The Pennsylvania State University

The Graduate School

Eberly College of Science

DEVELOPMENT OF NOVEL METHODOLOGY TO DETERMINE THE FUNCTIONAL ROLE OF DISORDER IN THE FCP1/RAP74 BINDING INTERACTION

A Dissertation in

Chemistry

by

Chad W. Lawrence

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

December 2013

The dissertation of Chad W. Lawrence was reviewed and approved* by the following:

Scott Showalter
Associate Professor of Chemistry
Dissertation Advisor
Chair of Committee

Mary Beth Williams
Professor of Chemistry
Associate Dean for Undergraduate Education, Eberly College of Science

Carsten Krebs
Professor of Chemistry
Associate Professor of Biochemistry and Molecular Biology

John H. Golbeck
Professor of Biochemistry and Biophysics
Professor of Chemistry

Barbara J. Garrison
Shapiro Proffesor of Chemistry
Head of the Chemistry Department

*Signatures on file in the Graduate School

Abstract

Modern structural biology currently holds the paradigm that structure determines function. Intrinsically disordered proteins are a unique class of proteins that forces the structure/function paradigm to be expanded. IDPs, unlike globular proteins, have no unique secondary or tertiary structure and yet carry out vital biological functions such as transcription, translation, and cellular signal transduction. The unique structural properties associated with IDPs are due to a sequence bias towards charged, hydrophilic amino acids and away from hydrophobic amino acids. Given of this bias, there are little to no hydrophobic amino acids to be removed from the protein/solvent interface to form a hydrophobic core, which is the primary motivation for protein folding. However, IDPs are known to undergo a cooperative folding upon binding process during protein-protein or protein-nucleic acid interactions. Here we will study FCP1, which is one such IDP that gains helical structure upon its interaction with RAP74, a globular protein.

TFIIF-associating RNA Polymerase II C-terminal domain phosphatase (FCP1) is an IDP that plays a vital role in the RNA Polymerase II (RNAPII) transcription cycle. During the elongation cycle of transcription the C-terminal domain (CTD) of RNAPII becomes increasingly phosphorylated and until it becomes dephosphorylated RNAPII cannot undergo another round of transcription. Interaction with the winged helix domain of RAP74, a subdomain of TFIIF, brings FCP1 into contact with CTD so that it may carry out this dephosphorylation. The residues that compose the binding interface of FCP1 have been shown in the literature to adopt an α-helical structure while in complex with RAP74. However, there is little information regarding the structure and dynamics of FCP1 in its disordered unbound state and what role that disorder plays in the overall binding mechanism.

^{13}C-direct detected NMR provides an alternative to traditional ^{1}H-detected NMR for the study of IDP structure and dynamics. Due to the low sequence complexity known for IDPs and

the low chemical shift dispersion associated with the ^{1}H nucleus, the NMR spectra of IDPs suffer from severe overlap. This property has prevented many IDPs from being structurally and dynamically characterized by NMR. ^{13}C-direct detected NMR, on the other hand, has been shown by our lab and others to have a large chemical shift dispersion. Using ^{13}C direct-detected NMR methods our lab has been able to acquire full chemical shift assignment FCP1 in the unbound state. Here we will show that even in the apo-state of FCP1, the residues that lie within the RAP74 binding interface show nascent helical structure through CD and recently developed ^{13}C-direct detected NMR. We will also show that FCP1 remains disordered while in 30% Dextran and HeLa cell extract, representing a cellular environment. Given the highly flexible nature of FCP1 in its disordered state, it is necessary to be able to characterize the local dynamic information of FCP1. We will show through the development of novel ^{13}C-direct detected experiments CON(T1)-IPAP and CON(T2)-IPAP accurate T_1 and T_2 relaxation times can be measured for FCP1 in both the apo and holo-state. After analysis of the relaxation data, only the residues of FCP1 in the RAP74 binding region experience a change in dynamics. Additional ^{15}N spin relaxation of RAP74, studied in parallel to FCP1, show that RAP74 experiences only a limited amount of ordering upon association with FCP1. In order to determine what role disorder plays in the binding interaction, the FCP1/RAP74 binding interaction was characterized by ITC.

ITC is a powerful method that allows one to determine the thermodynamic parameters that govern a binding interaction. Here we will show that the FCP1/RAP74 binding interaction is dominated by the removal of hydrophobic surface from solvent, characterized by the large, negative ΔC_P associated with complex formation. Additionally, regardless of the entropic penalty associated with FCP1 acquiring structure, the overall interaction is entropically favorable. We will also show through the use of osmolytes to stabilize secondary structure, that pre-formed helical structure in the binding region of FCP1 has a large effect on the overall binding interaction.

To fully understand the behavior of IDPs in solution one must be able to generate a structure. However, IDPs do not exist as a statically folded globular domain, but, instead, exist as an ensemble of rapidly interconverting conformers. The experimental data reported is only reflective of the ensemble average of FCP1 and not of any of the individual conformers that make up its ensemble structure. So, ensemble modeling then becomes the preferred approach to structurally characterize the FCP1 ensemble. we will show that through a combination of SAXS and the ensemble modeling software EOM, flexible-meccano, and ENSEMBLE that the overall ensemble structure is dominated by the by completely disordered N-terminus and that a random coil model, with a locally fixed structural propensity is sufficient to describe it.

Contents

List of Figures

List of Tables

List of Equations

List of Abbreviations

Å – angstrom

aa – amino acid

BRCT – BRCA1 C-terminal domain

C' – backbone carbonyl carbon of a given amino acid

C^{α} – alpha carbon of a given amino acid

C^{β} – beta carbon of a given amino acid

CD – circular dichroism

CK2 – Casein kinase 2

CSI – chemical shift index

CTD – C-terminal domain

ctFCP1 – C-terminal domain of FCP1

$\Delta\delta$ – change in 2D peak postion

ΔC_p – the change in heat capacity of a binding interaction

ΔG – the change in Gibbs free energy of a binding interaction

ΔH – the change in enthalpy of a binding interaction

ΔS – the change in entropy of a binding interaction

ΔT – difference in temperature

DNA – deoxyribonucleic acid

EOM – ensemble optimization method

FCP1 – TFIIF associating C-terminal domain phosphatase

FCPH – FCP1 homology domain

GB3 – B3 domain of protein G

GTF – general transcription factor

HSQC – heteronuclear single quantum coherence

IDP – intrinsically disordered protein

IPAP – in-phase anti-phase

ITC – isothermal titration calorimetry

$J_{C'-C\alpha}$ – one bond scalar coupling between C' and C^{α} carbons

$J(\omega)$ – spectral density function

K_D – dissociation constant

$[L]_t$ – total ligand concentration

[L] – free ligand concentration

$[M]_t$ – total macromolecule concentration

MD – molecular dynamics

miRNA – micro RNA

MoRF – molecular recognition feature

mRNA – messenger RNA

n – the binding stoichiometry

NMR – nuclear magnetic resonance

NOE – Nuclear Overhauser Effect

PIC – pre-initiation complex

PRE – paramagnetic relaxation enhancement

Q – total heat of a binding interaction

R_1 – longitudinal relaxation rate

R_2 – transverse relaxation rate

RDC – residual dipolar coupling

R_g – radius of gyration

RNA – ribonucleic acid

RNAPII – RNA polymerase II

RAP30 – RNA polymerase associating protein 30

RAP74 – RNA polymerase associating protein 74

RMSD – root mean square deviation

S^2 – a general order parameter to describe the amplitude of motion of a given N-H bond vector

SAXS – small angle X-ray scattering

T – a given temperature

T_1 – longitudinal relaxation time

T_2 – transverse relaxation time

T_R – a reference temperature

τ_e – rate of internal motion of a given N-H bond vector

τ_m – global rotational rate of a protein

TBP – TATA binding protein

TFIIB – transcription factor IIB

TFIIF – transcription factor IIF

TFE – 2,2,2-trifluroethanol

TMAO – trimethylamine N-oxide

V – active volume of the sample cell

X – fraction of sites occupied

Chapter 1 Introduction

Intrinsically disordered proteins are relatively recent discoveries that have challenged the preconceptions of structural biology and the structure-function paradigm. This class of proteins contains little to no unique secondary structure and yet carries out important biological function. One such intrinsically disordered protein is FCP1, a phosphatase of the C-terminal domain of RNA Polymerase II, which binds to RAP74, a subdomain of Transcription Factor IIF, causing a conformational change in the binding interface of FCP1; acquiring helical structure. It is our goal, through the development of novel NMR experiments, ITC, and molecular modeling programs, to characterize this coupled folding and binding mechanism and determine what role disorder plays in the interaction as described in more detail below.

Extending the Structure-Function Paradigm

The paradigm that has dominated protein science for decades is that structure determines function. Over the course of determining the structures of thousands of proteins it has generally been found true that those sharing similar structure will most likely share a similar function, however, these findings are heavily biased. Most protein structures that are deposited in the Protein Databank have been determined by X-ray crystallography, while only a small number have had their structure solved via alternative methods such as NMR. This has created a selection bias for those proteins amenable to forming a well diffracting crystal. However, with advances in alternative methods for studying protein structure, those proteins that are not amenable to crystallography can now be studied. In particular, we can now study intrinsically disordered proteins, which challenge the structure-function paradigm, and uncover transient structural elements that these proteins may possess in their highly dynamic native state.

Intrinsically disordered proteins (IDPs) are a class of proteins that, unlike globular proteins, have little to no unique secondary or tertiary structure yet carry out biological functions[1,2,2–5]. Despite the perceived rarity of proteins that incorporate such a high degree of disorder in their structure, the human proteome is predicted to be ~30% disordered[3,6]. IDPs are characterized as having a primary sequence bias towards charged/hydrophilic amino acids and a lack of hydrophobic amino acids[7,8]. For cooperatively folding proteins, which have a higher enrichment in hydrophobic amino acids, the physical chemical properties of the protein-solvent interface result in the burial of non-polar residues, forming what is called the hydrophobic core[9,10]. The burial of hydrophobic amino acids away from the protein surface is considered one of the primary driving forces behind protein folding[9]; thus it follows that IDPs, which lack this significant force, do no fold while free in solution. Given this significant sequence bias and generally low sequence complexity, one can confidently identify whether a protein or regions of a protein will not adopt a regular structure. Also, given their lack of structure these proteins cannot be thought of in the traditional sense of a singular structure that governs the behavior of the protein. Instead IDPs must be considered as an ensemble of structures that the protein is rapidly interconverting between, in which a small subset of structures may or may not be functionally relevant[11–14]. Therefore methods need to be developed that allow the measurement of structural information for the ensemble that can then utilized to generate a model, in order to provide a visual representation of how this ensemble may appear in solution.

IDPs are regularly found to function in transcription, translation, and cellular signaling[14–17]; suggesting that disorder must be a functional necessity for these processes. There are several advantages to having a high degree of disorder that may have led to the evolutionary selection of these proteins. Disorder may provide a range of useful properties such as greater surface area for binding, a high specificity/low affinity binding interaction, inducibility by posttranslational modification, structural plasticity for binding multiple partners, and high dissociation rates[1,4,5].

However, the structural characteristics that make these proteins exciting in terms of expanding the structure-function paradigm also make them difficult to study by traditional structural biology methods, such as X-ray crystallography. Due to the highly dynamic nature of IDPs and the lack of regular structural elements all attempts at producing apo-state crystal structures would end in some non-biologically relevant structure induced by crystal formation. It is necessary that methods be developed to study these proteins in their natively disordered state. Otherwise the chemical mechanisms of biological processes that are governed by these proteins will remain obscured.

Coupled Folding and Binding

A coupled folding/binding event is very common among protein interactions involving IDPs[18–20]. For some IDPs, multiple, specific structures may be present in their ensemble state that are suitable for binding different partners[21]. This property has been considered as one of the reasons IDPs constitute a larger portion of the proteome for higher organisms. This of course motivates the hypothesis that IDPs such as p53 can bind multiple different proteins, with simultaneous high specificity, due to the unique properties that emerge from its natively disordered state[15,22]. Given the high degree of ambiguity associated with how IDPs behave in solution there are multiple proposed mechanisms that describe these coupled folding/binding events, one such mechanism is the "fly-casting" model[23,24]. Given that IDPs lack a hydrophobic core, these proteins exist in a more extended state, thus occupying more space than traditional globular protein when in solution. Under the "fly-casting" model, this creates a larger capture radius for an IDPs particular ligand and allow it to encounter said ligand sooner than a globular protein of similar size would. The initial interaction is predicted to stimulate a chain reaction of sorts, causing the IDP to progressively fold and bring the ligand closer to the IDP until the binding interface is fully folded and the interaction is complete[23]. Two additional binding models that attempt to describe the interactions of IDPs may be incorporated within the fly-casting

model: induced fit and conformational selection[25,26]. Induced fit describes a model much like the general "fly-casting" model. First the IDP binds in its disordered state and once sufficiently bound the protein folds into its complex structure[20,27]. Conformational selection, on the other hand, proposes that the ensemble state of an IDP must at least partially sample the structure of the bound state, forming molecular recognition features or "MoRFs"[25,26,28,29]. The binding protein recognizes these MoRFs and initiates binding. However, these models put very specific limits on how these highly dynamic proteins must behave which is why a more general model, such as the extended conformational selection model, is more preferred.

In the extended conformational selection model the ensemble state of an IDP is expressed in terms of an energy funnel where there are multiple minima that the IDP is interconverting between[30]. Using the idea that binding affects the energy landscape of an IDP, the previous models can be redefined with this concept. For induced fit, the IDP is occupying some position within its energy landscape before its partner binds. Binding then induces a conformational change upon the IDP forcing it to the energy minimum of the complex state[27]. For conformational selection, the IDP is changing position within the energy landscape where sometimes it is in the minimum associated with the complex and it is only when the IDP is in this position that binding will occur[31]. However, these models are limited because they focus only on the IDP and not how binding affects the binding partner. The extended model attempts to improve upon this by stating that both the IDP and its partner exist in various positions within their own individual energy landscapes. As they approach each other their energy landscapes change until they both reach a global minimum and binding occurs.

It is important to consider that even though IDPs may gain secondary structure upon binding that does not require them to be static while in a complex. In fact, there are many IDPs that still retain a rather dynamic nature in their bound state. The binding interactions formed by IDPs have been termed "fuzzy complexes" as a means to describe the structural ambiguity that

may be associated with them. The term is meant to encompass a full dynamic range, where the IDPs can adopt more static structures to those that retain a fully disordered state.

FCP1 and RAP74 as a Model System

RNA Polymerase II (RNAPII) is a eukaryotic enzyme that is responsible for transcribing messenger RNA (mRNA)[32–34]. The mRNA transcription cycle is composed of three stages: initiation, elongation, and termination. In order for RNAPII to enter these stages it comes into contact with various other proteins known as general transcription factors (GTFs)[35,36]. These GTFs are responsible for identifying transcription promoters in the DNA, bringing RNAPII into contact with the DNA and orienting it into a position for unidirectional transcription, and causing conformational changes within RNAPII[37]. One such GTF, transcription factor IIF (TFIIF) takes part in both the initiation and elongation stages of the transcription cycle[38].

TFIIF is a general transcription factor that forms a heterodimer composed of two subdomains: RNA Polymerase Associating Protein 30 (RAP30) and RAP74[39–41]. It has been previously determined that the RAP30/74 complex binds to RNAPII and escorts it to the TATA binding protein (TBP) and promoter in order to form the pre-initiation complex (PIC)[37,38]. RAP30 alone is capable of bringing RNAPII into the PIC as well as facilitating elongation. However, RAP74 provides full functionality of the GTF in both the PIC and elongation complex and prevents RNAPII binding nonspecifically to non-promoter regions[42,43]. RAP74 also helps to regulate elongation via phosphorylation and its interaction with a phosphatase of the C-terminal domain (CTD) of RNAPII, FCP1[36]. The FCP1 binding region of RAP74 is at its C-terminal domain and consists of 4 alpha helices and three beta strands arranged in an anti-parallel sheet forming a winged-helix structure[44–46].

FCP1, TFIIF associating C-terminal domain phosphatase, acts in the regulation of transcription by dephosphorylating the CTD of RNAPII[33,36,47–49]. The CTD of RANPII consists of

a repeated heptad with the consensus sequence YSPTSPS[33,50,51]. Posttranscriptional modification of this tail, particularly phosphorylation, regulates the activity of RNAPII. During the initiation

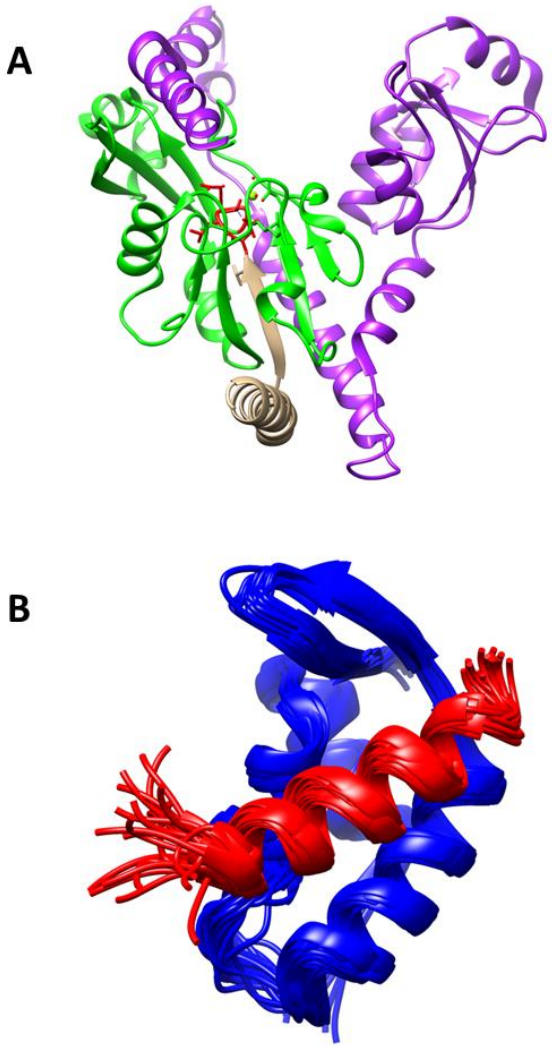

Figure 1.1: A X-ray crystal structure of the FCPH (green) and BRCT (purple) domains of FCP1 in Schizosaccharomyces pombe (A). The DxDxT binding motif that acts as the active site for dephosphorylation is shown in red. PDB: 3EF1. In (B) are the 20 lowest energy NMR structures for the C-terminal domain of RAP74 (blue) bound to the C-terminal domain of FCP1 (red). Note that this is the structure of FCP1$^{879\text{-}961}$, but a lack of long range, medium range, and intermolecular NOEs caused residues 879-940 to not be included in NMR structure calculations. PDB:1ONV

cycle the CTD is hypophosphorylated and upon entering into the elongation cycle the CTD becomes increasingly phosphorylated, particularly on the Ser 2 and Ser 5 of the heptad[33]. RNAPII cannot then enter back into the initiation cycle to start another round of transcription until it is then dephosphorylated[47]. FCP1 achieves this via its three functional domains that consist of an FCP homology domain (FCPH), a BRCA1 C-terminal domain (BRCT), and a disordered C-terminal tail (ctFCP1) shown in figure 1.1(A)[52]. FCP1 acts a phosphatase for RNAPII where the FCPH domain is the catalytic domain belonging to the DxDxT aspartylphosphotransferase superfamily[50,52]. The BRCT domain is thought to behave as a stabilizing agent for the catalytic domain through its interaction with the CTD[52]. ctFCP1 behaves as the regulator for phosphatase activity through its interactions with TFIIB and TFIIF as well as containing a nuclear transport motif allowing FCP1 to be concentrated in the nucleus[53,54]. RAP74 is the subdomain of TFIIF that interacts directly with ctFCP1 through its winged helix domain. Based on previous crystal and NMR data helices 2 and 3 and strand 2 come together to form a hydrophobic/basic cleft at the binding interface for FCP1[55,56]. When bound, ctFCP1 adopts an α-helical structure where the hydrophobic/acidic face of the ctFCP1 helix is complemented by the basic cleft of RAP74 ultimately forming two salt bridges between D947 and D959 of FCP1 and K471 and K498 of RAP74, respectively, figure 1.1(B)[56]. Since binding of RAP74 is necessary for FCP1 to be fully functional as a phosphatase for RNAPII CTD, it is necessary to gain insight into the disorder to order transition that is associated with this interaction.

In order to study this transition, acquisition of any structural information for FCP1 while free in solution or in complex with RAP74 is necessary. The very nature of IDPs does not make this a trivial matter, because it renders impractical the use of a standard set of techniques that are regularly used to study cooperatively folding systems. Therefore, it becomes imperative to develop a methodology that can be used to study FCP1, while it is disordered in its apo-state, as

well as while it is partially folded in complex. Here we will show that traditional NMR methods used to study globular proteins with ^{1}H-detection fail for IDPs due to their low sequence complexity and low chemical shift dispersion. However, newly developed NMR techniques that utilize ^{13}C-direct detection overcome the obstacles that are associated with ^{1}H detection. Using this detection method, we will show that there are transient structural elements that can be observed even while free in solution for FCP1 and that the amount of structure present can be fine-tuned using both kosmotropes and chaotropes as cosolutes. ^{13}C-direct detection can also be applied to the measurement of backbone dynamics, and we will show through modification of a standard CON-IPAP experiments both T_1 and T_2 relaxation rates can be measured. Combination of both the structural and dynamic NMR data with previously done molecular dynamics simulations will provide insight into what role "functional disorder" may play in the FCP1/RAP74 interaction. Also, we will use ITC to measure the physical parameters that describe the interaction. Solvent conditions will be adjusted using osmolytes to determine what effect disorder has on binding and how this may relate the overall binding mechanism. Finally, preliminary work towards generating an accurate ensemble structure of FCP1 has been under gone incorporating SAXS and molecular modeling software designed to generate an ensemble of conformers to describe IDP structure.

Technical Aspects

^{13}C-Direct Detection NMR Spectroscopy

As previously mentioned, the study of IDPs provides unique challenges because they are, for the most part, not amenable to traditional methods that study protein structure. In particular X-ray crystallography fails due to the nature of an IDP and its inability to form a biologically relevant crystal structure[57]. NMR then becomes the preferred method to study IDPs on the atomic scale. However, traditional NMR techniques used to study globular proteins that depend on the ^{1}H nucleus for detection are also inadequate to study IDPs, for which the traditional ^{15}N,^{1}H-

HSQC displays very low chemical shift dispersion and significant overlap[58–60]. Alternative NMR techniques that utilize heteronuclear detection represent a successful solution to this problem; here ^{13}C-direct detection is the preferred method to study this class of proteins, both structurally and dynamically[60,61,61]. ^{1}H has a very large gyromagnetic ratio compared to the biomolecular heteroatoms, ^{13}C and ^{15}N, which provides a significant increase in sensitivity making it a preferred nucleus for detection[62,63]. It is this fact that has caused most NMR method development to be aimed towards exploiting this sensitivity. Even though the enhanced sensitivity over other nuclei is an attractive advantage for ^{1}H detection there are significant disadvantages. Because of proton's high gyromagnetic ratio it is also a known source of large dipolar interactions that may broaden resonances in large or paramagnetic proteins[58,64]. Also, ^{1}H is known phenomenologically to have an intrinsically low chemical shift dispersion in IDPs (*vide infra*), making it a particularly limited detection nucleus for IDPs[59,65]. As I will show, ^{13}C-direct detection allows us to circumvent these problems.

^{13}C has a significantly lower gyromagnetic ratio making it less susceptible to large electronic magnetic moments associated with paramagnetic biomolecules, and is characterized by our laboratory and others as having a large chemical shift dispersion in IDPs[58–60,66]. While each carbon nucleus in a protein can be used as a detection platform, carbonyl (C') and alpha (C^{α}) carbon are typically used because their coupling topologies are not tied to amino acid type. In general, we find that C'-detection is particularly well suited for the study of IDPs.

Unlike in ^{1}H and ^{15}N experiments, which do not experience homonuclear couplings during data acquisition, it is important when designing ^{13}C-direct detect experiments to remove the strong one bond scalar coupling between C' and C^{α} along the protein backbone during acquisition[58,60]. For ^{13}C-detected experiments this can at least double the amount of signals in the direct acquisition dimension if special precaustions are not taken to achieve homodecoupling during acquisition. For example, C' signals are split by C'-C^{α} couplings while C^{α} is split by both

C'-C^{α} and C^{α}-C^{β} couplings, making the resultant spectra more complex. In order to collapse these multiplets into a single line shape, here, we have utilized the IPAP detection method as a means to "virtually" decouple C'-C^{α} one bond scalar coupling post acquisition, described schematically in figure 1.2[58]. This is accomplished by allowing the C'-C^{α} coupling to evolve and then separately collect the in-phase and anti-phase components in a single experiment. Post-acquisition, a linear sum and difference spectra is produced, such that there are two in-phase peaks separated by $J_{C'-C\alpha}$[58,60]. Subsequently, each peak is mathematically frequency shifted by $\pm (1/2)J_{C'-C\alpha}$, resulting in a final spectrum with a two dimensional absorption mode singlet line shape for each ^{13}C resonance. All NMR data that incorporated ^{13}C-direct detection came directly from or experiments built from a standard ^{13}C,^{15}N-CON-IPAP[58] shown in figure 1.3. Previously, we have used ^{13}C detection to fully assign over 90% of ctFCP1[67,68]. Here we will show how these chemical shifts can be used to probe the secondary structure accessed by FCP1.

Proteins are not static in solution, IDPs especially, and local motions may be directly tied to their functions[69,70]. Therefore, it is imperative to characterize these motions in order to completely determine what is necessary for a proteins function. NMR provides the opportunity to not only characterize structure but the dynamics of proteins through NMR spin relaxation and dispersion experiments[71–75]. To provide a physical picture of the internal motions of a protein, one typically considers the N-H bond as a proxy for motion of the rigid peptide plane while the rest of the protein is represented as a sphere[76–78]. While clearly not an appropriate model for IDPs, the spherical model will be used merely as an example of underlying principles, shown in figure 1.4. Both global and internal motions can be represented through what is known as a spectral density function, $J(\omega)$, which describes the frequency populations of a spin system. However $J(\omega)$ is not directly accessible by any experimental method[79]. Therefore, one needs to create a model from which $J(\omega)$ can be estimated in order to predict how these proteins may act in solution. One such model assumes that when the N-H bond motion is seen relative to a fixed reference

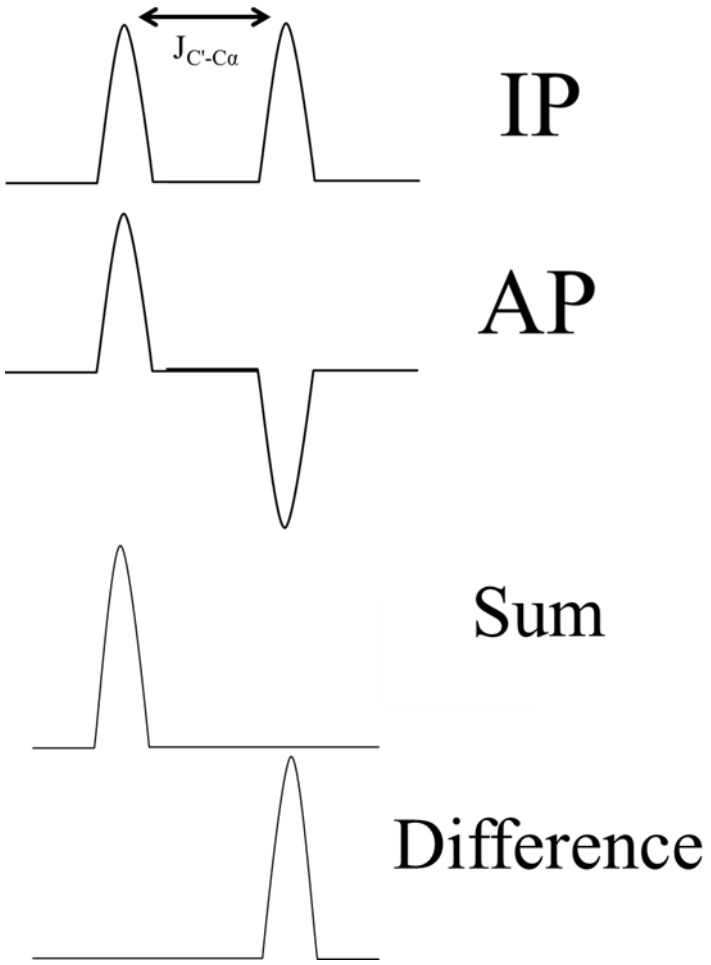

Figure 1.2: A 1D representation of the IPAP virtual decoupling scheme. The 1D in-phase spectrum (IP) and the anti-phase 1D spectrum (AP) are acquired experimentally. Post-acquisition linear sum and difference spectra are produced that are offset by $1/2J_{C'-C\alpha}$, which when removed from each spectra combine to form a single in-phase peak.

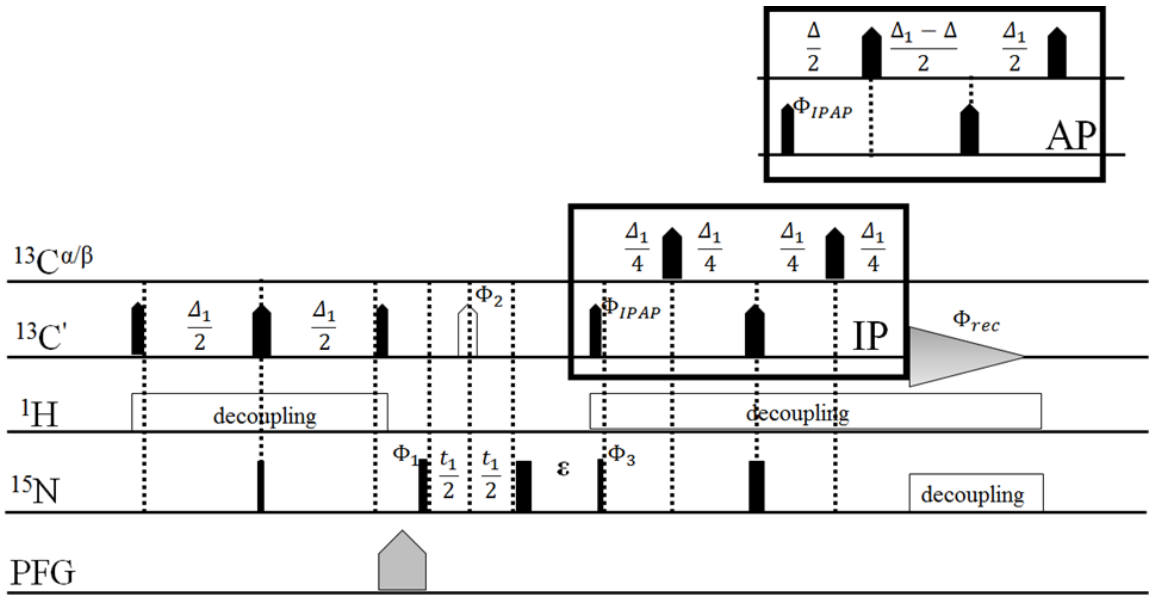

Figure 1.3: Above is the pulse scheme for the CON-IPAP experiment including the variants to obtain in-phase (IP) and anti-phase (AP) components. Pulses that are shown to be narrow of wide depict π/2 and π pulses respectively on the nucleus indicated by the line the pulse is on. The pulse with φ_2 phase is an adiabatic inversion pulse on C' and C^{α}. Pulsed field gradients are located on the line named "PFG". The delays are: Δ = 9ms ($1/2J_{C'-C\alpha}$), Δ_1 = 25ms, $\varepsilon = t_1(0)$. The phase cycle is: φ_1 = x, -x; φ_2 = x,x,-x,-x; φ_3 = x, x, x, x, -x, -x, -x, -x; $\varphi_{IPAP}(IP)$ = x; $\varphi_{IPAP}(AP)$ = -y; φ_{rec} = x, -x, x, -x, -x, x, -x, x; all other pulses have x phase.

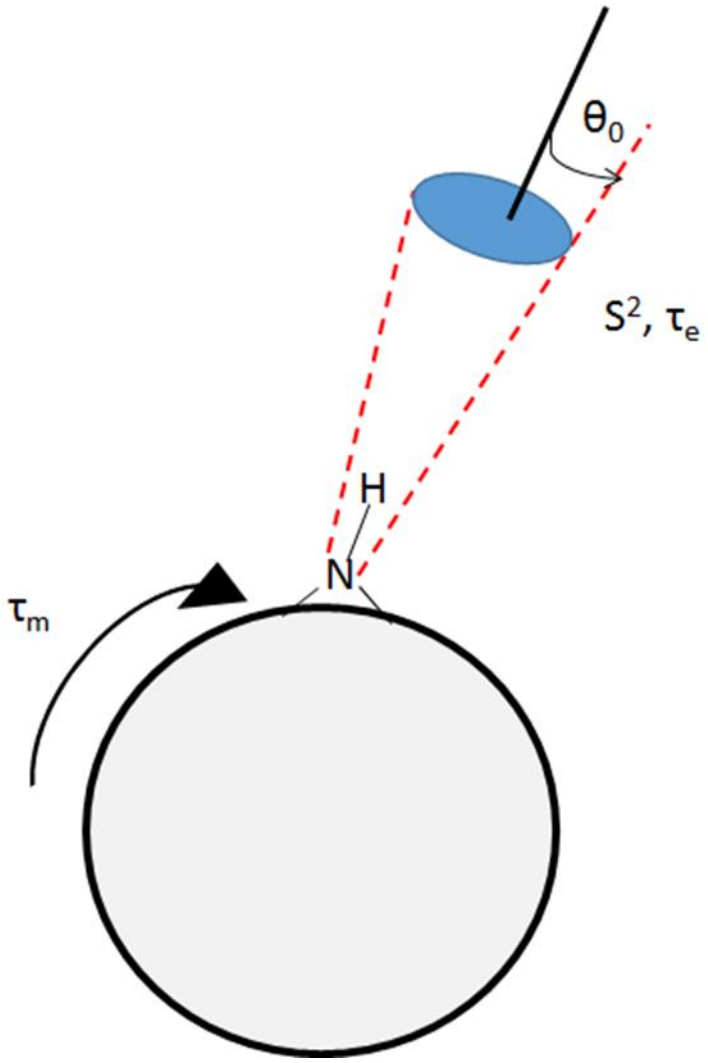

Figure 1.4: A cartoon representation of how a single NH bond vector is viewed in Model-Free. The protein as a whole is seen as a sphere tumbling in solution at a rate of τ_m. The NH bond vector is then rigidly attached to the sphere rotating at a rate of τ_e, with an amplitude of S^2about a principle axis at some angle θ_0.

frame there are three pieces of information that can be obtained: the rate of overall motion of the protein (τ_m), the rate of the internal motion of the NH bond (τ_e), and the amplitude of that N-H bond motion (S^2)[80].

