Spring 2021

Structural and Binding Studies of the Polar Organizing Protein Z (PopZ) Using NMR Spectroscopy

Christopher Taylor Nordyke
University of New Hampshire, Durham

Structural and Binding Studies of the Polar Organizing Protein Z (PopZ) Using NMR Spectroscopy

By

Christopher Taylor Nordyke

Bachelor of Science in Chemistry (ACS Approved), University of Wyoming, 2014

DISSERTATION

Submitted to the University of New Hampshire

in Partial Fulfillment of

the Requirements for the Degree of

Doctor of Philosophy

in

Biochemistry

May, 2021

This dissertation was examined and approved in partial fulfillment of the requirements for the degree of Doctor in Philosophy in Biochemistry by:

Dissertation Director, Dr. Krisztina Varga, Associate Professor, Department of Molecular, Cellular, and Biomedical Sciences

Dr. Xuanmao Chen, Assistant Professor, Department of Molecular, Cellular, and Biomedical Sciences

Dr. Rick Cote, Professor, Department of Molecular, Cellular, and Biomedical Sciences

Dr. W. Kelley Thomas, Professor, Department of Molecular, Cellular, and Biomedical Sciences

Dr. John Tsavalas, Associate Professor, Department of Chemistry

On April 9, 2021

Approval signatures are on file with the University of New Hampshire Graduate School.

TABLE OF CONTENTS

DEDICATION

I would like to dedicate this dissertation to my close friends and family. I would like to thank my parents, Kirk and Lisa, and my sisters, Kassie and Abbie, for their love and support throughout my graduate career. I would like to thank Jillian for her love and support during my time at the University of New Hampshire. Finally, I would like to thank my friends, Ben, Ian, Jen, Jenny, and Jon for all their friendship and the good times we've had that helped keep me sane.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Dr. Krisztina Varga, for her guidance and support during my time as both an undergraduate and graduate student. She has been exceptionally patient, understanding, and kind at every step of my graduate education, and none of this would have been possible without her. I would also like to thank my committee members, Dr. Xuanmao Chen, Dr. Rick Cote, Dr. Thomas Kelley, and Dr. John Tsavalas, for their participation and generosity with their time.

I would like to thank former Varga group member Dr. K. Wade Elliott for his friendship, input, and intellectual discussions surrounding the project. I would like to thank Mr. Jonathan Sreter, Ms. Katarina Jovic, Mr. Logan Brown, and Mr. Ryan Puterbaugh for their assistance and camaraderie in the lab. I would also like to thank Dr. Patricia Wilkinson for her support, especially for all her help setting up the new NMR instrumentation. Additionally, I would like to thank Dr. James Aramini at the City University of New York for data collection and Dr. Woonghee Lee from the University of Colorado Denver for his consultation.

Finally, I would like to thank our collaborator, Dr. Grant Bowman, from the University of Wyoming. This research would not have been possible without the assistance or knowledge he provided throughout the completion of this project.

This research was supported in part by grants from the National Institute of General Medical Sciences (1R01GM118792) at the National Institutes of Health, the National Science Foundation Major Research Instrumentation Award (DBI-1828319), and the University of New Hampshire startup funds.

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

The polar organizing protein Z (PopZ) is an intrinsically disordered protein from *Caulobacter crescentus* that is necessary for the formation of three-dimensional microdomains at the cell poles, where it functions as a hub protein that recruits multiple regulatory proteins. Although a large portion of the protein is predicted to be disordered, PopZ can self-assemble into polymeric superstructure scaffolds that directly bind to at least ten different proteins. Here, we report the solution NMR structure of $PopZ^{\Delta 134–177}$, a truncated variant of PopZ that does not self-assemble but retains the ability to interact with heterologous proteins. We show that the unbound form of $PopZ^{\Delta 134–177}$ is unstructured in solution, with the exception of a small amphipathic α-helix encompassing residues M10-I17, which is included within a highly conserved region near the N-terminus. In applying NMR techniques to map the interactions between $PopZ^{\Delta 134–177}$ and one of its binding partners, RcdA, we find that the α-helix and neighboring residues extending to position E23 serve as the core of the binding motif. Consistent with this, a point mutation at position I17 perturbs the binding region and severely inhibits interaction with RcdA. Our results show that a partially structured Molecular Recognition Feature (MoRF) within an intrinsically disordered domain of PopZ contributes to the assembly of polar microdomains, revealing a structural basis for complex network assembly in Alphaproteobacteria that is analogous to those formed by intrinsically disordered hub proteins in other kingdoms.

PROJECT AIMS AND HYPOTHESIS

Recently, protein science has been transformed by investigations into intrinsically disordered proteins (IDPs), which are characterized by a prominent lack of ordered tertiary structures. This is in stark contrast to the classical protein paradigm stating that a given protein has a well-defined, tertiary structure and an associated function. However, due to their structural disorder, IDPs have remarkable flexibility in their functional states. They are responsible for a variety of cellular processes, many of which center around signaling, regulation, and control. As a result, defects in IDPs can lead to detrimental effects in both prokaryotic and eukaryotic organisms. Due to this, IDPs have also garnered interest in medicine, particularly because they account for 79% of all proteins associated with human cancer.

Given their biological relevance, we have aimed to structurally and functionally characterize an IDP called polar organizing protein Z (PopZ). PopZ has a vital role in cellular organization where it forms polymeric scaffolds at both cell poles of the bacterium *Caulobacter crescentus* and is required for the recruitment of regulatory proteins from the cytoplasm. The bacterial cytoplasm contains a dense packing of many different proteins and macromolecules. *Escherichia coli*, for example, has an estimated two million soluble protein molecules in its cytoplasm[1]. Surprisingly, the cytoplasm is not a chaotic mixture, but instead highly organized by the formation of complex macromolecular assemblies despite the notable lack of cytoskeletal motors for direct transport and internal membranes for compartmentalization. As a result, organization is highly important, and the ability of PopZ to bind to a wide variety of partners to maintain organization in *C. crescentus* is mediated by its lack of a well-defined tertiary fold.

The results from this project are likely to have broad relevance in bacterial organization, the field of IDPs, how IDPs function despite lacking an ordered structure, and may hold relevance in medical applications. The goals of this project are reflected in the following project aims:

Aim 1: Structural characterization of PopZ by solution NMR spectroscopy. *Working hypothesis: PopZ adopts a structure of flexible disorder with a short helix on both the N- and C-termini. This was the predicted structure of PopZ based on protein information.* We have utilized NMR spectroscopy to probe chemical shifts and supplemented the data with distance and orientation constraints to define the protein structure.

Aim 2: Characterization of the interaction of PopZ with one of its binding partners, RcdA. *Working hypothesis: PopZ adopts a disorder-to-order shift upon binding with protein partners and binding occurs near the sites of proposed helicity at the N- and C-termini.* We have utilized NMR spectroscopy to probe the PopZ binding domain by examining PopZ spectral changes when complexed with a protein binding partner.

CHAPTER 1

INTRODUCTION

This project focused on the characterization of an intrinsically disordered protein (IDP) called Polar Organizing Protein Z (PopZ), a recruitment protein that localizes to the cell poles of the bacterium *Caulobacter crescentus*. PopZ self-associates into polymeric superstructures and organizes other proteins, many of which are necessary for cellular division, to the poles. Its ability to bind with multiple proteins is a hallmark characteristic of IDPs, a protein class named as such for their extreme structural flexibility. This flexibility is distinct from other types of proteins, which have much more static conformations and much fewer binding partners. The aim of our project was to determine the structure of PopZ, identify its potential binding partners, and characterize the binding site of PopZ with these partners. In this introduction, I will first describe *C. crescentus* **(section 1.1)**, followed by a background on PopZ **(section 1.2)**, and then conclude with a discussion on the IDP protein class **(section 1.3)**.

1.1. *Caulobacter crescentus*

Complex structures can be found at bacterial cell poles. In many species, cell poles are sites for the assembly of flagella, pili, and stalks[2-5], and they can also serve as locations for multiprotein complexes that regulate chromosome replication and the directionality of chromosome segregation[6,7]. A well-characterized example of complex cell pole

organization is *Caulobacter crescentus*, in which at least 100 different proteins are localized to one or both cell poles over the course of the cell cycle[8]. *C. crescentus* is a Gram-negative, *Alphaproteobacterium* commonly found in lakes, rivers, and oceans[9]. It has been coined a model organism for investigations surrounding asymmetric cell division in bacteria, where two distinct progenies are formed following this event. The first is a swarmer cell that searches for resources in its surrounding environment. Swarmers are characterized by the presence of a flagellum and multiple pili at a single pole and are inherently unable to reproduce until their transition into the second progeny: stalked, sessile cells responsible for reproduction[3,9-11]. Stalked cells contain a polar stalk tipped with a polysaccharide holdfast organelle needed for attachment to surfaces[11,12], effectively immobilizing the bacterium even under high shear conditions[13]. This surface adhesion is essential for the formation of biofilm, a multicellular community that confers many advantages to the population that can include antibiotic resistance, nutrient absorption, and gene transfer. In *C. crescentus*, adhesion is permanent and tightly controlled[12].

Asymmetric division is highly regulated in *C. crescentus*. It begins with the initiation of chromosome replication followed by the formation of a multi-protein complex called a divisome responsible for cytokinesis control[9]. Figure 1 is a representation of the *Caulobacter* life cycle and shows the spatial and timing distribution of many proteins responsible for cell division. The first protein to assemble in this complex is the filamenting temperature-sensitive mutant Z (FtsZ), a GTPase that forms a Z-ring at midcell and serves as the divisome scaffold by self-assembling into long filaments[9,14]. FtsZ is anchored to the cell membrane through anchor proteins and is thought to provide the

constrictive force necessary for cell division[14]. *C. crescentus* uses a variety of master regulators, such as CtrA, GcrA, DnaA, CcrM, and SciP, to make sure cell division occurs at the correct time relative to other cell cycle processes[9,15-18]. One of these regulators, the aptly named DnaA, is responsible for the initiation of DNA replication[16,19]. The ParB partitioning protein is bound to the chromosomal centromere located at the stalked pole beforehand, and after DNA duplication, segregates to the opposite pole with one of the two chromosomes in tow[19].

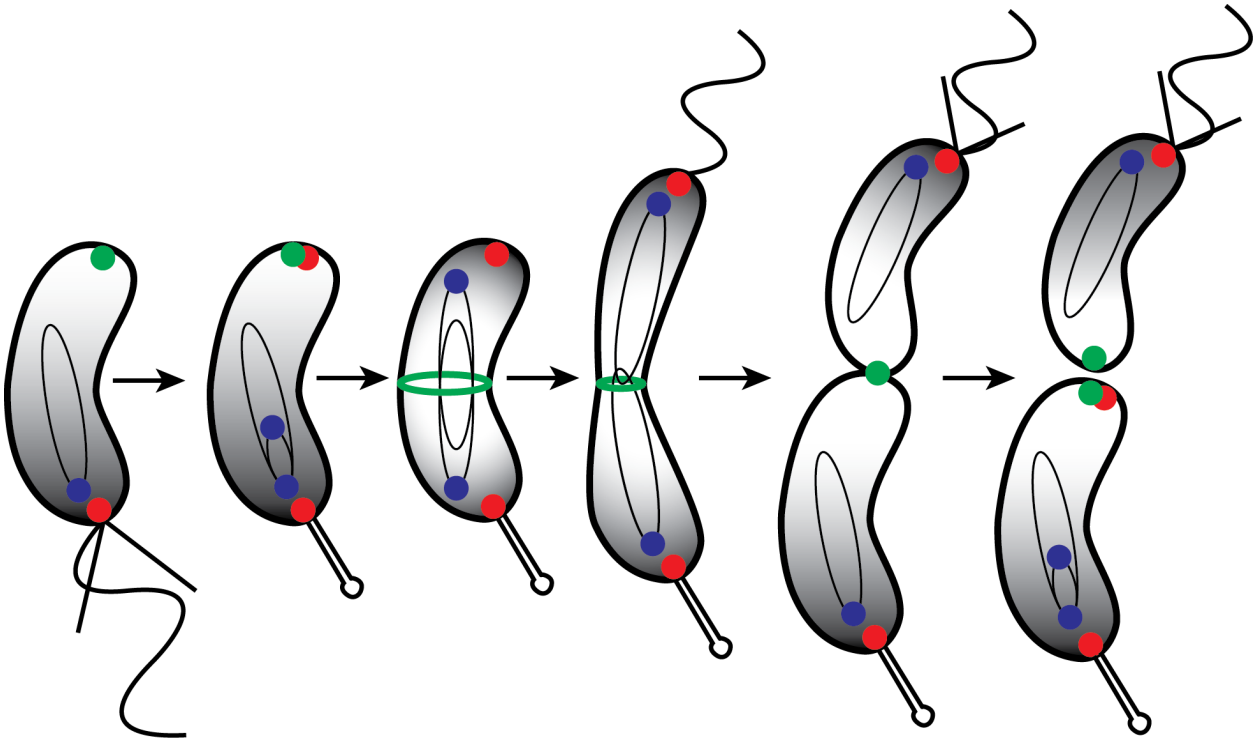


Figure 1. *Caulobacter crescentus* life cycle. Asymmetric division of *C. crescentus* results in a swarmer cell and a stalked cell. Colored regions represent essential proteins necessary for successful division. ParB (blue) binds to the chromosome origin and chaperones the chromosome to the cell pole where it is received by PopZ (red). FtsZ (green) forms a Z-ring at midcell and constricts to divide the cell into two distinct progenies. The timing of FtsZ is controlled by MipZ which interacts with ParB and forms a concentration gradient in the cell. MipZ concentration is shown by the grayscale gradient in the cell schematics, where dark is more concentrated and light is less concentrated.

The amount of coordination associated with the cell cycle of *C. crescentus* and other bacteria is indicative of a necessity for high organization within the cell. How this level of organization is achieved has been the subject of increasing interest, especially given the fact that bacteria tend to be lacking in higher functioning organelles traditionally associated with organization, such as cytoskeletal motors and internal membranes for compartmentalization[20]. One protein responsible for cellular organization in *C. crescentus* is the polar organizing protein Z.

1.2. Polar Organizing Protein Z (PopZ)

Polarity is important across many bacterial species, with polar proteins functioning in almost all cellular processes[19], including growth, division, motility, signal transduction, and cell cycle regulation[21]. This polarity is achieved through a variety of different classes of polar organizers which include both nonpolymer-forming landmark proteins and self-assembling polymer-forming landmark proteins[21], the latter of which includes PopZ. PopZ is an intrinsically disordered protein (IDP) found in the bacterium *C. crescentus*. Like many other IDPs, PopZ has been shown to bind with a variety of protein binding partners, many of which are crucial for bacterial cell division. PopZ acts as a bacterial hub protein and, as its name suggests, organizes binding partners at the bacterial cell poles. Organization by PopZ is achieved through the formation of large, polymeric scaffold superstructures around which the PopZ binding partners are organized. These scaffolds are found in liquid-liquid phase droplets, which are protein condensates with a denser population relative to the rest of the cytosol. Assembly of the scaffold itself is suggested to be hierarchal (Figure 2), where monomeric subunits interact to form trimeric rods, which

then dimerize laterally to form hexamers, and completed with the hexamers stacking end to end to form long filaments that accumulate at cell poles[22]. The C-terminal region (residues 134-177) is highly conserved and essential for this higher-order assembly. Mutation studies in this region showed both inhibited oligomerization and localization to cell poles, suggesting PopZ polar accumulation is dependent on higher-order assembly[22].

While PopZ is necessary for other protein localization, how PopZ itself localizes to cell poles remains unclear. Expression in *Escherichia coli* suggests the ability to localize is intrinsic to PopZ, considering *E. coli* is evolutionarily distant from *C. crescentus* and produces no homologous proteins[22-24]. Some polar organizers have been shown to localize to cell poles based on the negative curvature of the membrane, as seen with DivIVA from *Bacillus subtilus*[25], but it is largely unclear if PopZ does the same. Alternatively, being a cytoplasmic protein, it has been suggested that PopZ is anchored to the membrane by other proteins[19]. It is possible that when PopZ complexes with these transmembrane proteins they localize to membrane curvature at cells poles based on the intrinsic shape of the protein complex itself, as observed with TlpA of *B. subtilis*[19,26]. Most PopZ homologues are localized in a bipolar manner, with the exception of some homologues that show monopolar localization of PopZ at the new pole, such as from *Agrobacterium tumefaciens*, a plant pathogen, and *Brucella abortus*, a human and animal pathogen[19,27]. The reason for the monopolar localization in these two species remains unsettled.

Proteins that have been shown to interact with PopZ and are associated with its spatial location include ZitP and SpmX, both of which have membrane-spanning domains[28,29], and could serve as PopZ anchorage sites. There also appears to be

redundant modes of localization for PopZ. PopZ mutants lacking the N-terminal domain require ZitP for polar localization in *C. crescentus*, but wildtype (WT) PopZ will still localize in the absence of ZitP[19,28]. Additionally, ZitP requires PopZ for localization to the stalked pole which would suggest that polar localization is somewhat mutually dependent between some proteins[19,30]. Another protein associated with PopZ localization is MreB, whose deletion causes delocalization of PopZ in both *C. crescentus* and *E. coli*[24].

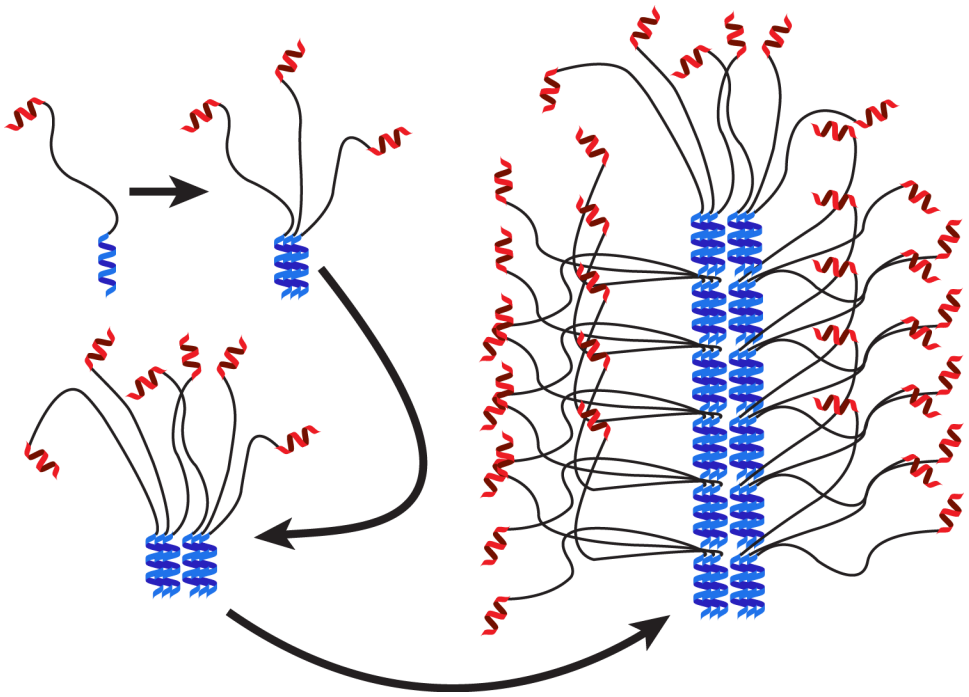


Figure 2. Schematic of the oligomerization of PopZ. PopZ monomers self-assemble into trimers that interact via the C-terminal domain. Trimers assemble into a hexamer by forming dimers. Hexamers interact with one another to form long filaments that localize at *C. crescentus* cell poles *in vivo*.

In the cell cycle of *C. crescentus*, PopZ is first localized to the old pole prior to cell division. Upon initiation of DNA replication in early S phase, a new PopZ foci at the

opposite pole is formed to bind the ParB partitioning protein, which is bound to the newly duplicated DNA, as it arrives to the pole[22]. Other proteins that rely on direct association with PopZ for polar recruitment were determined using *E. coli* co-expression assays, which include ParA, CpdR, ChpT, CckA, RcdA, and DivL[20]. Some proteins have an indirect dependency on PopZ for polar localization. For example, MipZ ATPase binds to ParB at the cell poles following chromosome replication and inhibits the FtsZ divisome formation through a concentration gradient[9,31]. The MipZ ATPase concentration is highest at the cell poles and lowest at the cell center, thereby restricting FtsZ to the cell center for division[9,31,32]. Deletion of PopZ disrupts the MipZ gradient and results in abnormal cell division[22]. Other proteins with an indirect dependence on PopZ include MopJ and DivJ[20,33,34]. Some proteins exhibit a weak interaction that results in only partial co-localization at cell poles, as seen with DnaK, TipN, PopA, PleD, and PdeA[20]. It is also worth noting that many of the binding partners exhibit dependence on PopZ for localization to only one pole, as seen for ZitP and PopA which require PopZ for localization to the old pole but rely on other factors for localization to the new pole[30,35]. Regardless, the wide range of direct and indirect protein partners exhibits the level of importance PopZ has in bacterial organization.

Previously, PopZ has been predicted to have helical character at both the N- and C-termini separated by a long region of intrinsic disorder[22]. The N-terminus has a highly conserved stretch of residues necessary for helix formation, interaction with partner proteins, and control of PopZ polar distribution. The C-terminus, required for PopZ accumulation and oligomerization, is also highly conserved and predicted to contain three helices[22]. The linker region between the termini is highly disordered and rich in negatively

charged residues and proline residues[22]. It is necessary for PopZ function, but can be shortened to half its length and scrambled while still maintaining function[36]. The disordered nature of PopZ has led to some struggle in determining its structure.

For this project, we used a truncated version of PopZ (PopZ$^{\Delta 134\text{-}177}$) that excluded the C-terminal oligomerization domain (residues 134-177). Thus, PopZ$^{\Delta 134\text{-}177}$ remained soluble in solution. This truncated version was necessary for studies utilizing solution NMR spectroscopy because the polymeric superstructures formed by full length PopZ would be too large to observe using this technique. We have previously shown that this

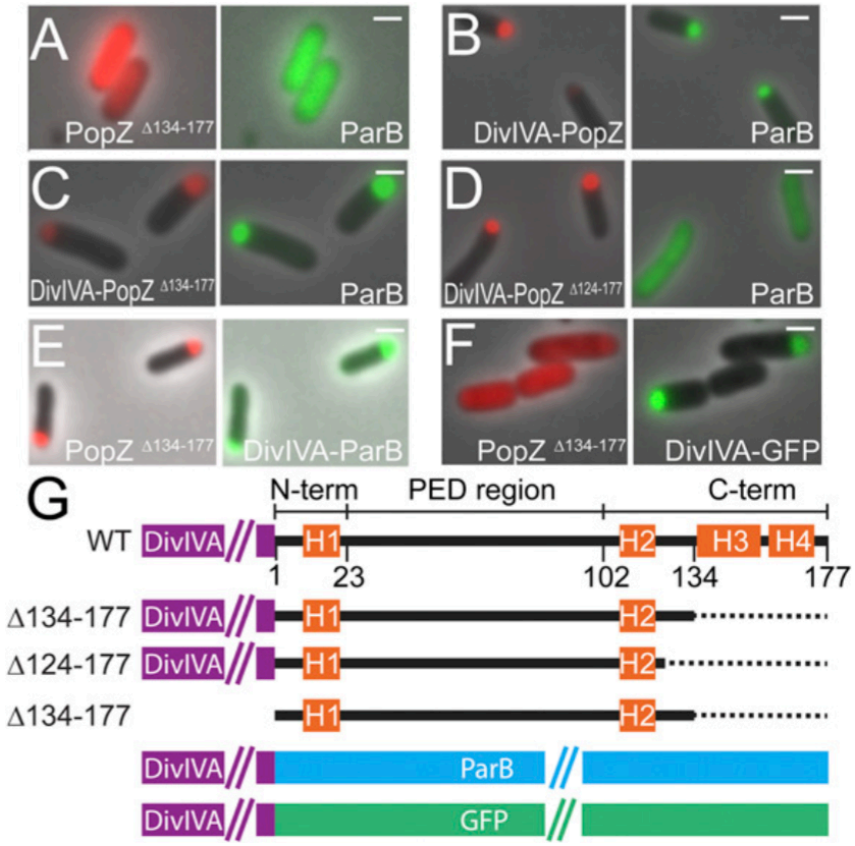


Figure 3. PopZ binding in *E. coli*. (A-F) Localization of mCherry-PopZ fusion variants (red, left) and ParB-GFP (green, right) co-expressed in *E. coli*. Scale bars are 1 μm. (G) Diagrams of the fusion proteins in A-F. Figure adapted from Holmes, *et. al*.

truncated variant is sufficient for interacting with multiple heterologous binding partners, even in the absence of homo-oligomeric assembly (Figure 3)[36].

1.3. Intrinsically Disordered Proteins (IDPs)

1.3.1. What are intrinsically disordered proteins?

Ordered proteins retain their stability due to the presence of van der Waals forces and hydrogen bonding which promote the formation of structural motifs such as α-helices and β-sheets. Loss of ordered structure almost unanimously results in loss of function. Alternatively, intrinsically disordered proteins are characterized by a lack of secondary and/or tertiary structure under native physiological conditions despite having biological activity[37]. It has been previously accepted that protein function is dependent on that protein's structure, but the existence of IDPs shows that a lack of structure can be important for vital cellular function as well. This new protein class was first identified in the 1950s with the discovery of specific IDPs including casein, phosvitin, fibrinogen, trypsinogen, and calcineurin[38]. The amount of disorder in proteomes has been extensively analyzed. Disorder is most prevalent in eukaryotes, with 20.5% averaged disorder content for eukaryotic proteomes, followed by 13.2% for viruses (considered a domain of life for the purposes of the study), 8.5% for bacteria, and 3.6% for archaea[39].

Biological systems are inherently complex, often involving multistep biochemical pathways and multiple levels of regulation among a large number of physically and functionally interacting components. In many cases, interactive networks are highly dynamic, changing rapidly in response to regulatory signals and adjusting their connectivity to accommodate a range of activities. Interestingly, there is a positive

correlation between higher levels of complexity and higher levels of structural disorder in networking components[40,41]. At the core of these networks are IDPs and their intrinsically disordered domains (IDDs) that have little or no native structure outside of their interactions with other proteins. Moreover, the disordered nature of the protein binding interface is inherent to the function of dynamic, multi-partner binding interfaces[42], also known as interaction hubs.

1.3.2. Amino acid composition and "fuzziness" of intrinsically disordered proteins

IDPs lack structural motifs such that they have a constantly fluctuating tertiary structure, thus they exist as a structurally random ensemble in solution. The tendency towards order or disorder in a protein is heavily influenced by amino acid composition. Proline residues have sidechains that loop and connect to the backbone nitrogen (Figure 4), thus removing a degree of freedom experienced by other residues. As such, they are unable to form many common secondary structure elements. For this reason, proline residues are famously known as helix-breakers and have the greatest propensity towards disorder. Figure 4E shows the typical Ramachandran plot for amino acids, which displays the favorable conformations of phi (Φ) and psi (ψ) torsion angles. The proline Ramachandran plot is much more conservative compared to the other residues. Additional residues that promote disorder are typically hydrophilic and charged, including P, E, S, Q, and K. Order-promoting residues tend to be hydrophobic[38,43] and include C, W, I, Y, F, L, M, H, and N.

Following proline, the second most common residue associated with disorder is glutamate[44]. Accordingly, PopZ contains a proline-glutamate rich domain (PED) at the

center of the sequence. The variant discussed in this dissertation was comprised of 25 proline and 20 glutamate residues, which accounted for 18% and 14% of the protein sequence, respectively.

In some cases, IDP binding partners have been shown to induce structured domains in IDPs, or conversely induce disorder in structured proteins. However, this is not always the case. An IDP in its bound state can still adopt a range of structures (polymorphic) or remain unstructured to varying degrees and be segmentally or fully disordered across the sequence. This variance has been described as "fuzziness"[45], of which there are four main categories (Figure 5). The first category describes the ability of the protein to adopt multiple conformers, previously mentioned as polymorphic. The second is the extreme end of fuzziness called random complexes where binding does not induce any ordering in the interacting regions. Between polymorphic and random complexes are the categories of flanking and clamp complexes, where disordered segments either neighbor or connect ordered binding regions[46]. Therefore, this fuzziness leads to a range of potential conformations across a dimpled binding energy landscape with no clear minimum, leading to appropriately named "fuzzy complexes"[47].

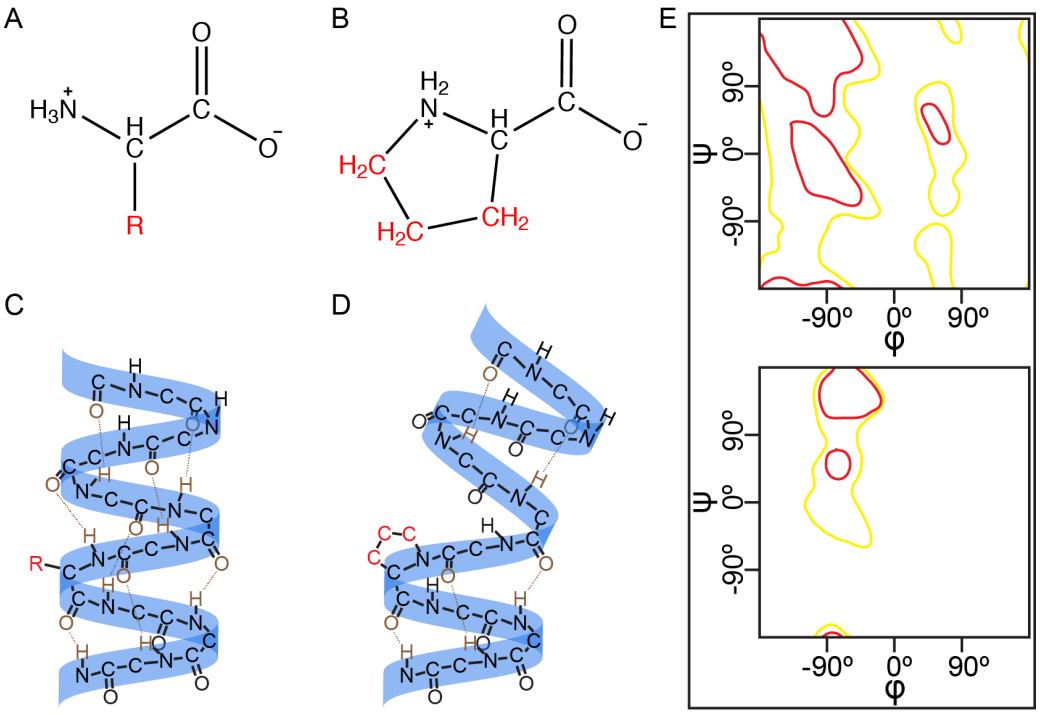


Figure 4. Proline residues. Proline residues are unique among amino acids. (A) Amino acid representation in stick form. The R group that distinguishes different amino acids apart is highlighted red. (B) Proline amino acid in stick form. The proline R group is highlighted red. (C) Schematic of an α-helix. Brown atoms and bonds represent hydrogen bonding interactions. A representative R-group (red) is shown on one amino acid. (D) The R-group is replaced with a proline side chain showing the effects a proline residue has on a helix. The missing hydrogen bond causes a kink to occur in the helix, leading to the proline "helix-breaker" namesake. In (C) and (D), most R groups and the hydrogen atoms bonded to carbons are excluded for simplicity. (E) Ramachandran plots showing the favorable (red lines) and allowed (yellow lines) Φ and ψ dihedral angles for the majority of amino acids (top) and for proline (bottom). Proline residues have a much more restricted range of possible dihedral angles.

1.3.3. Binding mechanism of intrinsically disordered proteins

Interestingly, due to their flexibility, the majority of IDPs have a multitude of binding partners, whereas ordered proteins have much fewer partners[38,48]. The mechanism for multi-partner interaction has two proposals: (i) one-to-many and (ii) many-to-one binding. In the one-to-many mechanism, an IDP has either a single binding region that can bind to multiple partners or have multiple binding sites that bind to different partners. For

example, CD28 is a phosphorylated IDP that binds to four different partners: both the PI3K p85 N-terminal and C-terminal SH2 domains, the Gads SH2 domain, and the Grb2 SH2 domain, all of which bind to the same site on the CD28[37,49,50]. Despite high structural similarity between the four partner proteins, the CD28 binding site recognizes the partners differently, where it adopts a bent conformation with Gads and Grb2 and an extended conformation with the PI3K p85 N- and C-terminal domains[49,50]. Alternatively, in the many-to-one binding mechanism, a protein can be targeted by multiple IDPs or intrinsically disordered regions (IDRs). The binding sites on the partner can be the same, overlapping, or separate, and the IDRs can be from the same IDP[38]. A well-known IDP that has been shown to have multiple modes of binding is p53[38].

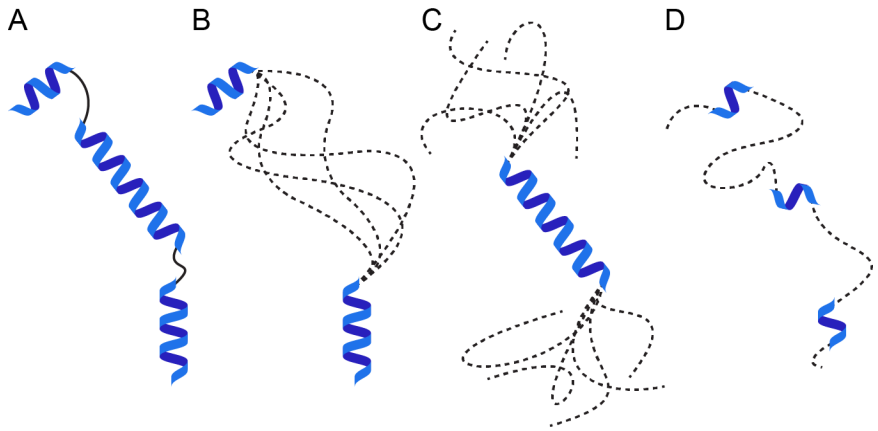


Figure 5. Categories of fuzziness. Dotted lines represent regions of disorder. (A) Polymorphic model: the binding IDP adopts a few or many conformations. (B) Clamp model: ordered binding domains are connected by nonbinding disordered segments. (C) Flanking model: an ordered binding domain has nonbinding disordered domains on either side. (D) Random model: the entire IDP is disordered with the exception of possible transient secondary structure characteristics.

IDD binding sites are referred to as molecular recognition features (MoRFs), and are typically 10-50 amino acids in length[51]. In general, MoRFs adopt a disorder to order

transition upon binding with a partner[52,53], otherwise known as coupled folding-binding[52], but the mechanism behind this is not well understood[37]. The extent of disorder before binding varies heavily across IDPs, wherein some have an inherent or residual structure before binding while others have complete disorder[53]. Binding mechanisms of inherent structure have the MoRF fluctuating through various conformational states with a bias towards nonrandom, preformed secondary structure that is then stabilized in the bound conformation[53,54], a process known as conformational selection[54]. Alternatively, some mechanisms involve MoRFs with initially complete disorder adopting an induced secondary structure upon binding[53]. For any given IDP binding event, it is likely a combination of induced fit and conformational selection that controls binding, but the degree of each is determined by protein concentrations, association rate, and conformation transition kinetics[52]. p53 is a prominent example of an IDD hub protein, which physically interacts with numerous binding partners via IDDs with MoRFs located in the N-and C-terminal sections of the protein[55-57]. There are a number of other examples of IDD hubs in eukaryotic systems[58-60], and their inherent networking capability may have been important in supporting the expansion of organismal complexity in this kingdom[40,61].

Structured MoRF domains can be categorized into three classes: α-helices, β-sheets, and irregular secondary structure[52,53], or a combination thereof[53]. The existence of α-helices and β-strands are predicted at frequencies of approximately 20% and 5%, respectively[62]. Commonly, binding between the MoRF of an IDP and its partner occurs in a singular, short segment of the protein, which is in stark contrast to ordered proteins which tend to bind partners with several domains at once and rarely ever with just one[63].

1.3.4. Biological functions of intrinsically disordered proteins

IDPs carry out a variety of cellular functions, most of which are some form of signaling, regulation, and control. These functions include biomolecular recognition, post-translational modification, ligand binding, alternative splicing, transportation through pores and channels, and cellular organization[38,43]. IDPs are implicated in the formation of membraneless organelles such as Cajal bodies, nucleoli, processing bodies, and germ line granules. The mechanism of formation by IDPs is largely unknown, but these types of organelles have a high prevalence of IDPs. P granules from *Caenorhabditis elegans*, for example, are enriched with RNA-binding proteins such as RNA helicase LAF-1 which has a long disordered segment of arginine and glycine residues necessary for both RNA-binding and the formation of the granules *in vitro*[64,65]. Well-known IDPs are the tumor suppressors FOXO3a and p53, both of which regulate cell cycle and apoptotic genes. They have similar binding partners, despite binding through different modes. For example, both bind to the KIX domain of CREB binding protein (CBP) and p300 at the MLL and c-Myb binding sites[66]. FOXO3a binds to KIX via specific orientations of its CR2C and CR3 domains[66]. Alternatively, p53 binds to the same KIX sites with its AD1 and AD2 domains, but with much more freedom in the domain orientations[67]. Chaperone activity has also been implicated with IDPs. Hsp33 is an oxidative stress-activated chaperone that undergoes an order-to-disorder transition upon activation. Under oxidative conditions, two disulfide bonds are formed that cause unfolding of Hsp33 which is necessary for the oxidative protection of hundreds of proteins in bacterial cytosol[68,69]. Similarly, the HdeA chaperone undergoes a partial order-to-disorder transition under acidic conditions where it protects other denatured proteins from aggregation. Upon a

return to neutral pH, HdeA releases its binding substrates slowly, which is thought to minimize the formation of aggregation-prone intermediates and promote passive refolding of the substrate to its natural state[68,70].

Research into IDPs has direct implications in the use of these proteins as therapeutic drug targets for cancer due to the fact that 79% of cancer-associated proteins are classified as IDPs[71]. IDP binding sites tend to be relatively weak, so disrupting the binding interface between an IDP and its partner with small molecules can be achieved[64,72]. Interaction can occur through orthosteric binding, where the small molecule directly blocks the IDP binding site, or allosteric binding, where the small molecule binds at an inactive site on the IDP and prevents the IDP from reaching the active conformation. In most cases, small molecule binding takes advantage of hydrophobic residues[72]. For example, AF9 and AF4 interact with one another and are associated with leukemia. A peptide containing the AF4 recognition site disrupts the AF9-AF4 complex by binding directly with AF9, thereby inhibiting the leukemia cell proliferation[72]. In another case, tyrosine phosphatase 1B is associated with breast cancer and can be functionally inactivated when allosterically bound by trodusquemine at the disordered C-terminus[72].

