Kinetic Characterization of the Coupled Folding and Binding Mechanism of Bacterial RNase P protein: An Intrinsically Unstructured Protein

by

Yu-Chu Chang

Department of Biochemistry
Duke University

Date: _______________________

Approved:

___________________________
Terrence G. Oas, Supervisor

___________________________
Michael D. Been

___________________________
Gordon G. Hammes

___________________________
Gary J. Pielak

___________________________
David C. Richardson

___________________________
Pei Zhou

Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biochemistry in the Graduate School of Duke University

2009
ABSTRACT

Kinetic Characterization of the Coupled Folding and Binding Mechanism of Bacterial RNase P protein: An Intrinsically Unstructured Protein

by

Yu-Chu Chang

Department of Biochemistry
Duke University

Date: _______________________

Approved: _______________________

Terrence G. Oas, Supervisor

Michael D. Been

Gordon G. Hammes

Gary J. Pielak

David C. Richardson

Pei Zhou

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biochemistry in the Graduate School of Duke University

2009
Abstract

Understanding the interconversion between the thermodynamically distinguishable states present in a protein folding pathway provides not only the kinetics and energetics of protein folding but also insights into the functional roles of these states in biological systems. The protein component of bacterial RNase P holoenzyme from *Bacillus subtilis* (P protein) was used as a model system to elucidate the general folding/unfolding of an intrinsically unstructured protein (IUP) both in the absence and presence of ligands.

P protein was previously characterized as an intrinsically unstructured protein, and it is predominantly unfolded in the absence of ligands. Addition of small anions can induce the protein to fold. Therefore, the folding and binding are tightly coupled. Trimethylamine-N oxide (TMAO), an osmolyte that stabilizes the unliganded folded form of the protein, enabled us to study the folding process of P protein in the absence of ligand. Transient stopped-flow kinetic time courses at various final TMAO concentrations showed multiphase kinetics. Equilibrium “cotitration” experiments were performed using both TMAO and urea to obtain a TMAO-urea titration surface of P protein. Both kinetic and equilibrium studies show evidence of an intermediate state in the P protein folding process. The intermediate state is significantly populated and the folding rate constants involved in the reaction are slow relative to similar size proteins.

NMR spectroscopy was used to characterize the structural properties of the folding intermediate of P protein. The results indicate that the N-terminal (residues 2-
19) and C-terminal regions (residues 91-116, 118 is the last residue) are mostly unfolded. 
$^1$H-$^{15}$N HSQC NMR spectra were collected at various pH values. The results suggest that 
His 22 may play a major role in the energetics of the equilibria between the unfolded, 
intermediate, and native states of P protein.

Ligand-induced folding kinetics were also investigated to elucidate the 
overall coupled folding and binding mechanism of P protein and the holoenzyme 
assembly process. Stopped flow fluorescence experiments were performed at various 
final ligand concentrations and the data were analyzed using a minimal complexity model 
that included three conformational states (unfolded, intermediate and folded) in each of 
three possible liganding states (0, 1 and 2 ligands). The kinetic and equilibrium model 
parameters that best fit the data were used to calculate the flux through each of the six 
possible folding/binding pathways. This novel flux-based analysis allows evaluation of 
the relative importance of pathways in which folding precedes binding or vice versa. The 
results indicate that the coupled folding and binding mechanism of P protein is strongly 
dependent on ligand concentration. This conclusion can be generalized to other protein 
systems for which ligand binding is coupled to conformational changes.
To my Mom, in my heart
## Table of Contents

Abstract........................................................................................................................................ iv

Dedication...................................................................................................................................... vi

Table of Contents.......................................................................................................................... vii

List of Tables ................................................................................................................................ xi

List of Figures ................................................................................................................................. v

Acknowledgements...................................................................................................................... vii

Chapter 1 : Introduction .............................................................................................................. 1

1.1 Protein folding problem .......................................................................................................... 1

1.2 Two-state and multi-state protein folding and misfolded protein aggregation.............. 3

1.3 Intrinsically unstructured proteins ...................................................................................... 5

1.3.1 Experimental characterization of IUPs ............................................................... 7

1.4 Linked conformational change and ligand binding ......................................................... 8

1.5 Ribonuclease P holoenzyme ............................................................................................. 10

1.5.1 P RNA subunit ......................................................................................................... 12

1.5.2 RNase P protein subunit......................................................................................... 16

1.5.3 The bacterial holoenzyme complex ....................................................................... 20

1.6 Overview of this work ........................................................................................................... 23

Chapter 2 : Using Osmolyte to Study Folding Kinetics and Equilibria of Bacillus subtilis RNase P Protein ........................................................................................................... 25

2.1 Introduction ......................................................................................................................... 25

2.2 Results and Discussion .......................................................................................................... 30
2.2.1 Tryptophan variants of P protein as the probe to study kinetics and thermodynamics of P protein ................................................................. 30
2.2.2 TMAO stopped flow kinetic studies and chevron plots ......................... 34
2.2.3 Urea-TMAO fluorescence and CD co-titration ........................................ 38
2.2.4 $\Delta G_{UI}^0$ and $m_{UI}^0$ from NMR urea titration experiments ................ 41
2.2.5 Global fitting of kinetic and equilibrium data ......................................... 45
2.2.6 Utility of TMAO (renaturant) stopped flow and cotitration experiments as a general method to determine IUP folding mechanisms ............ 48
2.2.7 Population of the intermediate state ....................................................... 50

2.3 Materials and Methods ............................................................................. 51
   2.3.1 Chemicals and reagents ....................................................................... 51
   2.3.2 Expression and purification of P protein .............................................. 52
   2.3.3 Fluorescence and CD equilibrium cotitration experiments ................... 52
   2.3.4 Stopped-flow fluorescence kinetic experiments .................................... 53
   2.3.5 Data analysis ....................................................................................... 54

Chapter 3 : Probing the Folding Intermediate of Bacillus subtilis RNase P protein by NMR ................................................................. 56

3.1 Introduction .................................................................................................. 56

3.2 Results ........................................................................................................ 59
   3.2.1 Presence of extra peaks in folded P protein HSQC spectrum and native peaks in the unliganded P protein spectrum ................................ 59
   3.2.2 Sulfate and pH titration of folded P protein ......................................... 61
   3.2.3 Assignment of extra peaks in the folded P protein HSQC spectrum ...... 63
   3.2.4 Effect of histidine replacement in P protein ........................................ 69
4.3.4 Model for RNase P holoenzyme assembly process \textit{in vivo} ................. 117

4.4 Materials and Methods ........................................................................................................ 119

4.4.1 Protein expression and purification ............................................................. 119

4.4.2 Circular dichroism and fluorescence equilibrium experiments ................. 120

4.4.3 Kinetic stopped flow fluorescence experiments ........................................ 121

4.4.4 NMR spectroscopy experiments ...................................................................... 122

4.4.5 Data analysis .............................................................................................................. 122

Chapter 5 : Conclusions and Future Directions .............................................................. 123

5.1 Conclusions ...................................................................................................................... 123

5.1.1 Osmolyte-induced folding/unfolding is a general method to study the thermodynamics and kinetics of IUPs ................................................................. 123

5.1.2 The electrostatic interactions in the P protein are important to the equilibrium between different conformational states ......................................................... 124

5.1.3 Flux concept is general in all coupled conformational changes and ligand binding reactions ................................................................................................. 126

5.2 Future experiments ......................................................................................................... 128

5.2.1 The role of P protein stability in the holoenzyme assembly process and catalysis ................................................................. 128

5.2.2 Structural studies of the folding intermediate in P protein and the interactions with P RNA ........................................................................................................ 130

5.2.3 Coupled folding and binding mechanism for P protein with different ligands ...................................................................................................................... 131

5.2.4 Kinetics of exchange between ligand-bound forms of intermediate state and folded state by NMR relaxation dispersion experiments ........................................ 132

Reference ........................................................................................................................ 134

Biography ........................................................................................................................ 152
List of Tables

Table 2-1: Kinetic and equilibrium parameters of three-states folding and unfolding of P protein ............................................................................................................................... 46

Table 3-1: Backbone chemical shifts of the extra peaks in the folded P protein HSQC spectrum ........................................................................................................................... 81

Table 4-1 Best fit kinetic and equilibrium parameters for the P protein folding/binding mechanism .............................................................................................................. 102
List of Figures

Figure 1-1: Ribbon diagram of the crystal structure of the B. subtilis RNase P protein . 19
Figure 1-2: The structure of B. stearothermophilus RNase P ........................................... 22
Figure 2-1: Model of Trp 107 sidechain contacts with surrounding sidechains........... 31
Figure 2-2: Circular dichroism and fluorescence spectrum of the F107W variant of P protein in the presence of different TMAO concentrations. ......................................................... 33
Figure 2-3: Folding/unfolding of P protein induced by changes in TMAO concentration and monitored by stopped-flow fluorescence. ................................................................. 36
Figure 2-4: Folding and unfolding kinetics of F107W P protein as a function of final TMAO concentration. ................................................................................................................ 37
Figure 2-5: Fluorescence and circular dichroism detected equilibrium co-titration surface of the P protein.................................................................................................................. 39
Figure 2-6: Urea titration of P protein in the absence of sulfate monitored by HSQC spectrum and the estimation of free energy difference between U and I state. ............. 43
Figure 2-7: Population of U, I, and N state as a function of the TMAO concentration.... 51
Figure 3-1: $^1$H-$^{15}$N HSQC spectrum of $^{15}$N-labeled P protein in 5 mM sulfate, pH 5.0 at 25°C. ............................................................................................................................ Error! Bookmark not defined.
Figure 3-2: $^1$H-$^{15}$N HSQC spectrum of $^{15}$N-labeled P protein in 20 mM sodium cacodylate, pH 5.0, at 25°C. ......................................................................................................................... 62
Figure 3-3: Sulfate titration of the $^1$H-$^{15}$N HSQC spectrum of $^{15}$N-labeled P protein, at pH 6.0 and 25°C. .............................................................................................................................. 64
Figure 3-4: pH titration of $^1$H-$^{15}$N HSQC spectrum of $^{15}$N-labeled P protein in 10 mM sulfate at 25 °C.............................................................................................................................. 65
Figure 3-5: Mapping of the assigned extra peak to the crystal structure of P protein..... 67
Figure 3-6: Chemical shift analysis of the extra peaks in the folded P protein HSQC spectrum................................................................................................................................. 69
Figure 3-7: $^1$H-$^{15}$N HSQC spectrum of $^{15}$N-labeled F107W variant of P protein in 20 mM sodium cacodylate, pH 6.0, at 25°C .................................................................................. 72

Figure 3-8: HSQC spectra of various histidine variants of P protein under unfolded conditions at different pH values ........................................................................................................ 73

Figure 4-1: Fluorescence and circular dichroism spectra of sulfate and TMAO folded F107W P protein .................................................................................................................................. 88

Figure 4-2: Sulfate-induced equilibrium folding of F107W P protein ........................................................................................................ 89

Figure 4-3: Stopped-flow fluorescence transient kinetic traces ........................................................................................................ 91

Figure 4-4: Sulfate-dependence of observed rate constants of F107W P protein obtained from the fitting results of the stopped-flow kinetic traces .................................................. 92

Figure 4-5: The population of each state calculated from the equilibrium binding constants estimated from the fitting ........................................................................................................ 99

Figure 4-6: Fluxes for the six possible pathways of P protein coupled folding and binding .................................................................................................................................. 101

Figure 4-7: $^1$H-$^{15}$N HSQC spectrum of folded P protein in various concentrations of $[\text{Cr(CN)}_6]^3-$ .................................................................................................................................. 105

Figure 4-8: PRE results mapped on the P protein crystal structure ........................................................................................................ 107

Figure 4-9: The simulated population of each state and the fluxes through different pathways as a function of pyrophosphate and chloride ........................................................................... 116
Acknowledgements

I want to thank many people that helped me complete the work throughout my graduate career. First I would like to thank my advisor, Terry, for sharing all his valuable experience of the way to solve a problem and the way to think about science. He used his knowledge and patience to guide me through my graduate studies. I also want to thank my committee members Dr. Gordon Hammes, Dr. Gary Pielak, Dr. Pei Zhou, Dr. Dave Richardson, and Dr. Mike Been for all the critical and helpful suggestions. The former and present members of the Oas lab, especially Dr. Stacy Chamberlin, Dr. Chris Henkels, Dr. Pooja Arora, Dr. Preeti Chugha, and Billy Franch, are all great people to work with, and we share all the happiness and frustration in the lab.

Finally I want to thank my wife Yen-Ping Hsueh and all my family for their constant support during these years.
Chapter 1: Introduction

1.1 Protein folding problem

Proteins are one of the major components of the cell and perform almost all the structural and enzymatic functions. The notion is widely accepted that in order for proteins to exert their biological functions, they need to form specific atomic interactions with targets or substrates through definitive three-dimensional structures. The protein folding problem is an effort to resolve how a protein’s amino acid sequence dictates its three-dimensional atomic structure. In the 1960’s, Anfinsen showed that denatured ribonuclease A could refold spontaneously in vitro (Anfinsen, Haber et al. 1961), demonstrating that the proteins’ primary amino acid sequence contained all the information needed to reach the native conformation (Anfinsen 1973). Subsequently it has been shown that some proteins need modulators, such as chaperones, to facilitate proteins’ folding process (Fink 1999 for review), but Anfinsen’s principle has been confirmed by many other single domain, multiple domains proteins and membrane protein as well (Lind, Höjeberg et al. 1981).

There are now three main questions in the protein folding research area (Dill, Ozkan et al. 2007; Dill, Ozkan et al. 2008): (1) The folding code: defining how the native structure is thermodynamically stable based on a given amino acid sequence. The folding code is predominantly viewed as a sum of many different small interactions including hydrogen bonding, ion pairs, Van der Waals interactions, and hydrophobic interactions, and these are mainly expressed through secondary structure and local interactions in the sequence (Anfinsen and Scheraga 1975). However, from more
statistical mechanical models emerged a different view highlighting the dominant role of hydrophobic interactions in the folding code and their local and non-local distribution in the sequence (Dill 1999).

(2) The folding process: determining kinetically what routes or pathways proteins use to fold so quickly. Levinthal first speculated on how long the process would take if a protein must fold by a random search of all possible conformations in order to converge to its native state (Levinthal 1968). For a small protein of one hundred residues, Levinthal estimated that the time for this random search process would take approximately $10^{11}$ years, which is physically impossible. Studies have demonstrated that proteins can search and converge to their native state, sometimes in milliseconds and even microseconds (Arora, Oas et al. 2004). Thus, real proteins do not fold through a random search. Theoretical development and simulation of protein folding energy landscape attempts to resolve this discrepancy. The model describes folding not as a single microscopic pathway, but rather as a funnel-shaped energy landscape (Dill and Chan 1997).

(3) The structure prediction: determining how to predict the native structure computationally. Both bioinformatics based and physics based approaches have had some success on protein structure prediction (Duan and Kollman 1998). Successful structure prediction also facilitates the development of the protein design research (Kaplan and Degrado 2004; Lassila, Privett et al. 2006). Novel proteins are now being designed as variants of existing proteins or *de novo* (Kuhlman, Dantas et al. 2003).
1.2 Two-state and multi-state protein folding and misfolded protein aggregation

The folding kinetics and equilibria of many small proteins of less than 100 amino acid residues have been extensively studied and most of these model proteins contain single globular domain in the folded structures (Jackson 1998). For many of these systems, kinetic and equilibrium data were interpreted with a simple two-state model in which only the unfolded state (U) and the folded state (N) are populated on the folding pathway. For such reactions, displacement of the folding equilibrium is a monophasic processes, and the natural logarithm of the unfolding rate constants and refolding rate constants versus denaturant concentration are frequently linear. Moreover, the thermodynamic parameters, $\Delta G_{UN}^0$ and $m$ value determined from the kinetic results should agree with those obtained from the equilibrium results.

Although a two-state model of protein folding provides a simple and elegant analytical description of the process for some proteins, other proteins require a multi-state folding mechanism. In these cases, single or multiple intermediate states exist in the folding process. In some cases even a small, monomeric domain of a protein can involve a multi-state folding process (Khorasanizadeh, Peters et al. 1996; Dalby, Clarke et al. 1998), so the presence of the intermediate states do not have a strong correlation with the size of the proteins. Other studies also demonstrate that under different experimental conditions, proteins that normally fold through a two-state mechanism can fold through an intermediate state (Gorski, Capaldi et al. 2001). Alternatively, mutations that destabilize the intermediate state relative to the unfolded state can switch the kinetics
from a three-state to a two-state mechanism. Changes in pH or temperature can have a similar effect (Khorasanizadeh, Peters et al. 1996; Dalby, Clarke et al. 1998).

The search for the intermediate states in the folding process led to the development of many techniques to identify and characterize the properties of these intermediate states. Baldwin and colleagues (Tsong, Baldwin et al. 1971) used temperature jump experiments to demonstrate a kinetic intermediate with a millisecond time scale lifetime for the unfolding of RNase A. Ikai and Tanford used stopped-flow techniques to demonstrate a kinetic intermediate in the guanidine hydrochloride induced unfolding transition of cytochrome C (Ikai and Tanford 1971). More recently, Kiefhaber and coworkers used an interrupted refolding and unfolding stopped-flow method to explore the folding mechanism of lysozyme (Kiefhaber 1995). This method enables one to monitor specifically the amounts of native and intermediate states that accumulate during the protein folding process. The results allowed the authors to distinguish between parallel kinetic partitioning and sequential pathways in a three-state folding mechanism. Maki and Roder used continuous-flow rapid mixing techniques in conjunction with stopped-flow methods to detect transient formation of intermediate states in the folding of staphylococcal nuclease (Maki, Cheng et al. 2007). Marion and Wüthrich revolutionized studies of kinetic intermediates by developing two-dimensional NMR methods, leading to hydrogen-deuterium exchange studies to investigate conformational change in proteins (Marion and Wüthrich 1983; Strop, Wider et al. 1983). The observation of an intermediate in a folding mechanism raises the question: is it on or off the pathway from the unfolded to folded state? Many studies have attempted to answer this question in a
specific protein system (Baldwin 1996 for review). Radford and coworkers showed that Im7 forms a transient on-pathway intermediate (Capaldi, Shastry et al. 2001) and Mierlo and Bollen demonstrated that both on-pathway and off-pathway intermediates are present in the apoflavodoxin folding process (Bollen, Sánchez et al. 2004).

In the search for folding intermediate states, especially for the off-pathway ones that often have non-native structural interactions, some workers have discovered cases of protein misfolding. Some partially folded or misfolded intermediate conformations are prone to form large oligomeric aggregations and subsequently amyloid fibrils (Dobson 2003). Elucidation of the process of amyloid fibril formation from these intermediates has provided insights into their role in neurodegenerative diseases (Horwich 2002; Chiti and Dobson 2006) and is the basis of drug development efforts to mitigate the health effects of protein misfolding (Guerois and López de la Paz 2006).

1.3 Intrinsically unstructured proteins

Intrinsically unstructured proteins (IUP) are an expanding class of proteins that lack compact globular structure or contain long disordered regions under physiological conditions (Tompa 2002; Uversky 2002). They were initially identified theoretically by sequence analysis (Romero, Obradovic et al. 1997) as well as experimentally through spectroscopic experiments (Kriwacki, Hengst et al. 1996). Prediction of disordered regions through bioinformatic analyses has reached a high degree of accuracy and numerous databases are available for the prediction of disordered regions from protein sequence (Radivojac, Iakoucheva et al. 2007). These theoretical and
bioinformatic analyses also showed that IUPs comprise possibly as many as 30% of all eukaryotic proteins (Dunker, Brown et al. 2002), and generally exhibit low hydrophobicity and high net charge. Moreover, most predicted disordered regions are characterized by low sequence complexity and an amino acid compositional bias (Tompa 2005).

Experimental studies have shown that IUPs are involved in many crucial regulatory mechanisms in vivo (Dyson and Wright 2002) and can be generally divided into four categories (Dunker, Brown et al. 2002). (1) Molecular recognition by disordered proteins may be primarily used for signaling. The ability of an IUP to bind to multiple targets with low affinity is an ideal property for a signaling pathway component. (2) Molecular assembly processes that utilize partially folded subunits appear to have significant advantages compared to those which exclusively utilize ordered subunits (Namba 2001). (3) Sites of protein modification occur with strong preference for regions of disorder, which are presumably more accessible to chemical modification or protease cleavage. (4) Entropic chains perform functions that depend directly on the properties of a disordered conformation like the flexibility of a domain linker/spacer.

Most IUPs participate in transcriptional and translational regulation, signal transduction, cell cycle regulation and the regulation of large multiprotein complex self-assembly (Wright and Dyson 1999; Dyson and Wright 2005). The flexibility of IUPs enables them to bind to multiple targets, which is important in cellular regulation. It is possible that their unstructured property might facilitate rapid turnover, providing another mechanism for rapid regulatory control when cells respond to environmental changes
(Fink 2005). One prominent feature of many IUPs is the disorder-to-order transition they undergo when they bind to their biological targets. For example, the high mobility group (HMG) domain of the lymphoid enhancer-binding factor LEF-1 is disordered in the absence of its cognate DNA (Love, Li et al. 1995). The KID (kinase-inducible transcriptional-activation domain) of CREB is intrinsically unstructured but when it binds to its target domain in CBP (CREB-binding protein) it folds to a pair of orthogonal helices (Dyson and Wright 2005). The transition is a thermodynamic consequence of the entropy that is lost during the folding process being compensated by the binding enthalpy.

### 1.3.1 Experimental characterization of IUPs

Structural information about conformations of IUPs both free in solution and in complexes with targets has been obtained primarily by CD, fluorescence, and NMR spectroscopy (Dyson and Wright 2005). Many other experimental techniques have been employed to investigate the structural content and conformational ensemble of the disordered states. These methods include small angle X-ray scattering (SAXS) (Bernadó, Mylonas et al. 2007), paramagnetic spin labeling (Dedmon, Lindorff-Larsen et al. 2005), ensemble molecular dynamics and single molecule fluorescence (Mukhopadhyay, Krishnan et al. 2007). Recently, Wells, et al. (2008) combined the NMR-based residual dipolar coupling technique and SAXS to investigate the structural ensemble of IUPs. The results yielded a model of the conformational ensemble of disordered N-terminal transactivation domain of full-length p53 tumor repressor alone and in a complex with a specific DNA response element. High-speed atomic force microscopy (Miyagi, Tsunaka
et al. 2008) and Raman optical activity (Zhu, Tranter et al. 2006) have also been used to characterize and classify IUPs and their dynamic properties.

IUPs not only perform variety biological functions in the cell, but many of them are also intimately related to human disease (Fink 2005; Dunker, Silman et al. 2008). Well known examples are the Aβ peptide and tau protein in Alzheimer’s disease, PrP in prion disease. These are neurodegenerative diseases caused by aggregation of misfolded proteins. Therefore, characterization of the folding kinetics and structural properties of IUPs is as important as studies of well-ordered proteins.

1.4 Linked conformational change and ligand binding

As described in the previous section, the disorder-to-order transition is an important feature of IUPs. This transition is often accompanied by the binding of cognate ligands or targets in vivo, so the folding and binding events are coupled to each other. One of the most intriguing questions related to the coupled folding and binding process of IUPs or disordered domains in proteins concerns the mechanism of this process. Does folding occur before binding or does binding occur before folding? Two extreme schemes can describe these two processes respectively: conformational selection or induced folding. In the conformational selection mechanism, target ligands select a conformation closely approximating that of the bond form from the ensemble of conformations populated by the disordered protein when free in solution. In the induced folding mechanism, the protein binds to its target ligand in a fully disordered state, and folding is induced by association with the ligand. Various studies of IUP systems have
demonstrated that each mechanism is valid in different systems (Ferreira, Hermann et al. 2005; Sugase, Dyson et al. 2007; Onitsuka, Kamikubo et al. 2008; Song, Guo et al. 2008). Examples have been reported where increasing the stability of pre-formed secondary structure fails to enhance the binding affinity or can even slow the kinetics of complex formation (Bienkiewicz, Adkins et al. 2002; Zor, Mayr et al. 2002). This observation suggests that disorder in the unbound protein may actually be advantageous for the binding process.