In order to evaluate the effectiveness of the model, experimental parameters that directly depend on the spectral density function are measured. These parameters are the longitudinal relaxation time (T_1), the transverse relaxation time (T_2), and the heteronuclear NOE[76,79].

However, these three observables do not provide enough information to determine the spectral density function on their own. An iterative approach must be taken where the values of the general motional parameters are assumed and J(ω) is calculated[71,77]. From that assumed spectral density function T_1, T_2, and NOE values are determined and compared to the values that were obtained experimentally. If these values do not match then new motion parameters are assumed and the process is repeated until likely values of τ_m, τ_e, and S^2 are found. Qualitatively one can determine S^2 from the T_1 and T_2 relaxation times and τ_e from the NOE[77]. However, these traditional experiments that monitor the motions of N-H bonds through the ^1H are not adequate when studying intrinsically disordered proteins, therefore, making the development of ^{13}C direct detect methods necessary.

Isothermal Titration Calorimetry

Molecular recognition interactions that are involved in protein binding include the intermolecular interactions of the ligand binding to a macromolecule and also intramolecular interactions that can range from subtle structural adjustments to refolding of entire domains[81,82]. Using methods like crystallography or NMR to gain structural insights into proteins and their complexes we can answer how these interactions come together; however, to answer why these interactions occur we need to rationalize these structures in terms of their energetics[83,84]. In order to define the energetics of a reaction additional experimental results are necessary. For the case of IDPs, the energy of binding involves a complex mixture of favorable and unfavorable interactions, arising from the formation of the complex and the induced folding upon binding[85,86]. The change of free energy of the interaction (ΔG) can be defined by the following equation:

$$\Delta G(T) = \Delta H(T_R) + \int_{T_R}^{T} \Delta C_P dT - T\Delta S(T_R) - T\int_{T_R}^{T} \Delta C_P dlnT \quad (1.1)$$

where ΔH is the change in enthalpy ΔS is the change in entropy, ΔC_P is the change in heat capacity and T_R is a reference temperature[81,87]. Therefore to measure the thermodynamics of a

binding interaction requires the ΔG, ΔH, and ΔS at particular temperature and the ΔC_P to predict their change with temperature. Isothermal titration calorimetry (ITC) is a method that monitors small changes in heat as a means of measuring the thermodynamics of a protein-ligand, protein-protein, and protein-nucleic acid interactions.

ITC is a powerful method for determining the thermodynamic parameters that describe a binding interaction which has some considerable advantages over other methods that could be used to extract binding information. One advantage is that ITC measurements allow for the study of proteins while in solution over a range of concentrations without any tags (e.g., fluorophores), or other modifications to the proteins to be studied unless desired[84,88]. Also, unlike other methods for monitoring protein interactions, such as immunoprecipitations, ITC measurements at each injection, due to rapid mixing, reach equilibrium[88]. Instrumentally, the calorimeter is constructed with two identical cells that are housed in an adiabatic jacket. The jacket is cooled so that heat is required to keep the cells at the desired temperature for the experiment. The two cells, where one contains the sample to be studied and the other a reference; typically water, are kept a thermal equilibrium with each other[87,89], shown in figure 1.5A. The purpose of the experiment is to determine the binding characteristics of the system being studied. This is accomplished by measuring the heat per unit time that must be adjusted in order to maintain a fixed ΔT between the two cells. Once the titrant is added to the cell as a series of fixed injection volumes, approximately 5-25 μL, there is a measured enthalpic change[87,90], an example experiment is shown in figure 1.5B. If the interaction between the two molecules is endothermic then there is an observed increase in the amount of heat required to maintain ΔT and if the interaction is exothermic the opposite is true. During each injection there is a mixing of the species by stirring to ensure a rapid return to thermal equilibrium.

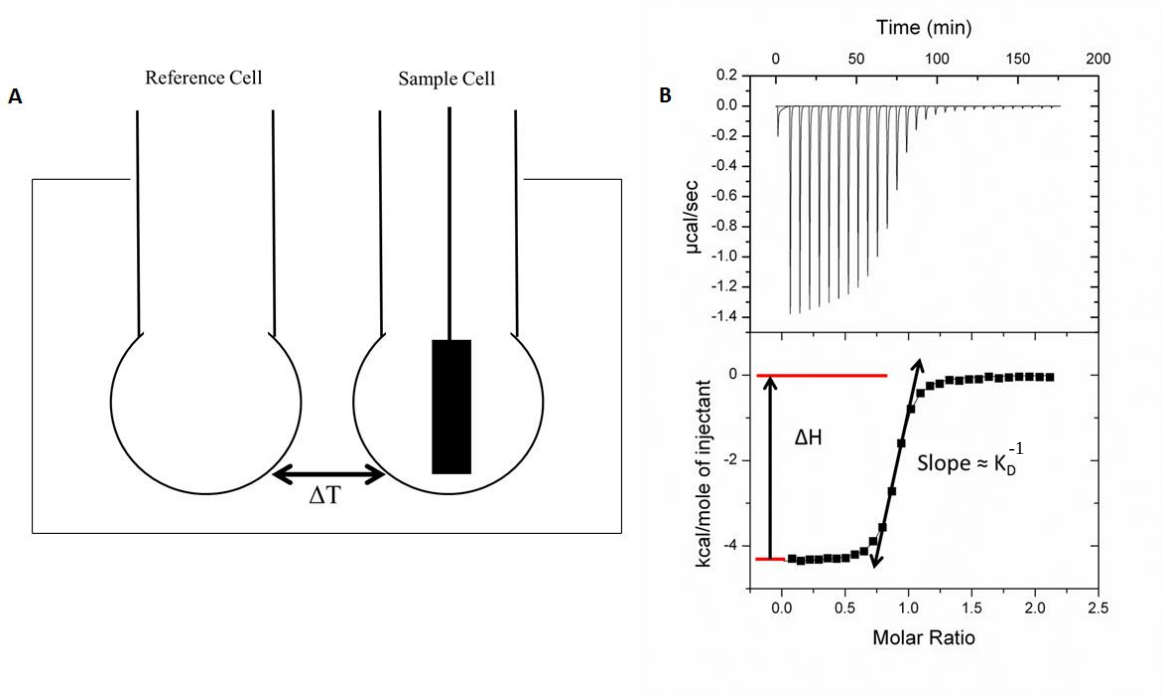

Figure 1.5: (A) A cartoon representation of the ITC instrumentation where there is a reference cell and a sample cell. During the experiment the injector and is placed into the sample cell were it spins to mix the solution at each injection. The difference in temperature between the reference cell and sample cell is kept constant over the course of the experiment, which is what generates the changes in heat at each injection. B(top) shows the raw experimental data obtained from an ITC experiment depicting the change in power required to maintain ΔT between the reference and sample cell. B(bottom) shows the resulting isotherm after the raw data has been integrated. Note that ΔH is determined directly from the change in molar heat observed over the course of the experiment and K_D is calculated from the slope of the transition.

Prior to starting an ITC experiment it is important to calculate what is known as the "c-value"[88,90]. The c-value is described by the equation below:

$$c = \frac{n[M]_t}{K_D} \quad (1.2)$$

Where K_D is the dissociation constant $[M]_t$ is the total macromolecule concentration and *n* is the binding stoichiometry for the interaction. The point of the c-value is that it provides the macromolecule concentration that creates an ideal line shape for a given binding reaction. This is integral for setting up a useful ITC experiment. If $[M]_t$ is too large then all of the ligand will immediately bind until saturation is achieved creating a binding curve that is too steep the

accurately resolve the dissociation constant. While a low $[M]_t$ concentration will result in a shallow curve. This makes it difficult to resolve all three parameters, ΔH, K_D, and n[88,90]. If the dissociation constant lands in the extreme at either end of the spectra then the user may have concerns about the amount of heat produced at low macromolecule concentrations or the possibility of aggregation and poor protein stability at higher concentrations[88]. These obstacles can be worked around if the binding constant is temperature dependent by adjusting the temperature such that K_D is in a useful range. However, due to strong enthalpy/entropy compensation for most cases, the K_D is practically temperature independent. ΔH on the other hand may be near zero for some interactions at a given temperature, which will be shown in data presented later[81,88,90]. This can be avoided by adjusting the temperature in either direction so that a larger molar heat can be observed at each injection.

Accurate measurement of the binding enthalpy and free energy are highly dependent on how well one can sample the binding transition from mostly unbound to fully saturated

$$K_D = \frac{(1-X)[L]}{X} \quad (1.3)$$

Since over the course of the titration the system is at equilibrium it can be described by its dissociation constant. The K_D depends on the free ligand concentration and the fraction of sites occupied, X[81,88,89].

$$[L]_t = [L] + nX[M]_t \quad (1.4)$$

The bulk concentration of ligand is distributed between bound and unbound which depends on the $[M]_t$ and the total number of sites on M available for L. The ultimate goal of ITC is to measure the total heat, Q, of the total titration of L into M[81,88].

$$\mathrm{Q} = nX[M]_t * \Delta H * V \quad (1.5)$$

During an ITC experiment $[M]_t$, $[L]_t$, and V_t are known, and the changes in Q as a function of increasing ligand concentration is directly measured, thus the distribution of L between the free and bound states, *n*, θ, and ΔH are calculable from the experiment. This is first accomplished by combining equations 3 and 5 which yield an expression that accounts for the distribution of M and L between the free and bound states as it is related to the K_D[81,88].

$$X^2 - X\left(1 + \frac{[L]_X + K_D}{n[P]_X}\right) + \frac{1}{n[P]_X} = 0 \quad (1.6)$$

Solving this in terms of fraction of bound and then substituting it back into equation 5 constructs an expression for the total amount of heat, Q, contained within a volume, V, for known concentrations $[M]_t$ and $[L]_t$. From this expression for total heat one can solve for the K_D, *n*, and ΔH[81,88].

$$Q = \frac{n[M]_t * \Delta H * V_0}{2}\left(\left(1 + \frac{[L]_t + K_D}{n[M]_t}\right) - \sqrt{\left(1 + \frac{[L]_t + K_D}{n[M]_t}\right)^2 - \frac{4[L]_t}{n[M]_t}}\right) \quad (1.7)$$

In order to fit ITC data one uses nonlinear regression analysis with an iterative approach[89]. The first step in the fitting process requires creating an initial estimate of each of the parameters in the model equation and then generating a theoretical curve based on these values. The theoretical curve is then compared to the experimental titration curve using an error function that calculates the sum of the squared deviations between the experimental and theoretical curves. The theoretical parameters are then adjusted in order to move the theoretical curve closer to the actual data. This process is then repeated until the error function is minimized and any subsequent iterations show insignificant changes in the error, however, if the error function cannot be minimized then it is necessary to use a different model. Also, it is important to note that the final fit parameters may be dependent on the initial estimates.

Chapter Summaries

Chapter 2

All previously published structural data about ctFCP1, acquired by NMR and X-ray crystallography, was with it bound to RAP74 and for the most part only included the ≈17aa binding domain of the over 80aa c-terminal construct. This demonstrated that there was a severe lack of structural information about the disordered interaction domain of FCP1 in the literature. Also, this lack of knowledge stemmed from the use of traditional ^{1}H-detected NMR experiments for structural studies. Utilizing ^{13}C-direct detection, our lab built an ≈90% complete set of chemical shifts for ctFCP1 both in the free and bound states. With the favorable experience using ^{13}C-direct detection, we wanted to explore just how much structure, if any, ctFCP1 had in the free state and what role that may play in RAP74 binding. An initial attempt to identify possible RAP74 interaction sites on ctFCP1 involved using ANCHOR, bioinformatic binding site prediction software. ANCHOR determines probable interaction sites on an IDP by predicting which regions are likely to fold if in the contact with binding protein. This showed that there were two possible binding sites which included the previously reported RAP74 binding region and an approximately 35-residue region nearer to the N-terminus. However, comparing the chemical shifts for the free and bound state only the previously reported RAP74 binding domain showed any significant change in chemical shift indicating it was the only RAP74 interaction site.

Knowing that the far C-terminus of FCP1 is the only region that interacts with RAP74 and based on previous literature results that it adopts an α-helix upon this interaction, determining whether or not this region had any propensity for helicity in the free state was the next logical step. A chemical shift index (CSI) was built using C', C^{α}, and C^{β} chemical shifts, and from their consensus, residues 945-949 have a propensity for preformed structural elements outside of the FCP1:RAP74 complex. Intrigued by this result we wanted to determine if it were possible to

observe any form of secondary structure with an alternative spectroscopic method, such as circular dichroism (CD).

CD provided a means to monitor whether or not there was a significant population within the ensemble structure of FCP1 that had regularly formed helical elements. Two constructs were used for this study: one the full ctFCP1 construct and the other a 17aa construct containing only the RAP74 binding region (944-961). The smaller construct was used in case approximately 60 residues that precede the binding domain, which show no tendency to form any structural elements, has the ability dominate any signal that may come from the α-helical structures that FCP1 may adopt. The resulting CD spectra showed for both constructs mostly random coil character with some helical content evidenced by subtle minima at 220nm. In an attempt to exaggerate this minima a small molecule osmolyte, 2,2,2-trifluoroethaol, (TFE), was used.

Osmolytes are small molecules, some of which naturally occur in the cell, that are known to stabilize protein structure. TFE is one such osmolyte and is historically known to induce helical structure on peptides regardless of their native state. The resulting spectra for both constructs show a more well resolved minima at 220 with increasing TFE concentration. In order to determine the extent of TFE's effect on FCP1, this same titration was applied while monitoring chemical shifts. We calculated the changes in C'-chemical shifts with increasing TFE concentration and the result was quite striking. The increase in α-helicity that was observed in the CD spectra was limited entirely to the RAP74 binding region of FCP1. The change in chemical shift showed a positive trend as more and more TFE was added to the solution for residues 944-961, however, for the rest of the protein the changes in chemical shift were random and nowhere near to the extent of the binging region, even at 40%(v/v) TFE. This provided very clear evidence that the disorder to order transition upon RAP74 binding is isolated to those residues that make up the binding interface.

While we were confident that FCP1 was disordered in the apo-state prior to binding RAP74 under aqueous conditions this does not necessarily represent cellular conditions. Macromolecular crowding agents have been used to simulate this environment. We used a high molecular weight PEG solution, due to its likelihood to not interact with FCP1, and 70mg/mL HeLa cell extract. In both NMR studies, while changes in chemical shift were observed, the changes were uniform throughout the entire protein and not nearly to the degree seen for RAP74 binding or while in the presence of TFE. Thus, we came to the conclusion that based on the NMR results that FCP1 does at least transiently acquire structural elements but mostly remains disordered even in the cell.

Chapter 3

Since IDPs in particular are highly dynamic in solution, accurately describing their motions and how they change upon binding are critical to understanding their ensemble state and, NMR spin relaxation measurements are an attractive biophysical technique to characterize these motions. Typical spin relaxation measurements use the N-H bond vector to report on backbone dynamics via ^{1}H detection. As was previously discovered, this is not convenient to study IDPs. The poor spectral dispersion in the proton dimension associated with IDPs has limited the number of IDPs that have been subjected to rigorous NMR spin relaxation studies. Having ourselves used ^{13}C-direct detection to assign the chemical shifts for FCP1 and seen its advantages over tradition ^{1}H we decided to attempt to apply ^{13}C spin relaxation.

The development of the CON(T1)-IPAP and CON(T2)-IPAP experiments were relatively straight forward starting from the conventional ^{13}C,^{15}N-CON-IPAP. These $^{1}H^{N}$-start experiments begin with a refocused INEPT transfer starting on the $^{1}H^{N}$ to generate in-phase ^{15}N magnetization. The INEPT transfer leads directly into the ^{15}N T_1 and T_2 measurement periods constructed just as they would be for spin relaxation experiments that utilize a ^{1}H,^{15}N-HSQC detection scheme. As these experiments require a ^{15}N,^{13}C-labelled protein there is ^{15}N-^{13}C scalar coupling evolves

during the T_2 relaxation period complicating the relaxation measurement. The scalar coupling is refocused in the CON(T2)-IPAP by including alternating $^{13}C^{\alpha}$, ^{13}C'selective 180° inversion pulses. After the relaxation period the ^{15}N chemical shift evolves under semi-constant time creating anti-phase ^{15}N-^{13}C coherence. The ^{13}C' chemical shift is evolves during the detection period and the IPAP method removes the one bond $^{13}C^{\alpha}$-^{13}C' scalar coupling. The resultant relaxation data was then compared to the same data which comes from using the typical spin relaxation experiments built off of the ^{1}H,^{15}N-HSQC.

Initially ubiquitin was used as a reference sample for our spin relaxation measurements because of its history in the literature as a standard and previously reported N-H T_1 and T_2 relaxation rates. Our first priority was to replicate the previously reported T1 and T2 relaxation rates using traditional ^{1}H,^{15}N-HSQC based relaxation experiments. However, the experimental relaxation rates did not match the previously reported values. A plot of T_2 rates per residue showed similar features to previous results but the overall baseline rate was not the same. One possible explanation was that the ubiquitin concentration was too high and aggregation was affecting the overall relaxation rate. To test if this was the case a serial dilution was carried out measuring the T2 relaxation at each dilution. From this set of experiments we saw the baseline elevate at each dilution until it matched the previously reported rate. Ubiquitin's tendency to aggregate caused us to use an alternate reference protein, GB3.

GB3 was used as a new reference to develop the novel spin relaxation experiments because its internal dynamics have been previously determined over multiple field strengths using the traditional ^{1}H-detection methods. First we had to determine if our experiments yielded fittable results. As will be seen later, the changes in peak intensity as a function of T_1 or T_2 delay length can be fit with a single exponential without any oscillation. The relaxation rates were compared for those acquired from an non-^{13}C-labelled GB3 sample using the traditional ^{1}H detected experiments and showed good agreement with each other without any significant deviations.

After successful implementation of these experiments on GB3 they were then used to study FCP1. In initial experiments, the resulting data analysis went as expected showing a well fit exponential decay of peak intensity with increasing relaxation delays. However, the binding region showed an almost linear trend for T_2. This simply meant that the delay period was not long enough for the signals of these residues to fully relax. So, the T_2 experiments were rerun with a set of longer delay periods to be added to the final data set. These additional delay periods resulted in changes in intensity that could be fit by a single exponential decay for these residues in the binding region. Upon plotting the relaxation rates as a function of residue, most of the protein was rather featureless, behaving much like a random coil, but for those residues in the binding region there was a significant change in relaxation rates from baseline. Repeating these relaxation experiments under different solvent conditions similar to those tested in chapter 2, 10% TFE and 30% Dextran, yielded interesting results. We previously saw that 10% TFE induces a significant increase in the amount of secondary structure sampled by FCP1 in the unbound state, while the rest of the protein remains widely unaffected. The relaxation rates for those residues outside the binding region remain relatively unaffected, as would be expected, but those in the binding region also remained unaffected as well. This observation is rather unexpected as the acquisition of more secondary structure in that region would cause the backbone to become more rigid, signifying an increase in the order parameter S^2, which would have an effect on both T_1 and T_2 relaxation rate; which is not observed.

In order to gain a complete understanding of the complex the local dynamics of RAP74 were also studied. Since RAP74 is a globular protein with regular structural elements traditional ^{1}H-detected methods were used to study the protein which allowed us to calculate S^2 and gain insight into the flexibility of the protein backbone. A comparison of the order parameter S^2 for both apo and holo-RAP74 showed that the majority of the backbone does not experience a change in dynamics upon interaction with FCP1. Helix 2.5 was the one region that showed notable

change by becoming more ordered upon FCP1 binding. Residues in this region of RAP74 have previously been proposed to mediate interactions with CK2 phosphorylated FCP1, suggesting that phosphorylation acts to adjust binding affinity of pre-existing interactions. Measurements made in 30% Dextran yielded similar results, with the exception of a significant change in baseline relaxation rates due to the difference in solvent viscosity from aqueous.

Combining our relaxation data of FCP1 and RAP74 with previously published molecular dynamics simulations we were able to observe a possible functional role for disorder in the FCP1/RAP74 complex. Creating a pictorial view of the FCP1/RAP74 complex where those regions of FCP1 and RAP74 that become more ordered upon complex, we were able to suggest possible contact between helix 2.5 on RAP74 and several residues on FCP1 that extended beyond the canonical helix but still gained an amount of order. As residues S942 and S944, which have been previously shown in work outside of our lab to increase RAP74 binding affinity via proposed phosphorylation of CK2, are in this buffer region, phosphorylation could restrict the conformations available to FCP1 while in the complex thus strengthening binding affinity.

Chapter 4

ITC was used to fully characterize the thermodynamic parameters that describe this folding upon binding interaction. ITC provides information on the strength of bimolecular interactions, their specificity, and what intermolecular and intra molecular interactions are present before binding and how they are changed by the event by yielding ΔG, ΔH, ΔS all within a single experiment. The process of acquiring the data presented provided me the opportunity of really learning an instrument from the inside out and of dealing with the multiple challenges that are unique to that particular instrument. Initial experiments were run on an auto-ITC using the short peptide construct of FCP1 that contained only the binding region. After running the experiments over a range of temperatures an endothermic to exothermic transition was observed, where the transition occurred at approximately 20°C. Further analysis revealed that the free

energy of the reaction remained unchanged within error over the entire temperature range, with the value at 20°C being the exception due to its inability to be accurately fit. This thermodynamic profile is a perfect example of the hydrophobic effect dominating the overall binding reaction, which is further evidenced by the large, negative ΔC_P associated with the interaction.

Once initial trials were complete with the shorter construct experiments were repeated for the full ctFCP1 construct. The ctFCP1 construct showed the same endothermic to exothermic transition centered on approximately 20°C, and the binding constants showed no temperature dependence. Again, the binding interaction is described by a large, negative ΔC_P indicating that the energetics of complex formation is dominated by the hydrophobic effect. It is not clear from the data to what extent the removal of water from the binding interface or the folding of ctFCP1 plays a role in these observed enthalpies. In order to further investigate this small molecule osmolytes were used to determine what effect preformed helical structure has on the binding interaction between RAP74 and FCP1.

Two osmolytes were used over the course of this study, TFE and trimethylamine N-oxide (TMAO), each with their own unique challenges. TFE has historically been used in the literature to fold short peptides into helical conformations as well as to fold previously unfolded proteins regardless, of their native state. In the work described in the previous chapters, we used TFE as a means to study the folded state of FCP1 in the absence of RAP74 because of its highly localized effect at the binding region. Incorporating TFE into these ITC experiments proved to be difficult because both FCP1 and RAP74 had to be dialyzed into 10% TFE buffer for extended periods of time which would cause at least some protein precipitation prior to the actual experiment making multiple samples difficult and protein precipitation during the ITC experiment common. Instrument cleanliness played a critical role in getting reproducible ITC results. Once the data set was complete, the binding isotherms over the same temperature range as the aqueous state were strikingly different. What was immediately evident was that the endothermic to exothermic

transition from the previous titrations was no longer present. Instead the ΔH remains relatively constant over all temperatures studied and change in capacity was much smaller which suggests that folding is responsible for the thermodynamic characteristics previously observed in the aqueous state. Even with these differences, the binding constants still remain temperature independent making the burial of protein surface the likely dominant heat during the observed interaction.

TMAO was chosen as an alternative osmolyte to determine whether the observed effects of TFE addition were due to changes in FCP1's conformational state or due to solvent effects. Using NMR, we determined that TMAO does not have the same effect on FCP1's conformational state, showing that FCP1 remains mostly disordered under 1M TMAO conditions. This provided the perfect opportunity to determine to what degree the observed changes were due to preformed structure while in the presence of TFE. While in the presence of TMAO the enthalpies over the observed temperature range grew increasingly negative with increasing temperature, just as in previous studies under aqueous conditions, however, the endothermic to exothermic transition was no longer present. To determine if this was indeed the case another experiment was run at a lower temperature outside the previous range to see if the enthalpy would turn endothermic; which is what was observed. TMAO, in this case, kept FCP1 in a mostly disordered state and preserved the endothermic to exothermic transition, all be it shifted to a colder temperature.

To complete the thermodynamic analysis of the various solvent conditions the change in free energy and entropy were calculated. Under aqueous and 1M TMAO conditions ΔG was constant across all temperatures and ΔS was always favorable. This was an important result due to the literature bias towards the naïve assumption that IDP binding is typically weak because of the restriction of conformational freedom associated with forming a complex and the possible acquisition of structure if there is a folding upon binding transition. Our results show that not only is there no large entropic penalty, but, in the case of FCP1, binding is entropically favorable. This

implies that under dilute aqueous conditions the general disorder found in apo –FCP1 plays a role in binding RAP74. Interpretation of the ITC experiments run in the presence of 10% TFE was more difficult on the other hand. It was clearly evident that the observed heats of the binding interaction were being dominated by solvent effects due to TFE release from the binding interface.

Chapter 5

After obtaining a significant amount of experimental data that describes the FCP1/RAP74 binding interaction, it was a logical step to attempt to generate an ensemble model of FCP1 using the obtained data as constraints. Before we built a theoretical model we decided to measure small angle X-ray scattering (SAXS) to generate a 3D model of the overall shape of the ensemble in solution. Accurate SAXS data requires that the sample is monodisperse in solution. A concentration of 30 mg/mL was initially decided to be ideal because it would yield a strong scattering signal and dynamic light scattering determined it to be monodisperse. Analysis of the scattering data for this concentration yielded a molecular weight that would be associated with a trimer, which was further demonstrated in the constructed envelop structure. A serial dilution was then carried out to determine the highest concentration yeilding a monomeric state for FCP1 under the multiple solvent conditions that would be tested. Ultimately, it was decided that 5 mg/mL would be used as it yielded a monomer while 20 and 10 mg/mL yielded a dimer. Once it was clear that we had a monomeric form of FCP1 we also measured SAXS for FCP1 in the presence of both 10% TFE and 4M urea. Resulting envelopes revealed that TFE had a strong ordering effect in that we reproduced the dimer present in the higher concentrations. Given the already low concentration that was able to provide reliable measurement of the monomer state under aqueous conditions, we were unable to remeasure the scattering at lower concentrations while in the presence of 10% TFE. Urea on the other hand had little effect on the overall shape on the ensemble envelope.

SAXS may provide a good picture of how the ensemble behaves as a whole in solution, but it does not provide any information about individual conformers that compose that ensemble in order to generate the experimental constraints that we observe. In an attempt to generate an ensemble structure that takes these individual conformers into account we used the ensemble optimization method (EOM). EOM builds an ensemble structure of IDPs using SAXS as the only experimental constraint. Since SAXS was only collected on the $FCP1^{932-961}$ construct we could not compare directly with any of our other experimental data collected on the $FCP1^{878-961}$ construct, however, molecular dynamics (MD) simulations have been carried out previously in our lab on the same construct. In order to assess the generated ensemble the Rg and asphericity distribution for each conformer and the contact map for the overall ensemble were compared to the simulated data. The overall size and shape of the EOM ensemble is similar to the MD ensemble, however, the average Rg for both ensembles was more compact than the experimental Rg. The compactness associated with EOM ensemble may be because of the construct chosen for SAXS was mostly composed of the RAP74 binding region. Given this regions proclivity to fold upon interaction could have created a selection bias for more compact structures in EOM.The contact maps for the two ensembles both show mostly α-helical secondary structural elements present in the conformers that compose the ensembles, which is consistent with our previous data for that region.

Ensemble models were made of the full ctFCP1 construct using two separate modeling methods: flexible-meccano and ENSEMBLE. The ensemble generated using flexible-meccano use a random coil model based on the primary sequences to generate a set of conformers. In addition to the random coil pool, any presumed structural propensities that an IDP may adopt can be set by the user to further refine the ensemble structure. Previous work in other labs has demonstrated that a random coil model is sufficient to generate an ensemble model that reproduces experimental data for some IDPs. Ensembles were generated using helix propensities

of 0, 50, and 100% for residues 944-961. After ensemble generation, Rg, asphericity, and contact maps were calculated for each ensemble. The Rg was consistent with the experimental value of 24Å , as obtained by SAXS. Plots of asphericity vs Rg showed that regardless of the amount of structural propensity assigned to the RAP74 binding region of FCP1the overall size and shape of the ensemble remained unaffected. While at first glance, it may appear that the conformers that compose these ensembles are essentially the same, the contact maps for each ensemble shows different amounts of helical structure present for each given propensity. Given this observation it was concluded that due to the region of FCP1 that gains structure being so small in comparison to the region that remains disordered it has little to no effect on the overall ensemble structure.

ENSEMBLE differs from flexible-meccano by utilizing multiple experimental constraints to further refine an ensemble structure. Here we have used chemical shift information, as well as R_2 relaxation rates. Residual dipolar couplings (RDCs) were used to generate final ensemble structures. In order to determine the minimum amount of experimental data required to generate an accurate ensemble structure for ctFCP1, experimental parameters were added one at a time such that ensembles were refined using no experimental data, chemical shifts, chemical shifts and R_2 relaxation rates, and chemical shifts; R_2 relaxation rates; and RDCs. Again, the Rg, asphericity and contact maps were calculated for each ensemble. Each case performs equally well at reproducing the experimental Rg, but unlike the flexible-meccano ensembles there are small changes in the asphericity vs Rg plots, indicating that there is some structure refinement upon the addition of experimental constraints. However, each ensemble does fall within the conformational space occupied explored by the random coil model. This, yet again, suggests that a random coil model is sufficient in defining the ensemble structure of FCP1. The chemical shifts for those ENSEMBLE models that used experimental constraints were simulated and shown to reproduce experimental chemical shifts all equally well. The same procedure was then applied to the

flexible-meccano ensembles where only the ensemble with 50% helix propensity was shown report back experimental values as well as ensemble.

From the data provided, some conclusions can be drawn in regards to FCP1's ensemble structure and which method is best used to determine it. The disordered N-terminus of FCP1 is large compared to the RAP74 binding region that is located at the far C-terminus and remains disordered regardless of whether FCP1 is in complex with RAP74 or free in solution. This larger disordered region most likely dominates the overall ensemble structure based on these properties, which is consistent with a random coil model generating ensembles capable reproducing experimental results. However, these conclusions are only based on the data at hand. Future experimentation should be done to acquire PRE (paramagnetic relaxation enhancement) and intramolecular NOE values which may have a large effect on the ensemble refinement process of ENSEMBLE. Also, the perspective of these ensembles is limited to the unbound state. In order to understand if changes in ensemble structure play a role in the FCP1/RAP74 interaction ensembles need to be generated for FCP1 while in complex.

References

1. Dunker, A. K. *et al.* The unfoldomics decade: an update on intrinsically disordered proteins. *BMC Genomics* **9,** S1 (2008).

2. Sambi, I., Gatti-Lafranconi, P., Longhi, S. & Lotti, M. How disorder influences order and vice versa - mutual effects in fusion proteins containing an intrinsically disordered and a globular protein. *FEBS J.* **277,** 4438–4451 (2010).

3. Tompa, P. Intrinsically unstructured proteins. *Trends Biochem. Sci.* **27,** 527–533 (2002).

4. Dyson, H. J. & Wright, P. E. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* **6,** 197–208 (2005).

5. Gsponer, J. & Madan Babu, M. The rules of disorder or why disorder rules. *Prog. Biophys. Mol. Biol.* **99,** 94–103 (2009).

6. Tompa, P. *et al.* Close encounters of the third kind: disordered domains and the interactions of proteins. *BioEssays* **31,** 328–335 (2009).

7. Mao, A. H., Crick, S. L., Vitalis, A., Chicoine, C. L. & Pappu, R. V. Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 8183–8188 (2010).

8. Liu, J., Faeder, J. R. & Camacho, C. J. Toward a quantitative theory of intrinsically disordered proteins and their function. *Proc. Natl. Acad. Sci.* (2009). doi:10.1073/pnas.0907710106

9. Dill, K. A. Dominant forces in protein folding. *Biochemistry (Mosc.)* **29,** 7133–7155 (1990).

10. Southall, N. T., Dill, K. A. & Haymet, A. D. J. A View of the Hydrophobic Effect. *J. Phys. Chem. B* **106,** 521–533 (2002).

11. Kohn, J. E. *et al.* Random-coil behavior and the dimensions of chemically unfolded proteins. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 12491–12496 (2004).

12. Marsh, J. A. & Forman-Kay, J. D. Ensemble modeling of protein disordered states: Experimental restraint contributions and validation. *Proteins* (2011). doi:10.1002/prot.23220

13. Fisher, C. K. & Stultz, C. M. Constructing ensembles for intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **21,** 426–431 (2011).

14. Mittag, T., Kay, L. E. & Forman-Kay, J. D. Protein dynamics and conformational disorder in molecular recognition. *J. Mol. Recognit. JMR* **23,** 105–116 (2010).

15. Oldfield, C. J. *et al.* Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners. *BMC Genomics* **9 Suppl 1,** S1 (2008).

16. Dagkessamanskaia, A. *et al.* Functional dissection of an intrinsically disordered protein: understanding the roles of different domains of Knr4 protein in protein-protein interactions. *Protein Sci. Publ. Protein Soc.* **19,** 1376–1385 (2010).

17. Brocca, S. *et al.* Order propensity of an intrinsically disordered protein, the cyclin-dependent-kinase inhibitor Sic1. *Proteins* **76,** 731–746 (2009).

18. Higo, J., Nishimura, Y. & Nakamura, H. A free-energy landscape for coupled folding and binding of an intrinsically disordered protein in explicit solvent from detailed all-atom computations. *J. Am. Chem. Soc.* **133,** 10448–10458 (2011).

19. Zor, T., Mayr, B. M., Dyson, H. J., Montminy, M. R. & Wright, P. E. Roles of phosphorylation and helix propensity in the binding of the KIX domain of CREB-binding protein by constitutive (c-Myb) and inducible (CREB) activators. *J. Biol. Chem.* **277,** 42241–42248 (2002).

20. Receveur-Bréchot, V., Bourhis, J.-M., Uversky, V. N., Canard, B. & Longhi, S. Assessing protein disorder and induced folding. *Proteins* **62,** 24–45 (2006).

21. Hsu, W.-L. *et al.* Exploring the binding diversity of intrinsically disordered proteins involved in one-to-many binding. *Protein Sci. Publ. Protein Soc.* **22,** 258–273 (2013).

22. Xiong, K. *et al.* Direct observations of conformational distributions of intrinsically disordered p53 peptides using UV Raman and explicit solvent simulations. *J. Phys. Chem. A* **115,** 9520–9527 (2011).

23. Trizac, E., Levy, Y. & Wolynes, P. G. Capillarity theory for the fly-casting mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 2746–2750 (2010).

