2012-07-25

Protein Ligand Interactions Probed by NMR: A Dissertation

Jennifer M. Laine
University of Massachusetts Medical School

PROTEIN LIGAND INTERACTIONS PROBED BY NMR

A Dissertation Presented

By

JENNIFER MARY LAINE

Submitted to the Faculty of the
University of Massachusetts Graduate School of Biomedical Sciences, Worcester
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

JULY 25, 2012

BIOCHEMISTRY AND MOLECULAR PHARMACOLOGY

PROTEIN LIGAND INTERACTIONS PROBED BY NMR

A Dissertation Presented
By

JENNIFER MARY LAINE

The signatures of the Dissertation Defense Committee signify completion and approval as to style and content of the Dissertation

_____
Francesca Massi, Ph.D., Thesis Advisor

_____
Daryl Bosco, Ph.D., Member of Committee

_____
Tatyana Igumenova, Ph.D., Member of Committee

_____
Sean Ryder, Ph.D., Member of Committee

_____
Lawrence Stern, Ph.D., Member of Committee

The signature of the Chair of the committee signifies that the written dissertation meets the requirements of the Dissertation Committee

_____
Celia Schiffer, Ph.D., Chair of Committee

The signature of the Dean of the Graduate School of Biomedical Sciences signifies that the student has met all graduation requirements of the school.

_____
Anthony Carruthers, Ph.D.,
Dean of the Graduate School of Biomedical Sciences

Biochemistry and Molecular Pharmacology
July 25, 2012

This work is dedicated to all the great teachers in my life who taught me a passion for learning and the importance of curiosity and imagination, and to my husband and daughters who taught me the importance of hard work and perseverance.

ACKNOWLEDGEMENTS

Foremost, I want to express my sincere gratitude to my thesis advisor Francesca Massi. Without her knowledge or guidance this thesis would not be possible.

I want to thank the following professors and members of my committee: Osman Bilsel, Daryl Bosco, Bob Matthews, Bill Royer, Sean Ryder, Celia Schiffer, Larry Stern and Jill Zitzewitz, for their insight, advice and thoughtful scientific discussions; each of these professors was essential for my success.

I also want to thank my present and former lab-mates Miguel Amat, Laura Deveau, Brian Farley, Brittany Morgan, Keith Romano and Kavitha Sompalli. I especially want to thank Laura Deveau for her friendship and her scientific insights. It has been a joy to work with each of you and I will miss you all.

The Biochemistry and Molecular Pharmacology department of the University of Massachusetts Medical is a wonderful place to work and learn. I would like to thank the entire BMP for welcoming and sustaining me through this adventure in medical research. It has been a blessing to be part of a true scientific community.

Lastly, I want to thank my parents and siblings, my husband Scott and his parents for all of their love and support during this process. I love you all very much.

ABSTRACT

Molecular recognition, defined as the specific interactions between two or more molecules, is at the center of many biological processes including catalysis, signal transduction, gene regulation and allostery. Allosteric regulation is the modification of function caused by an intermolecular interaction. Allosteric proteins modify their activity in response to a biological signal that is often transmitted through the interaction with a small effector molecule. Therefore, determination of the origins of intermolecular interactions involved in molecular recognition and allostery are essential for understanding biological processes. Classically, molecular recognition and allosteric regulation have been associated to structural changes of the system. NMR spectroscopic methods have indicated that changes in protein dynamics may also contribute to molecular recognition and allostery. This thesis is an investigation of the contributions of both structure and dynamics in molecular binding phenomena.

In chapter I, I describe molecular recognition, allostery and examples of allostery and cooperativity. Then I discuss the contribution of protein dynamics to function with a special focus on allosteric regulation. Lastly I introduce the hemoglobin homodimer, HbI of *Scapharca inaequivalvis* and the mRNA binding protein TIS11d.

Chapter II is the primary focus of this thesis on the contribution of protein dynamics to allostery in the dimeric hemoglobin of *scapharca inaequivalvi*s, HbI. Thereafter I concentrate on the mechanism of adenine recognition of the Tristetraprolin-like (TTP) protein TIS11d; this study is detailed in Chapter III. In Chapter IV I discuss broader impacts and future directions of my research.

This thesis presents an example of the use of protein NMR spectroscopy to probe ligand binding. The studies presented in this thesis emphasize the importance of dynamics in understanding protein function. Measurements of protein motions will be an element of future studies to understand protein function in health and disease.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

ALL – acute lymphoblastic leukemia

AML – acute myeloid leukemia

AREs – AU rich elements

DTT – dithiothreitol

CAM – calmodulin

cAMP – cyclic AMP

CAP – catabolite activator protein

CBP – Creb binding protein

CD – circular dichroism

CTD – C-terminal domain

Gdn – HCl - guanidine hydrochloride

Hb – Hemoglobin

HbI – Homodimeric hemoglobin of scapharca inaequivalvis

HCV – Hepatitis C Virus

H/D – hydrogen/deuterium

HisF – the cyclase moiety of imidazole-3-glycerol phosphate synthase

HX – hydrogen exchange

IRF-3 – Interferon regulatory factor 3

ILV – isoleucine, leucine and valine

ITC – isothermal titration calorimetry

KO – Knock out

MAVS – Mitochondrial antiviral signaling protein

MES – morpholineethanesulfonic acid

Mb – Myogllobin

MWC – Jacques Monod, Jeffries Wyman and Jean-Pierre Changeux concerted cooperative model

NHs – main chain amide hydrogens

NF- κB – Nuclear factor κB

NMR – Nuclear magnetic resonance

NS – Nonstructural

NS-HX – native state hydrogen exchange

NTD – N-terminal domain

PEG – polyethylene glycol

RBD – RNA binding domain

RIG-1 – Retinoic acid-inducible gene I

SF-CD – stopped-flow circular dichroism

TCEP – tris(2-carboxyethyl)phosphine

TLR3 – Toll-like receptor 3

TRIF – Toll-interleukin-1 receptor domain-containing adaptor-inducing beta interferon

TSE – transition state ensemble.

TZF – tandem zinc finger

UTR – Untranslated Region

WT – wild type

PREFACE

For clarity this dissertation has been organized such that each chapter encompasses one complete idea. Therefore each research based chapter and appendix includes an abstract, introduction, discussion and references. As a result there are some minor redundancies; this allows for each chapter to contain a complete idea.

CHAPTER II is a manuscript in preparation:

Laine J.M., Royer, W.E., F. Massi "Insights into the cooperativity of *scapharca inaequivalvis* dimeric hemoglobin." *PNAS*

CHAPTER III includes work to be published at a later date:

Laine J.M., L. M. Deveau, F. Massi "Characterizing the binding specificity of TIS11d."

Appendix A has been published as:

Romano, K. P., J. M. Laine, L. M. Deveau, H. Cao, F. Massi, and C. A. Schiffer. "Molecular Mechanisms of Viral and Host Cell Substrate Recognition by Hepatitis C Virus Ns3/4a Protease." *J Virol* 85, no. 13: 6106-16.

Appendix B is currently under review:

Basavanapura G. N., J. M. Laine, S. V. Kathuria, F. Massi, C. R. Matthews. "Clusters of Branched Aliphatic Side Chains Serve as Cores of Stability in the Native State of the HisF TIM Barrel Protein." *J Mol Bio*

CHAPTER I

INTRODUCTION

MOLECULAR RECOGNITION AND ALLOSTERY

Molecular recognition is a specific interaction between molecules driven by noncovalent bonding.(*1*) Covalent chemical bonds are formed between atoms that share an electron pair. Noncovalent bonds arise from favorable electromagnetic interactions and include: ionic bonds, hydrogen bonds, hydrophobic interactions and van der Waals forces.(*2*) Since noncovalent bonds are inherently weaker than covalent bonds, multiple noncovalent interactions are necessary to drive a favorable interaction between molecules. Structural studies suggest that high affinity interactions arise as the sum of many close range noncovalent interactions, and therefore, the complementarity between molecules is essential for the formation of high affinity molecular recognition.(*2*) Molecular recognition defines the specificity of an intermolecular interaction and provides insights into the range of biological partners a molecule can bind. Molecular recognition is important in many biological processes including enzyme catalysis, metabolism, signal transduction and gene regulation and allostery.(*3*)

Allostery is the modification of function caused by an intermolecular interaction. Allosteric proteins modify their activity in response to a biological signal that is often transmitted through intermolecular interactions with small effector molecules. The simplest model for allostery assumes that a proteins may exist and function in two states: an active “relaxed” (R) state and an inactive “tense” (T) state.(*4*) In this model, allostery is achieved when exogenous signal promotes a structural transition from the T state to the R state.

Determination of the origins of intermolecular interactions involved in molecular recognition and allostery are essential for understanding and intervening in biological processes. Classically, crystallographic studies have provided a structural explanation for the origin of intermolecular interactions, molecular recognition and allosteric regulation.

However, new methods in NMR have indicated that a structural transition between low and high activity conformers is not the only way molecular recognition or allostery may be achieved. In 1984, Dryden proposed a model where allosteric communication could be achieved in the absence of a structural transition.(*5*) In this model, cooperativity is entropically driven. Changes in protein conformational entropy can differentiate high and low affinity states and drive the cooperative mechanism. As such, small changes in the conformational fluctuations of many atoms in the protein upon binding the effector molecule, could add up to a significant difference in free energy. The importance of changes in protein dynamic in ligand binding processes and in allostery has been recently highlighted from NMR relaxation studies.(*3, 6, 7*)

MOLECULAR RECOGNITION AND BINDING SPECIFICITY

Protein promiscuity defines a protein's ability to perform multiple functions. Differential expression and environmental conditions such as pH and temperature may drive protein promiscuity.(*8, 9*) Additionally, post-translational modifications, oligomeric states, the concentration of ligand and allostery can transform protein function.(*10-12*) Furthermore, proteins may conduct multiple functions through promiscuous molecular recognition. Characterizing the factors leading to a protein's promiscuity is essential for understanding the broad role of a protein's function in the cell. An important topic of this thesis is on promiscuous molecular recognition.

Protein molecular recognition of biological partners forms the biochemical foundation of life.(*13*) Fundamental biological processes such as DNA replication, transcription, translation, and signal transduction are mediated by the molecular recognition of biological molecules in the cell. The promiscuity of protein molecular recognition defines the binding specificity of a protein. Proteins can be highly selective for one binding partner or they may function promiscuously and bind many non-complementary partners.(*13*) The total range of partners a protein binds defines a protein's role in the cell under a set of conditions.

Classically, it is believed that promiscuous proteins are able to bind many different ligands using hydrophobic interactions, hydrogen bonds and perhaps most importantly conformational flexibility.(*13*) Protein flexibility allows promiscuous proteins to adopt several conformations to bind different ligands

with comparable affinities. The diverse array of protein complexes of a promiscuous protein naturally suggest that the protein is flexible enough to exchange between all the known conformations; and structural and dynamic studies strongly support the theory of conformational selection of a protein in a dynamic equilibrium.(*14, 15*) In addition to conformational exchange promiscuous proteins may use changes in protein flexibility to modulate the contribution of conformational entropy to bind each unique partner.(*16-18*)

THE CHARACTERIZATION OF HEMOGLOBIN COOPERATIVITY

Human hemoglobin (Hb) is the first and most thoroughly characterized allosteric protein. It has formed the basis of our understanding of allosteric regulation and was the focus of the first determination of the molecular origin of a disease – sickle cell anemia.(*19*) The following is a brief history of the discovery and modeling of Hb cooperativity.

Cooperativity is the simplest form of allosteric regulation. In a cooperative system the ligand acts as an effector molecule, such that ligand binding alters the protein affinity for additional ligands. There are two kinds of cooperativity, positive and negative. For positive cooperativity, ligand binding causes an increase in affinity for the ligand at distal binding sites; whereas, for negative cooperativity, ligand binding causes a decrease in affinity for the ligand at distal binding sites. Hemoglobin binds oxygen with positive cooperativity.(*20*) Each of the following models has been modified to better represent Hb cooperativity but can be applied in other systems as appropriate.

Myoglobin, Mb, is a monomeric non-cooperative homologue of Hb. Mb binds oxygen as described by the following equation

$$Mb + O_2 \rightleftharpoons MbO_2. \qquad (1)$$

The hyperbolic binding curve of Mb to oxygen is shown in **Figure 1.1**. The allosteric nature of Hb was first observed by the physiologist Christian Bohr in 1904.(*19*) He recognized that hemoglobin did not bind oxygen like myoglobin. For a simple equilibrium as shown in (1), the fraction of HbO_2 as a function of oxygen concentration would be hyperbolic. Instead, Bohr observed that Hb binds

oxygen according to a sigmoidal curve, **Figure 1.1**. He recognized that Hb has an increased affinity for oxygen as more oxygen is bound indicating that Hb binds oxygen cooperatively. Although the cooperative mechanism was unknown, Bohr determined that the cooperativity of Hb was essential for the efficient transfer of oxygen and carbon dioxide to the body and the lungs, respectively.(*21*)

Archibald Vivian Hill derived the Hill equation in order to model cooperative binding curve of Hb. He hypothesized that hemoglobin may bind several oxygen molecules as follows:

$$Hb + nO_2 \leftrightarrow Hb\left(O_2\right)_n . \qquad (2)$$

To model the equilibrium in Equation 3, Hill derived the following equation

$$\theta = \frac{\left[L\right]^n}{K_d{}^n + \left[L\right]^n} , \qquad (3)$$

such that L is free ligand, θ is the fraction of ligand bound, K_d is the dissociation constant and n is the Hill coefficient.(*22*)

Gilbert Adair determined the molecular weight of Hb and considering the stoichiometric ratio of iron to Hb he concluded that Hb contained four oxygen binding sites.(*23*) The Hill equation assumes that Hb binds four oxygen molecules simultaneously and is insufficient to model Hb cooperative binding. This observation indicated that the Hill equation is insufficient to model Hb cooperativity since oxygen binding was best fit with a Hill coefficient of 2.8.(*24*) From this, Adair proposed an equilibrium model where the binding affinity of Hb to O_2 increased for each successive binding event. **Figure 1.2** is a visual representation of the sequential model.

Figure 1.1: Oxygen dissociation curves of human myoglobin and human hemoglobin. The fraction of ligand bound, Y, is shown as a function of the partial pressure of oxygen.(*22, 25*) The myoglobin dissociation curve shown in black is hyperbolic as expected from equation 1. The hemoglobin curve in red is sigmoidal indicative of cooperative binding.

The plot is generated from the following references:

Reginald Garrett, C.M.G. *Biochemistry*, 1141 (Cengage Learning, 2010).

Bisswanger, H. *Enzyme Kinetics: Principles and Methods*, 320 (John Wiley & Sons, 2008).

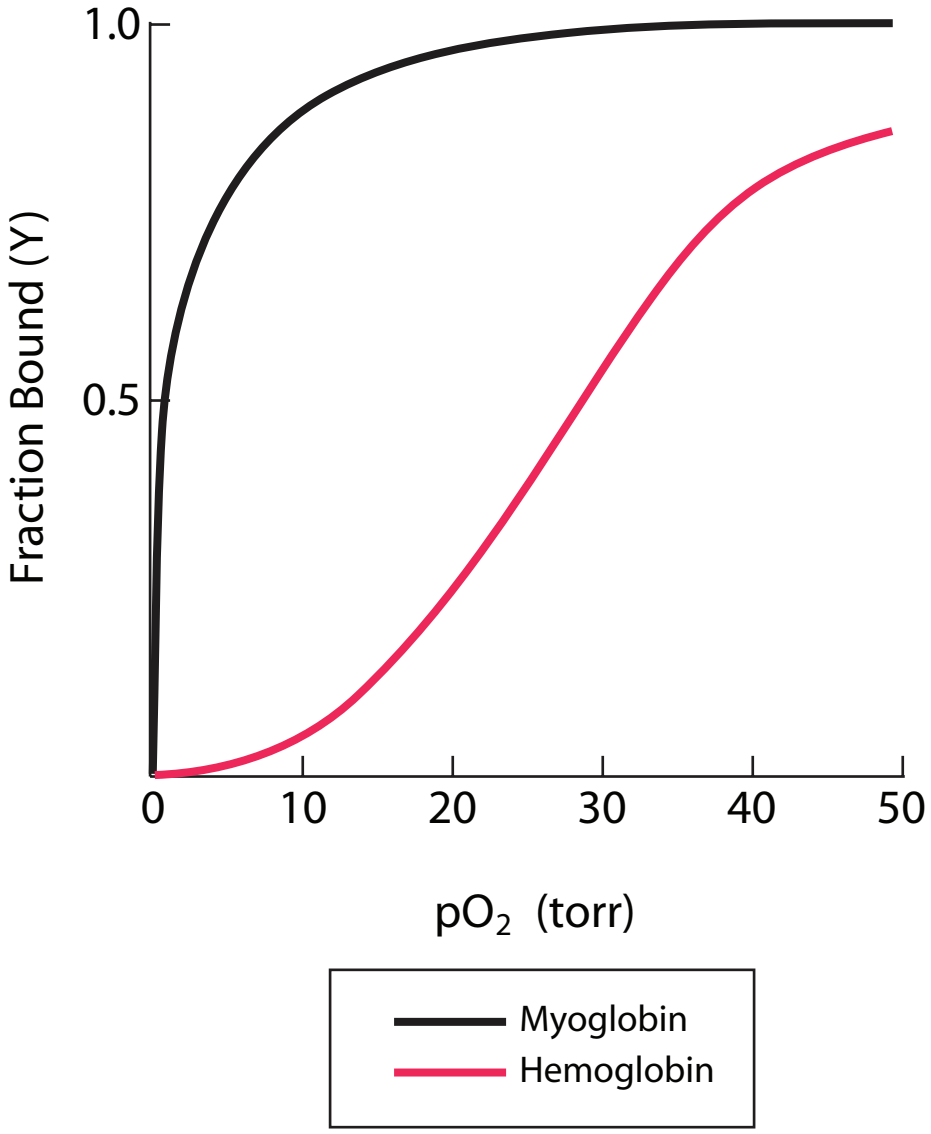

1.0
0.5
Fraction Bound (Y)
0
10
20
30
40
50
pO2 (torr)
Myoglobin
Hemoglobin

Figure 1.1

Figure 1.2: The Sequential Model and the MWC model. In both the sequential and MWC models, the T and R states are designated by a blue square or a red circle, respectively. For the sequential model, dissociation constants K_1, K_2, K_3, and K_4 decrease consecutively for a positive cooperative system like Hb. The MWC model assumes only two quaternary conformations of Hb exist. The ratio of T and R states at equilibrium is L. Without oxygen, the T state is favored and therefore L is greater than one. For positively cooperative Hb, the equilibrium L is shifted towards the R state by the constant c, which is the ratio of the dissociation constants of ligand by the R state, K_R and the T state K_T. Since the R state binds more tightly to oxygen than the T state, the ratio of K_R to K_T is less than one.

Image generated by J. M. Laine.

References:

Monod, J., Wyman, J., and Changeux, J. P. (1965) On the Nature of Allosteric Transitions: A Plausible Model, J Mol Biol 12, 88-118.

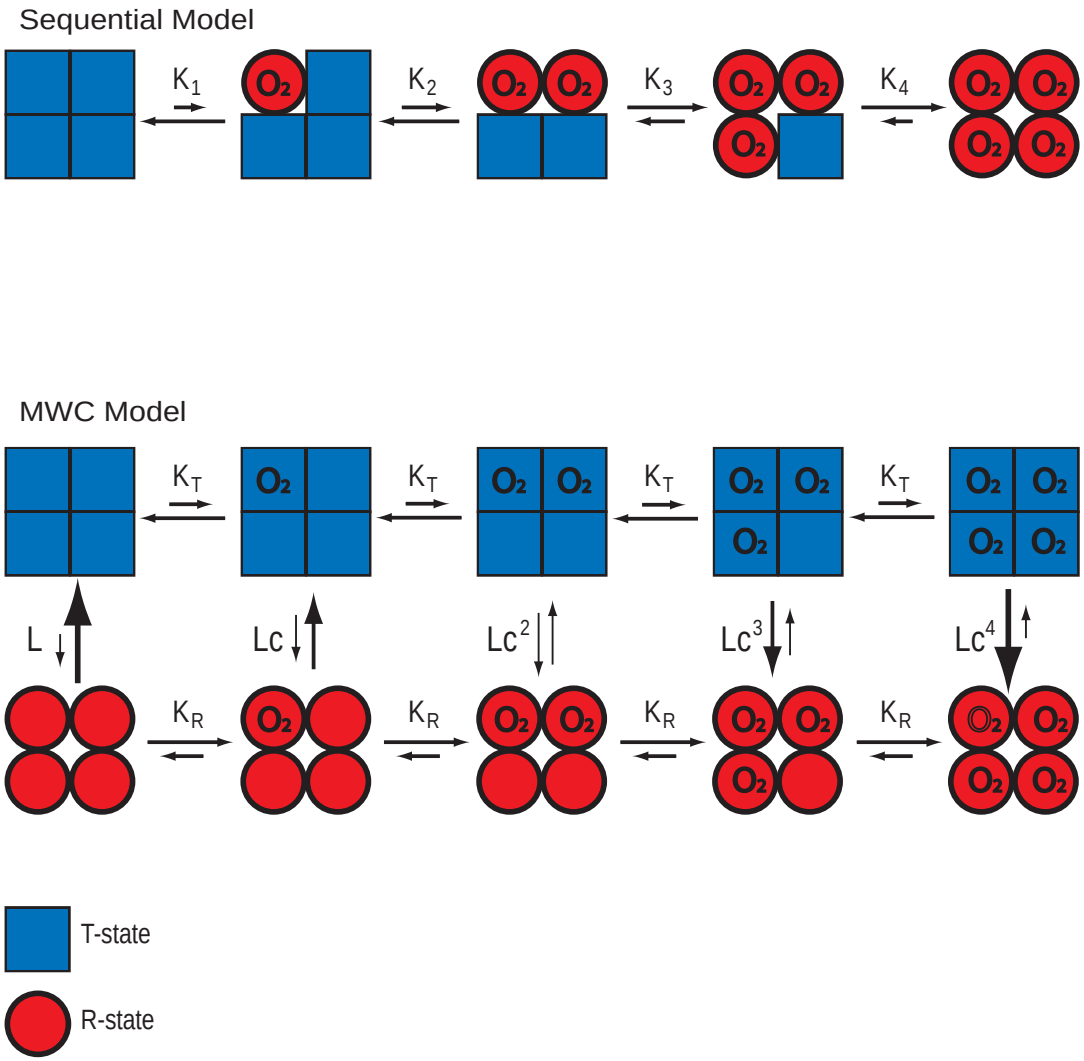

Sequential Model
K1
K2
K3
K4
O2
MWC Model
KT
KR
L
Lc
Lc2
Lc3
Lc4
T-state
R-state

Figure 1.2

Linus Pauling proposed the first structural model for Hb cooperative binding.(*26*) He postulated that through heme-heme interactions oxygen binding by a heme group could cause nearby hemes to increase their oxygen affinity. Assuming that Hb contains four equivalent hemes that can interact with two other hemes, Pauling was able to model the sigmoidal curve of Hb oxygen binding.(*26*)

In 1960, Max Perutz, the pioneer of protein structure determination, solved the crystal structure of horse hemoglobin at 5.5 Å.(*27*) The hemoglobin tetramer consists of two α and two β subunits; the complex is a dimer of $\alpha\beta$ heterodimers. In 1964 Perutz determined the unliganded form of horse hemoglobin and determined that the β subunits within the tetramer move closer together upon binding oxygen.(*27*) This discovery inspired Jacques Monod, Jean-Pierre Changeux and Jeffries Wyman to better model Hb cooperativity and allostery in general for multimeric proteins.(*19*)

Together, Monod, Wyman and Changeux devised the MWC model, wherein cooperativity arises due to concerted conformational rearrangements between subunits of multimeric proteins.(*28*) In this model Hb exists in one of two states, the high affinity R or low affinity T state. In this case oxygen drives the formation of the R state and cooperativity is driven by population shift. In this model, the quaternary conformation determines the affinity for ligand.(*28*) **Figure 1.2** shows the MWC model applied to Hb in detail.

In support of this model, Perutz developed a “stereo-chemical” mechanism to describe Hb cooperativity.(*29*) Perutz noticed that in the Hb T state salt bridges exist at the subunit interface; these contacts do not exist in the R state.

According to Perutz's mechanism binding oxygen disrupts the interface salt bridges, effectively destabilizing the T state and driving the formation of the R state.

The binding affinity of Hb for oxygen is altered by changes in pH; this is called the heterotropic effect. Perutz observed that a proton is released upon binding ligand and hypothesized that pH and the cooperative transition of Hb were linked and caused the heterotropic effect. Attila Szabo and Martin Karplus validated Perutz's mechanism and strengthened the MWC model by quantitatively reproducing the equilibrium properties of hemoglobin with respect to pH.(*30*)

Structural studies on Hb reveal how pH can affect binding affinity for oxygen, and they reveal how hemoglobin mutations distant from the heme stabilize a conformation and therefore, alter Hb affinity for ligand. Edelstein used structural studies to determine how protein mutations cause disease. The sickle cell mutation Glu-β6-Val forms a hydrophobic patch at the surface of Hb. Sickling occurs when Hb forms fibers, which consist of fourteen Hb molecules woven into helical strands formed from interactions at the hydrophobic patch.(*31*) These strands form from the Hb T conformation but not the R state.(*31*) This is consistent with the disease condition where cell sickling occurs in the blood upon deoxygenation and cells unsickle upon oxygenation at the lungs.(*31, 32*)

The determination of the structure of hemoglobin and the characterization of hemoglobin biochemically forged a new era in medicine. For the first time the structure of a protein was known to be critical in characterizing the protein's

function as well as a molecular disease. With only slight modifications, the two-state allosteric MWC model has quantitatively modeled equilibrium and kinetic studies of hemoglobin.(*19*) The investigation into hemoglobin cooperativity has formed the backbone for understanding allosteric regulation and cooperativity was explained by structural changes in one subunit that drive structural changes in the second.