CHAPTER 2

EXPRESSION AND PURIFICATION OF PROTEINS POPZ$^{\Delta 134\text{-}177}$ AND RCDA

This project required production of primarily two proteins: PopZ$^{\Delta 134\text{-}177}$ and its binding partner, RcdA. PopZ$^{\Delta 134\text{-}177}$ was used in both structural and binding studies, while RcdA was used only in binding studies. All proteins needed to be prepared in a way that was conducive towards both NMR experiments and protein stability. However, RcdA had much more stringent buffer requirements than PopZ$^{\Delta 134\text{-}177}$, so NMR binding studies were tailored to conditions favored by the binding partner (buffer pH 7.5). PopZ$^{\Delta 134\text{-}177}$ had much looser buffer requirements which allowed buffer conditions to be tailored towards experimental requirements for structural studies (buffer pH 5.5).

The expression and purification of PopZ$^{\Delta 134\text{-}177}$ and RcdA were performed following the protocols published in the literature[20]. However, while expression of PopZ$^{\Delta 134\text{-}177}$ did occur, overexpression of the protein was not apparent, therefore an optimization of the PopZ$^{\Delta 134\text{-}177}$ expression was performed. Overexpression and high purity are important for the production of protein sample required for solution NMR experiments (milligram quantities) due to the fact that NMR spectroscopy is an inherently insensitive technique, thus large quantities of protein sample are required to obtain high quality signal. Additionally, to observe PopZ$^{\Delta 134\text{-}177}$ using NMR spectroscopy, the protein must be

expressed with NMR-active isotopes, so the bacteria are grown in ^{15}N- and ^{13}C-enriched media.

This section will begin with a brief introduction into protein expression using bacterial systems (**section 2.1**). Next, a description of $PopZ^{\Delta 134\text{-}177}$ expression and purification (**sections 2.2** and **2.3**, respectively) is given followed by the expression, purification, and cleavage of the fusion Sumo-RcdA protein (**sections 2.4**, **2.5**, and **2.6**, respectively) to produce RcdA. An important distinction between the production of RcdA and $PopZ^{\Delta 134\text{-}177}$ is that RcdA is not isotopically enriched. Following this, preparation of the proteins for NMR studies (**section 2.7**) will be described before a brief description of the NMR experiments is provided (**section 2.8**). Finally, the section will conclude with a description of the methods used for a specific NMR experiment called residual dipolar coupling and the preparation of the RDC gel used in this experiment (**sections 2.9** and **2.10**, respectively).

2.1. Introduction to protein expression using bacterial systems

NMR studies require large quantities of protein, and the most efficient way to achieve this is through bacterial expression. The most common bacteria used are *E. coli* because they are easy to work with, have a high growth rate, have a large capacity for growing in different media[73], and the isotopic enrichment is affordable. As a result, there are a variety of laboratory strains that have been optimized for protein expression. Common strains include BL21 and Rosetta, the latter of which is a derivative of BL21 and was used for this project's PopZ variants. Rosetta cells are normally chosen to enhance expression of eukaryotic proteins, but the cell line also provides chloramphenicol resistance which we

used for this project. The production of a protein requires a plasmid encoding the desired genetic sequence, and typically the plasmid is engineered to allow the expression of the protein to take place in the presence of an inducer, such as isopropyl β-D-1-thiogalactopyranoside (IPTG). IPTG acts by binding to the lac repressor, allowing transcription of genes coding for the desired protein. Additionally, IPTG cannot be metabolized by the cell, so transcription does not stop, and overexpression occurs.

Furthermore, when protein expression is induced for NMR studies, it is done so in the presence of isotopically-labeled resources such as ^{13}C-glucose and $^{15}NH_4Cl$. NMR spectroscopy requires NMR-active isotopes, and in biomolecular NMR studies, these isotopes are typically ^{1}H, ^{13}C, and ^{15}N. However, only ^{1}H is naturally abundant in high quantities (~99.98%). ^{13}C and ^{15}N naturally exist at quantities of 1.1% and 0.37%, respectively. Therefore, isotopically-labeled resources ensure that the bacteria produce the protein with high abundance of NMR active isotopes. The most common labeling method is uniform labeling such that the produced protein has ≥95% of ^{13}C and/or ^{15}N isotopes. Less common methods include selective labeling (only specific amino acids are labeled)[74], stereo-array isotope labeling (SAIL; selected side chain positions are labeled)[74], and segmental isotopic labeling (only specific segments or domains of a protein are labeled)[75]. The PopZ$^{\Delta 134\text{-}177}$ induction time optimization is described below followed by the expression of PopZ$^{\Delta 134\text{-}177}$ using uniform labeling.

2.2. Expression of PopZ$^{\Delta 134\text{-}177}$

The induction time optimization was performed by growing Rosetta (DE3) *E. coli* (containing the PopZ$^{\Delta 134\text{-}177}$ expression vector) in Luria-Bertani (LB) broth media at 37 °C.

After the cells reached an optical density at 600 nm (OD_{600}) of ~0.5, they were centrifuged into pellets and resuspended into minimal media (MM). $PopZ^{\Delta 134\text{-}177}$ expression was induced with 0.8 mM IPTG at 37 °C. Aliquots of the bacterial solution were removed every two hours, pelleted, and stored at -80 °C. Pellets were lysed using lysozyme from chicken egg white (see protocol in Appendix) and the supernatant was separated from cellular debris using centrifugation. Following this, the supernatant components were separated using SDS-PAGE. Gels were visualized using Coomassie blue and Western blot staining (Figure 6). The bands representing $PopZ^{\Delta 134\text{-}177}$ were analyzed for their intensity, where the optimal induction time corresponded to the lanes with the most intense $PopZ^{\Delta 134\text{-}177}$ bands. The Western blot was imaged using a Li-Cor Odyssey CLx imager. The optimization indicated that a 12-hour induction course yielded the highest quantity of protein, and this time was used for all subsequent protein expressions.

For expression, wildtype $PopZ^{\Delta 134\text{-}177}$ was cloned into the *E. coli* expression vector pET28a (Novagen) with a 6-His HisTag and Leucine-Glutamate linker (see Table 1). Agar plugs containing *E. coli* transformed cells were surface-scraped with a pipette tip and dropped into 30 mL of LB supplemented with 30 µg/mL kanamycin, which was included to select for bacteria that retained the expression vector. The *E. coli* were shaken at 225 rpm and 37 °C overnight for bacterial growth. The next morning, *E. coli* freezer stocks were made by mixing together 0.75 mL of growth with 0.75 mL of 50% v/v glycerol in water for a final *E. coli* freezer stock in 25% glycerol. These were flash frozen in liquid nitrogen and stored at -80 °C. Freezer stocks were used for future growths in lieu of transformations as described below.

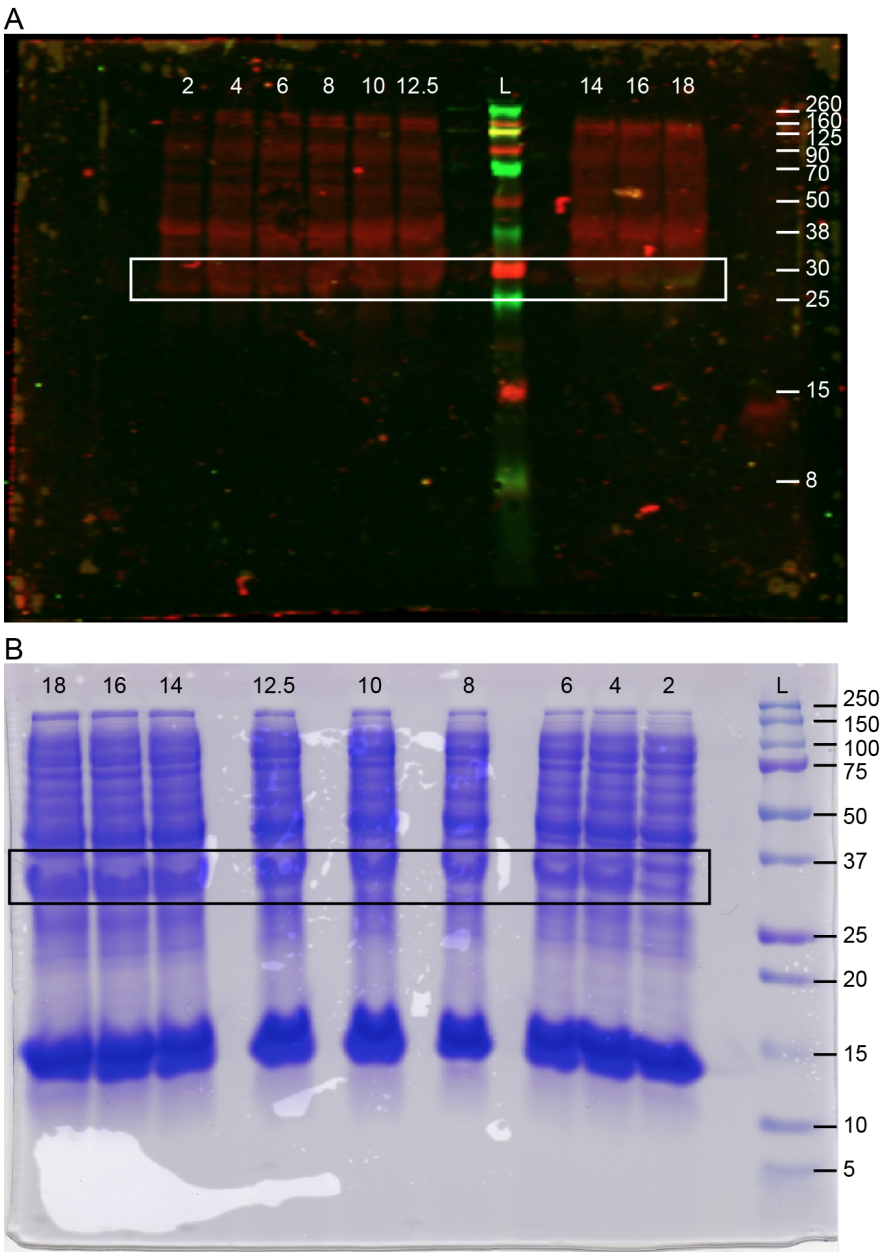


Figure 6. Western blot and Coomassie blue staining. Lanes are labeled at the times (hours) the aliquot was removed following induction with IPTG at 37 °C. The ladder standards are represented with an "L" and labeled with size markers (kDa). (A) Western Blot image. Thin bands (white box) at approximately 30 kDa correspond to PopZ. (B) Coomassie blue staining of a 12.5% SDS-PAGE gel. Bands (black box) between 25 and 37 kDa (approximately 30 kDa) correspond to $PopZ^{\Delta 134\text{-}177}$.

Table 1. Bacterial strain list.

Strains	Cell Background	Plasmid Description	Construction of Plasmid	Source
YA#134	Rosetta DE3	pET28a_PopZ$_{1-133}$ His6X	pET28a was cut with NcoI and XhoI and PopZ$_{1-133}$ His6X (MSDQSQEPTMEEILASIRRIISEDD APAEPAAEAAPPPPPEPEPEPVSF DDEVLELTDPIAPEPELPPLETVGDI DVYSPPEPESEPAYTPPPAAPVFD RDEVAEQLVGVSAASAAASAFGSL SSALLMPKDGLEHHHHHH) was inserted and ligated	This study
YA#129	Rosetta DE3	pET28a_PopZ$_{1-133}$ I17A His6X	pET28a was cut with NcoI and XhoI and PopZ$_{1-133}$ I17A His6X (MSDQSQEPTMEEILASARRIISED DAPAEPAAEAAPPPPPEPEPEPVS FDDEVLELTDPIAPEPELPPLETVG DIDVYSPPEPESEPAYTPPPAAPVF DRDEVAEQLVGVSAASAAASAFGS LSSALLMPKDGLEHHHHHH) was inserted and ligated	This study
YA#138	MG1655 DE3	pACYC mCherry PopZ(WT)	JH42	Holmes *et al*. 2016
		pBAD RcdAGFP	JH 64	Holmes *et al*. 2016
YA#142	MG1655 DE3	pACYC mCherry PopZ(I17A)	pACYC was cut with NcoI and XhoI and mCherry PopZ I17A (MVSKGEEDNMAIIKEFMRFKVHME GSVNGHEFEIEGEGEGRPYEGTQT AKLKVTKGGPLPFAWDILSPQFMY GSKAYVKHPADIPDYLKLSFPEGFK WERVMNFEDGGVVTVTQDSSLQD GEFIYKVKLRGTNFPSDGPVMQKK TMGWEASSERMYPEDGALKGEIK QRLKLKDGGHYDAEVKTTYKAKKP VQLPGAYNVNIKLDITSHNEDYTIVE QYERAEGRHSTGGMDELYKPAGA LINMHGTMSDQSQEPTMEEILASA RRIISEDDAPAEPAAEAAPPPPEP EPEPVSFDDEVLELTDPIAPEPELP PLETVGDIDVYSPPEPESEPAYTPP PAAPVFDRDEVAEQLVGVSAASAA ASAFGSLSSALLMPKDGRTLEDVV RELLRPLLKEWLDQNLPRIVETKVE EEVQRISRGRGA) was inserted and ligated	This study
		pBAD RcdAGFP	JH 64	Holmes *et al*. 2016

Freezer stocks were plated onto LB medium agar plates containing 50 μg/mL kanamycin. Plates were incubated at 37 °C overnight for approximately 14 hours. Single, isolated bacterial colonies were scraped from the plates and dropped into approximately 30 mL of liquid LB medium supplemented with 30 μg/mL kanamycin for preculture growths. Precultures were shaken in an orbital shaker at 225 rpm and 37 °C for 14-16 hours. 4 mL of preculture growth were transferred to 1 L fresh liquid LB supplemented with 30 μg/mL of kanamycin for a main growth. Main growths were shaken at 225 rpm and 37 °C until the OD_{600} reached 0.50-0.60. The cells were then centrifuged at 7,700 rcf and 4 °C for 30 minutes and the supernatant was poured off. Cell pellets were resuspended in 0.25 L of M9 minimal medium per 1 L of LB growth. The M9 minimal medium was prepared with ingredients listed in Tables A1-A3 in the Appendix, including vitamin and trace metal supplements. For uniformly labeled-^{15}N and both uniformly labeled-^{13}C, ^{15}N enriched expression of $PopZ^{\Delta 134-177}$, the M9 minimal medium was supplemented either with $^{15}NH_4Cl$ (for uniform labeling with ^{15}N) or $^{15}NH_4Cl$ and $^{13}C_6$-glucose (for uniform labeling with ^{15}N and ^{13}C), respectively. Cells were allowed to equilibrate under induction conditions for 30 minutes with shaking at 225 rpm and 37 °C before protein expression was induced with 0.8 mM of IPTG for 12 hours. Cells were centrifuged at 7,700 rcf and 4 °C for 30 minutes and the supernatant was discarded. Cell pellets were stored at -80 °C.

2.3. Purification of $PopZ^{\Delta 134-177}$

NMR studies require proteins at high purity, and there are multiple purification steps performed in order to ensure this. After the desired protein is expressed, the cells

harboring it are broken open through lysis to release the cytosol. The cytosol is separated from the cellular debris through centrifugation. The protein can then be separated from the rest of the cytosolic content through a variety of purification methods including size exclusion, ion exchange, glutathione-affinity, and nickel-affinity chromatography. Size exclusion separates proteins based on molecular weight. Ion exchange includes anion and cation exchange which separates proteins based on their intrinsic negative and positive charges, respectively. Glutathione affinity purifies proteins based on their ability to bind to glutathione. In this regard, target proteins are usually expressed as fusion proteins with glutathione-S-transferase (GSA), which has a high affinity towards glutathione. Often, multiple modes of purification need to be used to obtain high levels of protein purity.

For this project, purification of PopZ$^{\Delta 134\text{-}177}$ was performed using Nickel-affinity chromatography, wherein the target protein is engineered to include a HisTag which has high affinity towards nickel. Therefore, the protein binds to the nickel while the rest of the cytosol passes through the column. The protein is then eluted off the column by introducing a compound that has higher affinity for the nickel than the HisTag. The compound elutes with the protein but is later removed by exchanging the buffer through dialysis or spin columns. Protein purity is assessed using SDS-PAGE and staining. The details for PopZ$^{\Delta 134\text{-}177}$ purification from *E. coli* are described below.

Cell pellets were removed from the freezer and thawed on ice before being resuspended in 5 mL of Buffer A (25 mM Tris, 150 mM NaCl, 20 mM imidazole at pH 7.5) supplemented with 50 µL of Halt EDTA-free protease inhibitor and 5 µL of Benzonase nuclease. The cells were lysed using a French press in triplicate at 1,500 psi. Cell lysate

was centrifuged at 27,200 rcf and 4 °C for 40 minutes and the supernatant was kept and concentrated to 5 mL using a 10K Molecular Weight Cutoff (MWCO) centrifugal filter and then filtered with a 0.22 μm syringe filter to remove large particulates. The protein was purified using Ni-affinity Fast Protein Liquid Chromatography (FPLC; GE Healthcare ÄKTA purifier 900) equipped with a GE Healthcare HisTrap HP 1 mL column. Bound $PopZ^{\Delta 134\text{-}177}$ was eluted using an increasing imidazole concentration gradient up to 500 mM. A representative chromatogram showing the purification profile is shown in Figure 7. Fractions containing the peak were pooled together and buffer-exchanged back into Buffer A for a second round of purification. The protein fractions were pooled again and the buffer was exchanged to Buffer B (25 mM Tris, 150 mM NaCl pH 7.5). The purity of the sample was analyzed by SDS-PAGE and visualized with Coomassie blue. A picture of a typical gel is shown in Figure 8.

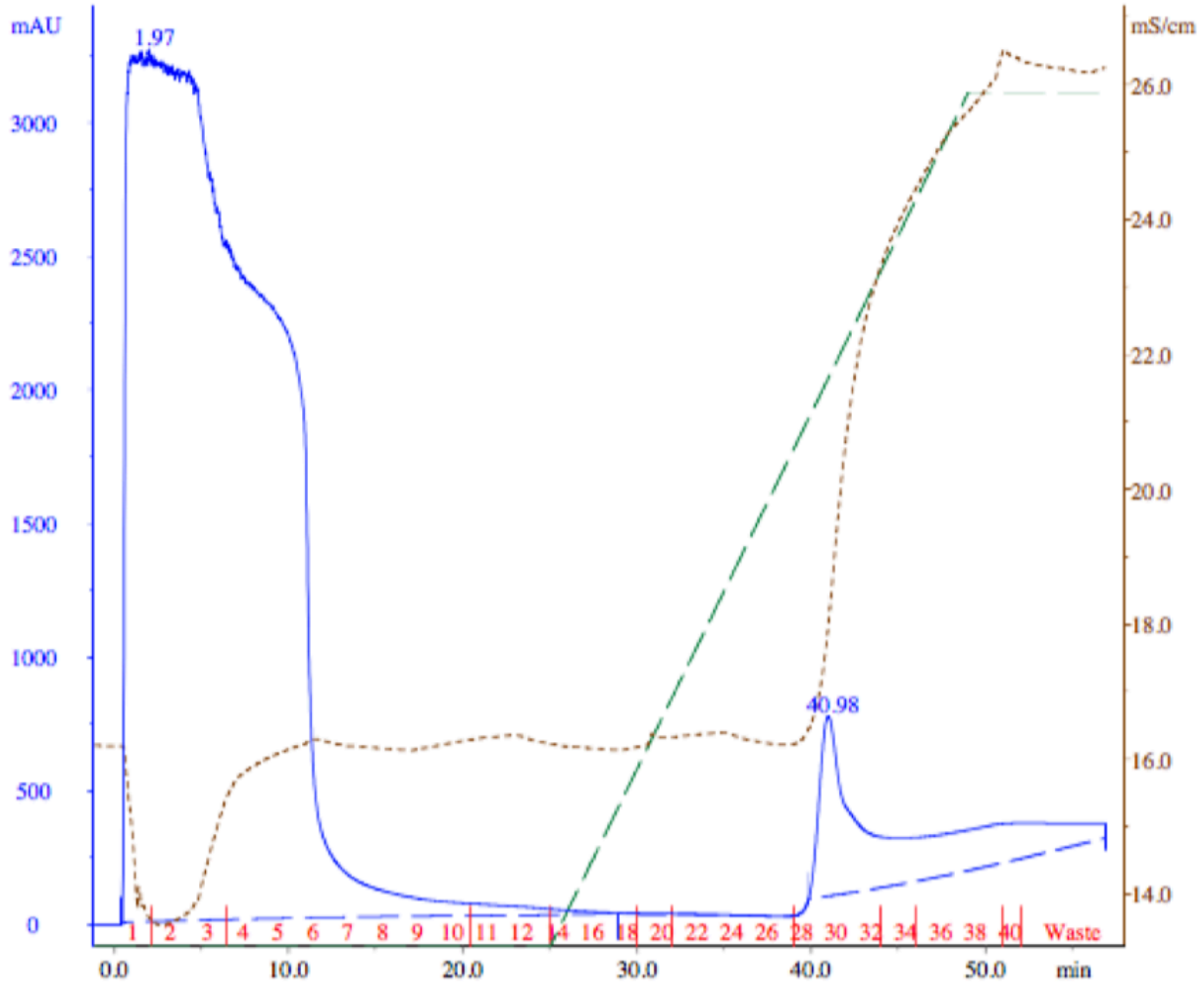


Figure 7. Chromatogram showing the PopZ$^{\Delta 134\text{-}177}$ purification profile. The colored/dashed lines represent the following measurements: (blue) absorption at 280 nm in mAU; (green) percent elution buffer used relative to the running buffer; (brown) conductivity of the solution in mS/cm. The red numbers show the fractionation of the eluent into different tubes. The blue peak between fractions 28-30 represents the elution peak of PopZ$^{\Delta 134\text{-}177}$. The sample in these fractions was collected and combined for a second round of purification. Blue peaks prior to 28-30 represent proteins without a HisTag, aggregated proteins, and cellular debris.

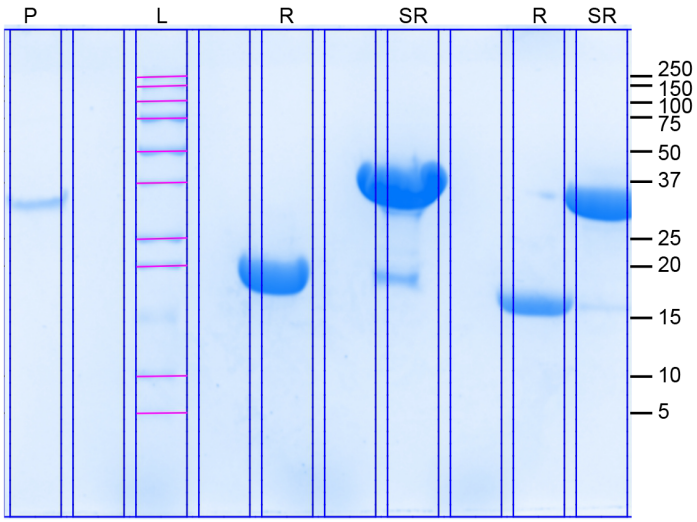


Figure 8. Representative gel used to assess protein purity. Lanes show bands representing PopZ$^{\Delta 134\text{-}177}$ (P), cleaved RcdA (R), and uncleaved Sumo-RcdA (SR). The ladder standard (L) is also shown with size markers (kDa) labeled on the right. PopZ$^{\Delta 134\text{-}177}$, RcdA, and Sumo-RcdA are found at approximately 30 kDa, 19 kDa, and 31 kDa, respectively.

2.4. Expression of Sumo-RcdA

Wild-type RcdA was cloned into the BL21 (DE3) *E. coli* expression vector pET28a (Novagen) as a fusion protein with Sumo and with a 6-His HisTag. Agar plugs containing *E. coli* transformed cells were surface-scraped with a pipette tip which was dropped into 50 mL conical tubes with 30 mL of LB supplemented with 30 µg/mL of kanamycin. The *E. coli* were shaken at 225 rpm and 37 °C overnight for bacterial growth. The next morning, *E. coli* freezer stocks were made by mixing together 0.75 mL of growth with 0.75 mL of 50% v/v glycerol in DI water for a final *E. coli* freezer stock in 25% glycerol. These were flash frozen in liquid nitrogen and stored at -80 °C. Freezer stocks were used for future growths in lieu of transformations as described below.

Freezer stocks were plated onto LB medium agar plates containing 50 µg/mL of kanamycin. Plates were incubated at 37 °C overnight for approximately 14 hours. Single,

isolated bacterial colonies were scraped from the plates and dropped into approximately 30 mL of liquid LB medium supplemented with 30 µg/mL of kanamycin for preculture growths. Precultures were shaken in an orbital shaker at 225 rpm and 37 °C for 14-16 hours. 4 mL of preculture growth were transferred to 1 L of fresh liquid LB supplemented with 30 µg/mL of kanamycin for a main growth. Main growths were shaken at 225 rpm and 37 °C until the OD_{600} reached approximately 0.50. Protein expression was induced with 1.0 mM of IPTG for 12 hours with shaking at 225 rpm and 21 °C. Cells were centrifuged at 7,700 rcf and 4 °C for 30 minutes and the supernatant was discarded. Cell pellets were stored at -80 °C.

2.5. Purification of Sumo-RcdA

Cell pellets were removed from the freezer and thawed on ice before being resuspended in 5 mL of Buffer C (20 mM HEPES, 100 mM KCl, 2 mM $MgCl_2$, 20 mM imidazole at pH 7.5) supplemented with 50 µL of Halt EDTA-free protease inhibitor and 5 µL of Benzonase nuclease. The cells were lysed using a French press in triplicate at 1,500 psi. Cell lysate was centrifuged at 27,200 rcf and 4 °C for 40 minutes and the supernatant was kept and concentrated to 5 mL using a 10K MWCO centrifugal filter and then filtered with a 0.22 µm syringe filter to remove large particulates. The protein was purified using Ni-affinity FPLC (GE Healthcare ÄKTA purifier 900) equipped with a GE Healthcare HisTrap HP 1 mL column. Bound Sumo-RcdA was eluted using an increasing imidazole concentration gradient up to 500 mM. A representative chromatogram showing the purification profile is shown in Figure 9. Fractions containing the peak were pooled and buffer exchanged back into Buffer C for a second round of purification.

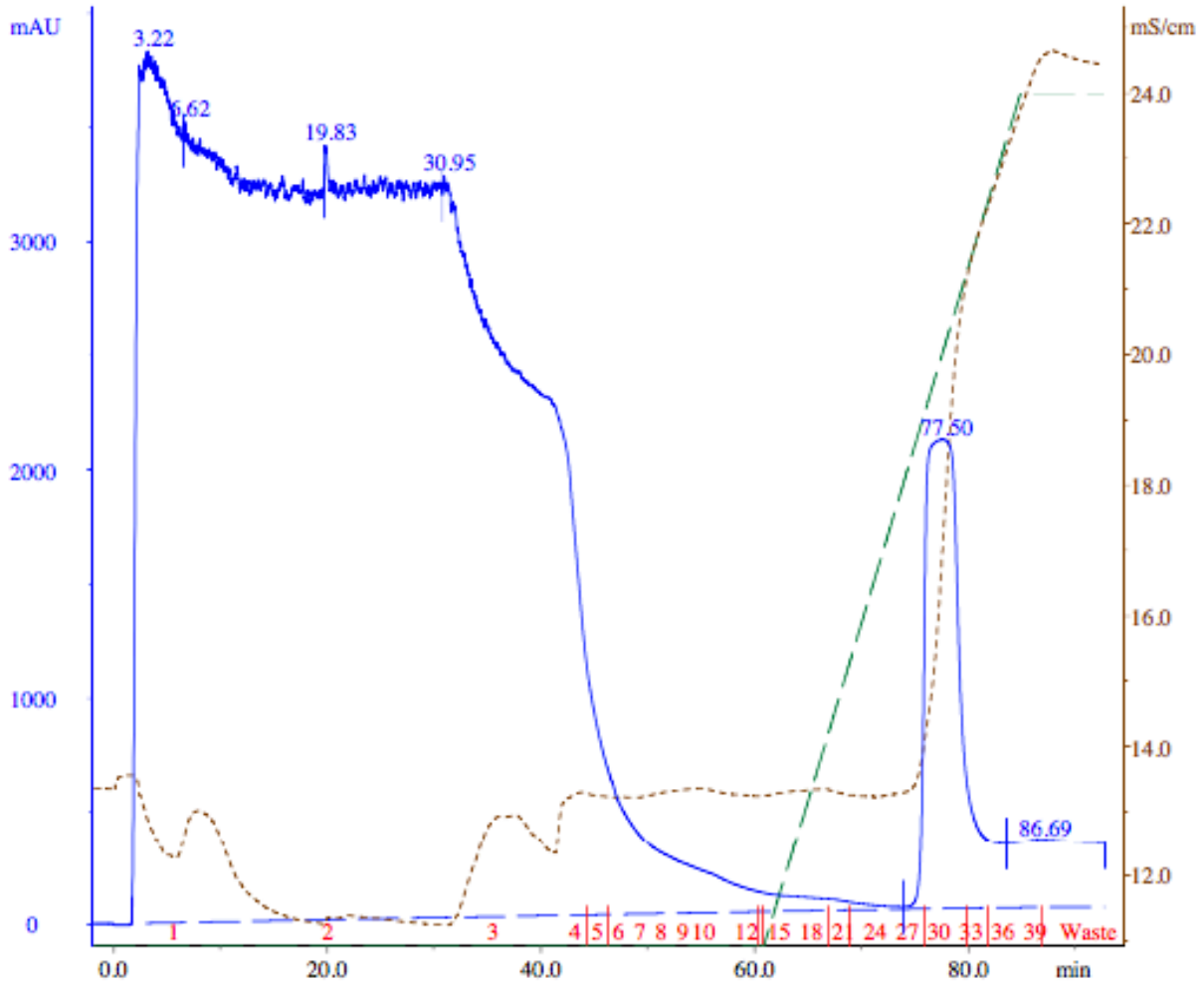


Figure 9. Chromatogram showing the Sumo-RcdA purification profile. The colored/dashed lines represent the following measurements: (blue) absorption at 280 nm in mAU; (green) percent elution buffer used relative to the running buffer; (brown) conductivity of the solution in mS/cm. The red numbers show the fractionation of the eluent into different tubes. The blue peak between fractions 28-32 represents the elution peak of Sumo-RcdA. The sample in these fractions was collected and combined for a second round of purification. Blue peaks prior to 28-32 represent proteins without a HisTag, aggregated proteins, and cellular debris.

2.6. Cleavage of Sumo-RcdA

Some recombinant proteins have difficulty folding and may form aggregates and precipitate as inclusion bodies if the bacterial machinery cannot properly express and fold them. For this reason, proteins can be expressed with stabilizing tags that assist in folding,

solubility, and purification of the target recombinant protein. For this reason, RcdA is expressed as a fusion protein with small ubiquitin-like modifier (Sumo)[76]. However, once expression of the fusion protein is achieved, we cleave the Sumo from RcdA in order to perform studies using only RcdA. Cleaving the fusion protein utilizes a Sumo protease, commonly ULP1. Sumo proteases are unique among cleavage enzymes in that they do not recognize a specific sequence, but instead recognize the tertiary structure of Sumo, resulting in extremely efficient cleavage at the C-terminal diglycine site, regardless of the identity of the amino acid directly following the diglycine motif (with the exception of proline)[77]. This ensures a cleavage without extraneous residues left on the target protein N-terminus[76,77]. The details of Sumo-RcdA cleavage for our studies is given below.

Fractions containing the Sumo-RcdA sample from the second HisTag purification were pooled together and buffer exchanged into Buffer D (25 mM Tris, 150 mM NaCl, at pH 7.5). To cleave Sumo from RcdA, protein yield was measured using UV-Vis spectrophotometry to determine the amount of Sumo protease that was required in the sample to get a cleavage reaction mixture of 1 Unit Sumo protease per 200 µg of Sumo-RcdA. The reaction mixture was diluted to 0.5 mL, dithiothreitol (DTT) was added to a concentration of 2 mM, and mixed gently on an oscillator for approximately 15 hours at 4 °C. The cleaved RcdA was separated from Sumo through a third round of Ni-affinity chromatography using an FPLC. A representative chromatogram showing the purification profile can be viewed in Figure 10. Fractions containing the cleaved RcdA were pooled together and buffer-exchanged into Buffer E (20 mM HEPES, 100 mM KCl, 2 mM $MgCl_2$ at pH 7.5). The sample was concentrated in preparation for NMR binding experiments with $PopZ^{\Delta 134\text{-}177}$ and the concentration was estimated using UV-Vis spectrophotometry

(Figure 11). The purity of the samples was analyzed by 12% SDS-PAGE and visualized with Coomassie blue (Figure 8).

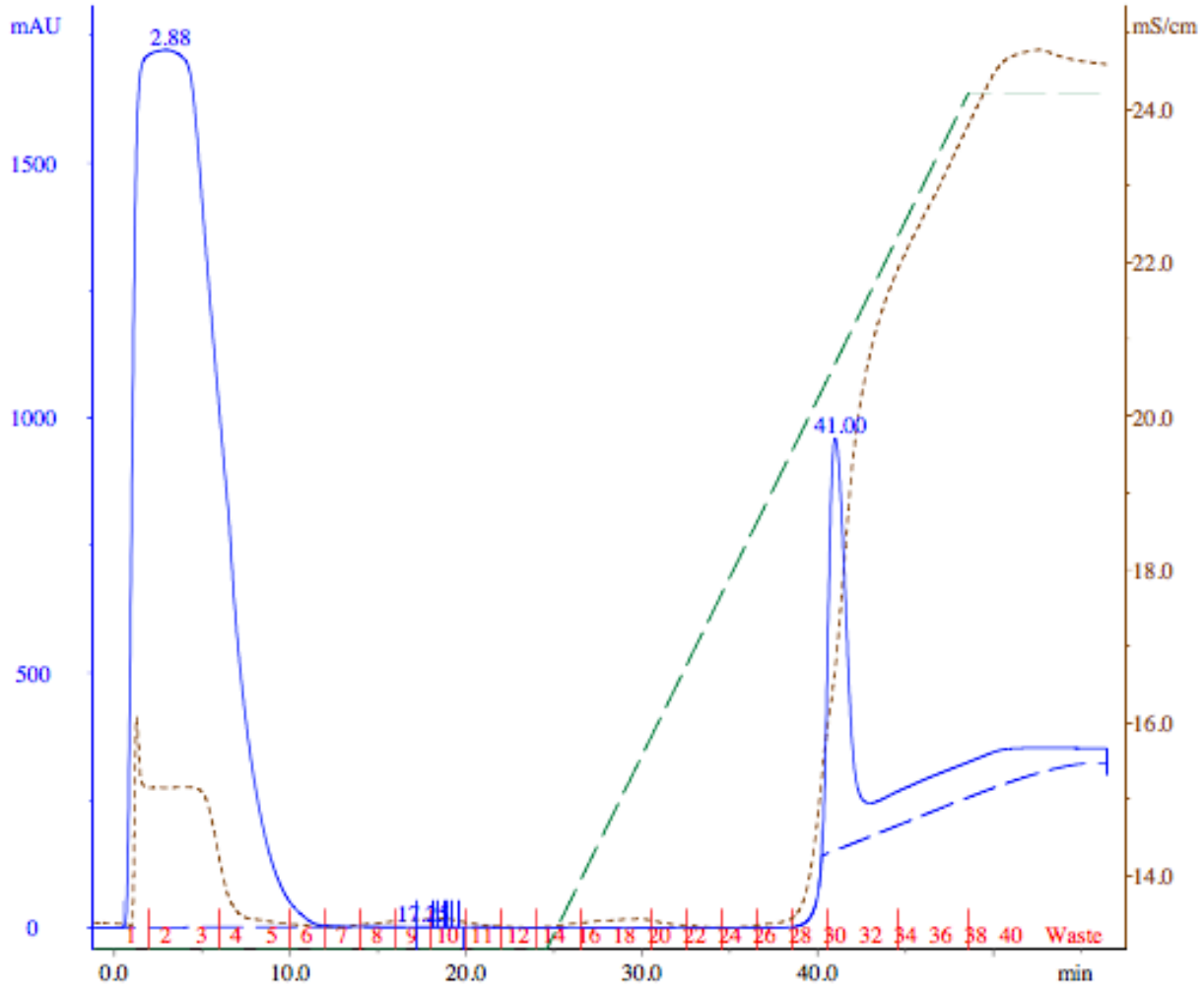


Figure 10. Chromatogram showing the cleaved RcdA purification profile. The colored/dashed lines represent the following measurements: (blue) absorption at 280 nm in mAU; (green) percent elution buffer used relative to the running buffer; (brown) conductivity of the solution in mS/cm. The red numbers show the fractionation of the eluent into different tubes. The tall blue peak between fractions 1-4 represents the elution peak of RcdA. The sample in these fractions was collected and combined. The blue peak following fractions 1-4 represent HisTag proteins, including Sumo, Sumo protease, and uncleaved Sumo-RcdA.

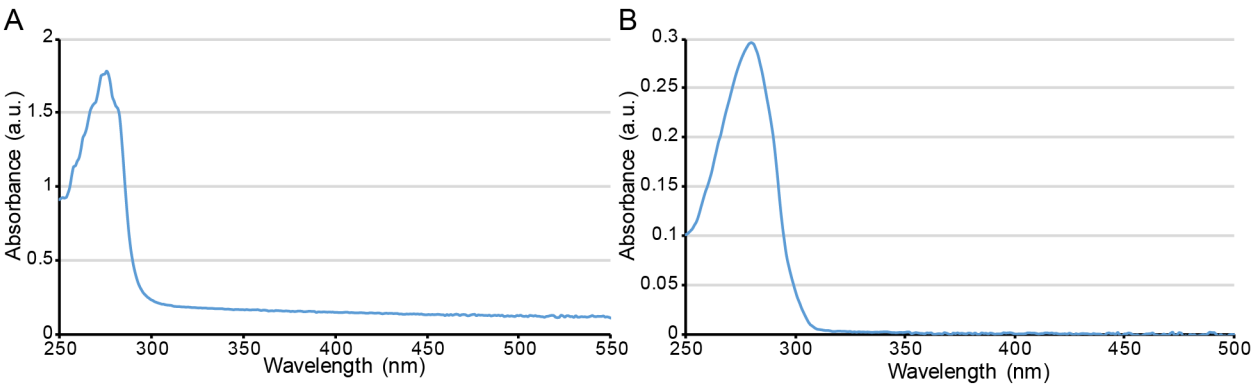


Figure 11. UV-Vis spectra of PopZ$^{\Delta 134\text{-}177}$ and RcdA. (A) Spectrum showing the UV-Vis profile of PopZ$^{\Delta 134\text{-}177}$. After performing a baseline correction, the observed absorbance corresponds to a concentration of 430 μM. (B) Spectrum showing the UV-Vis profile of RcdA. The observed absorbance corresponds to a concentration of 321 μM.