24. Levy, Y., Onuchic, J. N. & Wolynes, P. G. Fly-casting in protein-DNA binding: frustration between protein folding and electrostatics facilitates target recognition. *J. Am. Chem. Soc.* **129,** 738–739 (2007).

25. Hammes, G. G., Chang, Y.-C. & Oas, T. G. Conformational selection or induced fit: a flux description of reaction mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 13737–13741 (2009).

26. Mohan, A. *et al.* Analysis of molecular recognition features (MoRFs). *J. Mol. Biol.* **362,** 1043–1059 (2006).

27. Wang, J. *et al.* Multi-Scaled Explorations of Binding-Induced Folding of Intrinsically Disordered Protein Inhibitor IA3 to its Target Enzyme. *PLoS Comput. Biol.* **7,** (2011).

28. Pancsa, R. & Fuxreiter, M. Interactions via intrinsically disordered regions: what kind of motifs? *IUBMB Life* **64,** 513–520 (2012).

29. Knott, M. & Best, R. B. A Preformed Binding Interface in the Unbound Ensemble of an Intrinsically Disordered Protein: Evidence from Molecular Simulations. *PLoS Comput. Biol.* **8,** (2012).

30. Csermely, P., Palotai, R. & Nussinov, R. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends Biochem. Sci.* **35,** 539–546 (2010).

31. Bucher, D., Grant, B. J. & McCammon, J. A. Induced Fit or Conformational Selection? The Role of the Semi-closed State in the Maltose Binding Protein. *Biochemistry (Mosc.)* **50,** 10530–10539 (2011).

32. Jerebtsova, M. *et al.* Mass spectrometry and biochemical analysis of RNA polymerase II: targeting by protein phosphatase-1. *Mol. Cell. Biochem.* **347,** 79–87 (2011).

33. Meinhart, A., Kamenski, T., Hoeppner, S., Baumli, S. & Cramer, P. A structural perspective of CTD function. *Genes Dev.* **19,** 1401–1415 (2005).

34. McCracken, S. *et al.* The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature* **385,** 357–361 (1997).

35. Nikolov, D. B. & Burley, S. K. RNA polymerase II transcription initiation: A structural view. *Proc. Natl. Acad. Sci.* **94,** 15–22 (1997).

36. Kamenski, T., Heilmeier, S., Meinhart, A. & Cramer, P. Structure and mechanism of RNA polymerase II CTD phosphatases. *Mol. Cell* **15,** 399–407 (2004).

37. Orphanides, G., Lagrange, T. & Reinberg, D. The general transcription factors of RNA polymerase II. *Genes Dev.* **10,** 2657–2683 (1996).

38. Tan, S., Aso, T., Conaway, R. C. & Conaway, J. W. Roles for both the RAP30 and RAP74 subunits of transcription factor IIF in transcription initiation and elongation by RNA polymerase II. *J. Biol. Chem.* **269,** 25684–25691 (1994).

39. Gaiser, F., Tan, S. & Richmond, T. J. Novel dimerization fold of RAP30/RAP74 in human TFIIF at 1.7 A resolution. *J. Mol. Biol.* **302,** 1119–1127 (2000).

40. Chang, C., Kostrub, C. F. & Burton, Z. F. RAP30/74 (transcription factor IIF) is required for promoter escape by RNA polymerase II. *J. Biol. Chem.* **268,** 20482–20489 (1993).

41. Fang, S. M. & Burton, Z. F. RNA polymerase II-associated protein (RAP) 74 binds transcription factor (TF) IIB and blocks TFIIB-RAP30 binding. *J. Biol. Chem.* **271,** 11703–11709 (1996).

42. Lei, L., Ren, D. & Burton, Z. F. The RAP74 subunit of human transcription factor IIF has similar roles in initiation and elongation. *Mol. Cell. Biol.* **19,** 8372–8382 (1999).

43. Lei, L., Ren, D., Finkelstein, A. & Burton, Z. F. Functions of the N- and C-terminal domains of human RAP74 in transcriptional initiation, elongation, and recycling of RNA polymerase II. *Mol. Cell. Biol.* **18,** 2130–2142 (1998).

44. Kamada, K., De Angelis, J., Roeder, R. G. & Burley, S. K. Crystal structure of the C-terminal domain of the RAP74 subunit of human transcription factor IIF. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 3115–3120 (2001).

45. Abbott, K. L. *et al.* Interactions of the HIV-1 Tat and RAP74 proteins with the RNA polymerase II CTD phosphatase FCP1. *Biochemistry (Mosc.)* **44,** 2716–2731 (2005).

46. Gajiwala, K. S. & Burley, S. K. Winged helix proteins. *Curr. Opin. Struct. Biol.* **10,** 110–116 (2000).

47. Archambault, J. *et al.* FCP1, the RAP74-interacting subunit of a human protein phosphatase that dephosphorylates the carboxyl-terminal domain of RNA polymerase IIO. *J. Biol. Chem.* **273,** 27593–27601 (1998).

48. Palancade, B. & Bensaude, O. Investigating RNA polymerase II carboxyl-terminal domain (CTD) phosphorylation. *Eur. J. Biochem. FEBS* **270,** 3859–3870 (2003).

49. Bartkowiak, B. & Greenleaf, A. L. Phosphorylation of RNAPII. *Transcription* **2,** 115–119 (2011).

50. Mandal, S. S., Cho, H., Kim, S., Cabane, K. & Reinberg, D. FCP1, a phosphatase specific for the heptapeptide repeat of the largest subunit of RNA polymerase II, stimulates transcription elongation. *Mol. Cell. Biol.* **22,** 7543–7552 (2002).

51. Cagas, P. M. & Corden, J. L. Structural studies of a synthetic peptide derived from the carboxyl-terminal domain of RNA polymerase II. *Proteins* **21,** 149–160 (1995).

52. Ghosh, A., Shuman, S. & Lima, C. D. The structure of Fcp1, an essential RNA polymerase II CTD phosphatase. *Mol. Cell* **32,** 478–490 (2008).

53. Kobor, M. S. *et al.* A Motif Shared by TFIIF and TFIIB Mediates Their Interaction with the RNA Polymerase II Carboxy-Terminal Domain Phosphatase Fcp1p in Saccharomyces cerevisiae. *Mol. Cell. Biol.* **20,** 7438–7449 (2000).

54. Archambault, J. *et al.* An essential component of a C-terminal domain phosphatase that interacts with transcription factor IIF in Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci. U. S. A.* **94,** 14300–14305 (1997).

55. Nguyen, B. D. *et al.* NMR structure of a complex containing the TFIIF subunit RAP74 and the RNA polymerase II carboxyl-terminal domain phosphatase FCP1. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 5688–5693 (2003).

56. Kamada, K., Roeder, R. G. & Burley, S. K. Molecular mechanism of recruitment of TFIIF-associating RNA polymerase C-terminal domain phosphatase (FCP1) by transcription factor IIF. *Proc. Natl. Acad. Sci. U. S. A.* **100,** 2296–2299 (2003).

57. Mittag, T. & Forman-Kay, J. D. Atomic-level characterization of disordered protein ensembles. *Curr. Opin. Struct. Biol.* **17,** 3–14 (2007).

58. Bermel, W., Felli, I. C., Kümmerle, R. & Pierattelli, R. 13C Direct-detection biomolecular NMR. *Concepts Magn. Reson. Part* **32A,** 183–200 (2008).

59. Showalter, S. A. NMR assignment of the intrinsically disordered C-terminal region of Homo sapiens FCP1 in the unbound state. *Biomol. NMR Assignments* **3,** 179–181 (2009).

60. Bermel, W., Bertini, I., Felli, I. C., Piccioli, M. & Pierattelli, R. 13C-detected protonless NMR spectroscopy of proteins in solution. *Prog. Nucl. Magn. Reson. Spectrosc.* **48,** 25–45 (2006).

61. Bermel, W., Bertini, I., Felli, I. C., Peruzzini, R. & Pierattelli, R. Exclusively heteronuclear NMR experiments to obtain structural and dynamic information on proteins. *Chemphyschem Eur. J. Chem. Phys. Phys. Chem.* **11,** 689–695 (2010).

62. Bermel, W. *et al.* H-start for exclusively heteronuclear NMR spectroscopy: the case of intrinsically disordered proteins. *J. Magn. Reson. San Diego Calif 1997* **198,** 275–281 (2009).

63. Bermel, W., Bertini, I., Felli, I. C. & Pierattelli, R. Speeding Up 13C Direct Detection Biomolecular NMR Spectroscopy. *J. Am. Chem. Soc.* **131,** 15339–15345 (2009).

64. Arnesano, F. *et al.* A strategy for the NMR characterization of type II copper(II) proteins: the case of the copper trafficking protein CopC from Pseudomonas Syringae. *J. Am. Chem. Soc.* **125,** 7200–7208 (2003).

65. Dyson, H. J. & Wright, P. E. Unfolded proteins and protein folding studied by NMR. *Chem. Rev.* **104,** 3607–3622 (2004).

66. Wishart, D. S. & Sykes, B. D. The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data. *J. Biomol. NMR* **4,** 171–180 (1994).

67. Bermel, W. *et al.* Complete Assignment of Heteronuclear Protein Resonances by Protonless NMR Spectroscopy. *Angew. Chem. Int. Ed.* **44,** 3089–3092 (2005).

68. O'Hare, B., Benesi, A. J. & Showalter, S. A. Incorporating 1H chemical shift determination into 13C-direct detected spectroscopy of intrinsically disordered proteins in solution. *J. Magn. Reson. San Diego Calif 1997* **200,** 354–358 (2009).

69. Hall, J. B. & Fushman, D. Characterization of the overall and local dynamics of a protein with intermediate rotational anisotropy: Differentiating between conformational exchange and anisotropic diffusion in the B3 domain of protein G. *J. Biomol. NMR* **27,** 261–275 (2003).

70. Showalter, S. A. & Brüschweiler, R. Validation of Molecular Dynamics Simulations of Biomolecules Using NMR Spin Relaxation as Benchmarks: Application to the AMBER99SB Force Field. *J. Chem. Theory Comput.* **3,** 961–975 (2007).

71. Palmer, A. G., 3rd. Probing molecular motion by NMR. *Curr. Opin. Struct. Biol.* **7,** 732–737 (1997).

72. Zeng, L., Fischer, M. W. & Zuiderweg, E. R. Study of protein dynamics in solution by measurement of (13)C (α)- (13)CO NOE and (13)CO longitudinal relaxation. *J. Biomol. NMR* **7,** 157–162 (1996).

73. Csizmok, V., Felli, I. C., Tompa, P., Banci, L. & Bertini, I. Structural and dynamic characterization of intrinsically disordered human securin by NMR spectroscopy. *J. Am. Chem. Soc.* **130,** 16873–16879 (2008).

74. Otten, R., Villali, J., Kern, D. & Mulder, F. A. A. Probing microsecond time scale dynamics in proteins by methyl (1)H Carr-Purcell-Meiboom-Gill relaxation dispersion NMR

measurements. Application to activation of the signaling protein NtrC(r). *J. Am. Chem. Soc.* **132,** 17004–17014 (2010).

75. Palmer, A. G., 3rd & Massi, F. Characterization of the dynamics of biomacromolecules using rotating-frame spin relaxation NMR spectroscopy. *Chem. Rev.* **106,** 1700–1719 (2006).

76. Sapienza, P. J. & Lee, A. L. Using NMR to study fast dynamics in proteins: methods and applications. *Curr. Opin. Pharmacol.* **10,** 723–730 (2010).

77. 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 (2006).

78. Palmer, A. G., 3rd. NMR characterization of the dynamics of biomacromolecules. *Chem. Rev.* **104,** 3623–3640 (2004).

79. Luginbühl, P. & Wüthrich, K. Semi-classical nuclear spin relaxation theory revisited for use with biological macromolecules. *Prog. Nucl. Magn. Reson. Spectrosc.* **40,** 199–247 (2002).

80. Lipari, G. & Szabo, A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.* **104,** 4546–4559 (1982).

81. Jelesarov, I. & Bosshard, H. R. Isothermal titration calorimetry and differential scanning calorimetry as complementary tools to investigate the energetics of biomolecular recognition. *J. Mol. Recognit. JMR* **12,** 3–18 (1999).

82. Padlan, E. A. in *Adv. Protein Chem.* (Frederic M. Richards, D. E. E. and P. S. K.) **Volume 49,** 57–133 (Academic Press, 1996).

83. Demers, J.-P. & Mittermaier, A. Binding Mechanism of an SH3 Domain Studied by NMR and ITC. *J. Am. Chem. Soc.* **131,** 4355–4367 (2009).

84. Doyle. Characterization of binding interactions by isothermal titration calorimetry. *Curr. Opin. Biotechnol.* **8,** 31–35 (1997).

85. Ross, P. D. & Subramanian, S. Thermodynamics of protein association reactions: forces contributing to stability. *Biochemistry (Mosc.)* **20,** 3096–3102 (1981).

86. Spolar, R. S. & Record, M. T., Jr. Coupling of local folding to site-specific binding of proteins to DNA. *Science* **263,** 777–784 (1994).

87. Ladbury, J. E. & Chowdhry, B. Z. Sensing the heat: the application of isothermal titration calorimetry to thermodynamic studies of biomolecular interactions. *Chem. Biol.* **3,** 791–801 (1996).

88. Kathryn Malecek, A. R. Validation of histone-binding partners by peptide pull-downs and isothermal titration calorimetry. *Methods Enzymol.* **512,** 187–220 (2012).

89. Freyer, M. W. & Lewis, E. A. Isothermal titration calorimetry: experimental design, data analysis, and probing macromolecule/ligand binding and kinetic interactions. *Methods Cell Biol.* **84,** 79–113 (2008).

90. Leavitt, S. & Freire, E. Direct measurement of protein binding energetics by isothermal titration calorimetry. *Curr. Opin. Struct. Biol.* **11,** 560–566 (2001).

Chapter 2 The C-terminal Domain of FCP1 Retains Helical Structure in the Unbound State

[Published: Lawrence, C. W., Bonny, A. & Showalter, S. A. The disordered C-terminus of the RNA Polymerase II phosphatase FCP1 is partially helical in the unbound state. *Biochem. Biophys. Res. Commun.* 410, 461–465 (2011).]

The studies done in the presence of crowding agents, 30% dextran and HeLa cell extract, presented below were carried out and processed by Alain Bonny under my direct supervision.

Abstract

Intrinsically disordered proteins (IDPs) are a class of proteins that under native conditions have been shown to lack unique secondary or tertiary structure, but carry out vital functions. After transcription RNA Polymerase II is made ready for use in more rounds of transcription through the winged helix domain of RAP74, a subdomain of the general transcription factor IIF (TFIIF), and the C-terminus of TFIIF-associating RNA Polymerase II C-terminal domain phosphatase (FCP1). The residues that make up the binding interface of FCP1 have been shown to form an α-helix while in complex, but in the literature the protein is overall considered to be disordered. Here, we show that even though FCP1 is intrinsically disordered, those residues that interact with RAP74 show nascent helical structure in the unbound state through CD and recently developed ^{13}C-direct detected NMR. Also, we will show FCP1 retains disorder and nascent helical content in HeLa cell extract, representing behavior in a cellular environment. The observed conformational bias leads to mechanistic insights into the disorder to order transition associated with this interaction.

Introduction

The function of a protein has been attributed to its unique 3D structure for decades. However, intrinsically disordered proteins are a class of proteins that have no unique secondary or tertiary structure and are regularly associated with vital functions in transcription, translation, and cellular signaling. This amount of disorder can be attributed to a sequence bias towards charged hydrophilic amino acids[1–3]. Even though this class of proteins is natively disordered, IDPs are known to acquire order upon interaction with another protein or nucleic acid. The mechanism by which IDPs undergo this disorder-to-transition has been the subject of recent study and assumed critical to their function[4–9].

FCP1 (Transcription Factor IIF associating RNA Polymerase II C-terminal domain phosphatase) is an IDP that contains an approximately 80 amino acid disordered domain at its C-terminus[10]. This region is necessary for its recruitment to RNA Polymerase II (RNAPII) and its full functionality as a regulatory phosphatase during the termination step of transcription. The C-terminal domain (CTD) of RNAPII undergoes a phosphorylation/dephosphorylating cycle as the polymerase goes through messenger RNA transcription[11,12]. In order to promote FCP1 function the highly acidic C-terminal region of FCP1 (ctFCP1) binds with the C-terminal winged helix domain of RAP74 (RNAPII associating protein 74 kDa), a subunit of Transcription Factor IIF[13,14]. The structure of this complex has been determined by both crystallography and NMR spectroscopy, and shows that residues 944-961 of FCP1 adopt a helical conformation[15,16]. The remainder of ctFCP1 is presumed to remain disordered due to a lack of complete chemical shift assignments and electron density. It has also been suggested that all of ctFCP1 is disordered when free in solution[16].

In contrast to globular proteins the spectral properties of IDPs offer only limited NMR structural constraints; ultimately making the chemical shift itself the indicator of ensemble properties for IDPs[17,18]. By looking at chemical shift perturbations with the use of cosolutes such

as osmolytes, chaotropes, and crowding agents one can glean additional structural properties[19–21]. However, application of NMR to IDPs by traditional methods is limited due the unique challenges IDPs present[22,23]. Due to their highly dynamic nature and redundant sequences a typical ^{15}N, ^{1}H-HSQC shows a high degree of overlap and degeneracy. Advances in ^{13}C-direct detection spectroscopy from our lab and others with enhanced resolution and increased dispersion provide a reliable tool for the study of IDPs[24–26]. Here we report using a chemical shift based procedure for structural characterization of ctFCP1 in three states: free in solution, bound to RAP74, and perturbed by cosolutes. These studies show that the RAP74 binding region of ctFCP1 is partially helical free in solution while the rest of the protein still remains strongly disordered, even while bound to RAP74 or when in the presence of a crowded environment. How the transiently helical content found in apo-ctFCP1 may be involved in the folding-upon-binding mechanisms associated with the FCP1/RAP74 complex will be discussed.

Materials and Methods

Protein Generation

Recombinant FCP1 83aa (878-961) and RAP74 91aa (426-517) constructs with codon usage optimized for expression in *Escherichia coli* were cloned into the expression vector pET-47b (Novagen) and transformed into *E. coli* BL21 competent cells. The cells were grown in M9 media containing both $^{15}NH_4Cl$ and ^{13}C-glucose (Cambridge Isotope Laboratories, Inc.) and LB media until an optical density between 0.5 and 0.6 for M9 and 0.8-0.9 for LB at 600 nm was reached. ctFCP1 and RAP74 expression was induced using 0.5M IPTG followed by growth for an additional 3-4 hours and harvesting by sonication. Following centrifugation, the clarified cell lysates were run over a Ni-NTA column for purification of the His-tagged ctFCP1 and RAP74. The proteins were dialyzed in the presence of 3-C protease overnight at 4°C to remove the His-tag, leaving a non-native Gly-Pro-Gly tripeptide at the N-terminus of the protein. The His-tagged 3-C protease was removed by running the sample after dialysis over a Ni-NTA column resulting

in highly pure ctFCP1 and RAP74. An NMR sample was obtained by spin-concentration and buffer exchange in a VIVASPIN 20 centrifugal concentrator. The final buffer conditions for the ^{15}N, ^{13}C labeled sample were 20mM cacodylate, pH 7.0, 100mM NaCl, 0.002% w/v NaN3, and 90% H_2O/ 10% D_2O (v/v). CD samples were made from lyophilized protein to yield a final concentration of 20µM in 20mM cacodylate at a pH of 7.0.

CD* S*pectroscopy

CD was used to observe the effects of TFE on FCP1 peptide and the full C-terminal domain. All CD data were collected on a Jasco J-810 spectropolarimeter. Measurements were taken from 205-250nm every 1nm with a bandwidth of 2nm. TFE concentrations ranged from 0% to 40% in increments of 10%.

NMR Spectroscopy

All experiments were recorded on an 11.7T Bruker AVANCE-3 spectrometer operating at 500.13 MHz ^{1}H frequency equipped with a TCI cryoprobe, allowing high sensitivity acquisition of ^{13}C-direct detected spectra. All spectra were recorded at 298K. ^{15}N,^{13}C-CON-IPAP spectra were collected using the standard version of the pulse sequence found in the Topspin 2.1 library. Following virtual decoupling of the spectra in Topspin, all NMR spectra were processed in NMRpipe[27]. Peak picking and resonance assignment were done in Sparky[28].

Analysis of secondary structure through chemical shifts of backbone ^{13}C atoms was performed using the standard chemical shift indexing method and in-house software. The change of 2D peak position in ^{15}N,^{13}C-CON-IPAP spectra was used as an indication of chemical environment changes in response to RAP74 or cosolute addition and quantified as:

$$\Delta\delta = \sqrt{(N\Delta_{ppm}\alpha)^2 + (CO\Delta_{ppm})^2} \qquad (2.1)$$

where Δ_{ppm} is the change in chemical shift between the two states and α = 0.3 normalizes the ^{15}N and ^{13}C chemical shift ranges. TFE was titrated in increments of 5% (v/v) from 0 to 40%. Urea was titrated in increments of 0.5 M from 0.0 to 3.0M. Measurements of crowding effects were made through the mixture of 30% (w/v) dextran or 70 mg/mL HeLa cell extract with sufficient aqueous buffer to generate a standard 500 µL sample. The RAP74 bound sample was generated by mixing ctFCP1 and RAP74 winged helix domain in a 1:1.2 ratio in standard NMR buffer.

Results and Discussion

Utilizing Anchor to Predict Binding Sites

IDPs are often known to mediate protein-protein interactions, as demonstrated by ctFCP1's interaction with RAP74. Advances in bioinformatics have increased the number of predicted interactions, while experimental evidence is lagging behind. ANCHOR is one such bioinformatic method that predicts the likelihood that a region of an IDP will fold upon protein-protein interaction. In order to estimate this characteristic ANCHOR utilizes a general disorder prediction method, IUPred. Once regions of disorder are found it is determined whether or not they are isolated from other globular regions and that they do not have the ability to form favorable contacts that would cause the protein to fold except when in the presence of another protein[29]. Analysis of ctFCP1 by this software shows two sites that are likely to be involved in forming protein-protein -contacts, one of which is the RAP74 binding site and the other approximately residues 895-930, as seen in Fig 1.1A. In prior NMR studies of ctFCP1 bound to RAP74 there were very few assigned residues from the N-terminus to around 920 due to the severe overlap observed in the 2D ^{15}N,^{1}H-HSQC[16]. Therefore, it is possible that there may be at least some transient and previously unidentified interactions in this region with RAP74 as predicted by ANCHOR.

Before we can definitively characterize those residues that comprise the FCP1/RAP74 binding interface on ctFCP1 it is necessary to produce quantitative and complete NMR chemical

shifts for the whole protein. Incomplete chemical shifts assignments of IDPs are common due to low spectral dispersion in ^{1}H-detected NMR. IDPs are particularly challenging to study with traditional ^{1}H-detection methods because of their sequence bias and low sequence complexity. In order to avoid this problem our lab has utilized ^{13}C-direct detection methods in order to study IDPs. Using the carbonyl carbon as our detection nucleus makes these methods especially effective given its high chemical shielding anisotropy causes its isotropic chemical shift to become particularly sensitive to small changes in local structure[24]. Using ^{13}C-direct detected methods comprehensive chemical shift assignments of ctFCP1 have been generated in both the free state and bound to RAP74[30]. The spectral quality in the ^{15}N,^{13}C-CON-IPAP is clearly visible in Fig 1.1C, which is zoomed to show the resonances that shift upon binding RAP74. By comparing the peak positions of ctFCP1 in both the free and bound states it is very apparent that the only contacts that are formed with RAP74 are residues 944-961, Fig 1B, those found in previous NMR and crystal structures.

Building a CSI to Predict Structural Propensity

It has been previously documented that disordered protein segments transiently sample secondary structure associated with their bound states while free in solution[6]. Given that ctFCP1 adopts a helical structure when bound to RAP74 we built a ^{13}C chemical shift index (CSI), Fig 2.2, in order to explore whether or not ctFCP1 is likely to adopt any preformed structure in the apo state[31,32]. The consensus from the CSI shows that residues 945-949 show significant enough secondary shifts to indicate helical structure. The rest of ctFCP1 still shows a strong amount of disorder, however, there are still many residues in the binding region that show a tendency towards alpha helical character. The secondary structural elements associated with residues 945-949 have not been previously observed, most likely due to the enhanced sensitivity provided by the ^{13}C-direct detected methods and the more complete chemical shift assignments.

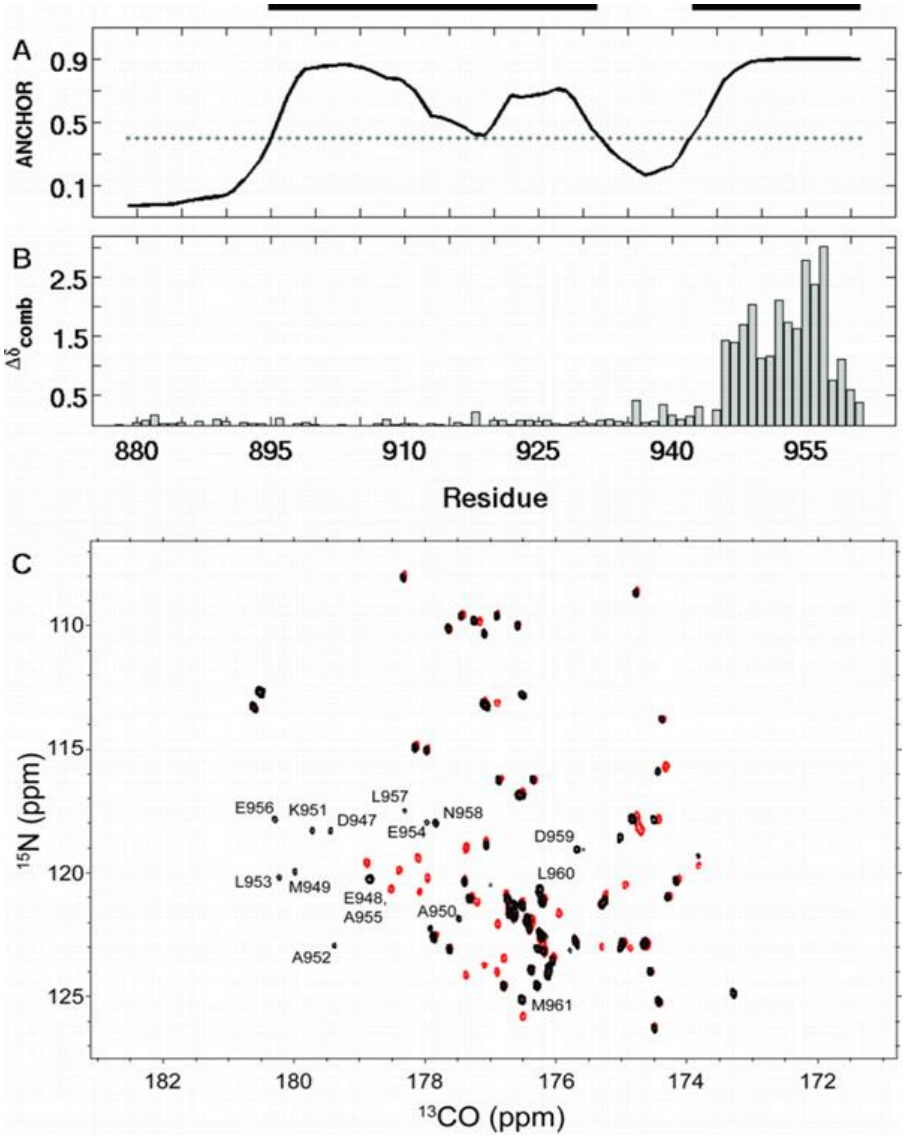

Figure 2.1*: Sites of interaction on ctFCP1 when in complex with RAP74. (A) Analysis of ANCHOR results shows two possible regions for protein-protein interaction on ctFCP1, indicated by horizontal bars above the figure. (B) Of those two predicted regions only the most C-terminal binding site shows significant change in chemical shift and corresponds to the known RAP74 interacting region. (C) An overlay of the* $^{15}N,^{13}C$*-CON spectra of apo-FCP1(red) and holo-FCP1 (black) zoomed into highlight those residues in the RAP74 interacting region (D947-M961 labeled in the bound state).*

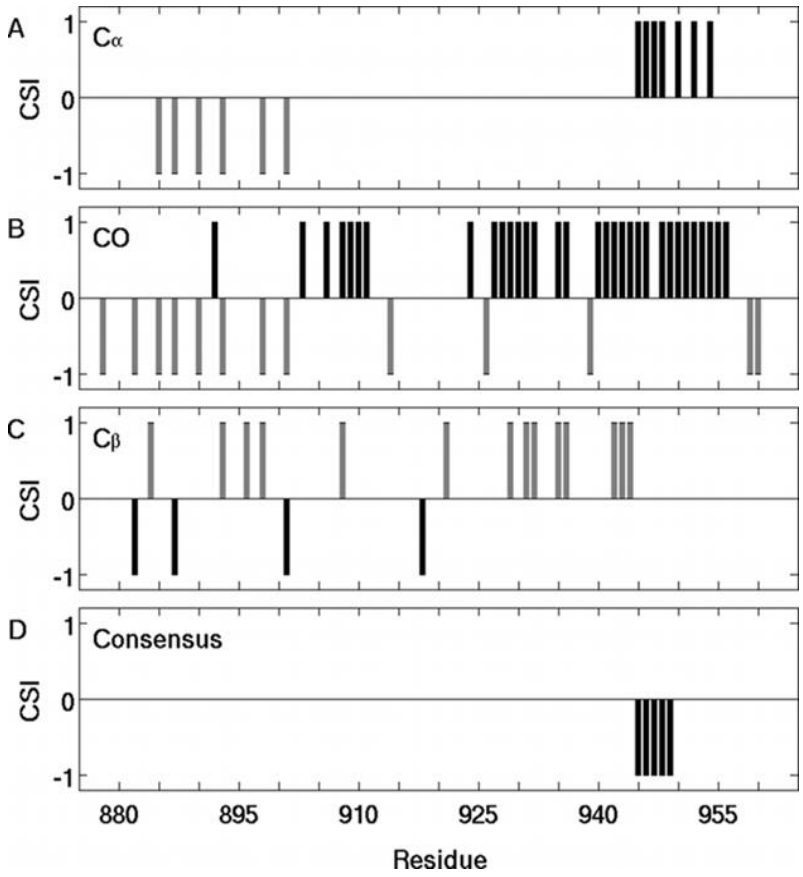

Figure 2.2*: Chemical shift index (CSI) analysis of ctFCP1. Analysis of 13Cα, 13CO, and 13Cβ chemical shifts show a consensus from 945-949 indicating helix nucleation in the RAP74 binding region while in the unbound state. Chemical shifts associated with helix formation are shown in black while those that indicate a more extended or β-strand structures are shown in grey.*

Using Cosolutes to Monitor Structural Changes as seen by CD and NMR

As we have observed a tendency towards helical structure in ctFCP1 while free in solution by ^{13}C-direct detect NMR, we sought to evaluate our findings via a second technique. CD is a spectroscopic technique that provides qualitative information about the secondary structure found within a protein. The technique works by shining plane-polarized light through the sample cell. The polarized light can be thought of being composed of two circular wave components, one turning counter-clockwise and the other clockwise. Based on the structural elements present in the protein one component will be more or less absorbed than the other giving

an “elliptical” character to the light that is expressed in terms of degrees[33]. The spectra provided by CD show secondary structure that is averaged over the entire peptide unlike the atomic resolution provided by NMR. The approximately 60 residues that from the NMR data show a more random coil like structure have the ability to dominate any signal that is provided by the helical region. In order to determine if this may be the case two samples were made: one short peptide that included just residues 944-961 and the full ctFCP1 region, Fig 2.3A and B respectively. The spectra are both characteristic of an IDP, however, they both show a subtle minimum at 220 nm that is characteristic of an α-helix . The short peptide still showed minimal helical character even though it lacks those residues shown to not have structural elements. These data are consistent with either a small region of helical tendency or only transient helix formation.

Small molecules called osmolytes are found in nature and are responsible for the stabilization of protein structure while under cellular stress[19]. This property is the reason osmolytes are often used as co-solutes to aid in studying protein folding processes and more recently in studying IDPs[20]. 2,2,2-trifluoroethanol (TFE) was chosen due to its history in the literature to cause short peptide chains to adopt helical structure regardless of a predisposition for the peptide to do so. At all TFE concentrations there was an increase in clarity and depth of the CD minimum at 222 nm for both the short peptide and ctFCP1 indicating an increase in overall helical content, Fig 2.3, colored curves. While ctFCP1 does show a similar change to the peptide it is not as dramatic. This is most likely because of the larger disordered region's signal convolution with helical signal with the binding region.