Theoretical models and simulation studies have suggested that the rate of macromolecular association can be increased by the presence of disordered regions in the molecules through the fly-casting mechanism (Pontius 1993; Shoemaker, Portman et al. 2000). The mechanism postulates that an unstructured protein has a greater “capture radius” and facilitates the diffusive search for a binding partner (Shoemaker, Portman et al. 2000). A concept implicit in the fly-casting mechanism is that flexibility of the intrinsically unstructured fragment enables an initially weak and non-specific binding, followed by the folding and specific binding to the cognate target. Several theoretical studies demonstrated that fly-casting plays an important role in coupled folding and binding of IUPs (Levy, Onuchic et al. 2007; Chen 2009), but few experimental studies have been carried out to support the model. A recent laser temperature jump analysis of the kinetics of an IUP system (Narayanan, Ganesh et al. 2008) suggested a coupled folding and binding mechanism consistent with the fly-casting mechanism. Based on the increase of the binding rate constants of unstructured proteins with their substrates from the fly-casting model, the authors concluded that IUPs should prefer the binding before
folding pathway in the coupled folding and binding mechanism. However, it needs to be pointed out that rate constant alone cannot properly describe the overall pathways in a mechanism. In this work, we use the concept of flux, which involves the rate constants and the population of each state in the mechanism to determine the predominant pathway in any coupled folding and binding process.

Generally speaking, protein folding coupled to ligand binding is an extreme example of conformational change upon ligand binding. Many proteins perform their biological functions through local structural rearrangements induced by ligand binding. The allosteric binding of oxygen to hemoglobin is a good example. Studies have also shown that the conformational change upon ligand binding is also important in the signal transduction pathway through G protein-coupled receptors (Hauser, Kauffman et al. 2007). The fundamental principles used to describe the coupled folding and binding process are readily applicable to any system involving conformational change coupled to binding equilibria.

1.5 Ribonuclease P holoenzyme

RNA molecules perform a variety of biological functions in the living cell. In addition to acting as a messenger in the transfer of genetic information from DNA to cellular proteins, RNA functions include gene regulation, RNA processing and protein translation. The catalytic center of the ribosome resides in the RNA component, which acts as a biocatalyst and performs multiple turnover activity on substrates like enzymatic proteins do. The first RNA molecules demonstrated to carry out a catalytic function were the Group I intron derived from tetrahymena (Cech, Zaug et al. 1981) and the RNA
subunit of Ribonuclease P (RNase P) (Guerrier-Takada, Gardiner et al. 1983). Since then several catalytic RNAs or ribozymes have been identified (Doudna and Cech 2002 for review). The biological role of the Group I intron is to operate in \textit{cis}, which means the catalyst and substrate are in the same RNA molecule. In contrast, RNase P performs its function in \textit{trans} and was the first \textit{trans} ribozyme to be identified.

RNase P holoenzyme is a ribonucleoprotein (RNP) complex, and RNPs are involved in a number of diverse and essential biological functions, such as protein synthesis (ribosome), RNA editing and post-translational processes (Weeks 1997; Perez-Canadillas and Varani 2001; Hall 2002). Other examples of RNPs include the spliceosome and telomerase. The spliceosome is involved in recognition of the intron-exon border and participates in the chemistry of splicing (Staley and Guthrie 1998). It includes five small nuclear RNAs (snRNA), each of which is complexed with at least seven protein subunits to form an snRNP (small nuclear ribonucleoprotein). Telomerase, which catalyzes the synthesis of telomeric DNA, consists of a reverse transcriptase protein subunit and an RNA subunit that acts as the template (Greider and Blackburn 1987). Among different ribonucleoprotein complexes, the associated RNA binding proteins are functionally diverse. These proteins serve to stabilize, protect, transport, catalytically modify RNA and can mediate RNA interactions with other macromolecules (Cusack 1999).

A general feature of the formation of RNP complexes is that it usually involves conformational changes in the RNA, protein or both. For example, when the ribosomal protein L11 binds to a four-way junction in 23S rRNA a disordered loop
region in the free protein becomes ordered at the RNA interface upon binding (Williamson 2000). The N-terminal RNA binding domain of the U1A protein binds to an internal loop in the 3’ untranslated region of its own mRNA. Binding of U1A protein to its mRNA changes the orientation and conformation of helix C in the protein (Varani, Gunderson et al. 2000). This conformational change forms the interface between U1A protein and poly(A) polymerase, which regulates the subsequent polyadenylation process. Many other RNPs also use this coupled conformational change/binding reaction to enhance their specificity and catalytic activity (Leulliot and Varani 2001).

1.5.1 P RNA subunit

Transfer RNA (tRNA) genes are transcribed as precursors and need to be processed by a series of post-transcriptional modifications to produce functional tRNA molecules. Ribonuclease P catalyzes the metal-dependent 5’ cleavage of the precursor tRNA (pre-tRNA) to mature tRNA with a monophosphate at the 5’ terminus after cleavage (Frank and Pace 1998; Kurz and Fierke 2000). RNase P is a ubiquitous ribozyme found in all kingdoms of life and, irrespective of origin, RNase P is composed of one RNA component and a varying number of protein subunits. There are ten protein subunits in mammalian RNase P, and nine in yeast. In archaea, there are more than four protein subunits (Hsieh, Andrews et al. 2004). In bacteria, RNase P holoenzyme contains only one RNA and one protein subunit. Recent studies also showed that RNase P in human mitochondria lacks the RNA component and is an entirely protein-based enzyme (Holzmann, Frank et al. 2008; Walker and Engelke 2008). Moreover, Randau et. al. (2008) showed that the RNase P gene could not be identified in the archaeon,
*Nanoarchaeum equitans*, through both extensive computational analysis of the genome and biochemical tests in the cell extracts. These studies also revealed that the tRNA genes in the organism are lacking a 5’ leading sequence and therefore do not require RNase P activity. These studies demonstrated how nature dealt with the loss of RNase P under the pressure of genome condensation through genomic rearrangement of tRNA genes. The diversity of the RNase P composition between different organisms is a subject of active interest in discussions of the evolution of a hypothesized “RNA world” to the “protein world”.

Phylogenetic analysis of bacterial RNase P RNA (P RNA) suggests that all forms fold to a common overall secondary structure with a conserved core. Based on these secondary structure comparisons, bacterial P RNA can be divided into two main types, type A (Ancestral type) and type B (*Bacillus* type) (Brown, Haas et al. 1991). Each type of P RNA can be divided into two structural domains: the substrate binding domain (S-domain) and catalytic domain (C-domain) (Pan 1995). The S-domain contains the binding site for the T-stem-loop region (TSL-region) of the tRNA precursor substrate, and the C-domain carries out the catalysis. These two domains can fold independently and the C-domain alone can perform the catalytic function. High resolution x-ray crystal structures of the S-domain of both type A (*Thermus thermophilus*) (Krasilnikov, Xiao et al. 2004) and type B (*Bacillus subtilis*) P RNA (Krasilnikov, Yang et al. 2003) have been solved. Although the S-domains of these two types have secondary and tertiary structures differences, the functionally important region is conserved and likely preserved for pre-tRNA recognition. More recently, the crystal structures of full length P RNA from type A
(Thermatoga maritima) (Torres-Larios, Swinger et al. 2005) and type B (Bacillus stearothermophilus) (Kazantsev, Krivenko et al. 2005) have been determined. Although only two-thirds of the P RNA structure is well resolved in these structural models, comparison of the catalytic domains of the two types of P RNA also reveals similar structural features in the core region of the domain.

Altman and co-workers showed that bacterial P RNA lacking its protein subunit is catalytically active in vitro in the presence of high salt concentrations (Guerrier-Takada, Gardiner et al. 1983). Subsequent studies demonstrated RNase P RNA from some Archeaea can also mediate correct cleavage in the absence of protein subunits under certain in vitro conditions (Pannucci, Haas et al. 1999). Kikovska and coworkers recently showed that P RNA from both human and the lower eukaryote, Giardia lamblia, can perform pre-tRNA substrate cleavage at low pH (Kikovska, Svärd et al. 2007) albeit at rates that are $10^6$-fold lower than that of E. coli RNase P RNA. These results collectively suggest that the RNA-based catalytic function of RNase P has been preserved during evolution.

Like many other RNA molecules, RNase P has a structure and catalytic activity tightly linked to metal ion concentration. The divalent metal ion magnesium is especially important (Beebe, Kurz et al. 1996) because it promotes the correct folding of the RNA, facilitates the interaction with substrate, and participates in the chemistry of the cleavage reaction. In the case of P RNA, approximately 100 Mg$^{2+}$ ions (Beebe, Kurz et al. 1996) bind to one molecule of RNA. The locations of Mg$^{2+}$ binding sites critical for catalysis are yet to be clearly determined. The optimal magnesium concentration for
RNase P holoenzyme activity is lower than that of P RNA ribozyme alone. It is also not clear how P protein enhances the affinity of P RNA for metal ions. Current structural models of RNase P holoenzyme (see below) place the P protein in the close proximity of P4 helix of P RNA, which has been proposed to participate in metal binding at the putative catalytic site of the enzyme. P protein can influence these critical metal ion interactions both directly by binding the ions and indirectly by binding P RNA. Another possible source of metal-dependent activity is non-specific binding of the metal ions to P RNA and pre-tRNA via electrostatic interactions. These interactions may stabilize the secondary and tertiary structure of P RNA and the formation of a negative charge in the cleavage reaction.

Precursor tRNAs are the main substrates for the RNase P, but there are a number of other RNA molecules that are processed by RNase P. These substrates possess a structural motif that has a short 5′ single stranded RNA flanking a short duplex RNA stem. Examples of in vivo RNase P substrates include: 4.5S RNA; tmRNA; bacteriophage M3 RNA and phage-derived antisense C4 RNA; mRNA; and E. coli RNA derived from non-coding intergenic regions (Bothwell, Garber et al. 1976; Bothwell, Stark et al. 1976; Komine, Kitabatake et al. 1994; Hartmann, Heinrich et al. 1995; Li and Altman 2003). Because pre-tRNAs are the preferred substrates for RNase P in vivo, three major productive interactions between the pre-tRNA and P RNA for recognition are listed below. (1) The interactions between the T-Stem-Loop (TSL) region in the pre-tRNA and the P7-P11 region in the S-domain. (2) The RCCA-P RNA interaction where the RCC-motif at the 3’ end of the pre-tRNA pairs with a conserved GGU-motif in the P15-loop in
the P RNA. (3) The A248/N-1 interaction where the residue immediately 5′ of the scissile bond interacts with residue A248 in P RNA. Until now, no co-crystal structure of any form of tRNA bound to P RNA complex has been determined. The putative interactions discussed above rely solely on genetic and biochemical data.

1.5.2 RNase P protein subunit

Although all P RNA subunits from different organisms show some structural homology, the absence of similarity in the protein subunits of bacterial P protein, archaeal P protein and eukaryotic P protein remains to be explained from an evolutionary point of view. The structure of the protein subunit of bacterial RNase P (P protein) from different species has been extensively studied. The crystal structure of P protein from *B. subtilis* (the protein used in all our studies) has been determined to 2.6 Å (Stams, Niranjanakumari et al. 1998). In addition, the solution structure of *S. aureus* P protein (Spitzfaden, Nicholson et al. 2000) and crystal structure of *T. maritima* P protein (Kazantsev, Krivenko et al. 2003) have been determined. Although the overall sequences of these P proteins are not highly conserved (20-30% amino acid identity), they all share highly structural similarity. This structural similarity results in functional similarity: P RNA and P proteins from these bacteria can be reconstituted in vitro to form active heterologous complexes (Guerrier-Takada, Gardiner et al. 1983). The general structure of the protein is shown in Figure 1-1: it is a mixed alpha-beta fold composed of four beta strands flanked by three alpha helices. Three RNA binding regions have been proposed in P protein (Niranjanakumari, Stams et al. 1998): (1) the central cleft formed by helix A and the face of the central beta sheet. Several conserved, solvent exposed, aromatic
residues are at the base of the cleft. Aromatic residues can form specific or nonspecific ring stacking interactions (Ishida, Doi et al. 1988; Kaushik, Singh et al. 1999) with the base of single-stranded RNA and DNA molecules. Studies indeed have shown that P protein has direct contact with the 5’ leader sequence of pre-tRNA substrate through this central cleft (Niranjanakumari, Stams et al. 1998). Other residues in and around the central cleft also showed crosslinking activity with the pre-tRNA leader sequence. (2) an ~18 residue RNR motif that connects beta strand 3 to helix B (colored in red in Figure 1-1). This stretch is the most conserved sequence in bacterial RNase P protein and has a consensus sequence of K-X_{4,5}-A-X_2-R-N-X_2-(K/R)-R-X_2-(R/K) (Brown 1999). The RNR motif, which encompasses the N-terminal half of helix B and its preceding loop, forms a left-handed crossover between parallel β-strands 3 and 4. This topology is unusual because most crossover connections between parallel β-strands are right-handed (Richardson 1985). (3) a metal binding loop (colored in blue in Figure 1-1) connecting β-strand 2 and 3 where aspartate and glutamate residues may mediate ion binding with RNA. Because of the highly basic nature of RNase P protein, this clustering of negatively charged residues is interesting.

Although bacterial RNase P RNA is catalytically active in the absence of P protein at high salt concentration *in vitro* the protein subunit is essential *in vivo* (Guerrier-Takada and Altman 1984), indicating that P protein can decrease the monovalent and divalent cation requirement to physiological conditions. Because of this essential property of P protein *in vivo* and its interactions with both P RNA and pre-tRNA, possible functions of P protein include: assisting P RNA in discriminating
between substrate and product by directly binding to the 5′ leader sequence of pre-tRNA; stabilization of the catalytically active conformation of P RNA (Hsieh, Andrews et al. 2004); mediation of holoenzyme dimerization (Fang, Yang et al. 2001); and activation of intrinsic RNA properties that perform holoenzyme functions (Buck, Dalby et al. 2005). These potential roles are not mutually exclusive and there is evidence to support each.

Eukaryotic RNase P holoenzymes from both yeast and human have been purified (Chamberlain, Lee et al. 1998; Jarrous 2002). Although the specific functions of the eukaryotic RNase P protein subunits are not currently known, potential roles include those proposed for the bacterial P protein. Because eukaryotic RNase P proteins represent a significant portion of the mass in the RNase P holoenzyme (~70%), specific interactions between among the protein components are likely. Protein-protein interactions of yeast and human RNase P have been studied by yeast two-hybrid genetic screens (Jiang and Altman 2001; Houser-Scott, Xiao et al. 2002). Each of the protein subunits of yeast RNase P has been found to make contact with at least one other protein component of the holoenzyme. Genetic knockout experiments have also shown that each of the nine yeast RNase P protein subunits is essential for viability (Chamberlain, Lee et al. 1998)

A bioinformatic search of several archaeal genome sequences for a gene encoding RNase P protein similar to that in bacteria was negative. On the other hand, open reading frames with homology to four eukaryal RNase P proteins, Rpr2, Pop4, Pop5 and Rpp1 (gene terminology from yeast), have been indentified and subsequently shown to be associated with RNase P activity. In contrast to the eukaryal RNase P proteins, four
RNase P protein subunits from different archaeal species have been determined using either NMR or X-ray crystallography (Sidote, Heideker et al. 2004; Takagi, Watanabe et
al. 2004; Kakuta, Ishimatsu et al. 2005; Wilson, Bohlen et al. 2006). The solution and
crystal structures of the protein subunit, Pop5, from *Pyrococcus furiosus* showed a high
degree of 3-D structural similarity to bacterial P protein. This similarity is particularly
remarkable considering the differences in the proteins’ primary sequence and secondary
structure topology, $\beta\alpha\beta\alpha\beta\alpha$ in Pfu Pop5 versus $\alpha\beta\beta\alpha\beta\alpha$ in the bacterial P protein,
suggests that the two proteins have different evolutionary origins. Biochemical studies
also showed that the four archaeal P proteins could be recombinantly expressed and
reconstituted *in vitro* with corresponding P RNA subunit to yield an active holoenzyme
(Boomershine, McElroy et al. 2003; Kouzuma, Mizoguchi et al. 2003).

1.5.3 The bacterial holoenzyme complex

Although the individual structures of P RNA and P protein from bacterial
RNase P and tRNA have been determined, the structure of holoenzyme alone and
holoenzyme-substrate complex are currently still unavailable. Because of P protein’s
functional importance and potential proximity to the holoenzyme active site, many
studies have attempted to position the P protein within the RNase P holoenzyme. These
studies have relied predominantly on Fe(II)-EDTA footprinting and chemical protection
(Chen, Nolan et al. 1998; Massire, Jaeger et al. 1998; Biswas, Ledman et al. 2000; Tsai,
Masquida et al. 2003). The orientation of P protein with respect to P RNA has also been
investigated using EPR and site-specific spin labeling (Gopalan, Kühne et al. 1999). In
addition to these P RNA-P protein interactions, the central cleft in P protein were shown
to crosslink to the -4 to -8 positions within the leading sequence of pre-tRNA substrate
(Niranjanakumari, Stams et al. 1998). Because the Fe atom is 14 Å from the $C_\alpha$ atom of
Cys on the backbone and Fe(II)-EDTA produces diffusible radicals, these biochemical results provide a low-resolution model for the position of P protein in the holoenzyme. Recently, two studies using in-gel phosphothioate-iodine footprinting (Buck, Kazantsev et al. 2005) and copper (II)-phenanthroline (Cu-OP) hydroxyl-radical cleavage (Niranjanakumari, Day-Storms et al. 2007) provide more information regarding the contact between the P RNA-P protein and holoenzyme-substrate complex. Because the phosphothioate is incorporated in the RNA backbone and Cu-OP generates localized hydroxyl-radicals through direct contact, these two methods facilitate identification of specific P RNA nucleotides located near a predefined P protein amino acid in the RNase P-pre tRNA complex. The results also provide important distance constraints complementary to photo-crosslinking data. Figure 1-2 shows an overall P RNA-P protein and P RNA-substrate complex model and some crucial interactions between P protein and P RNA based on the biochemical results. From the protein perspective, the RNR motif interact with P2, P3 and P4 regions of P RNA and is also close to the active site. The N terminus interacts with P3 and L3 regions of the P RNA and the metal binding loop interacts with L3, P19 and J18/2 regions. One interesting result is that K52 and K89 are close to the bases 65-67 in the pre-tRNA acceptor arm. Determination of the co-crystal structure of the RNase P holoenzyme-substrate complex will provide more insight regarding the function of each component and the process of the catalytic reaction.
Figure 1-2: The structure of *B. stearothermophilus* RNase P. (A) Tertiary model of the bacterial holoenzyme, shown in ribbon representation of P protein. The relative position between P RNA and P protein is modeled by the footprinting data from (Buck, Kazantsev et al. 2005). (B) Respective side and front view of the reconstruction of *B. stearothermophilus* P RNA with tRNA. The plots are adopted from the reference (Evans, Marquez et al. 2006).
1.6 Overview of this work

The main goal of this work is to characterize the kinetic and thermodynamic properties that govern the protein folding processes in the bacterial RNase P protein. Biophysical studies have characterized P protein as an intrinsically unstructured protein that can be induced to fold by addition of various small anions (Henkels, Kurz et al. 2001). Additionally, stoichiometric binding experiments showed that small anions have two binding sites in P protein. Since this coupled folding and binding property of P protein exhibits an extreme example of the conformational change/binding in the RNA-protein complex and most of the IUPs have the same features, fundamental understanding of the kinetic mechanism for this process can provide insight about how the protein folds and complexes assemble inside the cell.

In Chapter 2, osmolyte induced folding of P protein is carried out using trimethylamine-N oxide (TMAO) to dissect the P protein folding reaction out of the binding reaction, since TMAO favors the P protein in its folded state without involving any binding event. Stopped-flow experiments and equilibrium titration measurements were employed to obtain the kinetic and equilibrium information of the folding process of P protein. Global analysis of kinetic and equilibrium data enable the estimation of kinetic and thermodynamic parameters and the discovery of the intermediate state in the P protein folding mechanism.

The structural heterogeneity of P protein both in the absence and presence of ligand is interesting but is not fully understood. In Chapter 3, NMR spectroscopy is employed to characterize the intermediate state found in the kinetic and equilibrium
studies. The results provide a structural description of the intermediate state and its potential roles in the holoenzyme assembly process.

In Chapter 4, a detailed model of the coupled folding and binding mechanism of P protein is developed to determine which of the six possible pathways predominate. A novel flux calculation is used to interpret the results of ligand-induced folding kinetics of P protein. This flux concept cannot only shed light on the RNase P holoenzyme assembly process in vivo but also many coupled conformational change and substrate binding processes in the biology.
Chapter 2: Using Osmolyte to Study Folding Kinetics and Equilibria of Bacillus subtilis RNase P Protein

2.1 Introduction

A fundamental understanding of the protein folding process will help to predict protein structure from sequence, to elucidate the biological function of proteins in vivo and to understand how and why proteins misfold. Previous studies have demonstrated that many small proteins exhibit simple two-state folding kinetics (Jackson 1998). In this simple two-state model, the two thermodynamically distinguishable ensembles (native and unfolded states) are in equilibrium and can undergo cooperative conformational change as they interconvert. Because this conformational transition usually involves a significant change in solvent accessible surface area (ASA), chemical denaturants perturb the equilibrium and rate constants associated with the reaction. Because of its linear dependence on denaturant concentration, the stability of a two-state protein can be estimated in terms of relative free energy difference. The logarithms of rate constants for folding and unfolding of a two-state protein are also linear functions of denaturant, which allows extrapolation to 0 M denaturant. Thus, proteins that fold via a two-state mechanism are particularly amenable to thermodynamic and kinetic characterization.

However, many proteins do not exhibit two-state folding but instead fold via one or more partially folded intermediates. Previous studies have shown that intermediate states can attain the native state with two types of mechanisms (Baldwin 1996). One class of intermediate states develops native-like interactions, which accelerate
the search of the native state, and can be called an on-pathway folding intermediate (Capaldi, Shastry et al. 2001). The other kind of intermediate is a kinetically trapped misfolded species, which can slow the folding process and is usually referred to as an off-pathway intermediate. Although intermediate states can be crucial in the overall folding process, it is sometimes challenging to detect and characterize them. Many intermediates form rapidly within the dead-time of a conventional stopped-flow instrument. Transient kinetic intermediates are generally not stable enough to be observed at equilibrium under strongly denaturing conditions. The intermediates may also be spectroscopically silent during the kinetic experiments. Absence of the intermediate states from the experimental results do not guarantee that there is truly no intermediate state in the folding process, so finding conditions that maximize the population and stability of the intermediate states become critical to study their thermodynamic, kinetic and even structural properties.

Although previous kinetic studies have focused primarily on natively structured globular proteins, a biologically important class of proteins called intrinsically unstructured proteins (IUP) have recently become widely recognized (Dyson and Wright 2005). It had long been accepted that the folded structure of a protein determines its function, but most IUPs lack definite structure in the absence of bound ligand(s). They adopt active functions through disorder-to-order transitions coupled to binding of other cellular components and this coupled folding and binding phenomenon is a major feature of the IUP family (Dyson and Wright 2005). Coupling a folding transition to a molecular interaction has been considered functionally advantageous because it may allow a protein to recognize different target substrates, could enhance the rate of a intermolecular
interactions, and could provide for large intermolecular interfaces with a relatively small protein. Recent studies have shown that many IUPs participate in the transcriptional and translational regulation (Wright and Dyson 2009). Because of the structural flexibility of IUPs, studies also show they also could work as a multimolecular hub that interacts with multiple target proteins (Dunker, Oldfield et al. 2008).

Although there have been many studies of biological functions of IUPs, there have been many fewer fundamental biochemical and biophysical studies of properties such as the folding kinetics. The folding/unfolding processes of IUPs and natively structured proteins are equally important because they each favor the opposite side of the folding reaction. From the chemical reaction point of view, the theory and methods behind these two kinds of protein folding processes are fundamentally the same. The main difference is the experimental design: whereas natively structured proteins require denaturant such as urea or guanidinium chloride to perturb the equilibrium, IUPs require “protection osmolyte.” In the present study, an osmolyte, trimethylamine-N oxide (TMAO), accompanied with the denaturant, urea, was used to perturb the equilibria between the folded and unfolded states of an IUP.