2.7. NMR sample preparation

There are a variety of conditions to be optimized for NMR studies. The buffer pH is important and needs to be adjusted such that the protein is structurally stable and does not aggregate at 25-35 °C for a few days or weeks. Slightly acidic pH values are favored for protein NMR experiments (e.g. pH 5.0-6.5) to reduce the chemical exchange rate of the amide protons with water. At basic pH, due to the fast exchange with the solvent, the intensity of the HN signals drops. The salt content of the buffer needs to be minimized, as high salt concentrations are detrimental to NMR probe sensitivity, especially when using a CryoProbe, and thus affect the signal-to-noise (S/N) ratio. Additionally, many proteins are temperature sensitive, with the majority of proteins remaining stable for longer periods of time when kept at cold temperatures. However, NMR studies might require higher temperatures (25-35 °C) for optimal signal acquisition, so it is important to make sure a protein sample is stable for days or even weeks at NMR experiment temperatures. Fortunately, stability of a protein can be monitored using NMR

spectroscopy as the experiments progress by watching for distinct changes in spectra that are indicative of sample breakdown. These changes include losses in peak resolution, spectral profile changes, and deterioration of S/N. NMR sample conditions for both $PopZ^{\Delta 134\text{-}177}$ and RcdA experiments are given below.

$PopZ^{\Delta 134\text{-}177}$ was buffer exchanged into Buffer F (50 mM phosphate, 50 mM citric acid, 20 mM NaCl, and 3 mM NaN_3 at pH 5.5) or Buffer E (20 mM HEPES, 100 mM KCl, 2 mM $MgCl_2$ at pH 7.5) for NMR structural studies and RcdA binding studies, respectively. A lower pH of 5.5 was used for structural studies to reduce the exchange of amide protons with water, and to ensure that all NH resonances are observable in the spectra. Binding studies were performed at a higher pH because the binding partners were not stable at pH 5.5. Samples were concentrated to 280 µL using a 10K MWCO centrifugal filter and then mixed with deuterium oxide (D_2O), 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS), and sodium azide (NaN_3) to final concentrations of 5% v/v D_2O, 200 µM DSS, and 1 mM NaN_3 (or 4 mM NaN_3 if the sample is in Buffer F). This gave a final volume of 300 µL (the ideal volume for Bruker Shigemi NMR tubes). D_2O is used for sample locking in the NMR spectrometer, DSS is used as an internal reference for the NMR experiments, and NaN_3 is used to inhibit bacterial growth in the sample. The sample was then packed into a 5 mm Shigemi NMR tube.

2.8. NMR spectroscopy

PopZ concentration was estimated using UV-Vis spectrophotometry (ε_{280} =2,980 cm^{-1} M^{-1}). The 2D ^{1}H-^{15}N Heteronuclear Single Quantum Correlation (HSQC) NMR spectrum and standard protein backbone and side-chain NMR spectra (including 3D HNCA, HNcoCA,

HNCO, HNcaCO, HNCACB, CBCAcoNH, CCcoNH, HBHANH, HBHAcoNH, and HcccoNH, ^{15}N- and ^{13}C-edited NOESY) were collected at 25 °C on a Bruker AVANCE III HD 800 MHz NMR spectrometer (City University of New York (CUNY) Advanced Science Research Center, New York, NY) or on a Bruker Avance NEO 700 MHz NMR (Bruker Biospin, Inc., Billerica, MA) equipped with 5 mm triple resonance inverse TCI CryoProbes. All NMR data were processed using NMRPipe.[78] Analysis and assignments of the 2D and 3D data sets were carried out using NMRFAM-Sparky.[79] The assignment process was facilitated by using the PINE server for initial automated assignments[80,81] before completing the assignments manually. The secondary chemical shift values were calculated by subtracting experimental chemical shift values from random coil values supplied by NMRFAM-Sparky. Relaxation and heteronuclear NOE data were analyzed using the Dynamics Center software package (Bruker BioSpin, Inc).

To characterize the binding site of PopZ$^{\Delta 134\text{-}177}$ to RcdA, a series of 130 μM ^{15}N-enriched PopZ$^{\Delta 134\text{-}177}$ samples were prepared in the presence of a range of concentrations for RcdA in Buffer E including 0 μM RcdA (PopZ$^{\Delta 134\text{-}177}$ only), 50 μM RcdA, 100 μM RcdA, 150 μM RcdA, 200 μM RcdA, 300 μM RcdA, 500 μM RcdA, 700 μM RcdA, 900 μM RcdA, and 960 μM RcdA. 2D ^{1}H-^{15}N HSQC spectra were collected on all samples at 25 °C on a Bruker AVANCE III HD 700 MHz NMR spectrometer (CUNY Advanced Science Research Center, New York, NY). Combined ^{1}H and ^{15}N chemical shift perturbations (ΔHN) were calculated using Equation 1,

(Equation 1) $$\Delta\mathrm{HN} = \sqrt{(\Delta\mathrm{H})^2 + (0.15 \times \Delta\mathrm{N})^2}$$

where ΔH and ΔN are the chemical shift perturbations in ppm for ^{1}H and ^{15}N, respectively, and 0.15 is a scaling factor corresponding to the relative chemical shift dispersion in the ^{1}H and ^{15}N dimensions.

2.9. Residual dipolar coupling (RDC)

Residual dipolar coupling provides relative orientations of protein domains in a sample and is used to facilitate structural calculations of a protein. To acquire this data, it is necessary for the sample to be present in partially aligning media such as polyacrylamide gels used in this project (see *section 3.3.6.1.* for more information on RDC). For $PopZ^{\Delta 134-177}$ RDC NMR experiments, the purified and NMR-prepared $PopZ^{\Delta 134-177}$ sample was used to rehydrate a 5.4 mm or 6.0 mm (outer diameter) gel previously cast and dehydrated (see RDC gel preparation below). The gel was stretched into a New Era Enterprises 4.2 mm (inner diameter) NMR tube for RDC NMR experiments. The ^{1}H-^{15}N HSQC NMR spectrum was collected at 25 °C on a Bruker AVANCE III HD 800 MHz NMR spectrometer equipped with a 5 mm triple resonance inverse TCI CryoProbe.

2.10. RDC gel preparation

RDC gel preparation and sample preparation were performed using New Era Enterprises gel kits. A 5.4% acrylamide RDC gel was created by mixing 40% bis-acrylamide purchased from Bio-Rad with water and polymerized with 0.1% w/v ammonium persulfate and 0.1% v/v TEMED. This was cast in a New Era Enterprises gel stretching chamber and allowed to polymerize overnight at room temperature. The chamber was sealed using

parafilm to prevent gel leakage and left upright to ensure that gel ends remained flat upon polymerization.

The gel was removed from the chamber by gentle prodding using the supplied support rod and then dialyzed in pure water for eight hours to remove unreacted chemicals. A second round of dialysis in fresh water was then performed. The gel was cut to approximately 2.1 cm in length using a razor blade and then dehydrated on a flat platform for 18-24 hours in a desiccator at room temperature. Gel surfaces became extremely sticky upon drying, so gels remained unmoved during the dehydration process to prevent surface tears. Following dehydration, the surface of the gel was wetted with protein sample to remove it from the drying platform. The dehydrated gel was placed back in the gel chamber and the gel was incubated overnight with the protein solution at 4 °C.

The gel stretching kit was assembled according to the schematic below (Figure 12). The end of the NMR tube (New Era Enterprises 4.2 mm inner diameter) that receives the gel was briefly submerged in NMR sample buffer to wet the insides. In doing so, it decreased the chances of gel friction fractures during gel stretching. The gel was stretched into the NMR tube using the assembled gel stretching kit. The top plug was inserted from the top of the NMR tube. The end of the gel was slowly pressed out of the bottom of the tube using the support rod until approximately 2.1 cm in length was left in the tube. This length of the gel corresponded to the height of the sample window in the NMR probe. The protruding section of the gel was cut off using a razor blade and the bottom plug was inserted into the bottom of the tube to seal the end.

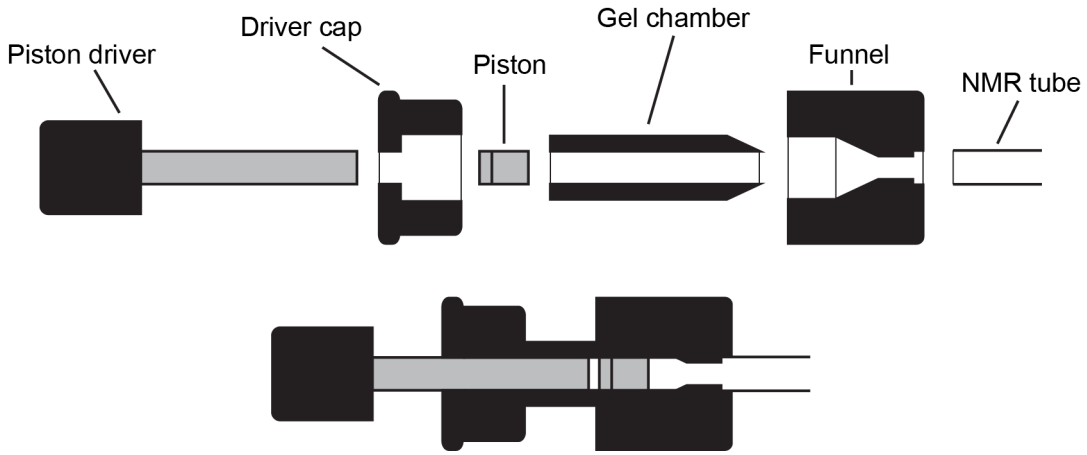


Figure 12. Schematic of the RDC gel stretching kit from New Era Enterprises. Individual components of the kit are shown on top. Assembled components are shown on bottom. The gel chamber houses the RDC gel sample and is driven by the piston through the funnel. Passage through the funnel axially stretches the gel to match the inner diameter of the NMR tube into which it goes.

CHAPTER 3

POPZ$^{\Delta 134\text{-}177}$ STRUCTURAL CHARACTERIZATION

This chapter will focus on the structural characterization of PopZ$^{\Delta 134\text{-}177}$ and introduce the techniques used to facilitate the calculations. First, the chapter will provide an introduction on protein structural characterization (**section 3.1**) and a brief overview of the theory of NMR spectroscopy **(section 3.2)**, followed by techniques used for PopZ$^{\Delta 134\text{-}177}$ structural characterization **(section 3.3),** including protein backbone and sidechain spectral assignments, long distance constraints, and orientation constraints. Lastly, PopZ$^{\Delta 134\text{-}177}$ dynamics studies (**section 3.4**), including T_1 and T_2 relaxation and heteronuclear NOE, will be discussed. For these studies, PopZ$^{\Delta 134\text{-}177}$ samples were enriched with the NMR-active isotopes ^{15}N and ^{13}C. All data was collected at both the City University of New York's (CUNY) Advanced Science Research Center (ASRC).

3.1. Introduction to protein structural characterization

The function(s) of a protein are dependent on that protein's structure. For a long time, the relationship between protein structure and function was accepted as a lock and key hypothesis, where a unique structure determined a unique function[82]. However, recent advancements in our understanding of proteomes has led to the conclusion that the relationship is not that simple, but instead consists of a continuum where a protein can exist as an ensemble of proteoforms with varying structural features and functional

potentials[82]. The extreme end of this continuum includes IDPs and IDRs which rapidly interconvert between multiple conformers and have been shown to be highly versatile in their binding interactions[41,42,82]. In many cases, structural data is reported along with functional data, but this is not always the case. In some instances, protein function remains unknown, so protein functions must be predicted. Functional predictions from protein structure is one avenue of approach, so determining protein structure is an important task[83,84].

Nuclear magnetic resonance (NMR) spectroscopy, X-ray crystallography, and Cryo-electron microscopy (CryoEM) are useful techniques for investigating the structure of biological macromolecules at atomic resolution. As of November, 2020, approximately 133,000, 11,500, and 4,200 protein structures have been deposited to the Protein Data Bank (PDB) using X-ray crystallography, NMR spectroscopy, and CryoEM, respectively[85]. X-ray crystallography was used to solve the first protein structure in 1957 for myoglobin from sperm whale[86]. This technique irradiates a crystallized protein sample with an X-ray beam and uses the resulting diffraction patterns and intensities to map out electron densities. These densities are used to computationally create a model of the protein structure[87]. The first protein structure solved with NMR spectroscopy was proteinase inhibitor IIA from bull seminal plasma[88]. NMR spectroscopy detects radio wave frequencies from atomic nuclei that are correlated to one another to generate distance, angle, and orientation constraints that are used to develop 3D models of a protein. CryoEM vitrifies an aqueous sample to freeze the protein in an amorphous, noncrystalline solid. A low-dose electron beam is applied to the sample to generate 2D projection images that are reconstructed into a 3D image[89].

For studies surrounding IDPs, X-ray crystallography is at a distinct disadvantage in relation to NMR spectroscopy due to the fact that IDPs rarely form crystalline units, an important requirement for crystallography[90]. Similarly, CryoEM is limited by the presence of disorder in IDPs[89]. Alternatively, solution NMR spectroscopy can observe IDPs with high quality signal despite the IDP sampling multiple conformers, making it the method of choice for this class of protein. Furthermore, proteins can be observed in conditions similar to their native environments using NMR spectroscopy, which increases the likelihood that solved structures are representative of true conformations. However, NMR spectroscopy is not without its own limitations. It is an inherently insensitive technique that requires large amounts of protein sample (milligram quantities). NMR spectroscopy also requires robust tumbling (rotation in solution) of the sample. Since tumbling rate is inversely correlated with protein size, the technique is best for smaller proteins that are typically less than 50 kDa in size. Larger globular proteins are best suited for X-ray crystallography or CryoEM, assuming the protein meets the required conditions (i.e. the ability to form crystals).

3.2. Theory of NMR: a brief discussion

Nuclei have the intrinsic property of spin, whereby they act like rotating spheres that spin on an axis. Nuclei are also magnetic and have a magnetic moment that points along the direction of the spin axis[91-93]. In the absence of a magnetic field, the nuclear spin states of all nuclei are energetically equivalent. However, in the presence of a magnetic field, the nuclei separate into two energetically distinct spin states (Figure 13): the α-spin state, where the magnetic moment aligns parallel to the magnetic field, and the

β-spin state, where the magnetic moment aligns antiparallel to the magnetic field[91-93]. The energy difference between the two states is directly proportional to the strength of the magnetic field.

The population difference between the two spin states is defined by the Boltzmann distribution (Equation 2)[91]:

(Equation 2) $$\frac{P_\beta}{P_\alpha} = e^{\left(\frac{-\Delta E}{k_b T}\right)}$$

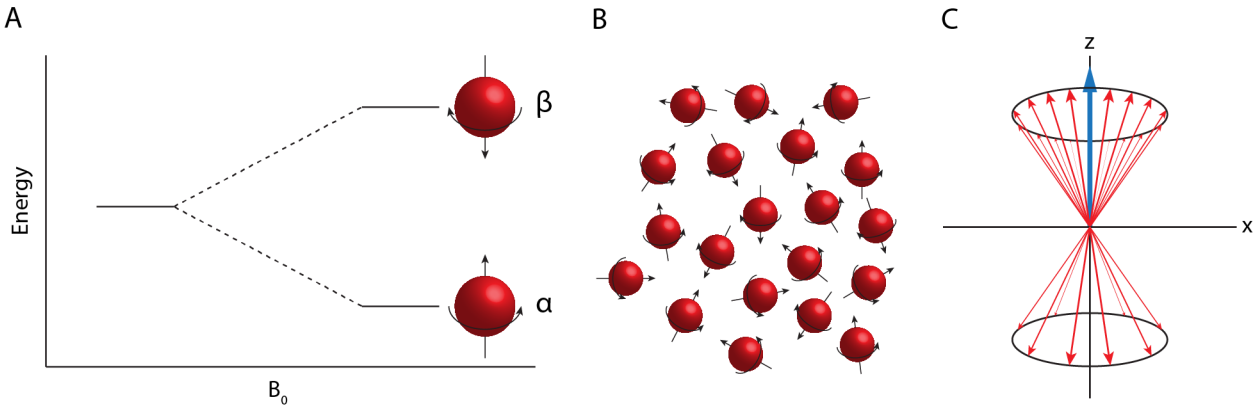


Figure 13. Establishment of net magnetization. (A) The difference in energy between α- and β-spin states is dependent on the strength of the external magnetic field (B_0). Spins whose magnetic moments are parallel to the magnetic field (α-spin state) are more favorable and have a lower energy than those that are antiparallel (β-spin state). This results in a very slight preference for α-spin states over β-spin states for the nuclei in a sample, as seen in (B). (C) This slight preference results in a small net magnetization (blue arrow) in the z-direction which is parallel to the magnetic field.

where Pα and Pβ are the populations of α-spin states and β-spin states, respectively, ΔE is the difference in energy between the two spin states, k_b is the Boltzmann constant, and T is the temperature in Kelvin[91]. The α-spin state is energetically more favorable than the β-spin state. Therefore, the population of the α-spin state is slightly higher than the

population of the β-spin state. Due to this difference, there is a net magnetization in the direction of the magnetic field, classically called the z-direction[91].

Related to the nuclear spin is nuclear precession, where the nucleus' axis of rotation revolves around the magnetic field in much the same way a spinning top will revolve around the Earth's gravitational field. This nuclear precession is called the Larmor frequency. Emission of a radiofrequency pulse at the Larmor frequency will excite the nuclei, resulting in a shift of the net magnetization from the z-direction into the xy-plane. Following this, the nuclei will relax back to equilibrium where the net magnetization returns to the z-direction by emitting radio waves that are detected by NMR instrumentation[91-93]. It is important to note that different nuclei in a protein sample will experience the magnetic field slightly differently than what is experienced by other nuclei. This is because the magnetic field is slightly modified by the microenvironment around the nuclei (neighboring amino acids, solvent molecules, salt ions, etc.). As a result, the energy difference between the α-spin state and β-spin state of a unique nucleus will be slightly different than other nuclei, so it will emit a slightly different radio wave frequency[91,92]. The unique frequencies (also called chemical shifts) detected from the nuclei in a protein sample can then be used to characterize the sample[92].

3.3. PopZ$^{\Delta 134\text{-}177}$ structural characterization

NMR spectroscopy allows us to observe proteins on an atomic basis and in near-native environments, the latter of which is an advantage over other spectroscopic methods used for structural elucidation[93,94]. It is particularly important for the study of IDPs because their

inherent conformational flexibility is incompatible with X-ray crystallography and CryoEM techniques but remain observable in NMR studies[89,90]. Given their size, proteins are best investigated using multidimensional NMR to observe specific atomic correlations, thereby reducing spectral overlap seen in 1D spectra. 2D experiments can provide important insights into protein stability, intramolecular dynamics, and domain orientations. 3D experiments allow for detailed sequence information and distance measurements. However, due to the volume of information obtained, 3D data acquisition is complex and time-consuming. As such, depending on the specific experiment, a single 3D NMR experiment can take hours or even days to perform.

3.3.1. Spectral assignments

All NMR-active nuclei in a protein sample emit a unique radio wave frequency signal after experiencing irradiation in the presence of a magnetic field. Nuclei can be coupled together to observe certain combinations of nuclear frequencies. For example, a ^{1}H-^{15}N heteronuclear single quantum correlation (HSQC) experiment observes the nuclear frequencies of ^{1}H and ^{15}N nuclei that are bonded together[95]. Importantly, each amino acid of a protein contains a ^{1}H-^{15}N pairing on its backbone, with the exception of proline which does not contain the amide ^{1}H. Therefore, the ^{1}H-^{15}N HSQC spectrum will theoretically show a peak for each amino acid in a protein sequence, as well as for side chain ^{1}H-^{15}N pairs in tryptophan, asparagine, glutamine, and arginine residues. This leads to a unique "fingerprint" spectrum for a protein sample[95].

The observed dispersion of peaks can immediately give important information regarding the type of protein being observed. Figure 14A shows the PopZ$^{\Delta 134\text{-}177}$ ^{1}H-^{15}N

HSQC spectrum characterized by a narrow dispersion of ^{1}H chemical shifts characteristic for an IDP[96], and resolution between the peaks indicates PopZ$^{\Delta 134\text{-}177}$ was in a functional state required for NMR study. Alternatively, a globular protein will have a much wider dispersion of peaks in the ^{1}H dimension[96,97], as seen with the antifreeze protein *Ap*AFP752 (Figure 14B).

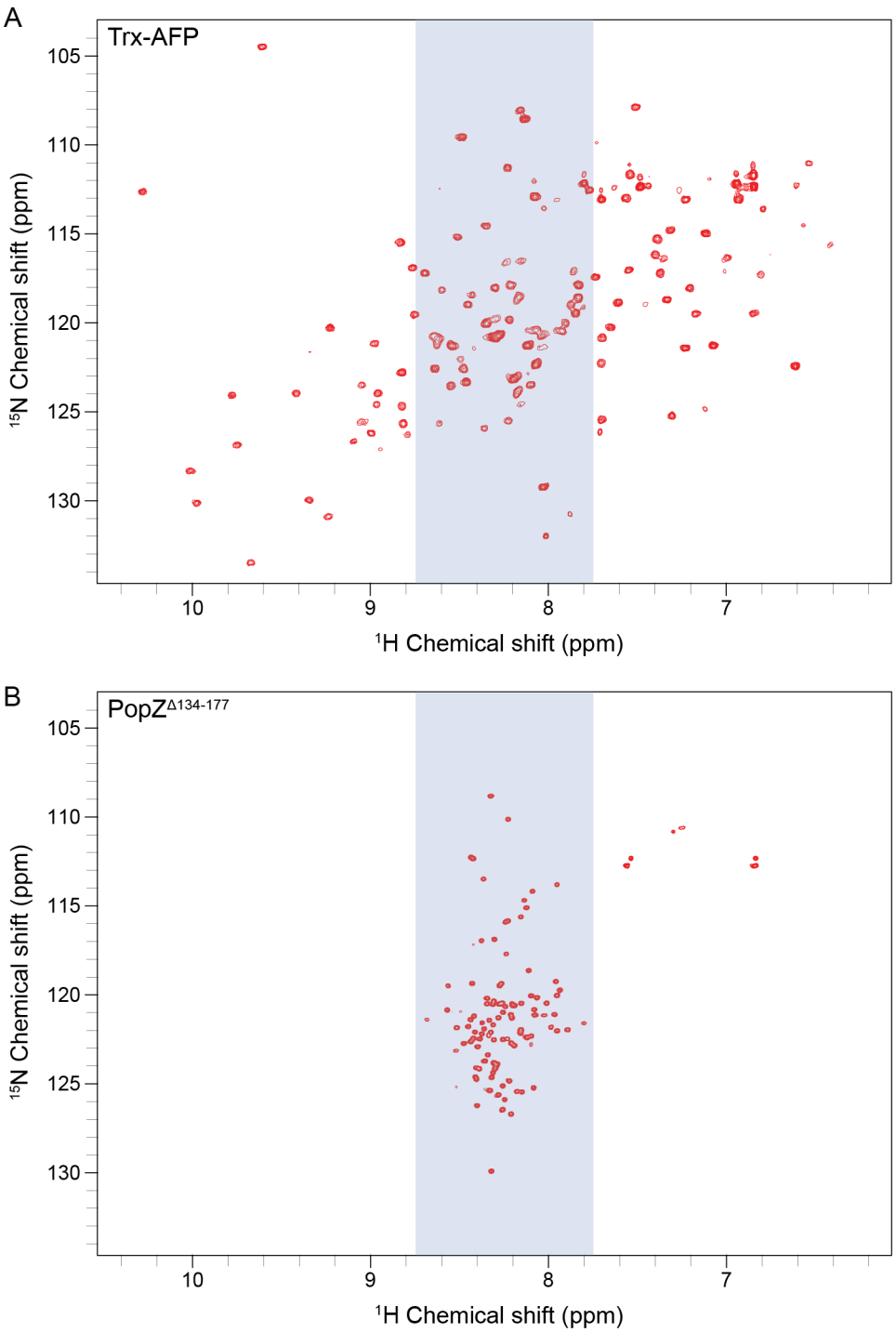


Figure 14. ^{1}H resonance dispersion. Structurally ordered and disordered proteins display characteristic peak dispersions in the backbone ^{1}H dimension. Ordered proteins tend to have a wide dispersion between 6 and 10 ppm, as seen in (A). Disordered proteins have a much narrower dispersion, typically between 7.5 and 9.0 ppm, as seen in (B). Blue shading represents the same area in each spectrum to exemplify the dispersion differences.

In 3-dimensional experiments, a correlation is observed between three different nuclei, such as ^{1}H, ^{15}N, and ^{13}C. The subsequent spectrum will be 3-dimensional. Multiple 2D and 3D NMR data are required for assigning peaks with the ultimate goal to have each peak across the spectra assigned to a protein residue. The 3D experiments utilized in making backbone and side chain assignments include HNCA, HNcoCA, HNCO, HNcaCO, HNCACB, CBCAcoNH, CCcoNH, HBHANH, HBHAcoNH, and HcccoNH (Figure 15). The acronyms of each experiment type describe which nuclei are used in the experiment, with capital letters showing which nuclei are observed in the spectra and lowercase letters showing which nuclei participate in the magnetization transfer but are not observed in the spectra. The strategy for making assignments centers around what is called a 'backbone walk,' which links together peaks between neighboring residues across the sequence of a protein. Assignments can then be transferred to the ^{1}H-^{15}N HSQC spectrum.

Chemical shifts are sensitive to backbone φ and ψ torsion angles. As a result, chemical shifts can be used for torsion angle predictions. A protein torsion angle is formed from three consecutive bonds linking four atoms and defined by the angle between the two outer bonds when observed along the length of the middle bond. φ is observed along the ^{15}N-$^{13}C\alpha$ bond and ψ is observed along the $^{13}C\alpha$-^{13}CO bond. Amino acid identity heavily influences the range torsion angles can adopt, allowing torsion angle predictions to be used as angular constraints in structure calculations. Furthermore, torsion angles observed in structural models are used to validate model accuracy.

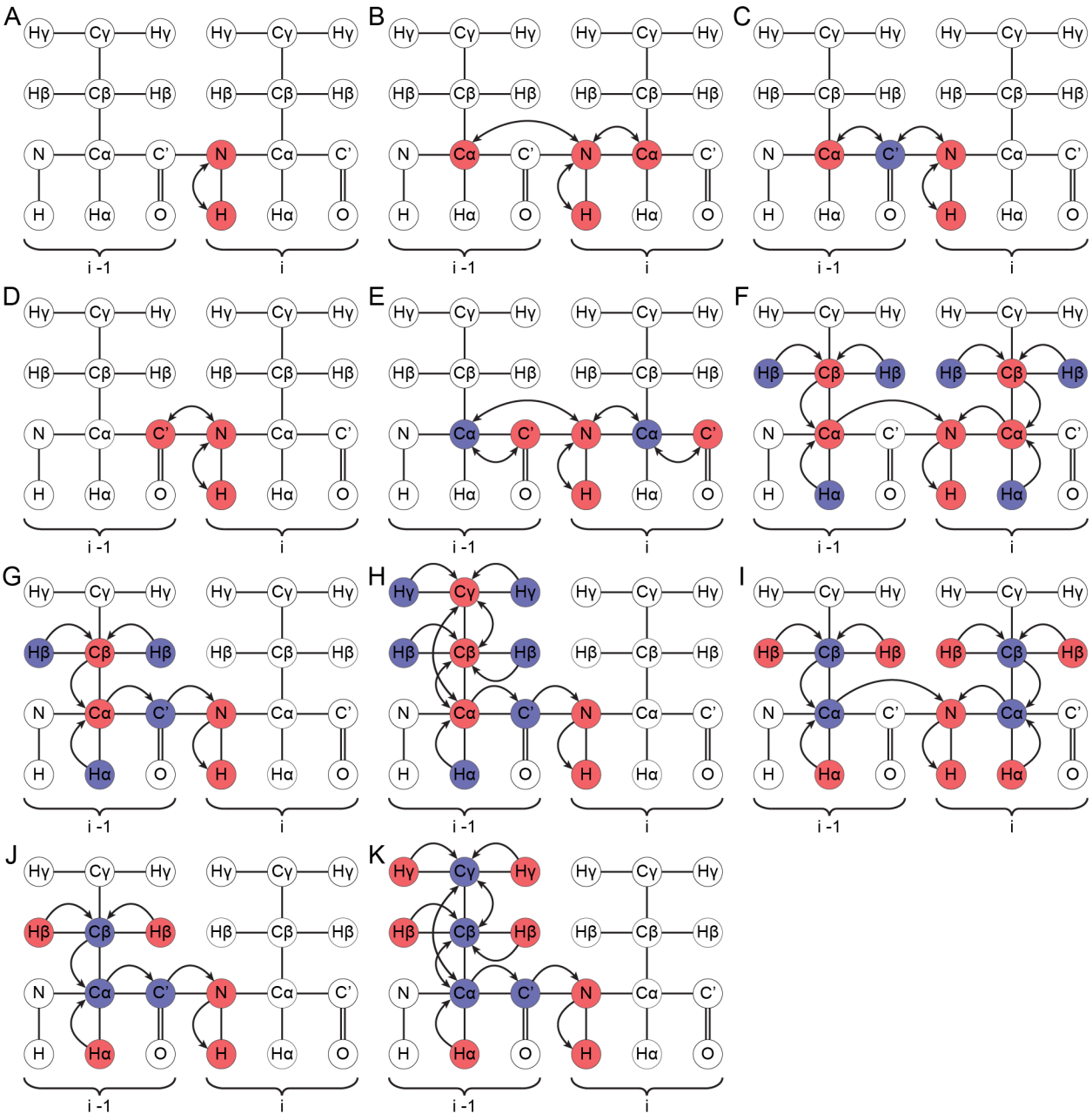


Figure 15. Diagrams of 2D and 3D experiments used to assign PopZ$^{\Delta 134\text{-}177}$. Red represents nuclei whose resonances are observed in the spectrum. Blue represents nuclei that participate in magnetization transfer but whose resonances are not observed in the spectrum. Arrows show the direction of magnetization transfer. (A) ^{1}H-^{15}N HSQC, (B) HNCA, (C) HNcoCA, (D) HNCO, (E) HNcaCO, (F) HNCACB, (G) CBCAcoNH, (H) CCcoNH, (I) HBHANH, (J) HBHAcoNH, and (K) HccoNH.

3.3.2. Assignment of PopZ$^{\Delta 134\text{-}177}$ NMR spectra

2D and 3D high resolution solution NMR spectra were acquired using a 700 or 800 MHz NMR spectrometer of the truncated PopZ$^{\Delta 134\text{-}177}$, which includes the first 133 residues followed by a Leu-Glu linker and 6xHis-tag. All data were processed using NMRPipe[78] and analyzed using Sparky from NMRFAM[79]. Uniform peak widths and intensities of NMR spectral peaks indicate PopZ$^{\Delta 134–177}$ was in a stable state necessary for NMR study. The intrinsic disorder of PopZ$^{\Delta 134–177}$ is shown by the 2D ^{1}H-^{15}N HSQC spectrum with the characteristic narrow ^{1}H chemical shift dispersion typically seen for IDPs, with the exception of side chain amine and amide resonances from arginine and glutamine residues, respectively. Comparatively, ordered proteins tend to have a wider dispersion of ^{1}H resonances than proteins without a well-defined fold[98]. The disordered nature of PopZ$^{\Delta 134–177}$ is further supported by the lack of long-range interactions in ^{1}H-^{15}N NOESY data.

Standard triple resonance experiments seen in Figure 15 above were performed for PopZ backbone and side chain resonance assignments such that 84.7% of the protein backbone was assigned, including the Leu-Glu linker on the C-terminus of PopZ$^{\Delta 134–177}$ (Figure 16). For simplicity, the linker residues are referred to as L134 and E135 in assignments.

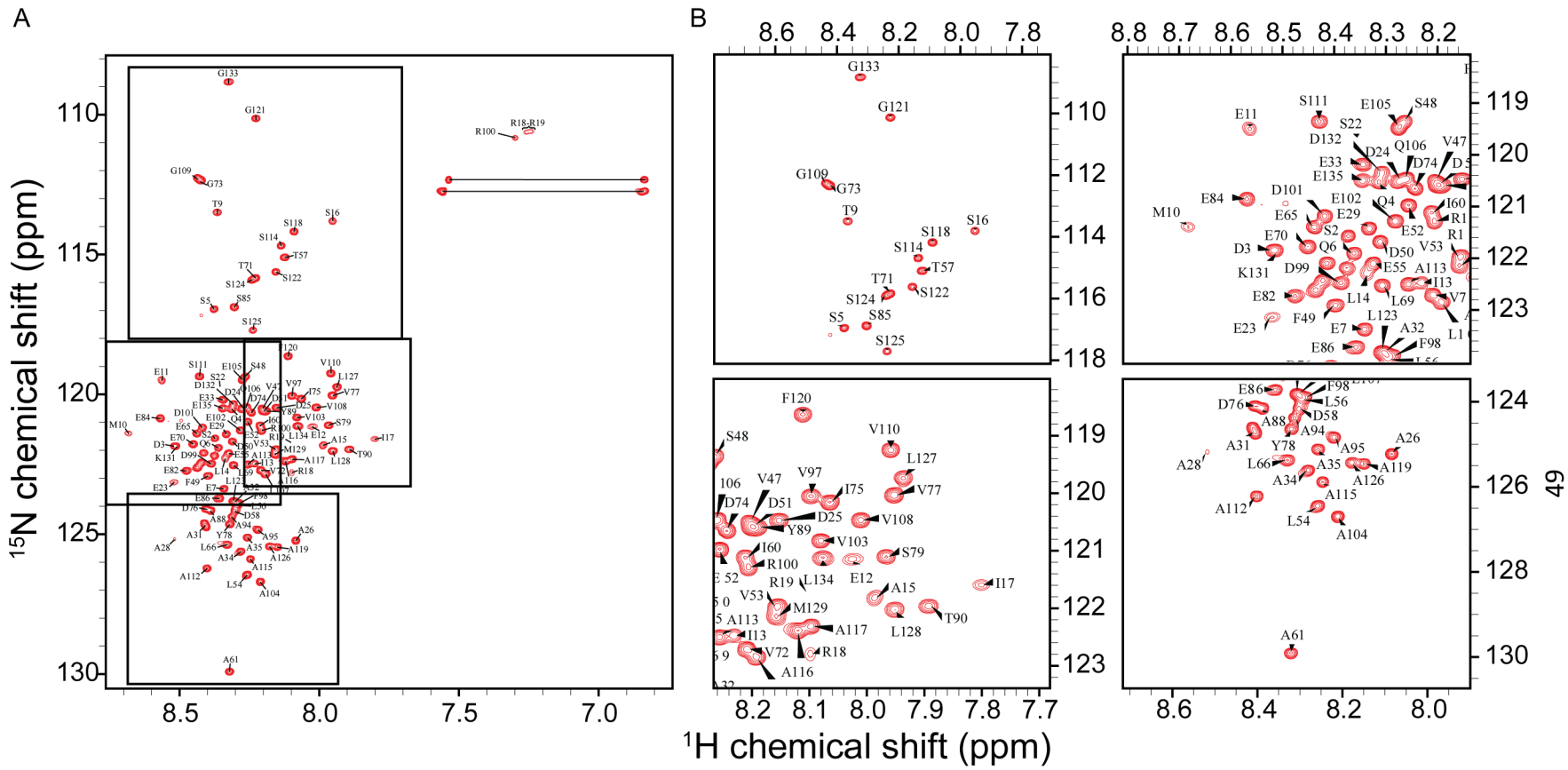


Figure 16. PopZ$^{\Delta 134\text{-}177}$ NMR resonance assignments. (A) 2D ^{1}H-^{15}N HSQC NMR spectrum of PopZ$^{\Delta 134\text{-}177}$ in its unbound state. Narrow dispersion of ^{1}H chemical shifts between 9.0 and 7.5 ppm is indicative of an IDP. Peaks connected by horizontal lines are glutamine side chain peaks. (B) The enlarged regions of the 2D ^{1}H-^{15}N HSQC NMR spectrum as highlighted in (A).

A full list of the assignments, representative strip plot for making assignments, and schematic showing backbone assignment completeness is seen in Table A4, Figure 17, and Figure 18, respectively. Most of the missing assignments were due to the abundance of proline residues, which account for 18% of the PopZ$^{\Delta 134–177}$ sequence. We were able to obtain partial conformational information on some of the proline residues by calculating the differences between proline $^{13}C\beta$ and $^{13}C\gamma$ chemical shifts[99]. This analysis indicates that 11 of the 25 proline residues are linked to the preceding amino acid by a peptide bond that is in the *trans* conformation. Additionally, we observed a number of low-intensity peaks in the ^{1}H-^{15}N HSQC spectrum that indicate small subpopulations of PopZ$^{\Delta 134–177}$ conformers. Many of the peaks corresponded to residues next to proline, suggesting that these subpopulations are comprised of species with peptide bonds in the *cis* configuration.

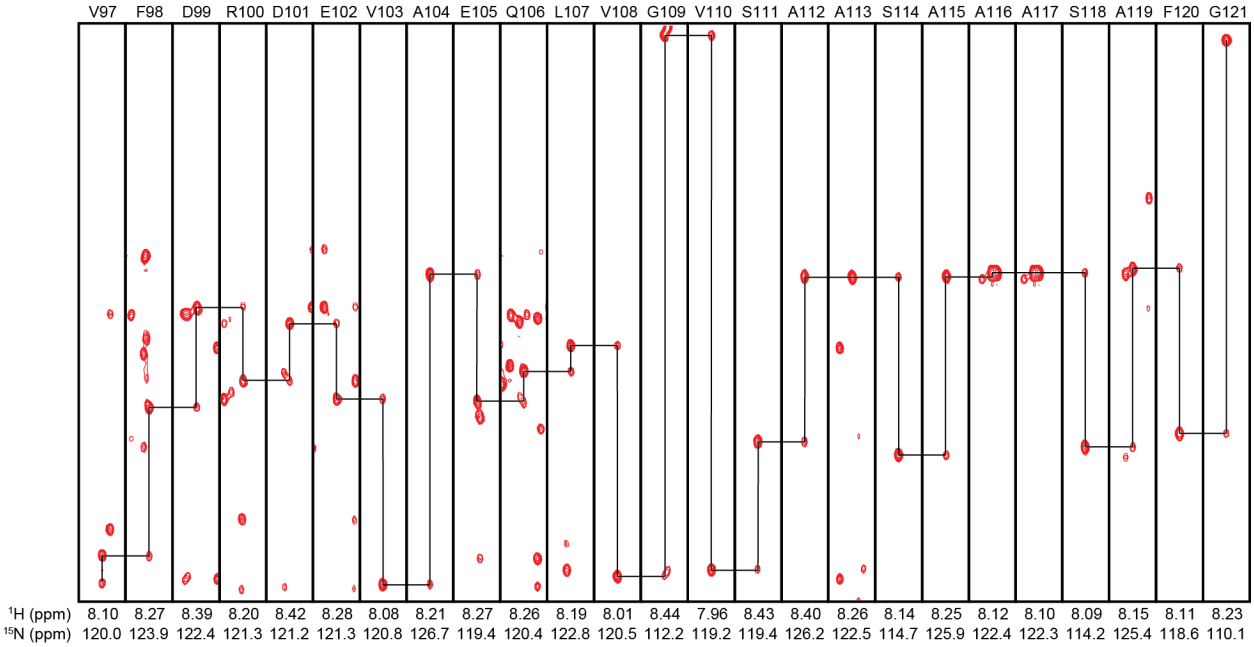


Figure 17. HNCA strip plot for PopZ$^{\Delta 134–177}$. Representative strip plot of a 3D HNCA spectrum for PopZ$^{\Delta 134–177}$ showing backbone correlations between residues V97 and G121. The $^{13}C\alpha$ peaks of the *i* and *i*-1 residues are connected to guide the eye.