CONTRIBUTION OF PROTEIN DYNAMICS TO MOLECULAR RECOGNITION AND ALLOSTERY

NMR spectroscopy is a powerful technique to study molecular recognition and allosteric proteins because it can be used to probe both structure and dynamics with atomic resolution. This section reviews examples of the contribution of fast (ps-ns) and slow (μs-ms) protein dynamics in protein function.

Fast time scale (ps-ns) Dynamics

Protein structures determined through x-ray crystallography define the mean conformation of a protein averaged over many identical molecules within the protein crystal.(*33*) Fast time scale motions corresponding to small amplitude conformational changes cannot be resolved by this technique.(*5*) NMR techniques provide evidence for the existence of intramolecular motions across many timescales including bond vibrations, side-chain rotations, bending at a hinge, local unfolding and global oscillations or protein "breathing".(*5, 33*) Bond vibrations and bond librations occur on the ps-ns time scale and establish the range of conformational microstates available to a protein.(*34*) Conformational entropy is the residual entropy of the mean conformation of a protein arising from the number of available microstates a protein explores and therefore is related to protein flexibility.(*35, 36*)

Classically binding processes are described according to a structural transition that occurs upon binding ligand. Ligand binding can stabilize one conformation of a protein over others without perturbing the dynamic fluctuations

of the atomic position about the mean conformation, **Figure 1.3(A)**.

Alternatively a ligand can also change the shape of the probability distribution of the dynamic fluctuations of the atomic positions about the conformational mean without perturbing its position, **Figure 1.3(B)**. In the latter case ligand binding occurs in the absence of a structural transition. In most cases, both changes in structure and dynamics contribute to ligand binding, **Figure 1.3C**. Because fast time scale dynamics provide information on conformational entropy,(*37*) studies of protein dynamics occurring in the ps-ns time scale can provide insight into the thermodynamic role of flexibility in protein binding and function.

At physiological temperature, proteins fluctuate between microstates within a conformational well on the ps-ns timescale. The total number of microstates available to the system determines the protein conformational entropy.(*37*) Fast time scale dynamics occurring in the ps-ns time scale are sensitive to the number of microstates a protein explores within the conformational well that make the largest contribution to the conformational entropy of the protein. Therefore, fast time scale dynamics can be used to estimate the conformational entropy of the system.(*38*)

Figure 1.3: Theoretical reaction coordinates of ligand binding. A) Ligand binding is driven by a conformational transition in protein, P. B) Ligand binding occurs in the absence of a conformational transition. Given that the width of an energetic well is proportional to the number of accessible microstates. The red bar indicates an increase in protein (*7*) and accessible microstates and is thermodynamically spontaneous. The blue bar indicates a decrease in protein flexibility and accessible microstates and binding is disfavored. C) Ligand binding causes a change in the mean conformation of P and simultaneously alters the available microstates the protein explores.

Figure created by J. M. Laine.

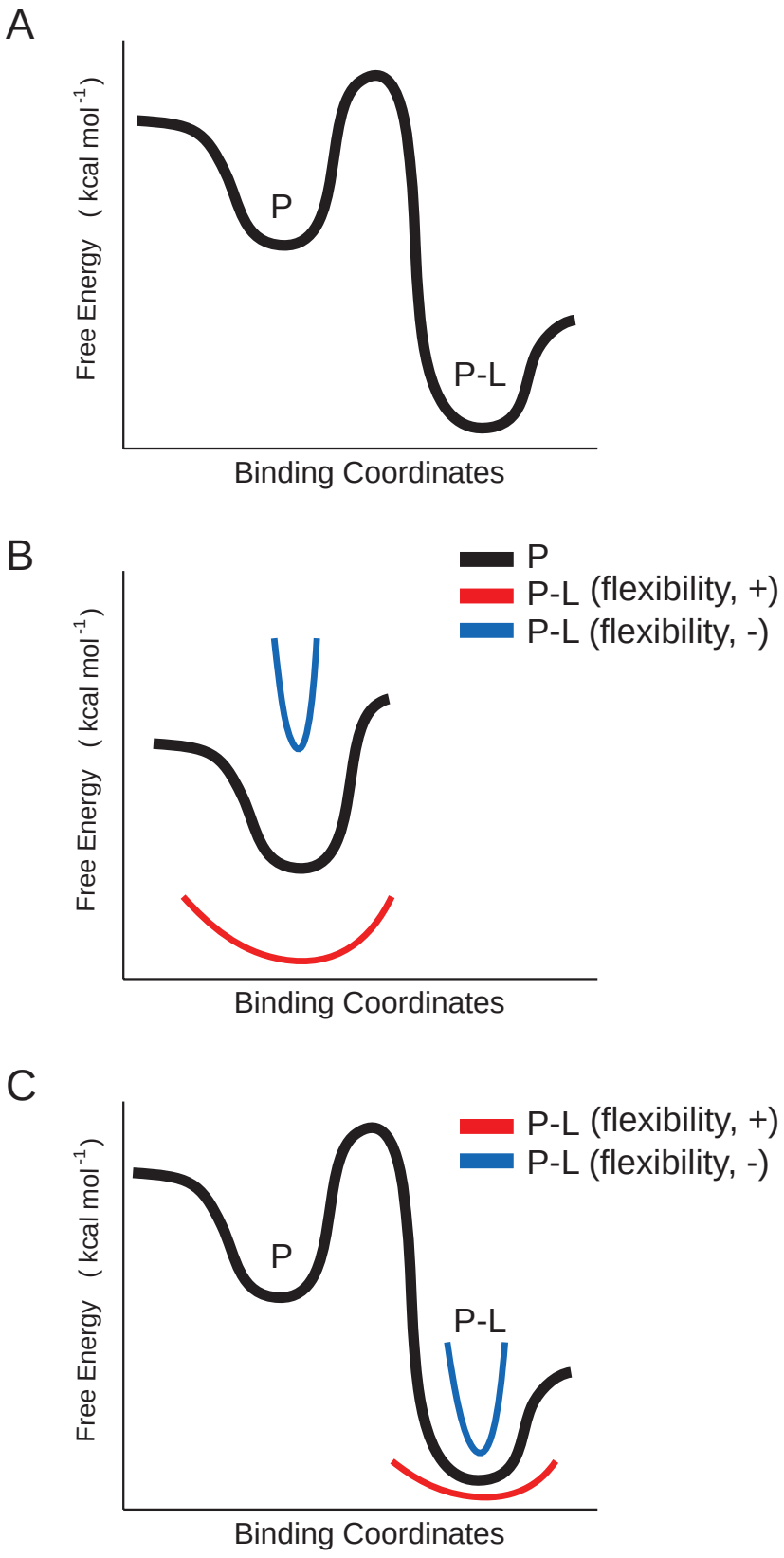

A
Free Energy (kcal mol-1)
P
P-L
Binding Coordinates
B
P
P-L (flexibility, +)
P-L (flexibility, -)
Free Energy (kcal mol-1)
Binding Coordinates
C
P-L (flexibility, +)
P-L (flexibility, -)
Free Energy (kcal mol-1)
P
P-L
Binding Coordinates

Figure 1.3

Using Lipari-Szabo model-free analysis the internal motion of a protein can be characterized by a generalized order parameter, S^2, and the internal correlation time, τ_e.(*39, 40*) S^2 and τ_e define the amplitude and timescale of the motion of a bond vector, respectively. **Figure 1.4** shows the stochastic motions of a bond vector within a cone. The change in internal dynamics observed upon ligand binding reflects changes in conformational entropy and can be calculated using

$$\Delta S = -k_B \sum_i \ln\left\{\frac{3-\left(1+8S_{R,i}\right)^{1/2}}{3-\left(1+8S_{T,i}\right)^{1/2}}\right\}, (37) \qquad (4)$$

for which k_B is the Boltzmann constant, and S_R and S_T are the square root of the order parameters of the R and T states respectively for spin *i*. This analysis is limited because only those order parameters that can be directly compared can be included. Secondly, all residues with $S^2 > 0.95$ must be excluded because the entropic changes associated with very rigid residues is underestimated by this equation. Lastly, this equation assumes that all motions are uncorrelated and the calculated value should be considered an upper limit of the conformational entropy difference for the residues included.(*37*)

Figure 1.4: Stochastic motions within a cone of a bond vector, μ. The Lipari-Szabo order paramater is related to θ°, the cone semi-angle, by the function $S^2 = \left[0.5\cos\theta\left(1+\cos\theta\right)\right]^2$.(*34*) When $S^2 = 1$ the bond is rigid, and $S^2 = 0$ the internal motion is isotropic. (*41, 42*)

Figure adapted from the following reference: Ishima, R. & Torchia, D.A. Protein dynamics from NMR. Nat Struct Biol 7, 740-3 (2000).

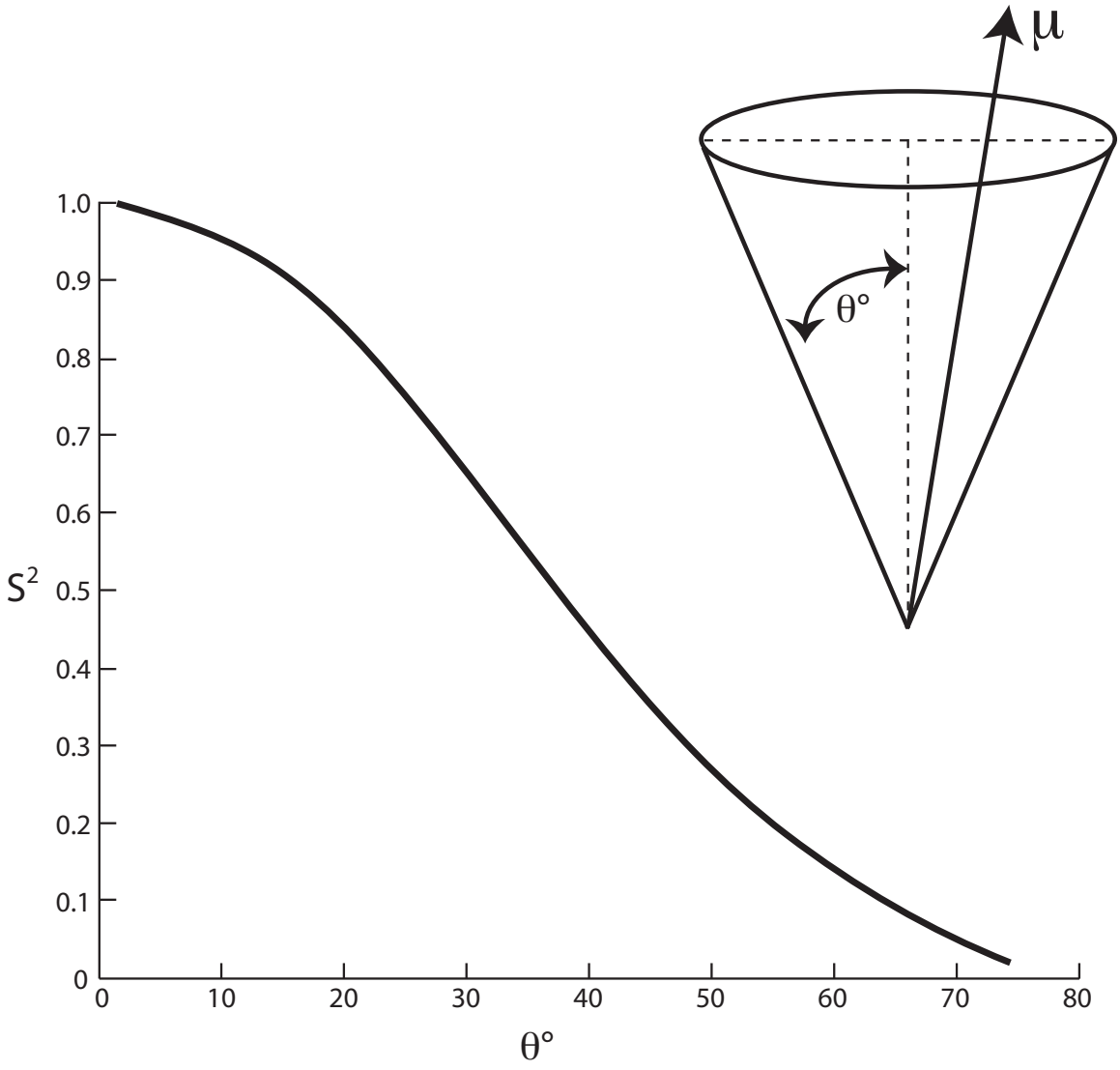


Figure 1.4

Protein flexibility is essential for proper protein function.(*43-45*) For example, fast time scale motions in the flaps of HIV protease allows for targets to gain access and bind the HIV active site.(*44*) DNA binding proteins utilize protein flexibility to coordinate non-specifically with DNA to search more effectively for target sequences.(*43*) Protein flexibility and changes in protein conformational entropy has been implicated in protein function, ligand binding and allosteric regulation.(*3, 6, 7, 18, 44, 46, 47*) Studies on calmodulin (CaM) peptide binding by Wand et al. provided the first evidence that NMR S^2 can be used to assess protein conformational entropy.(*17, 18*)

CaM binds hundreds of different α-helical peptides with similar affinities but with diverse entropic contributions.(*16*) Changes in methyl side chains order parameters of CaM were determined for each peptide and used to calculate an apparent change in CaM conformational entropy for each reaction. Remarkably, changes in CaM conformational entropy correlate linearly to the total change in entropy of binding, and CaM conformational entropy is a significant component to high affinity binding.(*17*)

Wand et al. used the calmodulin system to discern the magnitude of the contribution of each component of the total change in entropy, ΔS_{tot}.(*18*)

$$\Delta S_{tot} = \Delta S_{protein} + \Delta S_{ligand} + \Delta S_{solvent} + \Delta S_{RT} \qquad (5)$$

where $\Delta S_{protein}$, ΔS_{ligand} are the change in conformational entropy of the protein and ligand respectively, ΔS_{RT} is due to the change in rotational and translational entropy associated with the tumbling of CaM and ligand, $\Delta S_{solvent}$ is due to changes in the entropy of the water. The changes in conformational entropy of

each of the six CaM-peptide complexes are linearly related to the total entropy as follows.(*18*)

$$\Delta S_{tot} - \Delta S_{solvent} = m\left[n_{res}^{protein}\left\langle \Delta S_{protein}^{2}\right\rangle + n_{res}^{ligand}\left\langle \Delta S_{ligand}^{2}\right\rangle\right] + \Delta S_{RT} + \Delta S_{other} \qquad (6)$$

ΔS_{tot} is measured by isothermal titration calorimetery. $\Delta S_{solvent}$ is calculated from changes in the solvent accessible surface area of the free and bound structures of CaM and both $S_{protein}$ and ΔS_{ligand} are determined from changes in methyl side chain order parameters, $\Delta S^2_{protein}$ and ΔS^2_{ligand} normalized by residue, n. In this case the contribution of $\Delta S_{RT} + \Delta S_{other}$ is the y-intercept and is small. The slope, m is the relationship of the changes in conformational entropy of each CaM complex to the total entropic contribution of binding. For each complex the entropic contribution from the change in conformational entropy drives binding.(*18*) This study provides absolute evidence of the robust contribution of conformational entropy to high affinity binding and suggests that conformational entropy will contribute to all kinds of binding events including allosteric regulation.

The NMR study of the negative cooperative binding of catabolite activator protein, CAP to cyclic AMP, cAMP is a prime example of the importance of protein flexibility in cooperative binding. CAP is a symmetric homodimer capable of binding two cAMP molecules. In this study Kalodimos et al., characterizes changes in the structure and dynamics of CAP, cAMP-CAP and $cAMP_2$-CAP.(*6*) Upon binding one cAMP, fast time scale motions are unaffected in both monomers. Although the ligand bound monomer changes conformations, the unbound monomer does not undergo a structural transition until it binds ligand as well.(*6*) Binding two cAMP molecules dampens all fast time scale motions and

causes a robust conformational entropic penalty and causes the negative cooperativity of CAP.(*6*)

Slow time scale (μs-ms) Dynamics

Slow-time scale protein dynamics occurring in the μs-ms time scale are biologically significant because these motions occur at the same timescale as binding events, protein folding and conformational rearrangements.(*6, 48-51*) The kinetics of a two state chemical exchange process is defined as

$$R \underset{k_2}{\overset{k_1}{\rightleftarrows}} T , \qquad (7)$$

where k_1 is the forward rate constant and k_2 is the reverse rate constant. The equilibrium constant is defined as $K_{eq} = k_1/k_2$ and the apparent exchange rate is defined as $k_{ex} = k_1 + k_2$.

An atom's chemical shift is dependent on its local chemical environment. As a result chemical shift is very sensitive to the conformation of a protein and can be used to observe structural rearrangements. The observed chemical shift, δ_{obs}, is a population-weighted average of the conformational states the protein accesses. In a simple two state R and T model

$$\delta_{obs} = p_R\delta_R + p_T\delta_T , \qquad (8)$$

p_R and p_T are the population of the R and T conformers at equilibrium and δ_R and δ_T are the chemical shifts of each state. If δ_R and δ_T are different, k_{ex} can be observed using NMR techniques. Using the Lipari-Szabo model free analysis the chemical exchange contribution to relaxation relaxation, R_{ex}, can be measured.

To probe these motions further, relaxation dispersion methods can be used to measure the rate of exchange, the population distribution of the states and the chemical shift of lowly populated and “invisible” species.(*39, 52, 53*) The importance of lowly populated conformations has been highlighted in binding processes, enzyme catalysis and allostery.(*14*)

Using relaxation dispersion techniques, Loria et al. discovered that the inherent slow time scale motions and the enzymatic function of RNase A are coupled.(*54*) Regions of RNAse A fluctuates between two conformations open and closed, and the rate of slow time scale protein motions of regions of RNase A are the same as the rate of catalysis, k_{cat} and the off rate, k_{off}.(*55*) Binding substrate or product shifts the population of RNase A to the closed conformation, however the rate of exchange of RNase A is the same as the enzyme-substrate complex and the enzyme-product complex.(*56*) RNase A is always at the equilibrium between bound and free conformations and the population of each conformation changes in each step of the catalytic cycle. By sampling the next relevant conformation of the catalytic cycle slow motions of RNase A drive catalytic turnover.(*56*)

In addition to enzyme function, Tollinger et al. used relaxation dispersion techniques to determine the role of slow time scale dynamics in the allosteric regulation of the KIX domain of CREB binding protein, CBP.(*53*) KIX is a modular domain of CBP and it is capable of binding many different transcription factors at two distinct sites. At one locus KIX can bind the mixed lineage leukemia, MLL protein. KIX binds the transcription factor c-Myb at a second

distal locus. Binding to MLL activates KIX and doubles it's affinity for cMyb.(*57*) Upon binding MLL, KIX undergoes a population shift of 7% from the T state to the cMyb conformation, R state.(*53*) In this case, the allosteric mechanism is driven by conformational selection consistent with the MWC model.

MOLECULAR RECOGNITION OF RNA BY TIS11D

The human protein TIS11d regulates mRNAs involved in mammalian hematopoiesis.(*58, 59*) Knock out, KO, mice of the TIS11d mouse homologue gene *Zfp36l2* hemorrhage and die two weeks from birth.(*58*) The mice have fewer hematopoietic progenitor cells and as a result have reduced red and white blood cells, hemoglobin, hematocrit and platelets.(*58*) Recent screens of patient suffering from acute myeloid leukemia, AML, and acute lymphoblastic leukemia, ALL, have implicated TIS11d in the pathogenesis of these diseases.(*59*) In patients and in cancerous immortalized cell lines, both single-point mutations as well as frameshift mutation and point mutations have been identified throughout the *TIS11d* gene locus.(*59*)

TIS11d is one of the three known human TTP-like proteins, TTP, TIS11b and TIS11d. Trisetraprolin is also known as TTP, TIS11, ZFP36, or Nup475. TIS11b is also known as ERF1, ZFP36L1, or cMG1. TIS11d is also known as ERF2, ZFP36L2, or the gene encoding butyrate response factor-2. In mice a fourth member has been identified in placenta, ZFP36L3.(*60*) TTP and TTP-like proteins are involved in the regulation of important cytokines and growth factors associated with chronic inflammation and cancer.(*61, 62*)

TTP-like proteins contain three domains types, N-terminal domain, NTD, C-terminal domain, CTD, and an RNA binding domain, RBD, **Figure 1.5**.

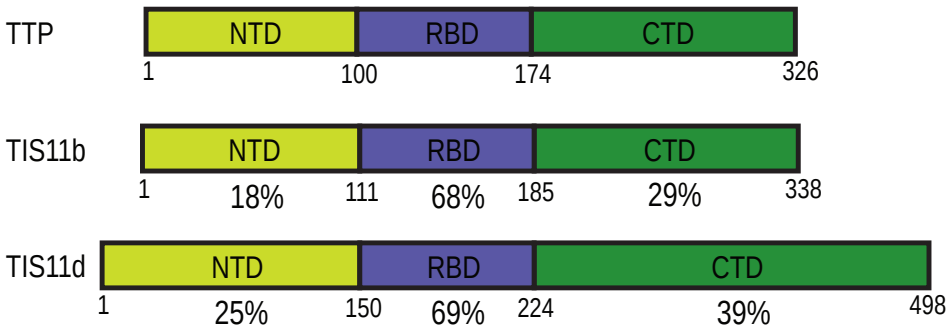
Figure 1.5: TTP-like proteins bind RNA through a CCCH TZF domain. TTP-like proteins contain a N-terminus domain (NTD) an RNA binding domain (RBD) and a C-teriminus domain (CTD). The NTD is shown in pea green, the RBD is purple and the CTD is dark green. The homology of each domain of TIS11b and TIS11d is calculated with respects to TTP, the prototype of the TTP-family. B) The structure of the RBD of TIS11d bound to ARE_{13} is shown within the cartoon. Zinc atoms are shown as purple spheres coordinated by the CCCH residues (pink sticks) of the first and second zinc fingers of TIS11d. The RNA cognate sequence bound to TIS11d is shown as light purple sticks. C) The RBD of the TTP family is highly identical. The amino acid sequences of the RBD of TTP, TIS11d and TIS11b are shown. CCCH residues are highlighted in pink. Residues that differentiate each sequence from the family are bold and underlined.

The structure of TIS11d-ARE_{13} is generated from the PBD file 14GO.

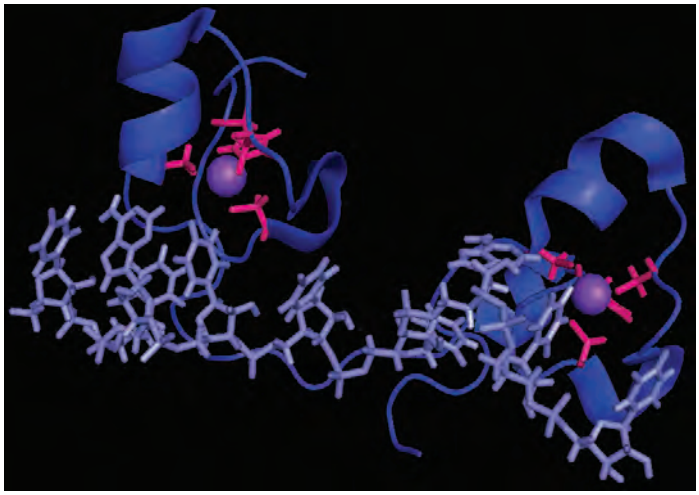
Reference:

Hudson, B.P., Martinez-Yamout, M.A., Dyson, H.J. & Wright, P.E. Recognition of the mRNA AU-rich element by the zinc finger domain of TIS11d. *Nat Struct Mol Biol* **11**, 257-64 (2004).

A



B



C

Figure 1.5

The NTD and CTD coordinate mRNA degradation by recruiting and binding mRNA degradation machinery.(*10, 63*) The RBD recognizes and binds to AU rich elements, (AREs) of single stranded RNA through a $CX_8CX_5CX_3H$ type tandem zinc finger (TZF). The RBD of TTP-like proteins binds to an ARE cognate sequence 5′-UUAUUUAUU-3′.(*64, 65*) The primary sequence homology of the TTP family is listed for each protein in **Figure 1.5**. AML and ALL related point mutations have been discovered in each domain of the *TIS11d* gene locus.(*59*) Specifically, the missense mutations H62Q and D219E are located in the NTD and the RBD, respectively, are linked to AML. These residues are critical for proper function as they are conserved within the TTP family.(*59*)

The RBD of TIS11d consists of two zinc fingers (ZF1 and ZF2) connected by an 18 amino acid linker. The length of the linker is conserved within the TTP family. Two sequences preceding each ZF in the TTP family are precisely conserved; the sequence RYKTEL directly precedes ZF1 and the sequence KYKTEL directly precedes ZF2.

The NMR structure of TIS11d bound to a 5′-UUAUUUAUU-3′ reveals that each ZF of TIS11d binds 5′-UAUU-3′ RNA element exclusively through hydrogen bonding, van der Waals forces, and base stacking.(*66, 67*) Each R/KYKTEL sequences makes significant contacts with the RNA to form a pocket and bind U6 and U2 at ZF1 and ZF2 respectively. The glutamate and leucine of R/KYKTEL and R160 and R198 of ZF1 and ZF2 recognize the adenine of each UAUU element through hydrogen bonds. The amide backbone of TIS11d forms hydrogen bonds with the Watson-Crick edge of each nucleotide. The conserved

aromatic residues Y170, F176, Y208 and F214, base stack with RNA bases.