Osmolytes are small organic molecules such as certain amino acids, polyols, and methylamines that are synthesized in different organisms (Bolen and Baskakov 2001) to protect proteins against urea denaturation \textit{in vivo} under environmental stress. Many studies have used different osmolytes to stabilize the folded state of proteins (Lin, Zarrine-Afsar et al. 2009) and RNA (Lambert and Draper 2007) secondary and tertiary structures. The major mechanism of protein stabilization by osmolyte is through
unfavorable interaction with the backbone atoms, which increases the free energy of the
denatured state (Bolen and Rose 2008). Most osmolytes are interchangeable and additive,
which means the protecting effects can be achieved by a variety of osmolytes and can
counteract the destabilizing effects of urea on protein stability. Studies also show that the
free energy of unfolding is linearly dependent on osmolyte concentration, analogous to
the linear dependence of the unfolding free energy on chemical denaturant (Baskakov and
Bolen 1998). This property is particularly useful for estimation of unfolding free energy
of IUPs and partly folding proteins (Mello and Barrick 2003). IUPs can be titrated with
osmolyte to generate a refolding curve from which the (positive) folding free energy can
be estimated using the same linear extrapolation method used to analyze a urea or
guanidine denaturation curve. The stability of a partly folded protein can be measured by
mixing both urea and osmolyte solvent in protein sample to generate complete denatured
or native state baselines in the titration curve (Mello and Barrick 2003).

Ribonuclease P holoenzyme (RNase P) is the endonuclease that catalyzes
the maturation of pre-tRNA by removal of the 5’ leader sequence (Frank and Pace 1998;
Kurz and Fierke 2000). The protein subunit (P protein) in *Bacillus subtilis* RNase P has
119 residues, and the crystal structure of P protein has been determined to 2.6Å
(Niranjanakumari, Stams et al. 1998). The general structure of the protein is a mixed
alpha-beta fold composed of four beta strands flanked by three alpha helices. Three RNA
binding regions have been proposed in the P protein (Stams, Niranjanakumari et al.
1998): 1) the central cleft formed by helix A and the face of the central beta sheet; 2) an
~18 residue conserved RNR motif which connects beta strand 3 to helix B; and 3) a metal
binding loop connecting beta strand 2 and 3 where aspartate and glutamate residues may mediate ion binding with RNA. Several structural roles of P protein have also been proposed including stabilizing the structure of P RNA or compensating for charge repulsion (Hsieh, Andrews et al. 2004). Recent studies have shown that P protein has direct contact with the 5' leader sequence of pre-tRNA substrate (Crary, Niranjanakumari et al. 1998; Kurz, Niranjanakumari et al. 1998; Niranjanakumari, Kurz et al. 1998; Niranjanakumari, Stams et al. 1998). Therefore, the role of P protein in RNase P is to enhance substrate recognition and stimulate the intrinsic RNA activity that performs holoenzyme functions (Buck, Dalby et al. 2005). Our previous biophysical studies of P protein demonstrated that it is intrinsically unstructured in the absence of ligand and at low ionic strength. The highly basic amino acid composition (11 Arg, 19 Lys, and 3 His out of 119 residues) of P protein might be part of the reason for its shift to the unfolded state under these conditions due to greater charge-charge repulsion in the more compact folded state.

In this study, P protein was used as a model system to elucidate the folding/unfolding process of intrinsically unstructured proteins. Stopped flow fluorescence was used to monitor TMAO-induced folding of P protein. Urea and TMAO co-solvent equilibrium co-titration monitored by circular dichroism and fluorescence was developed to estimate the parameters defining the equilibrium between the unfolded and native state of P protein. Both kinetic and equilibrium data show the presence of an intermediate state in the P protein folding pathway. A sequential three-state mechanism was modeled to fit the experimental data. Kinetic and equilibrium data were fitted
globally to extract the rate constants, thermodynamic parameters and relative spectroscopic properties between these three states. A significantly populated intermediate state was present under the unfolding conditions and the folding rate constants between unfolded and intermediate state. To our knowledge, these are the first folding and unfolding rate constants determined by TMAO induced folding kinetics for any intrinsically unstructured protein.

2.2 Results and Discussion

2.2.1 Tryptophan variants of P protein as the probe to study kinetics and thermodynamics of P protein

Wild type B. subtilis P protein does not contain a tryptophan residue, which is often used as an intrinsic fluorescent probe to monitor protein conformational changes. To utilize this property for fluorescence-detected folding kinetic and equilibrium studies, a tryptophan containing variant of P protein, F107W, was generated. The site for mutagenesis was selected based on a sequence comparison between B. subtilis P protein and the highly homologous (similarity) E. coli P protein, C5. In E. coli P protein, tryptophan 107 is highly conserved among different gram-negative bacteria. The Trp residue fits in a hydrophobic pocket and has a fluorescence change upon urea unfolding (Gopalan, Golbik et al. 1997). Figure 2-1 shows a model of the tryptophan sidechain using crystal structure of B. subtilis P protein. The Trp sidechain fits well and
Figure 2-1: Model of Trp 107 sidechain contacts with surrounding sidechains. The model was constructed using the crystal structure of B. subtilis P protein (PDB 1A6F) and the Sidechain Mutator Tool in King (Davis, Murray et al. 2004). The dots generated by Probe (Word, Lovell et al. 1999) indicate the contact surface of the Trp sidechain with the surrounding atoms. Red spikes represent steric clashes between atoms. The blue to green dots indicate a favorable Van der Waals contacts between each atom. The model shows that the engineered Trp fit well in this hydrophobic pocket and probably does not greatly perturb the overall structure. The modeled Trp sidechain rotamer is m0 with a minor adjustment of the chi2 angle.

Molprobity analysis (Davis, Murray et al. 2004) indicates good Van der Waals contacts with surrounding sidechains. The F107W variant shares the IUP property with wild type P protein and could be induced to fold upon small anion binding as observed in previous studies of wild type P protein. Addition of TMAO to the protein sample also shifts the folding equilibrium to the folded state. The fluorescence emission spectrum (285 nm excitation) of F107W from 0 M to 1 M TMAO is shown in Figure 2-2A. Unfolded
F107W has a \( \lambda_{\text{max}} \) at 358 nm, and the peak is blue-shifted to 308 nm when the protein folds to the native state, indicating that the tryptophan sidechain is located in a nonpolar region of the folded protein. (Szabo, Stepanik et al. 1983; Lakowicz 1999). The total fluorescence also decreases upon folding, suggesting that there is additional quenching in the native protein. The wild type P protein crystal structure supports these conclusions, but the Trp is not in a completely nonpolar environment. As shown in Figure 2-1, the modeled Trp sidechain is surrounded by 4 Leu, 1 Phe, 2 Tyr, 1 Gln, and 1 Glu residues. Adjacent aromatic sidechains are known to quench Trp fluorescence. Figure 2-2B shows the CD spectrum of the variant from 0M up to 1 M TMAO. The appearance of distinct peaks around 220 and 208 nm indicates the formation of \( \alpha \)-helical secondary structure at [TMAO] > 0.2 M. The CD spectrum of P protein in 1 M TMAO is similar to that found for 20 mM sulfate, in which P protein is predominantly folded (Henkels, Kurz et al. 2001). The NMR HSQC spectrum of F107W (data not shown) is similar to the wild type spectrum, and more than 80% of the peaks in the F107W spectrum can be assigned by comparing with the wild type spectrum. Most of the shifted resonances correspond to amides in close proximity to the site of the Trp substitution.

Taken together, these experiments show that TMAO can induce F107W P protein to fold to a structure similar to that of folded wild type P protein. The fluorescence of the Trp residue is sensitive to the structural change between folded and unfolded states, making it useful for equilibrium and kinetic studies of the P protein folding mechanism.
Figure 2-2: Circular dichroism and fluorescence spectrum of the F107W variant of P protein in the presence of different TMAO concentrations. (A) Far-UV CD spectrum of F107W variant of P protein (10µM, 1mm cuvette) in 20mM cacodylate buffer in 0, 0.2, 0.4, 0.6, 0.8 and 1M TMAO. The spectra were scanned from 260 to 200 nm in 1nm increments. (B) Fluorescence spectrum of F107W variant of P protein (5µM, 1cm cuvette) in the same TMAO concentrations used for CD. The excitation wavelength was 285nm and the emission spectra were scanned from 400 to 300 nm in 1nm increments.
TMAO induced transient kinetic traces were collected using the stopped-flow method. As expected from the equilibrium results, the fluorescence signal decreases when folding is induced by mixing protein in buffer with TMAO solutions (up-jump) while the signal increases when protein in TMAO is mixed with buffer alone (down-jump.) Figures 2-3A and 2-3B show representative up-jump and down-jump traces, respectively. The kinetic traces were fitted to multiple exponential functions using a nonlinear least squares method. The best fit results of both up-jump and down-jump kinetic traces can be described by the sum of triple exponentials (Equation 6 in materials and methods), so there are three phases in both the folding and unfolding processes. The quality of the fit was judged by the fitting residuals showed in the inset panel of Figure 2-3A and 2-3B. The observed rate constants ($k_{obs}$) of the fast (ranging from 2 to 7 s$^{-1}$) and middle (ranging from 0.8 to 3 s$^{-1}$) phases are TMAO concentration-dependent, as are the amplitudes associated with each phase, as shown in Figures 2-4A and 2-4B, respectively. Since the summation of the fast and the middle phase amplitudes is more than 95% of the total raw amplitude, these two phases describe the major folding kinetics of P protein. The slowest phase, which is around 0.10 ~ 0.15 s$^{-1}$ is TMAO-independent and the amplitude of the phase is small (less than 5% of total raw amplitude). We assigned this latter slow phase to cis/trans isomerization of Pro 39 and/or Pro 90 based on the absence of the third phase in folding and unfolding transients of the P39A/P90A variant of P protein (data not shown.) All subsequent studies employed the proline-containing F107W
variant, whose transients were fit to three exponentials but we did not include the slowest (third) phase in the subsequent data analysis.

In many multiphasic protein folding systems, the observed rate constants are usually at least one order of magnitude different with each other. Under these conditions, the nonlinear lease squares fitting method is robust and capable of accurately estimating both the rates and the associated amplitudes. However, one intrinsic challenge in data analysis of P protein system originates from the similarity of the observed rate constants associated with each of the two fastest phases. The fitting results of the observed rates constants for the fast and medium phases in P protein folding processes are relatively close to each other (between 2 to 6 fold). When the two rates are close, the uncertainty of the parameters obtained from nonlinear square fits is large because the fitting parameters have a correlation coefficient close to 1. In order to best represent the uncertainty for each data point in Figure 2-4A, the same kinetic experiment was performed at least three times. The uncertainty was calculated for each condition from the fitting results of these independent experiments.

\[
\begin{align*}
\text{U} & \quad \text{I} \\
\text{I} & \quad \text{N}
\end{align*}
\]

Scheme 2-1

Since there are two major phases in the folding process, the simplest mechanism is a three-state model with an on-pathway intermediate state, as shown in scheme 2-1. Four kinetic rate constants, \( k_{UI}, k_{IU}, k_{IN}, \text{ and } k_{NI} \) are associated with the scheme, and the TMAO concentration dependence of these rate constants are described
Figure 2-3: Folding/unfolding of P protein induced by changes in TMAO concentration and monitored by stopped-flow fluorescence. (A) Unfolding (down-jump) of P protein from 1 M to 0.09 M TMAO. The kinetic trace of the unfolding process monitored by fluorescence is the black line. The red line through the data points is the best fit to a triple exponential function. The inset shows the residual of the triple exponential curve fits of the fluorescence data. (B) Folding (up-jump) of P protein from 0 M to 0.54 M TMAO. The red line through the black data points is the best fit to a triple exponential function. The inset shows the residual of the fitting. Both data sets were collected at a rate of 500 points/sec.
Figure 2-4: Folding and unfolding kinetics of F107W P protein as a function of final TMAO concentration. (A) TMAO concentration dependence of the natural logarithm of observed rate constants for both up-jump and down-jump experiments. The open circles and closed circles are the fast and slow observed rate constants obtained from the folding (up-jump) experiments, respectively. (B) Raw amplitudes associated with fast and slow observed rate constants are plotted as a function of TMAO concentration. Only the amplitudes of folding (up-jump) experiments are shown. Smooth lines through the observed rate constants are best fits to scheme 1 using the analytical solutions described by Bujalowski and Jezewska (Bujalowski and Jezewska 2000). The best fits kinetic parameters of the model are listed in Table 2-1.
using Equation 7. The initial fitting analyses from the kinetic data alone did not yield convergent estimations of the parameters because the two observed rate constants were similar and the uncertainty in the fastest rate constant was high. Equilibrium titration experiments were performed to provide more constraints in the fitting process.

2.2.3 Urea-TMAO fluorescence and CD co-titration

Urea and guanidinium chloride denaturation titrations are widely used to estimate the protein stability. The method is suitable only for stable proteins when both native and denatured baselines are present in the titration curve. When applying this method to IUPs, the absence of a complete cooperative transition and native baseline in the titration curve makes it impossible to estimate the stability using denaturant alone. In the titrations described here, we used mixtures of urea and TMAO as the denaturant and a renaturant, respectively. Titrations were initiated with a starting sample consisting of 5 µM protein in various urea or urea/TMAO mixtures. Protein at the same concentration, in 1.4 M TMAO was then added using an automated tirator, which resulted in progressive dilution of the starting urea concentration while the TMAO concentration increased. Six such titration curves allowed us to construct a three-dimensional urea-TMAO cotitration surface. Each titration step yielded both fluorescence and CD data from the identical protein sample using the same instrument. This method eliminated errors from discrepancies in protein, urea, or TMAO concentration that might occur if the two types of data were collected separately. Figures 2-5A and 2-5B show the TMAO dependence of the equilibrium fluorescence signal (\(\lambda_{ex} = 285\) nm, \(\lambda_{em} > 320\) nm) and CD signal (\(\lambda = 222\) nm) for F107W variant, respectively. Taken separately, fluorescence and CD titration
Figure 2-5: Fluorescence and circular dichroism detected equilibrium co-titration surface of the P protein (5mM, 10mm path length cuvette). The gray meshed surfaces on the two data sets are the best global fit results using a three-state model described by equation 5. (A) The fluorescence titration surface constructed by six titration curves was plotted as a function of [urea] and [TMAO]. The protein sample was excited at 285nm and the fluorescence signal was collected through a 320nm high passed filter. (B) The CD titration surface constructed by six titration curves was plotted as a function of [urea] and [TMAO]. The CD signal was monitored at 222nm. The best-fit equilibrium parameters from the model were listed in Table 2-1.
curves can be fit to a two-state model but the resulting fits give statistically significant
different $C_{\beta_2}$ values. This observation indicates the presence of one or more significantly
populated equilibrium intermediates in the folding process, consistent with the kinetic
data. Our previous reports on P protein folding described the mechanism as two-state,
based on statistically insignificant differences in $C_{\beta_2}$ values (Henkels, Kurz et al. 2001),
but this conclusion was based on titration with sulfate to induce folding. Under these
conditions the population of the intermediate is lower than when TMAO is used to induce
folding. This difference, combined with the fact that the near- and far-UV CD data were
collected on different samples, led us to erroneously suggest that equilibrium
intermediate is not present, when in fact it is (see below.). By combining the equilibrium
titration data with the stopped flow data, we can globally fit the equilibrium data to a
model based on the three-state mechanism shown in scheme 2-1.

The experimental data were analyzed using equations 1-5 (materials and
methods), assuming that the free energy of each transition is linearly dependent on both
urea and TMAO concentrations. In general, the CD and fluorescence of each state may
vary with the urea or TMAO concentration, especially for fluorescence if the fluorophore
is solvent exposed. In accordance with convention, we assumed that the CD and
fluorescence signals of U, I and N states vary linearly with urea and TMAO
concentrations:

$$S_i^{\text{Sig}}(\text{urea,TMAO}) = S_i^{\text{Sig}}(0) + s_i^{\text{Sig, urea}} \cdot \text{urea} + s_i^{\text{Sig, TMAO}} \cdot \text{TMAO}$$

(8)

where $i = U,I,N$; $\text{Sig}$ is the CD or fluorescence signal; and $s_i^{\text{Sig, urea}}, s_i^{\text{Sig, TMAO}}$ are the slopes of
the baseplanes describe the urea and TMAO dependence of the signal of each state. Therefore, nine parameters for the fluorescence signals and nine parameters for the CD signals for a total of 18 spectroscopic parameters were used in the model. Fluorescence and CD titration data were simultaneously fitted using a three-state equilibrium model. Six thermodynamic parameters, $\Delta G_{\text{UI}}^0$, $\Delta G_{\text{IN}}^0$, $m_{\text{UI}}^T$, $m_{\text{U}}^T$, $m_{\text{IN}}^T$, and $m_{\text{N}}^T$, were used as global fitting parameters. Initial fitting of the data equilibrium data alone did not yield well-determined parameter values, especially the spectroscopic parameters associated with the intermediate state. For this reason, the free energy $\Delta G_{\text{UI}}^0$, and $\Delta G_{\text{IN}}^0$ also varied dramatically. The poor estimation of these parameters mainly results from the large parameter space (total 24 parameters) and lack of two distinct transitions. This latter limitation indicates that either the intermediate state is poorly populated throughout the titration range or that its spectroscopic signals are very similar to U or N. In order to overcome these limitations in the data analysis, we estimated $\Delta G_{\text{UI}}^0$ and $m_{\text{UI}}^0$ from NMR spectra collected at various urea concentrations as described below. These parameters were held fixed in a global non-linear least squares fit of the remaining parameters to the combined kinetic and equilibrium data.

### 2.2.4 $\Delta G_{\text{UI}}^0$ and $m_{\text{UI}}^0$ from NMR urea titration experiments

NMR experiments were performed to estimate the equilibrium constant between U and I states. More than 95% of the backbone amides of the folded state P
protein were assigned in a previous study (Henkels 2005). The presence of at least one alternative state was suggested by the observation of around 50 extra peaks in the $^{15}\text{N-}^{1}\text{H}$ HSQC spectrum of sulfate folded P protein. In addition, the $^{15}\text{N-}^{1}\text{H}$ HSQC spectrum of P protein in the absence of sulfate (20 mM sodium cacodylate buffer, pH 7) has some resonances whose positions are similar to those found in sulfate folded protein (see detail in Chapter 3). Both of these results suggest the existence of an alternative state in the P protein folding transition, which gives rise to extra HSQC peaks under both folding and unfolding conditions. As a working hypothesis, we assumed that this alternative state corresponds to the intermediate state observed in the kinetic and equilibrium experiments. In addition, we assumed that the native state peaks observed in the sulfate-free sample correspond to residues that are in a native-like magnetic environment in the intermediate state. If correct, this interpretation suggests that the intermediate state has a significant population, comparable to that of the unfolded state. The equilibrium data suggest that the population of the native state is negligible under these conditions, so addition of urea alters the NMR spectrum by perturbing the equilibrium between the U and I state. Thus, $^{15}\text{N-}^{1}\text{H}$ HSQC spectra were collected at various urea concentrations. The relative populations of U and I states were estimated by peak intensity or volume change and used to determine the $\Delta G_{ui}$ as a linear function of urea concentration.

Figure 2-6A shows a region of the unfolded P protein HSQC spectrum, which contains some of the observed crosspeaks of folded structure of P protein. These peaks were monitored in spectra obtained at urea concentrations ranging from 0 to 1 M. As urea concentration increases, the intensities of these peaks decrease, indicating a
Figure 2-6: Urea titration of P protein in the absence of sulfate (20mM cacodylate buffer, pH 7) monitored by HSQC spectrum and the estimation of free energy difference between U and I state. (A) A region of the HSQC spectrum is shown in the plot, and two peaks corresponding to N28 and R97 in a native like environment are labeled. The urea concentrations are shown in each plot. (B) The estimation of free energy between U and I state from different residues against urea concentration was plotted. The free energy was calculated from the population of U and I state from HSQC urea titration experiment. The red line is the best-fit results of all the data points in the plot using the linear function.
decrease in the population of the intermediate state to which they are tentatively assigned. To estimate the population of the intermediate from these spectra, the native state (in 20 mM sulfate, pH 7.0) HSQC spectrum was used as the reference to determine the intensity of each native-like peak in fully folded P protein. The peak intensities in all spectra were normalized to the intensity of the K119 amide peak, whose position is independent of sample conditions, making it an excellent internal standard. The normalized intensity of each intermediate peak was divided by the normalized intensity of the corresponding native spectrum peak to produce an estimate of the fractional population of the intermediate at each urea concentration. These populations were used to calculate $K_{UI}$, which was converted to free energy using Equation 2. To avoid line shape artifacts, peak volume was used to estimate peak intensity. Figure 2-6B shows the free energy calculated in this fashion for each monitored peak plotted against urea concentration and fit to a linear function. The intercept of each line is the free energy $\Delta G_{UI}^0$, and the slope of urea m-value, $m_{UI}^0$, between U and I transition. The data points from 4 peaks gave $\Delta G_{UI}^0$ ranging from 1.0 to 1.3 kcal/mol. The fact that urea m-values show larger variation between different peaks might be due to the large variation in estimation of the peak volume for different peaks. Combined estimates of $\Delta G_{UI}^0$ (1.0 kcal mol$^{-1}$) and $m_{UI}^0$ (1.5 kcal mol$^{-1}$ M$^{-1}$) were obtained from a linear regression of all the data points, shown as the red line in the Figure 2-6B. These two values were used as fixed parameter values in the global fitting of the combined equilibrium and kinetic data.
2.2.5 Global fitting of kinetic and equilibrium data

Because the kinetic or equilibrium data alone cannot provide an accurate estimate of all of the model parameters, a global analysis of the two data sets was performed to estimate the kinetic and spectroscopic parameters together. The experimental data included the two observed rate constants and the two associated amplitudes at 33 different TMAO concentrations and the fluorescence and CD titration surfaces. Because NMR studies indicated that the magnetic environment of the Trp residue in the intermediate state is similar to that of the unfolded state, we further simplified the fit by assuming that the TMAO dependence of fluorescence of the intermediate state is the same as the unfolded state. The lines in Figure 2-4A are the best-fit model for the observed rate constants. As depicted in Figure 2-4A, the usual chevron shape of $k_{obs}$ vs. [TMAO] is not observed because the two reaction steps in the scheme 2-1 have similar rate constants and are coupled to each other over most of the [TMAO] range. The amplitudes were also analyzed using the model, but the results do not fit well to the data points. Therefore the fitted results are not plotted with the amplitude data on Figure 2-4B. The reason for this discrepancy might result from the large uncertainty of fluorescence signal of each state from the fitting (Table 2-1). As shown in Figures 2-5, the best-fit model fit the equilibrium fluorescence and CD data quite well, although the denatured fluorescence baseplane did not fit as well as the CD. Inspection of the data in the denatured baseplane region shows differences in the denatured state fluorescence in different titrations. Although these differences could be caused by day-to-day variations in the fluorescence detection electronics, they also could result from small perturbations.
in the actual (not read) pH of the buffer caused by urea sensitivity of the pH electrode (Acevedo 2002). The fluorescence of F107W variant is pH sensitive between 5 and 7 (data not shown), presumably because it is near the His 105 residue whose ionization state could affect the fluorescence signal of tryptophan fluorphore. Differences in the starting fluorescence of each titration curve could be due to slightly different pH values at each starting urea concentration.