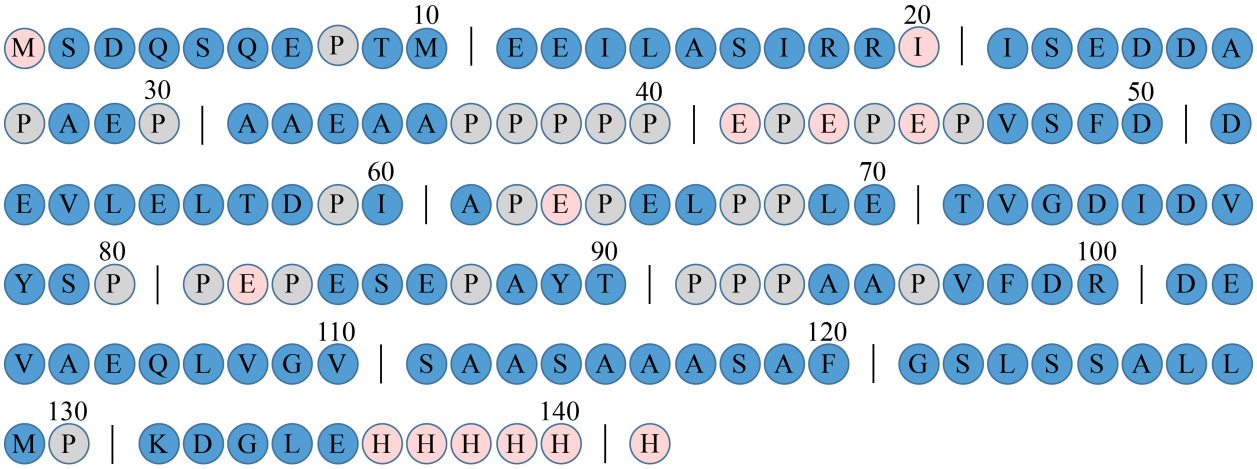


Figure 18. Schematic showing backbone ^{1}H and ^{15}N assignment completeness across the PopZ$^{\Delta 134–177}$ sequence. 70.3% of ^{1}H and ^{15}N backbone nuclei were assigned. Blue residues are assigned, pink residues are unassigned, and grey residues (proline) cannot be assigned.

3.3.3. Secondary chemical shifts

Assignments can be used to give information regarding the secondary structure of a protein. Chemical shifts are very sensitive to their local environments, so temperature, salinity, pH, and secondary structures heavily influence the chemical shift range at which nuclei will be observed[90]. When the average chemical shift of a random coil for a particular nucleus is subtracted from the observed chemical shift of the sample, the difference is called the secondary chemical shift and its value is indicative of the likelihood a residue is adopting a particular secondary structure. ^{13}Cα, ^{13}CO (carbonyl), and ^{1}Hα nuclei are particularly sensitive to their secondary structure environment and, by extension, their φ/ψ dihedral angles[90]. For example, positive secondary chemical shifts in ^{13}Cα and ^{13}CO are indicative of helical structure while negative secondary chemical shifts are indicative of β-sheet structure[90,95,100], as seen in Table 2.

Table 2. Trends observed between secondary chemical shift sign and secondary structure for protein nuclei.

	Nucleus					
Secondary Structure	$^{13}C\alpha$	$^{13}C\beta$	^{13}CO	^{15}N	^{1}HN	$^{1}H\alpha$
α-helix	+	–	+	–	–	–
β-sheet	–	+	–	+	+	+

3.3.4. PopZ$^{\Delta 134\text{-}177}$ secondary chemical shifts

$^{13}C\alpha$ secondary chemical shifts in PopZ$^{\Delta 134\text{-}177}$ spectra showed a continuous range of positive values between N-terminal residues T9-D25, while $^{13}C\beta$ and $^{1}H\alpha$ secondary chemical shifts showed a range of negative values in this same region. These data were plotted in a single structural propensity (SSP) plot[101] that indicates a helical motif in this part of the protein (Figure 19). Secondary chemical shifts tended to be more randomly oriented in the middle of the sequence (residues 46-100), as expected for an IDP. In the C-terminal region (residues 101-134), we observed a consistent, but small, α-helical propensity. In applying our data to the structural analysis program Ponderosa-C/S[102], we did not find evidence for a well-defined α-helix in the C-terminal region, although the consistent pattern of secondary shifts could indicate transient helical character.

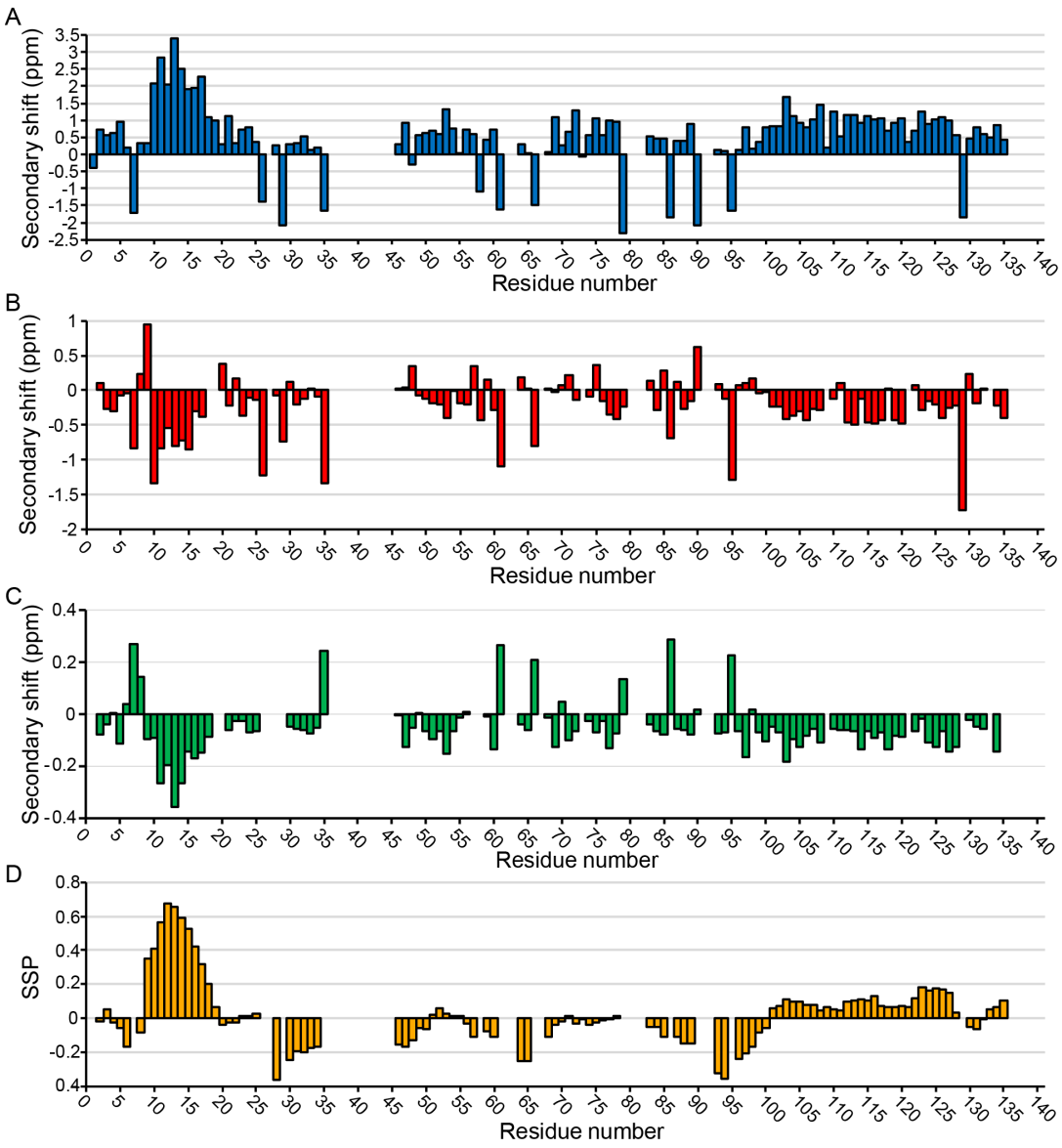


Figure 19. NMR secondary chemical shifts of PopZ$^{\Delta 134\text{-}177}$. The backbone (A) Cα, (B) Cβ, and (C) ^{1}Hα secondary chemical shifts observed for PopZ$^{\Delta 134\text{-}177}$. Large positive values for Cα and negative values for Cβ and ^{1}Hα secondary chemical shifts near the N-terminus are indicative of an α-helix. Values were generated using NMRFAM-Sparky. (D) Single structure propensity (SSP) plot derived from the data presented in (A-C) showing predicted secondary structure domains across the PopZ$^{\Delta 134\text{-}177}$ sequence, where positive values represent helical propensity and negative values represent sheet propensity.

3.3.5. Long-distance constraints

3.3.5.1. Nuclear Overhauser Effect spectroscopy (NOESY)

Chemical shifts observed in spectra arise from the coupling of nuclei through bonds. Therefore, the information obtained is representative of intra-residue or short inter-residue correlations which gives good data on secondary structure characteristics. However, this does not reveal the higher order folding experienced by the protein. The measurement of long-distance interactions (i.e. across several residues or domains) can be obtained using through-space, as opposed to through-bond, experiments such as the ^{15}N-edited Nuclear Overhauser Effect Spectroscopy (NOESY)-HSQC and ^{13}C-edited NOESY-HSQC experiments[92]. NOESY-HSQC experiments, and others like them, use the NOE to observe interactions between nuclei through space.

NOE relies on the dipolar coupling and subsequent dipolar relaxation between two nuclei. If two nuclei are close to one another spatially, even if they are far apart sequentially, a spectral cross-peak representing their interaction will appear in the spectrum[94]. The NOE phenomenon is dependent on the distance between the two nuclei, and is therefore observed up to a distance of approximately 5 Å between nuclei[100,103]. Past that distance, the signal becomes increasingly difficult to detect. Combining all the long-distance interactions creates long-distance constraints in the protein structure that help to elucidate how secondary structure domains observed from backbone and side chain experiments organize into the tertiary structure of the protein[94,104]. However, an IDP is characterized by a lack of long-range interactions because it lacks a constrained tertiary structure and samples through a variety of conformers that differ in shape. As expected, few long-range interactions were observed in PopZ$^{\Delta 134\text{-}177}$ NOESY data. However,

multiple correlations observed in the ^{15}N- and ^{13}C-edited NOESY-HSQC spectra support the existence of the N-terminal helix (Figure 20) including 45 sequential ($|i - j|=1$) and 13 medium-range ($|i - j| \leq 4$) correlations. The short-range and medium-range distances observed were used as constraints for structural calculations.

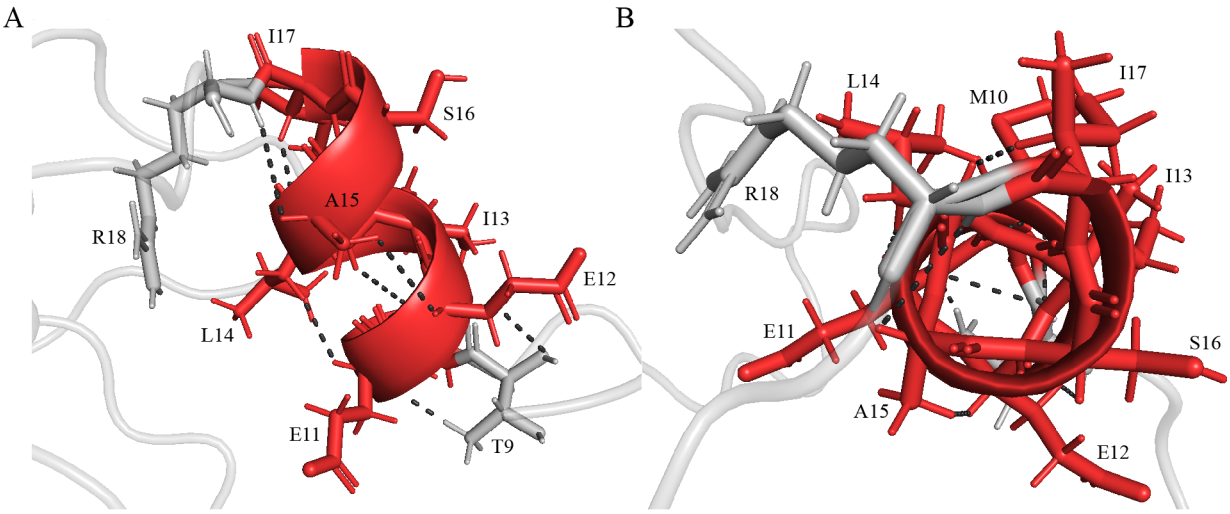


Figure 20. PopZ$^{\Delta 134-177}$ NOE correlations in the helical region. Through-distance correlations observed in ^{15}N- and ^{13}C-edited NOESY-HSQC spectra support the existence of the α-helix. Red represents the α-helix and grey represents random coil. Helix residues and the residues on either side of the helix are shown in stick form. Dashed lines represent NOE correlations between residues greater than *i*+1 away from one another. Correlations are observed from the perspective of (A) perpendicular to the helix and (B) parallel to the helix.

3.3.5.2. Paramagnetic relaxation enhancement (PRE)

Despite IDPs lacking global tertiary structure, they can still feature transient long-range contacts or molten globule-like structures that NOESY experiments are unable to detect[105]. The dipolar coupling that gives rise to the NOE phenomenon is not only dependent on distance, but on the identity of the particle being observed. Nuclei are

inherently weak magnets which lead to the 5 Å distance limitation previously mentioned for NOESY experiments[103] (Figure 21). Electrons, on the other hand, are much stronger magnets with a gyromagnetic ratio over 600 times greater than proton[106]. Utilizing an unpaired electron in paramagnetic relaxation enhancement (PRE) studies can increase the distance limitation up to 25-35 Å depending on the paramagnetic label being used[100,103,107], which include nitroxyl radicals, Mn^{2+}, Gd^{3+}, and Cu^{2+}, among others.

PRE arises from the dipolar interactions between a nucleus and the unpaired electron, resulting in an increase in the nuclear relaxation rate. Therefore, the relaxation experienced by a nucleus near the spin label will be greater than the relaxation rate if that spin label was not present[108]. The spin label used in protein studies is generally a nitroxyl radical introduced by forming a covalent bond with a cysteine residue, often times requiring a cysteine single point mutation beforehand. Due to the enhanced relaxation experienced by nuclei, the observed ^{1}H-^{15}N HSQC spectrum peak intensities will be weaker for the activated spin label (paramagnetic) sample than what is observed in spectra of the inactive spin label (diamagnetic) sample[103]. Consistent with this, residues near the spin-label will exhibit weaker peaks than those further from the spin-label. The peak intensity ratios from the two samples provide the information needed to calculate long-range intramolecular and intermolecular distance constraints[106,109]. Extending the distance observed using PRE methods can supplement NOESY data, sample transient tertiary structures, and detect weakly populated interactions[105,106]. For example, using PRE the intrinsically disordered γ-subunit of cGMP phosphodiesterase was shown to have a flexible N-terminus region (residues 1-20), a transient tertiary structure between residues 21-30, and a compact tertiary core for the rest of the sequence[96]. PRE can also

be used to probe docking sites between a protein and its binding partner, as observed with G-actin and thymosin β4[110].

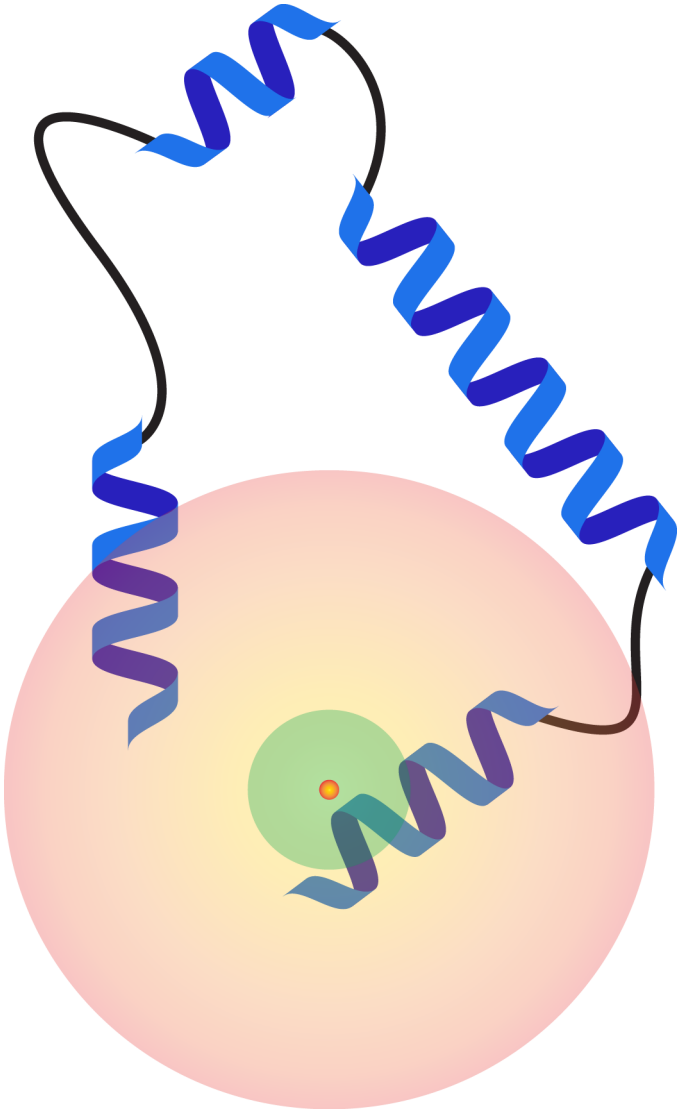


Figure 21. Distance constraints. Distance limitations observed in PRE experiments (largest sphere) are approximately four times larger than those observed in standard NOE interactions (green sphere). The center dot (red/yellow) represents a potential location of a spin label used in PRE experiments.

3.3.5.3. PopZ$^{\Delta 134\text{-}177}$ paramagnetic relaxation enhancement

We attempted to perform PRE experiments for PopZ$^{\Delta 134\text{-}177}$ to supplement distance constraints for the structure calculations. Our collaborators, Dr. Bowman's group, from the University of Wyoming provided *E. coli* harboring plasmids coding for PopZ$^{\Delta 134\text{-}177}$ cysteine single point mutations ranging across the protein sequence, including S16C, S48C, S85C, and S118C. The point mutation tested for our studies was the S48C variant. Bacterial expression and FPLC purification went as expected, though the S48C mutant exhibited higher levels of impurities than its wildtype PopZ$^{\Delta 134\text{-}177}$ variant. The PRE label used for our studies was 3-maleimido-PROXYL (mPROXYL). Attachment of the spin label to the cysteine site consisted of a three-hour reaction of PopZ$^{\Delta 134\text{-}177}$ S48C with 20-fold excess mPROXYL followed by dialysis to remove unreacted spin label. Unfortunately, the reaction caused severe aggregation of the sample as determined by UV-Vis spectrophotometry. Even after removing the aggregates using spin columns, the sample exhibited poor NMR spectra that could not be used to generate additional distance constraints.

The failure of the PRE experiment does not negate its usefulness in potential future studies. Only one of the four cysteine point mutation variants was tested, so it is possible the S48C mutant reacted unfavorably towards the mPROXYL spin label and that future attempts with the other mutants would have more success. Alternatively, a different spin label could be used, such as 5DSA, 16DSA, GD(DTPA-BMA), 4-(2-iodoacetamido)-TEMPO, other TEMPO derivatives, or other PROXYL derivatives, many of which have been used in other IDP structural studies.

3.3.6. Orientation constraints

3.3.6.1. Residual dipolar coupling (RDC)

Few long-range interactions observed in NOESY data can also be supplemented with orientation data obtained from residual dipolar coupling (RDC) experiments. Dipolar coupling, as stated above, is dependent on the distance between two nuclei, but orientation of the vector between the two nuclei with respect to the magnetic field is also important. Equation 3 describes the dipolar interaction between two magnetically active nuclei[111,112]:

(Equation 3) $$D_{ij} = -\frac{\gamma_i \gamma_j h \mu_0}{8\pi^3 r^3} \left\langle \frac{3cos^2\theta - 1}{2} \right\rangle$$

where γ_i and γ_j are the gyromagnetic ratios of nuclei *i* and *j*, respectively, h is Planck's constant, μ_0 is the permeability of a vacuum (the capability of a magnetic field to permeate a vacuum), *r* is the distance between the two nuclei, and θ is the angle of the vector between the two nuclei with respect to the external magnetic field. RDC experiments negate the distance dependence by measuring the coupling of two nuclei with fixed bond lengths (^{1}H-^{15}N), thus the measured RDC constants depend only on the orientation component[100,107]. However, under isotropic conditions the dipolar effects sample all possible orientations, so the couplings cancel out to zero[103]. To account for this, RDC experiments use an orientation aligning media, such as polyacrylamide gels, filamentous phage, liquid crystalline phase, or large bicelles, to create slightly anisotropic conditions for the protein[103,113,114] (Figure 22A). This results in a small population (~0.1%) of the protein adopting a biased orientation in the media which has a small, measurable effect

on ^{1}H-^{15}N bond vectors, hence the "residual" part of RDC[115]. These changes can be measured across the sequence of the protein from NMR spectra and the information gives orientation data relative to the direction of the magnetic field.

In NMR spectra, RDC manifests as small, finite perturbations in the perceived splitting of coupled spectra (Figure 22B)[103]. In a decoupled ^{1}H-^{15}N HSQC spectrum, each peak represents one amino acid. When undecoupled in the ^{15}N dimension, the single peaks are split into two peaks, and the distance between the peaks is the ^{1}H-^{15}N J-coupling. In aligned media, RDC constants will add or subtract from this distance. For IDPs, positive RDC constants are characteristic of order while negative constants are characteristic of disorder[71,96]. Negative couplings result from the tendency of the backbone to align with the magnetic field[103], thus making ^{1}H-^{15}N bond vectors orthogonal to the magnetic field[100]. Positive couplings are often indicative of helical character, which aligns parallel with the magnetic field and subsequently changes the direction of the ^{1}H-^{15}N vector to be aligned parallel as well[100,103]. Other bond vector RDCs can be used to describe the length and population of different helical structures among conformers, as was seen for the disordered C-terminal domain of a nucleoprotein from the *Sendai* virus, which showed that the MoRF sequence populated primarily three helical conformers[106,116].

RDCs can be used to compare structural states of a protein from NMR studies with structures derived from X-ray crystallography. The HIV-1 protease, for example, adopts a wide range of conformations consisting of open, semi-open, and closed states according to crystallography structures, but NMR RDC data showed that in solution the preferential state is closed[117]. Similarly, crystallographic structures of arginine kinase from

Limulus polyphemus shows distinct substrate-induced changes in structure between substrate-free and substrate-bound forms, but NMR RDC data indicated that in solution the substrate-free form adopts a mixed population of the two crystal structures with a preference towards the substrate-free form. This study also highlighted the ability for NMR and X-ray techniques to be used in parallel, such that NMR-generated structures refined crystallography structures and vice versa[118]. RDCs can also be used to determine the overall population distribution among conformers by comparing experimental RDCs with computational values generated from the conformer ensemble[119]. Molecular dynamics studies can also be supplemented with experimental NMR data, including RDCs, for more robust simulations[120]. Furthermore, RDCs can be induced by internal alignment using paramagnetic centers where the label is attached to one domain, and the induced RDCs on neighboring domains report on the flexibility of those domains relative to the labeled domain. In this case, flexible domains near the labeled domain will exhibit a narrow dispersion of RDCs and rigid domains will exhibit a broad distribution of RDCs (comparable to RDCs observed on the labeled domain)[107]. For example, when the N-terminal domain of Calmodulin is labeled with a lanthanide paramagnetic center, the C-terminal domain shows a narrow dispersion of RDCs, indicative of flexibility relative to the labeled domain, but a broad dispersion of RDCs when bound to its partner α-synuclein[121].

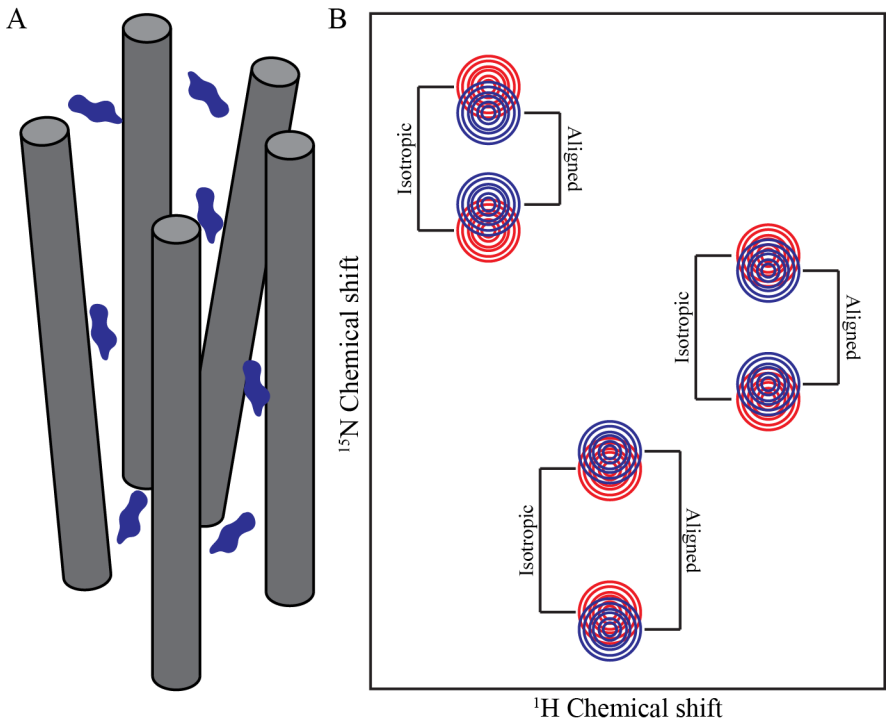


Figure 22. Residual dipolar coupling in NMR. (A) Representation of alignment media (gray cylinders) responsible for inducing partial orientation alignment in a protein sample (blue). (B) An example of a ^{1}H-^{15}N HSQC spectrum that is undecoupled in the ^{15}N dimension allowing for the measurement of J-coupling (shown as brackets). Partial alignment of a sample (blue peaks) causes changes in the measured J-coupling when compared to isotropic, or unaligned, samples (red peaks). The calculated difference in J-coupling is the RDC constant.

3.3.6.2. PopZ$^{\Delta 134\text{-}177}$ residual dipolar coupling

We obtained additional information on long-range orientation by measuring residual dipolar coupling in axially stretched polyacrylamide gels. We observed mostly negative couplings in the center of the sequence and a cluster of positive couplings in the suspected α-helical binding region, a pattern that is commonly observed for α-helices in this type of analysis[122]. The C-terminal region (residues 100-134) showed weaker couplings that were both positive and negative. We obtained dipolar coupling information

for all residues except for 25 proline residues and 20 other residue couplings that could not be measured due to overlapping NH resonances (Figure 23).

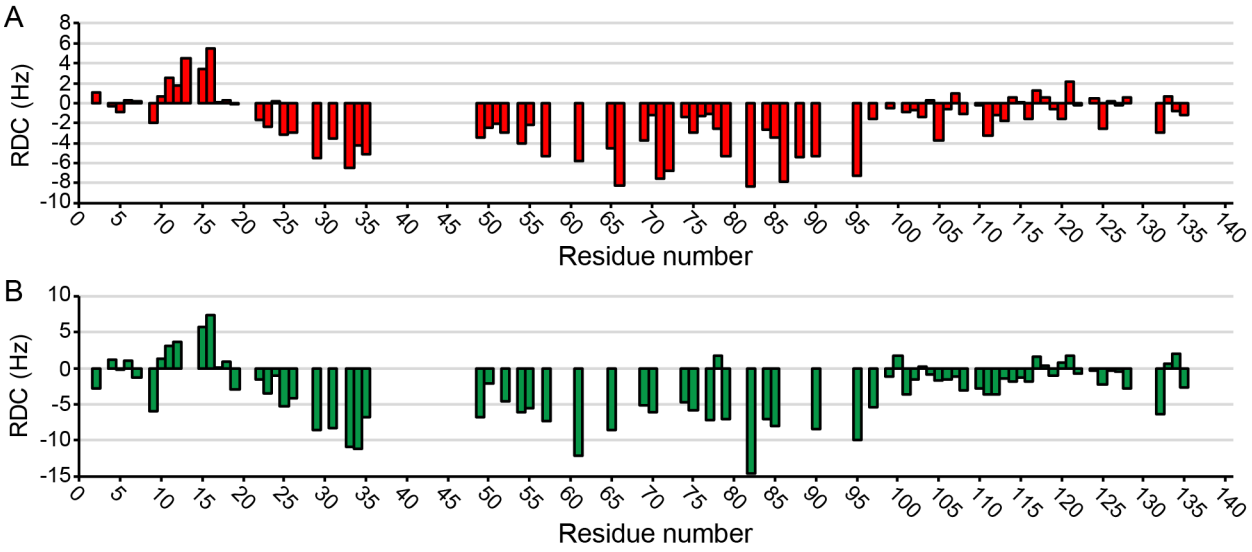


Figure 23. PopZ$^{\Delta 134\text{-}177}$ residual dipolar coupling. RDC measurements as a function of amino acid residue for a 5.4 mm (A) and 6.0 mm (B) outer diameter (OD) polyacrylamide gel axially stretched to 4.2 mm OD.

3.3.7. Structure of PopZ$^{\Delta 134\text{-}177}$ by NMR spectroscopy

Torsion angle predictions for φ and ψ were generated from TALOS-N (Table A5). Chemical shift assignments, ^{15}N-NOESY-HSQC and ^{13}C-NOESY-HSQC peak lists, and RDC restraints were uploaded to the Ponderosa-C/S server[102] for PopZ$^{\Delta 134\text{-}177}$ structure calculations. The 20 most favored structures were uploaded to the Protein Structure Validation Software[123] suite for validation. Global quality scores were good, and Ramachandran plots from MolProbity found 85.6% of all residues in most favored regions, 9.5% in allowed regions, and 4.9% in disallowed regions. 97.2% of ordered residues were in favored regions. A more detailed analysis of the validation is presented in Table 3. The

20 predicted structures show no other well folded secondary structure elements besides an N-terminus α-helix (M10-I17) (Figure 24).

Table 3. Summary of NMR and structural statistics for PopZ$^{\Delta 134\text{-}177}$.

Number of structures[a]	20		
RMSD Values[b]			
	All	Ordered[c]	Selected
All backbone atoms	18.9 Å	7.0 Å	18.9 Å
All heavy atoms	19.0 Å	7.3 Å	19.0 Å
Structure Quality Factors – overall statistics[b]			
	Mean score	SD	Z-score
Procheck G-factor (phi/psi only)	-0.79	N/A	-2.79
Procheck G-factor (all dihedral angles)	-0.44	N/A	-2.60
Verify3D	-0.01	0.0183	-7.54
ProsaII (-ve)	0.09	0.0722	-2.32
MolProbity clashscore	5.74	1.5876	0.54
Ramachandran Plot Summary from Procheck[b]			
Most favored regions	69.4%		
Additionally allowed regions	25.0%		
Generously allowed regions	3.9%		
Disallowed regions	1.7%		
Ramachandran Plot Statistics from Richardson's lab[b]			
Most favored regions	85.6%		
Allowed regions	9.5%		
Disallowed regions	4.9%		
Completeness of resonance assignments[d]			
N, H	84.08%		
N, H, C, CA, CB	84.37%		
All atoms except aromatic atoms	64.55%		
Distance constraints			
Total	976		
Short range [i – j <= 1]	918		
Medium range [1 < i – j <= 4]	42		
Long range [i – j > 4]	16		

[a]Structural statistics were calculated for the top 20 energetically favored models.
[b]Calculated using Protein Structure Validation Software (PSVS) Suite 1.5
[c]Ordered residues: 10-17
[d]Calculated using NMRFAM-Sparky
Notes: Structural validation for the top 20 most favorable models was determined using the Protein Structure Validation Software (PSVS) Suite.

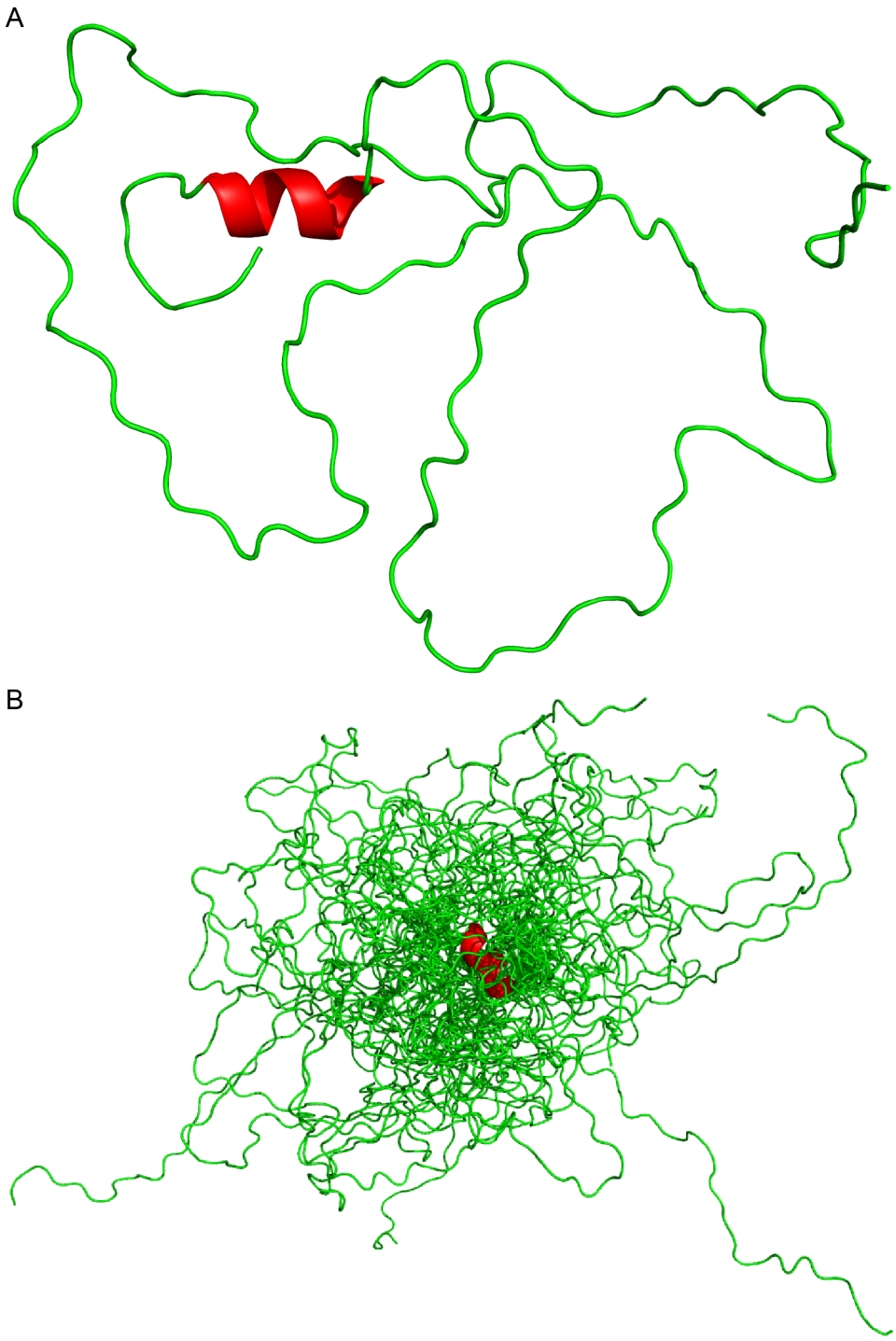
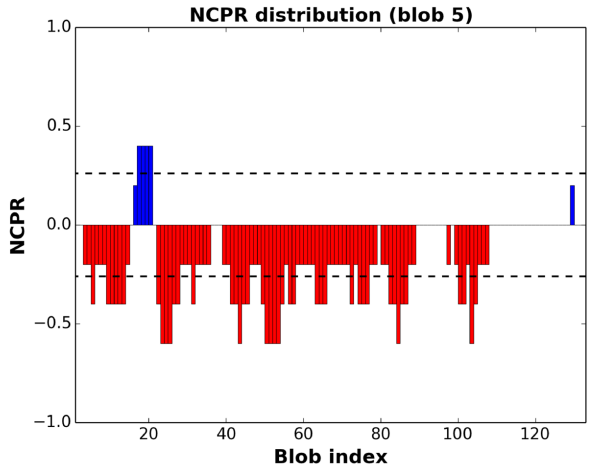


Figure 24. PopZ$^{\Delta 134\text{-}177}$ structural ensemble. (A) PopZ$^{\Delta 134\text{-}177}$ cartoon structure generated from Ponderosa-C/S and visualized using PyMOL. Green represents disorder along the sequence and red represents a short α-helical secondary structure on the N-terminus. (B) PopZ$^{\Delta 134\text{-}177}$ cartoon overlay of the top 20 best-evaluated structures generated from Ponderosa-C/S.