Molecular dynamic simulations of the RBD of TIS11d reveal that without RNA the structure of the ZFs of TIS1d remain intact but reorient such that TIS11d becomes more compact.(*68*) To bind RNA, TIS11d must open globally and the fingers must reorient.(*68*) This study shows that the backbone flexibility of TIS11d in complex with ARE is more rigid than when bound to polyuridine.(*68*)

HbI Cooperativity

The hemoglobin homodimer of *Scapharca inaequivalvis*, HbI, binds oxygen cooperatively and is driven by changes in both structure(*69*) and dynamics. Like Hb, HbI binds gaseous ligands such as oxygen, O_2, nitric oxide, NO, and carbon monoxide, CO at its coordinated Fe(II)-heme (ferroprotopophyrin IX). The tertiary fold of homodimeric HbI is homologous to human hemoglobin although its quaternary arrangement of subunits is unique. In the case of HbI the E and F helices of the globin fold form the dimeric interface; in Hb these helices are solvent exposed.(*70*) At the HbI interface, the hemes are directly connected to the opposite symmetric monomer through hydrogen bonds in both the R and T states, **Figure 1.6**. In the T state Asn100 and Lys96 hydrogen bond with the heme of the opposite monomer at heme propionates 1 and 2; in total 6 hydrogen bonds are formed at the interface. Upon binding ligand, the interface broadens and Asn100 and Lys96 hydrogen bond to propionate 1.(*70*) In the R state, Lys96 is indirectly bonded to the heme of the opposite monomer at propionate 2 but this connection is mediated through a water molecule. In total, there are 6 links in the HbI R and T states.

The difference in subunit orientation between oxygen bound and deoxy hemoglobin is a 3.4° rotation.(*69, 71*) Similarly, the distance between iron atoms of the heme groups only shifts 1.8 Å, such that the distance between hemes is 16.6 Å in the unliganded state and 18.4 Å in the liganded state. In the HbI unliganded T state conformation, a key residue, Phe97, packs against the proximial histidine, His101, within the heme pocket.(*70*) This packing maintains

the concave conformation of the heme. The HbI interfaces of the R and T states are shown in **Figure 1.6**. Upon binding ligand, F97 is released from the heme pocket into the homodimer interface where it disrupts coordinated water molecules. The motion of the Phe side chain to the interface allows the proximal His to form a strong hydrogen bond with the carbonyl oxygen of F97 and the iron is free to move into the plane of the heme.

Studies by Royer et al., imply that water is important for communication between monomers at the interface of HbI. A cluster of 17 crystallographic waters are found at the HbI dimer interface of deoxy HbI; CO-HbI has 11 crystallographic waters, **Figure 1.7**.(*72*) HbI binds oxygen more tightly in the presence of high concentrations of solutes such as glycerol, glucose and sucrose; solute causes an increase in binding affinity by stabilizing the bound state of HbI.(*72*) Fits of HbI affinity for O_2 with respect to osmotic pressure estimate that HbI releases 6 water molecules upon binding O_2.(*72*) The osmotic pressure alters the water interface between HbI subunits and simultaneously disrupts communication between subunits. This suggests that water is essential for communication between HbI subunits and is important for HbI cooperativity.

Figure 1.6: Local conformational changes at the interface of HbI. Here are the interfaces of T (top) and R (bottom) states of HbI. The interfacial E helices are gray and the F helices are black, hemes are blue and CO is magenta. The propionate groups of the hemes are labeled 1 and 2. Important residues side-chains are colored. The proximal histidine, His101 is pink. The distal histidine, His69, is cyan. The interface residues Lys96, Phe97 and Asn100 are orange, red and green respectively.

Structures of unliganded and CO-HBI are generated from the PBD files 4SDH and 3SDH respectively.

Reference:

Royer, W.E., Jr. High-resolution crystallographic analysis of a co-operative dimeric hemoglobin. J Mol Biol 235, 657-81 (1994).

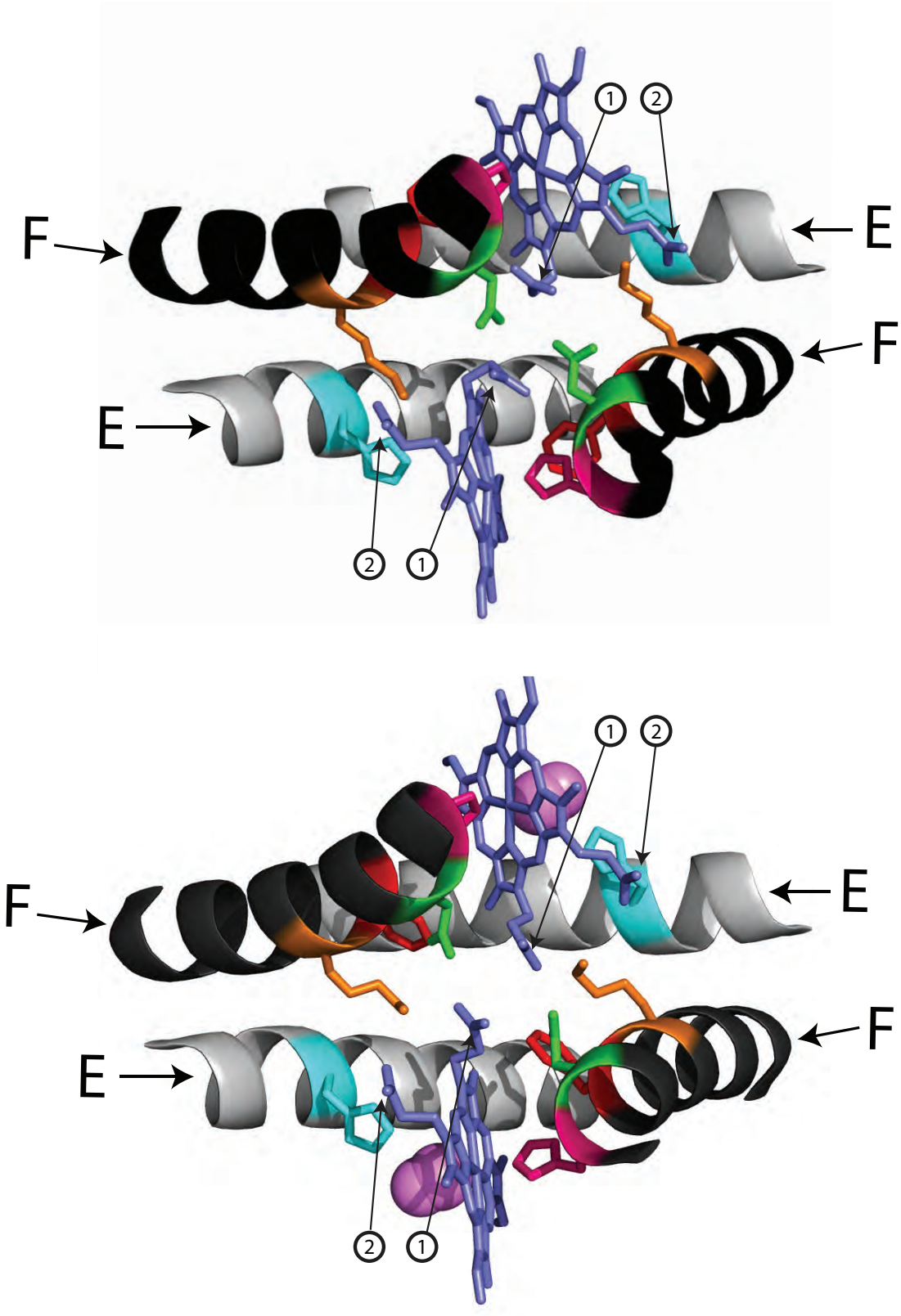


Figure 1.6

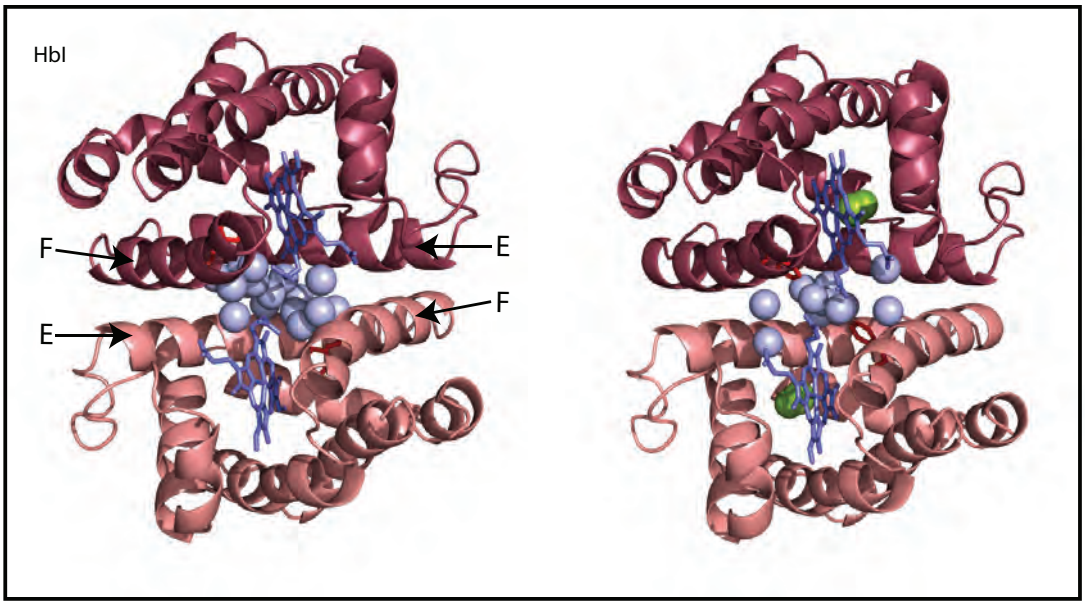
Figure 1.7: Crystal structures of the CO-bound HbI R and T conformations of HbI and F97Y. The HbI homodimer monomers are in dark and light purple, blue sticks represent the prosthetic heme, CO is green, residue F97 is shown in red, and water molecules are shown as light blue spheres. The top panel contains the structures of free and bound wild type HbI. The bottom panel contains the structures of free and bound F97Y HbI. The E and F interface helices are labeled in the top panel of deoxy wild type HbI.

Structures of unliganded and CO-HBI are generated from the PBD files 4SDH and 3SDH respectively. Structures of unliganded and CO-F97Y are generated from the PDB files 2AUP and 2AUO respectively.

References:

Royer, W.E., Jr. High-resolution crystallographic analysis of a co-operative dimeric hemoglobin. J Mol Biol 235, 657-81 (1994).

Knapp, J.E., Bonham, M.A., Gibson, Q.H., Nichols, J.C. & Royer, W.E., Jr. Residue F4 plays a key role in modulating oxygen affinity and cooperativity in Scapharca dimeric hemoglobin. *Biochemistry* **44**, 14419-30 (2005).



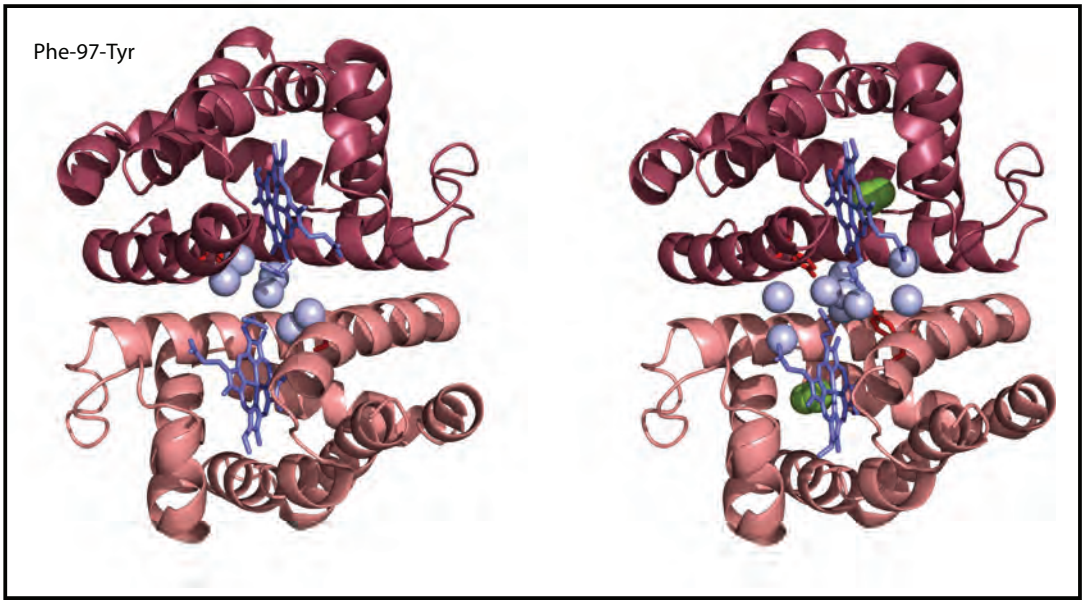


Figure 1.7

Although local structural rearrangements are important for HbI cooperativity they are not sufficient to describe the thermodynamics of cooperative binding. For HbI, the ligand binding affinity of the R state is 300-fold higher than the T state.(*69*) HbI binds oxygen with positive cooperativity and with a Hill coefficient of 1.5.(*11*) Ikeda-Saito et al. determined the thermodynamic properties of HbI cooperative ligand binding using a Van't Hoff analysis.(*11*) They measured the HbI oxygen binding equilibria in 0.1 M phosphate buffer at pH 7.8 for 10 μM – 100 μM HbI.(*11*) The cooperativity of HbI is unaffected by changes in the concentration of HbI or by changes in pH. However, the effect of salt or buffer concentration was not tested. The oxygen equilibria were measured at six different temperatures: 10.0, 15.1, 20.0, 25.1, 29.8, and 35.0 °C.(*11*) Assuming that HbI does not undergo a conformational change between 10 °C and 35.0 °C, they determine that the cooperativity of HbI is entropically driven. Consider **Table 1.1** which lists the thermodynamics of HbI cooperative ligand binding to oxygen at standard temperature and pressure.(*11*) In human hemoglobin cooperativity is enthalpically driven and entropically disfavored.(*73*) In contrast, entropy drives HbI positive cooperativity. Binding of the second oxygen is thermodynamically more favorable and is driven by a decrease in entropic restraint ($-T\Delta S$). This effect could result from a change in the conformational entropy of HbI and a change in the water surrounding HbI.(*74*)

Table 1.1: Thermodynamics of HbI Cooperative Binding to O_2 at 298 K.(*11*)

ΔG is the free energy change of the system, ΔH is the change in enthalpy, ΔS is the change in entropy, and T is the temperature in Kelvin.**(*11*)**

Table 1.1

First O_2	Second O_2	Energy Differences
ΔH_1 = - 11.4 kcal mol^{-1}	ΔH_2 = - 7.4 kcal mol^{-1}	$\Delta\Delta H_{2-1}$ = 4.0 kcal mol^{-1}
$-T\Delta S_1$ = 9.4 kcal mol^{-1}	$-T\Delta S_2$ = 4.1 kcal mol^{-1}	$-T\Delta\Delta S_{2-1}$ = - 5.3 kcal mol^{-1}
ΔG_1 = - 2.0 kcal mol^{-1}	ΔG_2 = - 3.3 kcal mol^{-1}	$\Delta\Delta G_{2-1}$ = - 1.3 kcal mol^{-1}

Although cooperativity in HbI is entropically driven, local structural rearrangements are essential for HbI cooperativity. In particular, residue F97 is critical for these rearrangments. Upon mutating F97 to tyrosine, cooperativity is reduced. The Hill coefficient of F97Y binding is 1.1.(*75*) In this case Y97 does not pack against His101 in the unliganded state and does not transition between the heme pocket but instead is located in the interface for both the R and T conformations.(*75*) William Royer determined the structure of both the R and T states of Hb and F97Y.(*70, 75*) He hypothesized that the hydroxyl group of tyrosine may either streically hinder packing in the heme pocket and or stabilize the tyrosine in the hydrophillic interface.(*75*)

Despite the importance of F97Y in local rearrangements, the F97Y HbI mutant does not dramatically alter the overall tertiary structure of the HbI R or T states. When aligned with residues 10-146 of wild type HbI the bound and deoxy forms differ by 0.07 and 0.18 root mean square deviations respectively.(*75*) The subunits rotation upon binding ligand is also similar to wild type at a 3.0° rotation. The R state assemblage differs from wild type, by a 0.15° rotation and the T state differs by 0.31° rotation from wild type.(*75*) The R and T state conformations of F97Y are shown in **Figure 1.7**. Although the crystallographic interface waters of both the R and T state of F97Y are different from wild type, the interface waters at the interface of the T state of F97Y are more different than the interface waters of the F97Y R state. The Asn100 and Lys96 links in wild type HbI are conserved in the T state of F97Y. In the R state of F97Y, the water-mediated link by Lys96 is lost, but otherwise the connective network is maintained.

Several additional mutants of HbI have been designed to probe the role of local structural rearrangements at the homodimeric interface and the contribution of water to wild type cooperativity. These mutants have been characterized by flash photolysis and structurally through x-ray crystallography, **Table 1.2.**

Table 1.2: HbI Mutants. Flash photolysis is used to measure kinetics near physiological conditions as reported.(*71, 72, 76-80*) These mutants have not yet been characterized thermodynamically. The crystallographic waters in each of the bound and unliganded mutant structures differs from wild type. The high affinity mimic adopts the conformation of the bound state in the presence or absence of ligand. The low affinity mimic adopts the conformation of the unliganded state in the presence or absence of ligand. Those mutants that have high affinity favor undergo a conformational change upon binding ligand, however the interface crystallographic waters of both the bound and unliganded states is most similar to the crystallographic water molecules observed in the bound interface of wt HbI.

Table 1.2

Variants of HbI	Ligand	Hill Coefficient	Qualities
Wild type	O_2	1.5 ± 0.05	Wild type
F97Y	O_2	1.1 ± 0.05	High-affinity Mimic
I114F	O_2	1.04 ± 0.04	Low-affinity Mimic
T72I	O_2	~1.2	High-affinity Favor
T72V	O_2	~1.7	High-affinity Favor

The F97Y and I114F mutants do not undergo wild type like local structural rearrangements at the interface. For both the R and T states, the F97Y interface residues resemble the R state. For both R and T states, I114F interface residues resemble the T state. T72I and T72V mutants disrupt the water network between HbI subunits and alter cooperativity: T72I diminishes cooperativity and T72V enhances cooperativity. The cooperativity of each of these mutants is different from wild type cooperativity. This suggests that local structural rearrangements and changes in water dynamics at the interface are linked and are critical in HbI cooperativity.

Cooperativity can arise through changes in 1) protein and/or solvent structure, 2) protein and/or water dynamics. Crystallographic studies of HbI and the F97Y mutant have revealed the structural contributions to cooperativity and have also implicated a role of the interfacial water in HbI cooperativity. The thermodynamics of cooperative binding of oxygen by HbI reveal that entropy drives HbI cooperativity. A change in entropy can arise from changes in protein or water dynamics.

SCOPE OF THESIS

Currently the importance and relevance of protein dynamics for cooperative binding and molecular recognition is in its infancy. Typically hemoglobin cooperativity is attributed to a change in conformation. This thesis demonstrates how both structure and dynamics are important for cooperativity.

The primary aim of this thesis is to understand how changes in HbI flexibility contribute to cooperativity and how changes in dynamics are coupled with structural changes of the protein and the interfacial water. HbI becomes more flexible upon binding CO, and notably the E and F interface helices of HbI undergo the largest change in flexibility. Also, the conformational entropy of HbI contributes to cooperative binding. The HbI mutant F97Y has significantly reduced cooperativity and we observe that F97Y does not undergo wild type-like changes in flexibility upon binding ligand. This is a proof of principle that the backbone fast time-scale dynamics of HbI contribute to cooperative binding.

The second aim of this thesis is to study the mechanism of adenine recognition by TIS11d. This research demonstrates that TIS11d binds three high affinity RNA targets. The chemical shifts of backbone amides of TIS11d bound to RNA suggest that TIS11d binds each of these RNA sequences differently.

The study of Hb and TIS11d furthered our understanding of protein dynamics and can be applied to other systems as highlighted by my collaborative projects in the appendices.

REFERENCES:

1. Gellman, S. H. (1997) Introduction: Molecular Recognition, *Chem Rev 97*, 1231-1232.
2. Lodish H, B. A., Zipursky SL et al. (2000) *Molecular Cell Biology*, 4th ed.
3. Tzeng, S. R., and Kalodimos, C. G. (2009) Dynamic activation of an allosteric regulatory protein, *Nature 462*, 368-372.
4. Eaton, W. A., Henry, E. R., Hofrichter, J., Bettati, S., Viappiani, C., and Mozzarelli, A. (2007) Evolution of allosteric models for hemoglobin, *IUBMB Life 59*, 586-599.
5. Cooper, A., and Dryden, D. T. (1984) Allostery without conformational change. A plausible model, *Eur Biophys J 11*, 103-109.
6. Popovych, N., Sun, S., Ebright, R. H., and Kalodimos, C. G. (2006) Dynamically driven protein allostery, *Nat Struct Mol Biol 13*, 831-838.
7. Petit, C. M., Zhang, J., Sapienza, P. J., Fuentes, E. J., and Lee, A. L. (2009) Hidden dynamic allostery in a PDZ domain, *Proc Natl Acad Sci U S A 106*, 18249-18254.
8. Wang, J. T., Yang, P. W., Lee, C. P., Han, C. H., Tsai, C. H., and Chen, M. R. (2005) Detection of Epstein-Barr virus BGLF4 protein kinase in virus replication compartments and virus particles, *J Gen Virol 86*, 3215-3225.
9. Lutz, S., Lichter, J., and Liu, L. (2007) Exploiting temperature-dependent substrate promiscuity for nucleoside analogue activation by thymidine kinase from Thermotoga maritima, *J Am Chem Soc 129*, 8714-8715.
10. Sandler, H., and Stoecklin, G. (2008) Control of mRNA decay by phosphorylation of tristetraprolin, *Biochem Soc Trans 36*, 491-496.
11. Ikeda-Saito, M., Yonetani, T., Chiancone, E., Ascoli, F., Verzili, D., and Antonini, E. (1983) Thermodynamic properties of oxygen equilibria of dimeric and tetrameric hemoglobins from Scapharca inaequivalvis, *J Mol Biol 170*, 1009-1018.
12. Parkison, C., Ashizawa, K., McPhie, P., Lin, K. H., and Cheng, S. Y. (1991) The monomer of pyruvate kinase, subtype M1, is both a kinase and a cytosolic thyroid hormone binding protein, *Biochem Biophys Res Commun 179*, 668-674.
13. Nobeli, I., Favia, A. D., and Thornton, J. M. (2009) Protein promiscuity and its implications for biotechnology, *Nat Biotechnol 27*, 157-167.
14. Boehr, D. D., McElheny, D., Dyson, H. J., and Wright, P. E. (2006) The dynamic energy landscape of dihydrofolate reductase catalysis, *Science 313*, 1638-1642.
15. Lange, O. F., Lakomek, N. A., Fares, C., Schroder, G. F., Walter, K. F., Becker, S., Meiler, J., Grubmuller, H., Griesinger, C., and de Groot, B. L. (2008) Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution, *Science 320*, 1471-1475.
16. Frederick, K. K., Kranz, J. K., and Wand, A. J. (2006) Characterization of the backbone and side chain dynamics of the CaM-CaMKIp complex reveals microscopic contributions to protein conformational entropy, *Biochemistry 45*, 9841-9848.

17. Frederick, K. K., Marlow, M. S., Valentine, K. G., and Wand, A. J. (2007) Conformational entropy in molecular recognition by proteins, *Nature 448*, 325-329.
18. Marlow, M. S., Dogan, J., Frederick, K. K., Valentine, K. G., and Wand, A. J. (2010) The role of conformational entropy in molecular recognition by calmodulin, *Nat Chem Biol 6*, 352-358.
19. Eaton, W. A., Henry, E. R., Hofrichter, J., and Mozzarelli, A. (1999) Is cooperative oxygen binding by hemoglobin really understood?, *Nat Struct Biol 6*, 351-358.
20. Cui, Q., and Karplus, M. (2008) Allostery and cooperativity revisited, *Protein Sci 17*, 1295-1307.
21. Edsall, J. T. (1980) Hemoglobin and the origins of the concept of allosterism, *Fed Proc 39*, 226-235.
22. Bisswanger, H. (2008) *Enzyme Kinetics: Principles and Methods*, 2 ed., John Wiley & Sons.
23. Adair, G. S. (1925) A critical study of the direct method of measuring osmotic pressure of hemoglobin., *Proc. R. Soc. London Ser. A*, 627-637.
24. Bezkorovainy, A. (1996) Concise Biochemistry, (Rafelson, M. E., Ed.), Dekker.
25. Reginald Garrett, C. M. G. (2010) *Biochemistry*, 4 ed., Cengage Learning.
26. Pauling, L. (1935) The Oxygen Equilibrium of Hemoglobin and Its Structural Interpretation, *Proc Natl Acad Sci U S A 21*, 186-191.
27. Perutz, M. F., Bolton, W., Diamond, R., Muirhead, H., and Watson, H. C. (1964) Structure of Haemoglobin. An X-Ray Examination of Reduced Horse Haemoglobin, *Nature 203*, 687-690.
28. Monod, J., Wyman, J., and Changeux, J. P. (1965) On the Nature of Allosteric Transitions: A Plausible Model, *J Mol Biol 12*, 88-118.
29. Perutz, M. F. (1970) Stereochemistry of cooperative effects in haemoglobin, *Nature 228*, 726-739.
30. Szabo, A., and Karplus, M. (1972) A mathematical model for structure-function relations in hemoglobin, *J Mol Biol 72*, 163-197.
31. Rodgers, D. W., Crepeau, R. H., and Edelstein, S. J. (1987) Pairings and polarities of the 14 strands in sickle cell hemoglobin fibers, *Proc Natl Acad Sci U S A 84*, 6157-6161.
32. Edelstein, S. J. (1971) Extensions of the allosteric model for haemoglobin, *Nature 230*, 224-227.
33. Cooper, A. (1976) Thermodynamic fluctuations in protein molecules, *Proc Natl Acad Sci U S A 73*, 2740-2741.
34. Stone, M. J. (2001) NMR relaxation studies of the role of conformational entropy in protein stability and ligand binding, *Acc Chem Res 34*, 379-388.
35. Wand, A. J. (2001) Dynamic activation of protein function: a view emerging from NMR spectroscopy, *Nat Struct Biol 8*, 926-931.
36. Igumenova, T. I., Frederick, K. K., and Wand, A. J. (2006) Characterization of the fast dynamics of protein amino acid side chains using NMR relaxation in solution, *Chem Rev 106*, 1672-1699.

