the circular dichroism spectra. *Nucleic Acids Res* **46**, W315–W322 (2018).

93. Lee, W., Tonelli, M. & Markley, J. L. NMRFAM-SPARKY: Enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **31**, 1325–1327 (2015).

94. Delaglio, F. *et al.* NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**, 277–293 (1995).

95. Vallejo, D. D. *et al.* A modified drift tube ion mobility-mass spectrometer for charge-multiplexed collision-induced unfolding. *Anal Chem* **91**, 8137–8146 (2019).

96. Polasky, D. A., Dixit, S. M., Fantin, S. M. & Ruotolo, B. T. CIUSuite 2: Next-Generation Software for the Analysis of Gas-Phase Protein Unfolding Data. *Anal Chem* **91**, 3147–3155 (2019).

97. Stow, S. M. *et al.* An interlaboratory evaluation of drift tube ion mobility-mass spectrometry collision cross section measurements. *Anal Chem* **89**, 9048–9055 (2017).

98. Strohalm, M., Hassman, M., Košata, B. & Kodíček, M. mMass data miner: An open source alternative for mass spectrometric data analysis. *Rapid Communications in Mass Spectrometry* **22**, 905–908 (2008).

99. Strohalm, M., Kavan, D., Novák, P., Volný, M. & Havlíček, V. MMass 3: A cross-platform software environment for precise analysis of mass spectrometric data. *Anal Chem* **82**, 4648–4651 (2010).

100. Dignon, G. L., Best, R. B. & Mittal, J. Biomolecular phase separation: From molecular driving forces to macroscopic properties. *Annu Rev Phys Chem* **71**, 53–75 (2020).

101. Shin, Y. & Brangwynne, C. P. Liquid phase condensation in cell physiology and disease. *Science (1979)* **357**, (2017).

102. Lyon, A. S., Peeples, W. B. & Rosen, M. K. A framework for understanding the functions of biomolecular condensates across scales. *Nature Reviews Molecular Cell Biology* vol. 22 215–235 Preprint at https://doi.org/10.1038/s41580-020-00303-z (2021).

103. Mitrea, D. M. *et al.* Nucleophosmin integrates within the nucleolus via multi-modal interactions with proteins displaying R-rich linear motifs and rRNA. *Elife* **5**, (2016).

104. Yang, P. *et al.* G3BP1 Is a Tunable Switch that Triggers Phase Separation to Assemble Stress Granules. *Cell* **181**, 325-345.e28 (2020).

105. Zhang, H. *et al.* RNA Controls PolyQ Protein Phase Transitions. *Mol Cell* **60**, 220–230 (2015).

106. Lin, Y., Protter, D. S. W., Rosen, M. K. & Parker, R. Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins. *Mol Cell* **60**, 208–219 (2015).

107. Putnam, A., Cassani, M., Smith, J. & Seydoux, G. A gel phase promotes condensation of liquid P granules in Caenorhabditis elegans embryos. *Nat Struct Mol Biol* **26**, 220–226 (2019).

108. Clemson, C. M. *et al.* An Architectural Role for a Nuclear Noncoding RNA: NEAT1 RNA Is Essential for the Structure of Paraspeckles. *Mol Cell* **33**, 717–726 (2009).

109. Peran, I. & Mittag, T. Molecular structure in biomolecular condensates. *Current Opinion in Structural Biology* vol. 60 17–26 Preprint at https://doi.org/10.1016/j.sbi.2019.09.007 (2020).

110. Kar, M. *et al.* Phase-separating RNA-binding proteins form heterogeneous distributions of clusters in subsaturated solutions. *Proc Natl Acad Sci U S A* **119**, 1–12 (2022).

111. Martin, E. W. *et al.* A multi-step nucleation process determines the kinetics of prion-like domain phase separation. *Nat Commun* **12**, (2021).

112. Shimobayashi, S. F., Ronceray, P., Sanders, D. W., Haataja, M. P. & Brangwynne, C. P. Nucleation landscape of biomolecular condensates. *Nature* **599**, 503–506 (2021).