The CD data provided evidence that TFE can fold ctFCP1 and possibly allow the observation of an intermediate along the RAP74 binding pathway. Again, the CD data provides a signal that is averaged over the entire backbone, so it is difficult to determine how best to interpret the data. Given the CSI analysis first showed residue specific helical tendencies in the

apo state, we returned to utilizing NMR chemical shifts to study this folding transition. TFE was titrated from 0 to 40% using 2D $^{15}N,^{13}C$-CON-IPAP experiments. A full CSI could not be

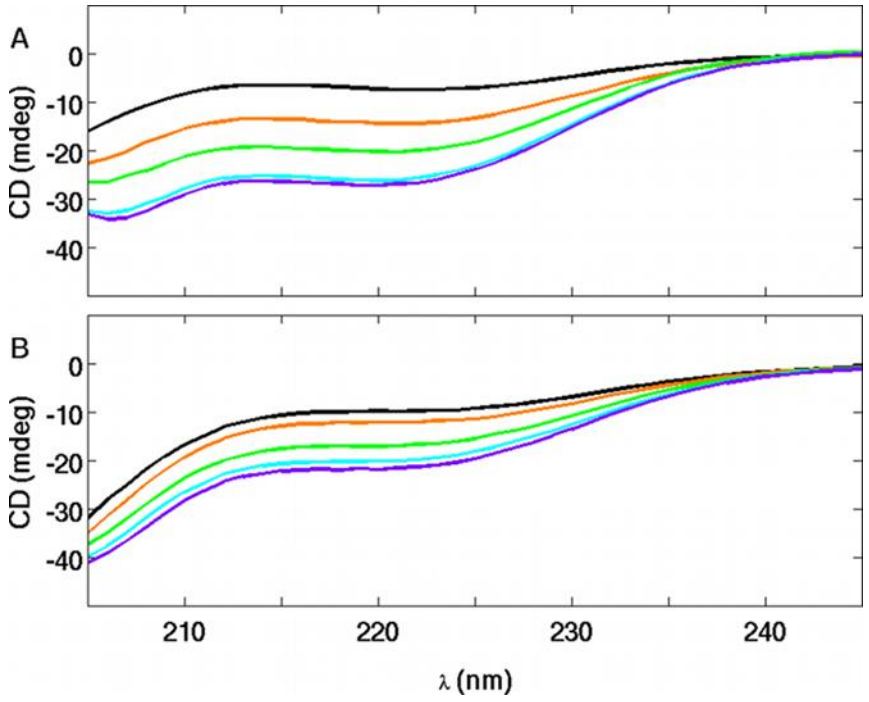

Figure 2.3*: Structural changes induced by TFE as monitored by circular dichroism. CD spectra of (A) a short FCP1 peptide (944-961) and (B) ctFCP1 (878-961) in 0% (black), 10% (orange), 20% (green), 30% (blue), and 40% (purple) TFE(v/v). There is an increasingly negative CD signal as the TFE volume percent increases with a minimum at 222 indicating an increase in α-helical structure.*

completed due to a sample lifetime too short to be amenable to 3D experiments. Fig 2.4 shows the difference in ^{13}CO chemical shift from the aqueous state for each TFE titration. The ^{13}CO chemical shift was chosen for its sensitivity to local structure. The chemical shift data shows a trend towards an increasingly positive deviation in RAP74 binding region indicating an increase in helicity. This trend is also isolated to the RAP74 binding region, and any deviations in any other area of the protein are small and insignificant compared to those in the binding region. This shows that the rest of ctFCP1 is largely unaffected by the presence of TFE and most likely remains disordered. Given that the chemical shift indexing has already show that the apo state at least transiently samples helical structure without the presence of TFE, we chose to titrate urea

from 0 to 3M and monitor the same ^{13}CO chemical shift deviations. These changes trended in the opposite direction as those with TFE indicating a loss in helical structure. Again chemical shifts throughout the rest of the protein show only small deviations indicating little to no structural changes due to the addition of urea. In combination with CSI analysis this demonstrates that ctFCP1 does have a native tendency to form helical secondary structure in the RAP74 binding region while the rest of the protein retains a significant amount of disorder.

The TFE and urea titrations have confirmed that ctFCP1 samples helical conformations while alone in dilute solution. In order to test the implications of RAP74 binding to a nascent helical structure we wanted to confirm that these solution conditions were representative of how ctFCP1 would exist in the cell. Polymeric crowding agents have been used to simulate a crowded cellular environment to study protein folding, binding affinities, and association kinetics[21,34]. Dextran is considered as an inert polymer in regards to protein-polymer interactions[35]. At a concentration of 30% (w/v), dextran should provide an environment similar to the crowding and confinement of a cellular environment without any contamination due to protein-polymer interactions. The 2D peak positions of each $^{15}N,^{13}C$-CON resonance showed virtually no change in 30% dextran compared to the aqueous state. The small changes that are observed are most likely due to the small changes in solvent dielectric or other physical parameters. However, given that the chemical shift changes are relatively small ctFCP1 does not fold to adopt a helical or more orderly structure in the RAP74 binding upon entering a crowded environment. As a further demonstration that the conformations that we observe in our NMR sample are physiologically relevant we collected a $^{15}N,^{13}C$-CON spectrum in the presence of 70 mg/mL HELA cell extract. The chemical shift changes were again relatively small and uniform across the entire protein. The only residue to show a chemical shift change of 0.2 ppm, His-918, was most likely reporting on changes in pH rather than changes in structure. Protease degradation over a timescale of hours prevented more detailed investigation. From these observations we can conclude our NMR

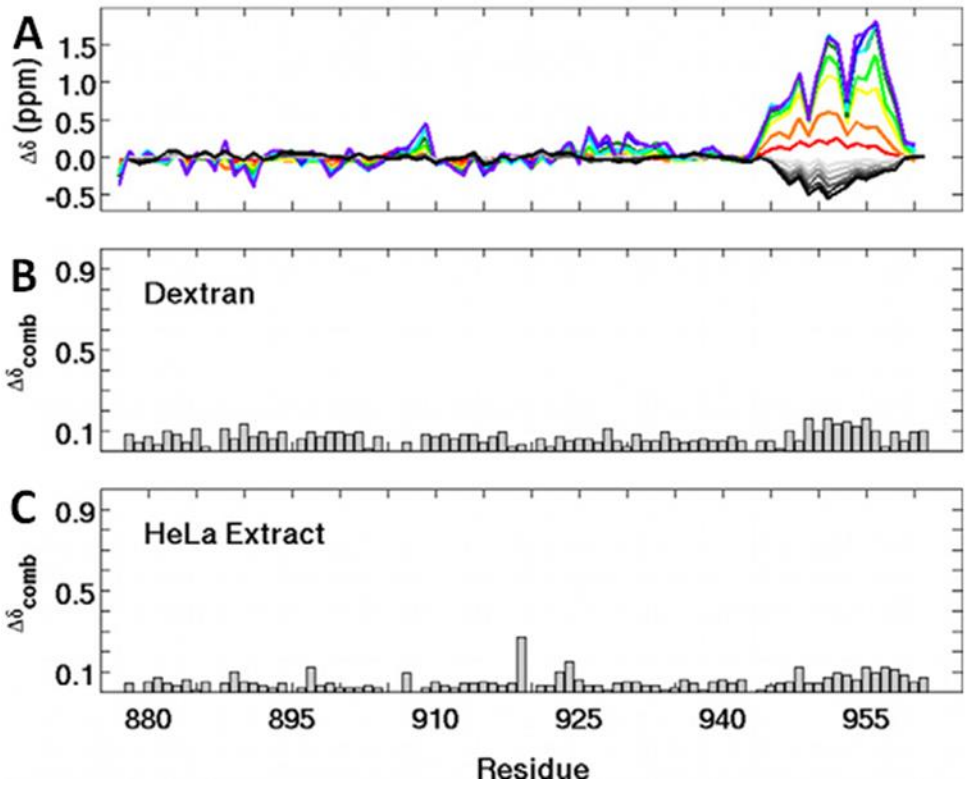

***Figure 2.4**: Structural response of ctFCP1 to co-solute addition. (A) Changes in ^{13}CO chemical shift during the TFE titration (5% red, 10% orange, 15% yellow, 20% light green, 25% dark green, 30% light blue, 35% dark blue, 40% purple) show a positive deviation indicating an increase in helicity. During the urea titration (increasing from 0.5M in light grey to 3.0M in black) there is an increasing negative deviation indicating a loss in structure. With the addition of crowding agents the backbone chemical shifts are largely unresponsive whether they are in the presence of 30% (w/v) dextran or (C) 70 mg/mL HeLa cell extract. This shows that FCP1 remains disordered even while under crowded conditions.*

structural analysis does provide a realistic picture of ctFCP1 prior to encountering RAP74 or RNAPII.

The paradigm in molecular biophysics that a protein's function is directly tied to its 3D structure is being rewritten by the existence of IDPs. Their prevalence in signaling and regulation of transcription only adds to the importance of modifying this model. It is now generally accepted that disordered 3D structure is just as functional as an orderly, cooperatively folded globular structure. IDPs still are found to adopt one or more temporally stable highly populated structures

while in complex with their binding partners. This suggests that conformational adaptation or selection may be a crucial aspect of IDP function just like dynamic structural rearrangement in enzymology. There is a large amount of interest in documenting the binding mechanism of IDPs and determining whether they fold prior to, during, or after the initial complex is formed[5]. Pre-existing structural elements encoded into the unbound state may provide an important reduction in the entropic penalty associated with folding-upon-binding, or it may be the case that these transient structures are not important at all if multiple conformations are involved in binding[6].

Conclusion

We have shown that apo-ctFCP1 has a tendency to adopt a helical structure in the unbound state and these structural elements are localized to the RAP74 binding region forming an α-helix while in complex. To the best of our knowledge that region of ctFCP1 only interacts with other proteins as an α-helix. Since this helix formation is highly localized and specific to protein interaction it is clearly important to the folding-upon-binding mechanism. Still the CD signal for ctFCP1 was weak and the minimal bias towards helical content in our NMR data suggests that these conformations are only temporally transient, or non-cooperative, or both. There are multiple ensembles consistent with the spectroscopic signatures for a-helicity which do not satisfy our physiochemical intuition about regarding the nature of the α-helix[36]. DSC measurements (data not shown) performed by the authors do not show a cooperative transition even in the presence of TFE, meaning that the transition should not be looked at as the two state transition associated with folding studies of globular proteins. The study in the presence of HeLa cell extract shows that the conformational tendencies observed are likely the same as those that would be encountered in the cell. Further studies aimed at defining the thermodynamics and kinetics of this folding upon binding interaction is critical.

References

1. Fuxreiter, M. *et al.* Malleable machines take shape in eukaryotic transcriptional regulation. *Nat. Chem. Biol.* **4,** 728–737 (2008).

2. Gsponer, J. & Madan Babu, M. The rules of disorder or why disorder rules. *Prog. Biophys. Mol. Biol.* **99,** 94–103 (2009).

3. Dunker, A. K. *et al.* The unfoldomics decade: an update on intrinsically disordered proteins. *BMC Genomics* **9,** S1 (2008).

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

5. Csermely, P., Palotai, R. & Nussinov, R. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends Biochem. Sci.* **35,** 539–546 (2010).

6. Uversky, V. N. Seven Lessons from One IDP Structural Analysis. *Structure* **18,** 1069–1071 (2010).

7. Hammes, G. G., Chang, Y.-C. & Oas, T. G. Conformational selection or induced fit: A flux description of reaction mechanism. *Proc. Natl. Acad. Sci.* **106,** 13737–13741 (2009).

8. Receveur-Bréchot, V., Bourhis, J.-M., Uversky, V. N., Canard, B. & Longhi, S. Assessing protein disorder and induced folding. *Proteins Struct. Funct. Bioinforma.* **62,** 24–45 (2005).

9. Jager, M. Structure-function-folding relationship in a WW domain. *Proc. Natl. Acad. Sci.* **103,** 10648–10653 (2006).

10. Ghosh, A., Shuman, S. & Lima, C. D. The Structure of Fcp1, an Essential RNA Polymerase II CTD Phosphatase. *Mol. Cell* **32,** 478–490 (2008).

11. Chapman, R. D., Heidemann, M., Hintermair, C. & Eick, D. Molecular evolution of the RNA polymerase II CTD. *Trends Genet.* **24,** 289–296 (2008).

12. Meinhart, A., Kamenski, T., Hoeppner, S., Baumli, S. & Cramer, P. A structural perspective of CTD function. *Genes Dev.* **19,** 1401–1415 (2005).

13. Abbott, K. L. *et al.* Interactions of the HIV-1 Tat and RAP74 Proteins with the RNA Polymerase II CTD Phosphatase FCP1 †. *Biochemistry (Mosc.)* **44,** 2716–2731 (2005).

14. Archambault, J. FCP1, the RAP74-Interacting Subunit of a Human Protein Phosphatase That Dephosphorylates the Carboxyl-terminal Domain of RNA Polymerase IIO. *J. Biol. Chem.* **273,** 27593–27601 (1998).

15. Kamada, K. Crystal structure of the C-terminal domain of the RAP74 subunit of human transcription factor IIF. *Proc. Natl. Acad. Sci.* **98,** 3115–3120 (2001).

16. Nguyen, B. D. NMR structure of a complex containing the TFIIF subunit RAP74 and the RNA polymerase II carboxyl-terminal domain phosphatase FCP1. *Proc. Natl. Acad. Sci.* **100,** 5688–5693 (2003).

17. Dyson, H. J. & Wright, P. E. Unfolded Proteins and Protein Folding Studied by NMR. *Chem. Rev.* **104,** 3607–3622 (2004).

18. Jensen, M. R., Salmon, L., Nodet, G. & Blackledge, M. Defining Conformational Ensembles of Intrinsically Disordered and Partially Folded Proteins Directly from Chemical Shifts. *J. Am. Chem. Soc.* **132,** 1270–1272 (2010).

19. Bolen, D. W. & Baskakov, I. V. The osmophobic effect: natural selection of a thermodynamic force in protein folding. *J. Mol. Biol.* **310,** 955–963 (2001).

20. Chang, Y.-C. & Oas, T. G. Osmolyte-Induced Folding of an Intrinsically Disordered Protein: Folding Mechanism in the Absence of Ligand. *Biochemistry (Mosc.)* **49,** 5086–5096 (2010).

21. Zhou, H.-X., Rivas, G. & Minton, A. P. Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences. *Annu. Rev. Biophys.* **37,** 375–397 (2008).

22. Mittag, T. & Forman-Kay, J. D. Atomic-level characterization of disordered protein ensembles. *Curr. Opin. Struct. Biol.* **17,** 3–14 (2007).

23. Jensen, M. R. *et al.* Quantitative Determination of the Conformational Properties of Partially Folded and Intrinsically Disordered Proteins Using NMR Dipolar Couplings. *Structure* **17,** 1169–1185 (2009).

24. Bermel, W., Bertini, I., Felli, I. C., Piccioli, M. & Pierattelli, R. 13C-detected protonless NMR spectroscopy of proteins in solution. *Prog. Nucl. Magn. Reson. Spectrosc.* **48,** 25–45 (2006).

25. Bermel, W. *et al.* Protonless NMR Experiments for Sequence-Specific Assignment of Backbone Nuclei in Unfolded Proteins. *J. Am. Chem. Soc.* **128,** 3918–3919 (2006).

26. O'Hare, B., Benesi, A. J. & Showalter, S. A. Incorporating 1H chemical shift determination into 13C-direct detected spectroscopy of intrinsically disordered proteins in solution. *J. Magn. Reson.* **200,** 354–358 (2009).

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

28. T. D. Goddard & D. G Kneller. SPARKY 3.

29. Dosztányi, Z., Mészáros, B. & Simon, I. ANCHOR: web server for predicting protein binding regions in disordered proteins. *Bioinformatics* **25,** 2745–2746 (2009).

30. Showalter, S. A. NMR assignment of the intrinsically disordered C-terminal region of Homo sapiens FCP1 in the unbound state. *Biomol. NMR Assignments* **3,** 179–181 (2009).

31. Wishart, D. S. & Sykes, B. D. in *Methods Enzymol.* **239,** 363–392 (Elsevier, 1994).

32. Wishart, D. & Sykes, B. The 13C Chemical-Shift Index: A simple method for the identification of protein secondary structure using 13C chemical-shift data. *J. Biomol. NMR* **4,** (1994).

33. Kelly, S. M., Jess, T. J. & Price, N. C. How to study proteins by circular dichroism. *Biochim. Biophys. Acta* **1751,** 119–139 (2005).

34. Uversky, V. N. Intrinsically Disordered Proteins and Their Environment: Effects of Strong Denaturants, Temperature, pH, Counter Ions, Membranes, Binding Partners, Osmolytes, and Macromolecular Crowding. *Protein J.* **28,** 305–325 (2009).

35. Jiao, M., Li, H.-T., Chen, J., Minton, A. P. & Liang, Y. Attractive Protein-Polymer Interactions Markedly Alter the Effect of Macromolecular Crowding on Protein Association Equilibria. *Biophys. J.* **99,** 914–923 (2010).

36. Johnson, E., Showalter, S. A. & Brüschweiler, R. A multifaceted approach to the interpretation of NMR order parameters: a case study of a dynamic alpha-helix. *J. Phys. Chem. B* **112,** 6203–6210 (2008).

Chapter 3 Development of Carbon-Detected ^{15}N Spin Relaxation to Measure the Dynamics of FCP1 Unbound and In Complex with RAP74

[Published: Lawrence, C. W. & Showalter, S. A. Carbon-Detected 15N NMR Spin Relaxation of an Intrinsically Disordered Protein: FCP1 Dynamics Unbound and in Complex with RAP74. *J. Phys. Chem. Lett.* 3, 1409–1413 (2012).]

Abstract

Intrinsically disordered proteins are a unique class of proteins that lack a unique 3D structure under native conditions and exist in as highly dynamic ensembles in solution. Here we present two experiments for the measurement of ^{15}N NMR spin relaxation, the CON(T1)-IPAP and CON(T2)-IPAP, that can be used to quantify the backbone dynamics on a per residue basis of IDPs free in solution. We have applied these experiments onto the C-terminal domain of FCP1 while free in solution and in complex with the winged helix domain of RAP74. After analysis of the acquired data most of FCP1 remains highly dynamic in both states, but the 20 residues that form direct contact with RAP74 become more ordered in the complex. The backbone ^{15}N NMR spin relaxation of RAP74 was studied in parallel and reveals that RAP74 only gains a limited amount of ordering upon interaction with FCP1. Using the data taken from both perspectives of this binding they show that folding-upon-binding is highly local in this system and there is a level of disorder that prevails even in the complex.

Introduction

Unlike globular proteins that typically have regular and persistent structural elements intrinsically disordered proteins (IDPs) lack any secondary or tertiary structure[1–3]. Also, rather than being a rare occurrence in the proteome, as once thought, IDPs are routinely found to function in signaling, transcription, and translation[4]. IDPs require an ensemble description for

their solution state and this structural ensemble as well as the disorder-to-order transition that often accompanies IDP binding, need to be well understood if there are to be any molecular insights into their function[5,6]. As proteins are not static in solution, their dynamics are integral for their function. Due to the highly dynamic nature of IDPs, measurements of their local motions are necessary to gain a useful ensemble description of their native states. This makes solution NMR spectroscopy a uniquely attractive technique for their characterization. Recent successes in the application of NMR spectroscopy[7,8] and molecular modeling[9,10] techniques have provided useful insights into the free energy landscapes and folding upon binding transitions of IDPs. Because these transitions always include a change in backbone conformation, performing a set of NMR relaxation experiments measuring backbone conformational dynamics can provide meaningful insight into the binding mechanism of IDPs.

NMR spin relaxation provides a physical picture of the internal motions of a protein. The N-H bond is typically a proxy for motion of the rigid peptide plane while the rest of the protein is represented as a sphere[11]. Both global and internal motions can be represented through what is known as a spectral density function, $J(\omega)$, which describes the frequency populations of the system. However, $J(\omega)$ is not directly accessible by any experimental method making it necessary to create a model from which $J(\omega)$ can be estimated[12,13]. One such model assumes that when the N-H bond motion is seen as relative to a fixed reference frame three constraints can be used to describe its motion: the molecular correlation time (τ_m), the effective rotational correlation time for local motions (τ_e), and a general order parameter to describe the amplitude of local motion within the reference frame (S^2)[14]. From these parameters one can calculate a spectral density function that may describe the dynamics of the protein. In order to evaluate the effectiveness of the model, experimental parameters that directly depend on the spectral density function are measured. These parameters are the longitudinal relaxation time (T_1), the transverse relaxation time (T_2), and the heteronuclear NOE, although these three observables alone do not provide

enough information to uniquely determine the spectral density function. Therefore, an iterative approach must be taken where the values of the general motional parameters are assumed and $J(\omega)$ is calculated[12,14]. From that assumed spectral density function T_1, T_2, and NOE values are determined and compared to the values that were obtained experimentally. If these values do not match then new motion parameters are assumed and the process repeats until likely values of τ_e, τ_m, and S^2 are found. Qualitatively one can determine S^2 from the T_1 and T_2 relaxation times and τ_e from the NOE.

Monitoring backbone ^{15}N spin relaxation in IDPs requires the development of NMR techniques capable of making quantitative measurements with a read-out format that will provide comprehensive coverage of the residues along the polypeptide chain. This is not possible with conventional experiments, which tend to suffer from poor spectral quality for IDPs. Using typical NMR techniques has made the study of IDPs particularly difficult due to poor spectral quality which has limited the number of IDPs subjected to experimental NMR relaxation studies. ^{13}C-detected NMR spectroscopy has recently become a useful alternative to traditional ^{1}H-detected NMR[15–17]. Purely heteronuclear ^{13}C-detected and triple resonance assignment methodologies built around the detection of the backbone C' or C^{α} are now available and have become especially prominent in studying IDPs because of their desirable spectral properties compared to ^{1}H-detection[18–22]. Using ^{13}C-direct detection to characterize protein dynamics has been demonstrated in multiple reports of ^{13}C T_1 and T_2 measurements[23,24] and C^{α}- detected NOE[25], however, these methods were designed with lengthy INEPT transfer sequences around a probe specially designed for ^{13}C-direct detection. We have previously reported the use of ^{1}H-start triple resonance experiments built from ^{15}N,^{13}C-CON that work well with stock Bruker H-inner coil TCI cryoprobes[21]. These experiments have been applied to the C-terminal domain of FCP1, a phosphatase that acts on the C-terminal domain of RNA polymerase II (RNAPII), which

promotes the recycling of RNAPII in the mRNA biogenesis cycle[26]. FCP1 will also be used as a model system in the present study.

FCP1 interacts with the winged helix domain of RAP74, a subunit of transcription factor IIF (TFIIF), which is responsible for promoting FCP1's phosphatase activity. Upon interaction with RAP74 FCP1 undergoes a disorder-to-order transition into an α-helix that is isolated to residues 944-961[27]. Utilizing ^{13}C-direct detection we have shown that those residues that bind RAP74 have a tendency to be helical while free in solution[28], and subsequent molecular dynamics calculations have indicated that FCP1 may be less ordered when in complex than previously thought[29]. This has led to the hypothesis that FCP1 loses little conformational entropy upon binding to RAP74 and that the folding upon binding transition may yield a complex ensemble that is highly dynamic. To date the NMR spin relaxation for either RAP74 or FCP1 has yet to be reported leaving a need for the experimental validation of these hypotheses.

Here we report the use of $^{1}H^{N}$-start experiments built from the conventional ^{15}N,^{13}C-CON pulse programs, that readily measure T_1 and T_2 relaxation rates on standard ^{1}H-inner coil cryogenic probes, called the CON(T_1)-IPAP and CON(T_2)-IPAP. Application of these novel pulse sequences validates that interactions with RAP74 as well as any gain in FCP1 structure are isolated to the C-terminus. Also we will show that FCP1 retains a certain amount of flexibility in the FCP1/RAP74 complex and can then be described as “fuzzy” like other IDP complexes in the literature.

Materials and Methods

Recombinant FCP1 83aa (878-961) and RAP74 91aa (426-517) constructs with codon usage optimized for expression in *Escherichia coli* were cloned into the expression vector pET-47b (Novagen) and transformed into *E. coli* BL21 competent cells. The cells were grown M9 media containing both $^{15}NH_4Cl$ and ^{13}C-glucose (Cambridge Isotope Laboratories, Inc.) and LB media until an optical density between 0.5 and 0.6 for M9 and 0.8-0.9 for LB at 600nm was

reached. ctFCP1 and RAP74 expression was induced using 0.5M IPTG followed by growth for an additional 3-4 hours and harvesting by sonication. Following centrifugation, the clarified cell lysates were run over a Ni-NTA column for purification of the His-tagged ctFCP1 and RAP74. The proteins were dialyzed in the presence of 3-Cprotease overnight at 4°C to remove the His-tag, leaving a non-native Gly-Pro-Gly tripeptide at the N-terminus of the protein. The His-tagged 3-C protease was removed by running the sample after dialysis over a Ni-NTA column resulting in highly pure ctFCP1 and RAP74. An NMR sample was obtained by spin-concentration and buffer exchange in a VIVASPIN 20 centrifugal concentrator. The final buffer conditions for the ^{15}N, ^{13}C labeled sample were 20mM cacodylate, pH 7.0, 100mM NaCl, 0.002% w/v NaN3, and 90% H_2O/ 10% D_2O (v/v). CD samples were made from lyophilized protein to yield a final concentration of 20µM in 20mM cacodylate at a pH of 7.0.

All data were collected on Bruker Avance 3 spectrometers operating at 11.7 T or 14.1 T static magnetic field strength and equipped with a TCI-cryoprobes for maximum carbon sensitivity. In the H-Start-CON(T_1)-IPAP, CON(T_2)-IPAP ^{15}N, Figure 3.1, chemical shift evolves under semi constant time, there is a second Refocused-INEPT to transfer polarization to the carbonyl carbon. Decoupling of C^{α} from the C' was achieved through asymmetric shaped pulsed during the ^{15}N chemical shift evolution period. ^{1}H decoupling was achieved with the WALTZ-16 sequence applied at 3.1 kHz. The narrow and wide pulses correspond to 90° and 180° pulses respectively, and the same applies to narrow and wide shaped pulses. The 90° and 180° shaped pulses in each experiment were Q5 (384 µs 11.7T; 320 µs 14.1T) and Q3 (307 µs 11.7T; 256 µs 14.1T) respectively except for the two asymmetrically placed carbon pulses in during the T_1 and T_2 periods which were Chirp pulses (500 µs). T_1 and T_2 relaxation delays were incremented as follows: T_1-50, ^{15}0, 250, 350, 450, 550, 650, and 750ms; and T_2- 16, 48, 81, 112, 144, 176, 208, 240, 272, and 304ms. Pulse field gradients (PFG) are also indicated by shaped pulses. The sweep widths are $^{13}C'$ = 20 ppm and ^{15}N = 36 ppm. Quadrature detection was obtained in all

experiments by States-TPPI incrementing of Φ_1 and Φ_2. All relaxation data was processed in NMRpipe[30] and analyzed using Sparky[31] and Matlab (MathWorks). RAP74 relaxation data were fit using the Lipari-Szabo model-free formalism[14] as implemented in ModelFree 4.20[32], with diffusion tensor fitting performed using the quadric method[33]. The coordinates from the crystal structures of apo-RAP74 (1I27)[34] and RAP74 in complex with FCP1 (1J2X)[27] were used as structural references for axially symmetric diffusion tensor determination.

Results and Discussion

NMR spin relaxation measurements provide powerful insight into the biological function of globular proteins. However, this requires well resolved spectra in order to get any kind of quantitative analysis, which is why these methods have not been readily applied to intrinsically disordered proteins. The novel ^{13}C-direct detected methods presented here circumvent the difficulties associated with IDPs because of the enhanced resolution and spectral dispersion that is associated with the ^{15}N,^{13}C-CON (Figure 3.2A). Plots of T_1 and T_2 as function of residue number are far from featureless and include considerable variation from baseline, similar to other IDPs previously studied by related methods[18,35–37], as shown in Figure 3.2B and C. As an example, residues 900-915 show more flexibility than the baseline, as evidenced by a simultaneous increase in the T_1 and T_2. While FCP1 does not adopt a stable globular structure it is possible to calculate an apparent correlation time from the ratio of R2 to R1. The $t_{m\,app}$ for FCP1 remains relatively constant for residues 890-940, with means of 3.3±0.4 and 3.0±0.6 ns for FCP1 in the apo and bound state respectively. The binding region of FCP1 was similar to the rest of the protein, however when bound to RAP74 the t_{app} for that region increases to 12±2 ns.

Binding of IDPs, as previously mentioned, is typically associated with a folding transition which could lead to loss of conformational entropy if RAP74 confines FCP1 causing a loss in flexibility, thus, it is especially important to change the dynamics of the binding region. In a rigid

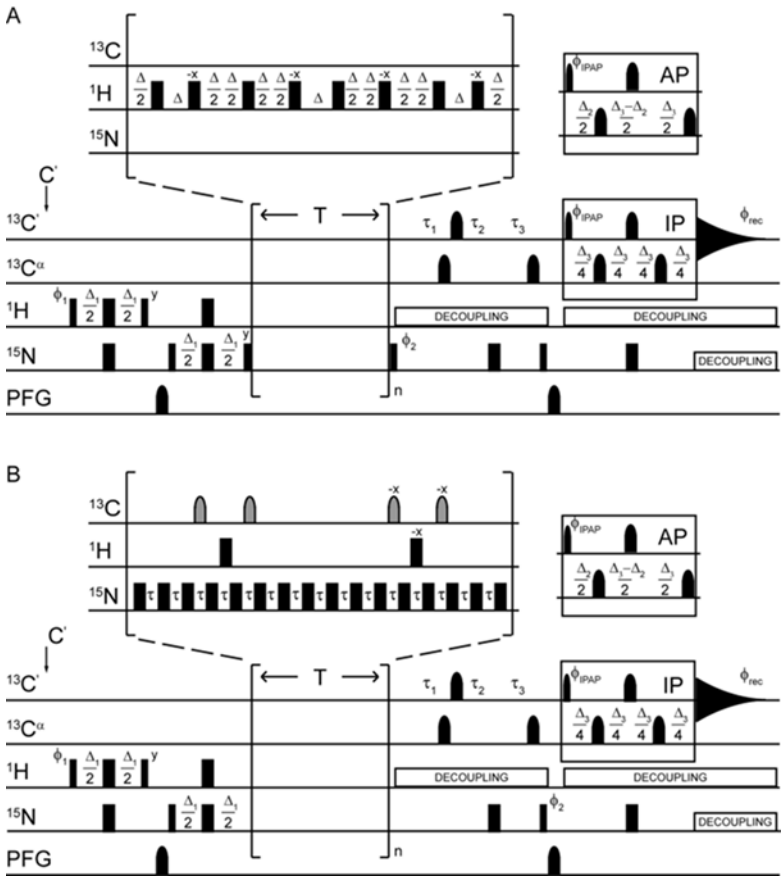

Figure 3.1: Pulse sequences for the A) CON(T1)-IPAP and B) CON(T2)-IPAP experiments. The delays are Δ = 5 ms, Δ_1 = 4.6 ms, Δ_2 = 9.0 ms, Δ_3 = 25 ms. $^{15}N(t_1)$ chemical shift evolution is performed using semiconstant time period with delays $\tau_1 = (\Delta_3 + t_1)/2$, $\tau_2 = (1- \Delta_3 / t_{1max})t_1 /2$, $\tau_3 = (1- t_1/t_{1max})\Delta_3 /2$. Pulses are applied at the frequency indicated at the left of each line , with narrow and wide rectangles or shapes representing 90 and 180° pulses, respectively. The insets in panels A and B show an expansion of the incremented relaxation delay periods and inclue grey pulses on ^{13}C to indicate band-selective $^{13}C'$ and $^{13}C^{\alpha}$ inversion pulses for refocusing of the ^{15}N-^{13}C coupling. In both cases, the phase cycle is φ_1 = x,-x; φ_2 = y, y, -y, -y; φ_{IPAP} (IP) =x, x, x, x -x ,-x, -x, -x; φ_{IPAP} (AP) = -y, -y, -y, -y, y, y, y, y; and φ_{rec} = x, -x, -x, x, x, -x, -x, x. Quadrature detection in the indirect dimension was obtained by States-TPPI incrementation of φ_2.

complex the average T_1 and T_2 values for both proteins would be dictated by the same global tumbling time. If this accurately describes the FCP1/RAP74 complex, the baselines for both proteins should match. This is not the case, the average T_1 for residues 944-960 in FCP1, the RAP74 binding region, are significantly higher than the average T_1 of the globular RAP74. The T_2 shows an opposite trend as FCP1 shows a lower baseline than expected in the complex (Figure 3.2B and C). The elevated T_1 baseline for FCP1 relative the RAP74 (discussed below), implies large amplitude local dynamics of those residues in the complex. This is consistent with previous molecular dynamics simulations[29], as well as the very high B-factors that are associated with the crystal structure[27]. This relatively high amount of dynamics compared the other complex should also elevate the T_2, relative to the baseline observed for RAP74; however, the T_2 for FCP1 shows an opposite trend. This is most likely due to chemical exchange on the intermediate time scale causing excess line broadening of the FCP1 resonances. NMR relaxation dispersion measurements will be necessary for a fully quantitative picture of the dynamics in this system to be developed.

The measured spin relaxation data validate that FCP1 does retain its dynamic nature even while in complex. In order to complete the characterization of this system, ^{15}N spin relaxation measurements were made of RAP74 both free in solution and in complex. The analysis of the data using the quadric method indicates that RAP74 tumbling is moderately anisotropic in solution as reported by $D_{\|}/D_{\perp}$ = 1.20 and 1.17 for apo-RAP74 and the complex respectively. Using the axially symetric diffusion tensor apo-RAP74 was fit to a t_{iso} that is higher than usual for protein of this size; equal to 6.8 ns. This can be explained by a 20 residue disordered tail at the N-terminus. NMR pulsed-field gradient diffusion measurements yielded a translational diffusion coefficient of 1.14x10-10 m^2/s for apo-RAP74, which is also consistent with the expected result for a species with RAP74's monomer weight. With the addition of FCP1's mass to the domain t_{iso} increases to 8.6 ns.