Additional analysis was carried out using the CIDER (Classification of Intrinsically Disordered Ensemble Regions) server[124], which is generally utilized to calculate and present the various sequence parameters commonly associated with disordered protein sequences. The analysis yielded low values for fraction of charged residues, FCR (0.256) and net charge per residue, NCPR (−0.195). Kappa (κ), a patterning parameter describing mixing vs. segregation of charged residues in the linear protein sequence was calculated to be 0.202, and omega (Ω), an analogous parameter that takes into consideration proline residues[125], was 0.244. The CIDER results indicated that (i) the net charge per residue was generally negative with the exception of a small net positive region close to the N-terminus (Figure 25A), (ii) the charged residues are relatively well-mixed with other residues throughout the protein primary sequence, and (iii) that the proline and charged residues were also well-mixed. The PopZ$^{\Delta 134–177}$ sequence is predicted by CIDER to fall within region 2 of the diagram of states (Figure 25B), the boundary region between weak and strong polyampholytes. The CIDER results are summarized in Table 4.

A



B

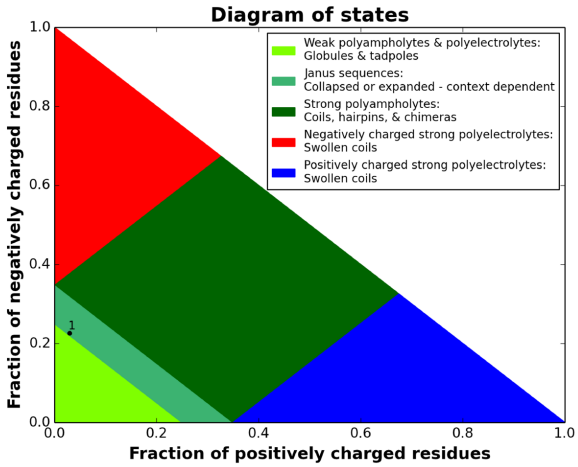


Figure 25. CIDER results for PopZ$^{\Delta 134\text{-}177}$. (A) Net charge per residue (NCPR) plot showing negative charge across most of the sequence with the exception of a small region with a net positive charge in the PopZ $^{\Delta 134–177}$ MoRF. (B) Diagram of states showing that the PopZ$^{\Delta 134–177}$ sequence (labeled as 1) falls within the boundary region between weak and strong polyampholytes. CIDER calculations were performed using the server http://pappulab.wustl.edu/CIDER/analysis/.

Table 4. Summary of CIDER results for PopZ$^{\Delta 134\text{-}177}$.

Parameter	Output
Length	133
f-	0.226
f+	0.030
FCR	0.256
NCPR	-0.195
κ	0.202
Ω	0.244
Σ	0.149
Δ	0.037
Max Δ	0.181
Hydropathy	4.145
Disorder Promoting	0.767
Plot region	2
Fraction of proline	0.19 (high)

3.4. Dynamics

3.4.1. Longitudinal and transverse relaxation

NMR relaxation is the phenomenon of a nuclear spin system returning to thermal equilibrium following some sort of perturbation (i.e. radiofrequency pulse)[126,127]. Two important kinds of relaxation are longitudinal and transverse relaxation. Following a radiofrequency pulse, longitudinal relaxation describes the process of spin populations returning to their Boltzmann distribution by dissipating their energy to their surroundings. This results in the bulk magnetization reestablishing to the z-dimension. Transverse relaxation describes the loss of spin coherence in the xy-plane between all the nuclei[91,92,126] (Figure 26). Longitudinal and transverse relaxation are characterized by the time constants T_1 and T_2, respectively, and the rate constants R_1 and R_2, respectively. R_1 and R_2 are the inverses of T_1 and T_2, respectively[126].

Global relaxation time and rate constants can be used to calculate the rotational correlation time (τ_c) of a protein, which is the amount of time it takes for a protein to rotate one radian[92], otherwise known as tumbling. The process is dependent on the size of the protein, the viscosity of the liquid it is in, and the temperature. This can give vital information on the oligomeric state of a protein. For example, if a protein has a molecular weight of 15 kDa, but a rotational correlation time corresponding to a protein twice that size, then this is indicative that the protein occurs as a dimer in solution. However, the calculation for τ_c assumes the protein to be rigid and spherical[128]. Therefore, these calculations are important for globular proteins, but are not reliably accurate for IDPs which rapidly transition between multiple conformers of varying shape.

T_1 and T_2 can still be measured on a per residue basis for IDPs. The information obtained can give insight into secondary structure characteristics. Structured domains tend to have shorter T_1 and T_2 values, and therefore longer R_1 and R_2 rate constants, than disordered domains[71,96]. This could be due to the more restricted flexibility or restricted conformational motions found in ordered regions[71]. This is consistent with the correlation of increased R_2 rates for bulkier residues, which could indicate local restrictions in proteins due to steric interactions between side chains and backbone[71,129]. Structural rigidity can also be indicated by an increase in R_2/R_1 ratios[130].

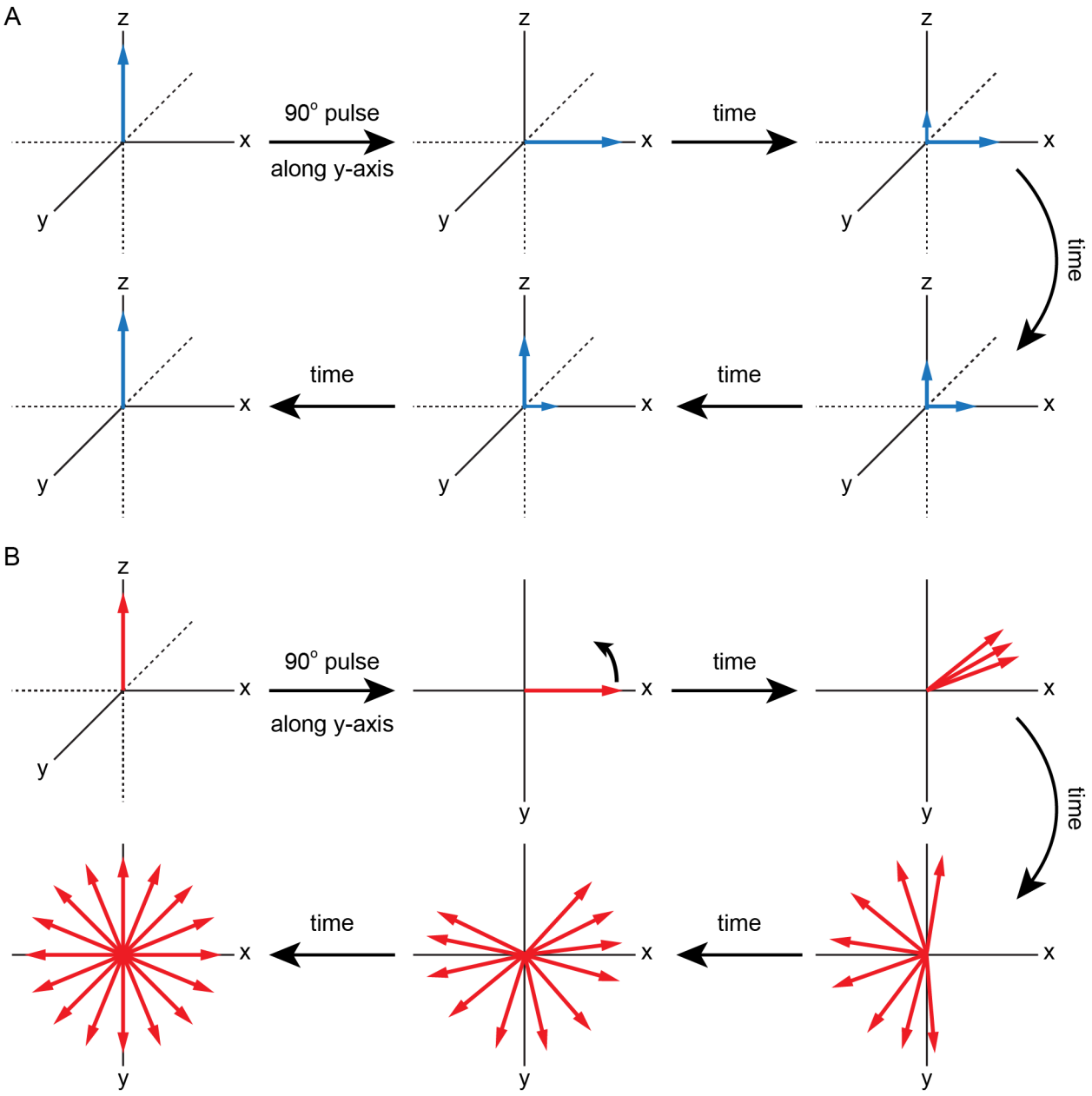


Figure 26. T_1 and T_2 relaxation. (A) T_1 relaxation is the process by which the net magnetization returns to the z-axis following a perturbation of some kind (i.e. 90° pulse). (B) T_2 relaxation where individual nuclear spins lose coherence over time until the magnetic moments in the xy plane cancel out. Note the change in perspective following the 90° pulse (looking down z-axis).

3.4.2. Hetero NOE

Relaxation is very closely associated with the Nuclear Overhauser Effect (NOE), the phenomenon by which nuclei transfer nuclear spin polarization to neighboring nuclei by an amount proportional to the inverse 6^{th} power of the distance between the nuclei[91]. The ^{1}H-^{15}N heteronuclear NOE (hetNOE) experiment can give important information regarding secondary structure conformational states, where small ratios that fluctuate between positive and negative values constitute more disorder and larger, positive ratios constitute more order in the protein backbone[95,96].

3.4.3. PopZ$^{\Delta 134–177}$ dynamics

To obtain information of structure dynamics, we determined hetero NOE ratios (Figure 27A) and longitudinal (T_1) and transverse (T_2) relaxation times, as well as their respective R_1 and R_2 relaxation rates (Figure 27B and 27C, respectively), for the majority of the PopZ$^{\Delta 134–177}$ sequence. Global T_1 and T_2 values could not be accurately determined due to the disordered nature of the protein. Hetero NOE values showed primarily small magnitudes of negative sign. However, a range of positive NOE values, typically seen with α-helical character, was observed between residues T9 and R19. Additionally, we observed shorter T_1 relaxation times for T9-R19 relative to other residues, and shorter T_2 times across T9-E23 were also observed, and these local decreases are indicative of α-helical character.

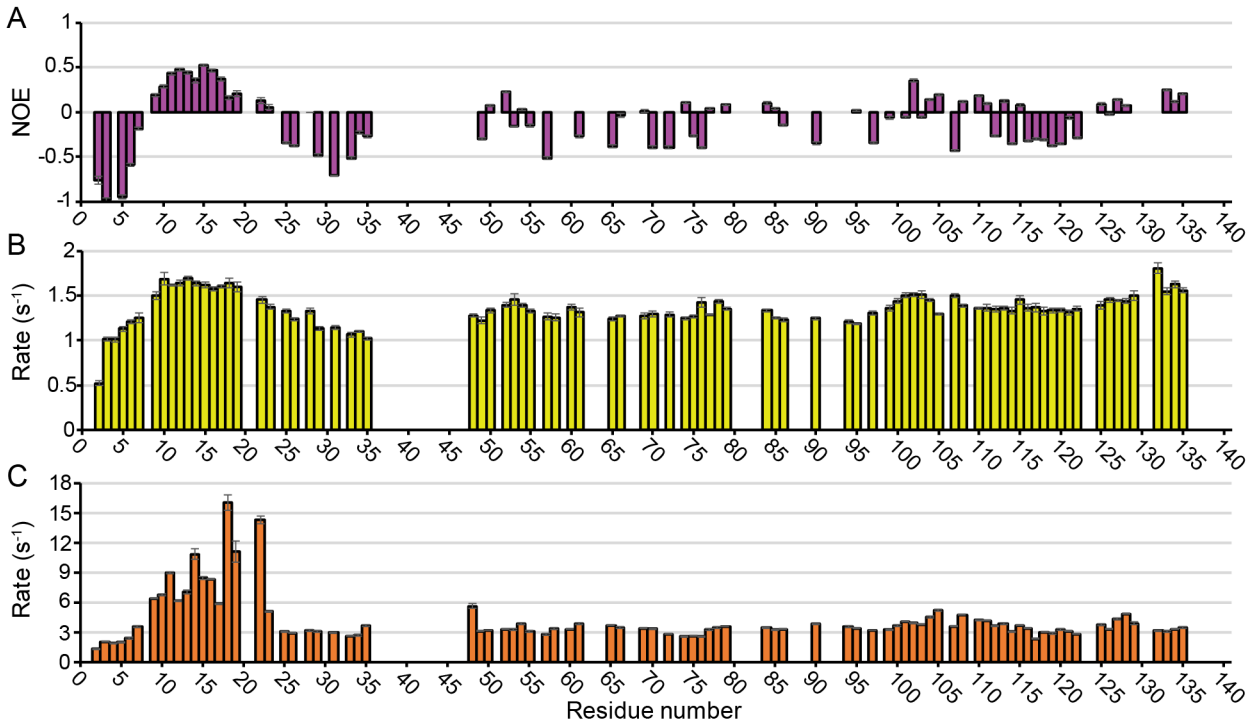


Figure 27. PopZ$^{\Delta 134\text{-}177}$ dynamics (A) Heteronuclear NOE ratios observed as a function of amino acid residue. Positive values near the N-terminus are indicative of an ordered secondary structure. (B) R_1 relaxation rates and (C) R_2 relaxation rates observed as a function of amino acid residue. Larger values near the N-terminus are indicative of ordered secondary structure.

CHAPTER 4

POPZ$^{\Delta 134\text{-}177}$ BINDING CHARACTERIZATION

This chapter will focus on binding studies performed of PopZ$^{\Delta 134\text{-}177}$ with the RcdA protein binding partner. It will begin with an introduction into protein interaction studies using NMR spectroscopy (**section 4.1**), followed by the binding interactions observed between PopZ$^{\Delta 134\text{-}177}$ and RcdA (**section 4.2**), and will conclude with binding interactions between the I17A PopZ$^{\Delta 134\text{-}177}$ point mutant and RcdA (**section 4.3**).

4.1. Introduction to protein binding interactions using NMR spectroscopy

Two phenomena are used to identify binding between a protein and a ligand: spectral shifts and spectral broadening. Chemical shifts are highly sensitive to their local microenvironment. If conditions around a nucleus change, there will likely be a change, or perturbation, in that nucleus' observed chemical shift[131]. When a protein binds to a ligand, the nuclei found in the binding domain will undergo a spectral shift to reflect the binding interaction, allowing spectroscopists to identify specific amino acid residues found in the binding motif. Peaks that shift more than others are generally considered to have stronger binding than those that shift less. Fast exchange between the bound and unbound state will result in a gradual, stepwise shift of the peaks with increasing concentration of ligand, resulting from the population average. Slow exchange will result in a one-step shift resulting in two peaks for each interacting residue whose intensities

are relative to the populations of bound and unbound states[95]. However, it should be noted that shift perturbations can also be a result of allosteric changes induced from conformational changes[131,132], and it can be difficult to distinguish spectral changes from true binding or from these indirect conformational changes.

If the binding ligand is sufficiently large (i.e. a protein), the second phenomenon will become much more significant in the NMR spectra. Sharp, distinct NMR peaks arise when adequate tumbling (rotation in solution) occurs for the sample. Tumbling rate is directly correlated with molecular size, so large proteins tumble slower than small proteins[128]. When a protein binds to a large ligand, the protein's binding domain effectively increases in size and tumbles slower. This will cause binding peaks to broaden and decrease in intensity in NMR spectra[133]. Oftentimes, the broadening occurs to such a degree that binding peaks will broaden into the noise and visually disappear from the spectra. Broadening can also be observed if the exchange rate between the bound and unbound state is intermediate (between fast and slow exchange)[95].

4.2. PopZ$^{\Delta 134\text{-}177}$ interaction with RcdA

We and others have previously shown that PopZ binds to at least ten different binding partners[36,134-136]. Additionally, when ^{15}N-enriched PopZ$^{\Delta 134\text{-}177}$ was mixed with saturating concentrations of two different unlabeled protein binding partners, we observed that both protein-bound conformations exhibited a nearly identical set of changes in 2D ^{1}H-^{15}N HSQC spectra. This suggests that PopZ$^{\Delta 134\text{-}177}$ interacts with these two binding partners via the same set of amino acids. In this study, we have continued the binding experiments with one of those binding partners, RcdA. In *C. crescentus*, RcdA is co-

localized with PopZ at one of the cell poles at the time of chromosome replication initiation[137], and it serves as an adaptor that interacts with specific protein substrates and links them to the ClpXP protease for degradation *in vitro*[138] and *in vivo*[136,137]. PopZ interaction appears to be functionally separable from protease adaptor activity, suggesting that RcdA has multiple binding interfaces[139,140]. By observing ^{1}H-^{15}N HSQC spectra of isotopically labeled PopZ$^{\Delta 134\text{-}177}$ under multiple RcdA concentrations and at a higher magnetic field (700 MHz), we obtained sufficient resolution for accurate peak identification even in congested regions of the spectra.

2D ^{1}H-^{15}N HSQC spectra were acquired from 130 μM ^{15}N-enriched PopZ$^{\Delta 134\text{-}177}$ in the presence of unlabeled RcdA binding partner at a range of concentrations between 0 and 960 μM (Figure 28). Given the different buffer conditions compared to structural studies, peak assignments were confirmed using 3D spectra for peaks that shifted under the binding conditions. Most PopZ$^{\Delta 134\text{-}177}$ NMR peaks remained unchanged in all binding conditions, indicating residues that do not interact with RcdA. For a minority of PopZ$^{\Delta 134\text{-}177}$ residues, increasing concentration of RcdA caused chemical shift perturbation and spectral broadening, and in some cases led to severe signal attenuation or loss of detection. Those peaks that display the greatest shifts and broadening likely indicate residues that interact directly with RcdA, while peaks that display moderate signal attenuation and chemical shift perturbation likely correspond to amino acids that participate in secondary or indirect binding interactions. Binding residues were determined by comparing combined ΔHN chemical shifts for each residue, where peaks undergoing a shift greater than the standard deviation were considered binding[131].

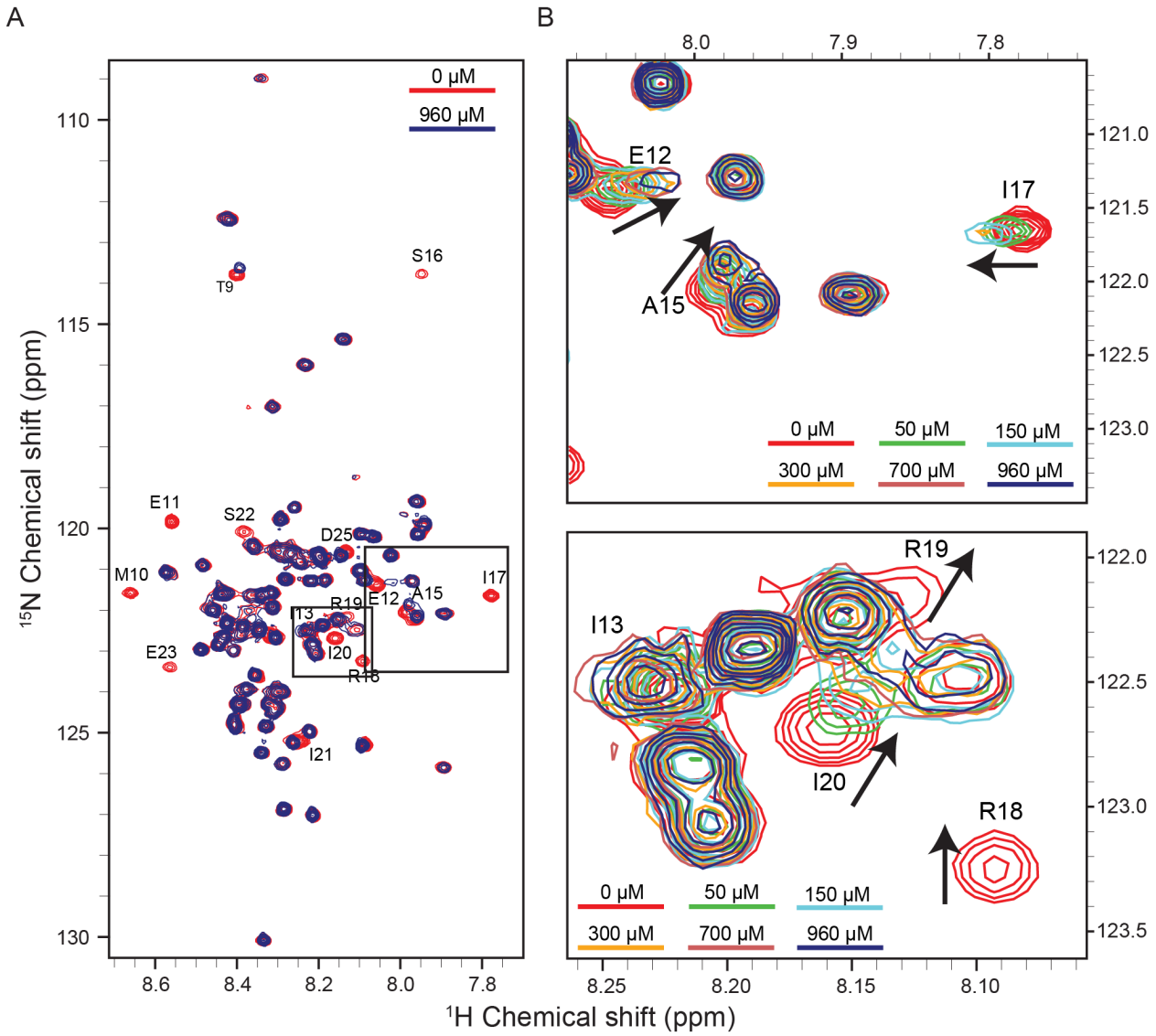


Figure 28. Binding of PopZ$^{\Delta 134\text{-}177}$ to RcdA. 2D ^{1}H-^{15}N HSQC spectra overlay of ^{15}N-enriched PopZ$^{\Delta 134\text{-}177}$ in solution with and without the RcdA binding partner exhibit spectral changes upon binding. Colored bars show RcdA concentration for each spectrum. (A) 130 μM PopZ$^{\Delta 134\text{-}177}$ with differing concentrations of RcdA: 0 μM (red) and 960 μM (blue). Peaks undergoing significant change are labeled. For simplicity, the spectra were shown without the sidechain region. Boxes represent regions seen in (B). (B) Enlarged regions highlight changes upon increasing RcdA concentration. Arrows indicate direction of peak shift. Binding of RcdA resulted in the chemical shift perturbation and significant broadening of a number of peaks. The contours were lowered in the spectra with high concentration of RcdA to show weak peaks.

The binding motif of PopZ$^{\Delta 134\text{-}177}$ was revealed to be between T9-E23 (Figure 29 and 30) using combined ΔHN chemical shift perturbations (Figure 31A) and peak intensity perturbations (Figure 31B) of PopZ$^{\Delta 134\text{-}177}$ upon binding to RcdA. Peaks corresponding to these residues undergo both significant chemical shift perturbations and line broadening most likely due to direct interaction with RcdA, although potential secondary structure changes cannot be excluded. Residue D25 is expected to experience indirect binding effects, as it exhibits chemical shift perturbation but no significant broadening. Notably, the α-helix (M10-I17) determined by our structural model is found within the RcdA binding region.

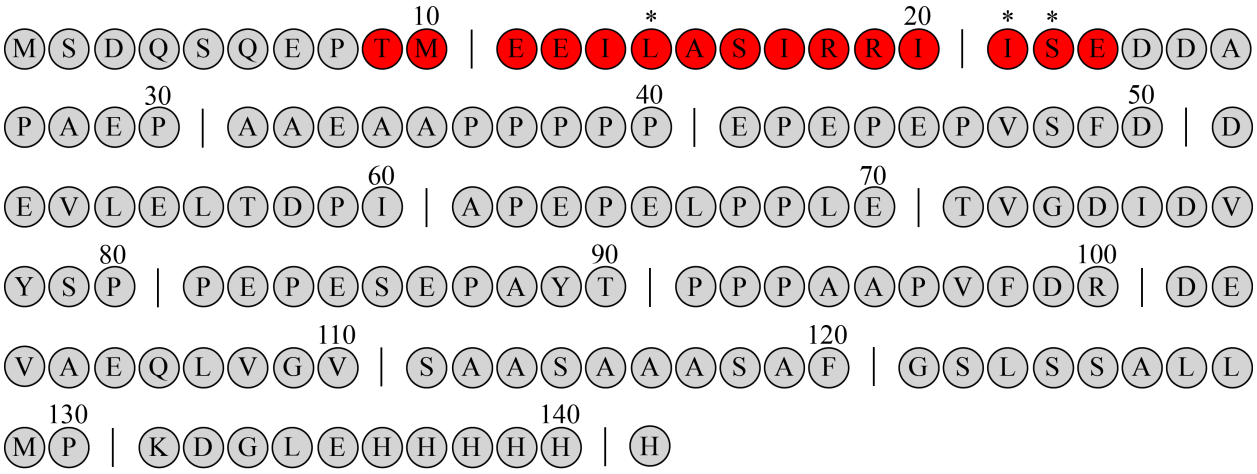


Figure 29. Schematic representation of the PopZ$^{\Delta 134-177}$ binding region. Asterisks indicate peaks that have undergone significant perturbation, but the exact value cannot be determined due to overlap.

While most perturbed peaks broadened beyond detection, some peaks were still detectable in the baseline when we significantly lowered the contour levels of the spectra (Figure 28B). Residues L14, I21, and S22 are likely binding as well, but their exact

chemical shift perturbation could not be determined due to extensive overlap with neighboring peaks. T9 and E23 (residues at the edge of the binding region) showed only a moderate amount of chemical shift perturbation and signal broadening, which is likely induced by secondary or indirect binding effects. D24 showed little change. The chemical shift perturbations that we observed in these experiments are closely matched with those produced by another binding partner, ChpT[36], indicating that the same subset of amino acids in PopZ$^{\Delta 134\text{-}177}$ participate in binding to both partners. We therefore conclude that residues T9-E23, which include an α-helix that spans M10-I17, act as a MoRF region that is directly responsible for interacting with at least two of its binding partners. Similar observations of chemical shift perturbations have been reported in the bound conformations of other MoRFs, including one in the C-terminal domain of p53[59,141,142]. Interestingly, CIDER results indicated that the net charge per residue across the PopZ$^{\Delta 134\text{-}177}$ sequence is generally negative with the exception of a small net positive region in the MoRF region due to residues R18 and R19 (Figure 25A). This could suggest that the dispersion of positive and negative charge within the binding region is critical for the electrostatic interaction between PopZ and its protein binding partners.

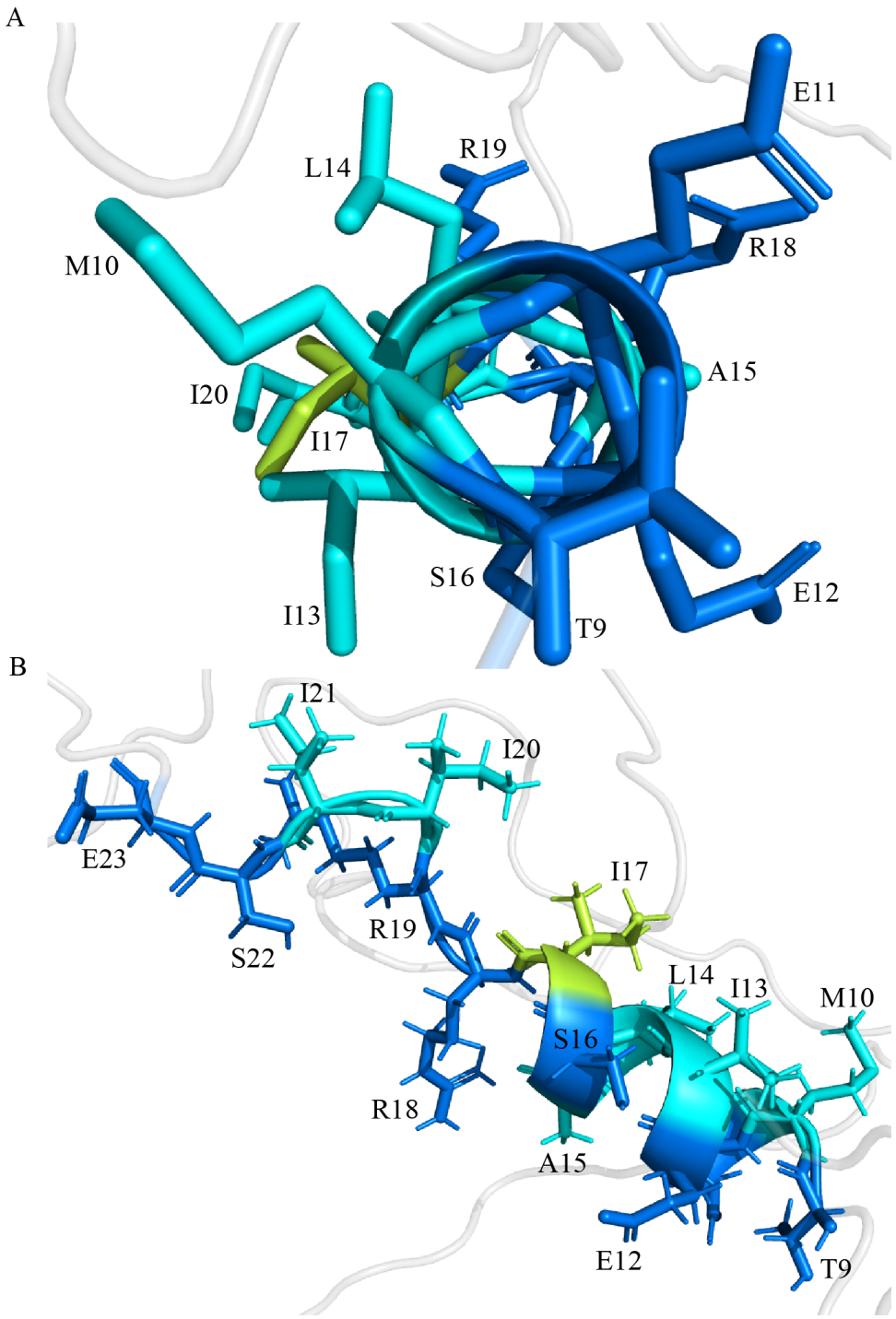


Figure 30. Binding motif of PopZ$^{\Delta 134–177}$. (A) The amphipathic nature of the helix is shown from this perspective, where hydrophobic residues are colored cyan and hydrophilic residues are colored blue. I17, a critical residue for binding, is colored limon for distinction. Hydrogen atoms have been removed for simplicity. (B) The binding motif as viewed from the side. Structures were generated from Ponderosa-C/S and visualized using PyMOL.

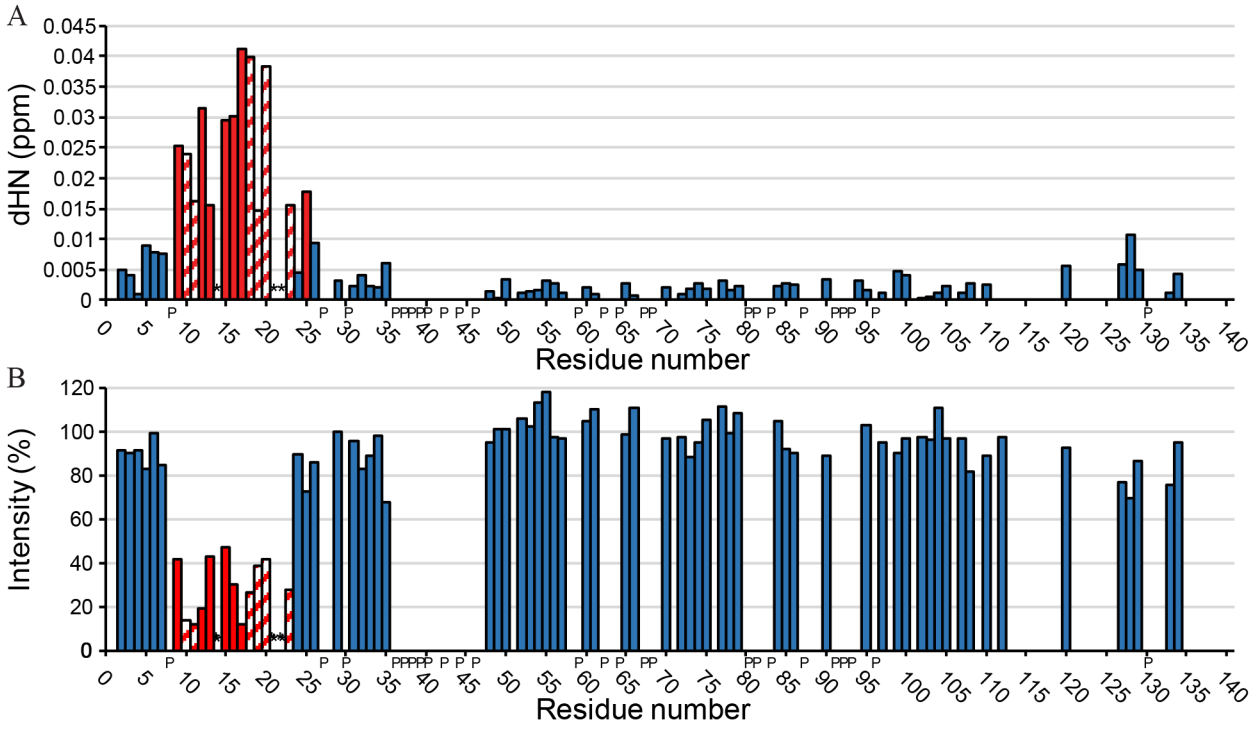


Figure 31. Spectral perturbations of PopZ$^{\Delta 134\text{-}177}$ upon binding to RcdA. (A) ΔHN combined chemical shift perturbations and (B) peak intensity perturbations of PopZ$^{\Delta 134\text{-}177}$ upon binding to RcdA. Binding is indicated by the most pronounced perturbations (red) found between residues T9-E23. The patterned columns represent residues that perturbed beyond detection before saturation was achieved. Therefore, the perturbation data from these patterned columns are from the last titration point where the peak was observed. Asterisks indicate residues with significant perturbations that we were not able to determine precisely due to severe peak overlap. "P" labels on the x-axis indicate proline residues which inherently cannot be observed in 2D ^{1}H-^{15}N HSQC spectra. Intensity perturbations were calculated as percentages of the 0 μM RcdA titration point peak intensities.

4.3. Mutation studies

To obtain genetic evidence for the functionality of the predicted MoRF in PopZ$^{\Delta 134\text{-}177}$, we introduced a point mutation into the most highly conserved residue within the T9-E23 region by replacing Ile 17 with Ala (I17A). Notably, the ^{1}H-^{15}N HSQC spectral peaks

corresponding to the α-helix were perturbed compared to wildtype PopZ$^{\Delta 134\text{-}177}$ spectra, indicating the I17A mutation disrupts the binding site. Adding excess RcdA (up to 1.0 mM) to 120 μM I17A PopZ$^{\Delta 134\text{-}177}$ did not induce spectral shifts or broadening (Figure 32), indicating that binding did not occur. This indicates that I17 of PopZ$^{\Delta 134\text{-}177}$ is critical for binding RcdA *in vitro*.

Dr. Bowman and colleagues (collaborators at the University of Wyoming) tested RcdA binding in an *E. coli* co-expression assay (Figure 33). Here, RcdA was produced as a fusion protein with green fluorescent protein (RcdA-GFP) and co-expressed with either full-length wildtype PopZ or the full-length I17A mutant (each produced as fusions with mCherry for visualization by fluorescence microscopy). Due to the presence of the C-terminal self-assembly domain in full-length PopZ, both of the PopZ variants accumulated at the cell poles of the bacteria[22]. RcdA-GFP exhibited strong co-localization with wildtype PopZ, but much weaker co-localization with the I17A mutant, confirming the interaction defect.

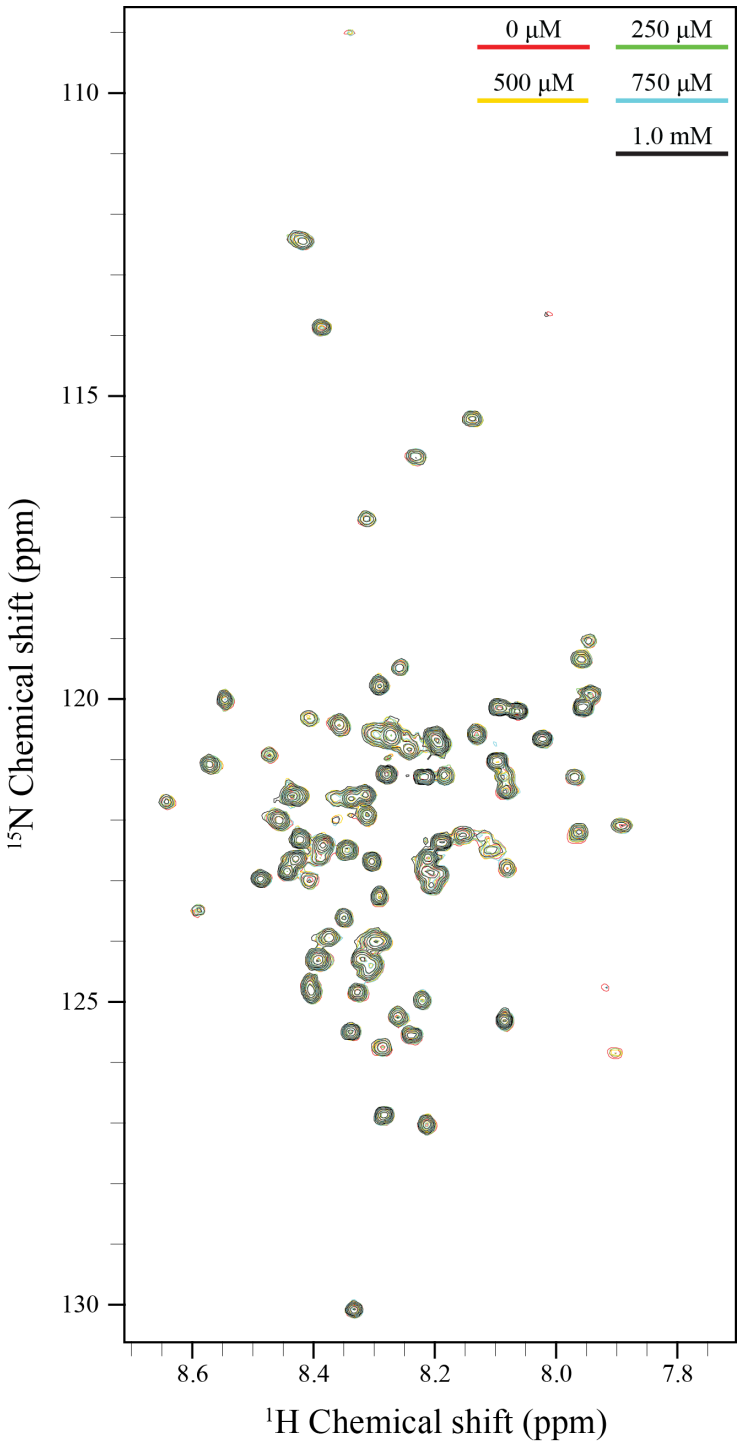


Figure 32. RcdA binding studies with I17A PopZ$^{\Delta 134\text{-}177}$. 2D ^{1}H-^{15}N HSQC NMR spectra overlay of I17A ^{15}N-enriched PopZ$^{\Delta 134\text{-}177}$ in solution with and without the RcdA partner exhibit no spectral changes, indicating that I17 is an essential residue of PopZ$^{\Delta 134\text{-}177}$ for RcdA binding. 120 µM PopZ$^{\Delta 134\text{-}177}$ was titrated with various concentrations of RcdA: 0 µM (red), 250 µM (green), 500 µM (yellow), 750 µM (blue), and 1.0 mM (black).