113. Wiedner, H. J. & Giudice, J. It's not just a phase: function and characteristics of RNA-binding proteins in phase separation. *Nat Struct Mol Biol* **28**, 465–473 (2021).

114. Nott, T. J. *et al.* Phase Transition of a Disordered Nuage Protein Generates Environmentally Responsive Membraneless Organelles. *Mol Cell* **57**, 936–947 (2015).

115. Ryan, V. H. *et al.* Mechanistic View of hnRNPA2 Low-Complexity Domain Structure, Interactions, and Phase Separation Altered by Mutation and Arginine Methylation. *Mol Cell* **69**, 465-479.e7 (2018).

116. Corbet, G. A., Wheeler, J. R., Parker, R. & Weskamp, K. TDP43 ribonucleoprotein granules: physiologic function to pathologic aggregates. *RNA Biology* vol. 18 128–138 Preprint at https://doi.org/10.1080/15476286.2021.1963099 (2021).

117. Murthy, A. C. *et al.* Molecular interactions underlying liquid−liquid phase separation of the FUS low-complexity domain. *Nat Struct Mol Biol* **26**, 637–648 (2019).

118. Yamazaki, T. *et al.* Functional Domains of NEAT1 Architectural lncRNA Induce Paraspeckle Assembly through Phase Separation. *Mol Cell* **70**, 1038-1053.e7 (2018).

119. Langdon, E. M. *et al. mRNA structure determines specificity of a polyQ-driven phase separation Downloaded from*. *Science* vol. 360 http://science.sciencemag.org/ (2018).

120. Daneshvar, K. *et al.* lncRNA DIGIT and BRD3 protein form phase-separated condensates to regulate endoderm differentiation. *Nat Cell Biol* **22**, 1211–1222 (2020).

121. Somarowthu, S. *et al.* HOTAIR Forms an Intricate and Modular Secondary Structure. *Mol Cell* **58**, 353–361 (2015).

122. Leontis, N. B., Lescoute, A. & Westhof, E. The building blocks and motifs of RNA architecture. *Current Opinion in Structural Biology* vol. 16 279–287 Preprint at https://doi.org/10.1016/j.sbi.2006.05.009 (2006).

123. Seim, I. *et al.* Dilute phase oligomerization can oppose phase separation and modulate material properties of a ribonucleoprotein condensate. *Proc Natl Acad Sci U S A* **119**, 1–11 (2022).

124. Falsone, S. F. *et al.* SERF Protein Is a Direct Modifier of Amyloid Fiber Assembly. *Cell Rep* **2**, 358–371 (2012).

125. Meinen, B. A., Gadkari, V. V., Stull, F., Ruotolo, B. T. & Bardwell, J. C. A. SERF engages in a fuzzy complex that accelerates primary nucleation of amyloid proteins. *Proceedings of the National Academy of Sciences* **116**, 23040–23049 (2019).

126. Sahoo, B. R. & Bardwell, J. C. A. SERF, a family of tiny highly conserved, highly charged proteins with enigmatic functions. *FEBS Journal* Preprint at https://doi.org/10.1111/febs.16555 (2022).

127. Corley, M., Burns, M. C. & Yeo, G. W. How RNA-Binding Proteins Interact with RNA: Molecules and Mechanisms. *Molecular Cell* vol. 78 9–29 Preprint at https://doi.org/10.1016/j.molcel.2020.03.011 (2020).

128. Herschlag, D., Bonilla, S. & Bisaria, N. The story of RNA folding, as told in epochs. *Cold Spring Harb Perspect Biol* **10**, (2018).

129. Gracia, B. *et al.* RNA Structural Modules Control the Rate and Pathway of RNA Folding and Assembly. *J Mol Biol* **428**, 3972–3985 (2016).

130. Leamy, K. A., Assmann, S. M., Mathews, D. H. & Bevilacqua, P. C. Bridging the gap between in vitro and in vivo RNA folding . *Q Rev Biophys* **49**, (2016).