37. Yang, D., and Kay, L. E. (1996) Contributions to conformational entropy arising from bond vector fluctuations measured from NMR-derived order parameters: application to protein folding, *J Mol Biol 263*, 369-382.
38. Henzler-Wildman, K., and Kern, D. (2007) Dynamic personalities of proteins, *Nature 450*, 964-972.
39. Palmer, A. G., 3rd. (2004) NMR characterization of the dynamics of biomacromolecules, *Chem Rev 104*, 3623-3640.
40. Lipari, G., and Szabo, A. (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity, *J. Am. Chem. Soc. 104*, 4546-4559.
41. Ishima, R., and Torchia, D. A. (2000) Protein dynamics from NMR, *Nat Struct Biol 7*, 740-743.
42. Sapienza, P. J., and Lee, A. L. (2010) Using NMR to study fast dynamics in proteins: methods and applications, *Curr Opin Pharmacol 10*, 723-730.
43. Kalodimos, C. G., Biris, N., Bonvin, A. M., Levandoski, M. M., Guennuegues, M., Boelens, R., and Kaptein, R. (2004) Structure and flexibility adaptation in nonspecific and specific protein-DNA complexes, *Science 305*, 386-389.
44. Freedberg, D. I., Ishima, R., Jacob, J., Wang, Y. X., Kustanovich, I., Louis, J. M., and Torchia, D. A. (2002) Rapid structural fluctuations of the free HIV protease flaps in solution: relationship to crystal structures and comparison with predictions of dynamics calculations, *Protein Sci 11*, 221-232.
45. Thorpe, I. F., and Brooks, C. L., 3rd. (2007) Molecular evolution of affinity and flexibility in the immune system, *Proc Natl Acad Sci U S A 104*, 8821-8826.
46. Jarymowycz, V. A., and Stone, M. J. (2008) Remote changes in the dynamics of the phosphotyrosine-binding domain of insulin receptor substrate-1 induced by phosphopeptide binding, *Biochemistry 47*, 13371-13382.
47. Barbato, G., Ikura, M., Kay, L. E., Pastor, R. W., and Bax, A. (1992) Backbone dynamics of calmodulin studied by 15N relaxation using inverse detected two-dimensional NMR spectroscopy: the central helix is flexible, *Biochemistry 31*, 5269-5278.
48. Song, X. J., Yuan, Y., Simplaceanu, V., Sahu, S. C., Ho, N. T., and Ho, C. (2007) A comparative NMR study of the polypeptide backbone dynamics of hemoglobin in the deoxy and carbonmonoxy forms, *Biochemistry 46*, 6795-6803.
49. Williams, J. C., and McDermott, A. E. (1995) Dynamics of the flexible loop of triosephosphate isomerase: the loop motion is not ligand gated, *Biochemistry 34*, 8309-8319.
50. Eisenmesser, E. Z., Millet, O., Labeikovsky, W., Korzhnev, D. M., Wolf-Watz, M., Bosco, D. A., Skalicky, J. J., Kay, L. E., and Kern, D. (2005) Intrinsic dynamics of an enzyme underlies catalysis, *Nature 438*, 117-121.
51. Eisenmesser, E. Z., Bosco, D. A., Akke, M., and Kern, D. (2002) Enzyme dynamics during catalysis, *Science 295*, 1520-1523.

52. Baldwin, A. J., and Kay, L. E. (2009) NMR spectroscopy brings invisible protein states into focus, *Nat Chem Biol 5*, 808-814.
53. Bruschweiler, S., Schanda, P., Kloiber, K., Brutscher, B., Kontaxis, G., Konrat, R., and Tollinger, M. (2009) Direct observation of the dynamic process underlying allosteric signal transmission, *J Am Chem Soc 131*, 3063-3068.
54. Loria, J. P., Berlow, R. B., and Watt, E. D. (2008) Characterization of enzyme motions by solution NMR relaxation dispersion, *Acc Chem Res 41*, 214-221.
55. Cole, R., and Loria, J. P. (2002) Evidence for flexibility in the function of ribonuclease A, *Biochemistry 41*, 6072-6081.
56. Beach, H., Cole, R., Gill, M. L., and Loria, J. P. (2005) Conservation of mus-ms enzyme motions in the apo- and substrate-mimicked state, *J Am Chem Soc 127*, 9167-9176.
57. Goto, N. K., Zor, T., Martinez-Yamout, M., Dyson, H. J., and Wright, P. E. (2002) Cooperativity in transcription factor binding to the coactivator CREB-binding protein (CBP). The mixed lineage leukemia protein (MLL) activation domain binds to an allosteric site on the KIX domain, *J Biol Chem 277*, 43168-43174.
58. Stumpo, D. J., Broxmeyer, H. E., Ward, T., Cooper, S., Hangoc, G., Chung, Y. J., Shelley, W. C., Richfield, E. K., Ray, M. K., Yoder, M. C., Aplan, P. D., and Blackshear, P. J. (2009) Targeted disruption of Zfp36l2, encoding a CCCH tandem zinc finger RNA-binding protein, results in defective hematopoiesis, *Blood 114*, 2401-2410.
59. Iwanaga, E., Nanri, T., Mitsuya, H., and Asou, N. (2011) Mutation in the RNA binding protein TIS11D/ZFP36L2 is associated with the pathogenesis of acute leukemia, *Int J Oncol 38*, 25-31.
60. Frederick, E. D., Ramos, S. B., and Blackshear, P. J. (2008) A unique C-terminal repeat domain maintains the cytosolic localization of the placenta-specific tristetraprolin family member ZFP36L3, *J Biol Chem 283*, 14792-14800.
61. Baou, M., Jewell, A., and Murphy, J. J. (2009) TIS11 family proteins and their roles in posttranscriptional gene regulation, *J Biomed Biotechnol 2009*, 634520.
62. Hodson, D. J., Janas, M. L., Galloway, A., Bell, S. E., Andrews, S., Li, C. M., Pannell, R., Siebel, C. W., MacDonald, H. R., De Keersmaecker, K., Ferrando, A. A., Grutz, G., and Turner, M. (2010) Deletion of the RNA-binding proteins ZFP36L1 and ZFP36L2 leads to perturbed thymic development and T lymphoblastic leukemia, *Nat Immunol 11*, 717-724.
63. Johnson, B. A., and Blackwell, T. K. (2002) Multiple tristetraprolin sequence domains required to induce apoptosis and modulate responses to TNFalpha through distinct pathways, *Oncogene 21*, 4237-4246.
64. Blackshear, P. J., Lai, W. S., Kennington, E. A., Brewer, G., Wilson, G. M., Guan, X., and Zhou, P. (2003) Characteristics of the interaction of a synthetic human tristetraprolin tandem zinc finger peptide with AU-rich element-containing RNA substrates, *J Biol Chem 278*, 19947-19955.

65. Brewer, B. Y., Malicka, J., Blackshear, P. J., and Wilson, G. M. (2004) RNA sequence elements required for high affinity binding by the zinc finger domain of tristetraprolin: conformational changes coupled to the bipartite nature of Au-rich MRNA-destabilizing motifs, *J Biol Chem 279*, 27870-27877.
66. Hudson, B. P., Martinez-Yamout, M. A., Dyson, H. J., and Wright, P. E. (2004) Recognition of the mRNA AU-rich element by the zinc finger domain of TIS11d, *Nat Struct Mol Biol 11*, 257-264.
67. Lai, W. S., Carballo, E., Thorn, J. M., Kennington, E. A., and Blackshear, P. J. (2000) Interactions of CCCH zinc finger proteins with mRNA. Binding of tristetraprolin-related zinc finger proteins to Au-rich elements and destabilization of mRNA, *J Biol Chem 275*, 17827-17837.
68. Morgan, B. R., and Massi, F. (2010) A computational study of RNA binding and specificity in the tandem zinc finger domain of TIS11d, *Protein Sci 19*, 1222-1234.
69. Knapp, J. E., Pahl, R., Srajer, V., and Royer, W. E., Jr. (2006) Allosteric action in real time: time-resolved crystallographic studies of a cooperative dimeric hemoglobin, *Proc Natl Acad Sci U S A 103*, 7649-7654.
70. Royer, W. E., Jr., Hendrickson, W. A., and Chiancone, E. (1990) Structural transitions upon ligand binding in a cooperative dimeric hemoglobin, *Science 249*, 518-521.
71. Royer, W. E., Jr. (1994) High-resolution crystallographic analysis of a co-operative dimeric hemoglobin, *J Mol Biol 235*, 657-681.
72. Royer, W. E., Jr., Pardanani, A., Gibson, Q. H., Peterson, E. S., and Friedman, J. M. (1996) Ordered water molecules as key allosteric mediators in a cooperative dimeric hemoglobin, *Proc Natl Acad Sci U S A 93*, 14526-14531.
73. Imai, K. (1979) Thermodynamic aspects of the co-operativity in four-step oxygenation equilibria of haemoglobin, *J Mol Biol 133*, 233-247.
74. Kern, D., and Zuiderweg, E. R. (2003) The role of dynamics in allosteric regulation, *Curr Opin Struct Biol 13*, 748-757.
75. Knapp, J. E., Bonham, M. A., Gibson, Q. H., Nichols, J. C., and Royer, W. E., Jr. (2005) Residue F4 plays a key role in modulating oxygen affinity and cooperativity in Scapharca dimeric hemoglobin, *Biochemistry 44*, 14419-14430.
76. Gambacurta, A., Piro, M. C., Coletta, M., Clementi, M. E., Polizio, F., Desideri, A., Santucci, R., and Ascoli, F. (1995) A single mutation (Thr72-->Ile) at the subunit interface is crucial for the functional properties of the homodimeric co-operative haemoglobin from Scapharca inaequivalvis, *J Mol Biol 248*, 910-917.
77. Knapp, J. E., and Royer, W. E., Jr. (2003) Ligand-linked structural transitions in crystals of a cooperative dimeric hemoglobin, *Biochemistry 42*, 4640-4647.
78. Pardanani, A., Gibson, Q. H., Colotti, G., and Royer, W. E., Jr. (1997) Mutation of residue Phe97 to Leu disrupts the central allosteric pathway in Scapharca dimeric hemoglobin, *J Biol Chem 272*, 13171-13179.

79. Pardanani, A., Gambacurta, A., Ascoli, F., and Royer, W. E., Jr. (1998) Mutational destabilization of the critical interface water cluster in Scapharca dimeric hemoglobin: structural basis for altered allosteric activity, *J Mol Biol 284*, 729-739.
80. Knapp, J. E., Gibson, Q. H., Cushing, L., and Royer, W. E., Jr. (2001) Restricting the ligand-linked heme movement in Scapharca dimeric hemoglobin reveals tight coupling between distal and proximal contributions to cooperativity, *Biochemistry 40*, 14795-14805.

CHAPTER II

INSIGHTS INTO THE ALLOSTERIC MECHANISM OF SCAPHARCA DIMERIC HEMOGLOBIN

AUTHOR CONTRIBUTIONS

This chapter is the main focus of my thesis and was done under the direction and with the help of my advisor Francesca Massi. William Royer provided materials, aided in the expression and purification of HbI and F97Y, and aided in the preparation of unliganded protein.

ABSTRACT

Upon ligand binding, the homodimeric hemoglobin of Scapharca inaequivalvis (HbI) undergoes a transition between a low affinity and a high affinity state characterized by a 300-fold increased affinity for oxygen with a Hill coefficient of 1.5.(*1*) A van't Hoff analysis of HbI indicates that cooperative binding of HbI to oxygen is due to a change in entropy.(*2*) The change in entropy of the system observed upon ligand binding may arise from changes in the protein, the ligand or in the water of the system. The focus of this study is on determining the contribution of the change in entropy of the protein backbone to HbI cooperative binding. NMR relaxation techniques reveal that the fast internal motions of HbI contributes to the cooperative binding to carbon monoxide in two ways, 1) by contributing favorably to the free energy of the system and 2) by participating in the cooperative mechanism at the HbI subunit interface. As a proof of principle we measured the fast internal motions of the weakly cooperative HbI mutant, F97Y. In this case, we observe that the fast backbone motions of F97Y are disrupted throughout the mutant protein and at the F97Y homodimeric interface. The protein conformational entropy change associated to CO binding estimated for F97Y HbI is of opposite sign than that estimated of wild type HbI, hence it does not contribute favorably to the very low cooperativity of this mutant protein. The results of this study suggest that interface flexibility and backbone conformational entropy of HbI participate and are important for the cooperative mechanism of carbon monoxide binding. In future studies, MD simulations and NMR relaxation techniques will be used to measure the entropic

contributions of the water and of HbI side-chains in HbI cooperativity. These studies will reveal the role of protein and water and how their coupling contributes to HbI cooperativity.

INTRODUCTION

Allosteric regulation is an essential function of many proteins that control a variety of different processes such as catalysis, signal transduction, and gene regulation.(*3-5*) Structural rearrangements have historically been considered the main means of communication between different parts of a protein. Recent studies have highlighted the importance of changes in protein flexibility as an effective way to mediate allosteric communication across a protein.(*6-8*) The dimeric hemoglobin from *Scapharca inaequivalvis* (a blood clam) (HbI) is the simplest possible allosteric system, with cooperative ligand binding between two identical subunits. HbI is unusual in this regard, as all other known allosteric hemoglobins invariably incorporate unlike subunits. Because of its simplicity, HbI is an excellent system for directly exploring key issues of allostery. HbI binds oxygen and CO cooperatively (the Hill coefficient, n, is equal to 1.5 for oxygen binding) at two chemically identical sites with no modulation of ligand affinity by non-heme ligands.(*9*) The dimer is formed by apposition of the E and F helices,(*1, 10*) bringing the heme groups into close proximity and permitting more direct communication in HbI than in mammalian hemoglobins, **Figure 2.1**. This assembly, which we term an “EF dimer”, has now been found in hemoglobins from echinoderms (*11*) and megadalton annelid hemoglobin complexes.(*12*) Ligand binding in HbI is associated with localized structural transitions (**Fig. 2.1**)(*13*) characterize by only a slight change in quaternary structure (3.3º subunit rotation) but with striking tertiary changes at the dimeric interface, including the movement of Phe 97 (F4) from the heme pocket into the interface.(*1*)

Figure 2.1: Local conformational changes at the homodimeric interface of HbI. The HbI interface of the unliganded low affinity state is in (A) and the CO bound high affinity state is in (B). The E helices are gray and the F helices are black, hemes are blue and CO is magenta. The propionate groups of the hemes are labeled 1 and 2. Important residues side-chains are colored. The proximal histidine, His101 is pink. The distal histidine, His69, is cyan. The interface residues Lys96, Phe97 and Asn100 are orange, red and green respectively.

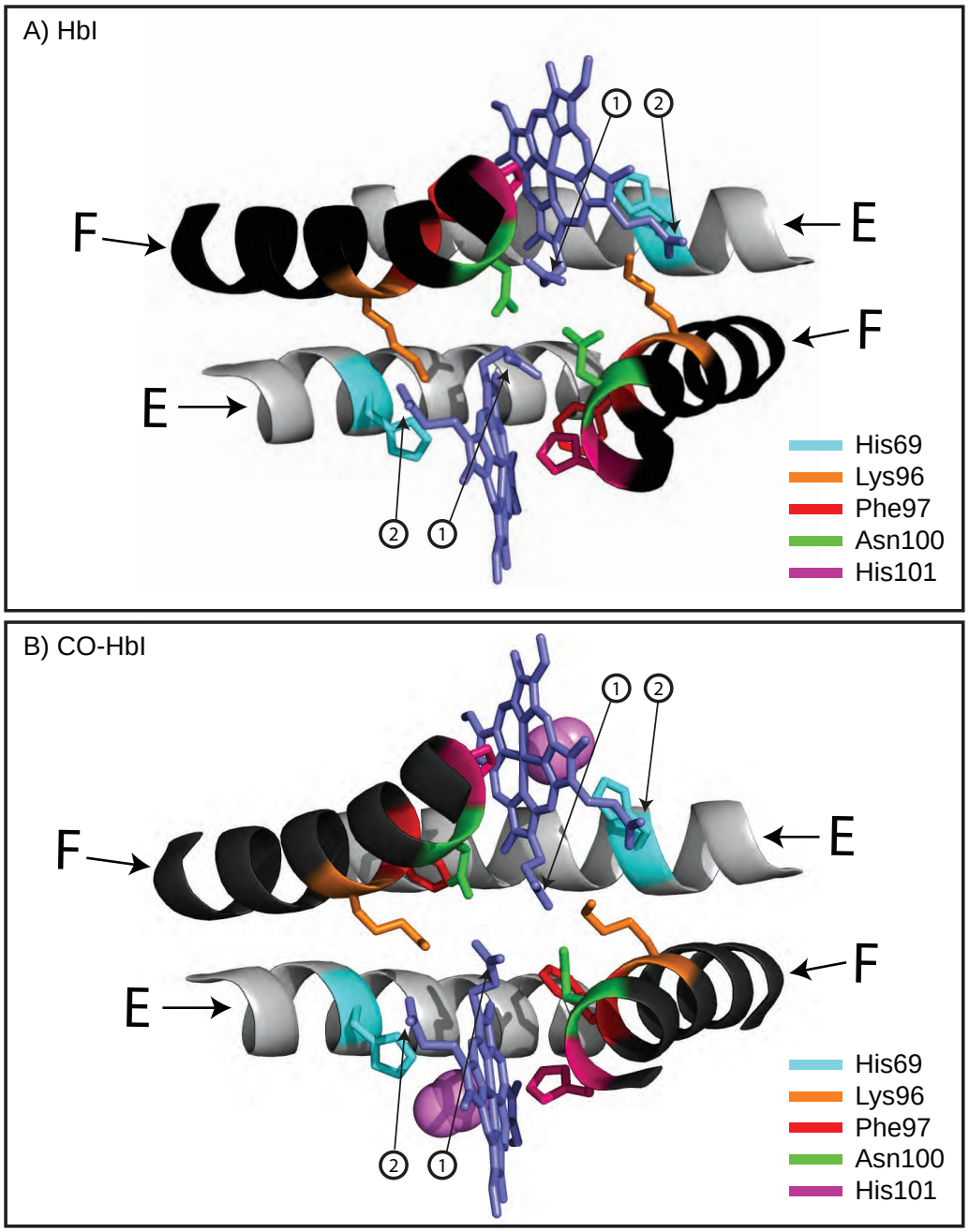

A) HbI
F
E
F
E
1
2
2
1
His69
Lys96
Phe97
Asn100
His101
B) CO-HbI
F
E
F
E
1
2
2
1
His69
Lys96
Phe97
Asn100
His101

Figure 2.1

(Individual residues of proteins that adopt the globin fold are named according to the helical position in sperm whale myoglobin, the first globin structure determined. Residue Phe 97 of HbI is at the fourth position of the F helix of the globin fold.) The side chain motion of Phe 97 disrupts a cluster of well-ordered water molecules. As a result, the observed water molecules in the crystal structures are more abundant and more ordered in the unliganded interface than the liganded form.(*1*)

Thermodynamic equilibrium studies of oxygen binding to HbI have shown that cooperativity in HbI is an entropically driven effect.(*2*) In particular, the differences in free energy, enthalpy and entropy between the first and second oxygen binding event are $\Delta\Delta G$ = -1.3 kcal/mol, $\Delta\Delta H$ = 4.0 kcal/mol and $-T(\Delta\Delta S)$ = -5.3 kcal/mol at 298 K, respectively. Hence, in HbI, binding of the second oxygen molecule is made thermodynamically more favorable by a less negative ΔS. This effect could result from a change in the conformational entropy of HbI or that of the water surrounding HbI when oxygen binding occurs. The importance of changes to protein dynamics in ligand binding processes has been recently highlighted from NMR relaxation studies.(*6, 14-19*)

HbI is unstable when bound to oxygen because oxygen oxidizes the iron of the HbI heme, but HbI is stable when bound to carbon monoxide. Although we know that HbI binds carbon monoxide with higher affinity than oxygen we hypothesize that the entropic contribution to cooperativity is the same. This is a reasonable hypothesis because the crystallographic structures of HbI bound to CO or O_2 are very similar; the average RMSD between CO-HbI and O_2-HbI is

less than 0.1.(*10, 20*)

All of this evidence points to a combination of the structure and dynamics of HbI and of the interfacial water in allosteric communication between the two binding sites. In order to fully understand the cooperative mechanism of HbI, each of these contributions must be measured. In this study, we investigated how changes in backbone HbI flexibility contribute to cooperativity using nuclear magnetic resonance (NMR) spectroscopy. To investigate how changes in protein dynamics are coupled to and correlated with structural changes upon ligand binding we compared the internal backbone dynamics of wild type HbI with those of the mutant HbI, F97Y. For the mutant protein the side chain at position 97 is locked at the dimeric interface, mimicking the high affinity conformation.(*21*) In F97Y, the structural transition associated with ligand binding, from the low affinity to the high affinity state is completely restricted at the interface.

RESULTS

HbI is a symmetric dimer in solution

HbI is a symmetric dimer in solution in both the liganded and unliganded states, as confirmed by the observation of only one set of signals corresponding to a monomeric subunit in two-dimensional (2D) $^{1}H–^{15}N$ correlation spectra. **Figure 2.2** shows the different spectra acquired for CO-bound and unbound HbI and F97Y. The chemical shift differences between these two spectra reflect the structural changes undergone by the protein upon CO binding. Even the subtle ligand-linked structural changes in HbI detected by x-ray crystallography (between 0.5 and 1 Å) result in significant differences in chemical shift values. In HbI the structural change, although relatively localized, affects the chemical environment of the entire dimeric interface because ligand binding is coupled with a change in quaternary structure (3.3º subunit rotation) (*22*) in addition to striking tertiary changes at the core of the interface.

Resonance assignments

The backbone assignment of unliganded and CO-liganded HbI and of unliganded F97Y HbI were determined using a three-dimensional triple resonance [^{1}H, ^{15}N]-TROSY experiments including, HNCO, HN(CA)CO, HNCA, HN(CO)CA, HNCACB, and CBCA(CO)NH (*23, 24*) collected at 14.1 T and at 298 K using uniformly labeled $^{13}C/^{15}N$ samples.

Figure 2.2: Amide Chemical shifts of HbI. This is an overlap of the [^{1}H, ^{15}N]-TROSY spectrum of (A) CO-HbI and unliganded HbI (B) and CO-F97Y and unliganded F97Y. In both A and B the CO bound conformations are red and unliganded conformations are blue. Unliganded peaks that undergo moderate chemical shifts upon binding ligand are labeled in both A and B. Residue D20 does not undergo a chemical shift change in each of the four spectra. Residue L54 is too low intensity to be assigned in B.

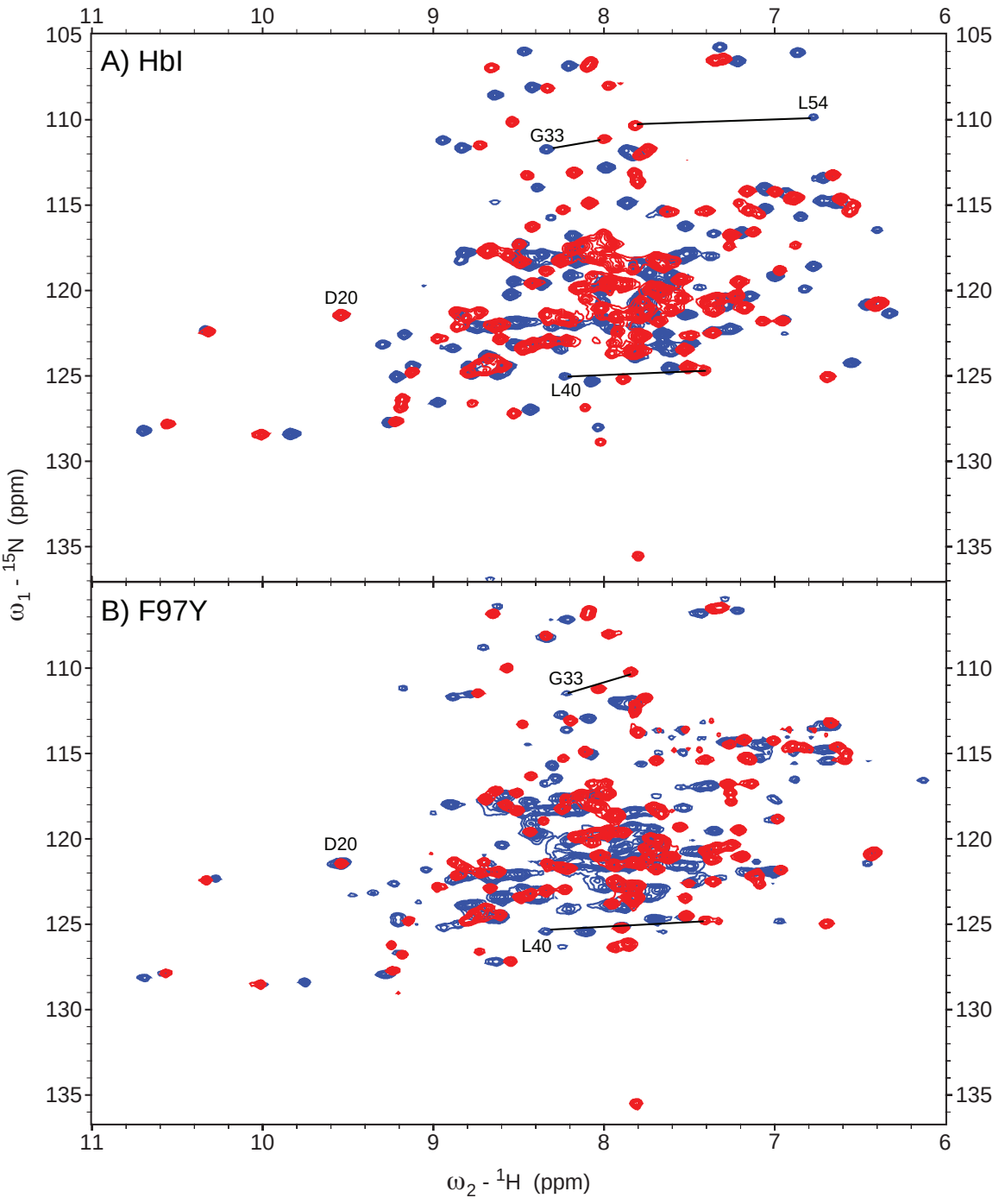

A) HbI
L54
G33
D20
L40
B) F97Y
G33
D20
L40
11
10
9
8
7
6
105
110
115
120
125
130
135
ω1 - 15N (ppm)
ω2 - 1H (ppm)

Figure 2.2

Assignment of the CO-bound F97Y mutant HbI was obtained by comparison with the spectrum of CO-bound WT HbI supplemented by the use of ^{1}H–^{1}H NOESY and HMQC-NOESY-HSQC experiments. Each subunit of HbI consists of 146 residues. In addition to the three proline residues that do not give rise to a signal in a 2D ^{1}H–^{15}N correlation spectrum, the following residues could not be assigned due to overlap or lack of signal; CO-HbI: 2, 12, 56, 81, and 130; unliganded HbI 2, 70, 71,73- 75, 77, 97, 99-103 105, 106, 111-114; CO-F97Y HbI: 2, 12, 56, 81, 99, 117, 130; unliganded F97Y HbI: 2, 12, 36, 42, 50, 51, 53, 56, 65, 69-77, 97-107, 113, 117, 130.