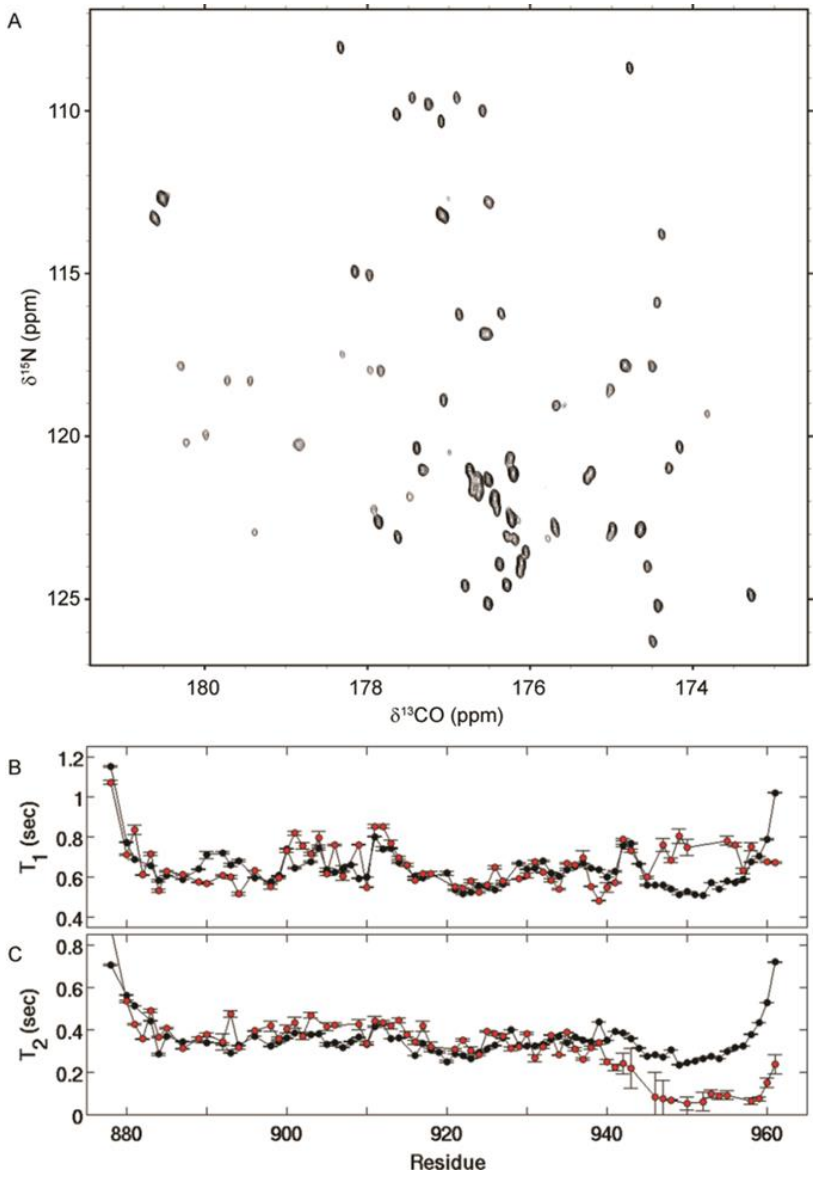

Figure 3.2*: A) $^{15}N,^{13}C$-CON(T2)-IPAP spectrum of FCP1 in complex with RAP74 collected with a relaxation delay of T = 16ms using the pulse program shown in Figure 1B. Panels B and C show the fitted values of T_1 and T_2, respectively, with error bars representing the uncertainty from fitting to an single exponential decay. In both cases apo-FCP1 is shown in black and holo-FCP1 in red.*

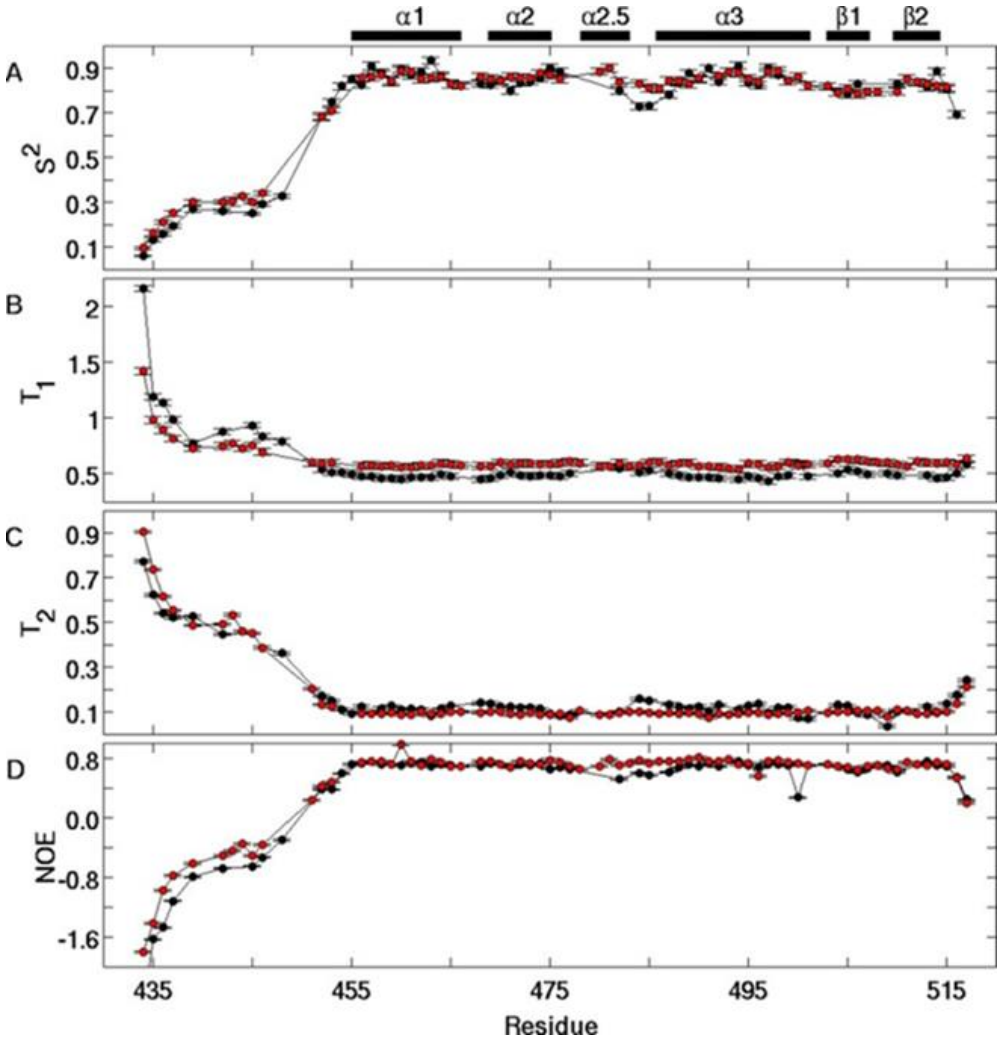

Figure 3.3: The backbone order parameters for RAP74 generated from Lipari-Szabo model-free analysis of the B) ^{15}N-T_1, C) ^{15}N-T_2, and D) [^{1}H-^{15}N]-NOE in the unbound and bound state, black and red respectively. The error bars in panel A represent uncertainties from fitting in Modelfree-4.20 and error bars from the relaxation rates comes from uncertainty in fitting to a single exponential. Error bars in the NOE represent uncertainty from averaging of two replicate experiments.

Given RAP74's globular nature it is meaningful to calculate generalized ordered parameters in order to gain a clear representation of its backbone flexibility. From a plot of S^2 as a function of residue number (Figure 3.3A) it is evident that RAP74 only experiences a minimal change in its backbone dynamics upon binding FCP1. There is one region of change at helix 2.5 and the loop connecting it with helix 3 of RAP74, showing that it is losing flexibility. This region of RAP74 has been proposed to mediate interactions with CK2 phosphorylated FCP1[38], thought

to tighten binding by stabilizing pre-existing interactions. Just like the region of FCP1 that remains disordered and does not interact in the complex, the N-terminus of RAP74 is largely unaffected by complex formation.

A visual representation of the data discussed above can be found in Figure 3.4. In the ribbon diagram of the complex, those residues shown through relaxation measurements not to be affected by FCP1 binding are colored white. Helix 2.5, which showed a loss in flexibility, is colored blue and the N-terminus is left out for clarity. FCP1, treated similarly, has those residues that become more rigid shaded pink and those that remain flexible shaded white. To provide a proper picture of our working model of the FCP1/RAP74 complex, 5 snapshots of FCP1, spaced one every 50 ns from our laboratory's previous molecular dynamics simulation[29], were superimposed upon the complex. As shown previously, the region that experiences conformational restriction upon forming the complex extends well beyond the construct shown in the crystal structure. As a result, the snapshots shown in the figure were extended by 10 residues from the crystal structure to provide a better representation of how the remaining disordered region is oriented as it leaves the complex. From this perspective several of the FCP1 snapshots come into contact with the part of helix 2.5 that becomes more rigid. The proposed CK2 phosphorylation sites of FCP1 have been shown to increase RAP74 binding affinity in vitro, which could restrict the conformations available to FCP1 while in the complex stabilizing protein-protein contacts. If this is the case, then the amount of disorder retained by FCP1 in the complex can be affected by post-translational modification as a means to regulate the affinity for this interaction. This is supported by a study done with an alternative peptide that showed enhanced binding affinity to RAP74 upon phosphorylation of analogous serines[39].

Conclusion

Through the application of ^{13}C-direct detected ^{15}N spin relaxation experiments reported here, we were able to investigate the conformational dynamics of FCP1, both free in solution and

in complex. These studies provided insight into IDP conformational sampling, as well as the nature of the folding-upon-binding mechanism. The measurement of ^{15}N spin relaxation in globular proteins has become routine as an aide in constructing binding mechanisms and rationalizing measured thermodynamics. For globular proteins, such as RAP74, large changes in the backbone dynamics are rarely observed, as evidenced by the highly localized response by helix 2.5. However, IDPs must undergo a considerable change in their backbone dynamics as they become fundamentally more restricted just by associating with another protein. This makes the measurement and analysis of spin relaxation necessary to understand such conformational restrictions. For example, recent work done by Kriwacki and colleagues demonstrates that dynamic associations mediated by the D2 subdomain of the IDP p27 modulate the affinity of association with a variety of Cdk/cyclin complexes that are present at temporally distinct stages of cell cycle progression[40].

From the data reported we have shown that there are very few residues in FCP1 that experience any change in local dynamics, but those 20 residues that are affected by association with RAP74 show a dramatic change. Also serines 942 and 944 which are possible targets of post-translational modification are located right at the boundary between the RAP74 binding region and the disordered N-terminal region. It is this region that may act as a regulator for RAP74 binding affinity.

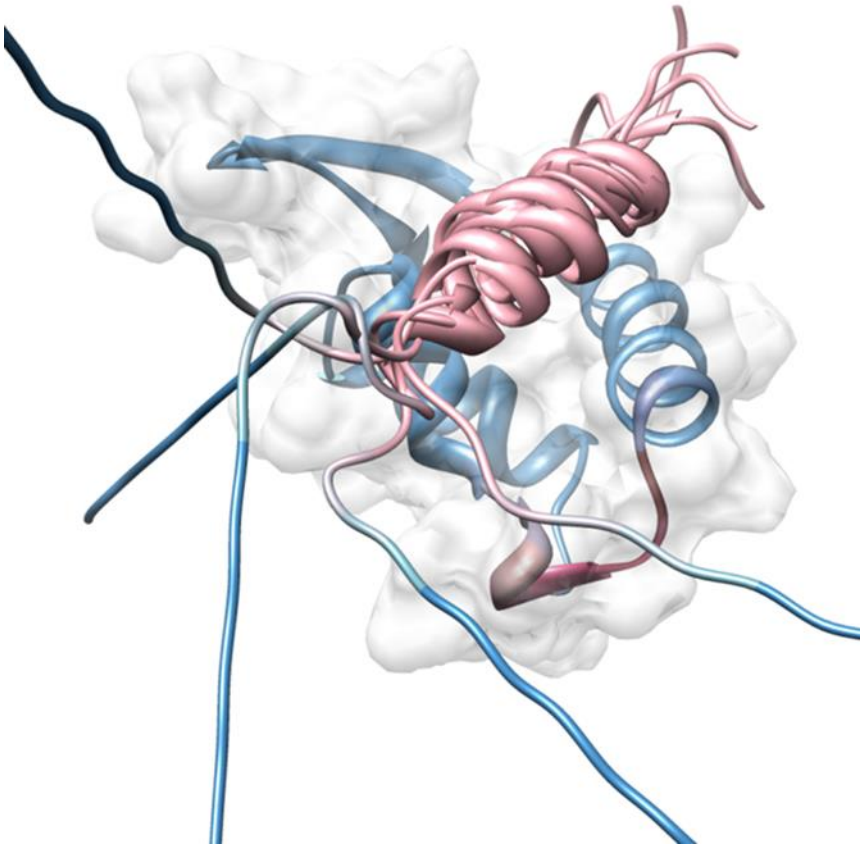

Figure 3.4: A ribbon diagram of the FCP1-RAP74 complex depicting the change in general order parameter experienced by both proteins through interaction. RAP74 is depicted as a backbone ribbon diagram embedded within a transparent van der walls surface. In RAP74 blue represents those sites that become more ordered upon binding FCP1 while those colored white represent those that experience no significant change in S^2. FCP1 is represented as a ribbon colored pink for sites that become more ordered upon binding RAP74 and white for those that do not experience a significant change in the relaxation rates Models of FCP1 are taken from snapshots spaced every 50 ns through a MD trajectory of the FCP1-RAP74 complex[29] (which modeled FCP1 944-61) with 10 extra residues added to the N-terminus in an extended conformation. Contact between this region of FCP1 and RAP74 is suggested by the increase in the order parameter for the helix 2.5 loop of RAP74 and the persistence of increased order in FCP1 past the end of the canonical α-helx.

References

1. Uversky, V. N. & Dunker, A. K. Understanding protein non-folding. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **1804,** 1231–1264 (2010).
2. Dunker, A. K. *et al.* The unfoldomics decade: an update on intrinsically disordered proteins. *BMC Genomics* **9,** S1 (2008).
3. Tompa, P. *et al.* Close encounters of the third kind: disordered domains and the interactions of proteins. *BioEssays* **31,** 328–335 (2009).
4. Dyson, H. J. & Wright, P. E. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* **6,** 197–208 (2005).
5. Mészáros, B., Simon, I. & Dosztányi, Z. The expanding view of protein–protein interactions: complexes involving intrinsically disordered proteins. *Phys. Biol.* **8,** 035003 (2011).
6. Csermely, P., Palotai, R. & Nussinov, R. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends Biochem. Sci.* **35,** 539–546 (2010).
7. Wells, M. *et al.* Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain. *Proc. Natl. Acad. Sci.* **105,** 5762–5767 (2008).
8. Mittag, T. *et al.* Structure/Function Implications in a Dynamic Complex of the Intrinsically Disordered Sic1 with the Cdc4 Subunit of an SCF Ubiquitin Ligase. *Structure* **18,** 494–506 (2010).
9. Ganguly, D. & Chen, J. Topology-based modeling of intrinsically disordered proteins: Balancing intrinsic folding and intermolecular interactions. *Proteins Struct. Funct. Bioinforma.* **79,** 1251–1266 (2011).
10. Liu, J., Faeder, J. R. & Camacho, C. J. Toward a quantitative theory of intrinsically disordered proteins and their function. *Proc. Natl. Acad. Sci.* **106,** 19819–19823 (2009).

11. Palmer, A. G., 3rd. Probing molecular motion by NMR. *Curr. Opin. Struct. Biol.* **7,** 732–737 (1997).

12. 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 (2006).

13. Luginbühl, P. & Wüthrich, K. Semi-classical nuclear spin relaxation theory revisited for use with biological macromolecules. *Prog. Nucl. Magn. Reson. Spectrosc.* **40,** 199–247 (2002).

14. Lipari, G. & Szabo, A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.* **104,** 4546–4559 (1982).

15. Z, S. *et al.* New carbon-detected protein NMR experiments using CryoProbes. *J. Am. Chem. Soc.* **122,** 3554–3555 (2000).

16. Bermel, W., Bertini, I., Felli, I. C., Piccioli, M. & Pierattelli, R. 13C-detected protonless NMR spectroscopy of proteins in solution. *Prog. Nucl. Magn. Reson. Spectrosc.* **48,** 25–45 (2006).

17. Bermel, W., Felli, I. C., Kümmerle, R. & Pierattelli, R. 13C Direct-detection biomolecular NMR. *Concepts Magn. Reson. Part* **32A,** 183–200 (2008).

18. Csizmok, V., Felli, I. C., Tompa, P., Banci, L. & Bertini, I. Structural and Dynamic Characterization of Intrinsically Disordered Human Securin by NMR Spectroscopy. *J. Am. Chem. Soc.* **130,** 16873–16879 (2008).

19. Bermel, W. *et al.* H-start for exclusively heteronuclear NMR spectroscopy: The case of intrinsically disordered proteins. *J. Magn. Reson.* **198,** 275–281 (2009).

20. Showalter, S. A. NMR assignment of the intrinsically disordered C-terminal region of Homo sapiens FCP1 in the unbound state. *Biomol. NMR Assignments* **3,** 179–181 (2009).

21. O'Hare, B., Benesi, A. J. & Showalter, S. A. Incorporating 1H chemical shift determination into 13C-direct detected spectroscopy of intrinsically disordered proteins in solution. *J. Magn. Reson.* **200,** 354–358 (2009).

22. Bermel, W. *et al.* Protonless NMR Experiments for Sequence-Specific Assignment of Backbone Nuclei in Unfolded Proteins. *J. Am. Chem. Soc.* **128,** 3918–3919 (2006).

23. Pasat, G., Zintsmaster, J. S. & Peng, J. W. Direct 13C-detection for carbonyl relaxation studies of protein dynamics. *J. Magn. Reson.* **193,** 226–232 (2008).

24. Bermel, W., Bertini, I., Felli, I. C., Peruzzini, R. & Pierattelli, R. Exclusively Heteronuclear NMR Experiments to Obtain Structural and Dynamic Information on Proteins. *ChemPhysChem* **11,** 689–695 (2010).

25. Bertini, I., Felli, I. C., Gonnelli, L., Kumar M., V. & Pierattelli, R. High-Resolution Characterization of Intrinsic Disorder in Proteins: Expanding the Suite of 13C-Detected NMR Spectroscopy Experiments to Determine Key Observables. *ChemBioChem* **12,** 2347–2352 (2011).

26. Archambault, J. *et al.* An essential component of a C-terminal domain phosphatase that interacts with transcription factor IIF in Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci. U. S. A.* **94,** 14300–14305 (1997).

27. Kamada, K., Roeder, R. G. & Burley, S. K. Molecular mechanism of recruitment of TFIIF- associating RNA polymerase C-terminal domain phosphatase (FCP1) by transcription factor IIF. *Proc. Natl. Acad. Sci.* **100,** 2296–2299 (2003).

28. Lawrence, C. W., Bonny, A. & Showalter, S. A. The disordered C-terminus of the RNA Polymerase II phosphatase FCP1 is partially helical in the unbound state. *Biochem. Biophys. Res. Commun.* **410,** 461–465 (2011).

29. Wostenberg, C., Kumar, S., Noid, W. G. & Showalter, S. A. Atomistic Simulations Reveal Structural Disorder in the RAP74-FCP1 Complex. *J. Phys. Chem. B* **115,** 13731–13739 (2011).

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

31. T. D. Goddard & D. G Kneller. SPARKY 3.

32. Mandel, A. M., Akke, M. & Palmer, I. Backbone Dynamics ofEscherichia coliRibonuclease HI: Correlations with Structure and Function in an Active Enzyme. *J. Mol. Biol.* **246,** 144–163 (1995).

33. Lee, L. K., Rance, M., Chazin, W. J. & Iii, A. G. P. Rotational diffusion anisotropy of proteins from simultaneous analysis of 15N and 13Cα nuclear spin relaxation. *J. Biomol. NMR* **9,** 287–298 (1997).

34. Kamada, K., Angelis, J. D., Roeder, R. G. & Burley, S. K. Crystal structure of the C-terminal domain of the RAP74 subunit of human transcription factor IIF. *Proc. Natl. Acad. Sci.* **98,** 3115–3120 (2001).

35. Olson, K. E. *et al.* Secondary Structure and Dynamics of an Intrinsically Unstructured Linker Domain. *J. Biomol. Struct. Dyn.* **23,** 113–124 (2005).

36. Bertoncini, C. W. *et al.* Structural Characterization of the Intrinsically Unfolded Protein β-Synuclein, a Natural Negative Regulator of α-Synuclein Aggregation. *J. Mol. Biol.* **372,** 708–722 (2007).

37. Libich, D. S. & Harauz, G. Backbone Dynamics of the 18.5 kDa Isoform of Myelin Basic Protein Reveals Transient α-Helices and a Calmodulin-Binding Site. *Biophys. J.* **94,** 4847–4866 (2008).

38. Abbott, K. L. *et al.* Enhanced Binding of RNAP II CTD Phosphatase FCP1 to RAP74 Following CK2 Phosphorylation†. *Biochemistry (Mosc.)* **44,** 2732–2745 (2005).

39. Nguyen, B. D. *et al.* NMR structure of a complex containing the TFIIF subunit RAP74 and the RNA polymerase II carboxyl-terminal domain phosphatase FCP1. *Proc. Natl. Acad. Sci.* **100,** 5688–5693 (2003).

40. Ou, L. *et al.* Incomplete Folding upon Binding Mediates Cdk4/Cyclin D Complex Activation by Tyrosine Phosphorylation of Inhibitor p27 Protein. *J. Biol. Chem.* **286,** 30142–30151 (2011).

Chapter 4 Utilizing Isothermal Titration Calorimetry to study the Interaction of FCP1 and RAP74: Order May Disrupt IDP Binding

Abstract

Intrinsically disordered proteins (IDPs) lack unique 3D structure when under native conditions. FCP1, a phosphatase of RNA Polymerase II (RNAPII), is one such IDP that has been shown to bind RAP74, a subunit of transcription factor IIF. This binding event is associated with a disorder-to-order transition in the C-terminal of FCP1 that forms an α-helix, which has been shown both by NMR and crystalographically. However, it has yet to be determined if a small population of the folded conformation is selected from the unbound ensemble as the initial stage of RAP74 binding or FCP1 is induced to fold after binding. Based on the results from our previous experimental NMR and CD work, the folded state of FCP1 can be enriched in the unbound ensemble through the addition of using small molecule osmolytes, such as TFE. We will show that the overall binding interaction is dominated by burial of hydrophobic surfaces from aqueous solvent, and regardless of FCP1 folding into an α-helix in the complex the interaction is still entropically favorable. Also, we will show through the use of structure stabilizing osmolytes that changes in the ensemble state of FCP1 have a large effect on the overall binding interaction.

Introduction

Within the cell proteins are responsible for most biological processes and their functions in these processes are directly tied to their native structure. Unexpectedly, intrinsically disordered proteins (IDPs) are a class of proteins that take part in important biological functions, such as transcription, translation, and cellular signal transduction, but lack any unique secondary or tertiary structure[1–3]. These IDPs force us to expand the definition of a biologically active native state to include those states that are both spatially and temporally disordered. IDPs are characterized as having a low sequence complexity as well as a compositional bias that has led to the use of empirical “charge-hydropathy” relationships in their identification[4,5]. The lack of

hydrophobic amino acids allows for disorder because, if present, those amino acids would be excluded from the aqueous interface, thus forming the hydrophobic core of a globularly structured protein. It is this low sequence complexity that makes for reliable prediction of disorder when compared to globular proteins[2]. Even though IDPs are highly disordered in their native state, many IDPs are known to undergo a disorder-to-order transition when interacting with a binding partner[6–8]. However, it is uncertain how these disorder-to-order transitions are incorporated into the overall binding mechanism.

As a model system to study the disorder-to-order transition associated with IDP binding we have chosen FCP1 (TFII-F associating component of the RNA Polymerase II C-terminal domain phosphatase). FCP1 is an IDP that carries out a vital function with regards to RNA Polymerase II (RNAPII) while in the termination complex[9,10]. RNAPII is responsible for the synthesis of mRNA and miRNA and consists of a folded multisubunit complex where the actual synthesis takes place and a highly dynamic C-terminal domain, CTD. During the elongation cycle RNAPII becomes increasingly phosphorylated. In order for RNAPII to be recycled it must be dephosphorylated, via FCP1[11,12]. RAP74 (RNA Polymerase II associating protein 74 kDa) mediates this dephosphorylation step by binding to the disordered C-terminal domain of FCP1, thus co-localizing the phosphatase with the RNAPII holoenzyme. Binding occurs along a hydrophobic groove in RAP74, formed between the second and third helices of the domain, which acts to stabilize amphipathic helix formation in FCP1[13,14]. Further stabilization occurs through complementary charges between the basic groove of RAP74 and the acidic helix of FCP1[15,16]. Despite knowledge of the bound state provided by a crystal structure[17], little is known about the coupling between the binding event and the folding of the C-terminal region of FCP1. Defining the energetics associated with both the binding interaction and the folding event is necessary for understanding what drives this binding transition, and isothermal titration calorimetry (ITC) provides an important tool for accomplishing this.

Defining the physical parameters that govern biomolecular interactions is key to understand the driving forces behind them. ITC defines the strength of biomolecular interactions and their specificity. In a single ITC experiment ΔH and ΔG can be directly determined and ΔS calculated, providing information on the binding affinity as well as whether a particular interaction is enthalpicly or entropicly driven[18,19]. Also, if a set of ITC experiments are run over a range of temperatures it is possible to derive the ΔC_P for an interaction. The ΔC_P helps to define the degree of hydrophobic surface burial associated with a protein-protein interaction[20,21]. Here we will show that utilizing ITC with the aid of small molecule osmolytes we can quantify the thermodynamic parameters and determine what drives the FCP1/RAP74 interaction.

Materials and Methods

Protein Generation

A recombinant FCP1 83aa (878-961) construct containing the C-terminal binding region, ctFCP1, and an 84aa (434-512) RAP74 containing the C-terminal region and an additional tryptophan mutated onto the N-terminus, with codon usage optimized for expression in *Escherichia coli* was cloned into the expression vector pET-47b (Novagen) and transformed into *E. coli* BL21 competent cells. The cells were grown in LB media until an optical density between 0.8-0.9 at 600nm was reached. ctFCP1 and RAP74 expression was induced using 0.5M IPTG followed by growth for an additional 3-4 hours and harvesting by sonication. Following centrifugation, the clarified cell lysates were run over a Ni-NTA column for purification of the His-tagged cterFCP1 and RAP74. The proteins were dialyzed in the presence of 3-Cprotease overnight at 4°C to remove the His-tag, leaving a non-native Gly-Pro-Gly tripeptide at the N-terminus of the protein. The His-tagged 3-C protease was removed by running the sample after dialysis over a Ni-NTA column resulting in highly pure cterFCP1. A sample was obtained by spin-concentration and buffer exchange in a VIVASPIN 20 centrifugal concentrator. The final buffer conditions for the sample were 50mM cacodylate, pH 6.5, 50mM KCl. Additionally, a synthetic 17 aa peptide of

FCP1 (944-961), purchased from Tufts University Core facility, was dissolved as lyophilized powder into final solution.

Isothermal Titration Calorimetry

ITC data was acquired on a MicroCal AutoITC 200 and VP-ITC 100. AutoITC samples were prepared by dissolving lyophilized RAP74 and FCP1 peptide in filtered 50mM cacodylate buffer (pH=6.5) and 50mM KCl. VP-ITC samples were prepared from FCP1 and RAP74 that had been buffer exchanged into 50mM Cacodylate buffer (pH=6.5) and 50mM KCl and then dialyzed overnight in that same buffer or buffer with 10% (v/v) TFE or 1M TMAO. The samples were degassed for 20 min prior to each experiment. The ITC measurements were acquired at a range of temperatures, 288-313K in increments of 5K. The AutoITC runs were set up where the concentration of RAP74 in the cell and the peptide in the syringe were 0.15mM and 1.5mM respectively. 2µL injections were carried out over 4s with a spacing of 300s between each injection. VP-ITC runs were set up where the concentration of FCP1 in the cell and RAP74 in the syringe were 0.02mM and 0.3mM respectively. 10µL injections were carried out over 20s with a spacing of 360s between each injection. All ITC data was fit to s single site binding model of n-identical sites described in Equations 1 and 2 below in MATLAB (MathWorks).

$$K = \frac{\theta}{(1-\theta)[X]} \qquad (4.1)$$

$$Q = nM_t \Delta H V_o \qquad (4.2)$$

Where K is the binding constant, θ is the fraction of sites occupied, [X] is the free ligand concentration, n is the number of sites, M_t is the bulk macromolecule concentration, ΔH is the molar heat of ligand binding, and V_o is the active cell volume.

Results and Discussion

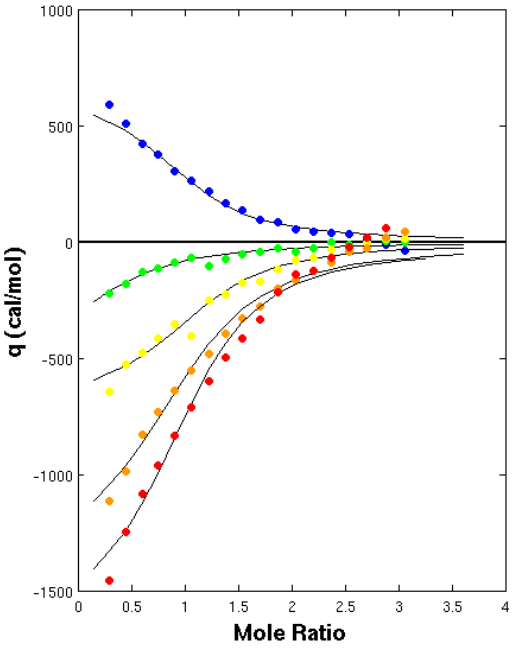

Figure 4.1: Above shows the ITC titrations of RAP74 with FCP1 peptide over a range of temperatures. The temperatures are 15(blue), 20(green), 25(yellow), 30(orange), and 35(red). Note that over the course of the study the interaction goes from endothermic to exothermic at around 20°C

Initial studies of the interaction between FCP1 and RAP74 included using both a synthetic peptide (944-961) of FCP1 and a construct containing the full C-terminal region (879-961). Initially, the peptide was studied to optimize conditions and eliminate any possible nonspecific interactions that FCP1 might have nearer to the N-terminus. Binding was monitored over a range of temperatures, 15-35°C. Over this range a clear endothermic to exothermic transition near room temperature was observed (Figure 4.1). Also, regardless of the nature of the binding enthalpy, the binding constants over the range of temperatures are all the same within experimental uncertainty (≈24μM; Table1). This analysis excludes fitting of the binding constant at 20°C because the event was accompanied by a ΔH too small to produce an accurate fit. A plot of the observed binding enthalpies against their respective temperatures reveals a large, negative change in heat capacity upon binding (ΔC_P = -160 ± 9 cal*mol^{-1}), Figure 4.2. Entropy-enthalpy

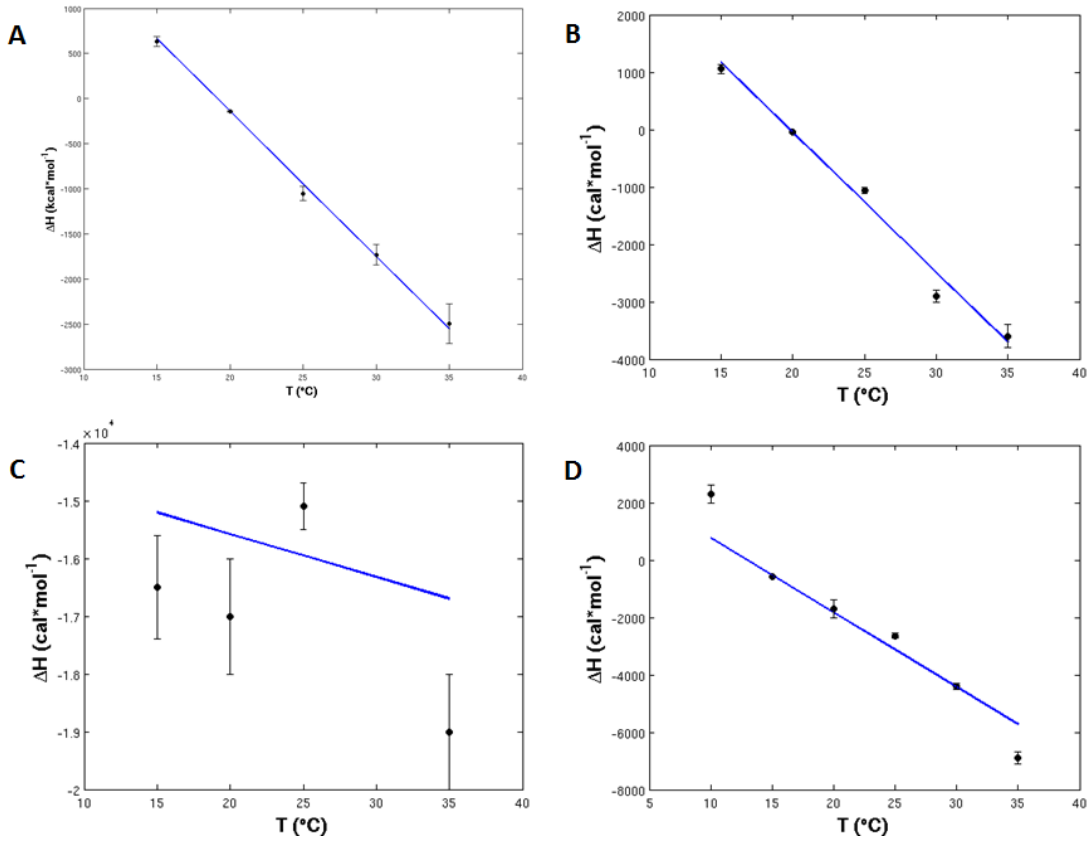

Figure 4.2: Plots of ΔH vs Temperature for each titration. The titrations run with the Fcp1 peptide (A) and ctFCP under aqueous conditions(B) and in the presence of 1M TAMO (D) show a large negative ΔC_P as indicated by the slope (peptide = -160 ± 9 cal/mol, ctFCP1(aqueous) = 227 ± 30cal/mol, ctFCP1(1M TMAO) = 260 ± 20 cal/mol).Note that the titration run in the presence of 10% TFE is characterized by a small ΔC_P (16 ± 100 cal/mol) indicating a loss of the burial of hydrophobic surface due to folding contribution to the enthalpy

compensation resulting in temperature independent binding constants combined with a large negative ΔC_P for the interaction shows that the binding thermodynamics for this interaction are dominated by the burial of hydrophobic surfaces[22,23].

Temperature-dependent titrations were repeated for the full C-terminal construct of FCP1. The same exothermic to endothermic trend is observed for the full construct (Figure 4.3) and the binding constants also show no significant temperature dependence (Table 4.1). However, the

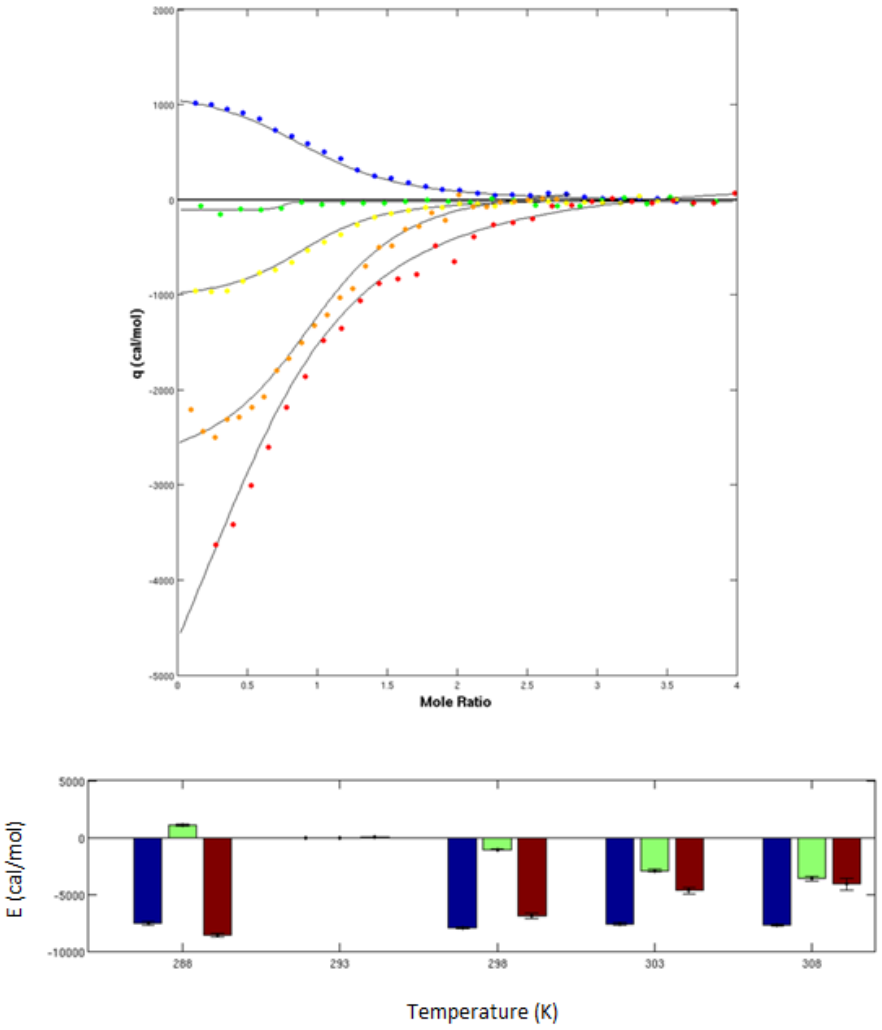

Figure 4.3: Above shows the ITC titrations of RAP74 with cterFCP1 over the same range of temperatures as the peptide (Top). The temperatures are 15(blue), 20(green), 25(yellow), 30(orange), and 35(red). Bar graphs depicting the ΔG (blue), ΔH (green), and -TΔS (red) for the interaction between FCP1 and RAP74 under aqueous conditions (Bottom). Note that the endothermic to exothermic transition is representative of the binding process transitioning from being entropically driven to enthalpically driven.