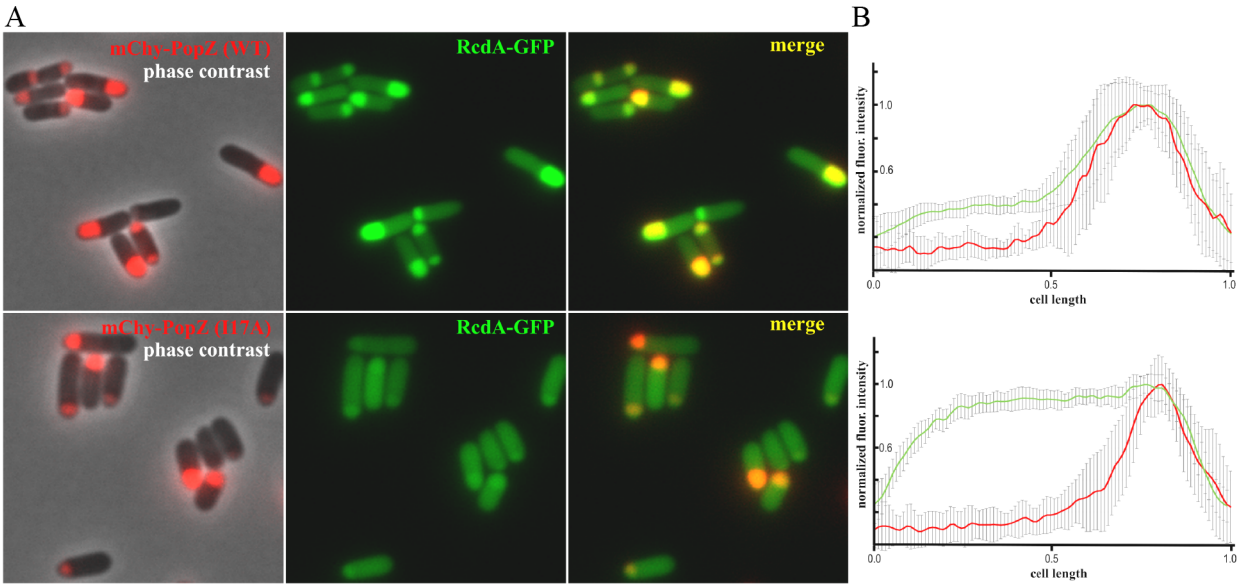


Figure 33. Localization of RcdA-GFP and wildtype or I17A mutant PopZ in an *E. coli* co-expression assay. (A) Full-length wildtype PopZ (upper panels) or the I17A point mutant (lower panels) was fused to the C-terminus of mChy (red signal) and co-expressed in *E. coli* cells with RcdA-GFP (green signal). Single channel fluorescence images are overlayed on a phase contrast image (left panels), shown independently (center panels), or overlayed to show co-localization of fluorescent proteins (right panels). (B) Distributions of mChy-PopZ and RcdA-GFP are shown by plotting normalized fluorescent pixel intensities (Y-axis) as a function of cell length (X-axis), with cells oriented such that 0 marks open poles and 1 marks poles with mChy-PopZ foci (n=10). Bars show standard deviation between cells. Figure from Grant Bowman.

Chapter 5

Discussion and Conclusion

5.1. Discussion

In this work, we have structurally characterized the multi-protein binding domain of the alphaproteobacterial hub protein PopZ. Native PopZ is 177 amino acid residues in length, and we have investigated a truncated version, $PopZ^{\Delta 134\text{-}177}$, which includes the first 133 residues, a short linker, and 6xHis-tag. In the unbound form of $PopZ^{\Delta 134\text{-}177}$, we find that the critical binding motif includes a short α-helical segment between residues M10 and I17, with additional unstructured residues on either side. To analyze a bound form of $PopZ^{\Delta 134\text{-}177}$ and probe the binding interface, we used NMR titration studies against a binding partner, RcdA. We found that residues T9-E23 bind either directly or indirectly to the RcdA, forming a MoRF that may include a mixture of α-helical and coiled features. Residue I17, which lies at the center of the binding motif, remains critical for the interaction, and disruption of this residue leads to significant perturbation of the MoRF spectral peaks and drastic loss in binding. Adjacent to I17 lie R18 and R19, which contrast with the generally negative charge of $PopZ^{\Delta 134\text{-}177}$ in forming a small island of positive charge. Positive charges at these positions are evolutionarily conserved and are likely to be important for electrostatic interactions between PopZ and its binding partners. Consistent with this, changing R19 to glutamate contributes to the loss of binding affinity to ParA and ParB[143].

In earlier work we used NMR spectroscopy to demonstrate binding between PopZ$^{\Delta 134\text{-}177}$ and a different binding partner, ChpT[36]. Comparison of the spectra from RcdA and ChpT binding showed a nearly identical pattern of chemical shift perturbations, suggesting that both proteins interact with the same set of amino acids in PopZ$^{\Delta 134\text{-}177}$, even though these two proteins exhibit no sequence or structural homology[140,144]. The only significant difference was in the peak corresponding to D25, located at the C-terminal end of the binding region, which did not show significant perturbation in complex with ChpT.

The cell cycle-dependent timing and intensity of polar localization of PopZ binding partners differ significantly and are not directly proportional to the localization pattern of PopZ. The transmembrane scaffolding protein SpmX is recruited by PopZ to the stalked pole[135], whereas transmembrane histidine kinases DivL and CckA, which are also direct binding partners of PopZ, are localized to the opposing swarmer pole or distributed between the two poles, respectively[145]. Other PopZ binding partners, including RcdA, ChpT, and the protease adaptor protein CpdR[146], are localized transiently to the stalked pole. Such wide variation in timing and localization of PopZ binding partners could not occur if polar localization were determined simply by the ability to interact with PopZ, and this suggests the existence of regulatory mechanisms that affect PopZ interaction. Kinases, their downstream effectors, and other factors that regulate the production and degradation of the secondary messenger c-di-GMP exhibit highly polarized localization and activity in *C. crescentus*[147-149]. PopZ binds directly to some of the elements within these signaling networks, and in doing so it could establish feedback loops that reinforce polar asymmetry through a signal that increases or decreases PopZ's affinity for certain

binding partners. Protein phosphorylation is one potential signal, as interactions involving eukaryotic IDPs can be regulated in this manner[150-152]. The symmetry-break in *C. crescentus* could be established by cell division, if the PopZ that accumulates at the newly formed pole establishes different signaling complexes than PopZ at the old pole.

We have shown that the PopZ side of the binding interface includes an α-helix. Binding helices are common in eukaryotic IDP hub domains, and there are multiple examples of interactions that are mediated by a single helix[153-156]. In many IDPs, relatively stable elements of secondary structure called αMoRFs form the core of the binding interface, and these are thought to aid binding by providing pre-formed structure that increases the stability of the bound complex[157-159]. Our structural analysis of PopZ$^{\Delta 134\text{-}177}$ demonstrates an αMoRF in *Alphaproteobacteria*, a feature that appears to be less common in bacterial proteomes than in eukaryotic proteomes [62]. Another bacterial protein with comparable structural and functional qualities is the divisome assembly protein FtsZ, which has a C-terminal IDP hub domain consisting of a disordered linker followed by a MoRF. In complex with different binding partners, FtsZ's MoRF region can be partially α-helical[160,161] or an extended linear motif[162]. For FtsZ, a destabilized αMORF may provide increased structural flexibility that allows binding to a wider range of binding surfaces. While our results suggest that the MoRF in PopZ is α-helical, transient unfolding is likely to occur, and some binding complexes may utilize other structural conformations of this region[163].

Since our analysis of secondary structure was limited to the unbound state of PopZ$^{\Delta 134\text{-}177}$, a remaining question is whether the helical region in the N-terminal MoRF of PopZ$^{\Delta 134\text{-}177}$ is altered upon interaction with binding partner proteins. For some IDPs,

the transition from unbound to bound state involves the formation of additional α-helices[164], extension of a preformed helix[54], or stabilization of coil[165]. Similarly, the bound form of PopZ$^{\Delta 134\text{-}177}$ could acquire more complex structural character through stabilization of features on either side of the M10-I17 helix. Consistent with this, we find that the set of amino acids that undergo chemical shift perturbation during interaction extends out to residue D25 for RcdA or E23 for ChpT. We note that M10-I17 forms an amphipathic helix (Fig. 4A), and that the amphipathic nature of this structure would continue if the helix were extended to E23. Furthermore, secondary structural prediction algorithms consistently predict that the helix extends to at least I21 or S22[166,167]. Together, the evidence suggests that the helical portion of the PopZ$^{\Delta 134\text{-}177}$ αMoRF could include several amino acids beyond I17 in the bound form of PopZ$^{\Delta 134\text{-}177}$, though further studies will be needed to confirm this hypothesis.

Our finding that the I17A mutation inhibits the interaction between PopZ and RcdA is consistent with earlier studies, which show that cell division defects in a *popZ* deletion strain can be rescued by expression of wildtype PopZ. The I17A mutant cannot rescue the cell division defects in a *popZ* deletion strain, and the I17A mutant PopZ protein fails to recruit a direct binding partner, SpmX, to the cell pole[22,135]. Similarly, PopZ bearing an I13A point mutation was found to be partially functional in rescuing *popZ* deletion phenotypes. We propose that the hydrophobic side of the amphipathic PopZ αMoRF (which includes I13 and I17) fits into a hydrophobic groove on RcdA, ChpT, SpmX, and other binding partners, in a manner that is analogous to the interaction of the amphipathic PUMA αMoRF with Mcl-1[155]. Hydrophobic residues have been shown to increase binding affinity in MoRF regions[168], as demonstrated for PUMA[169] and other binding helices in

IDPs[170,171]. Mutations in charged and polar amino acids in the N-terminal region of PopZ appear to have variable effects on binding. For example, the E12K R19E double mutation inhibits binding to ParB but not ParA, while the S22P mutant inhibits binding to ParA but not ParB[143]. It may be that all or most of PopZ's binding partners use contact with the hydrophobic side of the helix formed by M10-I17 to form a "fuzzy" binding intermediate[47], but that these interactions are not strong enough to lead to longer-term binding without additional contacts on other faces of the helix or peripheral contacts outside of the core helix.

In light of the fact that PopZ forms trimers, hexamers, and higher ordered structures *in vivo*[22,166], it is also possible that multiple M10-I17+ MoRF helices work together to form a complex interface that binds the target protein. Although our results do not rule this model out, we note that our data was collected with a truncated form of PopZ that does not self-assemble[22]. A related question is how the large, ~80 amino acid disordered region on the C-terminal side of the αMoRF region of PopZ contributes to binding. While total deletion of the disordered region results in loss of PopZ function, it can be reduced to half size and also scrambled without having strong effects on the ability to interact with other proteins[36] or complement the *popZ* knockout phenotype *in vivo*[22]. This region may be a flexible linker that separates the PopZ C-terminal oligomerization domain (residues 134-177) from the MoRF, while also contributing to the structural disorder that facilitates "fuzzy complexes" between IDP and target proteins[47,172]. We suggest that PopZ$^{\Delta 134\text{-}177}$ binds and reels the binding partners to the superstructure through sampling of its various disordered conformers, where a wide range of motion increases the likelihood of encountering and binding to other proteins in its local

environment. This is similar to the fly-caster model for other IDPs[173], although there is no clear evidence to suggest that PopZ would fold upon binding. Oligomerization of full-length PopZ via the C-terminal domain, which is predicted to be highly structured, may act as an anchor around which these binding events occur. Another possibility that is not mutually exclusive is that the disordered domain creates a phase-separated droplet[174], and this may provide a microenvironment that is conducive to hub binding activity, as has been proposed for the organization of transcription factor complexes in eukaryotes[175].

5.2. Project conclusions

Aim 1: Structural characterization of PopZ by solution NMR spectroscopy. *Working hypothesis: PopZ adopts a structure of flexible disorder with a short helix on both the N- and C-termini. This was the predicted structure of PopZ based on protein sequential information.* In this regard, we have found clear structural evidence for the existence of an α-helix on the N-terminus that encompasses residues M10-I17. Surrounding the helical domain are regions of remarkable disorder. Secondary chemical shifts, residue-specific relaxation rates (both R_1 and R_2), hetNOE ratios, RDC constants, and long-distance NOESY constraints all contribute to the structural elucidation of PopZ$^{\Delta 134\text{-}177}$. Regarding helicity near the C-terminus, we did not find strong evidence for a stable secondary structure. NMR evidence suggests there may be helical tendencies between residues 100-125, but the stability appears to be much weaker than what is observed for the N-terminal helix. It is possible the C-terminus adopts low population, preformed helices that stabilize upon interaction with a binding partner.

Aim 2: Characterization of the interaction of PopZ with one of its binding partners, RcdA. *Working hypothesis: PopZ adopts a disorder-to-order shift upon binding with protein partners and binding occurs near the sites of proposed helicity at the N- and C-termini.* In this regard, we determined a binding MoRF motif on the N-terminus between residues T9-E23, which includes the α-helix determined from the **Aim 1** studies. By comparing these results to previous studies[36], we determined that the RcdA-binding MoRF is the same binding region that interacts with the ChpT binding partner. Furthermore, we identified I17 as a critical residue in both helix stability and binding functionality. Point mutations at I17 led to severe perturbation of MoRF spectral peaks and almost total inhibition of binding as observed from NMR studies. Given the limitations of these studies, current results do not indicate that $PopZ^{\Delta 134\text{-}177}$ adopts a disorder-to-order shift upon binding, but this does not rule out the possibility of smaller structural changes occurring upon binding.

5.3. Future directions

The findings from this study open a wide variety of avenues that can be explored with future studies. We characterized the binding motif for RcdA which appears to be identical to the binding motif for ChpT[36]. However, PopZ binds to multiple partners, so future studies can include additional binding characterization of the $PopZ^{\Delta 134\text{-}177}$ variant with these binding partners. It would be interesting to see if the same binding motif (T9-E23) is used for all known binding partners, as it does with RcdA and ChpT, and if $PopZ^{\Delta 134\text{-}177}$ contains additional binding domains for other partners. This task is currently being undertaken with the CpdR binding partner. Previous studies have indicated that a region

between the PED and the C-terminus is necessary for binding[36], so it would be useful to investigate the binding partners for which this is true and determine the PopZ$^{\Delta 134\text{-}177}$ residues in this region that participate. This can be achieved using similar studies exemplified with the RcdA titration experiments presented in **section 4.5**. We did not see evidence of binding in this region with RcdA in the current study, but many peaks in this region were unobservable under the binding conditions used for NMR experiments. Another potential direction would be to investigate structural studies into the PopZ I17A mutant and other PopZ variants to reveal the structural changes that arise from these point mutations and help explain why binding activity becomes severely inhibited.

Furthermore, NMR binding studies of the binding partners with PopZ$^{\Delta 134\text{-}177}$ could reveal which residues in RcdA, and other binding partners, participate in binding with PopZ$^{\Delta 134\text{-}177}$. This approach would require isotopic labeling of the binding partners while PopZ$^{\Delta 134\text{-}177}$ remains unlabeled, giving data that would help characterize the binding interface and give valuable insight into the interactions between IDPs and their partners as a whole.

The most prominent and ambitious task would be to investigate the structure and binding interactions of full-length PopZ in its homo-oligomeric state. This was limited in the current study due to the size limitation inherent to solution NMR spectroscopy. The full-length PopZ polymeric superstructures are far too large to tumble effectively in an aqueous environment, causing broad, unresolved peaks in solution, and solid-state NMR spectroscopy will be utilized to overcome this limitation. The PopZ scaffold can be isolated via ultracentrifugation and PopZ can be assigned with standard solid-state NMR methods such as NCA, NCO, and NCOCX experiments. Structural information could then be

obtained and comparisons could be made between full-length PopZ and the variant used in this study ($PopZ^{\Delta 134\text{-}177}$). Since binding occurs in both full-length PopZ and the $PopZ^{\Delta 134\text{-}177}$ variant, the structural differences are likely small for residues 1-133. Expression and purification of full-length PopZ is currently being undertaken.

APPENDIX

Table A1. Recipe for Minimal Media.

Minimal Media	
Ingredient	Amount
Na_2HPO_4	48 mM
KH_2PO_4	22 mM
NaCl	9 mM
$MgSO_4$	2 mM
NH_4Cl (^{15}N-labeled or unlabeled)	18 mM
D-glucose (^{13}C-labeled or unlabeled)	0.4% w/v
[1]Trace metals stock	0.2% v/v
[2]Vitamin stock	1% v/v
Kanamycin	30 µg/mL
Note: bolded ingredients are mixed together first and autoclaved. Unbolded ingredients are added after cooling.	
[1]Ingredients are listed in Table ###.	
[2]Ingredients are listed in Table ###.	

Table A2. Recipe for Trace Metals stock used in minimal media.

Trace Metals Stock (100 mL in H_2O)	
Ingredient	Amount (mg)
$FeCl_3$	81.1
$CaCl_2$	22.2
$MnCl_2$	19.8
$ZnSO_4 \cdot H_2O$	28.75
$CoCl_2 \cdot 6H_2O$	4.76
$CuCl_2 \cdot 6H_2O$	3.41
$NiCl_2 \cdot 6H_2O$	4.75
$Na_2MoO_4 \cdot 2H_2O$	4.84
Na_2SeO_3	3.46
H_3BO_3	1.23
Once dissolved, filter sterilize the solution with a 0.22 µm syringe filter.	

Table A3. Recipe for Vitamin stock used in minimal media.

Vitamin Stock (100 mL in H_2O)	
Ingredient	Amount (mg)
Thiamin	50
Biotin	10
Choline chloride	10
Folic acid	10
Nicatinamide	10
D-pantothenic acid	10
Pyridoxal	10
Riboflavin	1
Once dissolved, filter sterilize the solution with a 0.22 µm syringe filter.	

Table A4. PopZ$^{\Delta 134\text{-}177}$ chemical shift table.

	C	CA	CB	CD	CD1	CD2	CG	CG1	CG2	H	HA	HA2	HB	HB2	HB3	HD11	HD2	HE	HG12	HG2	HG21	N	NE
M1	177.5	54.66	-	-	-	-	-	-	-	-	-	-	-	2.077	-	-	-	-	-	-	-	-	-
S2	174.8	58.6	63.96	-	-	-	-	-	-	8.371	4.433	-	-	3.863	-	-	-	-	-	-	-	121.6	-
D3	176.4	54.33	40.93	-	-	-	-	-	-	8.519	4.609	-	-	2.707	-	-	-	-	-	-	-	121.8	-
Q4	176.2	55.95	29.49	-	-	-	33.98	-	-	8.312	4.366	-	-	1.971	2.178	-	-	-	-	2.361	-	120.5	-
S5	174.5	58.83	63.77	-	-	-	-	-	-	8.377	4.396	-	-	3.88	-	-	-	-	-	-	-	117	-
Q6	175.7	55.52	29.75	-	-	-	33.89	-	-	8.362	4.4	-	-	1.953	2.141	-	-	-	-	-	-	121.9	-
E7	174.5	54.46	29.55	-	-	-	-	-	-	8.342	4.588	-	-	1.897	2.053	-	-	-	-	-	-	123.4	-
P8	177.1	63.01	32.23	50.76	-	-	27.47	-	-	-	4.572	-	-	1.908	2.289	-	3.851	-	-	1.993	-	-	-
T9	175.2	61.34	70.62	-	-	-	-	-	22.06	8.364	4.373	-	4.497	-	-	-	-	-	-	-	1.313	113.5	-
M10	177.5	57.15	32.02	-	-	-	-	-	-	8.682	4.37	-	-	2.097	-	-	-	-	-	2.613	-	121.4	-
E11	178.3	59	29.54	-	-	-	36.57	-	-	8.565	4.056	-	-	1.997	-	-	-	-	-	2.328	-	119.5	-
E12	178.5	58.23	29.83	-	-	-	36.71	-	-	8.028	4.126	-	-	2.012	2.183	-	-	-	-	2.257	-	121.1	-
I13	178.4	63.56	38.02	-	-	-	-	28.35	17.3	8.229	3.885	-	-	-	-	-	-	-	-	-	-	122.5	-
L14	178.5	56.84	41.81	-	25.15	23.15	27	-	-	8.336	4.157	-	-	1.549	1.75	0.8865	-	-	-	-	-	122.3	-
A15	179	53.89	18.64	-	-	-	-	-	-	7.987	4.196	-	-	1.461	-	-	-	-	-	-	-	121.8	-
S16	175.3	59.82	63.55	-	-	-	-	-	-	7.954	4.341	-	-	3.969	-	-	-	-	-	-	-	113.8	-
I17	176.7	62.45	38.44	-	-	-	-	-	-	7.8	4.095	-	1.931	-	-	-	-	-	-	-	-	121.6	-
R18	176.4	56.73	-	-	-	-	-	-	-	8.098	4.273	-	-	1.427	-	-	-	-	-	-	-	122.7	-
R19	-	56.63	-	-	-	-	-	-	-	8.101	-	-	-	-	-	-	-	-	-	-	-	121.7	-
I20	-	60.47	39.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
I21	176.2	61.28	38.6	-	-	-	-	-	-	8.553	4.181	-	1.892	-	-	-	-	-	-	-	-	125.4	-
S22	174.7	58.22	64.03	-	-	-	-	-	-	8.353	4.485	-	-	3.87	-	-	-	-	-	-	-	119.7	-
E23	176.3	56.92	30.01	-	-	-	35.87	-	-	8.521	4.295	-	-	1.946	2.096	-	-	-	-	2.297	-	123.1	-
D24	175.9	54.54	41.09	-	-	-	-	-	-	8.274	4.582	-	-	2.578	2.715	-	-	-	-	-	-	120.5	-
D25	175.5	54.11	41.06	-	-	-	-	-	-	8.152	4.583	-	-	2.651	-	-	-	-	-	-	-	120.5	-
A26	175.3	50.58	18.27	-	-	-	-	-	-	8.085	-	-	-	1.347	-	-	-	-	-	-	-	125.2	-
A28	177.6	52.23	19.41	-	-	-	-	-	-	8.518	-	-	-	-	-	-	-	-	-	-	-	125.2	-
E29	174.5	54.11	29.65	-	-	-	-	-	-	8.331	-	-	-	-	-	-	-	-	-	-	-	121.4	-
P30	176.7	62.98	32.12	50.66	-	-	27.45	-	-	-	4.384	-	-	1.9	2.279	-	3.755	-	-	1.994	-	-	-
A31	177.6	52.28	19.29	-	-	-	-	-	-	8.408	4.285	-	-	1.382	-	-	-	-	-	-	-	124.6	-
A32	177.9	52.49	19.36	-	-	-	-	-	-	8.3	4.278	-	-	1.386	-	-	-	-	-	-	-	123.8	-
E33	176	56.3	30.39	-	-	-	36.08	-	-	8.343	4.247	-	-	1.94	-	-	-	-	-	2.271	-	120.2	-
A34	177	52.15	19.4	-	-	-	-	-	-	8.279	4.287	-	-	1.362	-	-	-	-	-	-	-	125.6	-
A35	175	50.32	18.16	-	-	-	-	-	-	8.256	4.583	-	-	1.351	-	-	-	-	-	-	-	125.1	-
P46	176.8	62.98	32.03	50.78	-	-	27.45	-	-	-	4.429	-	-	1.864	2.261	-	-	-	-	2.021	-	-	-
V47	175.9	62.17	33.08	-	-	-	-	-	20.99	8.202	4.073	-	1.978	-	-	-	-	-	0.866	-	-	120.5	-
S48	174	57.6	64.2	-	-	-	-	-	-	8.264	4.456	-	-	3.769	-	-	-	-	-	-	-	119.4	-
F49	175.5	57.64	39.82	-	-	-	-	-	-	8.398	4.662	-	-	2.963	3.182	-	-	-	-	-	-	122.9	-
D50	175.7	54.4	41.07	-	-	-	-	-	-	8.311	4.584	-	-	2.556	2.703	-	-	-	-	-	-	121.7	-
D51	176.2	54.46	41.01	-	-	-	-	-	-	8.2	4.555	-	-	2.656	-	-	-	-	-	-	-	120.6	-
E52	176.4	56.78	30.18	-	-	-	35.88	-	-	8.256	4.256	-	-	1.99	-	-	-	-	-	2.257	-	120.9	-
V53	176.1	62.57	32.65	-	-	-	-	-	21.07	8.156	4.048	-	2.045	-	-	-	-	-	0.9247	-	-	122	-
L54	177.1	55.08	42.53	-	24.88	23.82	27.14	-	-	8.258	4.357	-	-	1.579	-	0.891	-	-	-	-	-	126.4	-
E55	176	56.2	30.19	-	-	-	35.86	-	-	8.324	4.306	-	-	1.951	-	-	-	-	-	2.254	-	122.1	-
L56	177.4	55.04	42.32	-	25.14	23.48	27.05	-	-	8.297	4.429	-	-	1.626	-	0.902	-	-	-	-	-	124	-
T57	174.1	61.59	70.03	-	-	-	-	-	21.71	8.123	-	-	4.183	-	-	-	-	-	-	-	1.182	115.1	-
D58	174.2	52.66	40.77	-	-	-	-	-	-	8.302	-	-	-	-	-	-	-	-	-	-	-	124.2	-
P59	176.8	63.11	32.15	50.73	-	-	27.3	-	-	-	4.423	-	-	1.914	2.26	-	-	-	-	2.006	-	-	-
I60	175.8	60.89	38.54	-	12.98	-	-	27.37	17.58	8.21	4.104	-	1.84	-	-	0.9	-	-	1.47	-	1.195	121.1	-
A61	175.3	50.33	18.39	-	-	-	-	-	-	8.32	4.603	-	-	1.35	-	-	-	-	-	-	-	129.9	-
P64	176.7	62.99	32.18	50.65	-	-	27.34	-	-	-	4.389	-	-	-	2.272	-	3.735	-	-	-	-	-	-
E65	176.2	56.23	30.39	-	-	-	36.11	-	-	8.44	4.258	-	-	1.942	-	-	-	-	-	2.284	-	121.4	-
L66	174.7	52.84	41.72	-	-	-	-	-	-	8.328	4.629	-	-	1.565	-	-	-	-	-	-	-	125.3	-
P68	177	62.75	32.01	50.47	-	-	27.41	-	-	-	4.418	-	-	1.926	2.279	-	-	-	-	-	-	-	-
L69	177.6	55.4	42.51	-	24.38	-	27.07	-	-	8.306	4.295	-	-	1.595	-	0.9075	-	-	-	-	-	122.5	-
E70	176.4	56.45	30.45	-	-	-	36.05	-	-	8.447	4.369	-	-	1.938	2.049	-	-	-	-	2.263	-	121.7	-
T71	174.5	61.66	69.9	-	-	-	-	-	21.74	8.228	4.372	-	4.186	-	-	-	-	-	-	-	1.175	115.8	-
V72	176.5	62.54	32.9	-	-	-	-	-	20.94	8.207	4.137	-	2.092	-	-	-	-	-	0.9495	-	-	122.7	-
G73	173.6	45.14	-	-	-	-	-	-	-	8.426	-	3.954	-	-	-	-	-	-	-	-	-	112.3	-
D74	176.3	54.31	41.11	-	-	-	-	-	-	8.243	4.624	-	-	2.588	2.693	-	-	-	-	-	-	120.6	-
I75	175.9	61.22	39.18	-	13.34	-	-	27.31	17.68	8.065	4.172	-	1.868	-	-	0.8885	-	-	1.435	-	1.17	120.1	-

	C	CA	CB	CD	CD1	CD2	CG	CG1	CG2	H	HA	HA2	HB	HB2	HB3	HD11	HD2	HE	HG12	HG2	HG21	N	NE
D76	176	54.32	41.04	-	-	-	-	-	-	8.406	4.625	-	-	2.647	-	-	-	-	-	-	-	124.1	-
V77	175.7	62.24	32.69	-	-	-	-	-	20.72	7.953	4.069	-	2.014	-	-	-	-	-	0.833	-	-	120	-
Y78	175.3	58.08	38.97	-	-	-	-	-	-	8.319	4.565	-	-	2.975	-	-	-	-	-	-	-	124.6	-
S79	171.2	55.58	63.62	-	-	-	-	-	-	7.967	4.643	-	-	3.673	3.737	-	-	-	-	-	-	121.1	-
P83	177	63.22	32.14	50.71	-	-	27.45	-	-	-	4.393	-	-	1.919	2.294	-	3.737	-	-	-	-	-	-
E84	176.5	56.65	30.1	-	-	-	36.14	-	-	8.57	4.254	-	-	1.978	-	-	-	-	-	2.306	-	120.9	-
S85	174.1	58.34	64.13	-	-	-	-	-	-	8.303	4.434	-	-	3.842	-	-	-	-	-	-	-	116.9	-
E86	174.4	54.35	29.68	-	-	-	-	-	-	8.356	4.608	-	-	1.907	2.055	-	-	-	-	-	-	123.7	-
P87	176.6	63.09	32.12	50.76	-	-	27.4	-	-	-	4.372	-	-	1.838	2.251	-	3.773	-	-	2.005	-	-	-
A88	177.3	52.34	19.23	-	-	-	-	-	-	8.389	4.278	-	-	1.348	-	-	-	-	-	-	-	124.2	-
Y89	175.3	57.99	39.24	-	-	-	-	-	-	8.188	4.561	-	-	2.865	3.048	-	-	-	-	-	-	120.6	-
T90	171.3	58.94	70.3	-	-	-	-	-	-	7.892	4.488	-	3.933	-	-	-	-	-	-	-	-	121.9	-
P93	176.6	62.81	32.08	50.46	-	-	27.36	-	-	-	4.357	-	-	-	2.276	-	3.625	-	-	-	-	-	-
A94	177.1	52.06	19.37	-	-	-	-	-	-	8.312	4.271	-	-	1.353	-	-	-	-	-	-	-	124.4	-
A95	175.4	50.3	18.2	-	-	-	-	-	-	8.218	4.564	-	-	1.349	-	-	-	-	-	-	-	124.8	-
P96	176.4	62.82	32.07	50.53	-	-	27.44	-	-	-	4.363	-	-	1.68	2.168	-	-	-	-	1.986	-	-	-
V97	175.8	62.05	33.14	-	-	-	-	-	21.13	8.097	4.037	-	1.948	-	-	-	-	-	0.8647	-	-	120	-
F98	175.2	57.26	40.06	-	-	-	-	-	-	8.286	4.677	-	-	2.952	3.14	-	-	-	-	-	-	123.9	-
D99	175.9	54.14	41.16	-	-	-	-	-	-	8.387	4.582	-	-	2.655	-	-	-	-	-	-	-	122.5	-
R100	176	56.43	30.91	43.48	-	-	27.02	-	-	8.205	4.257	-	-	1.815	-	-	3.18	7.298	-	1.614	-	121.3	110.8
D101	176.3	54.59	40.97	-	-	-	-	-	-	8.419	4.601	-	-	2.615	2.728	-	-	-	-	-	-	121.2	-
E102	176.8	57.02	30.15	-	-	-	36.06	-	-	8.281	4.249	-	-	1.992	-	-	-	-	-	2.256	-	121.3	-
V103	176.4	62.93	32.62	-	-	-	-	-	21.15	8.08	4.017	-	2.085	-	-	-	-	-	0.93	-	-	120.8	-
A104	178.3	53.07	19.13	-	-	-	-	-	-	8.21	4.246	-	-	1.394	-	-	-	-	-	-	-	126.7	-
E105	176.8	57.1	30.08	-	-	-	36.13	-	-	8.274	4.194	-	-	1.992	-	-	-	-	-	1.998	-	119.5	-
Q106	176.1	56.13	29.35	-	-	-	33.96	-	-	8.259	4.277	-	-	2.039	-	-	-	-	-	2.375	-	120.5	-
L107	177.4	55.33	42.27	-	25.13	23.36	27	-	-	8.191	4.363	-	-	1.559	1.66	0.874	-	-	-	-	-	122.8	-
V108	176.8	62.72	32.76	-	-	-	-	-	-	8.011	4.093	-	2.087	-	-	-	-	-	0.94	-	-	120.5	-
G109	174.3	45.38	-	-	-	-	-	-	-	8.435	-	3.964	-	-	-	-	-	-	-	-	-	112.3	-
V110	176.5	62.51	32.91	-	-	-	-	-	20.92	7.959	4.143	-	2.094	-	-	-	-	-	0.9327	-	-	119.2	-
S111	174.6	58.43	63.96	-	-	-	-	-	-	8.429	4.449	-	-	3.903	-	-	-	-	-	-	-	119.3	-
A112	178.1	53.11	19.03	-	-	-	-	-	-	8.401	4.279	-	-	1.401	-	-	-	-	-	-	-	126.2	-
A113	178.4	53.12	19	-	-	-	-	-	-	8.256	4.275	-	-	1.4	-	-	-	-	-	-	-	122.5	-
S114	175	58.82	63.72	-	-	-	-	-	-	8.136	4.376	-	-	3.907	-	-	-	-	-	-	-	114.7	-
A115	178.1	53.08	19.02	-	-	-	-	-	-	8.246	4.275	-	-	1.407	-	-	-	-	-	-	-	125.9	-
A116	178.1	52.99	19.02	-	-	-	-	-	-	8.118	4.251	-	-	1.405	-	-	-	-	-	-	-	122.4	-
A117	178.2	53.01	19.06	-	-	-	-	-	-	8.099	4.271	-	-	1.415	-	-	-	-	-	-	-	122.3	-
S118	174.6	58.59	63.86	-	-	-	-	-	-	8.09	4.377	-	-	3.86	-	-	-	-	-	-	-	114.2	-
A119	177.7	52.87	19.07	-	-	-	-	-	-	8.149	4.257	-	-	1.267	-	-	-	-	-	-	-	125.4	-
F120	176.6	58.14	39.41	-	-	-	-	-	-	8.112	4.575	-	-	3.031	3.189	-	-	-	-	-	-	118.6	-
G121	174.3	45.56	-	-	-	-	-	-	-	8.226	-	3.913	-	-	-	-	-	-	-	-	-	110.1	-
S122	174.9	58.58	63.92	-	-	-	-	-	-	8.155	4.446	-	-	3.891	-	-	-	-	-	-	-	115.6	-
L123	177.8	55.57	42.25	-	25.07	23.51	27.03	-	-	8.306	4.403	-	-	1.647	-	0.887	-	-	-	-	-	123.8	-
S124	175.1	58.77	63.69	-	-	-	-	-	-	8.239	4.4	-	-	3.879	-	-	-	-	-	-	-	115.9	-
S125	174.7	58.9	63.64	-	-	-	-	-	-	8.236	4.385	-	-	3.892	-	-	-	-	-	-	-	117.7	-
A126	177.8	53.05	19.1	-	-	-	-	-	-	8.177	4.274	-	-	1.381	-	-	-	-	-	-	-	125.4	-
L127	177.1	55.3	42.28	-	25.02	23.57	27.1	-	-	7.938	4.278	-	-	1.605	-	0.894	-	-	-	-	-	119.7	-
L128	176.9	54.87	42.31	-	25.05	23.52	27.05	-	-	7.952	4.296	-	-	1.595	-	0.875	-	-	-	-	-	122	-
M129	174.1	53.21	31.63	-	-	-	-	-	-	8.157	-	-	-	-	-	-	-	-	-	-	-	122.1	-
P130	177.2	63.16	32.23	50.72	-	-	27.54	-	-	-	4.41	-	-	1.92	2.291	-	-	-	-	-	-	-	-
K131	176.7	56.66	32.79	29.22	-	-	24.74	-	-	8.512	4.272	-	-	1.8	-	-	-	-	-	1.424	-	121.8	-
D132	176.8	54.35	41.2	-	-	-	-	-	-	8.308	4.595	-	-	2.712	-	-	-	-	-	-	-	120.3	-
G133	174.3	45.68	-	-	-	-	-	-	-	8.324	-	3.938	-	-	-	-	-	-	-	-	-	108.8	-
L134	177.4	55.19	42.31	-	25.07	23.18	27.03	-	-	8.078	4.279	-	-	1.564	-	0.856	-	-	-	-	-	121.1	-
E135	176.1	56.61	29.99	-	-	-	-	-	-	8.345	-	-	-	-	-	-	-	-	-	-	-	120.5	-

Table A5. PopZ$^{\Delta 134\text{-}177}$ φ and ψ torsion angle predictions generated from TALOS-N.