Effect of the paramagnetic Fe^{2+} in the unliganded state on the relaxation rates

The presence of a paramagnetic center in the unliganded state of HbI dramatically affects the relaxation properties of the nearby nuclei. Because of the increased transverse relaxation rates the signal of many of the residues on helix E and helix F is too low for detection in the unliganded state. The paramagnetic contribution, due to interaction with the metal ion, to the transverse and longitudinal relaxation rates is a function of the distance from the Fe^{2+} ion with $1/r^{6}$ dependence. As a consequence, the paramagnetic effect is the dominant contribution to R_1, R_2, and NOE for nuclei proximal to the Fe center, but will be negligible for nuclei that are >15 Å away from the paramagnetic center. Residues with a non-negligible paramagnetic contribution to the relaxation rates can be easily identified from a plot of R_1 and R_2 as a function of $1/r_{N\text{-}Fe}^{6}$, where $r_{N\text{-}}$

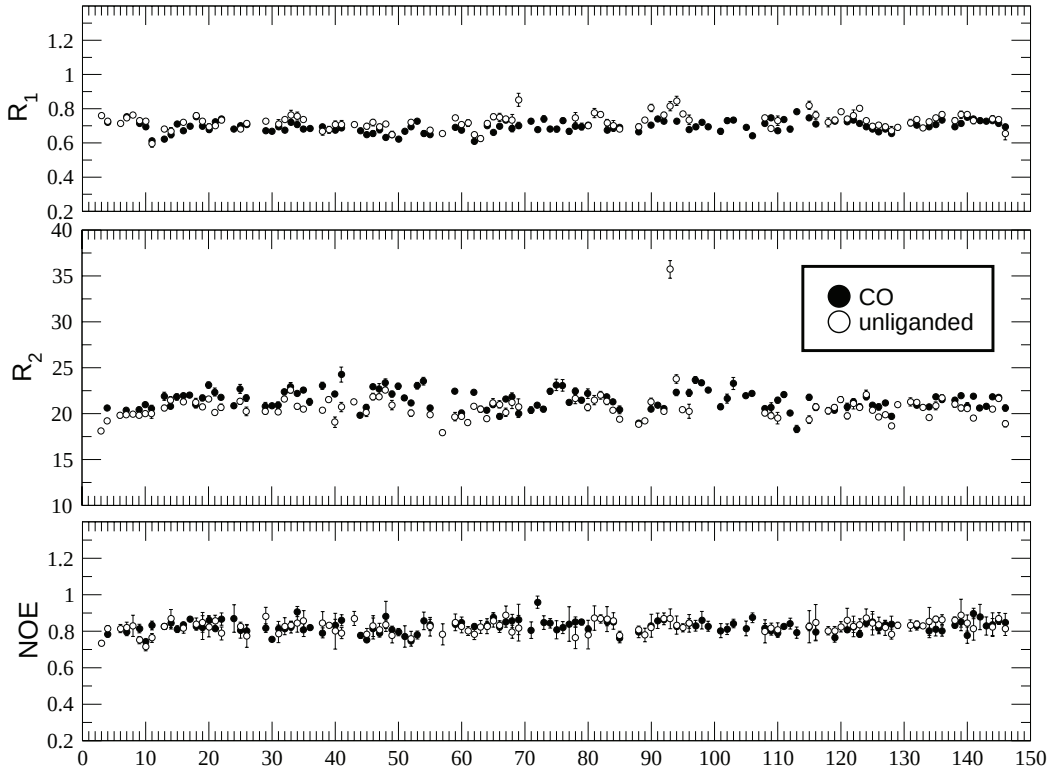
$_{Fe}$ is the distance between the amide N and the Fe^{2+} ion. The only residue that we identify as susceptible to the paramagnetic interaction with the Fe^{2+} ion was A98: this residue was excluded from the model free analysis.(*25*)

Internal dynamics of HbI

NMR ^{15}N longitudinal, R_1, and transverse, R_2, relaxation rate constants, and {^{1}H}-^{15}N nuclear Overhauser enhancement (NOE) (*26-28*) were measured for all the spectrally resolved amide backbone groups in WT and F97Y HbI in both the unliganded and unliganded states at 14.1 T and 298 K, (**Figure 2.3 A**). Relaxation parameters could not be accurately determined for the following residues because of spectral overlap: 3, 5, 6, 8, 23, 27, 28, 37, 39, 42, 43, 57, 58, 61, 63, 70, 82, 86, 89, 93, 95, 100, 104, 107, 114, 118, 120, 129, 131, 133, 137 in CO-HbI, 5, 12, 15, 17, 23, 24, 27, 28, 30, 36, 37, 42, 44, 53, 56, 58, 72, 91, 104, 107, 137 in unliganded HbI, 16, 21, 23, 27, 28, 29, 31, 39, 42, 43, 44, 49, 55, 57-60, 63, 69, 71, 72, 78, 83, 86, 89, 92-94, 97, 103, 104, 107, 114, 115, 118-120, 124, 131, 137 in CO-F97Y HbI and 5, 10, 11, 16, 18, 19, 21, 23, 27, 29, 39, 48, 62, 63, 65, 66, 78, 80-82, 89, 93, 114, 119-121, 123, 127, 129, 136, 141 in unliganded F97Y HbI. In addition, the following residues were excluded from the model free analysis due to the large relaxation parameter errors resulting from low signal intensities in one or more of the relaxation experiments, 50, 51, 54, 76, 79, 130, 142, 143 in unliganded HbI, 96, 101, 102, 142 in CO-F97Y and 37, 40, 47, 54, 79, 96, 112, 115, 142, 146 in unliganded F97Y.

Figure 2.3: Amide relaxation rates of HbI and Phe-97-Tyr. Relaxation rates of wt HbI are in (A) and F97Y are in (B). CO bound species are shown as closed circles, deoxy species are shown as open circles. Error bars represent the error of the fit.

A) HbI



B) F97Y

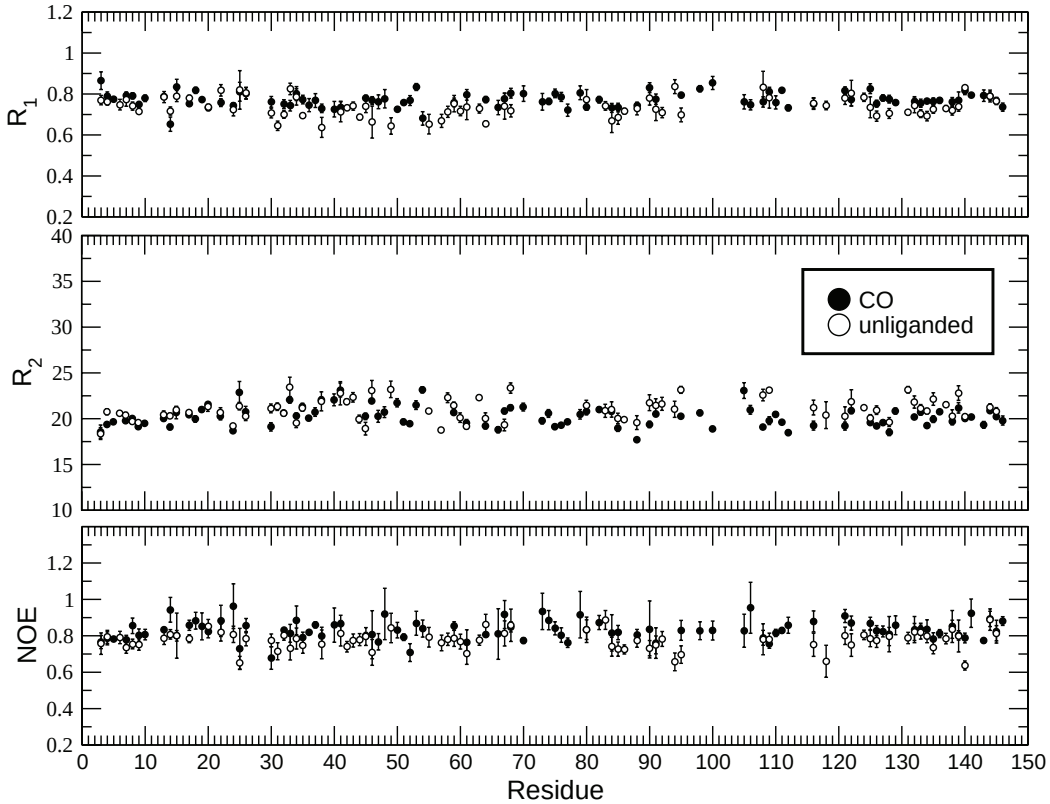


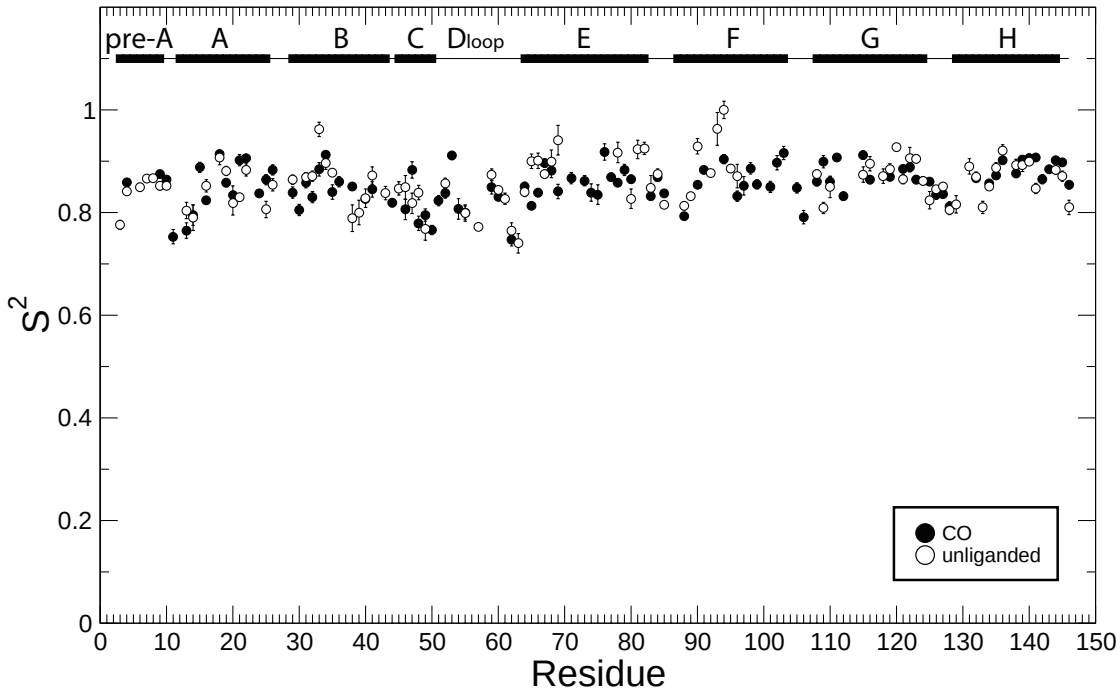
Figure 2.3

Laboratory-frame relaxation data were analyzed using the Lipari-Szabo model free formalism.(*29, 30*) Knowledge of the rotational diffusion properties of HbI is necessary for model free analysis. Analysis of the R_2/R_1 ratio (*31*) was used to estimate the overall rotational correlation time and determined the isotropic diffusion model to be the best representation for the rotational diffusion of WT HbI and F97Y HbI in both the liganded and unliganded states.

The order parameters, S^2, describing the amplitude of the backbone motion on the ps-ns time scale, were calculated for HbI in both the free and CO-bound states (**Figure 2.4 A**). The order parameter difference between the free and bound state represents the change in backbone dynamics that occurs upon ligand binding. CO-binding induces HbI to become, on average, more flexible, $S^2_{(bound)} - S^2_{(unliganded)} < 0$ (**Figure 2.5 A**). The interfacial helices E and F undergo the most dramatic change, becoming less rigid upon binding CO (**Figure 2.5 A**). Thermodynamic equilibrium studies of oxygen binding to HbI have shown that binding of the second oxygen molecule is made more favorable by a less negative entropic term, ΔS.(*2*) This suggests that ligand binding by HbI is entropically driven. From the order parameters, the protein conformational entropic contribution to the binding free energy of carbon monoxide was estimated to be -2.1 ± 0.4 kcal/mol. Although this result only accounts for the changes in conformational fluctuations of the residues with a calculated order parameter and neglects any correlation between spins, it supports the hypothesis that the increased protein flexibility associated with ligand binding contributes to cooperativity in HbI.

Figure 2.4: Backbone order parameters (S^2) of HbI and F97Y. Order parameters of wt HbI are in (A) and F97Y are in (B). CO bound species are shown as closed circles, deoxy species are shown as open circles. Structural elements are shown at the top of each plot. Solid bars represent α-helices and lines represent loops.

A) HbI



B) F97Y

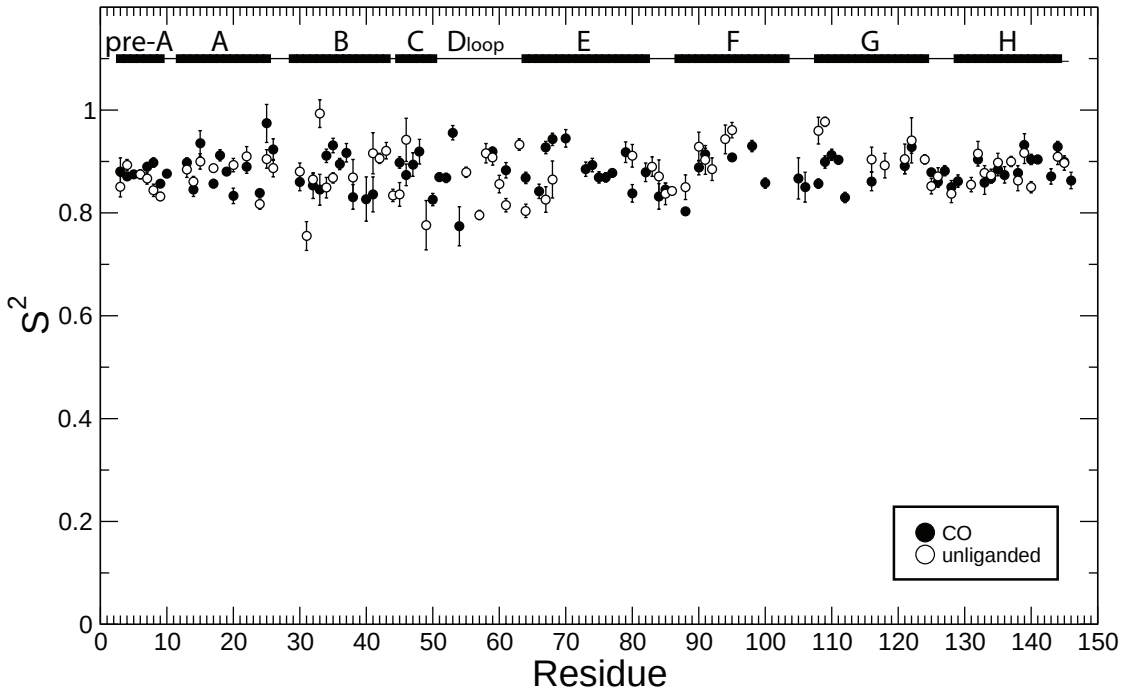


Figure 2.4

Figure 2.5: Ligand induced changes in HbI and Phe-97-Tyr backbone flexibility. Order parameters of wt HbI are in (A) and F97Y are in (B). For each available pair of backbone residues a black bar indicates the difference between bound and free order parameters. Error bars are the sum of the error in the fit. Structural elements are shown at the top of each plot. Solid bars represent α-helices and lines represent loops.

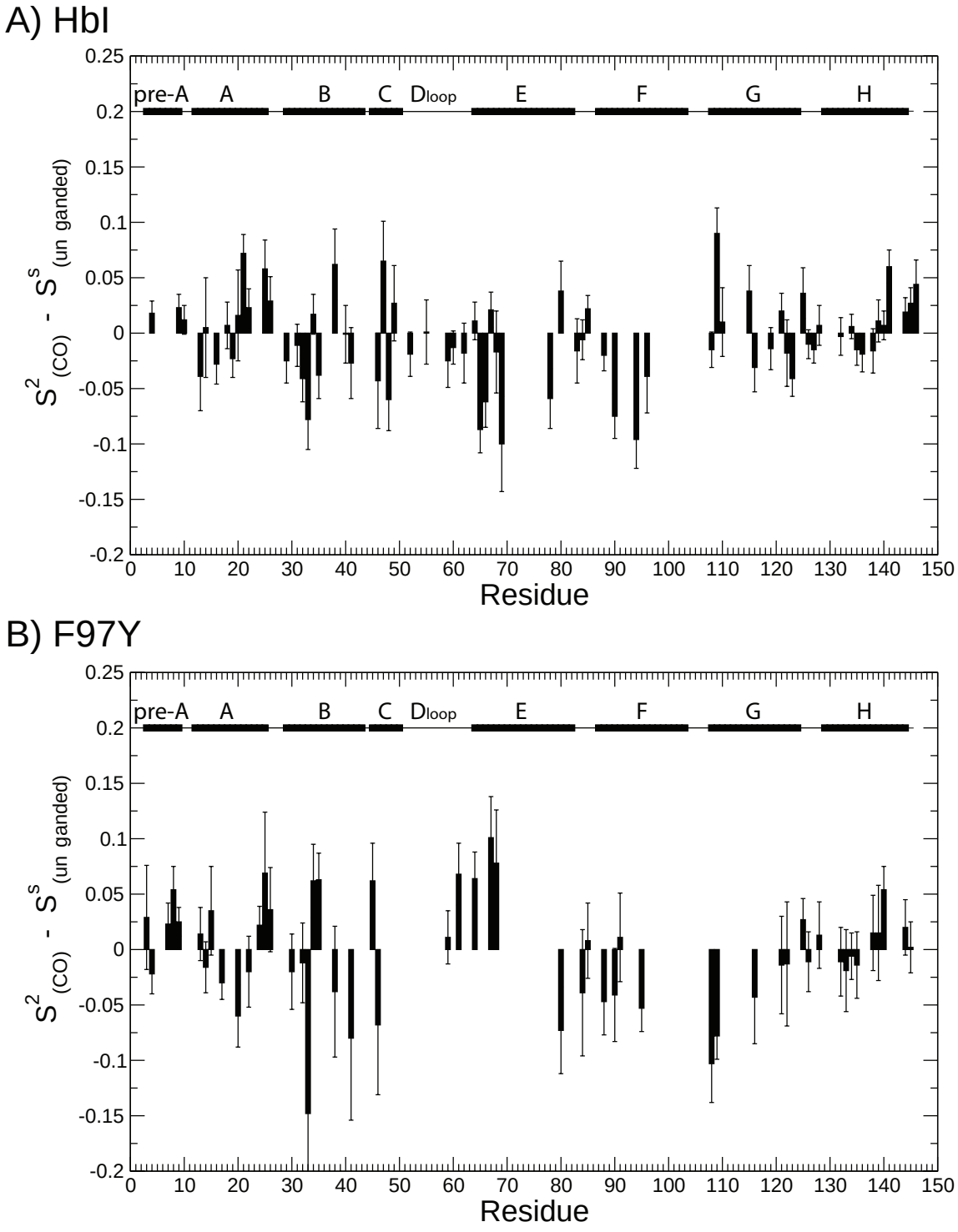

A) HbI
pre-A
A
B
C
Dloop
E
F
G
H
$S^2_{(CO)} - S^s_{(un\ ganded)}$
Residue
B) F97Y
pre-A
A
B
C
Dloop
E
F
G
H
$S^2_{(CO)} - S^s_{(un\ ganded)}$
Residue

Figure 2.5

To further test this hypothesis, the change in internal dynamics upon CO-binding was measured in the high affinity-state mimic F97Y. This mutant protein of HbI explores the effect of mutation on the F4 phenylalanine whose side-chain conformational transition, from the heme pocket to the dimeric interface, plays a central role in the allosteric mechanism of HbI and is characterized by a greatly reduced cooperativity: the Hill coefficient is 1.1.(*21*) If the increased flexibility observed upon binding is involved in HbI cooperativity, we expected this mutant to show a different change in dynamics upon ligand binding. Indeed, F97Y does show a different trend: on the average F97Y is more rigid in the CO-bound state than in the free state (**Figure 2.4 B**). The estimated contribution to the free energy of binding arising from the conformational restriction associated with carbon monoxide binding is 2.4 ± 0.6 kcal/mol. This mutant protein is generally more rigid than the WT, in both the free and bound state. The backbone dynamics of the mutant protein is restricted at the interface by the F97Y mutation, since HbI is cooperative it is likely that this interface mutation could yield an overall rigidification of the protein in both the CO bound and unliganded states of the protein.

When comparing the order parameters calculated in the liganded and unliganded states of each protein, the most obvious difference between the internal dynamics is observed at the interfacial helix E, which upon CO-binding becomes more rigid in F97Y but more flexible in wild-type HbI (**Figure 2.5 B**). The direction of changes in internal dynamics of wild-type and mutant HbI are mapped onto the structures of CO-HbI and CO-F97Y respectively (**Figure 2.6)**.

This result suggests that the backbone dynamics at the dimeric interface are involved in HbI cooperativity and are coupled with HbI allosteric regulation. However, a full understanding of the cooperative mechanism of HbI requires an assessment of the entropic contributions of the water and of protein side-chains.

Figure 2.6: Changes in average backbone flexibility of HbI and F97Y.

Backbone amide residues that undergo a significantly different change in order parameters outside the error of the sum of differences and undergo a change in order parameters greater than 0.02 are represented as blue and red spheres. Red spheres indicate a significant increase in flexibility upon ligand binding. Blue spheres indicate a significant decrease in flexibility upon ligand binding. Wild-type HbI is in A and F97Y is shown in B. The helical backbones of wild type and F97Y are shown as grey loops. Hemes and ligands are shown as grey sticks and grey spheres respectively.

Structures of CO-HBI are generated from the PBD file, 3SDH and structures of CO-F97Y are generated from the PDB file 2AUO.

References:

Royer, W.E., Jr. High-resolution crystallographic analysis of a co-operative dimeric hemoglobin. J Mol Biol 235, 657-81 (1994).

Knapp, J.E., Bonham, M.A., Gibson, Q.H., Nichols, J.C. & Royer, W.E., Jr. Residue F4 plays a key role in modulating oxygen affinity and cooperativity in Scapharca dimeric hemoglobin. *Biochemistry* **44**, 14419-30 (2005).