Table 4-1: A complete table of the dissociate constants and molar enthalpies for the binding of FCP1 to RAP74 under all solvent conditions and temperatures.

	Temperature (°C)	K_D (μM)	ΔH (kcal/mol)	n
Aqueous	15	2.5 ± 0.2	1.21 ± 0.07	1.01 ± 0.01
	20	0.8 ± 1.0	-0.08 ± 0.02	0.93 ± 0.01
	25	2.2 ± 0.1	-1.24 ± 0.02	0.99 ± 0.01
	30	2.3 ± 0.3	-2.94 ± 0.08	1.02 ± 0.02
	35	5.0 ± 0.7	-5.8 ± 0.4	0.96 ± 0.01
10% TFE	15	1.49 ± 0.07	-14.0 ± 0.2	0.94 ± 0.01
	20	2.29 ± 0.05	-16.4 ± 0.1	1.01 ± 0.01
	25	2.4 ± 0.1	-17.7 ± 0.3	0.89 ± 0.01
	30	3.7 ± 0.1	-16.5 ± 0.1	1.07± 0.05
	35	5.3 ± 0.3	-16.5 ± 0.3	0.89 ± 0.03
1M TMAO	10	1.0 ± 0.3	0.9 ± 0.04	0.92 ± 0.03
	15	0.3 ± 0.04	-0.6 ± 0.03	0.96 ± 0.05
	20	0.10 ± 0.04	-1.58 ± 0.07	1.08 ± 0.01
	25	0.34 ± 0.03	-3.24 ± 0.02	0.98 ± 0.01
	30	0.73 ± 0.02	-4.60 ± 0.01	0.93 ± 0.01
	35	0.80 ± 0.02	-6.4 ± 0.1	0.98 ± 0.01

observed K_D for this construct was an order of magnitude smaller, indicating tighter binding. This is most likely due to the extended N-terminal region, which has been previously shown to remain disordered while bound[10], and which could be influencing the binding thermodynamics by minimizing artifactual end effects. Also, the extended N-terminus for ctFCP1 most likely mediates additional contacts with RAP74 that stabilize the complex[24,25]. As with the peptide binding studies, ΔC_P was determined to be large and negative (-227 ± 30 cal*mol^{-1}), Figure 4.2, demonstrating that removal of hydrophobic surface from the aqueous solvent dominates the

thermodynamics of binding. The free energy and entropy for each temperature were calculated and compared with the directly observed enthalpies (Figure 4.3). From the bar graph, the binding interaction under aqueous conditions shows a transition from an entropically driven process to an enthalpically driven process as temperature increases. However, it could not be determined if the hydrophobic contribution is primarily due to the folding of FCP1 or removal of water from the FCP1/RAP74 binding interface; thus, the use of osmolytes was proposed.

As a probe to determine if the overall binding thermodynamics were dominated by protein folding, TFE, an osmolyte, was used. Historically, TFE has been used in protein folding studies to induce an α-helical structure on short peptides[26–29]. In our previous studies we have shown that TFE does induce a helical structure on ctFCP1 and that the transition is localized to the binding region[10]. Thus, we used TFE to isolate the folded form of FCP1 in the absence of RAP74. When ITC is run over the same temperature range as the previous experiments, the most obvious result is the loss of the endothermic to exothermic transition (Figure 4.4) and its replacement with a much smaller heat capacity (Figure 4.2), suggesting that folding of FCP1 may have been causative of the thermodynamic characteristics previously discussed for the pure aqueous case. Also, the binding constants still show relative temperature independence (Table 1). Based on the figure, one can qualitatively say that the binding enthalpies are relatively temperature independent; however, due to slight inconsistencies between TFE solutions, a reliable ΔC_P could not be determined. Because the binding constants still show temperature independence one can infer that the overall binding heat is being dominated by burial of protein surface. The same bar graph made for aqueous conditions was also made for the FCP1/RAP74 interaction in the presence of 10% TFE. The entropically to enthalpically driven process was not present in the 10% TFE, in fact, the entropy for the process over all temperatures was unfavorable.

While instructive, the analysis presented above does not reflect an unambiguous interpretation of the data. Without additional studies, it was not possible to definitively say if the

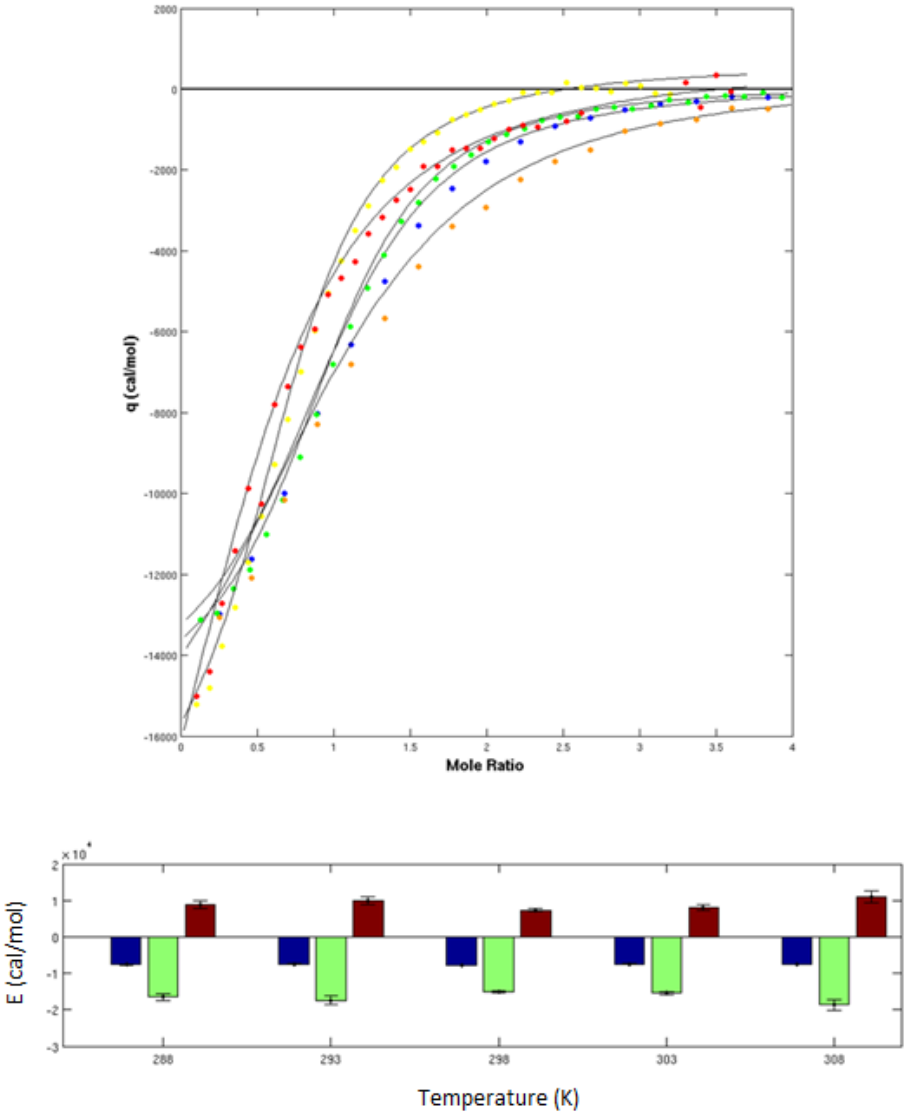

Figure 4.4 Above shows the ITC titrations of RAP74 with cterFCP1 in 10% TFE over the same range of temperatures as the peptide (TOP). The temperatures are 15(blue), 20(green), 25(yellow), 30(orange), and 35(red). Bar graphs depicting the ΔG (blue), ΔH (green), and -TΔS (red) for the interaction between FCP1 and RAP74 in the presence of 10% TFE (Bottom). Note that for 10% TFE both the endothermic to exothermic transition and the entropic to enthalpic transition are no longer present.

loss of the endothermic to exothermic transition is due to having only folded FCP1 or caused by solvent effects due to the presence of TFE. In order to ascertain if this is so, TMAO, an alternative osmolyte known to stabilize protein structure, replaced TFE in solution[30–32].

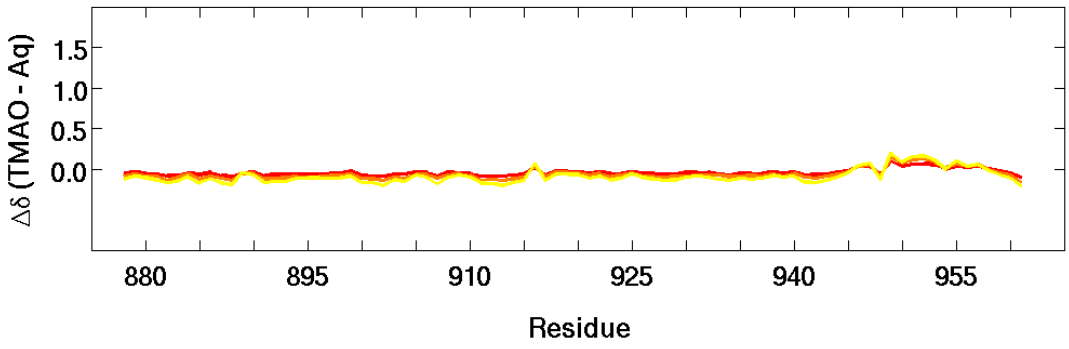

Figure 4.5: Above show the changes in chemical shift for FCP1 observed at 0.3M (red), 0.6M (orange), and 0.9M TMAO. Note that there are little to know changes in chemical shift indicating no change in secondary structure for FCP1.

An NMR study similar to that done for TFE in our previous work was carried with TMAO. Over TMAO concentrations of 0.3, 0.6, 0.9M the changes in chemical shift are insignificant for those residues outside the binding region and the chemical shift changes in the binding region are much smaller than those observed for TFE, Figure 4.5. These observations made us confident that at a 1.0M concentration FCP1 would exist in a state that more closely resembled unbound aqueous conditions. Thus, a full set of titration calorimetry experiments were carried out over a temperature range of 10-35°C. The change in the temperature range compared to all previous experiments was deemed necessary when looking at the overlay of all isotherms (Figure 4.6). The observed binding enthalpies while in the presence of TMAO became more negative as temperature increased following the trend found for FCP1 under aqueous conditions. However, the exothermic to endothermic transition was shifted to lower temperatures (Table 4.1) necessitating the expanded temperature range of this study. The presence of TMAO also preserved the entropic to enthalpic transition that drives the binding that, under 10% TFE conditions, the entopic penalty imposed by the FCP1/RAP74 binding interaction is due to the pre-folded helical structure in FCP1.

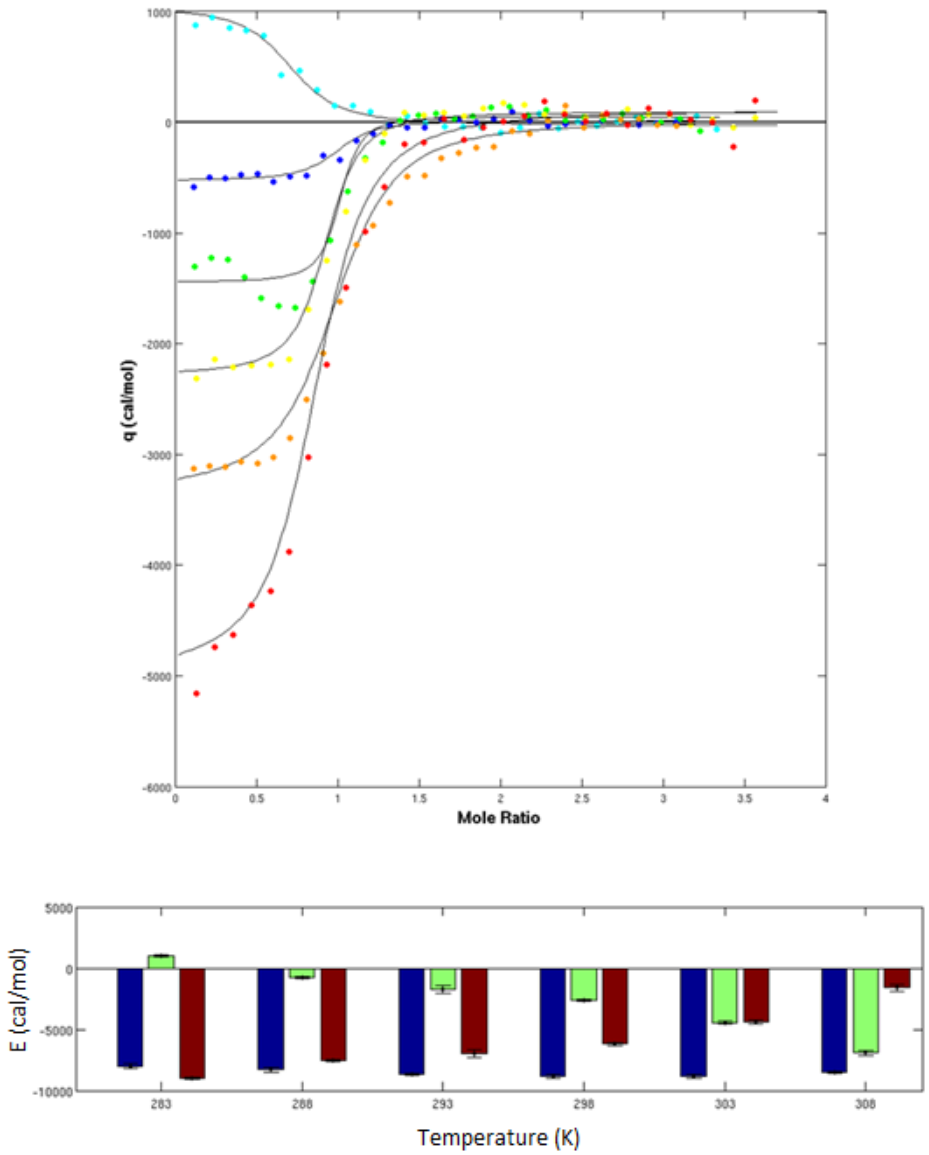

Figure 4.6 Above shows the ITC titrations of RAP74 with cterFCP1 with the addition of 1M TMAO over the same range of temperatures as the peptide. The temperatures are 10(cyan) 15(blue), 20(green), 25(yellow), 30(orange), and 35(red). Bar graphs depicting the ΔG (blue), ΔH (green), and -TΔS (red) for the interaction between FCP1 and RAP74 in the presence of 1M TMAO (Bottom). Note there is an endothermic to exothermic at transition, but unlike like in the aqueous samples the transition occurs at colder temperatures.

From the bar graphs, the binding interaction under aqueous conditions and 1M TMAO show a transition from an entropically driven process to an enthalpically driven process as temperature increases that is not present in the 10% TFE data set. It is clear that, over the

temperature ranges studied for aqueous and 1M TMAO conditions, the FCP1/RAP74 interaction is entropically favorable. This observation is contrary to the naïve assumption that IDP interactions are fundamentally entropically disfavored because of the loss of conformational freedom associated with binding. Also, this supports previous molecular dynamics simulation and NMR relaxation data from our lab that suggested interaction. This suggests that the FCP1 binding region still retains a dynamic nature even in complex. TFE conditions, on the other hand, show a thermodynamic profile where all three parameters are effectively the same over all temperatures. The most striking feature of each titration is that the entropy of interaction is unfavorable over the entire temperature range. This provides useful insight into how RAP74 recognizes FCP1. While in a disordered state under aqueous conditions at near biological temperature the binding interaction is driven both enthalpically and entropically. However, if FCP1 is in a folded state prior to binding RAP74 the interaction is enthalpically driven and entropically opposed. This suggests that initial contact with RAP74 takes place while FCP1 is in a mostly disordered state.

Conclusion

We have shown that the hydrophobic effect plays a dominate role in the process of FCP1 binding toRAP74. In order to determine whether or not this is due to FCP1 folding or simply the burial of hydrophobic residues that are associated with the binding site TFE was used to pre-fold FCP1. The addition of TFE is meant to force FCP1 to assume its folded form in the apo state so that the heats measured are purely the result of binding and desolvation not folding. We observed a significant change in the overall binding thermodynamics between the aqueous and TFE states. This was noted by the loss of an endothermic to exothermic transition over the experimental temperature range and by the overall interaction becoming entropically unfavorable over the entire temperature range. In order to determine whether or not the observed effect was due FCP1 having a preformed helix or solvent effects due to the addition of TFE, an alternative osmolyte was used, TMAO. Over the observed temperature range, we saw similar trends in the presence of

TMAO to those observed in the aqueous state, with the exception of the endothermic to exothermic transition being shifted to colder temperatures. From this we conclude that the conformational entropy of FCP1 plays a role in developing a favorable binding interaction. This observation follows other observations made in previous studies in our lab where it is predicted that disorder may even play a regulatory role after binding has already taken place. However, due to complex solvent effects associated with the binding interaction in the presence of TFE it is still difficult to tell whether this interaction is dominated by FCP1 folding into an α-helix or burial of the hydrophobic binding interface.

References

1. Dunker, A. K. *et al.* The unfoldomics decade: an update on intrinsically disordered proteins. *BMC Genomics* **9,** S1 (2008).

2. Gsponer, J. & Madan Babu, M. The rules of disorder or why disorder rules. *Prog. Biophys. Mol. Biol.* **99,** 94–103 (2009).

3. Tompa, P. *et al.* Close encounters of the third kind: disordered domains and the interactions of proteins. *BioEssays* **31,** 328–335 (2009).

4. Mao, A. H., Crick, S. L., Vitalis, A., Chicoine, C. L. & Pappu, R. V. Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. *Proc. Natl. Acad. Sci.* **107,** 8183–8188 (2010).

5. Liu, J., Faeder, J. R. & Camacho, C. J. Toward a quantitative theory of intrinsically disordered proteins and their function. *Proc. Natl. Acad. Sci.* (2009). doi:10.1073/pnas.0907710106

6. Csermely, P., Palotai, R. & Nussinov, R. Induced fit, conformational selection and independent dynamic segments: an extended view of binding events. *Trends Biochem. Sci.* **35,** 539–546 (2010).

7. Vogt, A. D. & Di Cera, E. Conformational Selection or Induced Fit? A Critical Appraisal of the Kinetic Mechanism. *Biochemistry (Mosc.)* **51,** 5894–5902 (2012).

8. Receveur-Bréchot, V., Bourhis, J.-M., Uversky, V. N., Canard, B. & Longhi, S. Assessing protein disorder and induced folding. *Proteins Struct. Funct. Bioinforma.* **62,** 24–45 (2006).

9. Ghosh, A., Shuman, S. & Lima, C. D. The Structure of Fcp1, an Essential RNA Polymerase II CTD Phosphatase. *Mol. Cell* **32,** 478–490 (2008).

10. Lawrence, C. W., Bonny, A. & Showalter, S. A. The disordered C-terminus of the RNA Polymerase II phosphatase FCP1 is partially helical in the unbound state. *Biochem. Biophys. Res. Commun.* **410,** 461–465 (2011).

11. Archambault, J. FCP1, the RAP74-Interacting Subunit of a Human Protein Phosphatase That Dephosphorylates the Carboxyl-terminal Domain of RNA Polymerase IIO. *J. Biol. Chem.* **273,** 27593–27601 (1998).

12. Chapman, R. D., Heidemann, M., Hintermair, C. & Eick, D. Molecular evolution of the RNA polymerase II CTD. *Trends Genet.* **24,** 289–296 (2008).

13. Abbott, K. L. *et al.* Interactions of the HIV-1 Tat and RAP74 Proteins with the RNA Polymerase II CTD Phosphatase FCP1 †. *Biochemistry (Mosc.)* **44,** 2716–2731 (2005).

14. Kamada, K. Crystal structure of the C-terminal domain of the RAP74 subunit of human transcription factor IIF. *Proc. Natl. Acad. Sci.* **98,** 3115–3120 (2001).

15. Nguyen, B. D. *et al.* NMR structure of a complex containing the TFIIF subunit RAP74 and the RNA polymerase II carboxyl-terminal domain phosphatase FCP1. *Proc. Natl. Acad. Sci.* **100,** 5688–5693 (2003).

16. Nguyen, B. D. *et al.* Solution structure of the carboxyl-terminal domain of RAP74 and NMR characterization of the FCP1-binding sites of RAP74 and human TFIIB. *Biochemistry (Mosc.)* **42,** 1460–1469 (2003).

17. Kamada, K., Roeder, R. G. & Burley, S. K. Molecular mechanism of recruitment of TFIIF- associating RNA polymerase C-terminal domain phosphatase (FCP1) by transcription factor IIF. *Proc. Natl. Acad. Sci.* **100,** 2296–2299 (2003).

18. Leavitt, S. & Freire, E. Direct measurement of protein binding energetics by isothermal titration calorimetry. *Curr. Opin. Struct. Biol.* **11,** 560–566 (2001).

19. Doyle, M. L. Characterization of binding interactions by isothermal titration calorimetry. *Curr. Opin. Biotechnol.* **8,** 31–35 (1997).

20. Jelesarov, I. & Bosshard, H. R. Isothermal titration calorimetry and differential scanning calorimetry as complementary tools to investigate the energetics of biomolecular recognition. *J. Mol. Recognit.* **12,** 3–18 (1999).

21. Ladbury, J. E. & Chowdhry, B. Z. Sensing the heat: the application of isothermal titration calorimetry to thermodynamic studies of biomolecular interactions. *Chem. Biol.* **3,** 791–801 (1996).

22. Dill, K. A. Dominant forces in protein folding. *Biochemistry (Mosc.)* **29,** 7133–7155 (1990).

23. Southall, N. T., Dill, K. A. & Haymet, A. D. J. A View of the Hydrophobic Effect. *J. Phys. Chem. B* **106,** 521–533 (2002).

24. Lawrence, C. W. & Showalter, S. A. Carbon-Detected 15N NMR Spin Relaxation of an Intrinsically Disordered Protein: FCP1 Dynamics Unbound and in Complex with RAP74. *J. Phys. Chem. Lett.* **3,** 1409–1413 (2012).

25. Wostenberg, C., Kumar, S., Noid, W. G. & Showalter, S. A. Atomistic simulations reveal structural disorder in the RAP74-FCP1 complex. *J. Phys. Chem. B* **115,** 13731–13739 (2011).

26. Sonnichsen, F. D., Van Eyk, J. E., Hodges, R. S. & Sykes, B. D. Effect of trifluoroethanol on protein secondary structure: an NMR and CD study using a synthetic actin peptide. *Biochemistry (Mosc.)* **31,** 8790–8798 (1992).

27. Gast, K., Siemer, A., Zirwer, D. & Damaschun, G. Fluoroalcohol-induced structural changes of proteins: some aspects of cosolvent-protein interactions. *Eur. Biophys. J.* **30,** 273–283 (2001).

28. Li, W., Qin, M., Tie, Z. & Wang, W. Effects of solvents on the intrinsic propensity of peptide backbone conformations. *Phys. Rev. E* **84,** (2011).

29. Chitra, R. & Smith, P. E. Molecular Association in Solution: A Kirkwood−Buff Analysis of Sodium Chloride, Ammonium Sulfate, Guanidinium Chloride, Urea, and 2,2,2-Trifluoroethanol in Water. *J. Phys. Chem. B* **106,** 1491–1500 (2002).

30. Baskakov, I. V. *et al.* Trimethylamine N-Oxide-induced Cooperative Folding of an Intrinsically Unfolded Transcription-activating Fragment of Human Glucocorticoid Receptor. *J. Biol. Chem.* **274,** 10693–10696 (1999).

31. Chang, Y.-C. & Oas, T. G. Osmolyte-Induced Folding of an Intrinsically Disordered Protein: Folding Mechanism in the Absence of Ligand. *Biochemistry (Mosc.)* **49,** 5086–5096 (2010).

32. Baskakov, I. & Bolen, D. W. Forcing thermodynamically unfolded proteins to fold. *J. Biol. Chem.* **273,** 4831–4834 (1998).

Chapter 5 A Comparative Study of Ensemble Modeling Methods to Construct an Accurate Ensemble Structure of the IDP FCP1

The residual dipolar coupling (RDC) data set used in this study was acquired and analyzed by Dr. Debashish Sahu

Abstract

Intrinsically disordered proteins (IDPs) are a class of proteins that lack a unique 3D structure under native conditions. Due to their high flexibility, IDPs do not form a singular structure like well folded globular proteins, but instead exist as an ensemble of rapidly interconverting conformers. Typical experimental techniques that report on protein structure, such as NMR, only provide information of the ensemble average and not of the individual conformers that make up the ensemble structure. Therefore, to fully understand the behavior of IDPs a molecular description is needed, and thus, we must look at ensemble modeling methods that are designed to generate an ensemble structure composed of multiple conformers that reproduce experimental results for the ensemble as a whole. Here we will report on ensembles generated using the ensemble modeling methods EOM, flexible-meccano, and ENSEMBLE. EOM and ENSEMBLE refine ensemble structures using experimental constraints, such as those from SAXS and NMR while flexible-meccano uses a random coil model approach, derived from the primary sequence. We will show that, given the experimental constraints available, a random coil model which has been adjusted to take into account the structural propensities of our model system, FCP1, is sufficient to reproduce experimental results. Future experimentation for generating more refined ensemble structures is also explored.

Introduction

In recent years the study into intrinsically disordered proteins (IDPs) has increased substantially as more and more of the proteome is found to be either fully or at least partially disordered[1,2]. IDPs do not adopt a unique 3D fold in their native state and as such exist outside the scope of traditional structural biology[3,4]. These proteins challenge the structure/function paradigm as they do have biologically relevant functions regardless of their lack of regular structure. IDPs are known to function in transcription, and translation, and cellular signaling and have been implicated in several neurological disorders[4–7], thus, there is a need to understand the behavior of these proteins. In order to do so and develop treatments of related diseases a molecular description is needed. One that is capable of assessing what role the high degree of flexibility found in IDPs plays.

The highly dynamic nature of IDPs requires an alternate description of their native state that goes beyond the static structures associated with well folded proteins. A single structure would only provide a description of an isolated sub-state that may or may not be biologically relevant. Instead IDPs must be treated as an ensemble state where the protein is rapidly interconverting between multiple conformations on a rugged potential energy landscape[8–11]. A description of the ensemble state of an IDP then becomes more about defining what properties govern the behavior of the ensemble and determine what areas in the conformational space will be more likely sampled[12]. Previously our lab has shown that NMR can work as a useful tool for extracting secondary structural information and local backbone dynamics for IDPs[13,14]. However, this is on an atomic scale and describes the average conformational state of the protein. It does not necessarily provide a description on the amount of space occupied by the ensemble state or what specific conformations could make up that ensemble state.

Small angle X-ray scattering (SAXS) is a powerful technique for structural characterization for both well folded and disordered biological systems. SAXS provides

information of the overall size and shape of a biomolecule in aqueous solution at a resolution of a few nm[15–17]. The technique has been used to study partially or completely disordered proteins as well as multi-domain proteins that have flexible linker to great effect to qualitatively study the their flexibility[11,18,19]. This high degree of flexibility IDPs can sample a large number of conformations and the scattering profile becomes representative of the average of these conformers. This property results in unique trends in the SAXS data of IDPs. The individual conformers generate scattering profiles with various features that describe its structure, but IDPs consist of a large pool of structures the scattering profile typically averages out to a very smooth and relatively featureless scattering curve[20]. Also, the radius of gyration (R_g), which can be obtained directly from a scattering curve using the Guinier approximation[16,21], is typically larger than that of a well folded protein of a similar molecular mass[19,20]. R_g is the most common descriptor of the size of molecules in solution[16,17] and is a useful means to determine whether or not a protein is disordered. SAXS has been used construct low resolution structures of IDPs; however, these are designed to be implemented on rigid body systems and not highly dynamic proteins like IDPs. The resulting 3D reconstructions show highly elongated shapes, which have become an indicator of large scale motion or flexibility[19,20]. This single conformation approach does not necessarily describe a highly dynamic system well, but it does aide in visualizing the ensemble state of IDPs.

Ensemble modeling has become a common approach for structurally characterizing ensemble structures of IDPs[10,22–24]. Many labs have approached this idea differently, but ultimately the process is the same: generate a set of structures that is consistent with experimental data. The structures that are generated this way should not be capable of reproducing all experimental constraints, but, instead, collectively as an ensemble should report back the same values[22,25]. As stated previously, there are multiple strategies to an ensemble modeling approach. One such strategy is to carry out molecular dynamics (MD) simulations where multiple

simultaneous trajectories are informed by experimental data[24,26]. An alternative strategy is to use a statistical coil model to generate a larege pool of conformers, which has been found to be fairly good at reproducing experimental values for some IDPs[11,27]. Another approach is to generate a set of structures and then apply population weights to each structure to report back experimental results[28,29]. Here we will use a combination of SAXS and multiple ensemble modeling strategies (EOM, Flexible-meccano, and ENSEMBLE) in order to evaluate these methods and generate a viable ensemble model of the IDP FCP1.

Materials and Methods

Sample Generation

A synthetic 30 aa peptide of FCP1 (932-961), purchased as a lyophilized powder from Tufts University Core facility, was dissolved into each of several final solutions. The buffer used for all samples was 50mM MOPS at a pH of 6.5 with a KCl concentration of 0, 50, and 100mM. Other solution conditions included 50mM MOPS at a pH of 6.5 and 50mMKCl in the presence of either 10%TFE or 2M urea. The protein concentration was 5, 10, 20, and 30 mg/mL while in the presence of 50mM MOPS and 50mM KCl and 5 mg/mL for all other solution conditions.

SAXS

Data was collected at the Cornell High Energy Synchrotron Source (CHESS) facility from beam line F2 and all measurements were made using a Pilatus 100K-S detector. Measurements of lysozyme scattering were used for absolute scale normalization. Initial data analysis was carried out using RAW where the R_g at each solvent was calculated using the Guinier approximation. 3D model reconstruction was accomplished using the ATSAS software package described in the following statements. Inverse Fourier transformation of the scattering profile was accomplished through GNOM[30]. DAMMIF[31] was then used to generate a set of 3D

reconstructions based on the output of GNOM which were then averaged into a single structure using DAMAVER[32].

Ensemble Modeling

Ensembles were generated using EOM, Flexible-meccano, and ENSEMBLE. EOM ensembles were generated using the SAXS scatting curve of FCP1$^{932\text{-}961}$ in 50mM MOPS at pH 6.5 and 50mM KCl solvent conditions. An initial pool was generated based on the primary sequence and a genetic algorithm was applied ten times to select a subset of conformers after 100 generations that best describe the experimental SAXS data. Flexible-meccano ensembles were generated after a pool of 10000 structures were generated directly from the primary sequence with helical propensity of 0, 50, and 100% for residues 944-961 being explicitly defined.

Results and Discussion

Building a SAXS informed Ensemble Model using EOM

Initial experiments used 30 mg/mL in order to provide maximum scattering signal, which was shown to be monodisperse by DLS (data not shown). Data at this concentration, after analysis, generated a predicted molecular weight consistent with a trimer; which was further evidenced by the constructed enveloped, figure 5.1C. Additional concentrations, 10 and 5mg/mL, were used to determine ideal solvent conditions. 10mg/mL generated a molecular weight and envelope consistent with a dimer, figure 5.1B, and a monomeric state was achieved at 5mg/mL, figure 5.1A. Once a monodisperse monomer state was observed, 5 mg/mL was used as the concentration for the duration of data acquisition.