Residue ID	Residue Name	φ (degree)	ψ (degree)	Δφ (degree)	Δψ (degree)	Class
1	M	9999	9999	0	0	None
2	S	-67.354	155.067	8.802	10.551	Dynamic
3	D	-73.076	-13.199	7.08	10.614	Dynamic
4	Q	-90.575	-4.837	9.436	10.934	Dynamic
5	S	-71.934	137.882	10.705	14.328	Dynamic
6	Q	-106.688	5.305	13.54	12.05	Dynamic
7	E	-82.053	137.091	19.682	19.41	Dynamic
8	P	-66.19	148.932	8.128	12.665	Dynamic
9	T	-78.208	164.759	8.123	5.617	Strong
10	M	-60.966	-35.807	4.514	4.232	Strong
11	E	-65.722	-41.934	4.18	6.146	Strong
12	E	-68.931	-39.002	3.371	5.025	Strong
13	I	-64.743	-43.134	3.373	5.127	Strong
14	L	-67.464	-34.764	4.395	5.294	Strong
15	A	-65.848	-34.764	4.535	6.961	Strong
16	S	-68.077	-29.907	5.306	7.31	Strong
17	I	-69.456	-21.814	7.452	10.252	Dynamic
18	R	-85.227	-7.474	8.667	11.647	Dynamic
19	R	-70.048	137.654	11.694	12.459	Dynamic
20	I	-90.942	131.678	13.287	10.747	Dynamic
21	I	-91.712	133.263	19.528	16.516	Dynamic
22	S	-73.073	144.424	10.863	11.488	Dynamic
23	E	-67.208	143.582	10.164	9.389	Dynamic
24	D	-75.421	-13.811	9.19	10.967	Warn
25	D	-94.569	-0.687	12.061	10.305	Strong
26	A	-69.756	137.467	13.197	12.226	Strong
27	P	-62.615	144.414	8.069	9.34	Dynamic
28	A	-68.829	144.51	10.451	13.71	Dynamic
29	E	-77.172	142.014	11.483	14.258	Dynamic
30	P	-64.743	148.039	7.485	8.237	Dynamic
31	A	-69.891	141.147	11.83	12.241	Dynamic
32	A	-70.243	155.271	10.833	15.085	Dynamic
33	E	-68.628	144.358	11.087	17.323	Dynamic
34	A	-71.627	140.779	9.747	14.744	Dynamic
35	A	-76.822	138.654	19.628	16.283	Dynamic
36	P	9999	9999	0	0	None
37	P	9999	9999	0	0	None
38	P	9999	9999	0	0	None
39	P	9999	9999	0	0	None
40	P	9999	9999	0	0	None
41	E	9999	9999	0	0	None
42	P	9999	9999	0	0	None
43	E	9999	9999	0	0	None
44	P	9999	9999	0	0	None
45	E	9999	9999	0	0	None
46	P	-64.443	149.178	7.132	7.321	Dynamic
47	V	-80.056	129.316	19.711	14.478	Dynamic
48	S	-85.04	139.765	23.424	14.108	Dynamic
49	F	-81.966	141.098	12.674	12.45	Dynamic
50	D	-70.718	130.141	20.238	17.049	Warn
51	D	-75.622	-19.711	11.793	14.066	Strong
52	E	-87.446	-7.951	12.382	10.049	Dynamic
53	V	-73.671	137.092	10.145	10.515	Dynamic
54	L	-78.701	138.827	11.106	17.568	Dynamic
55	E	-72.016	144.153	7.494	11.071	Dynamic
56	L	-73.875	142.611	13.015	10.719	Dynamic

Residue ID	Residue Name	φ (degree)	ψ (degree)	Δφ (degree)	Δψ (degree)	Class
57	T	-74.374	156.117	12.184	16.122	Dynamic
58	D	-71.217	138.585	12.64	11.714	Strong
59	P	-64.562	145.02	7.742	8.526	Dynamic
60	I	-81.688	130.043	14.909	14.819	Dynamic
61	A	-76.042	135.361	17.564	14.818	Dynamic
62	P	9999	9999	0	0	None
63	E	9999	9999	0	0	None
64	P	-63.183	144.305	7.403	9.812	Dynamic
65	E	-78.186	136.505	13.443	21.455	Generous
66	L	-79.892	135.325	14.92	14.652	Dynamic
67	P	-62.383	142.772	7.2	11.719	Dynamic
68	P	-61.936	147.588	6.212	8.502	Dynamic
69	L	-71.027	138.987	9.939	8.771	Dynamic
70	E	-72.272	145.399	10.082	12.259	Dynamic
71	T	-73.342	136.047	15.457	15.365	Dynamic
72	V	-73.74	142.774	17.334	10.024	Dynamic
73	G	89.585	-32.04	28.608	85.932	Dynamic
74	D	-69.605	148.612	5.782	10.708	Dynamic
75	I	-73.983	138.276	10.405	11.873	Dynamic
76	D	-88.067	125.046	26.881	25.903	Dynamic
77	V	-78.881	139.128	13.377	16.739	Dynamic
78	Y	-78.565	144.465	15.282	13.626	Dynamic
79	S	-99.672	141.731	22.911	17.7	Dynamic
80	P	9999	9999	0	0	None
81	P	9999	9999	0	0	None
82	E	9999	9999	0	0	None
83	P	-58.775	144.149	7.402	9.172	Dynamic
84	E	-67.991	135.975	15.953	13.726	Dynamic
85	S	-74.844	140.444	10.009	17.96	Dynamic
86	E	-85.005	125.628	28.549	26.724	Dynamic
87	P	-65.737	147.668	5.878	7.097	Dynamic
88	A	-79.664	141.333	13.662	20.173	Dynamic
89	Y	-83.751	131.663	21.899	13.694	Dynamic
90	T	-109.996	134.437	20.951	15.729	Strong
91	P	9999	9999	0	0	None
92	P	9999	9999	0	0	None
93	P	-63.005	146.53	7.262	9.451	Dynamic
94	A	-69.786	146.511	12.464	17.948	Dynamic
95	A	-72.221	134.557	13.641	13.507	Dynamic
96	P	-65.088	147.992	7.486	8.988	Dynamic
97	V	-79.968	132.499	14.325	14.224	Dynamic
98	F	-94.087	135.442	17.4	13.487	Dynamic
99	D	-74.059	133.094	11.57	22.292	Dynamic
100	R	-71.242	-19.8	17.391	13.565	Dynamic
101	D	-72.039	-20.648	7.663	10.658	Strong
102	E	-76.914	-21.805	10.532	13.505	Dynamic
103	V	-81.014	-28.616	18.463	12.741	Dynamic
104	A	-62.445	-30.928	7.082	6.899	Dynamic
105	E	-70.113	-20.015	7.42	7.445	Dynamic
106	Q	-93.659	-5.482	9.116	11.49	Dynamic
107	L	-91.018	-11.486	18.585	18.978	Dynamic
108	V	-62.113	129.579	9.71	8.493	Dynamic
109	G	91.285	-0.096	18.127	18.491	Dynamic
110	V	-76.191	146.807	10.595	11.219	Dynamic
111	S	-69.632	144.275	6.835	14.884	Dynamic
112	A	-63.288	-35.822	11.11	6.456	Dynamic
113	A	-67.328	-33.637	4.102	8.986	Dynamic

Residue ID	Residue Name	φ (degree)	ψ (degree)	Δφ (degree)	Δψ (degree)	Class
114	S	-70.553	-36.673	5.385	5.682	Dynamic
115	A	-68.477	-33.909	3.904	6.74	Dynamic
116	A	-68.045	-33.914	5.619	7.444	Dynamic
117	A	-66.518	-31.913	4.695	6.899	Dynamic
118	S	-73.073	-17.602	8.424	9.218	Dynamic
119	A	-80.884	-14.567	11.362	10.78	Dynamic
120	F	-95.574	0.261	15.545	19.974	Dynamic
121	G	117.507	-151.28	58.906	76.548	Dynamic
122	S	-90.342	121.543	20.367	25.97	Dynamic
123	L	-74.037	-19.545	13.051	15.572	Dynamic
124	S	-67.96	-28.236	10.034	10.351	Dynamic
125	S	-69.398	-26.502	8.305	8.939	Dynamic
126	A	-73.777	-20.17	9.901	10.505	Dynamic
127	L	-89.404	-5.53	10.428	7.352	Dynamic
128	L	-73.671	129.93	8.138	14.937	Dynamic
129	M	-77.42	136.061	9.382	12.997	Dynamic
130	P	-66.331	148.036	4.605	9.722	Dynamic
131	K	-68.294	152.962	8.026	14.914	Dynamic
132	D	-91.829	-10.324	19.1	14.74	Dynamic
133	G	83.513	9.154	8.389	8.815	Dynamic
134	L	-76.946	134.99	12.428	15.661	Dynamic
135	E	9999	9999	0	0	None

Notes: Predictions are classified as strong, generous, and dynamic. Predictions that are not trusted are classified as warn. Residues with no predictions are classified as none.

Protocol: Protein Expression Optimization

This protocol is written for optimizing the conditions for protein expression using an *E. coli* expression system. Common parameters that may need optimization include (A) induction time, (B) optical density at 600 nm (OD_{600}), and (C) induction temperature. If multiple parameters need to be optimized, then from personal experience I would recommend optimizing induction time first. I would start with an optical density of 0.4-0.5 and an induction temperature of 21 ºC.

(A) Induction Time Optimization

1. Assuming a transformation was performed, *E. coli* have been plated, and precultures have been grown, transfer 4 mL of preculture growth per 1 L fresh LB supplemented with the appropriate antibiotic. Grow these to the desired OD_{600}.
2. If you are optimizing your expression in LB, you can skip to step 4. If you are optimizing your expression in minimal media, then centrifuge the bacteria at 7,700 rcf for 30 minutes at 4 ºC. Pour off the supernatant and resuspend the pellet(s) in minimal media.
3. Allow the bacteria to equilibrate at induction conditions for 30 minutes before adding IPTG (Again, I would recommend 21 ºC, with shaking at 225 rpm).
4. Induce protein expression with IPTG (~1 mM).
5. Every 2 hours remove 20 mL of minimal media solution and pellet the cells with centrifugation at 7,700 rcf and 4 ºC for 30 minutes. The supernatant can be poured off and the pellet should be frozen at -80 ºC.

    -I would recommend doing a 24 hour time course.

6. After all time points have been collected and frozen, remove them from the freezer, thaw on ice, and then follow the lysozyme lysis protocol (separate document).

    -note: the lysozyme lysis protocol cannot be used if your target protein is the same size as lysozyme. A different lysis protocol will need to be used (i.e. French press).

7. Analyze the lysates with SDS-PAGE followed by Coomassie stain, silver stain, or Western blot. Band intensity is representative of the quantity of protein produced. Therefore, the most intense band corresponding to your protein will represent the optimized induction time.
8. If needed, perform an OD_{600} optimization or induction temperature optimization described below.

Troubleshooting tips:

- If there are no bands where you expect to see your protein, check for bands at oligomeric sizes (it may be showing as a dimer or trimer).
- If you have very low intensity bands for your protein, you may need to use a larger volume for your time point measurements. 20 mL was sufficient for the PopZ project, but you may need to increase your volume to 50 mL or more depending on the protein.
- If you are using Coomassie staining to analyze your gels and cannot see your desired bands, you may need to switch to a more sensitive visualization technique like Western blotting.

(B) OD_{600} Optimization

OD_{600} optimization will require a different setup from the induction time optimization since your *E. coli* growths will be induced at different times. I would recommend having 6 small growth batches of 50 mL each. If you can, try to stagger step 1 by about an hour between each growth. Otherwise you might get bottlenecked at the centrifuge.

1. Assuming a transformation was performed, *E. coli* have been plated, and precultures have been grown, make a 1:250 dilution of preculture growth to fresh LB supplemented with the appropriate antibiotic. Begin monitoring the OD_{600}.
2. Grow one batch to OD_{600} = ~0.3, the second to ~0.4, etc., up to ~0.8. The number of intervals you have will depend on how many batches you decided to do.
3. If you are optimizing your expression in LB, you can skip to step 5. If you are optimizing your expression in minimal media, then when each batch reaches its desired OD_{600}, centrifuge the bacteria at 7,700 rcf for 30 minutes at 4 ºC. Pour off the supernatant and resuspend the pellet in minimal media.
4. Allow the bacteria to equilibrate at induction conditions for 30 minutes before adding IPTG (Again, I would recommend 21 ºC, with shaking at 225 rpm).
5. Induce protein expression with IPTG (~1 mM) for the desired induction time.
6. Pellet the cells with centrifugation at 7,700 rcf and 4 ºC for 30 minutes. The supernatant can be poured off and the pellet should be frozen at -80 ºC.
7. After all OD_{600} batches have been collected and frozen, remove them from the freezer, thaw on ice, and then follow the lysozyme lysis protocol (separate document).

-note: the lysozyme lysis protocol cannot be used if your target protein is the same size as lysozyme. A different lysis protocol will need to be used (i.e. French press).

8. Analyze the lysates with SDS-PAGE followed by Coomassie stain, silver stain, or Western blot. Band intensity is representative of the quantity of protein produced. Therefore, the most intense band corresponding to your protein will represent the optimized OD_{600}.
9. If needed, perform an induction time optimization (described above) or induction temperature optimization (described below).

Troubleshooting tips:

- If there are no bands where you expect to see your protein, check for bands at oligomeric sizes (it may be showing as a dimer or trimer).
- If you have very low intensity bands for your protein, you may need to use a larger volume for your batches. Instead of 50 mL you may need to use 100.
- If you are optimizing in minimal media, you might be limited by the MM:LB ratio. For example, if you have a 1:4 MM to LB ratio, then your initial LB growth volume of 50 mL will end up having a volume of 12.5 mL in minimal media. This may not be enough to get decent bands in your gels or blots.

(C) Induction Temperature Optimization

The number of different induction temperatures tested may be limited due to the number of shakers and/or temperature setups you have available. For example, our lab has two

shakers, so we can test two temperatures at a time. If we use the warm room too we can do three. If you want more temperature points you may need to perform this optimization over multiple days.

1. Assuming a transformation was performed, *E. coli* have been plated, and precultures have been grown, transfer 4 mL of preculture growth per 1 L fresh LB supplemented with the appropriate antibiotic. Grow these to the desired OD_{600}.
2. If you are optimizing your expression in LB, you can skip to step 4. If you are optimizing your expression in minimal media, then centrifuge the bacteria at 7,700 rcf for 30 minutes at 4 ºC. Pour off the supernatant and resuspend the pellet(s) in minimal media.
3. Separate the minimal media into separate batches (one for each temperature point you are testing). Allow the bacteria to equilibrate at induction conditions for 30 minutes before adding IPTG (shaking at 225 rpm and the desired temperatures).
4. Induce protein expression with IPTG (~1 mM) for the desired induction time.
5. Pellet the cells with centrifugation at 7,700 rcf and 4 ºC for 30 minutes. The supernatant can be poured off and the pellets should be frozen at -80 ºC.
6. After all temperature batches have been collected and frozen, remove them from the freezer, thaw on ice, and then follow the lysozyme lysis protocol (separate document).

    -note: the lysozyme lysis protocol cannot be used if your target protein is the same size as lysozyme. A different lysis protocol will need to be used (i.e. French press).

7. Analyze the lysates with SDS-PAGE followed by Coomassie stain, silver stain, or Western blot. Band intensity is representative of the quantity of protein produced. Therefore, the most intense band corresponding to your protein will represent the optimized induction temperature.
8. If needed, perform an OD_{600} optimization or induction time optimization described above.

Troubleshooting tips:

- If there are no bands where you expect to see your protein, check for bands at oligomeric sizes (it may be showing as a dimer or trimer).
- If you are using Coomassie staining to analyze your gels and cannot see your desired bands, you may need to switch to a more sensitive visualization technique like Western blotting.

Protocol: Cell Lysis with Lysozyme

Lysozyme Buffer:

1. 10 mg lysozyme per 1 ml of buffer. Any of the following buffers will work: Tris HCl, $Na_xH_yPO_4$, HEPES, MOPS

    -The pH of the buffer needs to be between 6 and 9 (preferably 8) for the lysozyme to function properly.

2. Add Halt protease inhibitor cocktail to buffer, 1:100 v/v.
3. Add 2 – 5 µl of Benzonase nuclease (i.e. DNAse), depending on total buffer volume.
4. Completely dissolve lysozyme and protease inhibitor cocktail in buffer.

    -Do **NOT** Vortex! Remember lysozyme is a protein and can be damaged by heavy vortexing.

5. If you are **NOT** purifying the sample via Ni-affinity, the buffer should also contain 1 mM EDTA.

    -This will help with the percent yield as EDTA helps with lysozyme activity.

6. If you **ARE** purifying via Ni-affinity, **DO NOT** add EDTA.

    -EDTA will prevent His-Tag binding.

Lysis:

1. Completely resuspend cell pellets with lysozyme lysis buffer.

    -The volume buffer will depend on pellet size.

    -Assuming a small pellet (eg. 50 ml conical tube of growth centrifuged down to pellet), 1 ml of lysis buffer per pellet is sufficient.

-Do **NOT** Vortex! Remember protein can be damaged by heavy vortexing.

2. Freeze cell suspension in -80 °C freezer until completely frozen.
3. Thaw cells is a water bath.

   -Keep temperature below 45 °C.

4. Gently shake sample by hand to ensure homogeneity of cells in buffer.
5. Repeat steps 2 and 3.

   -If you are not using are not using a lysozyme susceptible cell line such as BL21(DE3)pLysS, you may want repeat an additional time or supplement with the use of gentle sonication at the end of step 3.

6. Centrifuge the lysate and separate the pellet and supernatant.

   -If you are running an optimization you will want to keep the pellet to analyze.

7. Collect supernatant for further use (i.e. purification, whole cell lysate gels and blotting, etc).

PopZ Expression and Purification Protocol

PopZ Expression

1. Plate PopZ freezer stocks using quadrant streaking onto Luria-Bertani (LB) medium agar plates containing 50 µg/mL kanamycin.

    -the freezer stock for PopZ is labeled as HW 225 in a 2 mL cryogenic tube.

    -Freezer stocks need to be kept on ice (buried so that only the cap is above the ice) to prevent freeze/thaw cycles of the glycerol stock. Minimize the amount of time the stocks are kept out of the freezer.

2. Incubate plates at 37 °C overnight for approximately 14 hours.

3. Scrape single, isolated bacterial colonies from the plates and drop in approximately 30 mL liquid LB medium supplemented with 30 µg/mL kanamycin for preculture growths.

    -Make a total of 100 mL LB separated into 3 conical tubes. Use the remaining 10 mL for a UV-Vis blank during the main growth.

4. Shake precultures in an orbital shaker at 225 rpm and 37 °C for 14-16 hours.

5. Transfer 4 mL of preculture growth to 1 L fresh liquid LB supplemented with 30 µg/mL kanamycin for a main growth.

    -It is recommended to do a total of 5 L of growth separated into 5 flasks.

6. Shake main growths at 225 rpm and 37 °C until the optical density at 600 nm (OD_{600}) reaches 0.50-0.60.

7. Centrifuge the cells at 7,700 rcf and 4 °C for 30 minutes. Pour off the supernatant and resuspend cell pellets in minimal media (MM) at a 1:4 ratio of MM:LB.

-For uniformly labeled-^{15}N and both uniformly labeled-^{13}C, ^{15}N enriched expression of $PopZ^{\Delta 134\text{-}177}$, supplement MM with either $^{15}NH_4Cl$ (for uniform labeling with ^{15}N) or $^{15}NH_4Cl$ and $^{13}C_6$-glucose (for uniform labeling with ^{15}N and ^{13}C), respectively.

8. Equilibrate cells under induction conditions for 30 minutes with shaking at 225 rpm and 37 °C before inducing protein expression with 0.8 mM isopropyl ß-D-1-thiogalactopyranoside (IPTG) for 12 hours.

9. Centrifuge cells at 7,700 rcf and 4 °C for 30 minutes. Discard the supernatant and store cells at -80 °C.

PopZ Purification

1. Thaw a cell pellet on ice and resuspend it in 5 mL Washing buffer (25 mM Tris, 150 mM NaCl, 20 mM imidazole at pH 7.5) supplemented with 50 µL Halt EDTA-free protease inhibitor and 5 µL Benzonase nuclease.

2. Lyse cells using a French press in triplicate at 1500 psi.

3. Centrifuge lysate at 27,200 rcf and 4 °C for 40 minutes. Keep the supernatant and concentrate it to 5 mL using a 10K Molecular Weight Cutoff (MWCO) centrifugal filter and then filter the lysate with a 0.22 µm syringe filter to remove large particulates.

4. Purify the protein with a round of Ni-affinity chromatography using a Fast Protein Liquid Chromatography instrument (FPLC; GE Healthcare ÄKTA purifier 900) equipped with a GE Healthcare HisTrap HP 1 ml column. Elute the column-bound $PopZ^{\Delta 134\text{-}177}$ using an increasing imidazole concentration gradient up to 500 mM with Elution buffer (25 mM Tris, 150 mM NaCl, 500 mM imidazole, pH 7.5).

5. Remove elution fractions containing the PopZ peak and pool them together. Buffer exchange the sample back into Washing buffer (using two Zeba columns) for a second round of purification (see step 4). Remove an aliquot for SDS-PAGE before purifying the second time.

6. Remove elution fractions containing the PopZ peak and pool them together. Buffer exchange the sample into either PopZ NMR buffer or Binding Partner NMR buffer (depending on what experiments are needed for NMR). The NMR buffer recipes are shown below. Remove an aliquot for SDS-PAGE.

7. Analyze the purity of the sample by SDS-PAGE and visualize with Coomassie blue

-PopZ$^{\Delta 134\text{-}177}$ separates at its dimer size (~30 kDa).

RcdA Expression and Purification Protocol

Sumo-RcdA Expression

1. Plate Sumo-RcdA freezer stocks using quadrant streaking onto Luria-Bertani (LB) medium agar plates containing 50 µg/mL kanamycin.

    -the freezer stock for Sumo-RcdA is labeled as HW 293 in a 2 mL cryogenic tube.

    -Freezer stocks need to be kept on ice (buried so that only the cap is above the ice) to prevent freeze/thaw cycles of the glycerol stock. Minimize the amount of time the stocks are kept out of the freezer.

2. Incubate plates at 37 °C overnight for approximately 14 hours.

3. Scrape single, isolated bacterial colonies from the plates and drop in approximately 30 mL liquid LB medium supplemented with 30 µg/mL kanamycin for preculture growths.

    -Make a total of 100 mL LB separated into 3 conical tubes. Use the remaining 10 mL for a UV-Vis blank during the main growth.

4. Shake precultures in an orbital shaker at 225 rpm and 37 °C for 14-16 hours.

5. Transfer 4 mL of preculture growth to 1 L fresh liquid LB supplemented with 30 µg/mL kanamycin for a main growth.

    -It is recommended to do a total of 5 L of growth separated into 5 flasks.

6. Shake main growths at 225 rpm and 37 °C until the OD_{600} reaches approximately 0.50.

7. Induce protein expression with 0.8 mM isopropyl ß-D-1-thiogalactopyranoside (IPTG) for 12 hours with shaking at 225 rpm and 21 °C.

9. Centrifuge cells at 7,700 rcf and 4 °C for 30 minutes. Discard the supernatant and store cells at -80 °C.

Sumo-RcdA Purification

1. Thaw a cell pellet on ice and resuspend it in 5 mL Washing buffer (20 mM HEPES, 100 mM KCl, 2 mM $MgCl_2$, 20 mM imidazole at pH 7.5) supplemented with 50 µL Halt EDTA-free protease inhibitor and 5 µL Benzonase nuclease.

2. Lyse cells using a French press in triplicate at 1500 psi.

3. Centrifuge lysate at 27,200 rcf and 4 °C for 40 minutes. Keep the supernatant and concentrate it to 5 mL using a 10K Molecular Weight Cutoff (MWCO) centrifugal filter and then filter the lysate with a 0.22 µm syringe filter to remove large particulates.

4. Purify the protein with a round of Ni-affinity chromatography using a Fast Protein Liquid Chromatography instrument (FPLC; GE Healthcare ÄKTA purifier 900) equipped with a GE Healthcare HisTrap HP 1 ml column. Elute the column-bound Sumo-RcdA using an increasing imidazole concentration gradient up to 500 mM with Elution buffer (20 mM HEPES, 100 mM KCl, 2 mM $MgCl_2$, 500 mM imidazole at pH 7.5).

5. Remove elution fractions containing the Sumo-RcdA peak and pool them together. Buffer exchange the sample back into Washing buffer (using two Zeba columns) for a second round of purification (see step 4). Remove an aliquot for SDS-PAGE before purifying the second time.

6. Remove elution fractions containing the Sumo-RcdA peak and pool them together. Concentrate the sample to about 3 mL using a 10K MWCO centrifugal filter. Then buffer exchange the sample into cleavage buffer (25 mM Tris, 150 mM NaCl, at pH 7.5). Remove an aliquot for SDS-PAGE.

Sumo-RcdA Cleavage

1. Calculate the concentration of Sumo-RcdA using UV-Vis spectrophotometry (ε = 19940 $M^{-1}cm^{-1}$). Then calculate the yield in mg.

2. Add 1 Unit Sumo Express Protease (Lucigen) per 200 µg Sumo-RcdA. Dilute the sample to 5 mL and then add 2 mM dithiothreitol (DTT). Gently mix the solution on an oscillator at 4 °C for 15 hours.

3. Purify the cleaved RcdA with a round of Ni-affinity chromatography (see step 4 from **Sumo-RcdA Purification** above). The cleaved RcdA will elute in the first few fractions because it does not contain a HisTag. Remove an aliquot for SDS-PAGE.

4. Buffer exchange the sample into HMK NMR buffer (20 mM HEPES, 100 mM KCl, 2 mM $MgCl_2$ at pH 7.5).

ADDITIONAL WORK

Structural Analysis of the Regulatory GAF Domains of cGMP Phosphodiesterase Elucidates the Allosteric Communication Pathway

Richa Gupta, Yong Liu, Huanchen Wang, Christopher T. Nordyke, Ryan Z. Puterbaugh, Wenjin Cui, Krisztina Varga, Feixia Chu, Hengming Ke, Harish Vashisth, and Rick H. Cote

Citation

Gupta, R. *et al*. Structural Analysis of the Regulatory GAF Domains of cGMP Phosphodiesterase Elucidates the Allosteric Communication Pathway. *J. Mol. Biol.* **432**, 5765-5783, doi:10.1016/j.jmb.2020.08.026 (2020).

Abstract

Regulation of photoreceptor phosphodiesterase (PDE6) activity is responsible for the speed, sensitivity, and recovery of the photoresponse during visual signaling in vertebrate photoreceptor cells. It is hypothesized that physiological differences in the light responsiveness of rods and cones may result in part from differences in the structure and regulation of the distinct isoforms of rod and cone PDE6. Although rod and cone PDE6 catalytic subunits share a similar domain organization consisting of tandem GAF domains (GAFa and GAFb) and a catalytic domain, cone PDE6 is a homodimer whereas rod PDE6 consists of two homologous catalytic subunits. Here we provide the x-ray crystal structure of cone GAFab regulatory domain solved at 3.3 angstrom resolution, in conjunction with chemical cross-linking and mass spectrometric analysis of conformational changes to GAFab induced upon binding of cGMP and

the PDE6 inhibitory gamma-subunit (P gamma). Ligand-induced changes in cross-linked residues implicate multiple conformational changes in the GAFa and GAFb domains in forming an allosteric communication network. Molecular dynamics simulations of cone GAFab revealed differences in conformational dynamics of the two subunits forming the homodimer and allosteric perturbations on cGMP binding. Cross-linking of P gamma to GAFab in conjunction with solution NMR spectroscopy of isotopically labeled P gamma identified the central polycationic region of P gamma interacting with the GAFb domain. These results provide a mechanistic basis for developing allosteric activators of PDE6 with therapeutic implications for halting the progression of several retinal degenerative diseases.

Contributions

In this publication, the section entitled "**Characterization of cone Pγ1-58 and its binding to GAFab by solution NMR spectroscopy"** was the focus of my contribution to this collaborative work. The work described therein is similar to the binding studies described in Chapter 4 of this dissertation. I collected 1D and 2D NMR data, processed 1D, 2D, and 3D NMR data, contributed to assignment of spectra, and analyzed binding data. Pγ1-58 was isotopically labeled with either ^{15}N or $^{15}N/^{13}C$ isotopes during protein expression in order to be observed with solution NMR spectroscopy. The ^{1}H-^{15}N HSQC spectral profile indicated that Pγ1-58 was intrinsically disordered. Spectral assignments were performed and confirmed by Richa Gupta and Ryan Puterbaugh such that 81% of the protein backbone was assigned. The unassigned residues consisted of nine proline residues and the first two amino acid residues of the protein. Assignments were used to investigate

binding residues when isotopically labeled Pγ1-58 was in solution with its unlabeled GAFab partner. Binding studies indicated that the primary binding interface was in the center of the Pγ1-58 sequence (residues F28-S38). This was indicated by pronounced signal attenuation exhibited by the representative peaks. Residues surrounding the F28-S38 region, as well as the C-terminus, exhibited smaller attenuation which is likely due to indirect or weaker binding by those regions. Finally, the N-terminal region showed very little spectral change.

We performed similar binding studies between Pγ1-58 and GAFab, but while in the presence of cyclic guanosine monophosphate (cGMP). cGMP binding to GAFab enhances binding activity of Pγ1-58. Consistent with this, we observed stronger universal signal attenuation of Pγ1-58 spectral peaks while in the presence of both GAFab and cGMP.

Functional Studies of the Translocator Protein TSPO from *Rhodobacter sphaeroides*

Christopher T. Nordyke, Nora Susanti, Jan Kubelka, Mark Gomelsky, and Krisztina Varga

Manuscript currently in preparation.

Abstract

The translocator protein (TSPO) is a transmembrane protein of high medical interest due to its association with a number of different cancers, neurological disorders, and other diseases. For this reason, TSPO is already exploited as a drug target for treating various diseases and conditions or as a biomarker for disease prognosis. Like TSPO, many other membrane proteins act as drug targets for diseases, with as many as 60% of all drugs currently targeting this class of protein. However, there is less information available on these proteins than on other protein classes. Here, we report functional data on the TSPO homologue from *Rhodobacter sphaeroides* (*Rs*TSPO). Other TSPO homologues have been shown to catalyze the photodegradation of protoporphyrin IX (PPIX), and we found that *Rs*TSPO likewise catalyzes this photodegradation. Additionally, we have expanded the investigation to include five other porphyrins and provide insight into the mechanism defining the porphyrin degradation.

Contributions

The *Rs*TSPO project was the initial project I worked on during my graduate career besides the research presented in the bulk of this dissertation. For this project I expressed and

purified *Rs*TSPO and performed experiments including fluorescence quenching to analyze binding between *Rs*TSPO and various porphyrins, *Rs*TSPO mediated degradation of porphyrins, and identification of the reactive oxygen species (ROS) participating in degradation reactions.

*Rs*TSPO was expressed and purified with a similar protocol to PopZ, though being a membrane protein required additional steps to extract *Rs*TSPO from the *E. coli* membrane into detergent micelles. Porphyrins used in binding and degradation experiments included protoporphyrin IX (PPIX), magnesium (II) protoporphyin IX (MgPPIX), zinc (II) protoporphyrin IX (ZnPPIX), N-methyl protoporphyrin IX (NMeP), copper protoporphyrin IX (CuPPIX), and hemin. Our fluorescence quenching experiments indicate that all six porphyrins bind to *Rs*TSPO. These porphyrins naturally degrade in the presence of light and oxygen, but the degradation reaction is catalyzed by *Rs*TSPO. Our results indicated that PPIX, MgPPIX, and ZnPPIX all degrade very quickly while in the presence of *Rs*TSPO, NMeP degrades slowly, and very little to no degradation is observed for CuPPIX and hemin. Calculations of binding constants and reaction constants are currently underway.

We aimed to identify the reactive oxygen species that participated in the degradation reactions using the ROS probe singlet oxygen sensor green (SOSG). Upon reaction with singlet oxygen, the unactivated SOSG, which does not show green fluorescence, becomes activated and exhibits green fluorescence emission at 533 nm when excited with 508 nm radiation. Our results showed enhanced SOSG activation during degradation experiments, indicating that the ROS necessary for porphyrin degradation is singlet oxygen.

Structural and Functional Characterization of Antifreeze Protein *Ap*AFP752

Korth W. Elliott, Christopher T. Nordyke, Jonathan A. Sreter, Katarina Jović, Logan M. Brown, Seth W. McNutt, Emily Asenath-Smith, Feixia Chu, Paul W. Baures, John G. Tsavalas, and Krisztina Varga

Manuscript currently in preparation.

Abstract

Antifreeze proteins (AFPs) are a unique class of proteins characterized by their ability to adsorb to the surface of ice crystals and prevent their growth. Antifreeze proteins are common in organisms that live in extremely cold climates (e.g. plants, fish, insects, etc.) and contribute to the freeze resistance of these organisms. Antifreeze proteins have great potential in the application of cryoprotection in medicine, agriculture, and food science. Antifreeze proteins have been studied for over five decades, however, their exact mechanism of action is still under investigation. In this study, we have characterized the antifreeze protein *Ap*AFP752 from the beetle *Anatolica polita* to better understand insect antifreeze protein activity. Additionally, we have structurally characterized the protein using solution NMR spectroscopy and found it to have a β-helical structure, similar to other known insect antifreeze proteins.

Contributions

My major contributions to this project consisted of NMR analyses that were similar to many already exhibited in Chapter 3 of this dissertation. Previous lab member, Korth W.

Elliott, performed the majority of sample preparation, NMR data collection and processing, and NMR spectral assignments of the *Ap*AF752 protein. I contributed to assignment confirmation. Additionally, I contributed to sample preparation for NMR RDC experiments and data analysis for RDC, T_1/T_2 relaxation, heteronuclear NOE, and ^1H temperature coefficients.

My contributions also included performing structural calculations using the structural generator server CS-Rosetta. The server utilized the chemical shift assignments, RDC constraints, and NOESY distance constraints as inputs for the calculations. The generated structural ensemble showed that the majority of the protein consisted of a β-barrel secondary structure with flexible N- and C-termini. The suspected ice-binding surface consists of TCT and TCI motifs that form parallel β-sheets. The cysteine residues at the center of these motifs form disulfide bonds across the center of the barrel with other cysteine residues on the opposing side, thus providing structural stability to the β-barrel. Multiple NOESY spectral crosspeaks representing correlations between barrel strands supported the existence of this secondary structure.

NMR Characterization of the Regulator of G-protein Signaling 8 (RGS8)

Ryan Z. Puterbaugh, Christopher T. Nordyke, Harish Vashisth, and Krisztina Varga

Abstract

G-protein coupled receptors (GPCRs) mediate signaling initiated by ligand binding and are coupled to G-proteins. They are involved in a wide range of organ systems, and consequently, are associated with a number of different cardiovascular, neurological, and visual disorders. Regulators of G-protein signaling (RGSs) are proteins that control the lifetime of G-protein signaling by binding to activated G-protein α-subunits, thereby accelerating GTP hydrolysis and inhibiting GPCR signaling. Thus, RGS proteins are a therapeutic target for enhancing GPCR signaling. Thiadiazolidinone (TDZD) inhibitors are a class of small molecule inhibitors that covalently bind to RGS proteins, making them viable candidates for therapies targeting these proteins. Here, we investigate the RGS isoform called RGS8 and its inhibition by two TDZD inhibitors through nuclear magnetic resonance (NMR) and molecular dynamics (MD) studies. We aimed to identify the role protein dynamics had in providing inhibitor access to the binding site and its participation in allosteric modulation. Additionally, we probed the roles of the two RGS8 cysteine residues in inhibitor recognition.

Contributions

One of the goals of this project was to characterize binding of the RGS8 protein with inhibitor ligands. Modeling indicated that domains of the protein would open to reveal a binding site that the ligands would then irreversibly bind to, and so NMR experiments

were performed in an effort to test this hypothesis. An NMR technique that can be utilized to probe this dynamic feature of the protein is called Carr-Purcell-Meiboom-Gill (CPMG) relaxation dispersion. My main role in this project was to contribute to experimental design and data analysis, with an emphasis on initiating and performing CPMG data analysis with a program called GUARDD (Graphical User-friendly Analysis of Relaxation Dispersion Data) and also to train other users. The program was designed to facilitate investigations into CPMG data by organizing, automating, and enhancing analytical processes.

REFERENCES

1 Mullineaux, C. W. How bacteria keep proteins moving. *Elife* **6**, doi:10.7554/eLife.33590 (2017).

2 Mignolet, J., Panis, G. & Viollier, P. H. More than a Tad: spatiotemporal control of Caulobacter pili. *Curr. Opin. Microbiol.* **42**, 79-86, doi:10.1016/j.mib.2017.10.017 (2018).

3 Ardissone, S. & Viollier, P. H. Interplay between flagellation and cell cycle control in Caulobacter. *Curr. Opin. Microbiol.* **28**, 83-92, doi:10.1016/j.mib.2015.08.012 (2015).

4 Kazmierczak, B. I., Schniederberend, M. & Jain, R. Cross-regulation of Pseudomonas motility systems: the intimate relationship between flagella, pili and virulence. *Curr. Opin. Microbiol.* **28**, 78-82, doi:10.1016/j.mib.2015.07.017 (2015).

5 Ringgaard, S., Yang, W., Alvarado, A., Schirner, K. & Briegel, A. Chemotaxis arrays in Vibrio species and their intracellular positioning by the ParC/ParP system. *J. Bacteriol.* **200**, e00793-00717, doi:10.1128/JB.00793-17 (2018).

6 Ramachandran, R., Jha, J. & Chattoraj, D. K. Chromosome segregation in Vibrio cholerae. *J. Mol. Microbiol. Biotechnol.* **24**, 360-370, doi:10.1159/000368853 (2014).

7 Kloosterman, T. G. *et al.* Complex polar machinery required for proper chromosome segregation in vegetative and sporulating cells of Bacillus subtilis. *Mol. Microbiol.* **101**, 333-350, doi:10.1111/mmi.13393 (2016).

8 Werner, J. N. *et al.* Quantitative genome-scale analysis of protein localization in an asymmetric bacterium. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 7858-7863, doi:10.1073/pnas.0901781106 (2009).

9 Collier, J. Cell division control in Caulobacter crescentus. *Biochim Biophys Acta Gene Regul Mech* **1862**, 685-690, doi:10.1016/j.bbagrm.2018.04.005 (2019).

10 Zhou, X., Wang, J., Herrmann, J., Moerner, W. E. & Shapiro, L. Asymmetric division yields progeny cells with distinct modes of regulating cell cycle-dependent chromosome methylation. *Proc Natl Acad Sci U S A* **116**, 15661-15670, doi:10.1073/pnas.1906119116 (2019).

11 Bodenmiller, D., Toh, E. & Brun, Y. V. Development of surface adhesion in Caulobacter crescentus. *J. Bacteriol.* **186**, 1438-1447, doi:10.1128/jb.186.5.1438-1447.2004 (2004).

12 Reyes Ruiz, L. M., Fiebig, A. & Crosson, S. Regulation of bacterial surface attachment by a network of sensory transduction proteins. *PLoS Genet.* **15**, e1008022, doi:10.1371/journal.pgen.1008022 (2019).

13 Nyarko, A., Barton, H. & Dhinojwala, A. Scaling down for a broader understanding of underwater adhesives - a case for the Caulobacter crescentus holdfast. *Soft Matter* **12**, 9132-9141, doi:10.1039/c6sm02163h (2016).

14 Meier, E. L., Razavi, S., Inoue, T. & Goley, E. D. A novel membrane anchor for FtsZ is linked to cell wall hydrolysis in Caulobacter crescentus. *Mol. Microbiol.* **101**, 265-280, doi:10.1111/mmi.13388 (2016).

15 Laub, M. T., Chen, S. L., Shapiro, L. & McAdams, H. H. Genes directly controlled by CtrA, a master regulator of the Caulobacter cell cycle. *Proc Natl Acad Sci U S A* **99**, 4632-4637, doi:10.1073/pnas.062065699 (2002).