### Table 2-1: Kinetic and equilibrium parameters of three-states folding and unfolding of P protein

<table>
<thead>
<tr>
<th>Kinetic Parameters</th>
<th>Equilibrium Parameters</th>
<th>Spectroscopic Parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_k_{UI}$ 0.4±0.05</td>
<td>$\Delta G_{UI}^0$ 1.0±0.2</td>
<td>$a\ Sigma_{UE}^U$ -23±5 (mdeg)</td>
</tr>
<tr>
<td>$b_k_{IU}$ 2.2±0.3</td>
<td>$\Delta G_{IN}^0$ 1.02±0.1</td>
<td>$b\ Sigma_{UE}^I$ -54±10 (mdeg)</td>
</tr>
<tr>
<td>$k_{IN}$ 1.1±0.1</td>
<td>$c_m_{TMAO}^U$ -3.2±0.2</td>
<td>$Sigma_{CD}^C$ -80±5 (mdeg)</td>
</tr>
<tr>
<td>$b_k_{NI}$ 6.1±0.3</td>
<td>$m_{TMAO}^{I}$ -3.9±0.3</td>
<td>$c\ Sigma_{FL}^I$ 5.9±0.5 (a.u.)</td>
</tr>
<tr>
<td>$\beta_{UI}$ 0.66±0.02</td>
<td>$m_{Urea}^U$ 1.5±0.5</td>
<td>$Sigma_{FL}^I$ 5.0±0.8 (a.u.)</td>
</tr>
<tr>
<td>$\beta_{IN}$ 0.88±0.02</td>
<td>$m_{Urea}^I$ 1.54±0.3</td>
<td>$Sigma_{FL}^N$ 1.46±0.5 (a.u.)</td>
</tr>
</tbody>
</table>

$^a$ Units are in s$^{-1}$

$^b$ Calculated using $k_{IU}/K_{IU}$ and $k_{IN}/K_{IN}$, respectively.

$^c$ Units are in kcal mol$^{-1}$ M$^{-1}$

$^d$ Units are in deg cm$^2$ dmol$^{-1}$

$^e$ Arbitrary units

The kinetic and equilibrium parameters for the best-fit model are listed in Table 2-1. The free energy differences between U and I ($\Delta G_{UI}^0 \equiv 1$ kcal/mol) and I and N ($\Delta G_{IN}^0 \equiv 1.05$ kcal/mol) are approximately the same and predict populations for unfolded, intermediate and native state populations of 83±5, 15±3, and 2±0.5%, respectively. The ratios of TMAO and urea m-values are -2.1±0.8 and -2.5±0.5 for U⇔I and I⇔N transitions, respectively. Auton and Bolen have proposed a method for calculating m-
values based on differences in the backbone and sidechain solvent accessible surface areas (Auton and Bolen 2007). Because this calculation depends on side chain composition, there is no single expected ratio of TMAO to urea m-value for different equilibria. Previous studies have shown that Barnase and Nank1-7* have the ratio of -1.1 and -2.4, respectively (Mello and Barrick 2003). However, in general TMAO has a stronger opposite effect on protein stability than urea (larger absolute m-value).

The best-fit values of both folding rate constants ($k_{ul}$=0.4 s$^{-1}$ and $k_{lu}$=1.1 s$^{-1}$) are relatively slow compared to other proteins of similar size (Jackson 1998). Contact order is a measure of the topological complexity of a given protein’s structure and a measurement of the number of contacts a residue makes with other residues that are local in sequence, relative to the number of contacts with residues distant in sequence (Plaxco, Simons et al. 1998). Thus, the contact order is related to secondary structure, but is also influenced substantially by the tertiary fold. Plaxco et al. showed a correlation between the contact order and natural logarithm of the refolding rate constant and the position of the transition state. These correlations suggested that the lower the contact order the faster the folding rate constant and less folded-like the transition state. From this correlation, the contact order of the P protein (14.3) predicts a folding rate constant (1000 s$^{-1}$) is at least 2 to 3 orders of magnitude faster than our experimental estimates. This major discrepancy suggests that other factors besides topology, such as the charge-charge interactions in the protein, can be a major determinant of the actual folding kinetics. Further kinetic studies on other IUPs are necessary to determine if the slow refolding rate constant is a general property of the IUP folding reactions.
The best-fit model parameters include the intermediate state CD and fluorescence signals, which provide information about the structure of I. The best-fit value for the fluorescence signal of I is close to that of U, indicating that the environment surrounding Trp107 is similar in the I and U states, in agreement with the NMR results described in Chapter 3. In contrast, the CD signal ($\theta_{222}$) of I is midway between U and N suggesting that approximately 50% of the native state secondary structure, particularly helix, is formed in the intermediate state. This conclusion is also in agreement with our NMR studies of I.

**2.2.6 Utility of TMAO (renaturant) stopped flow and cotitration experiments as a general method to determine IUP folding mechanisms**

Various protective osmolytes have been used as co-solutes to stabilize proteins and favor the folded state (Wang and Bolen 1997; Russo, Rosgen et al. 2003). Many studies have shown that IUPs can fold to their native state when encountering to their binding partners. Are the folded states of proteins induced by protective osmolytes the same as those induced by their in vivo binding target? From the experimental point of view in this study, the CD and fluorescence spectrum of the ligand induced native state and the TMAO induced native state are very similar. Furthermore, most of the recognizable peaks in the $^1$H-$^{15}$N HSQC spectrum of the native state from these two conditions have similar chemical shifts indicating the structures of the protein under the two conditions are very similar. Generally speaking, osmolytes favor the folded state of the protein, which represents a minimum free energy ensemble. An energy landscape can be used to conceptually depict the relative free energy of the conformational states of a
protein, based on its amino acid sequence (Dill and Chan 1997). Changing sample conditions (e.g., adding urea or TMAO) shifts the equilibrium between these states. Different solvents will not create new minima on the energy landscape but instead change the relative deepness of the existing minima on it. This explanation again rationalizes that the TMAO induced native state of P protein should be the same as the ligand induced one originally observed. These results also imply that TMAO does not dramatically change the structure of the folded P protein. This conclusion is shared by studies of other proteins that also used TMAO to study the effect on stabilities and structures of proteins and noted little perturbation on the folded structure (He, Chen et al. 2009; Lin, Zarrine-Afsar et al. 2009). Other studies have shown that TMAO can enhance the enzymatic activity, making TMAO perturbation of enzyme structure unlikely (Kumar, Serrette et al. 2005).

With the increasing number of reported IUPs and partially folded proteins, there is a need for systematic biophysical methods to measure the kinetic and thermodynamic properties of these proteins. Although some studies employed denaturant-induced protein unfolding kinetic experiments in the presence of various TMAO concentrations to investigate the stability effect (Banks and Gloss 2004; Topping and Gloss 2004), there have been no previous kinetic studies used TMAO to initiate IUP folding. The stopped-flow measurements of TMAO-induced IUP folding/unfolding described in the present study could become a standard experiment like traditional denaturant-induced unfolding/folding kinetics of natively structure proteins.

Several studies have demonstrated that TMAO combined with urea or guanidinium chloride denaturation can be used to properly estimate the free energy of
partially folded protein (Mello and Barrick 2003). The cotitration method we developed in this study should be able to measure the stability of other IUPs. The highest TMAO concentration in the titration was 1.4 M, but it can be as high as needed in the system that one is interested. The reason TMAO did not exceed 1.4 M in this study is that more addition of TMAO changed the pH value of the sample. A shown in the Figure 2-6A and 2-6B, the six titration curves still sampled the titration surface very well. The simultaneous detection of CD and fluorescence in the equilibrium titration is particularly useful for detecting intermediate states in protein folding equilibria.

2.2.7 Population of the intermediate state

Based on the parameter values listed in Table 2-1, the population of the P protein intermediate state is around 16-20% of the total protein concentration in the absence of urea and TMAO. The secondary structure of the intermediate state of P protein explains the shape of the CD signal vs. [urea] along the 0 M TMAO line in Figure 2-2B. The curvature from 0 M to 2 M urea can be explained by the equilibrium transition between the unfolded state and the intermediate state of the P protein. The population of the three states as function of TMAO concentration is plotted in Figure 2-7 using the equilibrium parameters from the fitting. The population of the intermediate state reaches its maximum around 0.3 M TMAO. At that TMAO concentration, the population of I is comparable to U and N state, and the relative amounts of three states are approximately equal.
Figure 2-7: Population of U, I, and N state as a function of the TMAO concentration. (A) The population of each state was calculated using equation 4 and the equilibrium parameters \((\Delta G_{UI}^0, \Delta G_{IN}^0, m_{UI}^{TMAO}, m_{IN}^{TMAO})\) listed in Table 2-1.

2.3 Materials and Methods

2.3.1 Chemicals and reagents

Ultra pure urea was purchased from Nacalai Tesque Inc. Trimethylamine-N-oxide (TMAO) was from Fluka. Cacodylic acid was from Sigma. In order to remove the background fluorescence from TMAO, a 3 M stock TMAO solution was mixed with activated charcoal (0.025g/ml) for at least 5 hours. The charcoal was then remove by filtration using 0.22 \(\mu\)m syringe filter. Urea and TMAO concentrations were measured by refractive index using the equation from (Warren and Gordon 1966) and (Wang and Bolen 1997), respectively.
2.3.2 Expression and purification of P protein

The gene for the P protein variant F107W used in this study was constructed by the QuickChange site direct mutagenesis protocol (Strategene). The variant was overexpressed in *E. coli* and purified as described (Niranjanakumari, Kurz et al. 1998) with the following modifications. The pooled fractions of P protein that eluted from the second CM-Sepharose column were concentrated using Centriprep until the volume of the pooled fractions was down to 1-2 ml. The buffer of the protein stock was changed to 6 M Guanidine-HCl, 10 mM Tris at pH 7.5 during this concentrating step. The concentrated stock was then loaded on an S-100 column (GE Healthcare) equilibrated with 6 M Guanidine-HCl, 10 mM Tris (pH 7.5) to remove EDTA bound P protein. Fractions containing P protein were pooled and stored at -80 °C. Before equilibrium titration or stopped flow kinetic experiments, the P protein sample was dialyzed against water extensively, then 20 mM sodium cacodylate, pH 7. In all experiments the protein concentration was determined using the Edelhoch method using an extinction coefficient of 11,200 M$^{-1}$ cm$^{-1}$ at 276nm (Edelhoch 1967).

2.3.3 Fluorescence and CD equilibrium cotitration experiments

The urea-TMAO equilibrium cotitration experiments were performed on an Applied Photophysics PiStar instrument equipped with both fluorescence and CD detection. A total of six renaturation equilibrium titration curves were collected in each set of experiments and all samples were in 20 mM sodium cacodylate buffer (pH 7.0). The starting sample for the protein concentration was 5 μM in 0 M, 2 M, 4 M, 6 M urea. The two additional starting samples were 5 μM protein in 6M urea, 0.5 M TMAO or 6 M
urea, 0.9M TMAO. The titrant solution is 5 µM protein in 1.4 M TMAO. An automated Hamilton titrator was used to perform all titrations. The CD signal was collected at 222 nm and fluorescence signal was collected using a 320 nm cut-off filter with an excitation wavelength of 285 nm and a 4 nm bandwidth. The CD and fluorescence signals were collected simultaneously from the same sample. Each urea-TMAO concentration pair data point was measured for 15 sec to obtain a mean signal. The equilibrium time between two titration points was 3 to 3.5 minutes.

2.3.4 Stopped-flow fluorescence kinetic experiments

The stopped-flow fluorescence kinetic experiments were performed on an Applied Photophysics SX20 instrument at 25 °C with a circulating water bath. The excitation wavelength was 285 nm with a slit width of 1 mm. The pathlength of the observation cell is 2 mm. The fluorescence signal was collected through a 320 nm cut-off filter. For single mixing folding and unfolding experiments, a 6-fold or 11-fold dilution in the stopped flow instrument was used to initiate the reactions. In the refolding experiments, protein samples are 6 µM in 20 mM sodium cacodylate buffer (pH 7.0), and TMAO concentrations range from 0.15 M to 1.4 M in 20 mM sodium cacodylate buffer (pH 7.0). The final protein concentration was 1 µM with TMAO concentrations ranging from 0.125-1.167 M. For unfolding experiments, protein samples were 11 µM in 1 M TMAO with 20 mM sodium cacodylate, and TMAO concentrations ranged from 0 to 0.5 M. By 11-fold dilution in the stopped flow instrument, the final protein concentration was 1 µM in 0.091 M to 0.55 M TMAO. The time traces of the fluorescence change for the
reaction were collected for 20 s in a split time base mode. Five thousand points were collected for the first 10s, and 1000 points from were collected from 10 to 20 s.

2.3.5 Data analysis

(a) Equilibrium titration data: Equilibrium urea-TMAO cotitration surface data were fit to a linear three-state model assuming that $\Delta G_{UI}$ and $\Delta G_{UN}$ are linearly dependent on urea and TMAO concentration (Mello and Barrick 2003), according to equations 1-5:

$$
\Delta G_{UI}(\text{TMAO}, \text{Urea}) = \Delta G_{UI}^0 + m^T_{UI} \cdot \text{TMAO} + m^U_{UI} \cdot \text{Urea}
$$

(1)

$$
\Delta G_{IN}(\text{TMAO}, \text{Urea}) = \Delta G_{IN}^0 + m^T_{IN} \cdot \text{TMAO} + m^U_{IN} \cdot \text{Urea}
$$

(2)

$$
K_{UI} = \exp\left(-\frac{\Delta G_{UI}}{RT}\right)
$$

(3)

$$
K_{IN} = \exp\left(-\frac{\Delta G_{IN}}{RT}\right)
$$

(4)

$$
F_U = 1 / (1 + K_{UI} + K_{UI}K_{IN})
$$

$$
F_I = K_{UI} / (1 + K_{UI} + K_{UI}K_{IN})
$$

$$
F_N = K_{UI}K_{IN} / (1 + K_{UI} + K_{UI}K_{IN})
$$

(4)

$$
CD^{obs} = S_U^{CD} \cdot F_U + S_I^{CD} \cdot F_I + S_N^{CD} \cdot F_N
$$

(5)

$$
FL^{obs} = S_U^{FL} \cdot F_U + S_I^{FL} \cdot F_I + S_N^{FL} \cdot F_N
$$

where $m^T_{ij}$ and $m^U_{ij}$ are the m values that describe the linear osmolyte concentration dependence of the free energy, $\Delta G_{ij}$; $K_{ij}$ is the equilibrium constant of the i-j equilibrium; $F_i$ is the fractional population of state i at a specific osmolyte concentration; $S_i^{CD}$ and $S_i^{FL}$ are the intensities of CD and fluorescence of the unfolded (U), intermediate (I), native (N) states, respectively; $CD^{obs}$ and $FL^{obs}$ are the observed CD and fluorescence signals.
(b) Kinetic folding/unfolding data: TMAO induced folding/unfolding kinetic time traces from stopped flow experiments were fitted by the non-linear least-squares method with multiple the exponential function:

\[ F(t) = F(\infty) + \sum_i A_i e^{\lambda_i t} \]  

where \( F(t) \) and \( F(\infty) \) are the observed fluorescence at time \( t \) and infinite time, respectively. \( A_i \) and \( \lambda_i \) are the fluorescence amplitudes and observed rate constants, respectively, of the different phases. The amplitudes and observed rate constants were plotted against the final TMAO concentration.

(c) Global fitting: The CD and fluorescence equilibrium titration surfaces were fit using equations 1-5. The amplitudes and observed rate constants were fit to a linear three-state model (scheme 2-1) using the analytical solutions described by Bujalowski (Bujalowski and Jezewska 2000). The TMAO dependence of the observed rate constants and associated amplitudes assumes a linear dependence of the logarithm of the microscopic rate constants in scheme 2-1 on TMAO concentration in the form:

\[ k_{ij} = \exp \left[ \log k_{ij}^0 + (\beta_{ij} - 1) \cdot \frac{m_{ij}^T \cdot TMAO}{R \cdot T} \right] \]  

where \( k_{ij} \) is the microscopic rate constant for the transition from state \( i \) to \( j \) at specific TMAO concentration. \( \beta_{ij} \) is the Tanford value (Fersht 1999) which is a measure of the average degree of solvent exposure of the transition state relative to the difference in solvent exposure between states \( i \) and \( j \). Equilibrium data, kinetic amplitudes and observed rates were then fit simultaneously using Mathematica (Wolfram) to obtain the equilibrium and kinetic parameters for the model.
Chapter 3 : Probing the Folding Intermediate of Bacillus subtilis RNase P protein by NMR

3.1 Introduction

An understanding of the processes by which a protein folds into its native conformation requires knowledge of the various states that populate the folding pathways, the kinetic rate constants between these states, and their relative stabilities, as well as their structural properties. A description of the first three properties of the P protein system were the main goals of Chapter 2. In this Chapter, we characterize the structural properties of one of these states, specifically the intermediate state.

Despite the observation that many proteins populate an intermediate state during the folding process (Capaldi, Shastry et al. 2001; Teilum, Maki et al. 2002; Sánchez and Kiefhaber 2003), the role of intermediates in productive folding is still the subject of fertile discussion (Clark 2008). Intermediates between the unfolded and folded states may enhance the rate of folding by decreasing the conformational space through which the polypeptide chain has to search, or may reduce the rate of folding by sequestering the polypeptide chain in a stable, partially folded state (Fersht 1995; Krantz, Mayne et al. 2002). For most small proteins, intermediates form within the dead-time of a stopped-flow instrument, so it has not been feasible to determine their role in the folding process (Baldwin 1996; Roder and Colón 1997). Recent development of an ultra rapid mixing-instrument enabled the initial folding phase and early folding events to be detected directly. It has been shown that the intermediate ensemble formed early in the folding of Im7 is on the folding pathway (Capaldi, Shastry et al. 2001), and similar
results have been found for the protein GB1 (Park, Shastry et al. 1999). In addition to the transient property of intermediate states, many of them are populated at very low levels, making their experimental characterization difficult in some systems.

Most of the mixing techniques that have been used to study protein folding are based on the use of a single probe or a single average property derived from many residues in the protein. It is thus difficult to obtain residue-specific information about the intermediate state. Therefore determination of the conformational properties of an intermediate state at high resolution is important for a full elucidation of the structural mechanism of folding. However, the transient nature and low population of the intermediate states relative to the unfolded and folded state are challenges for structural experiments.

A variety of NMR methods are available to study the conformational properties of intermediate states. Relaxation dispersion NMR spectroscopy provides a complementary approach to other biophysical probes of folding because the technique is extremely sensitive to processes that involve interconversion between a ground state (native or unfolded state) and excited states (intermediate state) of molecules, as long as the excited states are populated at levels of 0.5% or higher and the exchange kinetics are on the millisecond time-scale. The method also provides residue-specific information over the whole protein simultaneously. The results allow the extraction of the rate constants for interconversion between the two states, their population and information about their structure (Korzhnev, Salvatella et al. 2004; Korzhnev, Neudecker et al. 2005; Korzhnev, Neudecker et al. 2006). NMR analysis of partially folded proteins stabilized
by changing solution conditions or mutagenesis have also provided insights about the nature of their conformational ensemble (Feng, Zhou et al. 2005; Religa, Markson et al. 2005; Nishimura, Dyson et al. 2006). Long-range NOEs and other structural information for partially unfolded states could also be detected by NMR methods (Mok, Kay et al. 1999; Kortemme, Kelly et al. 2000). Another approach to investigate protein folding at the residue level is to record a series of $^1$H-$^{15}$N HSQC spectra at increasing denaturant concentrations under equilibrium conditions. The application of this method to the study of the unfolding of apoflavodoxin allowed for the determination of the stability of every native secondary structure element (van Mierlo, van den Oever et al. 2000). A similar study of the denaturation of a human $\alpha$-lactalbumin folding intermediate showed that the transition between the molten globule and the unfolded state is not cooperative (Schulman, Kim et al. 1997). Overall, the NMR technique offers several different ways to characterize the intermediate state kinetically, thermodynamically and structurally.

RNase P protein is the protein subunit in the RNase P holoenzyme. Many RNA binding proteins exhibit conformational dynamics, which promote specific recognition of cognate RNA partners (Deka, Rajan et al. 2005). The N-terminal RNA recognition domain (RRM) of human U1A binds to the 3'-UTR of U1A mRNA. In the absence of RNA, backbone and sidechain relaxation experiments reveal conformational heterogeneity on the $\mu$s-$ms$ timescale in the regions that bind RNA (Mittermaier, Varani et al. 1999; Shajani and Varani 2005). Previous backbone dynamic studies of P protein also found the presence of conformational heterogeneity (motions on $\mu$s-$ms$ timescale) in the native state ensemble of the protein. In addition, presence of extra peaks in the NMR
$^1$H-$^{15}$N HSQC spectrum of native state of P protein clearly shows evidence of alternative conformations in slow exchange with the native state.

For the work described in this chapter, standard multidimensional NMR experiments were performed to determine the residue identities of the extra peaks in the $^1$H-$^{15}$N HSQC spectrum of the native state of P protein. The assignment results showed that most of these residues are located in the vicinity of the N- and C-termini. The chemical shift index calculated from $^{13}$C$_{\alpha}$ and $^{13}$C$_{\beta}$ assignments show small deviations from the random coil chemical shifts (< 1 ppm), indicating that these residues are unfolded in the intermediate. Taken together, the NMR data suggests that the central $\beta$-sheet and $\alpha$-helix B are folded in the intermediate state and that the N-terminal $\alpha$-helix A and C-terminal $\alpha$-helix C are unfolded.

3.2 Results

3.2.1 Presence of extra peaks in folded P protein HSQC spectrum and native peaks in the unliganded P protein spectrum

Assignments for 95% of the backbone resonances of folded P protein have been reported previously (Henkels 2005). Although most of the peaks present in the $^1$H-$^{15}$N HSQC spectrum of the folded state of P protein were assigned in the previous studies, a few residues were missing and some extra peaks could not be assigned. In the previous study, the protein sample conditions of pH 6.0, 20 mM sodium sulfate were carefully chosen to minimize the number of the extra peaks thereby allowing unambiguous assignment of most peaks in the $^1$H-$^{15}$N HSQC spectrum of folded P protein. However,
Figure 3-1: $^1$H-$^{15}$N HSQC spectrum of $^{15}$N-labeled P protein in 5 mM sulfate, pH 5.0 at 25°C. The peaks in grey are the native spectrum resonances assigned in a previous study (Henkels and Oas 2006). The peaks in red are extra peaks that appear under these buffer conditions. All of the extra peaks are located between 7.9 and 8.5 ppm in the $^1$H dimension. The labels give the assignments of the extra peaks.

decreasing the sample pH from 6.0 to 5.0 and lowering the sulfate concentration to 5 mM increases the intensities of the unassigned extra peaks. As shown in Figure 3-1, there are approximately 50 unassigned peaks present in the spectrum, and most of them are located in the 8.0 to 8.5 ppm region of the $^1$H spectrum.

Figure 3-2 shows the HSQC spectrum of unliganded P protein at the same
pH, but in the absence of sulfate. Under these conditions, the protein is predominantly unfolded. This spectrum is characteristic of unfolded polypeptides, but there are two subsets of peaks (circled in Figure 3-2) whose positions are similar to assigned resonances in the HSQC spectrum of P protein (Henkels 2005). The presence of assigned native peaks in the unfolded P protein spectrum can be explained by the presence of a second partially folded species that exchanges slowly ($\tau_{ex} > 1$ sec) with the fully unfolded protein. The presence of extra peaks and missing residues in the native state spectrum are evidence of the existence of one (or more) alternative conformation that slowly exchanges with the native conformation. Taken together, both HSQC spectra are consistent with the presence of an alternative structural conformation that is in equilibrium between the unfolded and native state. This interpretation is supported by our folding kinetic studies of P protein, which demonstrated that there is an intermediate state in the P protein folding process. As a minimal complexity model that is consistent with all of the experimental data, we have assigned the alternative conformation found in the NMR experiments to the partially folded intermediate state in the P protein folding mechanism.