131. Solomatin, S. V., Greenfeld, M., Chu, S. & Herschlag, D. Multiple native states reveal persistent ruggedness of an RNA folding landscape. *Nature* **463**, 681–684 (2010).

132. Chu, V. B. *et al.* Do conformational biases of simple helical junctions influence RNA folding stability and specificity? *RNA* **15**, 2195–2205 (2009).

133. Bisaria, N. *et al.* Kinetic and thermodynamic framework for P4-P6 RNA reveals tertiary motif modularity and modulation of the folding preferred pathway. *Proc Natl Acad Sci U S A* **113**, E4956–E4965 (2016).

134. Al-Hashimi, H. M. NMR studies of nucleic acid dynamics. *Journal of Magnetic Resonance* **237**, 191–204 (2013).

135. Ganser, L. R., Kelly, M. L., Herschlag, D. & Al-Hashimi, H. M. The roles of structural dynamics in the cellular functions of RNAs. *Nature Reviews Molecular Cell Biology* vol. 20 474–489 Preprint at https://doi.org/10.1038/s41580-019-0136-0 (2019).

136. Mustoe, A. M., Brooks, C. L. & Al-Hashimi, H. M. Hierarchy of RNA functional dynamics. *Annual Review of Biochemistry* vol. 83 441–466 Preprint at https://doi.org/10.1146/annurev-biochem-060713-035524 (2014).

137. Dormann, D. & Haass, C. TDP-43 and FUS: a nuclear affair. *Trends Neurosci* **34**, 339–348 (2011).

138. Tauber, D., Tauber, G. & Parker, R. Mechanisms and Regulation of RNA Condensation in RNP Granule Formation. *Trends Biochem Sci* **45**, 764–778 (2020).

139. Gibbs, E. B. & Showalter, S. A. Quantitative biophysical characterization of intrinsically disordered proteins. *Biochemistry* **54**, 1314–1326 (2015).

140. Lawrence, C. W. & Showalter, S. A. Carbon-detected 15N NMR spin relaxation of an intrinsically disordered protein: FCP1 dynamics unbound and in complex with RAP74. *Journal of Physical Chemistry Letters* **3**, 1409–1413 (2012).

141. Sahu, D., Bastidas, M. & Showalter, S. A. Generating NMR chemical shift assignments of intrinsically disordered proteins using carbon-detected NMR methods. *Anal Biochem* **449**, 17–25 (2014).

142. Cook, E. C., Sahu, D., Bastidas, M. & Showalter, S. A. Solution Ensemble of the C-Terminal Domain from the Transcription Factor Pdx1 Resembles an Excluded Volume Polymer. *Journal of Physical Chemistry B* **123**, 106–116 (2019).

143. Bastidas, M., Gibbs, E. B., Sahu, D. & Showalter, S. A. A primer for carbon-detected NMR applications to intrinsically disordered proteins in solution. *Concepts Magn Reson Part A Bridg Educ Res* **44**, 54–66 (2015).

144. Camacho-Zarco, A. R. *et al.* NMR Provides Unique Insight into the Functional Dynamics and Interactions of Intrinsically Disordered Proteins. *Chemical Reviews* Preprint at https://doi.org/10.1021/acs.chemrev.1c01023 (2021).

145. Marsh, J. A., Singh, V. K., Jia, Z. & Forman-Kay, J. D. Sensitivity of secondary structure propensities to sequence differences between α- and γ-synuclein: Implications for fibrillation. *Protein Science* **15**, 2795–2804 (2006).

146. Abyzov, A. *et al.* Identification of Dynamic Modes in an Intrinsically Disordered Protein Using Temperature-Dependent NMR Relaxation. *J Am Chem Soc* **138**, 6240–6251 (2016).

147. Salvi, N., Abyzov, A. & Blackledge, M. Multi-Timescale Dynamics in Intrinsically Disordered Proteins from NMR Relaxation and Molecular Simulation. *Journal of Physical Chemistry Letters* **7**, 2483–2489 (2016).