In order to determine the effect different solvent conditions have on the ensemble structure of FCP1 a range of salt concentrations (0, 50, and 100mM KCl), 10% TFE, and 2M urea were used. Different salt conditions were used in attempt to change the compactness of FCP1's

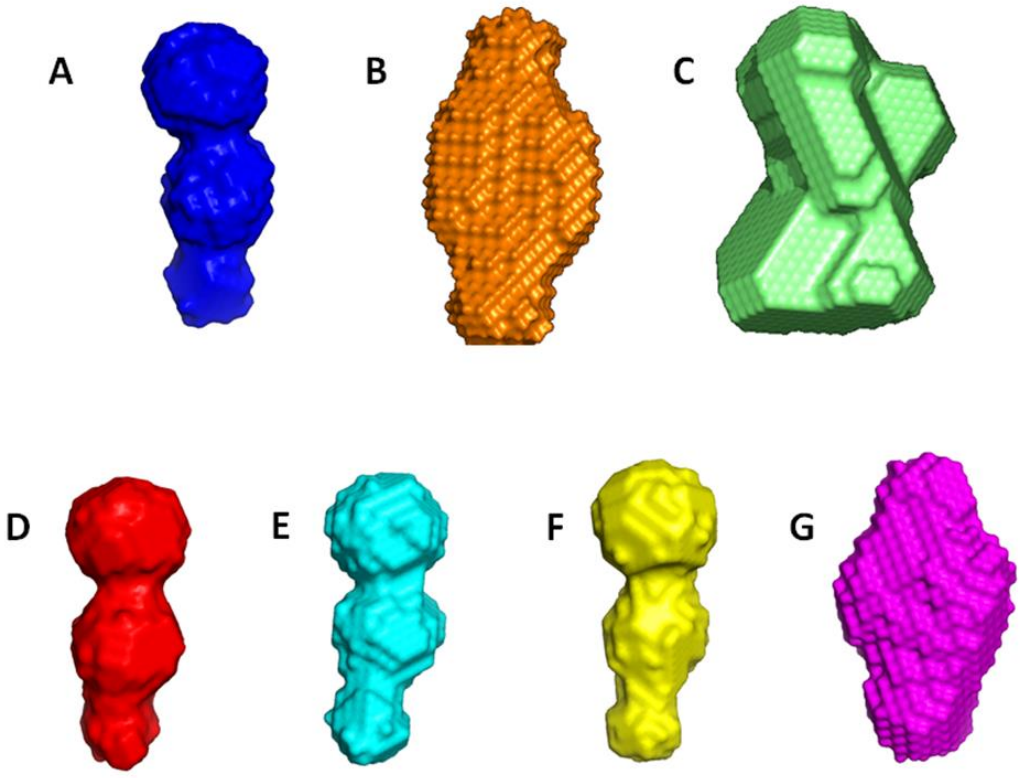

Figure 5.1: Above shows the SAXS 3D reconstructions of FCP1$^{932-961}$ at all concentrations and solution conditions: A) 5mg/mL FCP1 in 50mM KCL, B) 10mg/mL FCP1 in 50mM KCL, C) 30mg/mL FCP1 in 50mM KCL, D) 5mg/mL FCP1 in 0mM KCL, E) 5mg/mL FCP1 in 100mM KCL, A) 5mg/mL FCP1 in the presence of 2M urea, and A) 5mg/mL FCP1 in the presence of 10% TFE. Notice that as concentration increases there oligomerization occurs generating structures consistent with a dimer and trimer respectively. Also, Note that changes in salt conditions as well as the presence of 2M urea have no effect on the overall shape of the ensemble with 10% TFE being the only exception, showing reacquisition of a dimer.

ensemble state. Since FCP1 is a highly charged protein it was predicted that under low salt conditions it would collapse upon itself and become more compact while under high salt conditions it would become more extended. TFE and urea were chosen as cosolutes because of our labs previous successes in using TFE to stabilize secondary structure in FCP1 and using urea to melt any transient structural elements. With the exception of the 10% TFE sample, all conditions yielded constructed envelopes with the same overall shape, as shown in figure 5.1D-G. The SAXS data acquired with 5mg/mL FCP1 in 10% TFE solution fits to a shape and molecular

weight consistent with a dimer even though no oligimerization is observed at this concentration in the other solution conditions. One explanation is TFE stabilizes the formation of secondary structure to such a degree that it encourages the formation of dimers even under more dilute conditions. Folding may bring together several long chain aliphatic residues into a hydrophobic patch and dimer formation helps to build a hydrophobic core. Urea on the other hand, shows no overall change in shape compared to FCP1 under inorganic salt conditions. How these solvent effects help to describe the ensemble structure of FCP1 in solution may be better understood with the use of a computational model.

The SAXS data provides an excellent visual of how the FCP1 ensemble is behaving under various solvent conditions but it doesn't provide any information about the individual conformations that compose the ensemble. There are multiple methods available to generate ensembles of conformations to describe IDPs, which may or may not utilize a refinement process informed by experimental data. The Ensemble Optimization Method (EOM) is one such method that has been developed to generate an ensemble, based on experimental SAXS data[15,20]. EOM generates a large random pool of structures built directly from the primary sequence to describe the maximum flexibility of the IDP. The large initial pool of conformers is refined into a subensemble by a genetic algorithm[33] based on calculated scattering profiles of each conformer[34]. Here EOM was used to create a subensemble of 20 conformers from an initial pool of 1000 that have been selected to be the minimum number of conformers that describe the ensemble properties of FCP1 as seen through SAXS, shown in figure 5.2A. The properties of this ensemble were then compared to one generated through MD simulations of the same construct as a reference, figure 5.3. The R_g for the ensemble generated using EOM, 12.7Å, is similar to that for the MD simulation, 12.2Å, however, both show an ensemble structure that is much more compact than the R_g that is directly calculated from the experimental data, 15.8Å. The smaller R_g the EOM

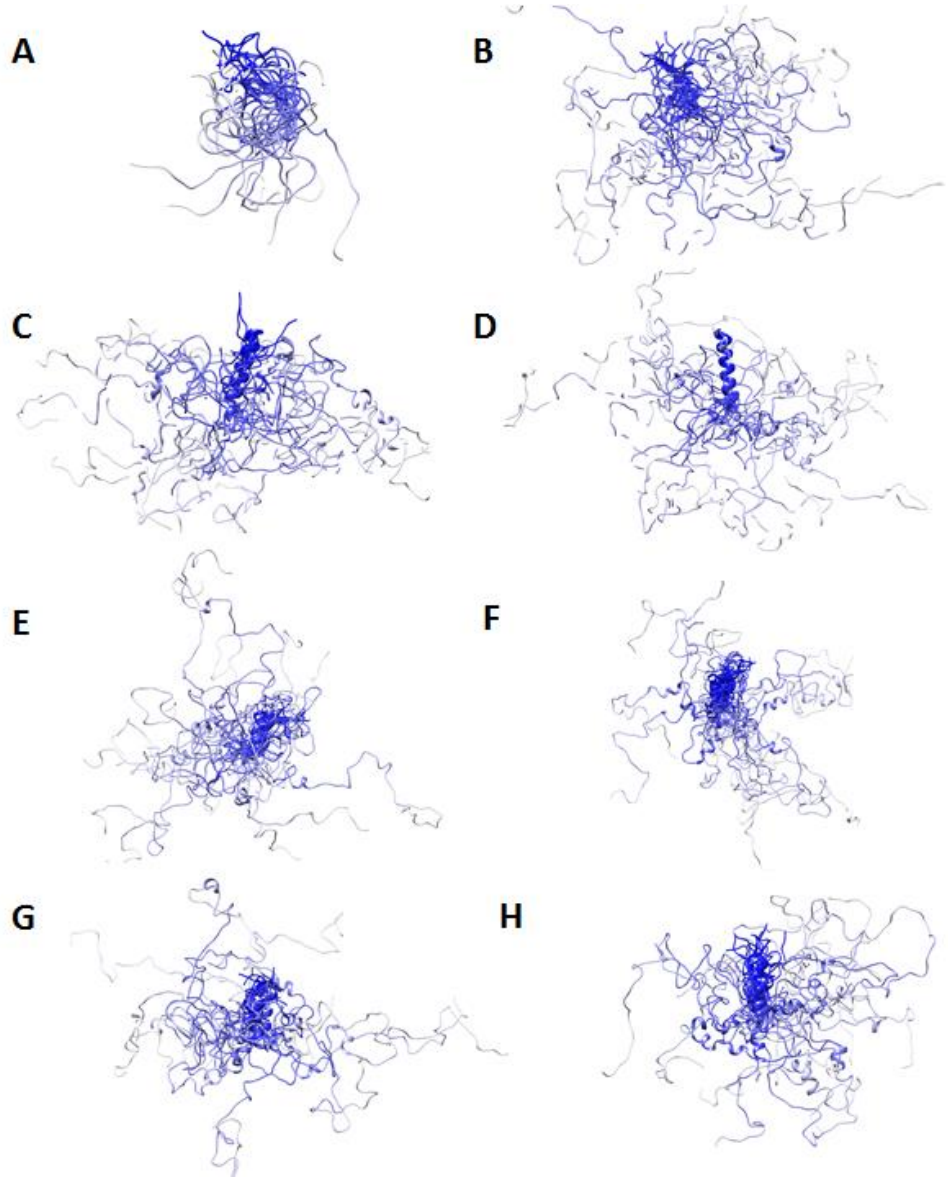

Figure 5.2: Above are the ensemble structures of FCP1 generated through the ensemble modeling methods discussed in this chapter. The EOM model (A), which uses the FCP1$^{932-961}$ construct, is very compact compared to the other ensemble structures generated. This is most likely due to the construct being composed of mostly of the RAP74 binding domain, which has a higher propensity to fold than other regions of ctFCP1. The ensembles generated using flexible-mecanno were defined by a 0% (B), 50% (C), and 100% (D) α-helical propensity are shown above. Note that regardless of the amount of structure assumed the disordered N-terminus is unaffected and samples the same amount of conformational space. Lastly, the ensembles generated with ENSEMBLE used no experimental restrains (E), chemical shifts (F), chemical shifts and R_2 relaxation rates (G), and chemical shifts; R_2 relaxation rates; and RDCs (H). For the ENSEMBLE structures there is relatively no visible difference for the binding region upon the addition of more constrains suggesting a random coil model is sufficient to describe the structure of FCP1.

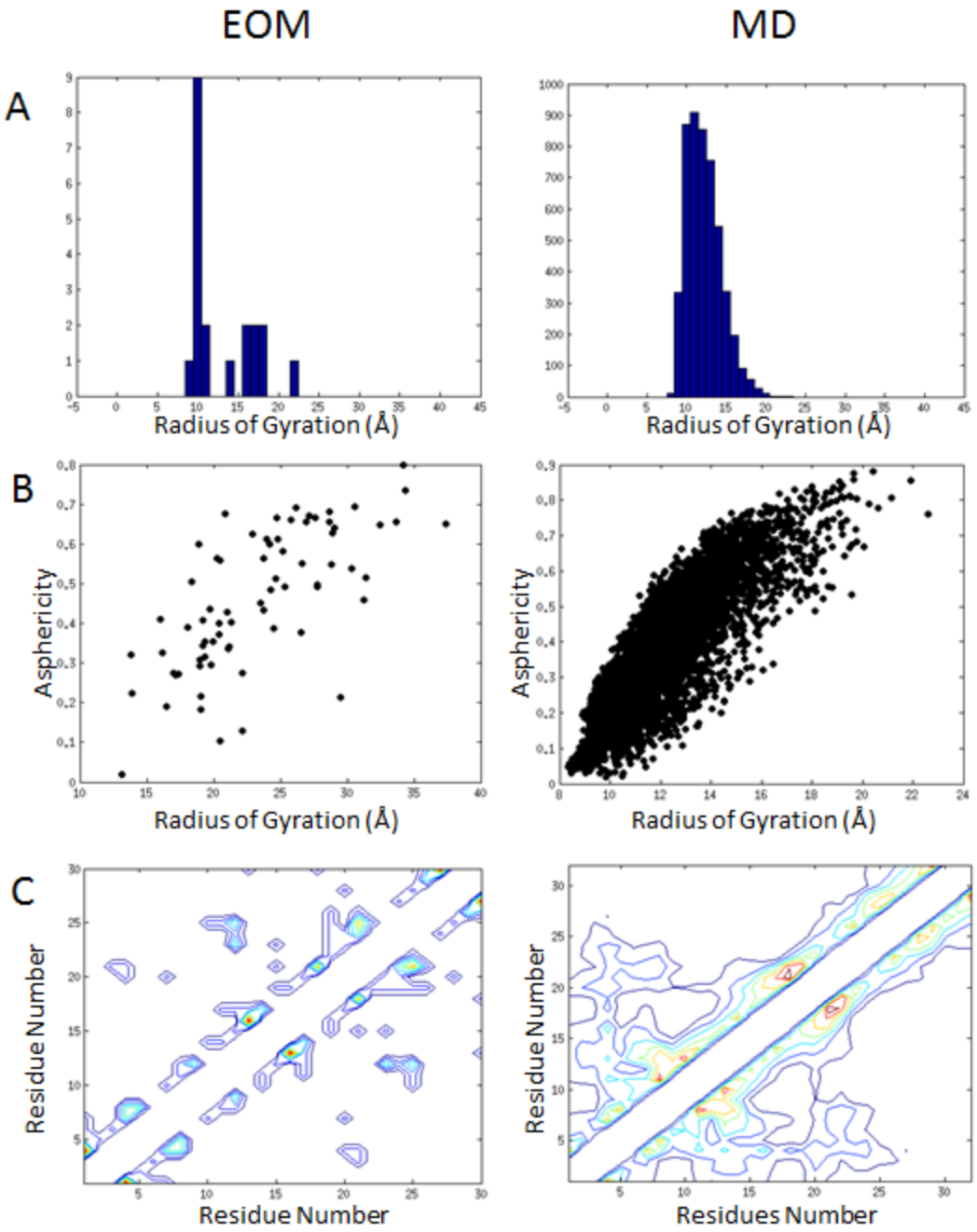

Figure 5.3: The R_g distributions (A), asphericity vs R_g plots (B), and contact maps (C) for the ensembles generated using EOM and MD simulations. The EOM conformers show an average R_g that is similar to those conformers generated using MD and from the asphericity vs R_g plots the EOM ensemble explores the same conformational space as the simulated ensemble. Note that while looking at the contact maps for each ensemble, the EOM ensemble does sample some helical structure, but the MD ensemble shows significantly more helical content.

structure may be due to a selection bias towards more collapsed structures based on the assumptions made for generating the initial pool. The difference in R_g aside, comparing plots of asphericity vs R_g for the EOM and MD ensemble shows that the overall size and shape of the conformations selected in EOM fall within the space sampled by the MD trajectories. This shows that while ensembles generated were more compact that the experimental data, these two methods, which generated structures by completely different means, were able to explore the same conformational space.

Comparing the Popular Ensemble Modeling Methods Flexible-Meccano and ENSEMBLE

Two popular alternate methods for generating IDP ensembles are flexible-meccano[35,36] and ENSEMBLE[25,37] which take different approaches to ensemble generation. Flexible-meccano builds multiple different copies of the same peptide chain by randomly sampling backbone dihedral angles through a highly efficient minimization algorithm[35]. The technique uses a simple model that makes structural ensemble generation highly efficient. Amino acid specific hard spheres are used to avoid steric clashes, no attractive forces are expressly used, and only particular backbone conformations are considered to avoid distortions due to additional potential energy terms[28,35]. Even though this statistical coil model lacks experimental constraints and is capable of predicting multiple experimental parameters (chemical shifts, RDCs, and small angle scattering curves) the user may replace or add to the statistical coil potentials and set specific population weights or conformational propensities[35]. Here we have used flexible-meccano to generate ensembles of ctFCP1 from the primary sequence using predetermined propensities for a-helix in the binding region (0%, 50%, and 100%). ENSEMBLE is a software package that is designed to use multiple experimental constraints to generate IDP ensembles. The initial step in the current scheme that describes the selection process for ENSEMBLE involves the creation of a large initial pool of random structures using a statistical coil model[22,25]. Then a Monte Carlo algorithm is used to select structures that are consistent with the experimental restraints and these

structures are then used as the basis for the generation of a new set of structures[25]. This process is repeated until a final ensemble is selected that is consistent with all experimental constraints.

Ensembles generated with these two methods are meant to be used as a comparison between the two as well as a means to determine which experimental parameters are essential to accurately describe the ensemble structure of FCP1. The ensembles generated in flexible-meccano are shown in figure 5.2B-D where they showed an average R_g that remained relatively unchanged regardless of the amount of helicity present (0% α-helix: 24.9Å, 50% α-helix: 24.8Å, 100% α-helix: 24.4Å), figure 5.4A. These values were similar to an experimentally determined R_g using SAXS, 24Å. As evident in the plots of asphericity vs R_g shown in figure 5.4B, the shape of each ensemble also remains unaffected by the addition of structure to the binding region. While it may appear that the conformers that compose these ensembles are very similar, the contact maps for each ensemble there clearly are different amounts of helical structure present, figure 5.4C. This is not necessarily to be unexpected. The previously shown SAXS reconstructions of $FCP1^{932-961}$ in the presence of 2M urea showed little effect. This may be because that even though ctFCP1 is known to at least transiently adopt helical structure in the unbound state, the amount structure acquired is small in relation to the rest of the protein It is likely that given the size of the binding region compared to the rest of the protein that remains disordered that regardless of the amount of order that is acquired in the binding region the overall ensemble remains unaffected.

Those ensembles generated in ENSEMBLE used a variety of NMR constraints including chemical shifts, R2 relaxation rates, and RDCs. Ensembles were made using adding each of these constraints one at time to determine the how much these constraints affect the overall ensemble to determine which are absolutely necessary to generate an accurate ensemble. Again, comparing the R_g values for each ensemble they remain relatively unaffected except for the ensemble that used no experimental constraints, shown in figure 5.5A. Once chemical shifts were incorporated as a constraint, the ensemble becomes more extended than the random coil model. Also, when

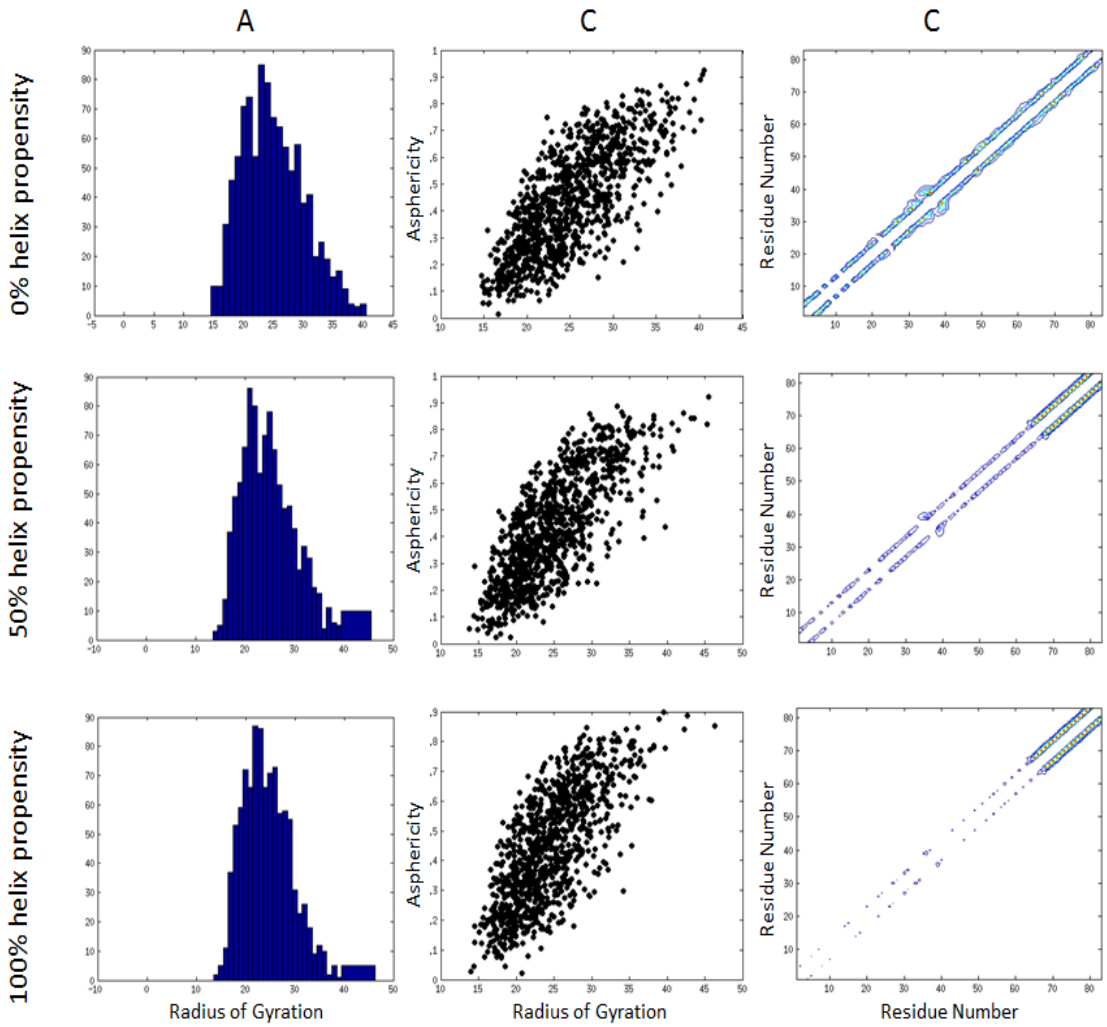

Figure 5.4: The R_g distributions (A), asphericity vs R_g plots (B), and contact maps (C) for ensembles generated using flexible-meccano with 0, 50, and 100% helical propensities. The average R_g for each ensemble does not change with increases in structure, but there is a clear change in the R_g distribution with acquisition of more and more structure. The aspericity vs R_g plots show that the overall size and shape of each is the same regardless of the amount of structure present. The contact maps show the increase in helical content with the increase of predetermined helical propensity, as to be expected.

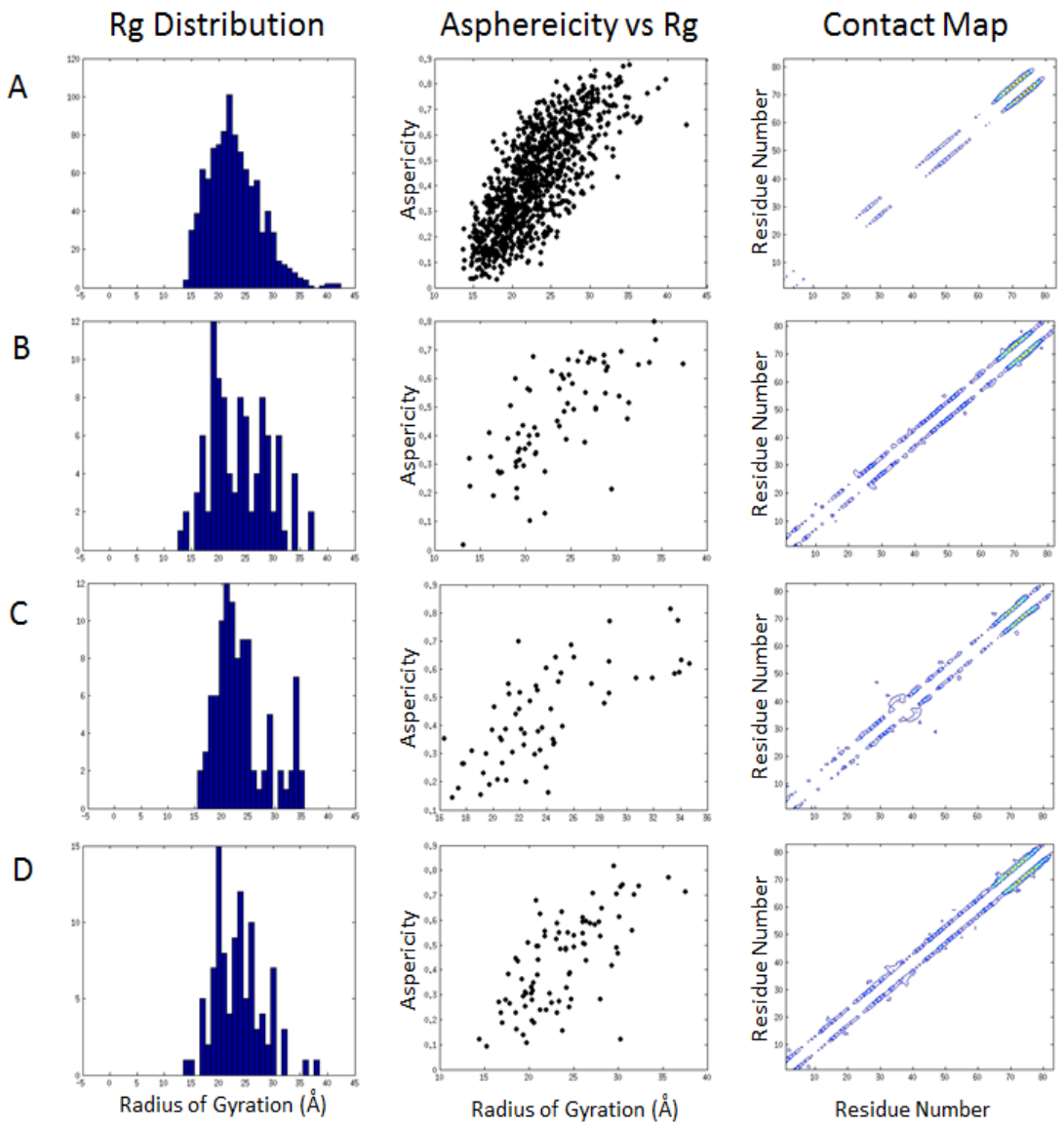

Figure 5.5: The R_g distributions, asphericity vs R_g plots, and contact maps for ensembles generated using no experimental restraints (A), chemical shifts (B), chemical shifts and R_2 relaxation rates (C), and chemical shifts; R_2 relaxation rates; and RDCs in ENSEMBLE. Above there are observed changes in the R_g distribution upon the addition of each experimental constraint, but the average R_g remains unchanged. The aspericity vs R_g plots show changes in the overall shapes of the ensembles between those that incorporate chemical shifts alone, R_2 rates, and RDCs. However, these all lay within the conformational space explored by the random coil model. Note that in the contact maps the random coil model is just as likely to predict

looking at the R_g vs ashpericity plots in figure 5.5B there are small changes in the aspericity for those ensembles that use experimental constraints, however, they all lie in the within the

conformational space that is occupied by the random coil model. The chemical shifts for each conformer within a given ensemble were simulated using SHIFTX[38] and compared to experimental chemical shifts, figure 5.6. The chemical shifts for each ensemble were all equally similar to the experimental data regardless of the amount of constraints used. The chemical shifts for flexible-meccano were also simulated the same way. Unlike ENSEMBLE, flexible-meccano did show a larger amount of variance from experimental results. Particularly, the ensembles generated with either a helical propensity of 0 or 100% showed marked differences from experimental results in the binding region. This indicated that a 50% helical propensity is a more accurate ensemble model.

Looking at all the data that were collected some inferences can be made about the ensemble structure of FCP1. Since the SAXS data were of a construct that contained mostly the binding region of FCP1, which becomes helical upon binding, it follows that the ensemble structures predicted by EOM have an R_g more compact than what was experimentally determined. The ensembles generated from flexible-meccano and ENSEMBLE, however, provide some insight how the disordered N-terminus affects the ensemble structure of the full C-terminal domain. Using both software packages, adding constraints meant to relate structural propensity do show increasing amounts of helical content that is clearly visible in the contact maps. Nonetheless, regardless of the amount of constraints imposed upon the ensemble structure, the overall size and shape does not change. This implies that given that the region of FCP1 which gains structure is small compared to the rest of the protein and highly localized the completely disordered N-terminus dominates the overall ensemble structure.

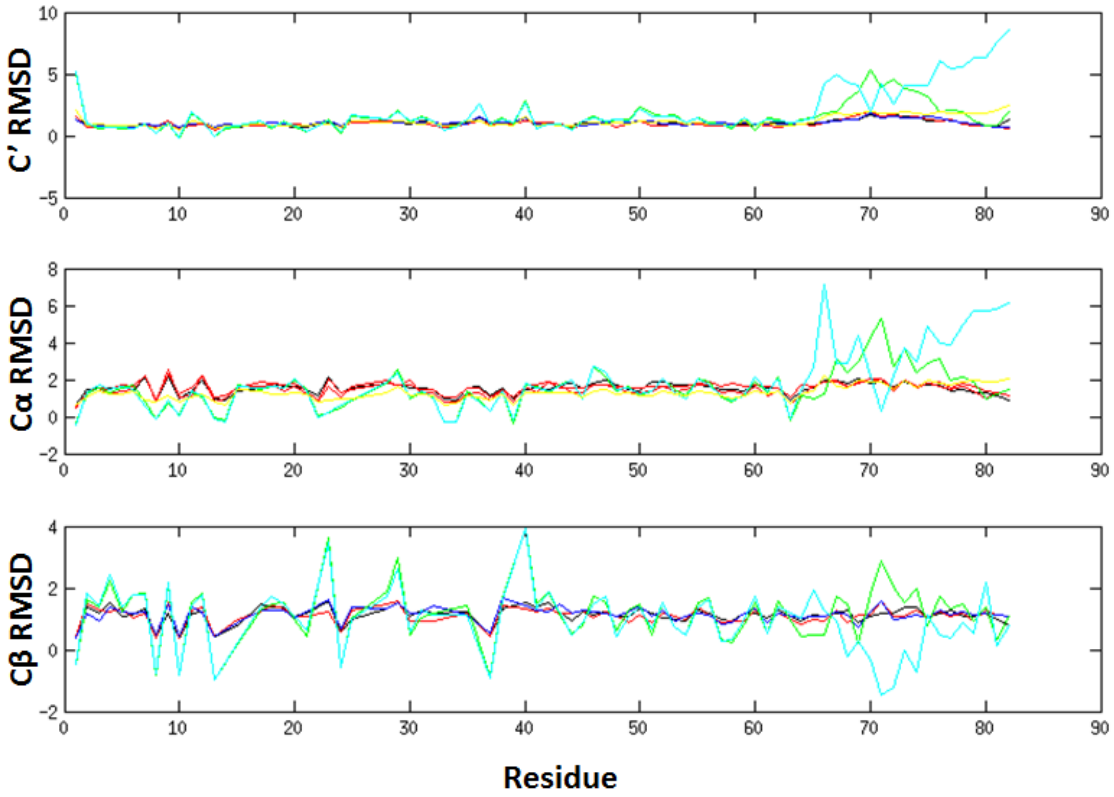

Figure 5.6: Above shows the root means square deviation (RMSD) of the simulated chemical shifts to experimental shifts for those ensembles generated by flexible meccano (0% helix: green, 50% helix: yellow, 100% helix: cyan) and ENSEMBLE (chemical shifts: black, chemical shifts + R_2 rates: red, chemical shifts + R_2 rates + RDCs: blue). Top shows the RMSD for C', middle shows the RMSD for C^{α}, and bottom shows the RMSD for C^{β}. Note that all ENSEMBLE ensembles reproduce experimental chemical shifts equally well, while only the flexible-meccano ensemble with 50% helix propensity has similarly low RMSD values.

Future Directions

From the data presented above it appears that the less computationally expensive flexible-meccano might be just as good at generating ensembles as ENSEMBLE, as long the correct amount of structural propensity is taken into account. However, more experimental data is necessary to say this for sure. For example, PRE (para magnetic resonance enhancement) and intramolecular NOE constraints need to be measured. This will provide long and medium range intraresidue distance information which will more fully refine the size and shape of the ctFCP1 ensemble. So far, modeling has only been done from the perspective of FCP1 while still free in

solution and not while it is complex with its binding partner, RAP74. While it is known that the C-terminal region of FCP1 does adopt a helical conformation in the complex it is not known, however, what affect this may have on the overall size and shape of the rest of the protein and what role that may have, if any, on binding. Thus, similar experimentation must also be carried out on FCP1 while in complex with RAP74 so that changes in the ensemble structure upon binding can be observed.

References

1. Tompa, P. Intrinsically unstructured proteins. *Trends Biochem. Sci.* **27,** 527–533 (2002).

2. Tompa, P. *et al.* Close encounters of the third kind: disordered domains and the interactions of proteins. *BioEssays* **31,** 328–335 (2009).

3. Dunker, A. K. *et al.* The unfoldomics decade: an update on intrinsically disordered proteins. *BMC Genomics* **9,** S1 (2008).

4. Gsponer, J. & Madan Babu, M. The rules of disorder or why disorder rules. *Prog. Biophys. Mol. Biol.* **99,** 94–103 (2009).

5. Dyson, H. J. & Wright, P. E. Intrinsically unstructured proteins and their functions. *Nat. Rev. Mol. Cell Biol.* **6,** 197–208 (2005).

6. Uversky, V. N. in *Protein Fold. Misfolding Neurodegener. Dis.* (Ovádi, J. & Orosz, F.) **7,** 21–75 (Springer Netherlands).

7. Uversky, V. N. Targeting intrinsically disordered proteins in neurodegenerative and protein dysfunction diseases: another illustration of the D2 concept. *Expert Rev. Proteomics* **7,** 543–564 (2010).

8. Mao, A. H., Crick, S. L., Vitalis, A., Chicoine, C. L. & Pappu, R. V. Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 8183–8188 (2010).

9. Kohn, J. E. *et al.* Random-coil behavior and the dimensions of chemically unfolded proteins. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 12491–12496 (2004).

10. Mittag, T. & Forman-Kay, J. D. Atomic-level characterization of disordered protein ensembles. *Curr. Opin. Struct. Biol.* **17,** 3–14 (2007).

11. Bernadó, P. *et al.* A structural model for unfolded proteins from residual dipolar couplings and small-angle x-ray scattering. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 17002–17007 (2005).

12. Schneider, R. *et al.* Towards a robust description of intrinsic protein disorder using nuclear magnetic resonance spectroscopy. *Mol. Biosyst.* **8,** 58–68 (2012).

13. Lawrence, C. W., Bonny, A. & Showalter, S. A. The disordered C-terminus of the RNA Polymerase II phosphatase FCP1 is partially helical in the unbound state. *Biochem. Biophys. Res. Commun.* **410,** 461–465 (2011).

14. Lawrence, C. W. & Showalter, S. A. Carbon-Detected 15N NMR Spin Relaxation of an Intrinsically Disordered Protein: FCP1 Dynamics Unbound and in Complex with RAP74. *J. Phys. Chem. Lett.* **3,** 1409–1413 (2012).

15. Putnam, C. D., Hammel, M., Hura, G. L. & Tainer, J. A. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q. Rev. Biophys.* **40,** 191–285 (2007).

16. Lipfert, J. & Doniach, S. Small-angle X-ray scattering from RNA, proteins, and protein complexes. *Annu. Rev. Biophys. Biomol. Struct.* **36,** 307–327 (2007).

17. Petoukhov, M. V. & Svergun, D. I. Applications of small-angle X-ray scattering to biomacromolecular solutions. *Int. J. Biochem. Cell Biol.* **45,** 429–437 (2013).

18. Bernadó, P., Mylonas, E., Petoukhov, M. V., Blackledge, M. & Svergun, D. I. Structural Characterization of Flexible Proteins Using Small-Angle X-ray Scattering. *J. Am. Chem. Soc.* **129,** 5656–5664 (2007).

19. Receveur-Brechot, V. & Durand, D. How random are intrinsically disordered proteins? A small angle scattering perspective. *Curr. Protein Pept. Sci.* **13,** 55–75 (2012).

20. Bernadó, P. & Svergun, D. I. Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering. *Mol. Biosyst.* **8,** 151–167 (2011).

21. Mertens, H. D. T. & Svergun, D. I. Structural characterization of proteins and complexes using small-angle X-ray solution scattering. *J. Struct. Biol.* **172,** 128–141 (2010).

22. Fisher, C. K. & Stultz, C. M. Constructing ensembles for intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **21,** 426–431 (2011).

23. Terakawa, T. & Takada, S. Multiscale ensemble modeling of intrinsically disordered proteins: p53 N-terminal domain. *Biophys. J.* **101,** 1450–1458 (2011).

24. Dedmon, M. M., Lindorff-Larsen, K., Christodoulou, J., Vendruscolo, M. & Dobson, C. M. Mapping long-range interactions in alpha-synuclein using spin-label NMR and ensemble molecular dynamics simulations. *J. Am. Chem. Soc.* **127,** 476–477 (2005).