### 3.2.2 Sulfate and pH titration of folded P protein

In order to gain more insight into how ligand concentration and pH affect the equilibria between the P protein intermediate state and the native state or unfolded state, a series of $^1$H-$^{15}$N HSQC spectra were collected at various ligand sulfate concentrations and pH values. The sulfate concentration ranged from 0 to 100 mM, but
Figure 3-2: $^1$H-$^{15}$N HSQC spectrum of $^{15}$N-labeled P protein in 20 mM sodium cacodylate, pH 5.0, at 25°C. Sulfate is not present in the sample, so P protein is predominantly unfolded. The peaks in the spectrum are poorly dispersed compared to those of folded P protein spectrum in Figure 3-1. The peaks in the two circles form patterns similar to those found in the folded P protein spectrum. The intensities of these peaks are weaker than those of the remaining unassigned peaks of unfolded P protein.

focused on the range around 2 to 100 mM. As shown in the Figure 3-3, the intensity of the extra peaks decreases as the ligand concentration increases, and the chemical shifts of some assigned native peaks, many of which are in the N-terminal region, are strongly dependent on ligand concentration. These two observations indicate that both slow exchange and fast exchange processes contribute to the observed ligand-dependence.
Similar effects are observed when sample pH is varied. As shown in Figure 3-4, the intensities of the unassigned extra peaks decreased and many native peaks shifted as the sample pH was increased from 5.0 to 7.5. The pKa of the free histidine imidazole group is 6.0 (Berg, Tymoczko et al. 2002) which is within the range of pH values used for this experiment. For this reason, it is likely that the changes in native resonance frequencies can be explained by the change of the protonation state of the three histidine sidechains in P protein. Figure 3-4 shows that the amide $^{15}\text{N}/^1\text{H}$ resonances assigned to His 22 and His 105 move as the pH changes, and the His 3 peak is missing in the spectrum, presumably due to intermediate chemical exchange (Henkels 2005). The $^1\text{H}$ or $^{15}\text{N}$ chemical shifts of the two histidine resonances plotted vs. can be fitted to the Henderson–Hasselbalch equation to estimate the pKa’s of His 22 and His 105 which are 6.08±0.02 and 5.65±0.05, respectively (data not shown). Changes in chemical shift were also observed for residues in close proximity to all three histidine residues, which is expected because the ionization state of the histidine sidechain can also affect neighboring residues.

3.2.3 Assignment of extra peaks in the folded P protein HSQC spectrum

Assignment of the extra peaks in the HSQC spectrum (Figure 3-1) allows identification of residues in the unfolded regions of the partially folded intermediate. To this end, we performed standard multidimensional backbone assignment experiments with $^{13}\text{C}/^{15}\text{N}$ labeled protein. A total of 50 extra peaks were observed in the HSQC spectrum of P protein in 20mM sodium cacodylate, 5 mM sodium sulfate, pH 5.0.
Figure 3-3: Sulfate titration of the $^1$H-$^{15}$N HSQC spectrum of $^{15}$N-labeled P protein, at pH 6.0 and 25°C. The spectra of P protein in 5 mM sulfate (red), 10 mM (orange), 20 mM (yellow), and 100 mM (green) are overlaid. The intensities of the extra peaks (between the two dashed line) decrease with increasing sulfate concentrations as evidenced by the disappearance of the yellow and green peaks. The labeled peaks are assigned residues in the native state P protein. The arrows show their movement as the sulfate concentration increases from 5 mM to 100 mM.
Figure 3-4: pH titration of $^{1}H$-$^{15}N$ HSQC spectrum of $^{15}N$-labeled P protein in 10 mM sulfate at 25 °C. The spectra of P protein at pH 5.3 (black), pH 5.5 (red), pH 6.0 (orange), pH 6.5 (yellow), and pH 7.0 (green) are overlaid. The labels for the two observable histidine residues are red. His3 is not observed in the spectrum. Residues next to histidine residues that exhibit significant chemical shift movement are labeled in black with arrows indicating their movement upon increasing pH. All extra peaks are located in the region between the two dashed lines and their intensities decrease with increasing pH.

The clustering of the peaks in the 8 to 8.5 ppm $^{1}H$ chemical shift range resulted in partial overlap of many of the extra peaks. This problem, combined with the relatively weak intensities of the extra peaks, made the assignments more challenging to determine than...
the native resonances. Combination of data from HNCO, HNCACO, and (HCA)CO(NH) experiments, allowed the sequential assignments through the $^{13}$CO nuclei. The HNCACB and CBCA(CO)NH experiments detecting $^{13}$C$_{\alpha}$ and $^{13}$C$_{\beta}$ resonance were also used for further validation of the $^{13}$CO assignment results. There were some short stretches consisting of three to five residues that could not be unambiguously assigned to definite fragments in the protein sequence. To resolve this ambiguity, an H(CCO)NH 3-dimensional experiment was performed to detect the sidechain proton spin system of the i-1 residue and subsequently each fragment was assigned to its correct sequence. This suite of experiments allowed for the assignment of 42 out of the observed 50 extra peaks. Five out of the remaining eight peaks exhibited no $^{13}$C chemical shift connectivity to each other or any of the native peaks. The chemical shifts for the backbone atoms of the 42 assigned peaks and the 8 unassigned peaks are shown in Table 3-1.

Most of the extra peaks were assigned to residues located near the N- and C-terminii in the protein sequence. There is one long N-terminal fragment from Leu 4 to Val 19, which is half way through helix A in the native structure (Figure 3-5). Four short fragments Ala 91-Thr 95, Glu 98- Gln 104, Leu 106-Arg 109, and Leu 112-Ser 116 are located in the C-terminal region, which includes the loop between $\beta$ sheet 4 and helix C and as well as most of the helix C. A short fragment from Lys 52 to Leu 54 is located outside of these two structural regions. As shown in the Figure 3-1, many extra peaks assigned to the same residue type are clustered near each other in the HSQC spectrum, which suggests that these residues are located in a highly solvated magnetic environment in the intermediate.
Figure 3-5: Mapping of the assigned extra peak to the crystal structure of P protein. The residues colored in light green are the extra peaks that are assigned in the study. The residues in dark grey are the residues that might contribute to the extra peaks but could not be unambiguously assigned. The remaining residues in dark green have a single crosspeak in the spectrum.

Chemical shift indices (CSI) for $^{13}C_\alpha$ and $^{13}C_\beta$ assignments can be used to deduce secondary structure from established empirical relationships (Spera and Bax 1991; Wishart and Sykes 1994). $^{13}C_\alpha$ resonances experience a downfield shift relative to
random coil shifts in α-helical conformations and the $^{13}\text{C}_\alpha$ chemical shift is typically upfield of random coil shifts when the residue is in a β-strand conformation. The chemical shift index values as a function of protein sequence for $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$ atoms of the extra peaks are shown in Figure 3-6. The N-terminal region of folded P protein is a flexible structure followed by helix A to residue S25. Most of the $^{13}\text{C}_\alpha$ CSI values of the stretch of 18 assigned extra peaks in this region are smaller than 1 ppm and there is no significant deviation from random coil shifts, indicating that these residues have a random-coil-like structure in the intermediate state. Extra peaks assigned to residues found in the C-terminal helix C region of the native state have $^{13}\text{C}_\alpha$ CSI values indicative of random coil, supporting the conclusion that this region of the intermediate state is also unfolded. The $^{13}\text{C}_\beta$ CSI values are consistent with these conclusions. For the extra peaks assigned to these two regions, the differences in chemical shifts from the native state assignments are greater than 3 ppm. The three residues in the middle of the sequence in Figure 3-6 are colored light green to indicate that their CSI values are not consistent between $^{13}\text{C}_\alpha$ and $^{13}\text{C}$ and some of them are larger than 2 ppm. Low chemical shift dispersion of the extra peak $^1\text{H}/^{15}\text{N}$ HSQC resonances and their low $^{13}\text{C}_\alpha$ and $^{13}\text{C}_\beta$ CSI values are consistent with the conclusion that these residues are unfolded in the partially folded intermediate.
Figure 3-6: Chemical shift analysis of the extra peaks in the folded P protein HSQC spectrum. Residues 52 to 54 are shown in light green because their CSI values are larger than 2 ppm.

3.2.4 Effect of histidine replacement in P protein

As shown in the Figure 3-2, the presence of assigned native peaks in the unfolded P protein spectrum can be explained by the presence of the intermediate state, which exchanges slowly with the unfolded state. In order to investigate how the histidine
residues in the protein perturb the equilibrium between U and I states, different single His to Lys variants (H3K, H22K, and H105K) and a triple His to Lys variant (HtriK) were constructed to study the effects. These variants were made in the background of the F107W variant to allow both NMR and fluorescence stopped-flow experiments. Since the F107W substitution has little effect on the overall structure and stability of P protein (the results in Chapter 2), the $^{1}$$H$-$^{15}$$N$ HSQC spectrum of unfolded F107W is similar to that of WT, including the native peaks observed in the unfolded WT spectrum (circled area in Figure 3-7). A series of pH titration experiments from 5.0 to 7.0 monitored by HSQC spectra were performed to investigate the pH effect on the equilibrium between the U and I states. As shown in the Figure 3-8A, when the pH is increased from 5.0 to 7.0, the intensities of the native peaks also increase. This means that the relative population of I increases as the pH increases. These results suggest that the ionization of histidine residues in the P protein effect the equilibrium transition between U and I, because the histidine side chain is the only likely titratable group in this pH range. The HtriK variant was constructed to mimic the charge state of the F107W variant at pH 5.0, since both histidines should be highly protonated and carry a positive charge at that pH. As shown in Figure 3-8B, the HtriK variant almost completely abolishes the intensities of the native peaks in the HSQC spectrum, which is now independent of pH. This is the expected result if the extra charge on the variant depopulates I. Because P protein is a highly basic protein containing 19 Lys and 11 Arg residues, introducing 3 more Lys residues into the protein increases electrostatic repulsion in the fully or partially folded forms, favoring the unfolded state of the protein.
Subsequent pH titration experiments were performed on each single H→K variant to delineate, which histidine residue(s) account for the observed effects. As shown in the Figures 3-8C to 8E, the spectrum of the H22K variant shows almost the same results as the HtriK variant suggesting that protonation of this histidine, located in the end of the helix A, is the source of the pH-dependence of the U⇌I equilibrium. The spectrum of the H105K shows a pH dependence similar to that of F107W, indicating that the protonation of His 105 has little effect on the stability of I. The pH dependence of peak intensities in the H3K variant spectrum shows that His 3 does have an effect on the population of I, but the effect is smaller than that of H22.

Based on the pH titration experiments performed with these different variants, we conclude that as the pH is raised from 5 to 7, the equilibrium between U and I is shifted toward I primarily by the deprotonation of His 22, with some contribution from His 3 and little or none from H105.
Figure 3-7: \(^1\text{H}^{15}\text{N}\) HSQC spectrum of \(^{15}\text{N}\)-labeled F107W variant of P protein in 20 mM sodium cacodylate, pH 6.0, at 25°C. The peaks in the spectrum are also poorly dispersed as the unfolded WT P protein shown in Figure 3-2. The peaks in the two circles are assigned native state P protein residues also present in the unfolded P protein spectrum. The intensities of these peaks are weaker than those of the remaining unassigned peaks of unfolded P protein.
Figure 3-8: HSQC spectra of various histidine variants of P protein under unfolded conditions at different pH values. The pH titration experiments were done at pH 5.0 (yellow), 6.0 (red) and 7.0 (black) for each variant. The name of the variant is in the bottom right corner of each plot. The spectra are overlaid and the peaks in the circles area are assigned native peaks of F107W. As shown in the plots, the intensities of the peaks in the circled area increase as the pH increases.
3.3 Discussion

3.3.1 Structural description of the intermediate state

Figure 3-5 shows the residues assigned to the extra peaks present in the folded P protein $^1$H-$^{15}$N HSQC spectrum mapped on the crystal structure of P protein. The extra peak residues are colored in light green on the structure and most are located in the N-terminal (residues 3 to 19) and C-terminal regions (residues 91 to 116). Three of the extra peaks were assigned to Lys 52 to Leu 54, which are located in the middle of loop 3. The low chemical shift dispersion of the extra $^1$H-$^{15}$N HSQC resonances and their low $^{13}$C$_\alpha$ and $^{13}$C$_\beta$ CSI values indicate that the associated residues are part of random-coil-like structure in I. It should be noted that each of the assigned residues in these three regions has two separate crosspeaks in the HSQC spectrum. One peak corresponds to the native state resonance assigned in previous studies (Henkels 2005; Henkels and Oas 2006), and the other corresponds to a resonance whose magnetic environment appears to be very similar to that of an unfolded protein. No extra peaks were assigned to the residues colored in dark green in Figure 3-5, suggesting that these residues retain a native-like magnetic environment in both the intermediate and native states. The residues colored dark grey in Figure 3-5 (Phe 20, Lys 21, His 22, Tyr 96, Glu97, His 105, Ser 110, and Ser 111) are near the N and C termini, but cannot be unambiguously assigned to any of the extra peaks. Residues unfolded in I would not have an extra peak if the difference between their folded and unfolded frequencies is small, i.e. if their native state frequencies happen to be near the random coil values. This might be the case for residues 20, 21, 22, 105, 110, and 111, since their native state $^1$H chemical shifts are between 8
and 8.5 ppm, near the random coil values. This explanation, however, is not valid for Tyr 96 and Glu 97, because their $^1$H chemical shifts are down field at 9.7 and 9.5 ppm, respectively. An alternative explanation could be that since only 42 out of 50 extra peaks have been assigned, the remaining unassigned eight peaks might correspond to these residues. The intensities of these unassigned peaks are weakest among the extra peaks in the HSQC spectrum and they correspond to even weaker or missing crosspeaks in the three-dimensional heteronuclear triple-resonance experiments. Therefore, it was difficult to obtain sequential assignments for these eight peaks. An ideal solution would be assignment of the spectrum of completely unfolded P protein allowing the identification of the remaining unassigned peaks. The last possibility is that these 8 peaks belong to residues outside the N and C-terminal regions and could be some of the residues colored dark green in Figure 3-5.

Based on the results of the NMR experiments, the likely structure of the intermediate includes a native-like central $\beta$-sheet and $\alpha$-helix B. Interestingly, helix B contains the conserved RNR motif in the P protein and has many crucial contacts with P RNA (Niranjanakumari, Day-Storms et al. 2007). The central $\beta$-sheet also has many interactions with pre-tRNA leading sequence during the catalysis (Crary, Niranjanakumari et al. 1998; Niranjanakumari, Stams et al. 1998). Given the preservation of the structural elements key to RNA interactions in the intermediate state, it may play a role in the holoenzyme assembly process. The structured RNR motif, which is preserved in the intermediate state, may facilitate the association of this partially folded form of P protein with P RNA.
3.3.2 Unstructured N and C termini of I may fold independently to N

As shown in Figure 3-5, α-helix A and α-helix C are spatially separated from each other. It is possible that the two α-helices fold independently so it is possible that there are three intermediate states: both helices unfolded, helix A unfolded, and helix B unfolded. There are some hints in the data that this might be the case. As shown in Figure 3-3, as additional sulfate is added to P protein all the extra peaks disappear at roughly the same sulfate concentration. However, a more detailed comparison of the extra peaks from the N- and C-termini, indicates that the C-terminal residues completely disappear between 10 and 20 mM, and N-terminal residues completely disappear above 20 mM. In addition, Figure 3-2 shows HSQC resonances the unfolded P protein spectrum that appear to correspond to peaks in the native spectrum previously assigned to residues E97, Leu 106, and Leu 112 found in the C-terminal helix and which are observed as extra peaks in the sulfate folded spectrum. Taken together, these results suggest that another intermediate conformation might exist with helix A unfolded and helix C folded in the structure. The kinetic mechanism developed in Chapter 2 is the simplest model that can describe the experimental data but it can’t exclude the possibility that the intermediate state is an ensemble of conformations that interconvert rapidly relative to the rates of complete folding or unfolding. If this were the case, the kinetic experiments would not be able to distinguish multiple intermediate conformations. For this reason, we will still treat I as a single thermodynamic state with as yet undetermined conformational heterogeneity.

3.3.3 Effect of histidine ionization on conformational equilibria

The results of the pH titration experiments of different histidine variants
indicate that both His 3 and His 22 perturb the relative populations of U and I between pH 5.0 and 7.0 and His 22 has the major effect on this equilibrium. Since only three different pH values were examined for each variant, the NMR experiments provide only qualitative information about the effect of histidine ionization on stability of U and I. As shown in Figures 3-4 and 3-8, the effect of increasing pH (i.e., deprotonating the histidines) is to shift the I ⇄ N conformational equilibrium toward N in the presence of sulfate, and the U ⇄ I equilibrium toward I in the absence of sulfate. These effects likely arise from unfavorable electrostatic interactions in more compact forms of the protein. At pH 5.0 in the absence of sulfate, extra repulsive forces due to additional net positive charge from the protonated histidine residues presumably raises the free energy of I relative to the less compact U, so P protein favors the U state. When the pH is increased, the neutralization of the histidines relieves some repulsive forces and lowers the free energy of I relative to U. The same principle can be used to explain the effect of pH on the I to N equilibrium in the presence of sulfate.

In order to obtain more quantitative estimates of the effect of pH on U, I and N free energies in F107W and the histidine variants, a more finely stepped pH titration followed by NMR is necessary. The relative populations of U and I could be estimated as a function of pH using the relative peak intensities in the HSQC spectrum for each variant. The results for each variant could then be compared with that of F107W, and the energetic contributions of each histidine quantitatively determined.
3.4 Materials and Methods

3.4.1 Chemicals and reagents

Ultra pure urea was purchased from Nacalai Tesque Inc. $^{15}$N ammonium chloride (99%), $^{13}$C glucose (U-13C6, 99%), and deuterated TMAO (d9, 98%) were purchased from Cambridge Isotope Laboratories Inc. Urea and TMAO concentration was measured by refractive index using the equation from (Warren and Gordon 1966) and (Wang and Bolen 1997), respectively.

3.4.2 Protein mutagenesis, expression and purification

Histidine variants of P protein were generated in a F107W background. These site-directed mutational plasmids were constructed using QuikChange® procedure (Stratagene) with the primer GGAGATATACCATGGCTAAGCTGAAAAAACGC and GCGTTTTTTTCAGCTTAGCCATGGGTATATCTCC for H3K, and CCAGAAAGTGTAAAAAAGGGGACATCAGTTGC and GCAACTGATGTCCCTTTTTAAACACTTTTCG for H22K, and CGAAAAAAGTCTGCAAGAAGCTATGGAGAAAGTCTTC and GAAGACTTTTCTCCATAGCTTCTGCAGACTTTTTCG for H105K. Single and triple histidine mutations in F107W background were overexpressed and produced P protein variants. The sequence of each variant was confirmed to assure it contains the desired substitutions.

Wild type P protein and variants were overexpressed in E. coli (BL21 (DE3) pLysS) cells. For preparation of $^{15}$N single label and $^{15}$N, $^{13}$C double labeled protein used in the NMR experiments, a single colony of positive transformants was used to inoculate
30mL LB media containing 50 µg/mL kanamycin and 17 µg/mL chloremphenicol till OD$_{600}$ around 0.8-0.9. The culture was diluted (1:200) into the M9 minimum media culture (1L) containing the same antibiotics and enriched with 2g/L $^{13}$C glucose and/or 1g/L $^{15}$N ammonium chloride. The culture was incubated until the OD$_{600}$ is 0.4-0.6, and the protein overexpression was induced by adding IPTG to a final concentration of 0.4 mM. After the induction, the culture was incubated for another 10-12 hour, and the cell was harvested by centrifugation. Overexpressed P protein was purified as described (Niranjanakumari, Kurz et al. 1998) with the following modifications. The pooled fractions of P protein that eluted from second CM-Sepharose column was concentrated using centriprep (Amicon) until the volume of the pooled fractions was reduced to 1~2 mL. The buffer in the sample was changed to 6 M Guanidine-HCl, 10mM Tris (pH 7.5) during this concentrating step. The concentrated sample was then loaded on an S-100 column (company?) equilibrated with 6 M Guanidine-HCl, 10 mM Tris (pH 7.5) to remove EDTA that bound to P protein. Fractions containing P protein were pooled and concentrated again using centriprep to desired concentration for NMR experiments. Protein purity was determined > 99% by SDS-PAGE. The molecular weight of each variant was confirmed by electrospray ionization mass spectrometer. Unlabeled proteins were ±0.5 Da, and labeled proteins were 3-4 Da different from the theoretical mass calculation, respectively. Protein was kept in 6 M Gdn-HCl at -80 °C for long term storage. The protein NMR sample was dialyzed extensively against water and then 20 mM sodium cacodylate (with proper pH value needed in the experiments). In all experiments, the protein concentration was calculated using Edelhoch method (Edelhoch
with an extinction coefficient of 11200 M$^{-1}$ cm$^{-1}$ at 276 nm.

### 3.4.3 NMR spectroscopy experiments

All three-dimensional NMR spectra were collected at 25 °C on a Varian INOVA 600 spectrometer with a triple resonance cryoprobe equipped with a z-field gradient coil. The two-dimensional spectra were either collected on a Varian INOVA 600 or a Varian INOVA 800 spectrometer with a triple resonance probe. The NMR sample used for assignment of the extra peaks consisted of 0.5 mM P protein in 5 mM sodium sulfate, 20 mM sodium cacodylate buffer at pH 5.0, 10% D$_2$O and 0.05% azide. The NMR sample used for the pH titrations of the unliganded protein consisted of 0.2-0.4 mM P protein in 20 mM sodium cacodylate buffer, 10% D$_2$O and 0.05% azide. The pH ranged from 5 to 7.5. Two-dimensional gradient-enhanced sensitivity-enhanced $^1$H-$^{15}$N HSQC experiments were collected with a spectral width in the $^1$H dimension of 11990 Hz and 1024 complex points and a spectral width in the $^{15}$N dimension of 1994 Hz and 192 complex points. The resonance assignments of extra peaks were made using a suite of triple resonance experiments including (HCA)CO(CA)NH, HNCACO, CBCA(CO)NH, HNCO and H(CCO)NH. In these three-dimensional NMR experiments, the spectral width in the $^1$H dimension was 7020 Hz with 1024 complex points and the spectral width of $^{15}$N dimension was 1944 Hz with 32,64 or 72 complex points. The spectral width in $^{13}$C dimension was 2000 Hz with 48,64 and 96 complex points for (HCA)CO(CA)NH, HNCACO, and HNCO, respectively. The spectral width of $^{13}$C dimension was 12066 Hz with 64 complex points for the CBCA(CO)NH experiment. The NMR spectra were processed using NMRpipe, and the processed spectra were viewed and examined by...
NMRDraw and NMRviewJ. Backbone sequential assignments were analyzed using CARA (CARA can be downloaded for free from www.nmr.ch.).

Table 3-1: Backbone chemical shifts of the extra peaks in the folded P protein HSQC spectrum.