148. Rezaei-Ghaleh, N. *et al.* Local and Global Dynamics in Intrinsically Disordered Synuclein. *Angewandte Chemie* **130**, 15482–15486 (2018).

149. Ullman, O., Fisher, C. K. & Stultz, C. M. Explaining the structural plasticity of α-synuclein. *J Am Chem Soc* **133**, 19536–19546 (2011).

150. Marius Clore, G. & Iwahara, J. Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes. *Chem Rev* **109**, 4108–4139 (2009).

151. Anthis, N. J. & Clore, G. M. Visualizing transient dark states by NMR spectroscopy. *Q Rev Biophys* **48**, 35–116 (2015).

152. Dinic, J., Marciel, A. B. & Tirrell, M. V. Polyampholyte physics: Liquid–liquid phase separation and biological condensates. *Curr Opin Colloid Interface Sci* **54**, 101457 (2021).

153. Suryawanshi, H., Sabharwal, H. & Maiti, S. Thermodynamics of peptide-RNA recognition: The binding of a tat peptide to TAR RNA. *Journal of Physical Chemistry B* **114**, 11155–11163 (2010).

154. Pitt, S. W., Majumdar, A., Serganov, A., Patel, D. J. & Al-Hashimi, H. M. Argininamide binding arrests global motions in HIV-1 TAR RNA: Comparison with Mg2+-induced conformational stabilization. *J Mol Biol* **338**, 7–16 (2004).

155. Aboul-ela, F., Karn, J. & Varani, G. The Structure of the Human Immunodeficiency Virus Type-1 TAR RNA Reveals Principles of RNA Recognition by Tat Protein. *J Mol Biol* **253**, 313–332 (1995).

156. Sugase, K., Dyson, H. J. & Wright, P. E. Mechanism of coupled folding and binding of an intrinsically disordered protein. *Nature* **447**, 1021–1025 (2007).

157. Arai, M., Sugase, K., Dyson, H. J. & Wright, P. E. Conformational propensities of intrinsically disordered proteins influence the mechanism of binding and folding. *Proc Natl Acad Sci U S A* **112**, 9614–9619 (2015).

158. Ottoz, D. S. M. & Berchowitz, L. E. The role of disorder in RNA binding affinity and specificity. *Open Biol* **10**, 200328 (2020).

159. Lester, E. *et al.* Tau aggregates are RNA-protein assemblies that mislocalize multiple nuclear speckle components. *Neuron* **109**, 1675-1691.e9 (2021).

160. Ditlev, J. A., Case, L. B. & Rosen, M. K. Who's In and Who's Out—Compositional Control of Biomolecular Condensates. *J Mol Biol* **430**, 4666–4684 (2018).

161. Garcia-Jove Navarro, M. *et al.* RNA is a critical element for the sizing and the composition of phase-separated RNA–protein condensates. *Nat Commun* **10**, (2019).

162. Smith, C. A. Estimation of sedimentation coefficients and frictional ratios of globular proteins. *Biochem Educ* **16**, 104–106 (1988).

163. Salvay, A. G., Communie, G. & Ebel, C. Sedimentation velocity analytical ultracentrifugation for intrinsically disordered proteins. *Methods in Molecular Biology* **896**, 91–105 (2012).

164. Manon, F. & Ebel, C. Analytical Ultracentrifugation, a Useful Tool to Probe Intrinsically Disordered Proteins. *Instrumental Analysis of Intrinsically Disordered Proteins: Assessing Structure and Conformation* 431–449 (2010) doi:10.1002/9780470602614.ch15.

165. Demeler, B. & Van Holde, K. E. Sedimentation velocity analysis of highly heterogeneous systems. *Anal Biochem* **335**, 279–288 (2004).

166. Dignon, G. L., Zheng, W., Best, R. B., Kim, Y. C. & Mittal, J. Relation between single-molecule properties and phase behavior of intrinsically disordered proteins. *Proc Natl Acad Sci U S A* **115**, 9929–9934 (2018).