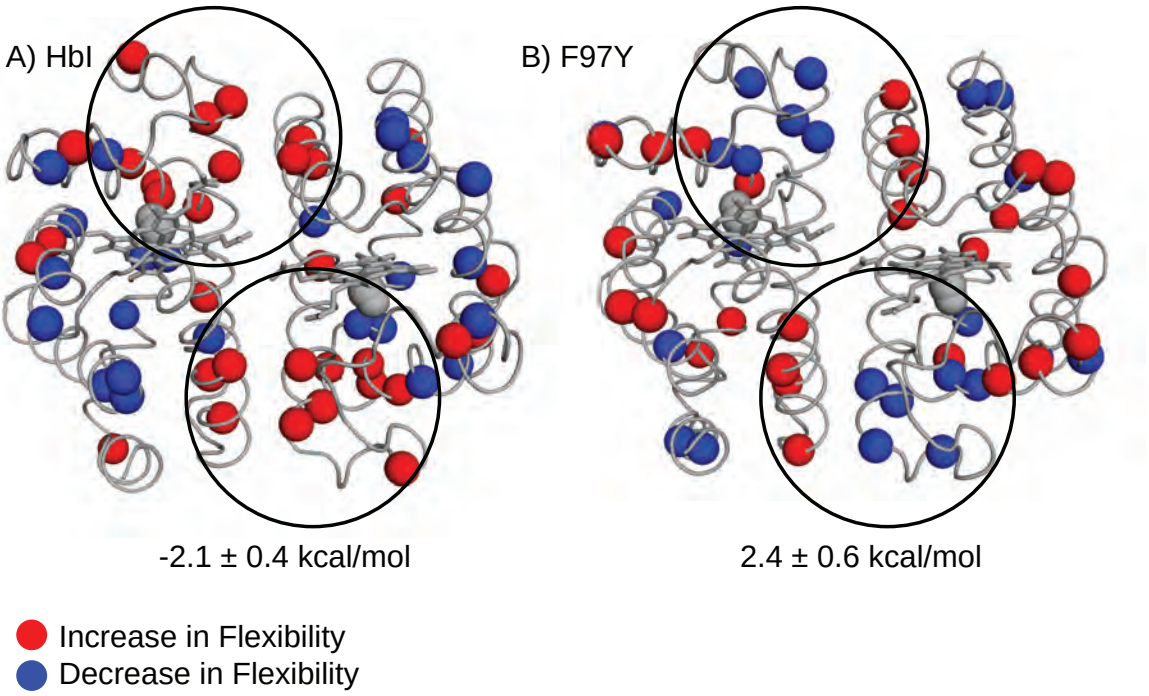

A) HbI
B) F97Y
-2.1 ± 0.4 kcal/mol
2.4 ± 0.6 kcal/mol
Increase in Flexibility
Decrease in Flexibility

Figure 2.6

DISCUSSION

Structural studies of HbI have been powerful in demonstrating the important role of local structural rearrangements in cooperative ligand binding. For the first time, using NMR, we demonstrate that the dynamics of HbI are also involved in HbI cooperativity. The backbone flexibility of HbI contributes to cooperativity in two ways, 1) by contributing favorably to the free energy of the system and 2) by participating in the cooperative mechanism at the HbI subunit interface. As a proof of principle we measured the fast internal motions of the weakly cooperative HbI mutant, F97Y. The change in backbone flexibility of F97Y is disrupted throughout this mutant protein and particularly at its interface. The conformational entropy change associated with ligand binding disfavors cooperativity in F97Y. These results demonstrate that changes in conformational entropy contribute together with structural changes to the cooperative mechanism of HbI.

X-ray crystallography has shown that the F-helix of HbI undergoes the largest conformational change upon binding ligand.(*1*) HbI undergoes a quaternary structural rearrangement upon binding ligand; the HbI subunits rotate by 3.3º. The conformational rearrangement by helix F including F97 is critical; as is evident by the loss of cooperativity in the F97Y mutant of HbI. Thus, both the structure and dynamics of the F-helix of wild type HbI are sensitive to ligand binding.(*1, 32*) F97Y undergoes a similar change in quaternary structure characterized by 3.0º subunit rotation.(*21*) However, for both the unliganded and CO bound states of F97Y, the tyrosine side-chain at position 97 remains in the

hydrophilic interface; this prevents helix F from undergoing a conformational transition that brings its backbone closer to the heme.(*21*) The ligand promoted change in the dynamics of helix F is not significantly perturbed in the mutant F97Y relative to wild type. This result suggests that the change in dynamics of helix F is a result of ligand binding may not be essential in HbI cooperativity.

Crystallographic data suggests that the conformation of the E helix is not perturbed by ligand binding. However, our data indicate that helix E becomes more flexible upon ligand binding and that this change in dynamics is important for the HbI cooperative mechanism. In fact, ligand binding has the opposite effect on the dynamics of the E helix in F97Y, where it becomes more rigid, and this mutant protein has greatly reduced cooperativity relative to the wild type.

Equilibrium thermodynamic studies of the thermodynamics of oxygen binding show that HbI cooperativity is driven by favorable changes in entropy. In this study we estimate that the difference in conformational entropy between bound and unliganded states of HbI favors HbI cooperativity. The change in conformational entropy of F97Y disfavors cooperativity. Previous studies of HbI have not been able to estimate the role of HbI conformational entropy to cooperativity. In this study we have learned that the changes in backbone conformational entropy favor HbI cooperativity and that changes in HbI dynamics of HbI are involved in HbI cooperative binding.

Molecular dynamics (MD) simulations of HbI have been used to investigate the role of interfacial water molecules in the cooperative mechanism of HbI.(*33*) The results of these studies indicate that water clusters in unliganded

HbI are less dynamic than in liganded HbI.(*33*) In addition, energy transfer between subunits is observed to be 10% more effective in the presence of interfacial water molecules.(*33*)

In future studies we plan to use MD simulations to understand the coupling between the change in structure and dynamics of the interfacial helices (E and F) and of the interfacial water and its role in cooperativity. If water dynamics mediate the cooperative mechanism at the interface, then comparison of MD simulations of wild type and mutant F97Y HbI will show changes in the dynamics of interfacial water molecules and reduced or lost coupling between the water and the protein. The results of MD simulations, together with studies of the HbI, the side-chain dynamics will reveal the full significance of the protein dynamics in HbI cooperativity.

MATERIALS AND METHODS

Protein expression and purification

HbI was expressed in *E. coli* Q-cells (*34*) under the control of a T7 promoter, in M9 minimal media supplemented with a standard vitamin mix and a trace metals solution.(*35*) To label HbI, the glucose or ammonium chloride of M9 was substituted with ^{13}C and ^{15}N labeled compounds respectively. HbI was induced for 6 hours with 1 mM Isopropyl β-D-1-thiogalactopyranoside, IPTG, and supplemented with 84 μM amino levulinic acid, ALA.

HbI was purified as previously described (*35*). Cell pellets were resuspended in 0.5 M Tris pH 8.0, 100 mM NaCl, 10 mM Ethylenediaminetetraacetic acid (EDTA), 1 mM phenylmethylsulfonyl fluoride (PMSF) and 5% (w/v) glycerol. Cells were lysed by sonication, duty cycle 7, for 30 seconds every minute for five minutes. The lysed cell solution was treated with DNase1, RNaseA and 6 mM $MgCl_2$ for thirty minutes. Cellular debri was pelleted by centrifugation at 15,000 rpms. Contaminates salt out of solution at 45% saturation with ammonium sulfate; HbI salts out at 95% saturation with ammonium sulfate. HbI was isolated using a DEAE column in 40 mM 2-(Cyclohexylamino)ethanesulfonic acid (CHES) pH 9.0 and a CM sepherose column in 40 mM 2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid (HEPES) pH 7.0.

Sample preparation

CO-HbI samples were degassed with CO and NMR tubes were filled with CO before being sealed with epoxy. Unliganded HbI samples were prepared from CO-HbI; to remove CO, samples were first agitated and flashed with white light in the presence of oxygen. Then to produce unliganded samples O_2-HbI was flashed with white light in a N_2 environment overnight. A trace amount of sodium dithionite was added to deoxy HbI samples before being sealed with epoxy within the anaerobic chamber. Absorbance at 534 nm, 416 or 422 nm light indicates the presence of deoxy HbI, O_2-HbI and CO-HbI respectively. Similar procedures have been successfully used on other hemoglobins for many years.(*36-38*)

NMR spectroscopy

3D sensitivity-enhanced gradient-selected [^{1}H,^{15}N]-TROSY (tranverse-relaxation optimized spectroscopy) triple-resonance experiments (HNCA, HN(CO)CA, HNCACB, CBCA(CO)NH, HNCO, and HN(CA)CO)(*23, 24*) were collected at 14.1 T and at 298 K using uniformly labeled ^{13}C/^{15}N 0.4 mM Hbl and 0.6 mM F97Y HbI (50 mM HEPES pH 7.0, 150 mM NaCl, and 95%H_2O/5%D_2O, T = 298 K). Additional 3D ^{15}N edited ^{1}H–^{1}H NOESY and HMQC-NOESY-HSQC experiments were collected at 14.1 T and 298 K and used to aid in the backbone resonance assignment of HbI and F97Y HbI in both the free and bound states. All NMR ^{15}N relaxation experiments were performed on [U-^{15}N] samples of HbI and F97Y HbI (50 mM Hepes, pH 7.0, 150 mM NaCl, 95%H_2O/5%D_2O, T = 298 K) at the following concentrations: 0.8 mM CO-HbI, 0.8 mM unliganded HbI, 0.7 mM CO-F97Y and 1.0 mM unliganded F97Y. All NMR triple-resonance and ^{15}N spin relaxation data were collected on a Varian Inova spectrometer operating at 600

MHz equipped with a triple-resonance cold probe. The temperature was calibrated using a sample of 100% methanol. ^{15}N R_1 and R_2 values and $\{^1H\}$-^{15}N NOE were measured using standard methods (*39-41*). Nine relaxation delays, two of which were duplicates, were used to measure R_1 and R_2. Relaxation delays ranged between 0 and 0.666 seconds for R_1 and between 0 and 0.090 s for R_2. The interval between 180° pulses in the CMPG experiment used to measure R_2 was 1.5 ms. Three sets of NOE data (proton saturated and non-saturated spectra) were collected in an interleaved manner. Data were acquired with 128 × 2048 (t_1× t_2) complex points and spectral widths of 2200 x 8,000 Hz. Data processing was performed using NMRPipe (*42*) and Sparky (*43*) software, Curvefit (www.palmer.hs.columbia.edu), along with in-house written programs. Intensities of cross peaks were used to quantify relaxation and uncertainties were estimated from duplicate (R_1, R_2) or triplicate (NOE) experiments.

Model free analysis

The dominant source of ^{15}N amide relaxation is through dipole dipole (DD) interactions with the attached protein and chemical shift anisotropy (CSA). The experimentally measured relaxation rates R_1, R_2, and the NOE depend on the reduced spectral density function J(ω) which describes internal motions: (*44*)

$$R_1 = \left(d^2/4\right)\left[3J\left(\omega_N\right) + 7J\left(0.921\omega_H\right)\right] + c^2 J\left(\omega_N\right) \quad (1)$$

$$R_2 = \left(d^2/8\right)\left[4J(0) + 3J\left(\omega_N\right) + 13J\left(0.955\omega_H\right)\right] + \left(c^2/6\right)\left[4J(0) + 3J\left(\omega_N\right)\right] + R_{ex} \quad (2)$$

$$NOE = 1 + \left(d^2/4R_1\right)\left(\gamma_H/\gamma_N\right)\left[5J\left(0.870\omega_H\right)\right] \quad (3)$$

where c = $\omega_N \Delta\sigma / \sqrt{3}$ and d = $(\mu_0 / 8\pi) h\gamma_H\gamma_X / (r_{NH})^3$ and, μ_0 is the permeability of free space, *h* is Planck's constant, γ_H and γ_X are the gyromagnetic ratios of the H and X nuclei respectively, r_{HX} is 1.04 Å.

Relaxation rates were analyzed using the Lipari-Szabo model free formalism.(*29, 30*) The Lipari-Szabo model free formalism assumes that the internal motion of the N-H vector is uncorrelated from the overall rotational motion of the protein. In the Lipari and Szabo formalism extended by Clore *et al (45)* the spectral density function can be described as

$$J(\omega) = \frac{2}{5}\left[\frac{S^2\tau_m}{1+(\omega\tau_m)^2} + \frac{(S_f^2 - S^2)\tau_s'}{1+(\omega\tau_s')^2} + \frac{(1-S_f^2)\tau_f'}{1+(\omega\tau_f')^2}\right] \qquad (4)$$

where $S^2 = S_f^2 S_s^2$, $1/\tau_f' = 1/\tau_f + 1/\tau_m$, and $1/\tau_s' = 1/\tau_s + 1/\tau_m$. The generalized order parameter, S^2, defines the amplitude of the motion and S_f^2 and S_s^2 are the square of the generalized order parameter of the fast and slow components. The order parameter is a measure of the degree of freedom of the motion of the intermolecular amide bond vector; S^2 is equal to 1 if the motion is completely restricted and is equal to 0 for isotropic motion τ_e is the effective correlation time of the generalized order parameter. τ_m is the isotropic rotational correlation time and τ_f and τ_s are the effective correlation times of the fast and slow internal motions. Both τ_f and τ_s are much faster than τ_m .

Values of the R_2/R_1 ratio were used to estimate the rotational diffusion tensor using the program r2r1_diffusion (www.palmer.hs.columbia.edu). Residues that have a value of the R_2 rate constant beyond one standard deviation from the mean and/or have NOE values less than 0.65 were eliminated

from this analysis.(*31*)

The relaxation data were analyzed using the program FAST Modelfree (*46*) (http://xbeams.chem.yale.edu/~loria) and Modelfree 4.20 (www.palmer.hs.columbia.edu). Five different motional models were used to facilitate the analysis of the relaxation data: 1, 2, 3, 4, 5 (*47*). For models 1-4 $\left(S^2 = S_f^2\right)$. Model 1 optimizes S_f^2 assuming internal motions are very fast $\left(\tau_f \approx 0\right)$ and slow internal motions are negligible $\left(S_s^2 = 1\right)$and $\left(\tau_s = 0\right)$. Model 2 optimizes S_f^2 and $\left(\tau_f = \tau_e\right)$ assuming slow internal motions are negligible $\left(S_s^2 = 1\right)$and $\left(\tau_s = 0\right)$. Model 3 and model 4 arise from models 1 and 2 respectively but including the chemical exchange term, R_{ex}. Model 5 optimizes the S_f^2 and S_s^2 assuming $\left(\tau_s = 0\right)$.

Order parameters have been used to estimate an upper bound for the change in entropy, ΔS, resulting from conformational restriction: (*48*)

$$\Delta S = -k_B \sum_i \ln\left\{\frac{3-\left(1+8S_{a,i}\right)^{\frac{1}{2}}}{3-\left(1+8S_{b,i}\right)^{\frac{1}{2}}}\right\} \qquad (5)$$

for which k_B is the Boltzmann constant, and S_a and S_b are the order parameters of the a and b states for spin *i*. This equation only accounts for the entropic effect of the part of the conformational ensemble characterized by the set of S^2 values but neglect any correlation between spins. The error in the conformational entropy is propagated from the error in the fits of the order parameters.

REFERENCES:

1. Royer, W. E., Jr. (1994) High-resolution crystallographic analysis of a co-operative dimeric hemoglobin, *J Mol Biol 235*, 657-681.
2. Ikeda-Saito, M., Yonetani, T., Chiancone, E., Ascoli, F., Verzili, D., and Antonini, E. (1983) Thermodynamic properties of oxygen equilibria of dimeric and tetrameric hemoglobins from Scapharca inaequivalvis, *J Mol Biol 170*, 1009-1018.
3. Kantrowitz, E. R. (2011) Allostery and cooperativity in Escherichia coli aspartate transcarbamoylase, *Arch Biochem Biophys 519*, 81-90.
4. Fuxe, K., Borroto-Escuela, D. O., Marcellino, D., Romero-Fernandez, W., Frankowska, M., Guidolin, D., Filip, M., Ferraro, L., Woods, A. S., Tarakanov, A., Ciruela, F., Agnati, L. F., and Tanganelli, S. (2012) GPCR heteromers and their allosteric receptor-receptor interactions, *Curr Med Chem 19*, 356-363.
5. Pardee, A. B., and Reddy, G. P. (2003) Beginnings of feedback inhibition, allostery, and multi-protein complexes, *Gene 321*, 17-23.
6. Popovych, N., Sun, S., Ebright, R. H., and Kalodimos, C. G. (2006) Dynamically driven protein allostery, *Nat Struct Mol Biol 13*, 831-838.
7. Tzeng, S. R., and Kalodimos, C. G. (2009) Dynamic activation of an allosteric regulatory protein, *Nature 462*, 368-372.
8. Petit, C. M., Zhang, J., Sapienza, P. J., Fuentes, E. J., and Lee, A. L. (2009) Hidden dynamic allostery in a PDZ domain, *Proc Natl Acad Sci U S A 106*, 18249-18254.
9. Chiancone, E., Vecchini, P., Verzili, D., Ascoli, F., and Antonini, E. (1981) Dimeric and tetrameric hemoglobins from the mollusc Scapharca inaequivalvis. Structural and functional properties, *J Mol Biol 152*, 577-592.
10. Condon, P. J., and Royer, W. E., Jr. (1994) Crystal structure of oxygenated Scapharca dimeric hemoglobin at 1.7-A resolution, *J Biol Chem 269*, 25259-25267.
11. Mitchell, D. T., Kitto, G. B., and Hackert, M. L. (1995) Structural analysis of monomeric hemichrome and dimeric cyanomet hemoglobins from Caudina arenicola, *J Mol Biol 251*, 421-431.
12. Royer, W. E., Jr., Sharma, H., Strand, K., Knapp, J. E., and Bhyravbhatla, B. (2006) Lumbricus erythrocruorin at 3.5 A resolution: architecture of a megadalton respiratory complex, *Structure 14*, 1167-1177.
13. Royer, W. E., Jr., Hendrickson, W. A., and Chiancone, E. (1990) Structural transitions upon ligand binding in a cooperative dimeric hemoglobin, *Science 249*, 518-521.
14. Fuentes, E. J., Der, C. J., and Lee, A. L. (2004) Ligand-dependent dynamics and intramolecular signaling in a PDZ domain, *J Mol Biol 335*, 1105-1115.
15. Kern, D., and Zuiderweg, E. R. (2003) The role of dynamics in allosteric regulation, *Curr Opin Struct Biol 13*, 748-757.

16. Frederick, K. K., Marlow, M. S., Valentine, K. G., and Wand, A. J. (2007) Conformational entropy in molecular recognition by proteins, *Nature 448*, 325-329.
17. Frederick, K. K., Kranz, J. K., and Wand, A. J. (2006) Characterization of the backbone and side chain dynamics of the CaM-CaMKIp complex reveals microscopic contributions to protein conformational entropy, *Biochemistry 45*, 9841-9848.
18. Igumenova, T. I., Frederick, K. K., and Wand, A. J. (2006) Characterization of the fast dynamics of protein amino acid side chains using NMR relaxation in solution, *Chem Rev 106*, 1672-1699.
19. Schnell, J. R., Dyson, H. J., and Wright, P. E. (2004) Structure, dynamics, and catalytic function of dihydrofolate reductase, *Annu Rev Biophys Biomol Struct 33*, 119-140.
20. Mozzarelli, A., Bettati, S., Rivetti, C., Rossi, G. L., Colotti, G., and Chiancone, E. (1996) Cooperative oxygen binding to scapharca inaequivalvis hemoglobin in the crystal, *J Biol Chem 271*, 3627-3632.
21. Knapp, J. E., Bonham, M. A., Gibson, Q. H., Nichols, J. C., and Royer, W. E., Jr. (2005) Residue F4 plays a key role in modulating oxygen affinity and cooperativity in Scapharca dimeric hemoglobin, *Biochemistry 44*, 14419-14430.
22. Knapp, J. E., and Royer, W. E., Jr. (2003) Ligand-linked structural transitions in crystals of a cooperative dimeric hemoglobin, *Biochemistry 42*, 4640-4647.
23. Cavanagh, J., Fairbrother, W. J., Palmer, A. G., 3rd, and Skelton, N. J. (1996) *Protein NMR Spectroscopy: Principles and Practice*, Academic Press, San Diego.
24. Loria, J. P., Rance, M., and Palmer, A. G., 3rd. (1999) Transverse-relaxation-optimized (TROSY) gradient-enhanced triple-resonance NMR spectroscopy, *J Magn Reson 141*, 180-184.
25. Bertini, I., Luchinat, C., and Piccioli, M. (2001) Paramagnetic probes in metalloproteins, *Methods Enzymol 339*, 314-340.
26. Weigelt, J. (1998) Single scan, sensitivity- and gradient-enhanced TROSY fo multidimentional NMR experiments, *J. Am. Chem. Soc. 120*, 10778-10779.
27. Zhu, G., Xia, Y., Nicholson, L. K., and Sze, K. H. (2000) Protein dynamics measurements by TROSY-based NMR experiments, *J Magn Reson 143*, 423-426.
28. Pervushin, K., Riek, R., Wider, G., and Wuthrich, K. (1997) Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution, *Proc Natl Acad Sci U S A 94*, 12366-12371.
29. Lipari, G., and Szabo, A. (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity, *J. Am. Chem. Soc. 104*, 4546-4559.

30. Lipari, G., and Szabo, A. (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results, *J. Am. Chem. Soc. 104*, 4559-4570.
31. Tjandra N., S. E. F., Firchard W. Pastor, Ad Bax. (1995) Rotational diffusion anisotropy of human ubiquitin from 15N NMR relaxation, *J. Am. Chem Soc. 117*, 12562-11566.
32. Ren, Z., Srajer, V., Knapp, J. E., and Royer, W. E., Jr. (2012) Cooperative macromolecular device revealed by meta-analysis of static and time-resolved structures, *Proc Natl Acad Sci U S A 109*, 107-112.
33. Gnanasekaran, R., Xu, Y., and Leitner, D. M. (2010) Dynamics of water clusters confined in proteins: a molecular dynamics simulation study of interfacial waters in a dimeric hemoglobin, *J Phys Chem B 114*, 16989-16996.
34. Mergulhao, F. J., Summers, D. K., and Monteiro, G. A. (2005) Recombinant protein secretion in Escherichia coli, *Biotechnol Adv 23*, 177-202.
35. Summerford, C. M., Pardanani, A., Betts, A. H., Poteete, A. R., Colotti, G., and Royer, W. E., Jr. (1995) Bacterial expression of Scapharca dimeric hemoglobin: a simple model system for investigating protein cooperatively, *Protein Eng 8*, 593-599.
36. Song, X. J., Yuan, Y., Simplaceanu, V., Sahu, S. C., Ho, N. T., and Ho, C. (2007) A comparative NMR study of the polypeptide backbone dynamics of hemoglobin in the deoxy and carbonmonoxy forms, *Biochemistry 46*, 6795-6803.
37. Ho, C. (1992) Proton nuclear magnetic resonance studies on hemoglobin: cooperative interactions and partially ligated intermediates, *Adv Protein Chem 43*, 153-312.
38. Volkman, B. F., Alam, S. L., Satterlee, J. D., and Markley, J. L. (1998) Solution structure and backbone dynamics of component IV Glycera dibranchiata monomeric hemoglobin-CO, *Biochemistry 37*, 10906-10919.
39. Kroenke, C. D., Loria J. P., Lee, L. K., Rance, M., Palmer, A. G. (1998) Longitudinal and transverse 1H-15N dipolar / 15N chemical shift anisotropy relaxation interference: unambiguous determination of rotational diffusion tensors and chemical exchange effects in biological macromolecules., *J Am Chem Soc 120*, 7905-7915.
40. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., and Kay, L. E. (1994) Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation, *Biochemistry 33*, 5984-6003.
41. Tjandra, N., Szabo, A., Bax, A. (1996) Protein backbone dynamics and 15N chemical shift anisotropy from quantitative measurement of relaxation interference effects., *J Am Chem Soc 118*, 6986-6991.
42. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes, *J Biomol NMR 6*, 277-293.

43. Goddard, T. D., and D.G. Kneller. . SPARKY 3, University of California, San Francisco, CA.
44. Farrow, N. A., Zhang, O., Szabo, A., Torchia, D. A., and Kay, L. E. (1995) Spectral density function mapping using 15N relaxation data exclusively, *J Biomol NMR 6*, 153-162.
45. Clore, G. M., Szabo, A., Bax, A., Kay, L. E., Driscoll, P. C., Gronenborn, A. M. (1990) Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins., *J Am Chem Soc 112*, 4989-4991.
46. Cole, R., and Loria, J. P. (2003) FAST-Modelfree: a program for rapid automated analysis of solution NMR spin-relaxation data, *J Biomol NMR 26*, 203-213.
47. Mandel, A. M., Akke, M., and Palmer, A. G., 3rd. (1995) Backbone dynamics of Escherichia coli ribonuclease HI: correlations with structure and function in an active enzyme, *J Mol Biol 246*, 144-163.
48. Yang, D., and Kay, L. E. (1996) Contributions to conformational entropy arising from bond vector fluctuations measured from NMR-derived order parameters: application to protein folding, *J Mol Biol 263*, 369-382.

CHAPTER III

MECHANISTIC STUDY OF ADENINE RECOGNITION BY TIS11D

AUTHOR CONTRIBUTIONS

This chapter is a collaborative project between Laura Deveau and myself. I cloned the TIS11d expression construct. Laura Deveau and I worked together to determine the RNA binding affinity of TIS11d. Laura Deveau optimized and performed the gel mobility shift assays shown and expressed and purified $^{15}N^{13}C$-TIS11d. I conducted and analyzed the NMR titrations of TIS11d with each of the RNAs and assigned the amide peaks of TIS11d, TIS11d-5′-UUUUUUUUAUUUU-3′, TIS11d-5′-UUUUAUUUUUUUU-3′, and TIS11d-U_{13}. This project was done under the direction of my advisor Francesca Massi.

ABSTRACT

The human protein TIS11d is an important regulator of mRNA transcripts involved in hematopoiesis.(*1, 2*) Mutant TIS11d is associated with the pathogenesis of acute myeloid leukemia and acute lymphoblastic leukemia; point mutations and frame shift mutations have been identified throughout the TIS11d gene locus of leukemic cells and immortalized cancer cell lines.(*2*) TIS11d binds 5′-UUAUUUAUU-3′ RNA sequences, although the mRNA targets of TIS11d are unknown. Studies of the TIS11d mouse homologue implicate TIS11d in the binding and regulation of Notch 1 mRNA at the 3′UTR (and other unknown targets) and demonstrate the potential tumor suppressor activity of TIS11d.(*3*) We show that TIS11d binds three sequences, 5′-UUUUAUUUAUUUU-3′ (ARE), 5′-UUUUUUUUAUUUU-3′ ($FM3_{13}$), and 5′-UUUUAUUUUUUUU-3′($FM4_{13}$) with high affinity. Backbone chemical shift differences of TIS11d in complex with each of these three high affinity targets suggest that TIS11d binds each RNA differently. NMR titration data shows that TIS11d binds different RNA sequences through different kinetic mechanisms and although 50% identical, zinc finger 1 (ZF1) and zinc finger 2 (ZF2) of TIS11d contribute to RNA binding differently.