<table>
<thead>
<tr>
<th>Residue</th>
<th>$^1$H (ppm)</th>
<th>$^{15}$N (ppm)</th>
<th>$^{13}$C$_\alpha$ (ppm)</th>
<th>$^{13}$C$_\beta$ (ppm)</th>
<th>$^{13}$CO (ppm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leu4</td>
<td>8.42</td>
<td>124.68</td>
<td>55.11</td>
<td>42.35</td>
<td>176.92</td>
</tr>
<tr>
<td>Lys5</td>
<td>8.40</td>
<td>123.25</td>
<td>56.04</td>
<td>32.93</td>
<td>176.33</td>
</tr>
<tr>
<td>Lys6</td>
<td>8.38</td>
<td>123.55</td>
<td>56.15</td>
<td>32.92</td>
<td>176.43</td>
</tr>
<tr>
<td>Arg7</td>
<td>8.45</td>
<td>122.78</td>
<td>56.02</td>
<td>30.75</td>
<td>175.95</td>
</tr>
<tr>
<td>Asn8</td>
<td>8.53</td>
<td>120.66</td>
<td>53.21</td>
<td>38.83</td>
<td>175.04</td>
</tr>
<tr>
<td>Arg9</td>
<td>8.38</td>
<td>121.97</td>
<td>56.10</td>
<td>30.59</td>
<td>176.02</td>
</tr>
<tr>
<td>Leu10</td>
<td>8.23</td>
<td>123.17</td>
<td>55.09</td>
<td>42.16</td>
<td>177.12</td>
</tr>
<tr>
<td>Lys11</td>
<td>8.25</td>
<td>122.73</td>
<td>56.04</td>
<td>32.92</td>
<td>176.25</td>
</tr>
<tr>
<td>Lys12</td>
<td>8.37</td>
<td>123.03</td>
<td>56.21</td>
<td>32.92</td>
<td>176.24</td>
</tr>
<tr>
<td>Asn13</td>
<td>8.50</td>
<td>120.16</td>
<td>53.50</td>
<td>38.60</td>
<td>175.25</td>
</tr>
<tr>
<td>Glu14</td>
<td>8.43</td>
<td>121.41</td>
<td>56.66</td>
<td>30.12</td>
<td>175.96</td>
</tr>
<tr>
<td>Asp15</td>
<td>8.25</td>
<td>121.18</td>
<td>54.47</td>
<td>40.84</td>
<td>176.21</td>
</tr>
<tr>
<td>Phe16</td>
<td>8.07</td>
<td>120.61</td>
<td>58.04</td>
<td>39.18</td>
<td>175.82</td>
</tr>
<tr>
<td>Gln17</td>
<td>8.17</td>
<td>120.65</td>
<td>55.91</td>
<td>29.16</td>
<td>175.76</td>
</tr>
<tr>
<td>Lys18</td>
<td>8.08</td>
<td>121.91</td>
<td>56.39</td>
<td>32.82</td>
<td>176.18</td>
</tr>
<tr>
<td>Val19</td>
<td>7.89</td>
<td>120.61</td>
<td>62.03</td>
<td>32.79</td>
<td>175.59</td>
</tr>
<tr>
<td>Gly23</td>
<td>8.45</td>
<td>110.18</td>
<td>45.19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thr24</td>
<td>8.12</td>
<td>113.70</td>
<td>61.65</td>
<td>69.74</td>
<td>174.10</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td></td>
</tr>
<tr>
<td>Lys52</td>
<td>8.15</td>
<td>121.82</td>
<td>61.96</td>
<td>32.58</td>
<td>175.56</td>
</tr>
<tr>
<td>Lys53</td>
<td>8.30</td>
<td>124.33</td>
<td>57.36</td>
<td>39.54</td>
<td>175.37</td>
</tr>
<tr>
<td>Leu54</td>
<td>8.20</td>
<td>123.39</td>
<td>55.99</td>
<td>33.05</td>
<td>175.86</td>
</tr>
<tr>
<td>Ala91</td>
<td>8.47</td>
<td>124.59</td>
<td>52.83</td>
<td>18.90</td>
<td>178.03</td>
</tr>
<tr>
<td>Ser92</td>
<td>8.24</td>
<td>114.30</td>
<td>58.30</td>
<td>63.51</td>
<td>174.54</td>
</tr>
<tr>
<td>Gln93</td>
<td>8.31</td>
<td>122.09</td>
<td>55.79</td>
<td>29.27</td>
<td>175.86</td>
</tr>
<tr>
<td>Leu94</td>
<td>8.14</td>
<td>123.06</td>
<td>55.43</td>
<td>42.28</td>
<td>177.62</td>
</tr>
<tr>
<td>Thr95</td>
<td>8.16</td>
<td>114.01</td>
<td>61.65</td>
<td>69.85</td>
<td>174.71</td>
</tr>
<tr>
<td>Glu98</td>
<td>8.21</td>
<td>121.16</td>
<td>57.52</td>
<td>30.09</td>
<td>177.73</td>
</tr>
<tr>
<td>Thr99</td>
<td>8.18</td>
<td>115.73</td>
<td>63.62</td>
<td>69.19</td>
<td>175.32</td>
</tr>
<tr>
<td>Lys100</td>
<td>8.10</td>
<td>122.92</td>
<td>57.32</td>
<td>32.29</td>
<td>177.37</td>
</tr>
<tr>
<td>Lys101</td>
<td>8.03</td>
<td>120.92</td>
<td>57.33</td>
<td>32.59</td>
<td>177.34</td>
</tr>
<tr>
<td>Ser102</td>
<td>8.11</td>
<td>115.70</td>
<td>59.02</td>
<td>63.27</td>
<td>175.07</td>
</tr>
<tr>
<td>Leu103</td>
<td>8.07</td>
<td>123.14</td>
<td>55.53</td>
<td>41.99</td>
<td>177.71</td>
</tr>
<tr>
<td>Gln104</td>
<td>8.04</td>
<td>119.27</td>
<td>56.67</td>
<td>28.96</td>
<td>176.23</td>
</tr>
<tr>
<td>Leu106</td>
<td>7.94</td>
<td>122.12</td>
<td>55.43</td>
<td>42.10</td>
<td>177.07</td>
</tr>
<tr>
<td>Phe107</td>
<td>8.13</td>
<td>120.48</td>
<td>57.75</td>
<td>39.25</td>
<td>175.73</td>
</tr>
<tr>
<td>Arg108</td>
<td>8.06</td>
<td>122.59</td>
<td>55.99</td>
<td>30.75</td>
<td>175.95</td>
</tr>
<tr>
<td>Lys109</td>
<td>8.30</td>
<td>124.19</td>
<td>54.21</td>
<td>30.81</td>
<td>175.97</td>
</tr>
<tr>
<td>Leu112</td>
<td>8.04</td>
<td>123.32</td>
<td>55.53</td>
<td>42.14</td>
<td>176.93</td>
</tr>
<tr>
<td>Tyr113</td>
<td>7.93</td>
<td>120.35</td>
<td>57.69</td>
<td>38.60</td>
<td>175.30</td>
</tr>
<tr>
<td>Lys114</td>
<td>7.99</td>
<td>123.83</td>
<td>55.84</td>
<td>33.22</td>
<td>175.84</td>
</tr>
<tr>
<td>Lys115</td>
<td>8.28</td>
<td>123.54</td>
<td>56.29</td>
<td>33.02</td>
<td>176.59</td>
</tr>
<tr>
<td>Ser116</td>
<td>8.28</td>
<td>116.62</td>
<td>58.39</td>
<td>63.63</td>
<td>174.80</td>
</tr>
<tr>
<td>Und1</td>
<td>8.09</td>
<td>121.45</td>
<td>62.27</td>
<td>32.64</td>
<td>176.21</td>
</tr>
</tbody>
</table>

82
<table>
<thead>
<tr>
<th>Und</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Und2</td>
<td>8.12</td>
<td>120.04</td>
<td>54.50</td>
<td>40.90</td>
<td>175.88</td>
</tr>
<tr>
<td>Und3</td>
<td>8.41</td>
<td>121.29</td>
<td>57.54</td>
<td>30.06</td>
<td>177.19</td>
</tr>
<tr>
<td>Und4</td>
<td>8.25</td>
<td>128.08</td>
<td>52.31</td>
<td>19.11</td>
<td>177.22</td>
</tr>
<tr>
<td>Und5</td>
<td>7.87</td>
<td>119.73</td>
<td>57.90</td>
<td>38.59</td>
<td>175.45</td>
</tr>
<tr>
<td>Und6</td>
<td>7.82</td>
<td>119.57</td>
<td>55.21</td>
<td>31.69</td>
<td>175.27</td>
</tr>
<tr>
<td>Und7</td>
<td>7.86</td>
<td>122.28</td>
<td>61.13</td>
<td>38.41</td>
<td>175.46</td>
</tr>
<tr>
<td>Und8</td>
<td>8.30</td>
<td>127.71</td>
<td>52.14</td>
<td>19.27</td>
<td></td>
</tr>
</tbody>
</table>
Chapter 4: Mechanism of Conformational Change Coupled to Ligand Binding by RNase P protein

4.1 Introduction

Proteins are flexible molecules that change conformation upon ligand binding (Hammes 2002; Teague 2003). From a thermodynamic perspective, this linkage is the mechanism by which the favorable free energy of ligand binding is reduced by the unfavorable free energy of conformational change. There are several reasons why a conformational change might be required for binding. (1) Binding-induced conformational changes are able to integrate signals from multiple ligands. For example, the conformational changes of R and T forms of hemoglobin induced by oxygen binding promote positive binding cooperativity, an example of allostery (Wyman 1964). (2) Binding-induced conformational changes may take place because the high affinity protein conformation is a high energy species in the absence of the ligand. This reason may be particularly relevant for nucleic acid binding proteins. These proteins are often highly positively charged on the protein surface and also in sequence composition, and usually form salt bridges with nucleotide phosphates (Frankel and Young 1998). (3) The binding site in the conformation of ligand bound protein may be completely surrounded by protein sidechains and backbone, so the binding site would be inaccessible if the conformation were formed prior to ligand binding (Fersht 1974).

The degree of conformational change in ligand binding proteins has significant variation. A database study has shown that movements in enzymes upon substrate binding are generally small (Gutteridge and Thornton 2005). However, another
recent study has shown that the extent of the movement may depend on the actual reaction mechanism (Koike, Amemiya et al. 2008). Many proteins also exhibit domain rearrangements upon binding to substrates (Qi and Hayward 2009). With the exponential increase of bioinformatic analyses of genome sequences and protein sequences from different organisms, many proteins have been labeled partially or completely disordered under physiological conditions (Dunker, Oldfield et al. 2008). These protein systems appear to exploit coupled ligand binding and protein conformational change to its extreme: they are entirely unfolded or contain long disordered fragments in the absence of their target ligands and undergo disorder-to-order transitions upon ligand or target binding (Dunker, Silman et al. 2008). Regardless of the extent of movement of the conformational change upon ligand binding in different systems, the fundamental principle that describes the thermodynamic properties and kinetic mechanisms of the processes are the same. In this chapter, we focus mainly on using RNase P protein to study the coupled protein folding and ligand binding process, although the results should be generalizable to coupled conformational changes and binding equilibria.

Although there is growing literature on the identification and characterization of IUPs and their complexes, coupled folding and binding mechanisms are poorly understood. The sequence of events in binding, the role of intermediates and the kinetics of the relevant transitions are generally not known. Recently, several studies have attempted to investigate whether IUP folding precedes binding or vice versa using different biophysical method such as NMR, stopped-flow, and laser temperature jump (Sugase, Dyson et al. 2007; Narayanan, Ganesh et al. 2008; Onitsuka, Kamikubo et al. 2008).
From the results of these studies, it is still unclear which mechanism is more common, and it is evident that the coupled folding and binding mechanisms for different IUPs may be qualitatively different.

RNase P holoenzyme is a ribonucleoprotein (RNP) complex that has diverse functions and carries out critical processes in the living cell. Given the biological importance of RNPs, it is crucial to understand the mechanisms of their assembly and degradation processes. Several studies have addressed the process of ribosome assembly (Talkington, Siuzdak et al. 2005; Williamson 2005), but there have been few on smaller RNP systems and almost no work on the role of protein folding in the process. In the present chapter, we utilize P protein as a model system to understand the kinetic mechanism of a coupled folding and binding process. Although the cognate binding partner of P protein is P RNA in vivo, a small anion (sulfate) is used as a surrogate ligand for the in vitro kinetic experiments described here. Building on the P protein ligand-free folding mechanism described in Chapter 2, a full mechanism of coupled P protein folding and small anion binding is constructed. The kinetic data were fitted to a minimal kinetic mechanism based on best-fit estimates of equilibrium association constants and kinetic rate constants. We use the concept of flux to interpret the experimental results and determine the predominant pathway in the mechanism. The results indicate that the coupled folding and binding pathway of P protein is highly ligand concentration dependent, so the sequence of the binding and folding events also depends on the ligand concentration. Finally, based on these results, a general mechanism describing the RNase P holoenzyme assembly process is proposed.
4.2 Results

4.2.1 Fluorescence spectrum and fluorescence equilibrium sulfate titration of F107W

As demonstrated in Chapter 2, the fluorescent F107W variant of P protein showed distinct fluorescence signal differences between the native and unfolded states when trimethylamine-N oxide (TMAO) was used as a renaturant. P protein can also be induced to fold by addition of various small anions. As shown in Figure 4-1A, the fluorescence emission spectra of P protein in 20 mM Na₂SO₄ and in 1M TMAO are nearly identical, indicating that the structural environment around the Trp residue is very similar under these folding conditions and very different from that of the unfolded protein. The circular dichroism spectra of sulfate- and TMAO-folded P protein are virtually indistinguishable and distinctly different from that of unfolded protein (Figure 4-1B). These results support the conclusion in Chapter 2 that sulfated and osmolyte folded P protein adopt the same folded structure.

The progressive addition of Na₂SO₄ to unfolded P protein shifts its equilibrium toward the folded form, as detected by tryptophan fluorescence. As shown in Figure 4-2, titration with sulfate causes the fluorescence of P protein to decrease and reaches saturation around 10 mM. The midpoint of the F107W variant’s titration curve (~0.8 mM) is similar to the midpoint of the wild type protein titration curve under similar conditions (Henkels, Kurz et al. 2001), indicating that the substitution of a tryptophan for phenylalanine at position 107 causes minimal destabilization of the folded
Figure 4-1: Fluorescence and circular dichroism spectra of sulfate and TMAO folded F107W P protein. (A) The continuous black line is the 20 mM sulfate folded F107W spectrum and short-dashed line is 1 M TMAO folded F107W. The long-dashed line is the spectrum of F107W P protein in buffer only.  (B) CD spectrum of F107W P protein in the same three conditions. The spectra of sulfate folded (filled circle) and TMAO folded (open rectangular) F107W P protein overlap and are nearly identical to each other. The open circle trace is the spectrum of F107W P protein in buffer alone.
structure. This conclusion is consistent with the thermodynamic analysis of F107W P protein folding described in Chapter 2. In addition, these results suggest that the tryptophan substitution does not significantly perturb the intrinsic sulfate binding affinity of P protein.

4.2.2 Ligand induced folding kinetics

Stopped-flow fluorescence experiments to measure ligand induced folding
were performed with unfolded P protein in 20 mM sodium cacodylate buffer, pH 7, mixed 1:1 with 0.1 – 400 mM Na₂SO₄ in the same buffer (sulfate up jump experiments). Data collected at final Na₂SO₄ concentrations in the range of 0.025–0.1 mM were obtained by mixing a solution of sulfate-bound protein with sulfate-free buffer at a protein:buffer ratio of 1:10 (sulfate down jump experiments). Figures 4-3A and 4-3B show representative kinetic traces for the sulfate up jump and down jump data, respectively. All kinetic traces were fitted to a triple exponential function using the nonlinear least squares fitting method (see Materials and Methods), and the quality of the fit was judged by visual inspection of the fitting residuals as shown in the inset panels of Figure 4-3. Among the three phases detected in the kinetic traces, the slowest always had the smallest amplitude and a rate constant of ~0.2 s⁻¹, which was relatively independent of sulfate concentration. This result is comparable to that observed during the TMAO induced folding analysis of P protein described in Chapter 2, so we attribute this slowest phase to proline isomerization which is kinetically separated from the other two major phases. Therefore, the analysis described here will focus on the two major phases (fast and slow), which account for more than 95% of the total fluorescence change in the transient. The observed rate constants ($k_{obs}$) for the fast and slow phases of the folding/binding reaction are plotted versus sulfate concentration in Figure 4-4. The $k_{obs}$ for the unfolding/folding reaction in the absence of sulfate (0 M sulfate points) constraints were obtained from the TMAO induced folding experiments described in Chapter 2. Observations obtained from sulfate up jump to 0.1 mM and sulfate down jump to 0.09 mM experiments are in good agreement (the lowest two concentration points of
Figure 4-3: Stopped-flow fluorescence transient kinetic traces. (A) Sulfate up-jump experiments. The red line is the best-fit results to a triple exponential result. The inset panel shows the residuals of the fitting. (B) Sulfate down-jump experiments. The red line is the best-fit results to also a triple exponential function. The inset panel shows the residuals of the fitting.
Figure 4-4: Sulfate-dependence of observed rate constants of F107W P protein obtained from the fitting results of the stopped-flow kinetic traces. Open circles and closed circles are the fast and slow observed rate constants, respectively. The data are plotted on a log-log scale. The smooth lines through the two data sets are the best-fit results using Equation Set 2, and the estimated parameters from the fitting are listed in Table 4-1.

both phases in Figure 4-4), which confirms the reversibility of the reaction.

4.2.3 Kinetic model for the coupled folding/binding of P protein

Detailed analysis of the TMAO-induced P protein folding kinetics reported in Chapter 2 showed the existence of an intermediate state. Previous experiments demonstrated that there are two anion-binding sites responsible for inducing the folding of P protein (Henkels, Kurz et al. 2001). In order to interpret the sulfate dependence of the $k_{obs}$ values reported here, we propose the minimal complexity kinetic mechanism of folding/binding shown in Scheme 4-1. Under the most general assumption,
each of the three conformational states of P protein contains two binding sites. In Scheme 4-1 the vertical pathways depict the sequential folding reaction of unliganded protein, singly liganded protein, and doubly liganded protein. The three horizontal pathways describe the binding events of each conformational state of P protein. The full kinetic mechanism involving the nine states depicted in Scheme 4-1 has eight associated rate constants. The observation of only two phases in our stopped flow transients suggests that six of the $k_{obs}$ anticipated by Scheme 4-1 are related to ligand association ($k_{on}$) and dissociation ($k_{off}$) rates and are too fast to be detected by the stopped-flow instrument under our experimental conditions. Thus, the two experimentally observed rate constants are related the rate constants associated with the vertical folding/unfolding transitions depicted in Scheme 4-1.

The number of parameters and the complexity of our kinetic model can be greatly simplified by assuming a priori that the rates of sulfate association and dissociation are much higher than those of P protein folding and unfolding. This is a reasonable assumption since $k_{on}$ is likely to be diffusion controlled with a second order
rate constant of $\sim 10^8 \text{M}^{-1}\text{s}^{-1}$, which converts to a pseudo-first order rate constant of $10^4 - 4 \times 10^7 \text{s}^{-1}$ for the Na$_2$SO$_4$ concentration range used in our experiments (100 µM – 400 mM). These rate constants are at least two orders of magnitude greater than the rate constants observed in the stopped flow transient. Thus, the horizontal reactions depicted in Scheme 4-1 are treated as rapid pre-equilibria in the analysis describe below. This approach has been used previously to analyze the O$_2$ binding kinetics of hemoglobin (Eigen and Hammes 1963).

4.2.4 Derivation and explanation of the kinetic model

The observed rate constants of complex kinetic mechanisms are straightforward to compute using the rate matrix method (Dill 1999). The solutions are the eigenvalues of the rate matrix associated with the kinetic mechanism. The analytical solutions of the eigenvalues of the 9 by 9 rate matrix in Scheme 1 are cumbersome and not informative in terms of explaining the fitting results. However, as pointed out above, it is reasonable to assume the sulfate binding events are orders of magnitude faster than folding of P protein. Under the pseudo first order condition ([Ligand]>>[Protein]), we can derive a simplified form of the equations that describe the ligand concentration dependence of the $k_{obs}$ and Scheme 1 can be re-written into the more compact form as Scheme 4-2:

$$
\begin{align*}
U^{tot} & \rightleftharpoons \frac{k_{agg}}{k_{agg}} U^I \\
I^{tot} & \rightleftharpoons \frac{k_{agg}}{k_{agg}} I^{N} \\
N^{tot} & \rightleftharpoons \frac{k_{agg}}{k_{agg}} N^{tot}
\end{align*}
$$

Scheme 4-2

In this scheme, $U^{tot}$ represents all forms of $U$ in the mechanism, and the same
applies $I^{tot}$ and $N^{tot}$. The four aggregate folding and unfolding rate constants in Scheme 4-2 can be determined by computing the population weighted averaged rate constant from Scheme 4-1 using Equation Set 1:

\[
\begin{align*}
    k_{ui}^{agg} &= P_{i}k_{ui} + P_{ul}k_{ui} + P_{ul2}k_{ui}^{2l} \\
    k_{iu}^{agg} &= P_{i}k_{iu} + P_{il}k_{iu} + P_{il2}k_{iu}^{2l} \\
    k_{in}^{agg} &= P_{i}k_{in} + P_{il}k_{in} + P_{il2}k_{in}^{2l} \\
    k_{ni}^{agg} &= P_{i}k_{ni} + P_{nl}k_{ni} + P_{nl2}k_{ni}^{2l}
\end{align*}
\]  

(Equation Set 1)

where $P_i$ are the population of each state as a fraction of all liganding forms of the corresponding conformation and $k_i$ are the rate constants associated with the state (reactant). For the linear three state mechanism depicted in Scheme 4-2, there are two expected observed rate constants (Eigen and Hammes 1963):

\[
\begin{align*}
    k_{obs1} &= \frac{1}{2} \left( k_{ui}^{agg} + k_{iu}^{agg} + k_{in}^{agg} + k_{ni}^{agg} + \sqrt{(k_{ui}^{agg} + k_{iu}^{agg} - k_{in}^{agg} - k_{ni}^{agg})^2 + 4k_{ui}^{agg}k_{in}^{agg}} \right) \\
    k_{obs2} &= \frac{1}{2} \left( k_{ui}^{agg} + k_{iu}^{agg} + k_{in}^{agg} + k_{ni}^{agg} - \sqrt{(k_{ui}^{agg} + k_{iu}^{agg} - k_{in}^{agg} - k_{ni}^{agg})^2 + 4k_{ui}^{agg}k_{in}^{agg}} \right)
\end{align*}
\]  

(Equation Set 2)

These two equations describe the two observed rate constants measured from the kinetic experiments in terms of the six intrinsic rate constants associated with the conformational transitions. The populations in Equation Set 1 can be calculated as a fraction of all forms of a particular conformation. For example, $P_{ul}$ can be calculated using the concentration of UL divided by the total concentration of all forms of U:

\[
P_{ul} = \frac{[UL]}{[U]+[UL]+[UL_2]}  
\]  

(Equation 3)

Similarly, the fractional population of all the other species can be obtained. The concentration of all species can be written in terms of macroscopic association constants.
and the concentration of the unliganded form of each species:

\[
[UL] = K_{UL} [U][L]
\]

\[
[UL_2] = K_{UL} K_{UL_2} [U][L]^2
\]

(Equation Set 4)

where \( K_{UL} \) is the macroscopic association constant for the binding of one ligand to either site of the unliganded protein and \( K_{UL_2} \) is the macroscopic binding constant for the second ligand binding to the remaining site. The equations for the I and N species can be expressed the same way. Substituting Equation Set 4 into Equation 3 and the result into Equation Set 1 gives the expression of all four aggregate rate constants in terms of all the equilibrium constants and rate constants in Scheme 4-1. Substituting Equation Set 1 into Equation Set 2 gives an expression for the ligand concentration dependence of the \( k_{obs} \) in terms of 18 parameters. However, not all of these are independent of each other. Four detailed balance equations can be used to express the four unfolding rate constants of liganded I and N species in terms of other parameters:

\[
k_{IU}^L = \frac{k_{IU}^L K_{UL}}{k_{UI} \ K_{IL}}
\]

\[
k_{NI}^L = \frac{k_{NI}^L K_{IL}}{k_{IN} \ K_{NL}}
\]

\[
k_{IU}^{2L} = \frac{k_{IU}^{2L} K_{UL_2}}{k_{UI}^L \ K_{IL_2}}
\]

\[
k_{NI}^{2L} = \frac{k_{NI}^{2L} K_{IL_2}}{k_{IN}^L \ K_{NL_2}}
\]

(Equation Set 5)

With these substitutions, Equation Set 2 was used to simultaneously fit the data for the two observed rate constants vs. \([Na_2SO_4]\) shown in Figure 4-4 using the NonlinearRegress function of Mathematica™ with weights based on the estimated \( k_{obs} \) uncertainties (see Methods.) The four folding/unfolding rate constants associated with U, I and N state transitions were determined in the folding kinetic studies of unliganded P protein described in Chapter 2. As a further strategy to reduce the number of adjustable
parameters, our previous equilibrium-based estimate of $2.2\pm1 \times 10^4 \text{ M}^{-1}$ for the microscopic binding constant ($K_N$) was used to estimate both $K_{NL} (=2K_N)$ and $K_{NL2} (=K_N/2)$. With these six previously determined parameters held fixed, the number of adjustable parameters was reduced to eight. Because initial fitting efforts did not converge to unique values for these parameters, further simplifications were made as follows. Because the U state anion binding sites are expected to be highly disordered, it is reasonable to assume that the microscopic binding constants of the two binding sites in the U state are similar and weak. In that case, the two macroscopic binding constants associated with U species will have the relationship $K_{UL} \cong 4K_{UL2}$. The NMR studies described in the Chapter 3 indicate that the N terminal region of the intermediate state is unfolded. This N terminal region forms part of the sulfate binding site observed in the x-ray crystal structure of P protein (Stams, Niranjanakumari et al. 1998). NMR results described in Chapter 3 and here (see below) also indicate that the second anion binding site is intact in the intermediate state. These observations support the assumption that the two microscopic binding constants for I are significantly different from each other and are reasonable approximations of the macroscopic binding constants $K_{IL}$ and $K_{IL2}$. Fixing these three binding constants to the ranges $K_{UL} \equiv 10 \sim 20 \text{ M}^{-1}, K_{IL} \equiv 2000 \sim 3500 \text{ M}^{-1}$, and $K_{IL2} \equiv 3 \sim 15 \text{ M}^{-1}$ produced convergent least squares fits of the $k_{obs}$ vs. $[\text{Na}_2\text{SO}_4]$ data. These parameters were also fixed during the final fitting process. Therefore, the weighted nonlinear least squares fitting was performed with just four adjustable parameters: $k_{UL}^1, k_{UL}^2, k_{IN}^1,$ and $k_{IN}^2$. Table 4-1 lists the best-fit values of these four parameters. Also
listed are the values of the fixed parameters discussed above and the values for $k_{U}^{L}$, $k_{N}^{IL}$, $k_{N}^{L}$, and $k_{N}^{IL}$ derived from Equation Set 5.