167. Schuster, B. S. *et al.* Identifying sequence perturbations to an intrinsically disordered protein that determine its phase-separation behavior. *PNAS* **117**, (2000).

168. Sing, C. E. & Perry, S. L. Recent progress in the science of complex coacervation. *Soft Matter* **16**, 2885–2914 (2020).

169. Boeynaems, S. *et al.* Spontaneous driving forces give rise to protein−RNA condensates with coexisting phases and complex material properties. *Proc Natl Acad Sci U S A* **116**, 7889–7898 (2019).

170. Jain, A. & Vale, R. D. RNA phase transitions in repeat expansion disorders. *Nature* **546**, 243–247 (2017).

171. Pak, C. W. *et al.* Sequence Determinants of Intracellular Phase Separation by Complex Coacervation of a Disordered Protein. *Mol Cell* **63**, 72–85 (2016).

172. Aumiller, W. M. & Keating, C. D. Phosphorylation-mediated RNA/peptide complex coacervation as a model for intracellular liquid organelles. *Nat Chem* **8**, 129–137 (2016).

173. Fay, M. M. & Anderson, P. J. The Role of RNA in Biological Phase Separations. *J Mol Biol* **430**, 4685–4701 (2018).

174. Peran, I. & Mittag, T. Molecular structure in biomolecular condensates. *Current Opinion in Structural Biology* vol. 60 17–26 Preprint at https://doi.org/10.1016/j.sbi.2019.09.007 (2020).

175. Dignon, G. L., Best, R. B. & Mittal, J. Biomolecular phase separation: From molecular driving forces to macroscopic properties. *Annu Rev Phys Chem* **71**, 53–75 (2020).

176. Choi, J. M., Holehouse, A. S. & Pappu, R. V. Physical Principles Underlying the Complex Biology of Intracellular Phase Transitions. *Annu Rev Biophys* **49**, 107–133 (2020).

177. Liquid-liquid phase separation is driven by large-scale conformational unwinding and fluctuations of intrinsically disordered protein molecules.

178. Zarin, T. *et al.* Proteome-wide signatures of function in highly diverged intrinsically disordered regions. *Elife* **8**, (2019).

179. Davey, N. E. *et al.* Attributes of short linear motifs. *Mol Biosyst* **8**, 268–281 (2012).

180. Kumar, M. *et al.* The Eukaryotic Linear Motif resource: 2022 release. *Nucleic Acids Res* **50**, D497–D508 (2022).

181. Henley, M. J. *et al.* Unexpected specificity within dynamic transcriptional protein–protein complexes. *Proceedings of the National Academy of Sciences* **117**, 27346–27353 (2020).

182. Bellay, J. *et al.* Bringing order to protein disorder through comparative genomics and genetic interactions. *Genome Biol* **12**, (2011).

183. Johnson, M. E. & Hummer, G. Nonspecific Binding Limits the Number of Proteins in a Cell and Shapes their Interaction Networks. *Biophys J* **100**, 32a (2011).

184. Corpet, F. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res* **16**, 10881–10890 (1988).

185. Tamiola, K., Acar, B. & Mulder, F. A. A. Sequence-specific random coil chemical shifts of intrinsically disordered proteins. *J Am Chem Soc* **132**, 18000–18003 (2010).

186. Lawrence, C. W. & Showalter, S. A. Carbon-detected 15N NMR spin relaxation of an intrinsically disordered protein: FCP1 dynamics unbound and in complex with RAP74. *Journal of Physical Chemistry Letters* **3**, 1409–1413 (2012).

187. Rosenzweig, R., Nillegoda, N. B., Mayer, M. P. & Bukau, B. The Hsp70 chaperone network. *Nat Rev Mol Cell Biol* **20**, 665–680 (2019).

188. Park, C., Zhou, S., Gilmore, J. & Marqusee, S. Energetics-based Protein Profiling on a Proteomic Scale: Identification of Proteins Resistant to Proteolysis. *J Mol Biol* **368**, 1426–1437 (2007).