INTRODUCTION

The human protein TIS11d is an important regulator of mRNA transcripts involved in hematopoiesis.(*1, 2*) Knock out, (KO) mice of the TIS11d mouse homologue gene *Zfp36l2* die two weeks from birth from hemorrhaging.(*1*) The mice have fewer hematopoietic progenitor cells and as a result have reduced red and white blood cells, hemoglobin, hematocrit and platelets.(*1*) In humans, misregulation of TIS11d is associated with the pathogenesis of acute myeloid leukemia, AML, and acute lymphoblastic leukemia, ALL.(*2*) Frameshift mutation and point mutations throughout the *TIS11d* gene locus were identified in the leukemic cells from patients with acute myeloid leukemia and in several cancerous immortalized cell lines.(*2*) AML- and ALL- related point mutations have been discovered in the NTD, CTD and RBD of the *TIS11d* gene locus.(*2*) The missense mutations H62Q and D219E are linked to AML; residues H62 and D219 are conserved in the TTP family.(*2*) The mutation at residue D219 is within the TIS11d RBD. The missense mutation, A329V is detected in ALL patients and is not conserved in the family.(*2*)

In cells, overexpression of wild type TIS11d suppresses cell proliferation and induces apoptosis.(*2*) The apoptosis-related proteins caspase-3, PARP and phosphorylated H2AX, are induced by overexpression of TIS11d.(*2*) Similar to the response of cells that are exposed to ultraviolet radiation, TIS11d activates the S phase checkpoint with increased amounts of p53 and diminished p21.(*2*) A frameshift mutation, I373fsX91 of the carboxy–terminal region has been used as a control in overexpression studies of TIS11d in HeLa cells. Overexpression of

I373fsX91 only has a limited effect on cell proliferation and does not mediate H2AX phorphorylation or activation of the S phase checkpoint.(*2*)

The presence of H2AX phosphorylation and activation of the S phase checkpoint indicate that DNA damage is occurring within the cell. Misregulation of proteins involved in histone biosynthesis and chromatin remodeling causes a similar DNA damage response as TIS11d. In addition, histone biosynthesis proteins have also been linked to AML.(*4*) If the function of TIS11d is linked to histone biosynthesis then misregulation of histone chaperones could be the source of TIS11d related leukemia. Reduction in p21 in the DNA damage response is a hallmark of inaccurate DNA repair by translesion DNA synthesis.(*5*) The overexpression of TIS11d activates the S phase checkpoint and reduces p21 suggesting that TIS11d may be involved in DNA repair.(*2*)

The RBD of the TTP family is highly conserved; the NTD and CTD of the TTP family are less homologous, **Figure 1.5**.(*6, 7*) However, in vivo, TTP, TIS11d and TIS11b may only be partially redundant; they do not bind all the same in vivo mRNA targets or proteins nor are they expressed in the same cell types.(*6, 8-10*) There is strong evidence that TTP regulates over 100 in vivo mRNA targets including TNFα.(*8, 11, 12*) Studies of the TTP and TIS11d lymphocyte double knock out mice implicate TIS11d and TTP in the binding and regulation of Notch 1 mRNA. Notch 1 mRNA is the only known in vivo mRNA target of TIS11d, and DII4 mRNA is the only known in vivo target of TIS11b.(*13, 14*) The search for in vivo mRNA targets is hampered by the lack of knowledge of the binding specificity of the RBDs of these proteins.

In a magnetic field, magnetically active nuclei have a characteristic resonance frequency. The resonance frequency relative to a standard is the chemical shift. Chemical shift is sensitive to the local chemical environment of a nucleus and therefore is related to the structure of the molecule. The chemical shift is then sensitive to structural transitions and ligand binding events that affect the distribution of electrons around a nucleus. Using only the observed chemical shift changes, it is difficult to identify which structural changes are associated with ligand binding. Therefore, in this study of TIS11d we cannot identify the structural differences of TIS11d in complex with three different RNA sequences based only on the difference between the two-dimensional NMR spectra of these complexes.

In this study we have used NMR and gel electrophoresis mobility shift assays (EMSA) to investigate the mechanism of adenine recognition by TIS11d. Using EMSA we have identified three RNA sequences that TIS11d binds with high affinity. NMR titration data shows that TIS11d binds different RNA sequences through different kinetic mechanisms. Although zinc fingers 1 (ZF1) and zinc finger 2 (ZF2) of TIS11d are each 50% identical each contribute to RNA binding differently. Future studies on the structure, flexibility and conformational entropy of RNA bound TIS11d will reveal the full significance of adenine recognition in the selection of mRNA targets.

MATERIALS AND METHODS

Reagents

All materials used were reagent grade except for the following purified chemicals. Deuterium oxide (>99%) was purchased from Spectra Stable Isotopes (Columbia, MD). ^{15}N ammonium chloride (98%), ^{13}C glucose (98%), and were purchased from Sigma-Aldrich (St Louis, MO).

Expression and Purification of TIS11d:

The RBD of TIS11d (residues 151-220) was expressed from the pet21b vector within BL21(DE3)*E. coli.* Isotopic labeling with ^{15}N was performed by growing the cells in M9 enriched with 1 g of $^{15}NH_4Cl$ per liter. Carbon labeling was performed by growing cells in ^{13}C-glucose. The cells were grown at 37℃ to an OD_{600} of 0.8 and then induced for 4 hours with 1 mM Isopropyl β-D-1-thiogalactopyranoside, IPTG, and 0.15 mM $ZnSO_4$ at the same temperature. In total 3 mgL^{-1} of TIS11d is purified, 50% of the protein is soluble and 50% is harvested from inclusion bodies. Cell pellets were resuspended in chilled lysis buffer: 25 mM HEPES pH 7.8, 100 mM NaCl, 1 protease inhibitor pellet, and 2 mM DTT. Cells were lysed by sonication, duty cycle 7, for 30 second increments every minute for five minutes. Inclusion bodies were pelleted by centrifugation for 30 minutes at 25,000 rpm and were dissolved in reducing buffer: 50 mM Tris pH 8.0, 150 mM DTT, 6.4 M guanidine for 30 minutes at 75 ºC. The cell slurry was titrated with 10% (v/v) trifluoroacetic acid, TFA to pH 2.0. TIS11d was isolated using a Sep-Pak reverse phase column and was eluted with a step-wise

gradient of acetonitrile and TIS11d fractions are lyophilized at -40 ºC overnight. The soluble fraction of TIS11d was purified as discussed previously.(*15*)

Refolding TIS11d:

TIS11d was refolded by dialysis in 10 mM TRIS pH 6.2, 20 mM KCl, 2.5 mM DTT at room temperature overnight. In the refolding buffer the concentration of $ZnSO_4$ was 2.5 times the concentration of unfolded TIS11d.

Gel Mobility Shift Assays:

Gel shift mobility assays were conducted as discussed previously.(*16*) Optimal binding of TIS11d to RNA was measured in the presence of 2 mM dithiothreitol, DTT. The concentration of TIS11d was measured using a Ninhydrin assay. The affinities of TIS11d, $K_{d\ (apparent)}$ for the sequence: 5′-UUAUUUAUU-3′ (ARE_{13}), 5′-UUUUUUAUU-3′ and 5′-UUAUUUUUU-3′ were measured by direct titration of 3 nM labeled RNA with increasing concentrations of TIS11d. As discussed previously, the $K_{d,app}$ of TIS11d-5′-UUUUUUUUAUUUU-3′ and TIS11d-5′-UUUUAUUUUUUUU-3′ were fit to the Hill equation.(*16*)

$$\theta = \frac{\left([P]_t\right)^n}{\left([P]_t\right)^n + \left(K_{d,app}\right)^n}\ , \qquad (1)$$

where θ is the measured fraction bound, $[P]_t$ is the total protein concentration, $K_{d,app}$ is the apparent binding constant and n is the hill coefficient. The binding affinity of TIS11d to ARE_{13} is a very high affinity interaction.

Therefore the $K_{d,app}$ of TIS11d to ARE_{13} is not fit to the Hill equation but to the quadratic equation

$$\theta = \frac{\left([L]_t + [P]_t + K_d\right) - \sqrt{\left([L]_t + [P]_t + K_d\right)^2 - 4[L]_t[P]_t}}{2[P]_t}, \quad (2)$$

where $[L]_t$ is the total ligand concentration. The lower limit of binding to U_{13} and ARE_7 was estimated by competition with ARE_{13}.

Chemical Shift Calculations:

TIS11d-ARE_{13} has been assigned previously.(*17*) The free form of TIS11d was assigned using standard triple resonance experiments: HNCA, HN(CO)CA, HNCACB, CBCA(CO)NH. The RBD of TIS11d contains four prolines, P161, P190, P207, P210. Using the assignments of free TIS11d and TIS11d in complex with ARE_{13} TIS11d-$U_{13,}$ TIS11d-5′-UUUUUUUUAUUUU-3′ and TIS11d-5′-UUUUAUUUUUUUU-3′ were assigned. Amide cross-peaks of the TIS11d RNA complexes wee assigned from the assignments of TIS11d and TIS11d-ARE_{13}. Residues that undergo large chemical shifts could not be assigned. Amide cross-peaks of the N-terminal residues S151, T152 and residue H181 are absent in the free form of TIS11d and in each of the TIS11d complexes. Residues R188, H189 of free TIS11d are also absent. Amide cross-peaks of residues: R153, Y154, T156, E157, C159, Y170, F176, A177, R184, L186, T187, K191, K193, L196, C197, T199, F200, R211, H213, F214 of TIS11d-U_{13} could not be assigned with certainty but are present in the spectrum. The residues: T156, E157, L158, C159, R160, T187, R189, of TIS11d-5′-UUUUUUUUAUUUU-3′ could not be assigned with certainty but are present in the spectrum. Residues

R153, Y154, T156, C159, R184, L186, T187, K193, T194, E195, R198, T199, F200, R211, H213, F214 of free TIS11d-5′-UUUUAUUUUUUUU-3′ could not be assigned and are absent from the TIS11d-5′-UUUUAUUUUUUUU-3′ spectrum.

Chemical shift differences between the free and bound states of TIS11d were monitored using a series of ^{1}H-^{15}N HSQCs. All experiments were collected on a 600 MHz Varian spectrometer at 25 °C. All spectra were processed with nmrPipe,(*18*) visualized with nmrDraw(*18*) and Sparky.(*19*) The chemical shift differences, Δ, were determined using $\Delta = (\delta_H^2 + (\delta_N\gamma_N/\gamma_H)^2)^{1/2}$, where δ_H is the chemical shift difference of proton, δ_N is the chemical shift difference of nitrogen and γ_N, γ_H are the gyromagnetic ratio of nitrogen and hydrogen atoms. Chemical shift differences were mapped onto the TIS11d-ARE_{13} NMR structure using PyMOL.(*20*) Chemical shift difference plots are generated using Grace.(*21*)

NMR titrations:

NMR titrations were conducted by titrating small volumes (µl) of 1-4 mM into a 600 µl sample of TIS11d. For each titration the starting concentration of TIS11d was less than 300 µM. Titrations were complete at stoichiometric excess. The titrations of TIS11d with either ARE_{13} or U_{13} were repeated twice and show the same results. Chemical shift changes between the free and bound states of TIS11d were monitored using a series of ^{1}H-^{15}N HSQCs.

RESULTS

The binding affinity of TIS11d for ARE cognate sequences

The affinity of TIS11d for three ARE-like sequences has been determined using quantitative gel electrophoretic mobility shift assays. TIS11d binds 5′-UUUUAUUUAUUUU-3′ (ARE_{13}), 5′-UUUUUUUUAUUUU-3′ and 5′-UUUUAUUUUUUUU-3′ with nM affinity. Out of the three, TIS11d binds the ARE_{13} cognate sequence with the highest affinity, **Figure 3.1**. The affinity of TIS11d for U_{13} and ARE_7 are estimated through competition gel shift. TIS11d binds U_{13} and ARE_7 weakly and therefore, only a lower limit can be determined, **Figure 3.2**. At higher concentrations of RNA, the RNA self associates and migrates within the agarose gel as a mixture of complexes. The affinity of TIS11d for ARE-like sequences and the free energy differences of binding are listed in **Table 3.1**.

The structure of TIS11d shows that each ZF of TIS11d binds the UAUU sequence element.(*17*) This suggests that the shortest RNA sequence TIS11d can bind using both ZFs is eight bases long.(*22*) In order to verify this we tested the affinity of TIS11d for 5′-UUUUAUU-3′, ARE_7. In this case the $K_{i\ (apparent)}$ is greater than 20 μM. If either ZF1 or ZF2 selected a UAUU recognition element out of context the affinity of ARE_7 would be comparable to the affinity of 5′-UUUUUUUUAUUUU-3′ or 5′-UUUUAUUUUUUUU-3′. Since the affinity of ARE_7 is too weak to be measured we conclude that TIS11d binds RNA only if both fingers bind RNA.

Figure 3.1: Direct titration of TIS11d with the ARE_{13} cognate sequence. The fraction bound is determined as a function of the concentration of TIS11d increasing from 70 pM to 986 nM designated by the black triangle as determined by gel mobility shift assays. Within the gel, fluorescently labeled RNA appears as dark band. The lower band is free RNA and the higher band is bound RNA. Binding to ARE_{13} was determined by direct titration and fitting to the quadratic equation.

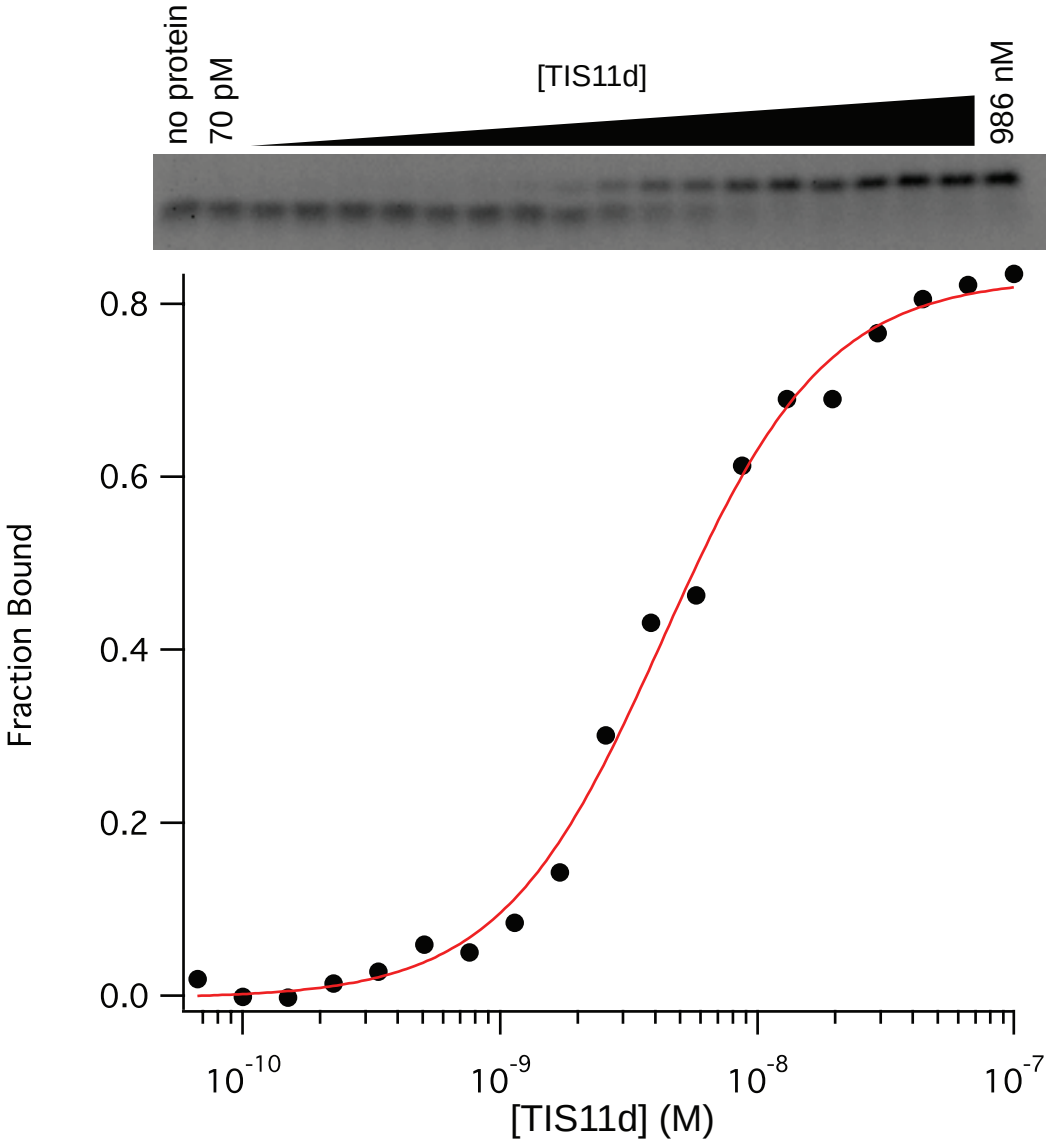

no protein
70 pM
[TIS11d]
986 nM
Fraction Bound
0.0
0.2
0.4
0.6
0.8
10^-10
10^-9
10^-8
10^-7
[TIS11d] (M)

Figure 3.1

Figure 3.2 RNA competition assays of TIS11d bound to RNA ARE_{13} cognate sequence. The fraction of TIS11d bound to ARE_{13} is competed off with increasing the concentration of RNA at maximum 20 μM designated by the black triangle. Within the gel, fluorescently labeled RNA appears as a dark band. The lower band is free RNA and the higher band is bound RNA. The competing RNAs are in the following panels: (A) ARE_{13}, (B) U_{13}.

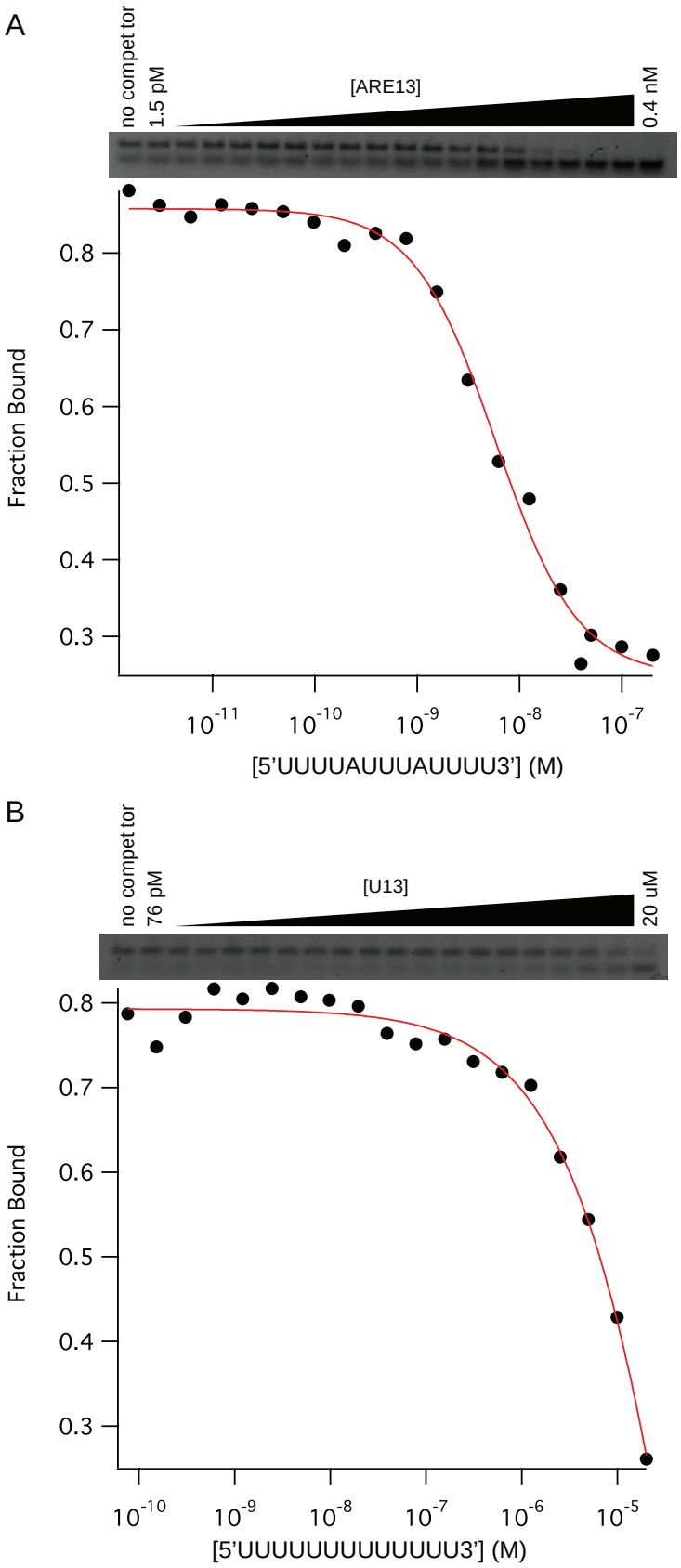

A
no competitor
1.5 pM
[ARE13]
0.4 nM
Fraction Bound
0.8
0.7
0.6
0.5
0.4
0.3
10^-11
10^-10
10^-9
10^-8
10^-7
[5'UUUUAUUUAUUUU3'] (M)
B
no competitor
76 pM
[U13]
20 uM
Fraction Bound
0.8
0.7
0.6
0.5
0.4
0.3
10^-10
10^-9
10^-8
10^-7
10^-6
10^-5
[5'UUUUUUUUUUUUU3'] (M)

Figure 3.2

Table 3.1: The specificity of TIS11d for ARE-like sequences. The RNA codes are listed in the first column with the corresponding RNA sequence listed 5′ to 3′ in the second; bases can be identified by positions numbered 1-13. The recognition element of ZF2 is shown in bold and the recognition element of ZF1 is underlined. For $ARE7_7$ and U_{13} the K_i apparent is > 20 μM. ΔG is the free energy of binding RNA and $\Delta\Delta G = \Delta G_{(RNAx)} - \Delta G_{(ARE13)}$, this indicates the thermodynamic preference for ARE_{13}. The fold ARE preference for ARE_{13} is listed in the final column. Errors listed are the standard deviation of three duplicates.

RNA	Sequence	$K_{d\ apparent}$	ΔG (kcal mol^{-1})	ΔΔG	ARE preference
ARE_{13}	UUU**UAUU**UAUUUU	3 ± 1 nM	-11.7 ± 0.1	0	1
$FM3_{13}$	UUU**UUUU**UAUUUU	24 ± 7 nM	-10.4 ± 0.2	1.3	8
$FM4_{13}$	UUU**UAUU**UUUUUU	16 ± 3 nM	-10.6 ± 0.1	1.1	6
ARE_7	UUU**UAUU**_____	> 20 μM			
U_{13}	UUU**UUUU**UUUUUU	> 20 μM			

TIS11d binds single stranded RNA in the 3′ to 5′ direction such that ZF1 binds the 3′ UAUU element and ZF2 binds the 5′ UAUU element.(*17*) The adenines at positions A5 and A9 are critical for generating the highest affinity recognition by TIS11d. Upon substituting both adenines with uridines TIS11d binding is more than 8,000 times weaker than for ARE_{13}, **Table 3.1**. However removal of one of the two adenines yields only a six fold or eight fold reduction in affinity, **Table 3.1**.

The adenine at position five and the adenine at position nine contribute similarly to TIS11d recognition. Removal of either the adenine at position five or the adenine at position nine causes a similar reduction in affinity. Therefore, ZF1 and ZF2 of TIS11d contribute similarly for the selection of ARE sequences. This is not surprising since ZF1 and ZF2 of TIS11d are more than 50% identical in primary sequence. Although the RBD of TIS11d is nearly identical to the RBD of TTP, the binding specificity of TTP and TIS11d are different. TTP and TIS11d bind ARE_{13} with similarly affinity.(*22*) However, TIS11d and TTP bind 5′-UUUUUUUUAUUUU-3′ and 5′-UUUUAUUUUUUUU-3′ with different affinities.(*22*) TIS11d is tolerant of U-A base substitutions at either positions 5 or 9. TTP is less tolerant of these substitutions. As a result, TTP prefers the ARE_{13} twenty times more than either the FM3 or FM4 sequences.(*22*) Removal of both adenine bases has an affect on the binding affinity of TIS11d, the binding is so weak it cannot be measured.

Chemical shift changes in TIS11d associated with high affinity binding to ARE_{13}

Chemical shift changes are observed for most of the residues of the RBD of TIS11d upon addition of RNA, **Figure 3.3 (C)**. The most significant changes come from those backbone residues that interact and recognize RNA bases, **Figure 3.4(A)**. Three residues undergo the largest change in chemical shift, R160, K191 and R198. The backbone amide of R160 and R198 forms hydrogen bonds with the A9 and A5 RNA bases, respectively.(*17*) K191 may make contact with the phosphate backbone of the RNA although the contact is not clear from the NMR structure.(*17*) The conserved residues of the R/KYKTEL sequence that precedes ZF1 and ZF2, undergo moderate changes in chemical shifts, these residues also form backbone hydrogen bonds with the 5′UA bases of each 5′-UAUU-3′ recognition element.