4.2.5 Populations and fluxes derived from the best-fit kinetic mechanism

The fractional population of each of the nine species depicted in Scheme 4-1 can be calculated as a function of the sulfate concentration using the best-fit parameter values listed in Table 4-1, as shown in Figure 4-5. Our kinetic model predicts that only six states, U, I, IL, N, NL, and NL2 are significantly populated over the sulfate concentration range from 0 to 400 mM. The populations of both liganded states of U and the doubly liganded form of I are predicted to never exceed 1%.

Another physical property of the mechanism, the flux through each direct pathway from U to NL2, can also be computed using the parameter estimates listed in Table 4-1. The flux of an elementary step is the product of the reactant population (or concentration) times the forward rate constant. At equilibrium the forward flux and the reverse fluxes are identical. The following two rules are used to calculate the total flux for a mechanism consisting of more than one elementary reaction:

$$F_{tot} = \sum_{i}^{n} F_{i},$$  
if the reactions are parallel \hspace{1cm} (Equation 6)

$$F_{tot} = \left(\sum_{i}^{n} 1/F_{i}\right)^{-1},$$  
if the reactions are sequential \hspace{1cm} (Equation 7)

where the $F_{tot}$ is the total flux of a multiple reaction pathway and $F_{i}$ is the individual flux for each elementary reaction. Equation 7 was used to calculate the fluxes of the six
Figure 4-5: The population of each state calculated from the equilibrium binding constants estimated from the fitting. Red, green, and blue lines are the U, I and N states, respectively. Green dashed and blue dashed lines are the singly liganded I and N states, respectively. Blue dot-dashed line is the doubly-liganded state of N. Each of these six states is significantly populated at some range of sulfate concentration examined. Since the lines representing the populations of the $UL$, $UL_2$ and $IL_2$ states are nearly zero throughout the entire sulfate concentration range tested, these three states are not labeled on the plot.

direct pathways depicted in Figure 4-6A. Each pathway includes two binding and two folding reactions. Because the binding reactions are much faster than the conformational transitions in all cases, the fluxes for binding steps are much higher and their reciprocals contribute negligibly to the total flux of each pathway. Figure 4-6B shows the total flux through each pathway as a function of sulfate concentration, calculated from the fluxes of the elementary folding steps. The plot also shows that the predominant pathway depends
Figure 4-6: Fluxes for the six possible pathways of P protein coupled folding and binding. (A) The top figure shows Scheme 1 with the rate constants, equilibrium constants and the ligands for mass balance in the reaction removed for clarification. The three vertical pathways colored in red, green and blue represent the unliganded, singly and doubly liganded forms of P protein, respectively. The bottom figure shows the six direct pathways connecting the $U$ and $NL_2$ states. The red, green and blue dots in the six pathways correspond to the state at that same position in the scheme above it. (B) Calculated fluxes through the six pathways using the rules described in the results. The three predominant pathways throughout most of the sulfate concentration range are pathways (1), (2) and (3) depicted in (A). At sulfate concentrations above 10 mM, the predominant pathways shift to (4) and (5). Pathway (6) does not become a significant pathway until the sulfate concentration exceeds 0.1 M. (C) The fractional flux calculation of each pathway as a function of ligand concentration. The red line is pathway (1), the blue line is pathway (2), the green line is pathway (3).
Table 4-1: Best fit kinetic and equilibrium parameters for the P protein folding/binding mechanism

<table>
<thead>
<tr>
<th>Kinetic parameters (units: s⁻¹)</th>
<th>Equilibrium parameters (M⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( a k_{UI} ) 0.4±0.05</td>
<td>( b k_{IN} ) 2.5±1.0</td>
</tr>
<tr>
<td>( a k_{IU} ) 2.2±0.3</td>
<td>( c k_{NI} ) 2±0.2</td>
</tr>
<tr>
<td>( a k_{IN} ) 1.1±0.1</td>
<td>( b k_{IL} ) 79±10</td>
</tr>
<tr>
<td>( a k_{NI} ) 6.1±0.3</td>
<td>( c k_{IL} ) 0.3±0.1</td>
</tr>
<tr>
<td>( b k_{UI} ) 32±5</td>
<td>( b k_{IN} ) 14±5</td>
</tr>
<tr>
<td>( c k_{IU} ) 0.4±0.1</td>
<td>( c k_{IL} ) 0.02±0.1</td>
</tr>
</tbody>
</table>

\( a \): Values are determined from folding kinetic studies in Chapter 2.
\( b \): Fitting results from the scheme 1 using equation 2.
\( c \): Calculated using the relationships in equation 5.
\( d \): Manual fitting to reduce the overall \( R^2 \).
\( e \): Using the previously reported values.

on the sulfate concentration, thus demonstrating that the coupled folding and binding mechanism of P protein is a function of ligand concentration. As discussed below, this conclusion is likely to apply to most reactions in which protein conformational change is coupled to ligand binding.

**4.2.6 NMR mapping of the high affinity ligand binding sites with paramagnetic \([\text{Cr(CN)}_6]^{3-}\)**

Our previous studies proposed the locations of the two high affinity anion binding sites responsible for the folding coupled binding transition of P protein (Henkels, Kurz et al. 2001). The electron density of one sulfate was clearly observed by x-ray crystallography near the N-terminus with sulfate coordinated by His3, Arg9 and Arg68 (Stams, Niranjanakumari et al. 1998). We later proposed that the second site is located...
near Arg88 and Lys89, although Stams, et al. had assigned the observed electron density at that site to a water molecule instead of a sulfate ion. In order to determine the anion binding sites more definitively, we used paramagnetic relaxation enhancement (PRE) NMR experiments. In these experiments, the PRE is caused by a magnetic dipolar interaction between a nucleus and the unpaired electron of a paramagnetic center. This interaction causes an increase in the relaxation rate of the nuclear magnetization (Iwahara and Clore 2006) that broadens the NMR resonance thus reducing its apparent intensity. Since PRE is a dipole-dipole interaction between nuclear and electron spins, its magnitude is proportional to \( <r^{-6}> \) (where \( r \) is the distance between the nucleus of interest and the paramagnetic center). As a result of the large magnetic moment of an unpaired electron, the effect of the PRE can be detected at a fairly long range (15-24 Å). Previous studies used the paramagnetic hexacyanochromate(III) ion (\([\text{Cr(CN)}_6]^{3-}\)) to probe the active site of RNase A (Inagaki, Watanabe et al. 1979) and the anion binding site of phosphoglycerate kinase (Fairbrother, Graham et al. 1990). We have used the same anion for the PRE studies of the P protein anion binding site.

The apparent affinity of P protein for \([\text{Cr(CN)}_6]^{3-}\) was determined by CD detected titration with this ligand yielding an apparent \( K_d \) of approximately 20 \( \mu \text{M} \) (data not shown), consistent with high affinity binding expected for a trivalent anion. Subsequently, \([\text{Cr(CN)}_6]^{3-}\) titration experiments monitored by \(^1\text{H}-^{15}\text{N}\) HSQC spectrum were performed to identify the P protein residues in close proximity to the anion binding sites. The initial NMR sample contained 200 \( \mu \text{M} \) of pre-folded P protein in 20mM sulfate, pH 6.5. \([\text{Cr(CN)}_6]^{3-}\) was added to the sample at concentrations ranging from 2.5 \( \mu \text{M} \) to
100 μM. Because $[\text{Cr(CN)}_6]^{3-}$ is more tightly bound than sulfate, 2.5 μM of $[\text{Cr(CN)}_6]^{3-}$ was sufficient to displace some of the bound sulfate. It is also reasonable to assume the on/off rates of the ligands will be in the fast exchange limit on the NMR chemical shift time scale. Under this condition, the apparent transverse relaxation rate ($R_2$) is the population weighted average of the relaxation rate of the nucleus in the sulfate- and $[\text{Cr(CN)}_6]^{3-}$-bound forms of the folded protein. Even if the population of the sites in $[\text{Cr(CN)}_6]^{3-}$ is small (<1%), this minor species can still strongly affect the apparent relaxation rate and cause line broadening of the $^1\text{H}-^{15}\text{N}$ HSQC resonances from the backbone amides of nearby residues.

Figure 4-7 shows overlapping HSQC spectra from the $[\text{Cr(CN)}_6]^{3-}$ titration experiments. With the addition of the $[\text{Cr(CN)}_6]^{3-}$ ion, the amide peak intensities of the residues that are in vicinity of the binding sites start to decrease or completely broaden out. These residues are labeled in Figure 4-7. The degree of the intensity decrease of a residue as a function of $[\text{Cr(CN)}_6]^{3-}$ concentration provides information about the distance between the amide and the paramagnetic center. The lower the $[\text{Cr(CN)}_6]^{3-}$ concentration needed to broaden the peak, the closer the residue is to the binding site. As seen in Figure 4-8A, mapping the affected residues to the structure of the P protein identified two anion binding sites as expected. In general, the PRE NMR results agree with the previously proposed anion binding sites, especially the one located in the N-
Figure 4-7: $^1$H-$^{15}$N HSQC spectrum of folded P protein in various concentrations of [Cr(CN)$_6$]$^{3-}$. The protein concentration was 200µM with 20mM sulfate in 20mM cacodylate buffer pH 6.5. The [Cr(CN)$_6$]$^{3-}$ concentration ranged from 2.5 µM to 50 µM. The color representations are as follows: black, 2.5µM; red, 5µM; orange, 10µM; yellow, 25µM; green, 50µM. The spectrum shows the residues that are affected most strongly by the paramagnetic source during the titration. Those residues are labeled in the figure. The spectra with [Cr(CN)$_6$]$^{3-}$ concentrations higher than 50µM are not shown in the plot. Higher concentrations of [Cr(CN)$_6$]$^{3-}$ broaden the resonances of many residues, which may result from nonspecific binding of [Cr(CN)$_6$]$^{3-}$ thereby masking the effects due binding at the two high affinity sites.
Figure 4-8: PRE results mapped on the P protein crystal structure. (A) The results of the NMR PRE experiments show two regions of P protein involved in the anion binding. The color of each residue indicates the $[\text{Cr(CN)}_6]^{3-}$ concentration at which the associated resonance intensity is affected: red is 2.5$\mu$M, orange is 5 to 10 $\mu$M, and yellow is 25 to 50 $\mu$M. Residues colored gray correspond to resonances already missing in the 0M $[\text{Cr(CN)}_6]^{3-}$ HSQC spectrum, prolines or unresolved peaks. The sidechain of the Lys, Arg, and His residues colored red are candidates for coordination of the anionic ligands. Residues colored green are unaffected by $[\text{Cr(CN)}_6]^{3-}$ between 2.5$\mu$M and 100$\mu$M concentration range. (B) Structure of P protein depicting backbone B-factors reported in the crystal structure (Stams, Niranjanakumari et al. 1998) The ribbon thickness is proportional to B-factor. The residue color scheme is the same as in (A) The loop3 region has overall higher B-factors than the rest of protein backbone. The sidechain of Lys52 is shown.
terminal region. The other site is also close to the previously proposed site, but Lys89 and Lys52 amide resonances were more strongly affected by the paramagnetic anion than the Arg88 amide. This suggests that the second anion binding site might be closer to the loop2 and loop3 region indicated in Figure 4-8A rather than loop1 and loop2 region as previously modeled.

4.3 Discussion

4.3.1 Sequence of events in the coupled folding and binding mechanism

The mechanism of folding coupled to binding has been widely discussed for a variety of intrinsically unstructured proteins. A fundamental question in many of these discussions relates to the order of the two events: which comes first, binding or folding? The mechanism in which binding precedes folding has been called “induced folding” implying that ligand binding promotes protein folding or conformational change to form the binding site. The alternative mechanism with folding prior to binding has been called “conformational selection” referring to the selection by the ligand of binding-competent protein conformations from a mixture of alternative forms. In the case of an IUP, the selected conformation is a high-energy state whose population can be quite small. However, because the ligand can bind to this conformation with high affinity, the equilibrium can be pulled to fully populate the bound protein conformation.

The distinction between these two limiting mechanisms can only be made with kinetic experiments performed over a range of ligand concentration that spans the apparent $K_d$, such as those described above. In our studies, all the possible mechanisms
were modeled in one complete scheme enumerating six possible pathways with the same two end points \((U \text{ and } NL_2)\). Based on the parameter values derived from fitting the kinetic data, the predominant pathway depends on the ligand concentration. In order to properly describe the mechanism, the concept of flux, described in the Results section, must be employed to quantitatively estimate the relative importance of each possible pathway. For the P protein system described here, the fluxes through each pathway are strongly depended on the ligand concentration.

Although coupled folding and binding is known to play a role in many protein/ligand interactions, few studies have provided insight about the sequence of the two events in the interactions. Some studies attempt to experimentally characterize the mechanism of coupled folding and binding in a particular system. (Sugase, Dyson et al. 2007; Narayanan, Ganesh et al. 2008; Onitsuka, Kamikubo et al. 2008; Sullivan and Holyoak 2008). Sugase, et al. found evidence that the intrinsically unstructured pKID domain of the transcription factor CREB forms an encounter complex with its target, the KIX domain of the CREB binding protein, and that this encounter complex contains numerous nonspecific or unfolded interactions (Sugase, Dyson et al. 2007). The authors suggested that this could be interpreted as evidence that weak nonspecific binding precedes actual folding. Onitsuka, et al. have recently reported (Onitsuka, Kamikubo et al. 2008) that intrinsically disordered mutants of staphylococcal nuclease (SNase) can be induced to fold upon ligand binding. The studies demonstrated that for an alanine addition mutant, 33A34, folding precedes binding, while for the other C-terminal truncation mutant, \(\Delta 140–149\), binding precedes folding. Sullivan et al. reported that the
lid domain of the PEPCK protein followed an induced fit model because the closed form of the lid domain hindered the binding site for ligand (Sullivan and Holyoak 2008). The authors modeled the two mechanisms separately and discussed them individually. Most of these studies have some drawbacks on the interpretations of the experimental data that we want to discuss below.

Comparing to the studies listed above, we define all the possible mechanisms in one complete scheme since the two end points, $U$ and $NL_2$, are the same in all the pathways. Based on the kinetic mechanism we developed in scheme 4-1, the order of the folding and the binding events can be clearly described at different ligand concentrations. In order to properly interpret our results, the concept of fluxes described in the results section through different pathways was used. As shown in Figures 4-6A and 4-6B, the fluxes through each pathway are strongly dependent on the ligand concentration. At ligand concentrations below 100 µM, the flux is predominantly through pathway (1) in which the two folding events precede the two binding events. From 0.1 mM to 1 mM sulfate, the greatest flux is through pathway (2) in which the folding of the $U$ to $I$ state precedes the binding of one ligand to the $I$ state. Subsequently, folding of $IL$ to $NL$ is followed by binding of the second ligand to $NL$ to form $NL_2$. When the ligand concentration is above 2 mM, the greatest flux is through pathway (3). In this pathway, the binding of ligand to $U$ precedes the two folding reactions from $UL$ to $IL$ to $NL$, followed by binding of the second ligand. At ligand concentrations above 200 mM, the greatest flux is through pathway (6) in which both binding steps precede the two folding steps. It should be emphasized that the above descriptions are based on the pathway with
the greatest flux. Figure 4-6C shows the fractional flux through each pathway and demonstrates that other pathways contribute significantly to the overall flux. For instance, at 1 mM sulfate, pathways (1), (3), (4), and (5) contribute a total around 65% of the flux, in addition to the predominant pathway (2). Thus, a single pathway description of the mechanism is inappropriate. This is likely to be the case in all other binding coupled systems as well, at least at some ligand concentrations.

Two conclusions from our results for the coupled folding and binding mechanism of P protein need to be emphasized. First, in order to appropriately study such a mechanism, a detailed investigation of the ligand concentration dependence of the folding/binding process is required. Narayanan et al. recently reported a coupled folding and binding mechanism using IA3 peptide inhibitor with its target YPrA protease. IA3 is intrinsically unstructured, but folds to an α helix as it binds to YPrA, leading to sub-nanomolar inhibition of the protease. The results of the studies favor a model in which IA3 first forms a weak complex with YPrA, prior to forming the α helix structure. The peptide and the protease concentrations used in the experiments were both ~50 µM, so the binding equilibrium strongly favored the bound/inhibited state. The kinetic mechanism under this condition may not be the same if the protease or peptide concentration were close to the binding constant. Moreover, the in vivo concentration of the inhibitor and protease is no doubt significantly lower than experimental conditions. Therefore, experimental conditions that cover a wide range of ligand concentration is crucial to determine the coupled folding and binding mechanism.

Secondly, the concept of flux is important to properly determine the actual
mechanism. To our knowledge, no previous studies of folding/binding mechanisms have used flux to deduce which pathway predominates. In some cases the predominant pathway is deduced by comparing rate constants of the first step in each pathway, based on the conclusion that whichever reaction is faster will happen first. For example, by comparing pathway (1) and (3) in Figure 4-6A, the first step in pathway (1) is folding and in pathway (3) is binding. If the statement above were true for P protein then rapid binding would be earlier than folding and pathway (3) should always be predominant. However, Figure 4-6C shows that either pathway 1 or 2 predominates at sulfate concentrations of 1 mM or lower, meaning that folding to I or N precedes binding in the mechanism. Instead of viewing the coupled folding and binding mechanism as either folding first or binding first reaction, the mechanism should be more appropriately interpreted as combinations of the two or more pathways. The sequence of the folding and binding events is best described as set of pathways whose fluxes determine their relative importance.

4.3.2 Two small anion binding sites in the P protein

The results of the PRE NMR experiments showed good agreement with the previously modeled binding sites in the P protein. Leu 4, Lys 6 and Arg 9 in the N-terminal region are strongly affected by the paramagnetic anions, and they are either the residue or close to the residues that coordinated the sulfate (His 4, Arg 9, and Arg 68) anion in the crystal structure. The cross peak of His 3 is missing in the $^1$H-$^{15}$N HSQC spectrum and the cross peak of Arg 68 is overlapped with other two strong peaks, Ser 116 and Ser 117. For these reasons, the effects of PRE on these two residues cannot be
confidently assigned. They are colored in gray in Figure 4-8A for clarification. The second binding region probed by paramagnetic anion is also close to the previously proposed region. Interestingly, Lys 89 (colored in red) is affected most strongly by the anion and is followed by Lys 52 (colored in orange). Arg 88 (colored in yellow) is actually has less effect compared to both Lys 89 and Lys 52. These results point out some discrepancies between the crystal structure shown in Figure 4-8A and the PRE NMR data. In the crystal structure, Lys 53 is closer to Lys 89 compared to Lys 52. However, the peak intensity of Lys 52 is broadened out at lower concentrations of paramagnetic anion compared to Lys 53 indicating that amide dipole of Lys 52 should spatially closer to Lys 89. As shown in the Figure 4-8B, the loop region between residue 51 to 60 has the highest B-factor values in the structure, suggesting that the loop has significant flexibility. If this region is in fact flexible in solution, a different angle combination between these residues could reorient the Lys 52 closer to Lys 89. The loop orientation in the crystal structure may be due to specific crystal packing and might not well represent the structural orientation in the solution. Based on the PRE NMR experiments, we conclude that the sidechains of Lys 89 and Lys 52 play an important role in forming the second anion binding site of P protein.

We emphasize these two particular residues because a recent cross-linking study of P protein with P RNA and pre-tRNA substrate showed that these two residues interact with pre-tRNA (Niranjanakumari, Day-Storms et al. 2007). Cross linking studies of P protein have shown that it interacts with the 5′ leading sequence of pre-tRNA through its central cleft RNA binding region. These new data showed that P protein
interacts through Lys 52 and Lys 89 with pre-tRNA bases 65-67 at the 3′ end of the acceptor stem. These two residues may not directly participate in the catalytic function, but might play roles in substrate recognition and holoenzyme assembly.

4.3.3 Effects of different ligands on the equilibrium states and fluxes in the coupled folding and binding mechanism

Besides the sulfate anion used in this study, many other small anions also can induce P protein to fold (Henkels, Kurz et al. 2001). These anions can be broadly divided into three groups (tight, intermediate, and weak binding) depending on the apparent binding affinity, which is proportional to the net charge of the anion. The greater the charge on the anion, the tighter it binds to the P protein. Sulfate has intermediate binding affinity with an apparent $K_d \sim 0.5$ mM. It is informative to see how the tight binding anions such as pyrophosphate or mononucleotide and weak binding anions such as chloride or acetate might change the population of each state and the fluxes through different pathways. Using Scheme 1 for the overall coupled folding and binding mechanism of P protein for different small anions, we simulated the population of each state and fluxes for different types of anions. For the purpose of this calculation, we assume that the values of all folding and unfolding rate constants are the same as those estimated from the sulfate kinetic data. Only the macroscopic association constants are adjusted in the simulation. For the tight binding ligands (assuming an apparent $K_d$ of 5 µM), the macroscopic association constants might be $K_{UL} = 4K_{UL_2} = 10M^{-1}$, $K_{IL} = 1.5 \times 10^4$, $K_{IL_2} = 450M^{-1}$, and $K_{NL} = 4K_{NL_2} = 1 \times 10^6 M^{-1}$. For the weak binding ligands (assuming an apparent $K_d$ of 50 mM), the values for the same parameters might
be $K_{UL} = 4K_{UL_2} = 10 M^{-1}$, $K_{IL} = 160 M^{-1}$, $K_{IL_2} = 15 M^{-1}$, and $K_{NL} = 4K_{NL_2} = 1000 M^{-1}$. These values were chosen to fulfill the condition that total population of $N$, $NL$, and $NL_2$ is 50% of the total protein population at a ligand concentration equal to the apparent $K_d$.