189. He, W. *et al.* Chaperone Spy Protects Outer Membrane Proteins from Folding Stress via Dynamic Complex Formation. *mBio* **12**, (2021).

190. Lyu, Z. X. & Zhao, X. S. Periplasmic quality control in biogenesis of outer membrane proteins. *Biochem Soc Trans* **43**, 133–138 (2015).

191. Schwalm, J., Mahoney, T. F., Soltes, G. R. & Silhavy, T. J. Role for Skp in LptD assembly in Escherichia coli. *J Bacteriol* **195**, 3734–3742 (2013).

192. Brehme, M. *et al.* A chaperome subnetwork safeguards proteostasis in aging and neurodegenerative disease. *Cell Rep* **9**, 1135–1150 (2014).

193. Watanabe, M. & Blobel, G. High-affinity binding of Escherichia coli SecB to the signal sequence region of a presecretory protein. *Proceedings of the National Academy of Sciences* **92**, 10133–10136 (1995).

194. Schiffrin, B. *et al.* Skp is a multivalent chaperone of outer-membrane proteins. *Nat Struct Mol Biol* **23**, 786–793 (2016).

195. Saio, T., Guan, X., Rossi, P., Economou, A. & Kalodimos, C. G. Structural basis for protein antiaggregation activity of the trigger factor chaperone. *Science* **344**, 1250494 (2014).

196. Moon, C. P., Zaccai, N. R., Fleming, P. J., Gessmann, D. & Fleming, K. G. Membrane protein thermodynamic stability may serve as the energy sink for sorting in the periplasm. *Proceedings of the National Academy of Sciences* **110**, 4285–4290 (2013).

197. Stenberg, G. & Fersht, A. R. Folding of barnase in the presence of the molecular chaperone SecB. *J Mol Biol* **274**, 268–275 (1997).

198. Sathish, H. A., Stein, R. A., Yang, G. & Mchaourab, H. S. Mechanism of Chaperone Function in Small Heat-shock Proteins. *Journal of Biological Chemistry* **278**, 44214–44221 (2003).

199. Jiang, Y., Rossi, P. & Kalodimos, C. G. Structural basis for client recognition and activity of Hsp40 chaperones. *Science (1979)* **365**, 1313–1319 (2019).

200. Jensen, M. R., Zweckstetter, M., Huang, J. & Blackledge, M. Exploring Free-Energy Landscapes of Intrinsically Disordered Proteins at Atomic Resolution Using NMR Spectroscopy. *Chem Rev* **114**, 6632–6660 (2014).

201. Murthy, A. C. & Fawzi, N. L. The (un)structural biology of biomolecular liquid-liquid phase separation using NMR spectroscopy. *Journal of Biological Chemistry* **295**, 2375–2384 (2020).

202. Conicella, A. E. *et al.* TDP-43 α-helical structure tunes liquid–liquid phase separation and function. *Proceedings of the National Academy of Sciences* **117**, 5883–5894 (2020).

203. Reichheld, S. E., Muiznieks, L. D., Keeley, F. W. & Sharpe, S. Direct observation of structure and dynamics during phase separation of an elastomeric protein. *Proc Natl Acad Sci U S A* **114**, E4408–E4415 (2017).

204. Brady, J. P. *et al.* Structural and hydrodynamic properties of an intrinsically disordered region of a germ cell-specific protein on phase separation. *Proc Natl Acad Sci U S A* **114**, E8194–E8203 (2017).

205. Wong, L. E., Kim, T. H., Muhandiram, D. R., Forman-Kay, J. D. & Kay, L. E. NMR Experiments for Studies of Dilute and Condensed Protein Phases: Application to the Phase-Separating Protein CAPRIN1. *J Am Chem Soc* **142**, 2471–2489 (2020).

206. Kim, T. H. *et al.* Interaction hot spots for phase separation revealed by NMR studies of a CAPRIN1 condensed phase. *Proc Natl Acad Sci U S A* **118**, (2021).