25. Marsh, J. A. & Forman-Kay, J. D. Ensemble modeling of protein disordered states: Experimental restraint contributions and validation. *Proteins Struct. Funct. Bioinforma.* **80,** 556–572 (2012).

26. Lowry, D. F., Stancik, A., Shrestha, R. M. & Daughdrill, G. W. Modeling the accessible conformations of the intrinsically unstructured transactivation domain of p53. *Proteins* **71,** 587–598 (2008).

27. Jha, A. K., Colubri, A., Freed, K. F. & Sosnick, T. R. Statistical coil model of the unfolded state: resolving the reconciliation problem. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 13099–13104 (2005).

28. Nathalie Sibille, P. B. Structural characterization of intrinsically disordered proteins by the combined use of NMR and SAXS. *Biochem. Soc. Trans.* **40,** 955–62 (2012).

29. Jensen, M. R. *et al.* Quantitative determination of the conformational properties of partially folded and intrinsically disordered proteins using NMR dipolar couplings. *Struct. Lond. Engl. 1993* **17,** 1169–1185 (2009).

30. Svergun, D. I., Semenyuk, A. V. & Feigin, L. A. Small-angle-scattering-data treatment by the regularization method. *Acta Crystallogr. A* **44,** 244–250 (1988).

31. Franke, D. & Svergun, D. I. *DAMMIF* , a program for rapid *ab-initio* shape determination in small-angle scattering. *J. Appl. Crystallogr.* **42,** 342–346 (2009).

32. Volkov, V. V. & Svergun, D. I. Uniqueness of *ab initio* shape determination in small-angle scattering. *J. Appl. Crystallogr.* **36,** 860–864 (2003).

33. Jones, G. in *Encycl. Comput. Chem.* (John Wiley & Sons, Ltd, 2002). at <http://onlinelibrary.wiley.com/doi/10.1002/0470845015.cga004/abstract>

34. Svergun, D., Barberato, C. & Koch, M. H. J. *CRYSOL* – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *J. Appl. Crystallogr.* **28,** 768–773 (1995).

35. Ozenne, V. *et al.* Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. *Bioinforma. Oxf. Engl.* **28,** 1463–1470 (2012).

36. Jie-Rong Huang, M. G. Residual dipolar couplings measured in unfolded proteins are sensitive to amino-acid-specific geometries as well as local conformational sampling. *Biochem. Soc. Trans.* **40,** 989–94 (2012).

37. Krzeminski, M., Marsh, J. A., Neale, C., Choy, W.-Y. & Forman-Kay, J. D. Characterization of disordered proteins with ENSEMBLE. *Bioinforma. Oxf. Engl.* **29,** 398–399 (2013).

38. Neal, S., Nip, A. M., Zhang, H. & Wishart, D. S. Rapid and accurate calculation of protein 1H, 13C and 15N chemical shifts. *J. Biomol. NMR* **26,** 215–240 (2003).

Chapter 6 Conclusion

Intrinsically disordered proteins are a class of proteins that have generated a large amount of interest, and have forced those who study them to think from the perspective of an ensemble structure. How the large amount of flexibility and lack of unique secondary or tertiary structure translates into function for these proteins is not fully understood. Ironically, in order to do so requires the generation of a structure; an ensemble structure that is composed of multiple conformers that collectively reproduce observed experimental data. This dissertation has outlined the process of developing a methodology to determine what role disorder plays in the binding mechanism of the FCP1/RAP74 interaction and the generation of an ensemble structure that describes the IDP FCP1.

The C-terminal Domain of FCP1 Retains Helical Structure in the Unbound State

It is well known in the literature that FCP1 undergoes a coupled folding upon binding transition where the very C-terminus forms an α-helix upon interaction with the winged helix domain of RAP74. Due to the highly disordered nature of FCP1's C-terminal domain, prior to this thesis structural information that describes ctFCP1 was limited to only those ≈17 residues that make up the RAP74 interaction domain of an ≈80aa construct; and even then, only while in complex. Using ^{13}C-direct detected NMR experiments our lab was able to generate a complete set of chemical shifts of FCP1 in both unbound and bound states. These chemical shifts became probes for the presence of secondary structure in the apo-state of FCP1 with the addition of kosmotropes and chaotropes. TFE, an osmolyte known to produce α-helical structure in peptides, increased the amount of helical structure present in FCP1, as confirmed by CD. However, this structure formation was limited to the binding region. Urea, on the other hand, showed a small

loss in structural elements, again only in the binding region. Collectively this data shows that FCP1 does indeed acquire at least some transient helical structure while alone in solution and that the N-terminal region of FCP1 retains a fully disordered state. Also, through the addition of macromolecular crowding agents, Dextran and HeLa cell extract, we simulated the crowded environment of the cell and ctFCP1 did not show chemical shift changes consistent with an increase in structure. This verified that the disordered state of FCP1 under dilute aqueous conditions was biologically relevant.

The observations made here shed light onto structural propensities accessible to an IDP that was previously thought to be completely disordered. Above we have outlined a very general set of experiments that could be applied to other IDPs in order to qualitatively assess the amount of natively accessible structural elements for a given protein. FCP1 has also provided an excellent case study for the amount structure that can be found in an IDP because through the TFE and urea studies a very stark transition from fully disordered to at least partially disordered is observed. From this it is evident that one cannot just the simply draw the conclusion that IDPs must regularly sample some structural elements or IDPs exist as a purely random coil because within the same intrinsically disordered protein there are clear variations in flexibility.

Development of Carbon-Detected ^{15}N Spin Relaxation to Measure the Dynamics of FCP1 Unbound and In Complex with RAP74

Since IDPs are highly flexible in solution, quantifying the local backbone dynamics of FCP1 is imperative to accurately describe its behavior both in and out of complex with RAP74. This was accomplished through the development of novel NMR ^{15}N spin relaxation experiments that incorporate ^{13}C-direct detection. Analysis of the relaxation data showed that the RAP74 binding region of FCP1displays different relaxation rates from baseline, consistent with the conclusions from chapter 2, where those same residues at least transiently adopt helical structure in the unbound state. Upon interaction with RAP74 the binding domain of FCP1 undergoes a

significant change in dynamics while the rest of the proteins remained unperturbed by the interaction. ^{15}N spin relaxation measurements were made of RAP74 in parallel to those done on FCP1 and show that, unlike FCP1, RAP74 experiences only a small loss in flexibility, isolated to helix 2.5. This provides a clear example for why it is necessary to characterize the dynamic behavior of IDPs. Well-folded globular proteins experience only a minimal change in their backbone dynamics upon interaction with a binding partner. IDPs, on the other hand, become much more conformationally limited. This significantly affects the local dynamics at the interaction site. Interestingly, the region of RAP74 that does show a change in dynamics upon interaction with FCP1 is not near the portion that becomes fully α-helical, but is located near two serines (942 and 944) at the boundary between the binding region and the disordered N-terminus of FCP1. These serines are possible targets for post-translational modification and may play a role in regulating RAP74 binding affinity. However, the relaxation data set is still not complete. A carbon-detected NOE needs to be developed so that a model-free analysis, similar to the one carried out on RAP74, can be done on FCP1 to obtain the general order parameter S^2. This will allow for a more quantitative description of the changes in backbone flexibility of FCP1 while in complex.

Through the set of carbon-detected ^{15}N spin relaxation experiments presented in this chapter we have made information of the backbone dynamics easily available to other labs studying IDPs where they were limited either because of ^{1}H-detected relaxation experiments or specialized equipment. Also, for the case of FCP1 we have shown that the backbone dynamics of IDPs that contain structural elements is discernible from the baseline given by a random coil. As many IDPs undergo a cooperative folding upon interaction with their binding partner they experience a large change in conformational flexibility. This can be directly observed from changes in T_1 and T_2 relaxation rates of the IDP as they are closely tied to the general order parameter S^2, which generally describe the overall flexibility of the backbone at a given residue.

By directly comparing the backbone dynamics of the IDP FCP1 with its globular binding partner we were able to hypothesize that the greater flexibility of the disordered region nearest the binding domain allows for it interact with the binding protein at a site distant from the binding interface to play a regulatory role. This could be applied more generally to other IDPs that bind at several locations on the same protein.

The Effect of Stabilizing Structure on IDP Binding using the FCP1/RAP74 Interaction

The previous chapters sought to generate data that would structurally and dynamically describe ctFCP1 in the apo and holo-state but they do not report on the interaction itself. With the use of ITC we sought to better define the interaction between FCP1 and RAP74 and what role the disorder-to-order transition may play in binding. Initial experiments were carried out under aqueous solvent conditions and showed an endothermic-to-exothermic transition as temperature increased. Analysis of titration data at each temperature reported that the binding affinity remained relatively constant and that the binding interaction displayed a large, negative ΔC_P. This was representative of burying protein surface away from solvent. However, it could not be determined whether or not burial of protein surface was due to the removal of solvent from the binding interface or FCP1 folding into a helical conformation. It is important to note, however, that the interaction was entropically favorable regardless of the amount of conformational restriction FCP1 underwent to form the complex. Since in previous chapters TFE has been shown to fold only the RAP74 binding domain, further experiments were carried out in its presence in an attempt to remove folding from the observed enthalpies. While in the presence of TFE, the representative endothermic-to-exothermic transition was no longer present and the interaction was no longer entropically favorable. This leads to the hypothesis that the conformational entropy of FCP1 in the unbound state plays a role in forming a favorable binding interaction. In the chapter relating to the backbone dynamics of FCP1, it was proposed that the disordered region of FCP1 directly near to the binding domain may have some regulatory role in RAP74 binding.

Thus, the idea that disorder may play a functional role in the FCP1/RAP74 binding interaction is not unique to this chapter.

By utilizing ITC we were able to observe that the burial of hydrophobic surfaces has a large effect on the binding interactions of IDPs. Since IDPs undergo what can be a large folding event upon protein-protein or protein-nucleic acid interaction the hydrophobic effect, at least in the case of FCP1, dominates the overall binding interaction and can provide an observable endothermic to exothermic transition. Over the course of this transition the binding event goes from being entropically favorable to entropically driven. The observation that an IDP interaction can be so entropically favorable contradicts a commonly held belief in the literature that IDP interaction are typically weak because a large entropic penalty must be paid in order to fold into the complex state. Lastly, with the use of cosolvents, such as TFE, we were able to show that pre-formed structural elements can make an IDP interaction entropically unfavorable, allowing us to hypothesize the role disorder plays in the overall binding mechanism.

Evaluation of Ensemble Modeling Methods Through the Incorporation of SAXS and NMR parameters

As previously mentioned IDPs do not exist as a singular well folded structure, but as an ensemble state where the protein is rapidly interconverting between multiple conformers. In previous chapters we have collected data that describes the structural and dynamic characteristics of this ensemble. This data, however, describes the average structure of the ensemble and not specific to any one conformer. In an attempt to generate an ensemble structure we turned to a combination of SAXS and ensemble modeling techniques (EOM, flexible-meccano, and ENSEMBLE).From SAXS we were able to create 3D reconstructions of a low resolution structure of a shorter FCP1 construct, residues 932-961, under various solvent conditions. Here we found that the only solvent condition that resulted in a change to the ensemble structure was the inclusion of 10% TFE, and this was most likely due to TFE facilitating dimerization of the

peptide at that concentration. The scattering profile for FCP1$^{932-961}$ in aqueous solvent was then passed onto the ensemble modeling program EOM. The resulting ensemble structure displayed properties that showed that it was more compact than the experimental data suggested. This may have been due to a selection bias resulting from the shorter peptide having a higher percentage of residues with a proclivity to gain secondary structure. Ensembles generated with flexible-meccano and ENSEMBLE yielded an R_g consistent with experimental data. For both methods the addition of parameters meant to relay structural information does not display any considerable change in the overall size or shape of the ensembles. In the case of ENSEMBLE, all structures reproduced experimental chemical shifts equally well, while only the flexible-meccano ensemble that had 50% helix propensity explicitly defined was capable of reproducing chemical shifts to the same degree as ENSEMBLE. Ultimately from the data that has been acquired one can infer that given that the RAP74 binding region is small in comparison to the rest of the protein that the much more disordered N-terminus dictates the overall ensemble structure, and that a random coil model with the appropriate amount of explicitly define structural propensities is sufficient in generating an accurate ensemble structure of FCP1.

Future Directions

From the data that has been described above, it is clear that still more experiments should be carried out to further understanding of this binding interaction. We anticipate that, with time, more IDPs will be found amenable to application of our set of carbon –detected NMR spin relaxation experiments, which are archived for the public on the Web and through Penn State ScholarSphere. Therefore, as previously mentioned, a carbon-detected NOE needs to be developed in order to expand the technique and so that the general order parameter S^2 can be determined. This could prove to be a useful parameter to obtain if software packages like ENSEMBLE further expand to use it.

As mentioned in chapter 5, measurement of PRE and intramolecular NOE values could prove useful in further refining the ensemble structure. Also, the ensemble modeling has been limited to FCP1 free in solution. In order to determine if there are significant changes in the ensemble structure upon binding similar experimentation must be done on FCP1 while in complex. In terms of further understanding the binding event itself additional ITC experiments should be carried out. The addition of TFE, while generating rather striking results, made data analysis rather difficult. A series of mutations in the RAP74 binding region of FCP1to other amino acid types designed to either enhance or disrupt helical formation should provide an alternate approach. Future directions aside, this dissertation lays a strong foundation for developing a set of experiments applicable over a wide range of IDPs which can be utilized to build an ensemble model for disordered proteins, thus, providing a molecular description of the behavior of this interesting class of proteins.

Appendix

Supplementary Information: Chapter 3

Table A1: Relaxation Rates of FCP1in the apo-state

Residue Number	T_1 (s)	T_1 (err)	T_2 (s)	T_2(err)
878	1.151	0.004	0.704	0.005
880	0.773	0.003	0.562	0.002
881	0.689	0.001	0.511	0.001
883	0.656	0.001	0.441	0.002
884	0.582	0.007	0.283	0.004
885	0.608	0.001	0.368	0.001
887	0.589	0.001	0.343	0.001
889	0.639	0.002	-	-
890	0.711	0.019	0.340	0.001
892	0.721	0.006	0.333	0.005
893	0.660	0.007	0.290	0.002
894	0.681	0.005	0.325	0.002
896	0.598	0.001	0.369	0.000
898	0.577	0.004	0.321	0.002
899	0.606	0.002	0.342	0.002
900	0.742	0.002	0.360	0.001
901	0.644	0.005	0.384	0.002
903	0.676	0.002	0.374	0.001
904	0.743	0.006	0.381	0.002
905	0.637	0.007	0.330	0.009
906	0.622	0.001	0.340	0.001
907	0.645	0.012	0.317	0.001
908	0.662	0.003	0.342	0.002
909	0.594	0.002	0.363	0.002
910	0.599	0.001	0.329	0.001
911	0.800	0.002	0.413	0.003
912	0.741	0.009	0.430	0.003
913	0.743	0.002	0.357	0.001
914	0.673	0.004	0.359	0.002
916	0.603	0.003	0.278	0.000
917	0.596	0.005	0.334	0.006
918	-	-	0.305	0.009
919	-	-	0.294	0.006
920	0.619	0.010	0.249	0.005

921	0.537	0.003	0.284	0.001
922	0.516	0.003	0.277	0.001
923	0.525	0.002	0.264	0.001
924	0.555	0.005	0.293	0.001
925	0.554	0.001	0.310	0.000
926	0.535	0.001	0.326	0.001
927	0.563	0.003	0.359	0.001
928	-	-	0.398	0.001
929	0.667	0.003	0.326	0.002
930	0.639	0.001	0.322	0.001
931	0.644	0.002	0.323	0.001
932	0.679	0.005	0.331	0.002
933	0.621	0.001	0.353	0.000
934	0.604	0.001	0.370	0.000
935	0.638	0.004	0.338	0.003
936	0.653	0.003	0.364	0.002
937	0.667	0.001	0.351	0.002
938	0.642	0.002	0.341	0.000
939	0.636	0.001	0.438	0.000
940	0.601	0.001	0.350	0.001
941	0.629	0.004	0.390	0.001
942	0.757	0.006	0.382	0.002
943	0.768	0.006	0.358	0.002
944	0.666	0.003	0.313	0.002
945	0.559	0.003	0.274	0.002
946	0.558	0.001	0.280	0.001
947	0.559	0.001	0.269	0.001
948	0.538	0.001	0.305	0.000
949	0.512	0.001	0.234	0.001
950	0.527	0.001	0.245	0.000
951	0.512	0.000	0.252	0.000
952	0.509	0.001	0.264	0.001
953	0.572	0.000	0.275	0.001
954	0.541	0.001	0.263	0.000
955	0.578	0.001	0.298	0.001
956	0.574	0.001	0.314	0.000
957	0.589	0.001	0.323	0.000
958	0.674	0.001	0.378	0.001
959	0.705	0.001	0.433	0.001
960	0.788	0.002	0.526	0.001
961	1.021	0.001	0.718	0.001

Table A2: Relaxation Rates of FCP1in the holo-state

Residue Number	T_1(s)	T_1(err)	T_2(s)	T_2(err)
878	1.074	0.011	0.894	0.038
880	0.712	0.004	0.534	0.006
881	0.836	0.023	0.425	0.003
882	0.610	0.003	0.357	0.002
883	0.717	0.007	0.489	0.008
884	0.532	0.008	0.364	0.005
885	0.629	0.003	0.407	0.002
887	0.611	0.004	0.310	0.003
889	0.575	0.003	0.356	0.007
890	0.566	0.003	0.374	0.010
892	0.609	0.004	0.341	0.039
893	0.601	0.009	0.474	0.014
894	0.515	0.008	0.316	0.010
896	0.632	0.001	0.394	0.002
898	0.551	0.011	0.416	0.024
899	0.596	0.003	0.358	0.007
900	0.732	0.009	0.402	0.018
901	0.819	0.010	0.431	0.026
902	0.756	0.018	0.369	0.018
903	0.715	0.008	0.468	0.012
904	0.796	0.034	-	-
905	0.617	0.008	0.413	0.020
906	0.759	0.004	0.423	0.008
907	0.604	0.021	-	-
909	0.762	0.002	0.424	0.024
910	0.549	0.002	0.336	0.008
911	0.854	0.010	0.441	0.023
912	0.854	0.010	0.431	0.011
913	0.767	0.017	0.417	0.015
914	0.695	0.007	0.445	0.011
915	0.661	0.007	0.377	0.014
916	0.586	0.002	0.340	0.010
917	0.618	0.001	0.416	0.026
918	0.615	0.012	0.319	0.024
921	0.550	0.009	0.307	0.007
922	0.544	0.005	0.351	0.007
923	0.580	0.008	0.302	0.016
924	0.524	0.005	0.283	0.007

925	0.562	0.006	0.390	0.004
926	0.648	0.006	0.378	0.004
927	0.578	0.006	0.368	0.012
928	-	-	0.310	0.010
929	0.590	0.003	0.321	0.006
930	0.608	0.012	0.381	0.006
931	0.674	0.008	0.265	0.019
932	0.622	0.003	0.320	0.008
933	0.583	0.007	0.372	0.002
934	0.541	0.003	0.281	0.002
935	0.669	0.005	0.387	0.002
936	0.659	0.003	0.308	0.007
937	0.698	0.035	0.260	0.005
938	0.553	0.002	0.315	0.007
939	0.481	0.002	0.339	0.007
940	0.548	0.023	0.248	0.011
941	0.572	0.005	0.222	0.015
942	0.786	0.008	0.238	0.049
943	0.732	0.008	0.218	0.094
945	0.598	0.005	-	-
947	0.760	0.030	0.081	0.116
947	-	-	0.072	0.088
948	0.684	0.012	0.066	0.002
949	0.803	0.037	-	-
950	0.747	0.039	0.050	0.030
952	-	-	0.060	0.042
953	-	-	0.097	0.020
954	-	-	0.085	0.014
955	0.782	0.021	0.089	0.021
956	0.761	0.010	-	-
957	0.631	0.014	-	-
958	0.752	0.022	0.061	0.016
959	-	-	0.073	0.010
960	0.677	0.005	0.149	0.022
961	0.673	0.003	0.236	0.047

Table A3: Relaxation Rates and order parameters of RAP74 in the apo-state

Residue Number	T_1(s)	T_1(err)	T_2(s)	T_2(err)	NOE	NOE (err)	S^2	S^2(err)
434	2.0109	0.1046	0.9425	0.2271	-2.628	0.013	0.06	0.005
435	1.1346	0.0219	0.6820	0.0972	-1.638	0.013	0.132	0.008
436	1.1685	0.0458	0.5751	0.0793	-1.476	0.013	0.158	0.008
437	0.9829	0.0215	0.5606	0.0649	-1.121	0.013	0.194	0.008
439	0.7693	0.0042	0.5447	0.0295	-0.798	0.013	0.269	0.013
442	0.9197	0.0235	0.4732	0.0552	-0.682	0.013	0.26	0.008
443	0.6999	0.0139	0.2876	0.0219	-	-	-	-
444	0.8490	0.0190	-	-	-	-	-	-
445	0.9150	0.0224	0.4714	0.0581	-0.654	0.013	0.249	0.008
446	0.8466	0.0162	0.4098	0.0385	-0.536	0.013	0.292	0.009
448	0.8188	0.0190	0.3827	0.0380	-0.300	0.013	0.326	0.009
449	-	-	0.4002	0.1144	-	-	-	-
450	0.5702	0.0067	0.1886	0.0098	-	-	-	-
451	0.6291	0.0058	0.2708	0.0121	-	-	-	-
452	0.5397	0.0004	0.1633	0.0014	0.378	0.013	0.682	0.016
453	0.5131	0.0003	0.1489	0.0008	0.376	0.013	0.748	0.016
454	0.5125	0.0001	0.1108	0.0004	0.588	0.013	0.819	0.016
455	0.4826	0.0010	-	-	0.713	0.013	0.85	0.016
456	0.4736	0.0009	0.1232	0.0013	0.727	0.013	0.826	0.016
457	0.4775	0.0003	0.0943	0.0006	0.760	0.013	0.909	0.017
458	0.4611	0.0002	0.1146	0.0007	0.713	0.013	0.879	0.016
459	0.4575	0.0001	0.1253	0.0002	0.715	0.013	0.838	0.016
460	0.4538	0.0001	0.1114	0.0002	0.699	0.013	0.897	0.016
461	0.4650	0.0001	0.1116	0.0003	0.730	0.013	0.867	0.016
462	0.4641	0.0000	0.1122	0.0002	0.741	0.013	0.884	0.016
463	0.4634	0.0002	-	-	0.683	0.013	0.935	0.016
464	0.4922	0.0005	0.1136	0.0011	0.717	0.013	0.866	0.017
465	0.4746	0.0001	0.1285	0.0003	0.697	0.013	0.828	0.016
468	0.4539	0.0002	0.1371	0.0001	0.690	0.013	0.829	0.015
469	0.4590	0.0002	0.1379	0.0003	0.727	0.013	0.826	0.015
470	0.4887	0.0017	0.1204	0.0017	0.707	0.013	0.846	0.016
471	0.4938	0.0007	0.1173	0.0017	0.678	0.013	0.8	0.017
472	0.4915	0.0004	0.1173	0.0010	0.708	0.013	0.834	0.016
473	0.4692	0.0001	0.1206	0.0003	0.696	0.013	0.84	0.016
474	0.4789	0.0001	0.1135	0.0003	0.730	0.013	0.854	0.016
475	0.4853	0.0000	0.0931	0.0001	0.650	0.013	0.901	0.017

476	0.4806	0.0002	0.0809	0.0006	0.671	0.013	0.882	0.016
477	0.4989	0.0002	0.0882	0.0002	0.653	0.013	0.8	0.016
482	0.5502	0.0038	0.0990	0.0035	0.514	0.013	-	-
483	0.9404	0.0247	0.4608	0.0535	-	-	-	-
484	0.5122	0.0031	0.1581	0.0039	0.595	0.013	0.728	0.016
485	0.5316	0.0041	0.1528	0.0052	0.566	0.013	0.73	0.016
487	0.5001	0.0024	0.1318	0.0034	0.609	0.013	0.781	0.016
488	0.4855	0.0015	0.1245	0.0020	0.671	0.013	0.835	0.016
489	0.4667	0.0005	0.1141	0.0007	0.709	0.013	0.876	0.015
490	0.4633	0.0002	0.1228	0.0006	0.689	0.013	0.85	0.015
491	0.4679	0.0002	0.1016	0.0002	0.731	0.013	0.899	0.016
492	0.4627	0.0001	0.1289	0.0000	0.685	0.013	0.84	0.016
493	0.4385	0.0001	0.1109	0.0001	-	-	-	-
494	0.4492	0.0000	0.1127	0.0000	0.754	0.013	0.911	0.016
495	0.4715	0.0000	0.1253	0.0001	0.701	0.013	0.834	0.016
496	0.4557	0.0000	0.1344	0.0001	0.673	0.013	0.823	0.016
497	0.4425	0.0002	0.0967	0.0002	0.714	0.013	0.901	0.015
498	0.4720	0.0000	0.1180	0.0000	0.716	0.013	0.87	0.016
499	0.4807	0.0000	0.1200	0.0000	0.698	0.013	0.842	0.016
501	0.4755	0.0002	-	-	0.267	0.013	-	-
503	0.4811	0.0000	-	-	0.695	0.013	-	-
504	0.4988	0.0000	0.1363	0.0001	0.689	0.013	0.793	0.016
505	0.5366	0.0002	0.1295	0.0001	0.648	0.013	0.783	0.016
506	0.5178	0.0011	0.1253	0.0002	0.602	0.013	0.827	0.016
507	0.5047	0.0013	0.0977	0.0007	0.662	0.013	-	-
509	0.4987	0.0009	-	-	0.698	0.013	-	-
510	0.4877	0.0002	0.1004	0.0006	0.606	0.013	0.831	0.015
513	0.4868	0.0004	0.1236	0.0007	0.748	0.013	0.814	0.016
514	0.4593	0.0001	0.0952	0.0003	0.700	0.013	0.886	0.015
515	0.4695	0.0001	0.1354	0.0001	0.687	0.013	0.806	0.016
516	0.4995	0.0002	0.1734	0.0004	0.545	0.013	0.693	0.016
517	0.5815	0.0000	0.2375	0.0005	0.240	0.013	-	-

Table A4: Relaxation Rates and order parameters of RAP74 in the holo-state

Residue Number	T1(s)	T1(err)	T2(s)	T2(err)	NOE	NOE (err)	S2	S2(err)
434	1.4167	0.0352	0.9053	0.1910	-1.810	0.0130	0.094	0.008
435	0.9798	0.0046	0.7367	0.0509	-1.416	0.0130	0.161	0.013
436	0.8921	0.0101	0.6148	0.0689	-0.982	0.0130	0.212	0.008
437	0.8149	0.0045	0.5538	0.0304	-0.785	0.0130	0.25	0.01
439	0.7257	0.0005	0.4860	0.0051	-0.616	0.0130	0.298	0.011
442	0.7459	0.0050	0.4921	0.0357	-0.512	0.0130	0.301	0.01
443	0.7719	0.0084	0.5329	0.0563	-0.447	0.0130	0.303	0.019
444	0.7245	0.0042	0.4615	0.0251	-0.354	0.0130	0.326	0.009
445	0.7521	0.0068	0.4520	0.0324	-0.514	0.0130	0.298	0.007
446	0.6928	0.0030	0.3865	0.0146	-0.373	0.0130	0.341	0.008
448	0.6496	0.0032	0.3039	0.0166	-	-	-	-
449	0.7074	0.0077	0.2798	0.0174	-	-	-	-
450	0.6284	0.0029	0.2871	0.0114	-	-	-	-
451	0.5985	0.0004	0.2024	0.0011	0.239	0.0130	-	-
452	0.5962	0.0001	0.1315	0.0006	0.415	0.0130	0.681	0.01
453	0.5983	0.0000	0.1224	0.0001	0.479	0.0130	0.71	0.011
456	0.5681	0.0002	0.0910	0.0003	0.737	0.0130	0.854	0.017
457	0.5780	0.0002	0.0915	0.0003	0.754	0.0130	0.86	0.017
458	0.5639	0.0002	0.0916	0.0006	0.757	0.0130	0.871	0.017
459	0.5758	0.0001	0.0952	0.0003	0.718	0.0130	0.844	0.017
460	0.5562	0.0000	0.0887	0.0001	0.985	0.0130	0.888	0.017
461	0.5592	0.0001	0.0869	0.0003	0.758	0.0130	0.883	0.017
462	0.5717	0.0001	0.0983	0.0005	0.701	0.0130	0.852	0.017
463	0.5756	0.0000	0.0918	0.0001	0.776	0.0130	0.855	0.017
464	0.5956	0.0004	0.0929	0.0012	0.734	0.0130	0.858	0.017
465	0.5877	0.0001	0.1001	0.0001	0.692	0.0130	0.828	0.017
466	0.5763	0.0001	0.1003	0.0001	0.686	0.0130	0.822	0.017
468	0.5678	0.0001	0.0974	0.0001	0.747	0.0130	0.862	0.017
469	0.5697	0.0001	0.1004	0.0003	0.748	0.0130	0.852	0.017
470	0.6018	0.0004	0.0987	0.0017	0.717	0.0130	0.842	0.017
471	0.5849	0.0001	0.0894	0.0003	0.677	0.0130	0.862	0.017
472	0.5951	0.0004	0.0863	0.0005	0.749	0.0130	0.856	0.017
473	0.5894	0.0001	0.0957	0.2566	0.721	0.0130	0.854	0.017
474	0.5860	0.0001	0.0872	0.0000	0.711	0.0130	0.876	0.017
475	0.5836	0.0001	0.0867	0.0001	0.773	0.0130	0.87	0.017
476	0.5952	0.0001	0.0908	0.0004	0.740	0.0130	0.852	0.017

477	0.6102	0.0003	0.0727	0.0001	0.687	0.0130	-	-
478	0.5944	0.0002	0.1056	0.0002	0.653	0.0130	-	-
480	0.5696	0.0004	0.0867	0.0007	0.691	0.0130	0.884	0.017
481	0.5648	0.0002	0.0867	0.0002	0.776	0.0130	0.899	0.017
482	0.5963	0.0007	0.0948	0.0005	0.706	0.0130	0.84	0.017
483	0.5691	0.0003	0.1006	0.0006	0.727	0.0130	-	-
484	0.5746	0.0008	0.1009	0.0014	0.768	0.0130	0.831	0.016
485	0.6032	0.0007	0.0952	0.0011	0.725	0.0130	0.811	0.016
486	0.6043	0.0035	0.0920	0.0026	0.748	0.0130	0.808	0.017
487	0.5788	0.0004	0.0952	0.0008	0.754	0.0130	0.841	0.017
488	0.5931	0.0007	0.0926	0.0011	0.759	0.0130	0.844	0.017
489	0.5949	0.0009	0.0964	0.0009	0.776	0.0130	0.83	0.017
490	0.5670	0.0002	0.0914	0.0003	0.808	0.0130	0.854	0.017
491	0.5676	0.0001	0.0726	0.0656	0.748	0.0130	-	-
492	0.5610	0.0002	0.0897	0.0006	0.757	0.0130	0.863	0.017
493	0.5467	0.0001	0.0884	0.0001	0.777	0.0130	0.882	0.017
494	0.5422	0.0003	0.0907	0.0007	0.709	0.0130	0.877	0.017
495	0.5891	0.0000	0.0952	0.0001	0.732	0.0130	0.853	0.017
496	0.5871	0.0000	0.0973	0.0016	0.558	0.0130	0.839	0.017
497	0.5562	0.0000	0.0876	0.0002	0.757	0.0130	0.882	0.017
498	0.5690	0.0000	0.0903	0.0001	0.771	0.0130	0.888	0.017
499	0.5992	0.0000	0.0942	0.0000	0.746	0.0130	0.842	0.017
500	0.5736	0.0000	0.0963	0.0001	0.732	0.0130	0.859	0.017
501	0.5820	0.0001	0.1055	0.0007	0.703	0.0130	0.82	0.017
503	0.5923	0.0002	0.0968	0.0006	0.717	0.0130	0.818	0.017
504	0.6279	0.0001	0.0994	0.0001	0.687	0.0130	0.791	0.017
505	0.6258	0.0000	0.1011	0.0005	0.671	0.0130	0.808	0.017
506	0.6256	0.0001	0.1046	0.0003	0.637	0.0130	0.785	0.017
507	0.6093	0.0007	0.1059	0.0003	0.688	0.0130	0.793	0.017
508	0.6027	0.0012	0.1053	0.0018	0.707	0.0130	0.793	0.017
509	0.5982	0.0001	0.0783	0.0002	0.656	0.0130	-	-
510	0.5836	0.0003	0.1070	0.0004	0.650	0.0130	0.794	0.016
511	0.5635	0.0005	0.1035	0.0011	0.740	0.0130	0.851	0.016
512	0.6081	0.0006	0.0921	0.0005	0.717	0.0130	0.836	0.017
513	0.5976	0.0003	0.0908	0.0005	0.701	0.0130	0.831	0.017
514	0.5964	0.0001	0.1011	0.0001	0.742	0.0130	0.818	0.017
515	0.6016	0.0001	0.1006	0.0001	0.716	0.0130	0.814	0.017
516	0.5866	0.0001	0.1370	0.0003	0.526	0.0130	-	-
517	0.6327	0.0000	0.2109	0.0004	0.188	0.0130	-	-

VITA

Education

Ph.D. in Chemistry, December 2013

The Pennsylvania State University, University Park, PA

GPA 3.72

B.S. in Chemistry, May 2008

Roanoke College, Salem, VA

GPA 3.40

Publications

Lawrence, C. W.; Showalter, S. A. "FCP1 Dynamics Unbound and in Complex with RAP74." J. Phys. Chem. Lett. 2012, 3, 1409−1413.

Lawrence, C. W.; Bonny, A.; Showalter, S. A. "The Disordered C-Terminus of the RNA Polymerase II Phosphatase FCP1 Is Partially Helical in the Unbound State." Biochem. Biophys. Res. Commun. 2011, 410, 461−465

Presentations

Lawrence, Chad W.; Noid, William G; Showalter, Scott A."The Effect of Induced Folding on IDP Binding" Symposium on Biomolecular Structure, Dynamics and Function, Memphis, TN, 27-29 April 2012

Lawrence, Chad W.; Showalter, Scott A. "Application of Carbon Detected NMR Methods to the Study of Intrinsically Disordered Proteins" The 52nd ENC, Experimental Nuclear Magnetic Resonance Conference, Pacific Grove, CA, 10-15 April 2011