Simulation results based on these assumptions are shown in Figure 4-9. Figures 4-9A and 4-9C show the population of each state with a tight binding ligand and weak binding ligand, respectively. As can be seen in the plots, only six states ($U$, $I$, $IL$, $N$, $NL$, and $NL_2$) are significant populated throughout the ligand concentration, which is similar to the sulfate results. In contrast, the simulated fluxes through each pathway as function of ligand concentration have very different patterns for the different ligand classes shown in Figures 4-9B and 4-9D. Pathways (1) and (2) have the only significant flux for the tight binding ligands. In these two pathways, folding of $U$ to $I$ state precedes the first ligand binding step. For the weak binding ligands, each of the six pathways is dominant over some range of ligand concentrations. Unlike tight and intermediate binding ligands, pathways (4), (5), and (6) actually play major roles in the weak ligand mechanism. The assumptions for the parameter values in the simulations are no doubt oversimplified, and actual values need to be verified through more kinetic experiments. Nevertheless, the simulation results still provide insight into the effect of ligand affinity on the kinetic mechanism of $P$ protein folding/binding. This information is particularly useful when we discuss the holoenzyme assembly process in vivo below.
Figure 4-9: The simulated population of each state and the fluxes through different pathways as a function of pyrophosphate and chloride. (A) and (C) The population of each state as a function of different ligand are simulated using the chosen parameters described in the discussion. Red, green, and blue lines are the U, I and N states, respectively. Red dashed, green dashed and blue dashed lines are the singly liganded U, I and N states, respectively. Red dot-dashed, green dot-dashed and blue dot-dashed line is the doubly-liganded state of U, I and N, respectively. (B) and (D) The total flux through each direct pathway is plotted as a function of ligand concentration. The mechanism diagram in the flux plot (B) indicates which pathway the flux refers to. The color of each line in (B) and (D) indicates different direct pathway, and the color scheme is the same as Figure 4-6B and C.
4.3.4 Model for RNase P holoenzyme assembly process *in vivo*

Other than the intrinsically unstructured proteins or fragments identified by bioinformatic analyses, most IUPs have been identified and characterized experimentally *in vitro*. Questions have been raised as to whether these disordered regions are unstructured *in vivo*, as they appear to be in the *in vitro* experiments. Aside from the issue of protein stability in the cell, it has been suggested that the crowded and viscous environment within the cell might induce IUPs to fold. This question has been addressed by NMR experiments. The intrinsically unstructured bacterial protein FlgM gains structure inside living *E. coli* (Dedmon, Patel et al. 2002), while the eukaryotic α-synuclein remain disordered within the cytoplasm (Li, Charlton et al. 2008). It is also valid to ask the same question for the P protein system. P protein can be induced to fold by different small anions so it is reasonable to assume that newly synthesized P protein *in vivo* will bind to small cellular anions first and fold to its native state before it encounters the cognate P RNA molecule. This conclusion raises further questions. What is the process by which a small ligand bound P protein is displaced to form the P RNA-P

Scheme 4-3

and viscous environment within the cell might induce IUPs to fold. This question has been addressed by NMR experiments. The intrinsically unstructured bacterial protein FlgM gains structure inside living *E. coli* (Dedmon, Patel et al. 2002), while the eukaryotic α-synuclein remain disordered within the cytoplasm (Li, Charlton et al. 2008). It is also valid to ask the same question for the P protein system. P protein can be induced to fold by different small anions so it is reasonable to assume that newly synthesized P protein *in vivo* will bind to small cellular anions first and fold to its native state before it encounters the cognate P RNA molecule. This conclusion raises further questions. What is the process by which a small ligand bound P protein is displaced to form the P RNA-P
protein holoenzyme complex? What is the holoenzyme assembly process?

Scheme 4-3 depicts a proposed mechanism for the RNase P assembly process. Since bacterial RNase P holoenzyme is composed of one RNA and one protein component, another vertical pathway is added to Scheme 1. The question is now: what is the predominant pathway going from the $NL_2$ state to $PRNA-N$ complex. We again use the fluxes through possible pathways to address the holoenzyme assembly process. Small anionic ligands play an important role in the mechanism because the dominant pathways are dependent on both concentration and type of ligand. For example, Figure 4-9B shows that pathway (2) is the major pathway for P protein in 5 µM pyrophosphate. When the small anion bound P protein encounters P RNA under these conditions, the reaction follows pathway (2) going from $NL_2 \rightarrow NL \rightarrow IL \rightarrow I$. $I$ can form a $PRNA-I$ complex that is subsequently followed by the folding of the protein to form the $PRNA-N$ holoenzyme complex; or $I$ can unfold to $U$, which subsequently forms the $PRNA-U$ encounter complex, followed by the folding of the protein to form the $PRNA-N$ complex.

The same principle can be used to explain other assembly pathways under different ligand conditions. Thus, the intermediate state in P protein could be important in the assembly process. From the results of the NMR studies in Chapter 3, the N and C-termini of the intermediate state are mostly unfolded, but the RNR motif and metal binding loop regions have a native-like structure. These two regions form crucial interactions between P protein and P RNA.

Scheme 4-3 is an over simplified model. For example, P RNA can also have some conformational rearrangement upon P protein binding which would further
complicate the mechanism. Moreover, P RNA presumably can bind to singly or doubly liganded I state instead of only I state. This also adds complexity to the mechanism. However, Scheme 4-3 defines a possible mechanistic framework for RNP complex assembly in vivo.

4.4 Materials and Methods
4.4.1 Protein expression and purification

The F107W variant of P protein was expressed and purified as previously described (Niranjanakumari, Kurz et al. 1998) with the following modifications. The pooled fractions of P protein that eluted from the second CM-Sepharose column were concentrated using a Centriprep (Amicon) until the volume of the pooled fractions was reduced to 1~2 ml. Then the buffer in the sample was then changed to 6 M Guanidine-HCl, 10 mM Tris, pH 7.5 using the same method. The concentrated sample was loaded onto an S-100 column (GE Helthcare) equilibrated with 6 M Guanidine-HCl, 10 mM Tris, pH 7.5 to remove any EDTA bound to the P protein. Fractions containing P protein were pooled and stored at -80 °C. Before equilibrium titration or stopped flow kinetic experiments, the protein was first extensively dialyzed against water, then dialyzed against 20 mM sodium cacodylate at the appropriate pH needed for the subsequent experiments. In all cases, the protein concentration was determined using the Edelhoch method with an extinction coefficient of 11,200 M⁻¹ cm⁻¹ at 276 nm.

For preparation of the ¹⁵N labeled protein used in the NMR experiments, a single colony of transformants was used to inoculate 30 ml of LB media containing 50
µg/ml kanamycin and 17 µg/ml chloramphenicol and grown with shaking at 37 °C. When the OD$_{600}$ reached 0.8-0.9, the culture was diluted 1:200 into the 1 liter of M9 minimum media culture containing the same antibiotics and enriched with 1 g/l $^{15}$N ammonium chloride. When the OD$_{600}$ was 0.4-0.6, protein overexpression was induced by the addition of IPTG to a final concentration of 0.4 mM. The culture was grown for another 10-12 hours then harvested by centrifugation. The $^{15}$N enriched P protein was purified as described above.

### 4.4.2 Circular dichroism and fluorescence equilibrium experiments

An Aviv model 202 CD spectrometer was employed to collect CD wavelength spectra. Samples were scanned from 260nm to 190 nm in 1 nm increments with a bandwidth of 1.5 mm and averaged for 3s at each wavelength. Protein concentration was 10 µM in 20 mM sodium cacodylate, pH 7, in the presence or absence of 20 mM sulfate or 1 M TMAO using a cuvette with a 1 mm path length at 25 °C.

Fluorescence spectra were collected on a Fluorolog-3 spectrometer from HORIBA Jobin Yvon using a 1 cm pathlength cuvette. The sample was 5 µM P protein in 20 mM cacodylate, pH 7. The temperature was controlled at 25°C by a circulating water bath. The excitation wavelength was 285 nm with a 4 nm bandpass, and emission spectra were collected from 400 to 300 nm at 1nm increments.

Sulfate titration fluorescence data were collected using a Amico-Bowman Series2 fluorimeter with an excitation wavelength of 285 nm with a 4nm bandpass. Fluorescence emission signals were collected at 345 nm with a 4 nm bandpass for 30 s at 0.5 s intervals for each sulfate concentration, then averaged. The cuvette had a 1 cm
pathlength and the temperature was controlled at 25 °C. Protein concentration was 10 µM in 20 mM sodium cacodylate, pH 7. Three different stock solutions containing 5 µM protein in the presence of 2 mM, 40 mM or 800 mM were prepared to cover full range of 10 uM-200 mM final sulfate concentration for these experiments.

4.4.3 Kinetic stopped flow fluorescence experiments

All kinetic traces were collected on an Applied Photophysics SX.20MV stopped flow instrument. The excitation wavelength was 285 nm with 1 mm slit width on the monochrometer and the path length of the optical detection cell was 2 mm. The fluorescence at wavelengths longer than 320 nm was collected using a high pass filter. Temperature was controlled at 25 °C using a circulating water bath. Protein samples were in 20 mM sodium cacodylate buffer pH 7.0 and protein concentrations were 5 µM. In sulfate induced refolding experiments, the concentration of sulfate ranged from 0.09 mM to 600 mM (0.09, 0.1, 0.2, 0.4, 0.6, 0.8, 1, 2, 3, 4, 5, 6, 8, 10, 20, 30, 40, 50, 60, 80, 100, 200, 300, 400, 500 and 600 mM). Sulfate induced refolding reactions were initiated by mixing equal volumes of P protein and sulfate buffer (total mixed volume is ~120 µl). Asymmetrical mixing experiments (1:10 volume ratio) were performed to acquire low sulfate concentration data (0.025~0.1 mM). In asymmetrical mixing experiments, protein concentration was 12.5 µM in the presence of 0.1 mM to 1 mM sulfate from a 250 µl syringe followed by mixing with buffer from a 2.5 ml syringe (mixed volume is around 250 µl). This setup yielded a final protein concentration of 1.25 µM and sulfate concentrations between 0.01 mM and 0.1 mM. Eight to twelve kinetic traces were averaged for each sulfate concentration.
4.4.4 NMR spectroscopy experiments

Two-dimensional gradient-enhanced sensitivity-enhanced $^1$H-$^{15}$N HSQC were collected on a Varian INOVA 600 spectrometer with a triple resonance probe. The P protein was 200 µM with 20 mM sulfate in 20 mM cacodylate buffer, pH 6.5, 10% D$_2$O, and 0.05% azide. Hexacyanochromate anion [Cr(CN)$_6$]$^{3-}$ was titrated into the protein sample from 2.5 µM to 100 µM (with the concentration steps of 5, 10, 20, and 50 µM). The HSQC spectra were collected at each [Cr(CN)$_6$]$^{3-}$ concentration with a spectral width in the $^1$H dimension of 7020 Hz and 1024 complex points and a spectral width in the $^{15}$N dimension of 1994 Hz and 128 complex points. All spectra were processed using nmrPipe. The processed spectra were analyzed with nmrDraw to verify the crosspeak phases, and the intensities of the crosspeaks were quantified by nmrView.

4.4.5 Data analysis

The stopped-flow kinetic traces were fitted by double or triple exponential function with form:

$$F(t) = F(\infty) + \sum_{i} A_i e^{\lambda_i t}$$

where $F(t)$ is the fluorescence at any given time $t$, $F(\infty)$ is the fluorescence signal at infinite time, $\lambda_i$ is the observed rate constant of phase $i$ and $A_i$ is its amplitude. Kinetic traces were analyzed by Pro-Data Viewer from Applied Biophysics Inc. The sulfate dependence of the observed rate constants was fitted by the nonlinear square method using Mathematica™ (Wolfram Research) software.
Chapter 5 : Conclusions and Future Directions

Fundamental understanding of the thermodynamic and kinetic parameters that govern the cooperative conformational changes in proteins is important because many biological phenomena are involved in this process including protein folding, allostERIC regulation, and enzymatic catalysis. Most of these biological processes exhibit complex kinetic mechanisms that involve multiple conformational states, substrates and pathways. Therefore, it is equally important to correctly interpret the overall mechanisms of different systems based on the estimated kinetic and thermodynamic parameters.

This work focuses on the folding kinetics of an intrinsically unstructured protein, the RNase P protein. We investigated the folding/unfolding equilibrium and kinetics of P protein, the electrostatic interactions in the proteins, the structural properties of a folding intermediate state present in the folding kinetics, and the coupled folding and ligand binding mechanism of P protein.

5.1 Conclusions

5.1.1 Osmolyte-induced folding/unfolding is a general method to study the thermodynamics and kinetics of IUPs

The kinetic and thermodynamic properties of an intrinsically unstructured protein, RNase P protein, have been characterized using stopped-flow and equilibrium titration experiments. In order to study these properties of an IUP using the conventional experimental techniques, one major modification needs to be made. Instead of chemical denaturants used for natively structured proteins, protecting osmolytes (renaturant) such
as TMAO or glycerol are necessary for the studies of intrinsically unstructured proteins.

The theory and data analysis process for denaturation and renaturation experiments are exactly the same, and they are complementary to each other in studies of proteins with different stabilities. Despite these similarities, few studies have used the osmolyte-induced folding/unfolding method to investigate folding kinetics and thermodynamics of IUPs. Therefore, the experiments we developed in Chapter 2 should be a general protocol to study the kinetic and thermodynamic properties of any partially folded proteins and intrinsically unstructured proteins.

5.1.2 The electrostatic interactions in the P protein are important to the equilibrium between different conformational states

The electrostatic interactions between amino acid sidechains often have effects on the kinetic and thermodynamic properties of proteins (Sato and Raleigh 2002). P protein is a highly basic (pI=10.3) and highly charged protein (11 Arg, 19 Lys, and 3 His out of 119 residues). This high charge could be part of the reason that P protein is intrinsically unstructured because the folded structure needs to overcome a strong repulsive force. Therefore, it is not a surprise that the equilibria between the U and I states and the I and N states shows a strong pH dependence. Figures 3-2 and 3-4 show the pH perturbation of P protein equilibrium monitored by NMR HSQC spectrum between pH 5 and 7 in the absence and the presence of ligand, respectively. Single mutation of each of the three histidines to lysine in the P protein shows that His 22 may play a major role perturbing the equilibria. Based on the crystal structure of P protein (Figure 1-1), the three histidines are quite solvent accessible (His 3 is at the N-terminus,
His 22 is at the end of helix A, and His 105 is at the end of helix C), and do not participate in hydrogen bonded network. This can explain why the pKa of these histidines determined in Chapter 3 are not significantly perturbed compared to the histidine alone. Thus, the perturbation of the equilibrium of P protein by pH should be mainly due to the charge-charge interaction between histidine sidechains and nearby residues. The sidechains of Lys 18 and Lys 21 are spatially close to His 22, so deprotonation of His 22 at pH 7 may resolve the repulsive force between the sidechain and favor the folded structure. This interpretation is consistent with the conclusion in Chapter 4 that the equilibrium shifts toward the I state in the absence of ligand and toward the N state in the presence of ligand as the pH is raised.

His 3 participates in the coordination of one anion binding site in P protein, and the protonation state of this histidine will affect the electrostatic interactions between the histidine sidechain and the ligand. Therefore, changing the pH also changes the ligand binding affinity of P protein. We cannot estimate the extent of the change in binding affinity based on the results in this study, but the change of the binding constant of each state, especially the I and N states, for the ligand will significantly perturb the population of each state in the mechanism.

Figure 3-3 shows the intensities of corresponding crosspeaks for the intermediate state on the HSQC spectrum decrease as the sulfate concentration increases indicating that the population of the intermediate state depends on the ligand concentration. As shown in the population plot in Figure 4-5, the singly liganded form of intermediate state (IL) is about 10 % of total protein concentration at 1 mM sulfate, and
decreases to approximate 0 % above 20 mM sulfate. These two results are consistent with each other, and show that the main species contributing to the HSQC signal of the intermediate state in the presence of ligand is the singly liganded form of the intermediate state. Moreover, this provides further evidence that the intermediate state discovered in NMR experiments is the same as that detected in the kinetic and equilibrium experiments.

5.1.3 **Flux concept is general in all coupled conformational changes and ligand binding reactions**

The concept of flux for chemical reactions introduced in Chapter 4 is a novel way to describe the predominant pathway in a mechanism if the overall reaction can take place through different routes. As emphasized throughout this study, the importance of the fundamental mechanism of coupled folding and binding reaction has become more prominent because a variety of biological reactions involve this process. Many theories and experimental methods have been developed to address this question. The theoretical fly-casting model is used by many experimental studies to describe the coupled folding and binding reaction in a particular system. The model suggests that binding should precede folding because the larger capture radius of an unstructured protein can make the rate constant for target ligand binding higher than that of a folded protein. It is true that the binding rate constant might increase for a binding site in an unstructured protein. However, the binding event, which has faster rate constant does not necessary occur first, especially at ligand concentrations below the apparent Kₐ. In
order to properly describe the overall coupled folding and binding mechanism, the calculation of flux includes the rate constants of each individual step and the population of each species.

Since the flux calculation is performed under equilibrium conditions, the ligand concentration is not restricted to the pseudo first order limit ([ligand]>>[protein]). This can be realized from Figure 4-6B. As shown in the Figure, when the ligand concentration is very low (<10^{-7} M), the total flux through each pathway is approximately zero. This is because at this ligand concentration P protein only favors the U and I states, and the populations of other states are negligibly small. Therefore, pathways involving any liganded forms of U have little or no flux. Using the same argument, under high ligand concentrations (> 0.1 M), the total flux through each pathway is also very small. Ligand concentration is a crucial determinant of a coupled folding and binding mechanism because the predominant pathway can switch under different ligand concentrations (Figure 4-6C).

The flux calculation is an appropriate method to interpret the overall mechanism of coupled folding and binding, and can apply to any mechanism that involved conformational change upon ligand binding. If the rate constants in the mechanism and the populations of each state under different conditions can be estimated, the predominant pathway can then be readily determined. Moreover, if
the correct *in vivo* protein and ligand concentration can be obtained, the pathway preferred in the cell can be determined.

In conclusion, this study provides a detailed analysis of the folding/unfolding process of an intrinsically unstructured protein and a general structural characterization of an intermediate state present in the process. The coupled folding and binding mechanism characterized in this study should be applicable to many other comparable systems in the biology.

### 5.2 Future experiments

#### 5.2.1 The role of P protein stability in the holoenzyme assembly process and catalysis

RNase P proteins from other bacterial organisms do not show high sequence similarity. However, the solution structure of *S. aureus* P protein (Spitzfaden, Nicholson et al. 2000) and crystal structure of *T. maritima* P protein have been solved (Kazantsev, Krivenko et al. 2003), and the three-dimensional protein structures from these two species exhibit overall structural similarity compared to P protein from *B. subtilis*. In addition, although without a determined three-dimensional structure, P protein from *E. coli*, C5 protein, was shown to be a mixture of unfolded and folded protein in the absence of ligand or P RNA, and folds to the native state upon addition of P RNA (Guo, Campbell et al. 2006; Sun and Harris 2007). The thermodynamic properties of P proteins from *T. maritima* and *S. aureus* are not fully understood. From a thermodynamic perspective, it would be interesting to investigate whether P proteins from different
bacteria also show IUP-like properties. P protein from thermophilic *T. maritima* would be particular interesting because its thermal stability should be significantly different. It would also be intriguing to characterize how the primary anion acid sequence perturbs the stability of the protein without affecting the overall structure of protein (Cuneo, Tian et al. 2008). From a kinetic perspective, if some of these proteins indeed have intrinsically unstructured properties, stopped-flow experiments like those described in Chapter 2 and 4 can be performed to study the folding/binding mechanism. The results would address whether intrinsically unstructured properties are be conserved in homologs among difference bacterial species.

The stability of an IUP could be important to its function and complex formation process. p27kip1 contributes to cell-cycle regulation by inhibiting cyclin-dependent kinase (Cdk) activity, and the unbound p27kip1 Cdk-inhibition domain is intrinsically unstructured. Recent studies show that stabilization of the helix with alanine hindered kinetically the formation of the inhibited complex, suggesting that p27kip1 derives a kinetic advantage from intrinsic structural disorder (Bienkiewicz, Adkins et al. 2002). For the P protein system, it is interesting to ask whether the activity of the *B. subtilis* RNase holoenzyme is affected *in vitro* and *in vivo* if the stability of P protein increases. It has been shown *in vitro* that the protein subunits from *B. subtilis* and *E. coli* are interchangeable in the holoenzyme complex with similar catalytic activity (Guerrier-Takada, Gardiner et al. 1983). Heterologous reconstitution of the P protein from *T. maritima* with P RNA from *E. coli* yielded a moderate decrease in holoenzyme catalytic activity (Paul, Lazarev et al. 2001). Although this result could be due to small defects in
the interface between P protein and a non-cognate P RNA, it could also be caused by inefficient holoenzyme assembly due to increased stability of the protein component.

5.2.2 Structural studies of the folding intermediate in P protein and the interactions with P RNA

Based on the experimental data presented here, the central β-sheet and helix B already form the secondary structure in both the intermediate and folded states of P protein. We cannot completely rule out that the eight remaining unassigned peaks do not belong to the residues in these two broad regions. Weak intensities of these unassigned peaks make the sequential assignment of them more challenging. In order to assign these remaining peaks, backbone assignment of P protein under weak to moderate denaturing conditions could be performed. A preliminary HSQC spectrum of P protein in 2 M urea showed that the spectrum has narrow chemical shift dispersion, but most of the crosspeaks are well resolved. This makes the assignment more feasible. Several studies have demonstrated successful backbone assignment and backbone dynamic studies in various denatured or unfolded protein systems (Wu, Kim et al. 2008) (Religa, Markson et al. 2005; Chugha, Sage et al. 2006). Many of the unassigned extra peaks are identifiable in the unfolded spectrum of P protein. The assignment may be possible in 2M urea because of the expected stronger peak intensity under these conditions. By collecting a series of $^1$H-$^{15}$N HSQC spectra under low to moderate urea concentrations (0.25-2 M), the movement of each peak under different urea concentrations can be followed. The identity of the unassigned extra peaks can be determined by tracing back to the assigned peak in the 2 M spectrum.
Structures of each component of the bacterial RNase P protein have been obtained through x-ray crystallography and NMR methods, but the structure of the holoenzyme complex is still undetermined. Many models for the holoenzyme complex have been proposed based on biochemical footprinting and hydroxyl radical cleavage data. Different NMR methods have the potential to investigate the crucial interactions between P RNA and P protein. Recent studies of the P RNA P4 helix, which is believed to bind catalytically important metals, using residual dipolar couplings and $^{13}\text{C}/^{15}\text{N}$ relaxation data showed that the plasticity of P4 helix is important for Mg$^{2+}$ binding (Getz, Andrews et al. 2006). An NMR chemical shift perturbation method was employed to study the interactions between the archaeal RNase P protein, Mth11, and the RNase P RNA, and the results showed some critical residues in the protein that may participate in the binding interface (Boomershine, McElroy et al. 2003). These types of experiments can also be applied to the bacterial P protein system for directly probing the interactions between the P protein and the P RNA.

### 5.2.3 Coupled folding and binding mechanism for P protein with different ligands

Flux simulations for weak and tight binding ligands for P protein have suggested that different kinds of ligands might have effects on the total flux and shift the predominant pathway in the coupled folding and binding mechanism. However, many assumptions are oversimplified in the simulation. For example, the folding/unfolding rate constants were the same using the fitting results from sulfate induced folding kinetics, but these rate constants could be different when P protein binds other kinds of ligands. The
same experiments and analyses on sulfate induced folding kinetics can be performed on other types of ligand, such as chloride, pyrophosphate and the cognate P RNA, and in principle the flux calculation can characterize how different ligands shift the predominant pathway in the folding/binding processes.

5.2.4 Kinetics of exchange between ligand-bound forms of intermediate state and folded state by NMR relaxation dispersion experiments

Previous NMR backbone dynamic studies of P protein suggested $\mu$s to ms timescale motions in the ensemble (Henkels, Chang et al. 2007). The residues that exhibit these slow internal motions are found in regions that have been previously identified as part of the P protein-P RNA interface, namely the N-terminal region and the RNR motif in the P protein. These results suggest that structural flexibility within the P protein may be important for proper RNase P holoenzyme assembly. The N-terminal region and part of the RNR motif are unfolded in the proposed structure of the intermediate state and are in the slow exchange with the folded structure. Taken together, these two observations suggest that the intermediate state participates in the holoenzyme assembly process and/or substrate catalysis. It would be intriguing to see if there is any correlation between the kinetics of the intermediate-folded state interconversion and the slow motions detected by the NMR dynamic studies. We intend to perform NMR relaxation dispersion experiments (Mulder, Mittermaier et al. 2001; Palmer, Kroenke et al. 2001; Wang, Grey et al. 2001) to estimate the chemical exchange rate between the intermediate and folded state of P protein and also the populations of the various protein species under different ligand concentrations. These results can be compared to the kinetic and equilibrium
parameters determined in Chapter 4. It would be interesting to investigate the correlation between the conformational heterogeneity revealed by NMR and the multiphase kinetics detected by transient kinetic techniques.
Reference


Biography

Yu-Chu Chang was born on August 27, 1979 in Taipei, Taiwan. He obtained his Bachelor of Science degree in Physics at National Taiwan University in Taipei. During his graduate training he was partially supported by the Structural Biology and Biophysics training grant form NIH.

Publications:


Chang Y. C. and Oas T. G. “Using Osmolyte to Study Folding Kinetics and Equilibrium of Bacillus subtilis RNase P Protein: An Intrinsically Unstructured Protein”. In preparation.

Chang Y. C., Franch W. and Oas T. G. “Probing the Folding Intermediate of Bacillus subtilis RNase P protein by NMR”. In preparation.

Chang Y. C. and Oas T. G. “Elucidation of Induced Fit Mechanism Requires Kinetic Measurements at Multiple Ligand Concentrations”. In preparation.