16 Hottes, A. K., Shapiro, L. & McAdams, H. H. DnaA coordinates replication initiation and cell cycle transcription in Caulobacter crescentus. *Mol. Microbiol.* **58**, 1340-1353, doi:10.1111/j.1365-2958.2005.04912.x (2005).

17 McAdams, H. H. & Shapiro, L. System-level design of bacterial cell cycle control. *FEBS Lett.* **583**, 3984-3991, doi:10.1016/j.febslet.2009.09.030 (2009).

18 Tan, M. H., Kozdon, J. B., Shen, X., Shapiro, L. & McAdams, H. H. An essential transcription factor, SciP, enhances robustness of Caulobacter cell cycle regulation. *Proc Natl Acad Sci U S A* **107**, 18985-18990, doi:10.1073/pnas.1014395107 (2010).

19 Berge, M. & Viollier, P. H. End-in-Sight: Cell Polarization by the Polygamic Organizer PopZ. *Trends Microbiol.* **26**, 363-375, doi:10.1016/j.tim.2017.11.007 (2018).

20 Holmes, J. A. *et al.* Caulobacter PopZ forms an intrinsically disordered hub in organizing bacterial cell poles. *Proc Natl Acad Sci U S A* **113**, 12490-12495, doi:10.1073/pnas.1602380113 (2016).

21 Treuner-Lange, A. & Sogaard-Andersen, L. Regulation of cell polarity in bacteria. *J Cell Biol* **206**, 7-17, doi:10.1083/jcb.201403136 (2014).

22 Bowman, G. R. *et al.* Oligomerization and higher-order assembly contribute to sub-cellular localization of a bacterial scaffold. *Mol. Microbiol.* **90**, 776-795, doi:10.1111/mmi.12398 (2013).

23 Ebersbach, G., Briegel, A., Jensen, G. J. & Jacobs-Wagner, C. A self-associating protein critical for chromosome attachment, division, and polar organization in caulobacter. *Cell* **134**, 956-968, doi:10.1016/j.cell.2008.07.016 (2008).

24 Bowman, G. R. *et al.* A polymeric protein anchors the chromosomal origin/ParB complex at a bacterial cell pole. *Cell* **134**, 945-955, doi:10.1016/j.cell.2008.07.015 (2008).

25 Lenarcic, R. *et al.* Localisation of DivIVA by targeting to negatively curved membranes. *EMBO J* **28**, 2272-2282, doi:10.1038/emboj.2009.129 (2009).

26 Strahl, H. *et al.* Transmembrane protein sorting driven by membrane curvature. *Nat Commun* **6**, 8728, doi:10.1038/ncomms9728 (2015).

27 Grangeon, R., Zupan, J. R., Anderson-Furgeson, J. & Zambryski, P. C. PopZ identifies the new pole, and PodJ identifies the old pole during polar growth in Agrobacterium tumefaciens. *Proc Natl Acad Sci U S A* **112**, 11666-11671, doi:10.1073/pnas.1515544112 (2015).

28 Berge, M. *et al.* Modularity and determinants of a (bi-)polarization control system from free-living and obligate intracellular bacteria. *Elife* **5**, doi:10.7554/eLife.20640 (2016).

29 Perez, A. M. *et al.* A Localized Complex of Two Protein Oligomers Controls the Orientation of Cell Polarity. *mBio* **8**, doi:10.1128/mBio.02238-16 (2017).

30 Mignolet, J. *et al.* Functional dichotomy and distinct nanoscale assemblies of a cell cycle-controlled bipolar zinc-finger regulator. *Elife* **5**, doi:10.7554/eLife.18647 (2016).

31 Thanbichler, M. & Shapiro, L. MipZ, a spatial regulator coordinating chromosome segregation with cell division in Caulobacter. *Cell* **126**, 147-162, doi:10.1016/j.cell.2006.05.038 (2006).

32 Kiekebusch, D., Michie, K. A., Essen, L.-O., Löwe, J. & Thanbichler, M. Localized dimerization and nucleoid binding drive gradient formation by the bacterial cell division inhibitor MipZ. *Mol. Cell* **46**, 245-259, doi:10.1016/j.molcel.2012.03.004 (2012).

33 Sanselicio, S., Berge, M., Theraulaz, L., Radhakrishnan, S. K. & Viollier, P. H. Topological control of the Caulobacter cell cycle circuitry by a polarized single-domain PAS protein. *Nat Commun* **6**, 7005, doi:10.1038/ncomms8005 (2015).

34 Bowman, G. R. *et al.* Caulobacter PopZ forms a polar subdomain dictating sequential changes in pole composition and function. *Mol. Microbiol.* **76**, 173-189, doi:10.1111/j.1365-2958.2010.07088.x (2010).

35 Ozaki, S. *et al.* Activation and polar sequestration of PopA, a c-di-GMP effector protein involved in Caulobacter crescentus cell cycle control. *Mol. Microbiol.* **94**, 580-594, doi:10.1111/mmi.12777 (2014).

36 Holmes, J. A. *et al. Caulobacter* PopZ forms an intrinsically disordered hub in organizing bacterial cell poles. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 12490-12495, doi:10.1073/pnas.1602380113 (2016).

37 Bhattarai, A. & Emerson, I. A. Dynamic conformational flexibility and molecular interactions of intrinsically disordered proteins. *J. Biosci.* **45** (2020).

38 Oldfield, C. J. & Dunker, A. K. Intrinsically Disordered Proteins and Intrinsically Disordered Protein Regions. *Annual Review of Biochemistry, Vol 83* **83**, 553-584, doi:10.1146/annurev-biochem-072711-164947 (2014).

39 Peng, Z. *et al.* Exceptionally abundant exceptions: comprehensive characterization of intrinsic disorder in all domains of life. *Cell Mol Life Sci* **72**, 137-151, doi:10.1007/s00018-014-1661-9 (2015).

40 Yruela, I., Oldfield, C. J., Niklas, K. J. & Dunker, A. K. Evidence for a Strong Correlation Between Transcription Factor Protein Disorder and Organismic Complexity. *Genome Biol. Evol.* **9**, 1248-1265, doi:10.1093/gbe/evx073 (2017).

41 Babu, M. M. The contribution of intrinsically disordered regions to protein function, cellular complexity, and human disease. *Biochem. Soc. Trans.* **44**, 1185-1200, doi:10.1042/bst20160172 (2016).

42 Cumberworth, A., Lamour, G., Babu, M. M. & Gsponer, J. Promiscuity as a functional trait: intrinsically disordered regions as central players of interactomes. *Biochem. J* **454**, 361-369, doi:10.1042/bj20130545 (2013).

43 Wright, P. E. & Dyson, H. J. Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* **16**, 18-29, doi:10.1038/nrm3920 (2015).

44 Uversky, V. N. The alphabet of intrinsic disorder: II. Various roles of glutamic acid in ordered and intrinsically disordered proteins. *Intrinsically Disord Proteins* **1**, e24684, doi:10.4161/idp.24684 (2013).

45 Tompa, P. & Fuxreiter, M. Fuzzy complexes: polymorphism and structural disorder in protein-protein interactions. *Trends Biochem Sci* **33**, 2-8, doi:10.1016/j.tibs.2007.10.003 (2008).

46 Fuxreiter, M. & Tompa, P. Fuzzy complexes: a more stochastic view of protein function. *Adv Exp Med Biol* **725**, 1-14, doi:10.1007/978-1-4614-0659-4_1 (2012).

47 Fuxreiter, M. Fold or not to fold upon binding - does it really matter? *Curr. Opin. Struct. Biol.* **54**, 19-25, doi:10.1016/j.sbi.2018.09.008 (2019).

48 Hsu, W. L. *et al.* Exploring the binding diversity of intrinsically disordered proteins involved in one-to-many binding. *Protein Sci.* **22**, 258-273, doi:10.1002/pro.2207 (2013).

49 Higo, K. *et al.* High resolution crystal structure of the Grb2 SH2 domain with a phosphopeptide derived from CD28. *PLoS One* **8**, e74482, doi:10.1371/journal.pone.0074482 (2013).

50 Inaba, S. *et al.* Crystal Structures and Thermodynamic Analysis Reveal Distinct Mechanisms of CD28 Phosphopeptide Binding to the Src Homology 2 (SH2) Domains of Three Adaptor Proteins. *J Biol Chem* **292**, 1052-1060, doi:10.1074/jbc.M116.755173 (2017).

51 Malhis, N., Jacobson, M. & Gsponer, J. MoRFchibi SYSTEM: software tools for the identification of MoRFs in protein sequences. *Nucleic Acids Res.* **44**, W488-W493, doi:10.1093/nar/gkw409 (2016).

52 Yang, J., Gao, M., Xiong, J., Su, Z. & Huang, Y. Features of molecular recognition of intrinsically disordered proteins via coupled folding and binding. *Protein Sci.* **28**, 1952-1965, doi:10.1002/pro.3718 (2019).

53 Mohan, A. *et al.* Analysis of molecular recognition features (MoRFs). *J. Mol. Biol.* **362**, 1043-1059, doi:10.1016/j.jmb.2006.07.087 (2006).

54 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, doi:10.1073/pnas.1512799112 (2015).

55 Furth, N., Aylon, Y. & Oren, M. p53 shades of Hippo. *Cell Death Differ.* **25**, 81-92, doi:10.1038/cdd.2017.163 (2018).

56 Grossman, S. R. p300/CBP/p53 interaction and regulation of the p53 response. *Eur. J. Biochem.* **268**, 2773-2778, doi:10.1046/j.1432-1327.2001.02226.x (2001).

57 Uversky, V. N. p53 Proteoforms and Intrinsic Disorder: An Illustration of the Protein Structure-Function Continuum Concept. *Int. J. Mol. Sci.* **17**, doi:10.3390/ijms17111874 (2016).

58 Nyqvist, I., Andersson, E. & Dogan, J. Role of Conformational Entropy in Molecular Recognition by TAZ1 of CBP. *J. Phys. Chem. B* **123**, 2882-2888, doi:10.1021/acs.jpcb.9b01343 (2019).

59 O'Shea, C. *et al.* Structures and Short Linear Motif of Disordered Transcription Factor Regions Provide Clues to the Interactome of the Cellular Hub Protein Radical-induced Cell Death1. *J. Biol. Chem.* **292**, 512-527, doi:10.1074/jbc.M116.753426 (2017).

60 Jespersen, N. *et al.* Systematic identification of recognition motifs for the hub protein LC8. *Life Sci. Alliance* **2**, doi:10.26508/lsa.201900366 (2019).

61 Nagulapalli, M., Maji, S., Dwivedi, N., Dahiya, P. & Thakur, J. K. Evolution of disorder in Mediator complex and its functional relevance. *Nucleic Acids Res.* **44**, 1591-1612, doi:10.1093/nar/gkv1135 (2016).

62 Yan, J., Dunker, A. K., Uversky, V. N. & Kurgan, L. Molecular recognition features (MoRFs) in three domains of life. *Mol. Biosyst.* **12**, 697-710, doi:10.1039/c5mb00640f (2016).

63 Meszaros, B., Tompa, P., Simon, I. & Dosztanyi, Z. Molecular principles of the interactions of disordered proteins. *J. Mol. Biol.* **372**, 549-561, doi:10.1016/j.jmb.2007.07.004 (2007).

64 Wallin, S. Intrinsically disordered proteins: structural and functional dynamics. *Res. Rep. Biol.* **8**, 7-16, doi:10.2147/rrb.S57282 (2017).

65 Elbaum-Garfinkle, S. *et al.* The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics. *Proc Natl Acad Sci U S A* **112**, 7189-7194, doi:10.1073/pnas.1504822112 (2015).

66 Wang, F. *et al.* Structures of KIX domain of CBP in complex with two FOXO3a transactivation domains reveal promiscuity and plasticity in coactivator recruitment. *Proc Natl Acad Sci U S A* **109**, 6078-6083, doi:10.1073/pnas.1119073109 (2012).

67 Lee, C. W., Arai, M., Martinez-Yamout, M. A., Dyson, H. J. & Wright, P. E. Mapping the interactions of the p53 transactivation domain with the KIX domain of CBP. *Biochemistry* **48**, 2115-2124, doi:10.1021/bi802055v (2009).

68 Bardwell, J. C. & Jakob, U. Conditional disorder in chaperone action. *Trends Biochem Sci* **37**, 517-525, doi:10.1016/j.tibs.2012.08.006 (2012).

69 Winter, J., Ilbert, M., Graf, P. C., Ozcelik, D. & Jakob, U. Bleach activates a redox-regulated chaperone by oxidative protein unfolding. *Cell* **135**, 691-701, doi:10.1016/j.cell.2008.09.024 (2008).

70 Tapley, T. L., Franzmann, T. M., Chakraborty, S., Jakob, U. & Bardwell, J. C. Protein refolding by pH-triggered chaperone binding and release. *Proc Natl Acad Sci U S A* **107**, 1071-1076, doi:10.1073/pnas.0911610107 (2010).

71 De Biasio, A. *et al.* p15PAF is an intrinsically disordered protein with nonrandom structural preferences at sites of interaction with other proteins. *Biophys. J.* **106**, 865-874, doi:10.1016/j.bpj.2013.12.046 (2014).

72 Santofimia-Castano, P. *et al.* Targeting intrinsically disordered proteins involved in cancer. *Cellular and Molecular Life Sciences* **77**, 1695-1707, doi:10.1007/s00018-019-03347-3 (2020).

73 Terol, G. L., Gallego-Jara, J., Martinez, R. A. S., Diaz, M. C. & Puente, T. d. D. Engineering protein production by rationally choosing a carbon and nitrogen source using E. coli BL21 acetate metabolism knockout strains. *Microbial Cell Factories* **18**, doi:https://doi.org/10.1186/s12934-019-1202-1 (2019).

74 Hoffmann, B., Lohr, F., Laguerre, A., Bernhard, F. & Dotsch, V. Protein labeling strategies for liquid-state NMR spectroscopy using cell-free synthesis. *Prog. Nucl. Magn. Reson. Spectrosc.* **105**, 1-22, doi:10.1016/j.pnmrs.2017.11.004 (2018).

75 Liu, D., Xu, R. & Cowburn, D. Segmental isotopic labeling of proteins for nuclear magnetic resonance. *Methods Enzymol.* **462**, 151-175, doi:10.1016/S0076-6879(09)62008-5 (2009).

76 Butt, T. R., Edavettal, S. C., Hall, J. P. & Mattern, M. R. SUMO fusion technology for difficult-to-express proteins. *Protein Expr. Purif.* **43**, 1-9, doi:10.1016/j.pep.2005.03.016 (2005).

77 Malakhov, M. P. *et al.* SUMO fusions and SUMO-specific protease for efficient expression and purification of proteins. *J Struct Funct Genomics* **5**, 75-86, doi:10.1023/B:JSFG.0000029237.70316.52 (2004).

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

79 Lee, W., Tonelli, M. & Markley, J. L. NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics (Oxford, England)* **31**, 1325-1327, doi:10.1093/bioinformatics/btu830 (2015).

80 Bahrami, A., Assadi, A. H., Markley, J. L. & Eghbalnia, H. R. Probabilistic Interaction Network of Evidence Algorithm and its Application to Complete Labeling of Peak Lists from Protein NMR Spectroscopy. *PLOS Computational Biology* **5**, e1000307, doi:10.1371/journal.pcbi.1000307 (2009).

81 Lee, W., Westler, W. M., Bahrami, A., Eghbalnia, H. R. & Markley, J. L. PINE-SPARKY: graphical interface for evaluating automated probabilistic peak assignments in protein NMR spectroscopy. *Bioinformatics (Oxford, England)* **25**, 2085-2087, doi:10.1093/bioinformatics/btp345 (2009).

82 Uversky, V. N. Protein intrinsic disorder and structure-function continuum. *Prog. Mol. Biol. Transl. Sci.* **166**, 1-17, doi:10.1016/bs.pmbts.2019.05.003 (2019).

83 Watson, J. D. *et al.* Towards fully automated structure-based function prediction in structural genomics: a case study. *J. Mol. Biol.* **367**, 1511-1522, doi:10.1016/j.jmb.2007.01.063 (2007).

84 Redfern, O. C., Dessailly, B. & Orengo, C. A. Exploring the structure and function paradigm. *Curr. Opin. Struct. Biol.* **18**, 394-402, doi:10.1016/j.sbi.2008.05.007 (2008).

85 Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235-242 (2000).

86 Kendrew, J. C. & Parrish, R. G. The Crystal Structure of Myoglobin. III. Sperm-Whale Myoglobin. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* **238**, 305-324 (1957).

87 Smyth, M. S. & Martin, J. H. x ray crystallography. *Mol. Pathol.* **53**, 8-14, doi:10.1136/mp.53.1.8 (2000).

88 Williamson, M. P., Havel, T. F. & Wuthrich, K. Solution conformation of proteinase inhibitor IIA from bull seminal plasma by 1H nuclear magnetic resonance and distance geometry. *J. Mol. Biol.* **182**, 295-315, doi:10.1016/0022-2836(85)90347-x (1985).

89 Nwanochie, E. & Uversky, V. N. Structure Determination by Single-Particle Cryo-Electron Microscopy: Only the Sky (and Intrinsic Disorder) is the Limit. *Int. J. Mol. Sci.* **20**, doi:10.3390/ijms20174186 (2019).

90 Borcherds, W. M. & Daughdrill, G. W. Using NMR Chemical Shifts to Determine Residue-Specific Secondary Structure Populations for Intrinsically Disordered Proteins. *Methods Enzymol.* **611**, 101-136, doi:10.1016/bs.mie.2018.09.011 (2018).

91 Jacobsen, N. E. *NMR Spectroscopy Explained: Simplified Theory, Applications and Examples for Organic Chemistry and Structural Biology*. (John Wiley & Sons, Inc., 2007).

92 Levitt, M. H. *Spin Dynamics: Basics of Nuclear Magnetic Resonance*. (John Wiley & Sons Ltd., 2001).

93 Cavanagh, J., Fairbrother, W. J., Palmer, A. G., Rance, M. & Skelton, N. J. *Protein NMR Spectroscopy: Principles and Practice*. (Elsevier, 2007).

94 Nerli, S., McShan, A. C. & Sgourakis, N. G. Chemical shift-based methods in NMR structure determination. *Prog. Nucl. Magn. Reson. Spectrosc.* **106-107**, 1-25, doi:10.1016/j.pnmrs.2018.03.002 (2018).

95 Prestel, A., Bugge, K., Staby, L., Hendus-Altenburger, R. & Kragelund, B. B. Characterization of Dynamic IDP Complexes by NMR Spectroscopy. *Methods Enzymol.* **611**, 193-226, doi:10.1016/bs.mie.2018.08.026 (2018).

96 Song, J. *et al.* Intrinsically disordered γ-subunit of cGMP phosphodiesterase encodes functionally relevant transient secondary and tertiary structure. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 1505-1510, doi:10.1073/pnas.0709558105 (2008).

97 Feuerstein, S., Plevin, M. J., Willbold, D. & Brutscher, B. iHADAMAC: a complementary tool for sequential resonance assignment of globular and highly disordered proteins. *J. Magn. Reson.* **214**, 329-334, doi:10.1016/j.jmr.2011.10.019 (2012).

98 Yao, J., Dyson, H. J. & Wright, P. E. Chemical shift dispersion and secondary structure prediction in unfolded and partly folded proteins. *FEBS Lett.* **419**, 285-289, doi:http://dx.doi.org/10.1016/S0014-5793(97)01474-9 (1997).

99 Schubert, M., Labudde, D., Oschkinat, H. & Schmieder, P. A software tool for the prediction of Xaa-Pro peptide bond conformations in proteins based on 13C chemical shift statistics. *J. Biomol. NMR* **24**, 149-154, doi:10.1023/a:1020997118364 (2002).

100 Kosol, S., Contreras-Martos, S., Cedeno, C. & Tompa, P. Structural Characterization of Intrinsically Disordered Proteins by NMR Spectroscopy. *Molecules* **18**, 10802-10828, doi:10.3390/molecules180910802 (2013).

101 Marsh, J. A., Singh, V. K., Jia, Z. & Forman-Kay, J. D. Sensitivity of secondary structure propensities to sequence differences between alpha- and gamma-synuclein: Implications for fibrillation. *Protein Sci.* **15**, 2795-2804, doi:10.1110/ps.062465306 (2006).

102 Lee, W., Stark, J. L. & Markley, J. L. PONDEROSA-C/S: client–server based software package for automated protein 3D structure determination. *J. Biomol. NMR* **60**, 73-75, doi:10.1007/s10858-014-9855-x (2014).

103 Karamanos, T. K., Kalverda, A. P., Thompson, G. S. & Radford, S. E. Mechanisms of amyloid formation revealed by solution NMR. *Prog. Nucl. Magn. Reson. Spectrosc.* **88-89**, 86-104, doi:10.1016/j.pnmrs.2015.05.002 (2015).

104 Sugiki, T., Kobayashi, N. & Fujiwara, T. Modern Technologies of Solution Nuclear Magnetic Resonance Spectroscopy for Three-dimensional Structure

Determination of Proteins Open Avenues for Life Scientists. *Comput Struct Biotechnol J* **15**, 328-339, doi:10.1016/j.csbj.2017.04.001 (2017).

105 Gibbs, E. B., Cook, E. C. & Showalter, S. A. Application of NMR to studies of intrinsically disordered proteins. *Arch Biochem Biophys* **628**, 57-70, doi:10.1016/j.abb.2017.05.008 (2017).

106 Schneider, R. *et al.* Towards a robust description of intrinsic protein disorder using nuclear magnetic resonance spectroscopy. *Mol. Biosyst.* **8**, 58-68, doi:10.1039/c1mb05291h (2012).

107 Chiliveri, S. C. & Deshmukh, M. V. Recent excitements in protein NMR: Large proteins and biologically relevant dynamics. *J. Biosci.* **41**, 787-803, doi:10.1007/s12038-016-9640-y (2016).

108 Kovermann, M., Rogne, P. & Wolf-Watz, M. Protein dynamics and function from solution state NMR spectroscopy. *Q. Rev. Biophys.* **49**, e6, doi:10.1017/S0033583516000019 (2016).

109 Konrat, R. NMR contributions to structural dynamics studies of intrinsically disordered proteins. *J. Magn. Reson.* **241**, 74-85, doi:10.1016/j.jmr.2013.11.011 (2014).

110 Huang, S. *et al.* Utilization of paramagnetic relaxation enhancements for structural analysis of actin-binding proteins in complex with actin. *Sci. Rep.* **6**, 33690, doi:10.1038/srep33690 (2016).

111 Prestegard, J. H., Bougault, C. M. & Kishore, A. I. Residual dipolar couplings in structure determination of biomolecules. *Chem Rev* **104**, 3519-3540, doi:10.1021/cr030419i (2004).

112 Bax, A. Weak alignment offers new NMR opportunities to study protein structure and dynamics. *Protein Sci.* **12**, 1-16 (2003).

113 Chen, K. & Tjandra, N. The use of residual dipolar coupling in studying proteins by NMR. *Top. Curr. Chem.* **326**, 47-67, doi:10.1007/128_2011_215 (2012).

114 Gobl, C., Madl, T., Simon, B. & Sattler, M. NMR approaches for structural analysis of multidomain proteins and complexes in solution. *Prog. Nucl. Magn. Reson. Spectrosc.* **80**, 26-63, doi:10.1016/j.pnmrs.2014.05.003 (2014).

115 Bax, A. & Grishaev, A. Weak alignment NMR: a hawk-eyed view of biomolecular structure. *Curr. Opin. Struct. Biol.* **15**, 563-570, doi:10.1016/j.sbi.2005.08.006 (2005).

116 Jensen, M. R. *et al.* Quantitative conformational analysis of partially folded proteins from residual dipolar couplings: application to the molecular recognition element of Sendai virus nucleoprotein. *J. Am. Chem. Soc.* **130**, 8055-8061, doi:10.1021/ja801332d (2008).

117 Roche, J., Louis, J. M. & Bax, A. Conformation of inhibitor-free HIV-1 protease derived from NMR spectroscopy in a weakly oriented solution. *ChemBioChem* **16**, 214-218, doi:10.1002/cbic.201402585 (2015).

118 Niu, X. *et al.* Arginine kinase: joint crystallographic and NMR RDC analyses link substrate-associated motions to intrinsic flexibility. *J. Mol. Biol.* **405**, 479-496, doi:10.1016/j.jmb.2010.11.007 (2011).

119 Vammi, V., Lin, T. L. & Song, G. Enhancing the quality of protein conformation ensembles with relative populations. *J. Biomol. NMR* **58**, 209-225, doi:10.1007/s10858-014-9818-2 (2014).

120 Chan-Yao-Chong, M., Durand, D. & Ha-Duong, T. Molecular Dynamics Simulations Combined with Nuclear Magnetic Resonance and/or Small-Angle X-ray Scattering Data for Characterizing Intrinsically Disordered Protein Conformational Ensembles. *J. Chem. Inf. Model.* **59**, 1743-1758, doi:10.1021/acs.jcim.8b00928 (2019).

121 Bertini, I. *et al.* Paramagnetism-based NMR restraints provide maximum allowed probabilities for the different conformations of partially independent protein domains. *J. Am. Chem. Soc.* **129**, 12786-12794, doi:10.1021/ja0726613 (2007).

122 Mohana-Borges, R., Goto, N. K., Kroon, G. J. A., Dyson, H. J. & Wright, P. E. Structural characterization of unfolded states of apomyoglobin using residual dipolar couplings. *J. Mol. Biol.* **340**, 1131-1142, doi:10.1016/j.jmb.2004.05.022 (2004).

123 Bhattacharya, A., Tejero, R. & Montelione, G. T. Evaluating protein structures determined by structural genomics consortia. *Proteins: Struct. Funct. Bioinform.* **66**, 778-795, doi:10.1002/prot.21165 (2007).

124 Das, R. K. & Pappu, R. V. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc Natl Acad Sci U S A* **110**, 13392-13397, doi:10.1073/pnas.1304749110 (2013).

125 Martin, E. W. *et al.* Sequence Determinants of the Conformational Properties of an Intrinsically Disordered Protein Prior to and upon Multisite Phosphorylation. *J. Am. Chem. Soc.* **138**, 15323-15335, doi:10.1021/jacs.6b10272 (2016).

126 Nicholas, M. P., Eryilmaz, E., Ferrage, F., Cowburn, D. & Ghose, R. Nuclear spin relaxation in isotropic and anisotropic media. *Prog. Nucl. Magn. Reson. Spectrosc.* **57**, 111-158, doi:10.1016/j.pnmrs.2010.04.003 (2010).

127 Jarymowycz, V. A. & Stone, M. J. Fast time scale dynamics of protein backbones: NMR relaxation methods, applications, and functional consequences. *Chem. Rev.* **106**, 1624-1671, doi:10.1021/cr040421p (2006).

128 Lee, D., Hilty, C., Wider, G. & Wuthrich, K. Effective rotational correlation times of proteins from NMR relaxation interference. *J. Magn. Reson.* **178**, 72-76, doi:10.1016/j.jmr.2005.08.014 (2006).

129 Cho, M. K. *et al.* Amino acid bulkiness defines the local conformations and dynamics of natively unfolded alpha-synuclein and tau. *J. Am. Chem. Soc.* **129**, 3032-3033, doi:10.1021/ja067482k (2007).

130 Zook, J. D. *et al.* High-resolution NMR reveals secondary structure and folding of amino acid transporter from outer chloroplast membrane. *PLoS One* **8**, e78116, doi:10.1371/journal.pone.0078116 (2013).

131 Williamson, M. P. Using chemical shift perturbation to characterise ligand binding. *Prog. Nucl. Magn. Reson. Spectrosc.* **73**, 1-16, doi:10.1016/j.pnmrs.2013.02.001 (2013).

132 Agback, P. *et al.* Structural characterization and biological function of bivalent binding of CD2AP to intrinsically disordered domain of chikungunya virus nsP3 protein. *Virology* **537**, 130-142, doi:10.1016/j.virol.2019.08.022 (2019).

133 D'Silva, L. *et al.* Monitoring the effects of antagonists on protein-protein interactions with NMR spectroscopy. *J. Am. Chem. Soc.* **127**, 13220-13226, doi:10.1021/ja052143x (2005).

134 Berge, M. *et al.* Modularity and determinants of a (bi-)polarization control system from free-living and obligate intracellular bacteria. *Elife* **5**, 31, doi:10.7554/eLife.20640 (2016).

135 Perez, A. M. *et al.* A Localized Complex of Two Protein Oligomers Controls the Orientation of Cell Polarity. *Mbio* **8**, 16, doi:10.1128/mBio.02238-16 (2017).

136 Wang, H. B. & Bowman, G. R. SpbR overproduction reveals the importance of proteolytic degradation for cell pole development and chromosome segregation in Caulobacter crescentus. *Mol. Microbiol.* **111**, 1700-1714, doi:10.1111/mmi.14247 (2019).

137 McGrath, P. T., Iniesta, A. A., Ryan, K. R., Shapiro, L. & McAdams, H. H. A dynamically localized protease complex and a polar specificity factor control a cell cycle master regulator. *Cell* **124**, 535-547, doi:10.1016/j.cell.2005.12.033 (2006).

138 Joshi, K. K., Berge, M., Radhakrishnan, S. K., Viollier, P. H. & Chien, P. An Adaptor Hierarchy Regulates Proteolysis during a Bacterial Cell Cycle. *Cell* **163**, 419-431, doi:10.1016/j.cell.2015.09.030 (2015).

139 Joshi, K. K., Battle, C. M. & Chien, P. Polar Localization Hub Protein PopZ Restrains Adaptor-Dependent ClpXP Proteolysis in Caulobacter crescentus. *J. Bacteriol.* **200**, doi:10.1128/jb.00221-18 (2018).

140 Taylor, J. A., Wilbur, J. D., Smith, S. C. & Ryan, K. R. Mutations that alter RcdA surface residues decouple protein localization and CtrA proteolysis in Caulobacter crescentus. *J. Mol. Biol.* **394**, 46-60, doi:10.1016/j.jmb.2009.08.076 (2009).

141 Xu, H. B. *et al.* The MDM2-Binding Region in the Transactivation Domain of p53 Also Acts as a Bcl-X-L-Binding Motif. *Biochemistry* **48**, 12159-12168, doi:10.1021/bi901188s (2009).

142 Kriwacki, R. W., Hengst, L., Tennant, L., Reed, S. I. & Wright, P. E. Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity. *Proc Natl Acad Sci U S A* **93**, 11504-11509, doi:10.1073/pnas.93.21.11504 (1996).

143 Ptacin, J. L. *et al.* Bacterial scaffold directs pole-specific centromere segregation. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E2046-2055, doi:10.1073/pnas.1405188111 (2014).

144 Blair, J. A. *et al.* Branched Signal Wiring of an Essential Bacterial Cell-Cycle Phosphotransfer Protein. *Structure* **21**, 1590-1601, doi:10.1016/j.str.2013.06.024 (2013).

145 Tsokos, C. G., Perchuk, B. S. & Laub, M. T. A dynamic complex of signaling proteins uses polar localization to regulate cell-fate asymmetry in Caulobacter crescentus. *Dev. Cell* **20**, 329-341, doi:10.1016/j.devcel.2011.01.007 (2011).

146 Iniesta, A. A., McGrath, P. T., Reisenauer, A., McAdams, H. H. & Shapiro, L. A phospho-signaling pathway controls the localization and activity of a protease complex critical for bacterial cell cycle progression. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 10935-10940, doi:10.1073/pnas.0604554103 (2006).

147 Lasker, K. *et al.* Selective sequestration of signalling proteins in a membraneless organelle reinforces the spatial regulation of asymmetry in Caulobacter crescentus. *Nat Microbiol* **5**, 418-429, doi:10.1038/s41564-019-0647-7 (2020).

148 Lori, C. *et al.* Cyclic di-GMP acts as a cell cycle oscillator to drive chromosome replication. *Nature* **523**, 236-239, doi:10.1038/nature14473 (2015).

149 Abel, S. *et al.* Regulatory cohesion of cell cycle and cell differentiation through interlinked phosphorylation and second messenger networks. *Mol. Cell* **43**, 550-560, doi:10.1016/j.molcel.2011.07.018 (2011).

150 Dahal, L., Shammas, S. L. & Clarke, J. Phosphorylation of the IDP KID Modulates Affinity for KIX by Increasing the Lifetime of the Complex. *Biophys. J.* **113**, 2706-2712, doi:10.1016/j.bpj.2017.10.015 (2017).

151 Yadahalli, S. *et al.* Kinetic and thermodynamic effects of phosphorylation on p53 binding to MDM2. *Sci. Rep.* **9**, 693, doi:10.1038/s41598-018-36589-5 (2019).

152 Sharma, S. *et al.* Affinity switching of the LEDGF/p75 IBD interactome is governed by kinase-dependent phosphorylation. *Proc Natl Acad Sci U S A* **115**, E7053-E7062, doi:10.1073/pnas.1803909115 (2018).

153 Kussie, P. H. *et al.* Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain. *Science* **274**, 948-953, doi:10.1126/science.274.5289.948 (1996).

154 Follis, A. V. *et al.* PUMA binding induces partial unfolding within BCL-xL to disrupt p53 binding and promote apoptosis. *Nat. Chem. Biol.* **9**, 163-168, doi:10.1038/nchembio.1166 (2013).

155 Day, C. L. *et al.* Structure of the BH3 domains from the p53-inducible BH3-only proteins Noxa and Puma in complex with Mcl-1. *J. Mol. Biol.* **380**, 958-971, doi:10.1016/j.jmb.2008.05.071 (2008).

156 Zor, T., De Guzman, R. N., Dyson, H. J. & Wright, P. E. Solution structure of the KIX domain of CBP bound to the transactivation domain of c-Myb. *J. Mol. Biol.* **337**, 521-534, doi:10.1016/j.jmb.2004.01.038 (2004).

157 Liu, X., Chen, J. & Chen, J. Residual Structure Accelerates Binding of Intrinsically Disordered ACTR by Promoting Efficient Folding upon Encounter. *J. Mol. Biol.* **431**, 422-432, doi:10.1016/j.jmb.2018.12.001 (2019).

158 Giri, R., Morrone, A., Toto, A., Brunori, M. & Gianni, S. Structure of the transition state for the binding of c-Myb and KIX highlights an unexpected order for a disordered system. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 14942-14947, doi:10.1073/pnas.1307337110 (2013).

159 Bienkiewicz, E. A., Adkins, J. N. & Lumb, K. J. Functional consequences of preorganized helical structure in the intrinsically disordered cell-cycle inhibitor p27(Kip1). *Biochemistry* **41**, 752-759, doi:10.1021/bi015763t (2002).

160 Schumacher, M. A., Huang, K.-H., Zeng, W. & Janakiraman, A. Structure of the Z Ring-associated Protein, ZapD, Bound to the C-terminal Domain of the Tubulin-like Protein, FtsZ, Suggests Mechanism of Z Ring Stabilization through FtsZ Cross-linking. *J. Bio. Chem.* **292**, 3740-3750, doi:10.1074/jbc.M116.773192 (2017).

161 Mosyak, L. *et al.* The bacterial cell-division protein ZipA and its interaction with an FtsZ fragment revealed by X-ray crystallography. *EMBO J.* **19**, 3179-3191, doi:10.1093/emboj/19.13.3179 (2000).

162 Schumacher, M. A. & Zeng, W. Structures of the nucleoid occlusion protein SlmA bound to DNA and the C-terminal domain of the cytoskeletal protein FtsZ. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 4988-4993, doi:10.1073/pnas.1602327113 (2016).

163 Hsu, W.-L. *et al.* Exploring the binding diversity of intrinsically disordered proteins involved in one-to-many binding. *Protein Science: A Publication of the Protein Society* **22**, 258-273, doi:10.1002/pro.2207 (2013).

164 Karlsson, E. *et al.* A structurally heterogeneous transition state underlies coupled binding and folding of disordered proteins. *J. Bio. Chem.* **294**, 1230-1239, doi:10.1074/jbc.RA118.005854 (2019).

165 Lacy, E. R. *et al.* p27 binds cyclin-CDK complexes through a sequential mechanism involving binding-induced protein folding. *Nat. Struct. Mol. Biol.* **11**, 358-364, doi:10.1038/nsmb746 (2004).

166 Bowman, G. R. *et al.* Caulobacter PopZ forms a polar subdomain dictating sequential changes in pole composition and function. *Mol. Microbiol.* **76**, 173-189, doi:10.1111/j.1365-2958.2010.07088.x (2010).

167 Laloux, G. & Jacobs-Wagner, C. Spatiotemporal control of PopZ localization through cell cycle-coupled multimerization. *J. Cell Biol.* **201**, 827-841, doi:10.1083/jcb.201303036 (2013).

168 Warfield, L., Tuttle, L. M., Pacheco, D., Klevit, R. E. & Hahn, S. A sequence-specific transcription activator motif and powerful synthetic variants that bind Mediator using a fuzzy protein interface. *Proc. Natl. Acad. Sci. U. S. A.* **111**, E3506-E3513, doi:10.1073/pnas.1412088111 (2014).

169 Rogers, J. M. *et al.* Interplay between partner and ligand facilitates the folding and binding of an intrinsically disordered protein. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 15420-15425, doi:10.1073/pnas.1409122111 (2014).

170 Tuttle, L. M. *et al.* Gcn4-Mediator Specificity Is Mediated by a Large and Dynamic Fuzzy Protein-Protein Complex. *Cell Rep.* **22**, 3251-3264, doi:10.1016/j.celrep.2018.02.097 (2018).

171 Dogan, J., Mu, X., Engstrom, A. & Jemth, P. The transition state structure for coupled binding and folding of disordered protein domains. *Sci. Rep.* **3**, 6, doi:10.1038/srep02076 (2013).

172 Crabtree, M. D. *et al.* Conserved Helix-Flanking Prolines Modulate Intrinsically Disordered Protein:Target Affinity by Altering the Lifetime of the Bound Complex. *Biochemistry* **56**, 2379-2384, doi:10.1021/acs.biochem.7b00179 (2017).

173 Shoemaker, B. A., Portman, J. J. & Wolynes, P. G. Speeding molecular recognition by using the folding funnel: the fly-casting mechanism. *Proc Natl Acad Sci U S A* **97**, 8868-8873, doi:10.1073/pnas.160259697 (2000).

174 Harmon, T. S., Holehouse, A. S., Rosen, M. K. & Pappu, R. V. Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins. *eLife* **6**, doi:10.7554/eLife.30294 (2017).

175 Kribelbauer, J. F., Rastogi, C., Bussemaker, H. J. & Mann, R. S. Low-Affinity Binding Sites and the Transcription Factor Specificity Paradox in Eukaryotes. *Annu. Rev. Cell Dev. Biol.* **35**, 357-379, doi:10.1146/annurev-cellbio-100617-062719 (2019).