207. Bracha, D. *et al.* Mapping Local and Global Liquid Phase Behavior in Living Cells Using Photo-Oligomerizable Seeds. *Cell* **175**, 1467-1480.e13 (2018).

208. Schwartz, J. C., Wang, X., Podell, E. R. & Cech, T. R. RNA Seeds Higher-Order Assembly of FUS Protein. *Cell Rep* **5**, 918–925 (2013).

209. Powers, S. K. *et al.* Nucleo-cytoplasmic Partitioning of ARF Proteins Controls Auxin Responses in Arabidopsis thaliana. *Mol Cell* **76**, 177-190.e5 (2019).

210. Tang, H. *et al.* Systematic Unraveling of the Unsolved Pathway of Nicotine Degradation in Pseudomonas. *PLoS Genet* **9**, (2013).

211. Tararina, M. A. *et al.* Fast Kinetics Reveals Rate-Limiting Oxidation and the Role of the Aromatic Cage in the Mechanism of the Nicotine-Degrading Enzyme NicA2. *Biochemistry* **60**, 259–273 (2021).

212. Kallupi, M., Xue, S., Zhou, B., Janda, K. D. & George, O. An enzymatic approach reverses nicotine dependence, decreases compulsive-like intake, and prevents relapse. *Sci Adv* **4**, (2018).

213. Xue, S. *et al.* An enzymatic advance in nicotine cessation therapy. *Chemical Communications* **54**, 1686–1689 (2018).

214. Pentel, P. R. *et al.* The nicotine-degrading enzyme NicA2 reduces nicotine levels in blood, nicotine distribution to brain, and nicotine discrimination and reinforcement in rats. *BMC Biotechnol* **18**, (2018).

215. Thisted, T. *et al.* Optimization of a nicotine degrading enzyme for potential use in treatment of nicotine addiction. *BMC Biotechnol* **19**, (2019).

216. Dulchavsky, M., Clark, C. T., Bardwell, J. C. A. & Stull, F. A cytochrome c is the natural electron acceptor for nicotine oxidoreductase. *Nat Chem Biol* **17**, 344–350 (2021).

217. Baron, R. *et al.* Multiple pathways guide oxygen diffusion into flavoenzyme active sites. *Proc Natl Acad Sci U S A* **106**, 10603–10608 (2009).

218. Eisenmesser, E. Z., Bosco, D. A., Akke, M. & Kern, D. Enzyme dynamics during catalysis. *Science (1979)* **295**, 1520–1523 (2002).

219. Gronenborn, A. M. Small, but powerful and attractive: 19F in biomolecular NMR. *Structure* vol. 30 6–14 Preprint at https://doi.org/10.1016/j.str.2021.09.009 (2022).

220. Groitl, B. *et al.* Protein unfolding as a switch from self-recognition to high-affinity client binding. *Nat Commun* **7**, (2016).

221. Picard, L. P. & Prosser, R. S. Advances in the study of GPCRs by 19F NMR. *Current Opinion in Structural Biology* vol. 69 169–176 Preprint at https://doi.org/10.1016/j.sbi.2021.05.001 (2021).

222. Guo, J., Melançon, C. E., Lee, H. S., Groff, D. & Schultz, P. G. Evolution of Amber Suppressor tRNAs for Efficient Bacterial Production of Proteins Containing Nonnatural Amino Acids. *Angewandte Chemie* **121**, 9312–9315 (2009).

223. Marius Clore, G. & Iwahara, J. Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes. *Chem Rev* **109**, 4108–4139 (2009).

224. Jurcik, A. *et al.* CAVER Analyst 2.0: Analysis and visualization of channels and tunnels in protein structures and molecular dynamics trajectories. *Bioinformatics* **34**, 3586–3588 (2018).

225. Hughes, T. S., Wilson, H. D., De Vera, I. M. S. & Kojetin, D. J. Deconvolution of complex 1D NMR spectra using objective model selection. *PLoS One* **10**, (2015).