Figure 3.3: Chemical shifts of TIS11d and TIS11d-ARE$_{13}$. (A) The ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) spectra of TIS11d and ARE$_{13}$-TIS11d are shown in red and blue, respectively. The largest chemical shifts are mapped by a grey bar and labeled. (B) Amide chemical shift differences, $\Delta\delta$, have been determined using $\Delta\delta = (\delta_H^2 + (\delta_N\gamma_N/\gamma_H)^2)^{1/2}$, where δ_H is the chemical shift difference of proton, δ_N is the chemical shift difference of nitrogen and γ_N, γ_H are the gyromagnetic ratio of nitrogen and hydrogen atoms. Increasing chemical shift differences mapped from white to red onto the NMR structure of ARE$_{13}$-TIS11d using PyMOL (C). Zinc is shown as grey spheres and RNA is shown as blue lines.

The structure of TIS11d-ARE$_{13}$ is generated from the PBD file 14GO.

Reference:

Hudson, B.P., Martinez-Yamout, M.A., Dyson, H.J. & Wright, P.E. Recognition of the mRNA AU-rich element by the zinc finger domain of TIS11d. *Nat Struct Mol Biol* **11**, 257-64 (2004).

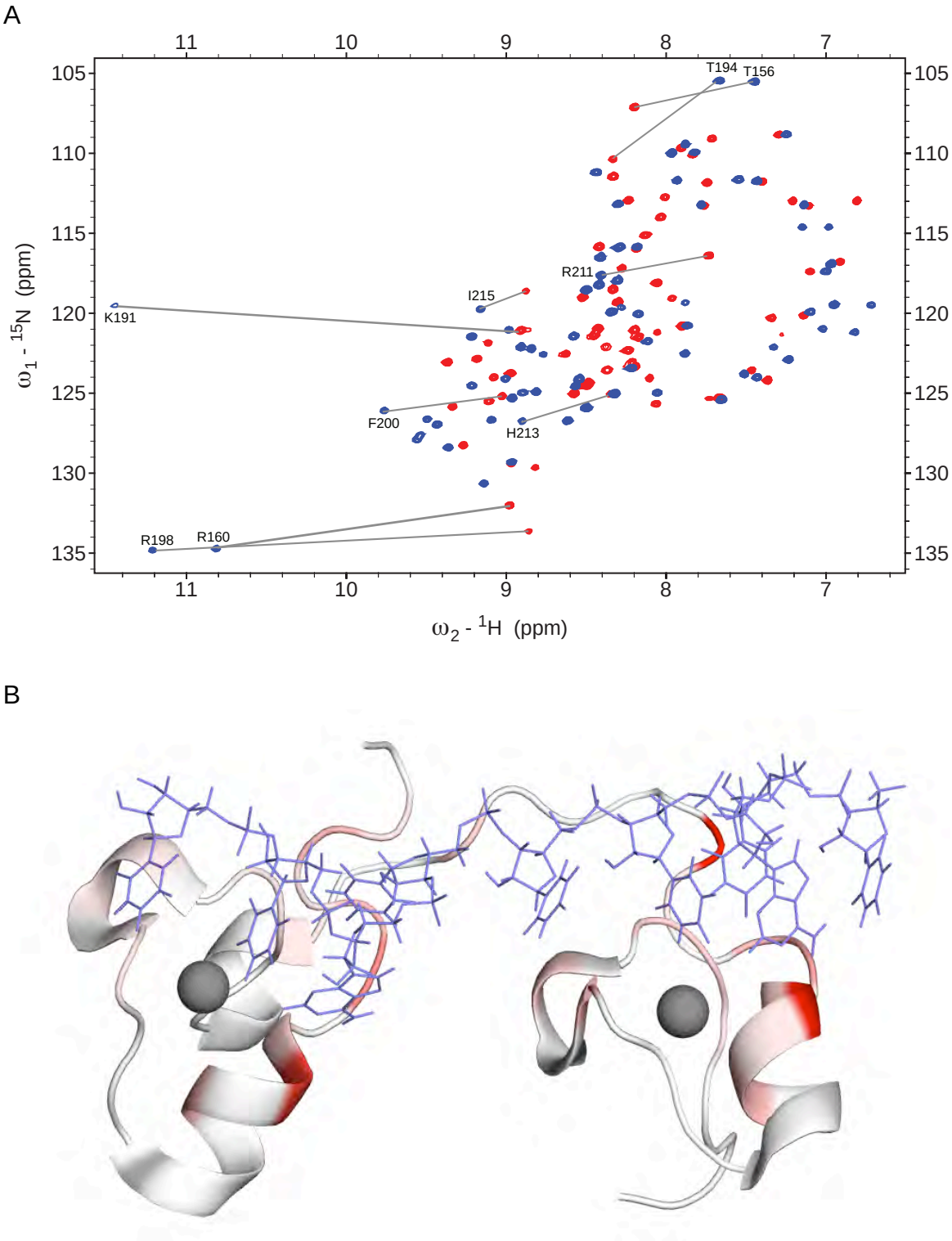

A
11
10
9
8
7
105
110
115
120
125
130
135
T194
T156
R211
I215
K191
F200
H213
R198
R160
ω1 - 15N (ppm)
ω2 - 1H (ppm)
B

Figure 3.3

Figure 3.4: Chemical shift differences of TIS11d bound to cognate RNA are larger than to pulyuridine. Atop each plot is a schematic representation of the TZD of TIS11d; the black horizontal bars represent ZF1 and ZF2, the blue horizontal bars are the preceding conserved RYKTEL and KYKTEL sequences, white circles represent the CCCH residues of ZF1 and ZF2, the red circles represent residues whose side chains form base stacks with RNA bases in the TIS11d-ARE_{13} complex, black dots denote a residue that hydrogen bonds to a RNA base in the TIS11d-ARE_{13} complex. For each plot, black bars designate the magnitude of amide backbone chemical shift differences of TIS11d between states. A) Chemical shift differences between free TIS11d and TIS11d in complex with the highest affinity cognate sequence ARE_{13}, ($\Delta\delta$ =TIS11d – TIS11d-ARE_{13}) B) The chemical shift differences of free TIS11d and TIS11d bound to pulyrudine, ($\Delta\delta$ =TIS11d – TIS11d-U_{13}). C) The chemical shift difference between TIS11d-ARE_{13} and TIS11d-U_{13}, ($\Delta\delta$ =TIS11d-ARE_{13} – TIS11d-U_{13}).

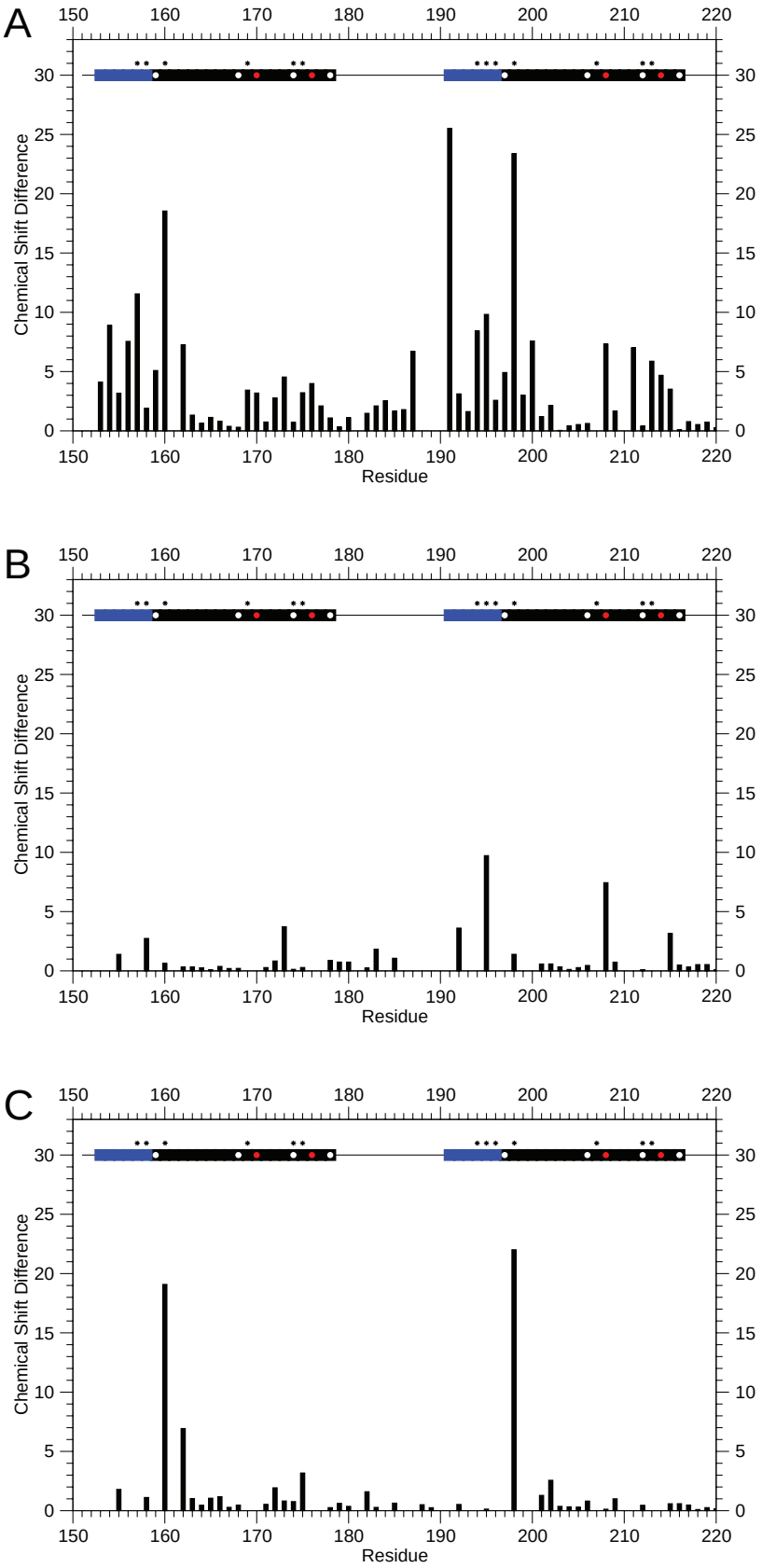

A
B
C
Chemical Shift Difference
Residue
150
160
170
180
190
200
210
220
0
5
10
15
20
25
30

Figure 3.4

Chemical shift changes in TIS11d associated with weak binding to U_{13}

The chemical shift differences between TIS11d and TIS11d-U_{13} highlight the changes that occur to TIS11d upon binding U_{13}**, Figure 3.4(B)**. The observed differences suggest that the complex of TIS11d with U_{13} maintains some of the structural features present in the TIS11d-ARE complex. Chemical shift changes occur throughout ZF1, ZF2 and the linker upon binding weakly to U_{13}.

Residues Y208 and E195 of TIS11d in ZF2 and the KYKTEL sequence preceding ZF2 respectively undergo the largest change in chemical shift upon binding U_{13}, **Figure 3.4(B)**. The chemical shifts of these residues overlap with the chemical shifts of TIS11d-$ARE_{13.}$ In the TIS11d-ARE_{13} conformation of TIS11d the backbone amide of E195 hydrogen bonds to U4 and the side-chain of Y208 intercalates between U6 and U7. This suggests that binding to U4 and coordinating U6 and U7 by TIS11d-U_{13} is similar to the base coordination in TIS11d-$ARE_{13.}$

The structure of TIS11d when bound to ARE_{13} is stabilized by intramolecular H-bonds and by a hydrophobic core made of hydrophobic residues of the linker and ZF1.(*17*) Upon binding U_{13} the largest measurable chemical shift differences occur for residues L158, K173, Y192, E195, Y208, and I215**, Figure 3.4(B)**. The chemical shift of these important residues is labeled in the $^{1}H^{15}N$-HSQC spectra shown in **Figure 3.5**. These residues are involved in intramolecular H-bonds or are part of they hydrophobic cluster that stabilizes the structure of TIS11d in the TIs11d-ARE_{13} complex.

Figure 3.5: Amide backbone chemical shifts of TIS11d bound to polyurdine. The ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) spectrum of TIS11d-5′-UUUUUUUUUUUUU-3′ is red, free TIS11d is blue, and TIS11d-ARE_{13} is black. Grey bars indicate large chemical shift differences between TIS11d (blue) and TIS11d-5′-UUUUUUUUUUUUU-3′ (red); for these peaks the TIS11d-5′-UUUUUUUUUUUUU-3′ cross-peaks are labeled.

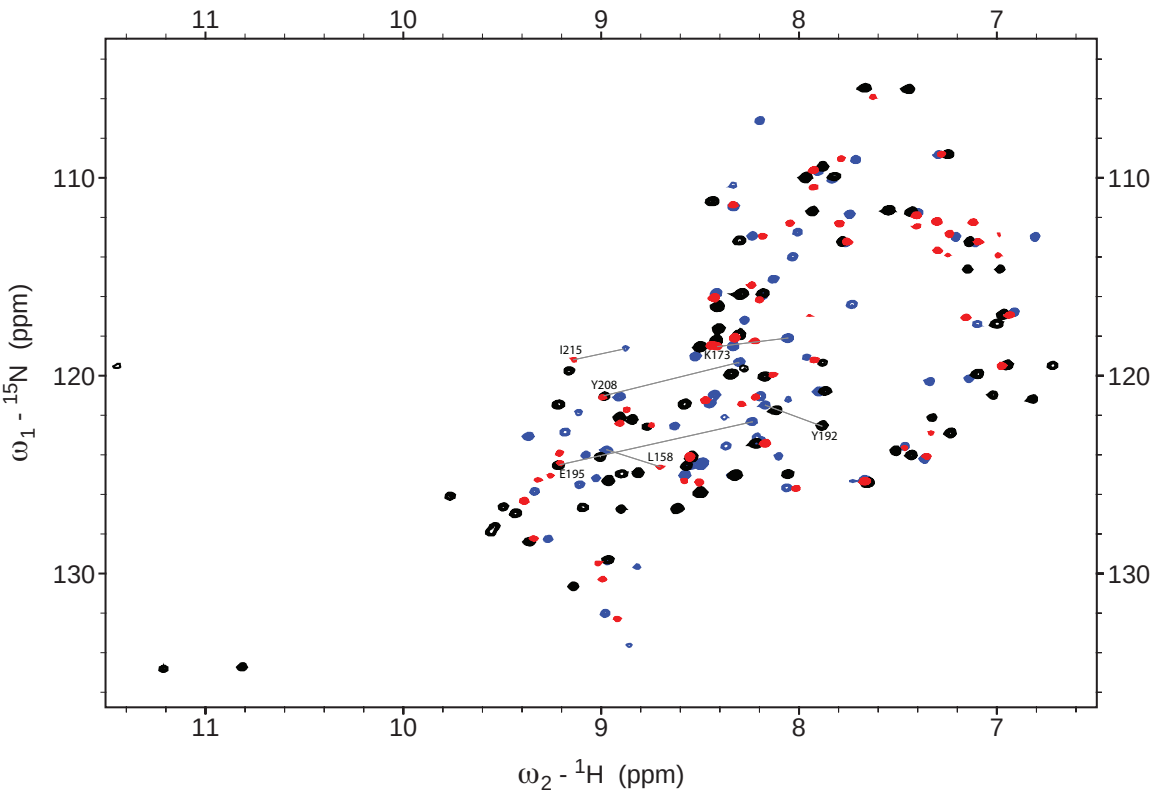

11
10
9
8
7
110
120
130
ω1 - 15N (ppm)
ω2 - 1H (ppm)
I215
K173
Y208
Y192
L158
E195

Figure 3.5

In particular, Y192 is part of a hydrogen-bonding network that stabilizes the C-terminus of TIS11d.(*17*) Y208 hydrogen bonds to the sulfur of C212 and I215 forms a hydrogen bond with the carbonyl of E195.(*17*) The side-chain of L158 participates in this hydrophobic core.(*17*) The chemical shift of K173 is unclear, Wright et al hypothesize it may be sensitive to the RNA phosphate backbone.(*17*)

The presence of large chemical shift differences between the spectra of TIS11d-U_{13} and free TIS11d suggests that many of the structural features of TIS11d-ARE_{13} are conserved in the TIS11d-U_{13} complex. However, not all of the chemical shift changes observed upon addition of ARE_{13} are present. For, instance backbone amides of residues K155 and F162 undergo a smaller transition upon binding U_{13} than upon binding ARE_{13}. This suggests that hydrogen bonds are absent, suboptimal or that the conformation of TIS11d is different at the N-terminus.

Differences between TIS11d-ARE_{13} and TIS11d-U_{13}, **Figure 3.4 (C)**, highlight differences between the spectra. Residues R160, Q175 and R198 undergo large changes in chemical shifts upon binding ARE_{13}. These changes are diminished upon binding TIS11d-U_{13}. Upon binding ARE_{13} backbone amides of R160, R198 and Q175 form hydrogen bonds to RNA. R160 and R198 form hydrogen bond to the A9 and A5, respectively and Q175 forms a hydrogen bond to U10. These diminished changes in chemical shifts suggest that hydrogen bonding to U9, U5, and U10 may be absent or weakened and that the chemical environment near these residues is different. It is also possible that TIS11d does form hydrogen bonds with these bases and that TIS11d is characterized by an

equilibrium between two or more conformational states. In general the chemical shifts of TIS11d-U_{13} are mores similar to free TIS11d than to the chemical shifts of TIS11d-ARE_{13}, **Figure 3.4 (B, C)**. Upon binding U_{13} backbone amides of TIS11d undergo fewer and smaller chemical shifts. We hypothesize that U_{13} promotes a transition to a conformation that is more compact than the unbound state but that is more dynamic than the ARE-bound state and that only some of the H-bonds and hydrophobic interactions known to stabilize the TIS11d-ARE complex are present.

Chemical shift changes in TIS11d associated with high affinity binding to $FM3_{13}$

5′-UUUUUUUUAUUUU-3′ has one adenine at position nine; this adenine is coordinated by ZF1. Upon binding 5′-UUUUUUUUAUUUU-3′, TIS11d undergoes moderate chemical shift changes throughout ZF1, ZF2 and linker, **Figure 3.6 (A)**. The chemical shift changes at ZF1 are larger than the chemical shift changes at ZF2; this suggests that ZF1 binds the 5′-UUAU-3′ and ZF2 binds 5′-UUUU-3′. This is consistent with 3′ to 5′ RNA recognition observed in the TIS11d-ARE NMR structure. The largest measured chemical shifts TIS11d undergoes upon binding 5′-UUUUUUUUAUUUU-3′ is at residues F162 and T194, **Figure 3.6 (A)**. In the TIS11d-ARE structure, the backbone amide of F162 forms a hydrogen bond to the sulfur of C159 and the backbone amide of T194 forms a hydrogen bond to the RNA base U8, **Figure 3.7**.

The chemical shifts of ZF2 of TIS11d-5′-UUUUUUUUAUUUU-3′ is nearly identical to the chemical shifts of TIS11d-U_{13}, **Figure 3.6 (B)**. The most significant chemical shift difference between TIS11d-5′-UUUUUUUUAUUUU-3′ and TIS11d-U_{13} is at residue F162. The NMR structure of TIS11d-ARE_{13} shows that the backbone amide of F162 forms a hydrogen bond to the sulfur of C159. The observed chemical shift differences suggest that this bond is present in TIS11d-5′-UUUUUUUUAUUUU-3′ and it is either absent or weakened in TIS11d-U_{13}. Perhaps the N-terminus of TIS11d-5′-UUUUUUUUAUUUU-3′ is more stable than TIS11d-U_{13}. In ZF2 residues R198 of TIS11d-5′-UUUUUUUUAUUUU-3′ is different than TIS11d-ARE_{13}, **Figure 3.6 (C)**. R198 forms a hydrogen bond to A5 in the TIS11d-ARE_{13} structure. ZF2 of the TIS11d-5′-UUUUUUUUAUUUU-3′

complex binds 5′-UUUU-3′ therefore the backbone amide R198 cannot form a hydrogen bond to an adenine.

Figure 3.6: Amide backbone chemical shift difference analysis of TIS11d-5′-UUUUUUUUAUUUU-3′ Atop each plot is a schematic representation of the TZD of TIS11d; the black horizontal bars represent ZF1 and ZF2, the blue horizontal bars are the preceding conserved RYKTEL and KYKTEL sequences, white circles represent the CCCH residues of ZF1 and ZF2, the red circles represent residues whose side chains form base stacks with RNA bases in the TIS11d-ARE_{13} complex, black dots denote a residue whose backbone amide hydrogen bonds to a RNA base in the TIS11d-ARE_9 complex. For each plot, black bars designate the magnitude of amide backbone chemical shift differences of TIS11d between states. A) Chemical shift differences between free TIS11d and TIS11d in complex with the high affinity sequence $FM3_3$, (Δδ =TIS11d – TIS11d-5′-UUUUUUUUAUUUU-3′) B) The chemical shift difference between TIS11d-5′-UUUUUUUUAUUUU-3′ and TIS11d-U_{13}, (Δδ =|TIS11d-5′-UUUUUUUUAUUUU-3′ – TIS11d-U_{13}|). This highlights the TIS11d chemical shift differences between the high affinity binding to 5′-UUUUUUUUAUUUU-3′ and the low affinity binding to $U_{13.}$ C) The chemical shift differences of TIS11d-5′-UUUUUUUUAUUUU-3′ and TIS11d bound to ARE_{13}, (Δδ =|TIS11d-5′-UUUUUUUUAUUUU-3′ – TIS11d-ARE_{13}|). This figure highlights the residues of TIS11d in complex with 5′-UUUUUUUUAUUUU-3′ that are different from the highest affinity binding to the cognate ARE_{13}.

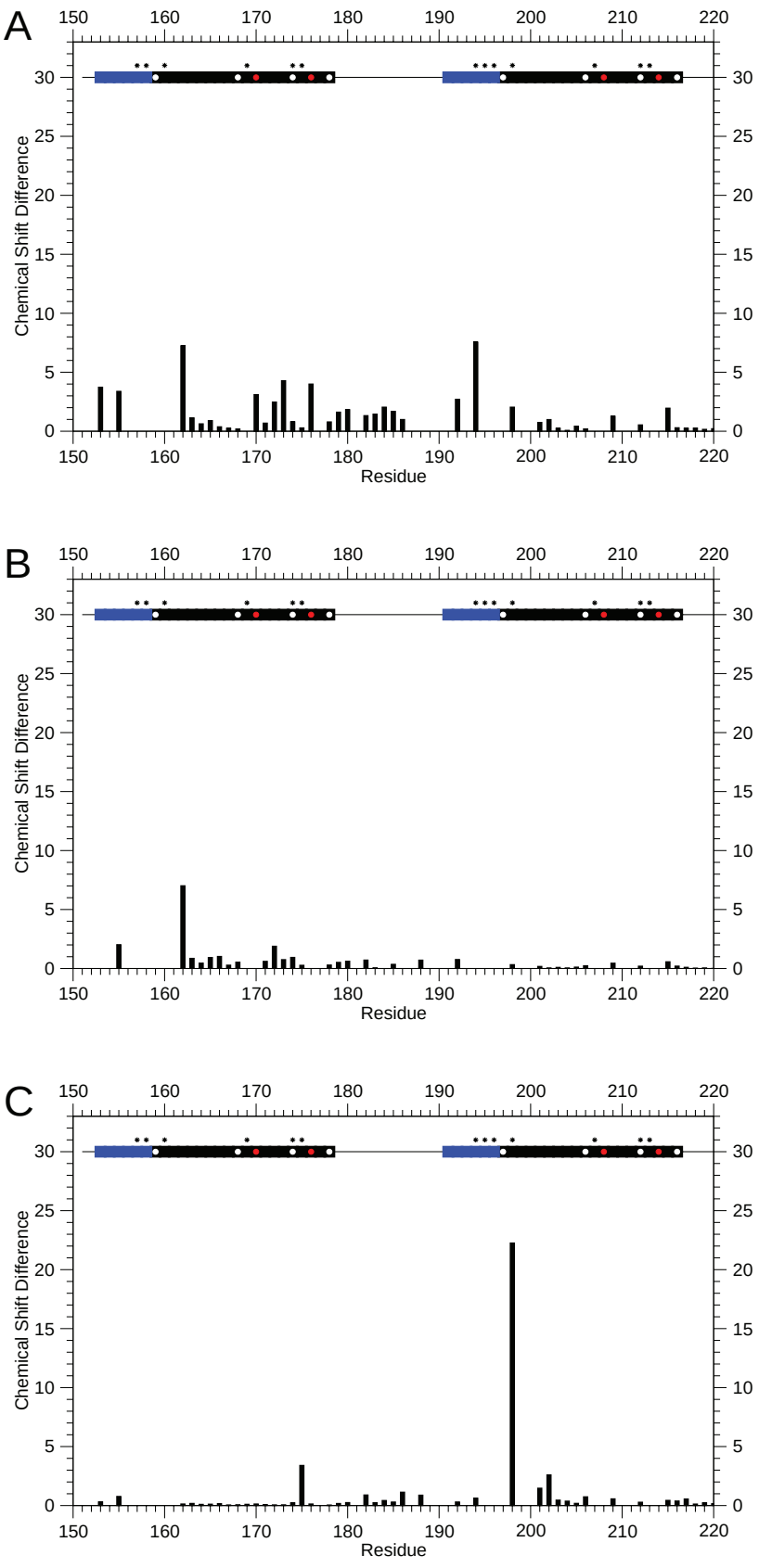

A
B
C
Chemical Shift Difference
Residue
150
160
170
180
190
200
210
220
0
5
10
15
20
25
30

Figure 3.6

Figure 3.7: Amide backbone chemical shifts of TIS11d bound to 5′-UUUUUUUUAUUUU-3′. The ^{1}H-^{15}N HSQC spectrum of TIS11d-5′-UUUUUUUUAUUUU-3′ is purple, free TIS11d is blue, and TIS11d-ARE$_{13}$ is black. Grey bars indicate large chemical shift differences between TIS11d (blue) and TIS11d-5′-UUUUUUUUAUUUU-3′ (purple); for these peaks the TIS11d-5′-UUUUUUUUAUUUU-3′ cross-peaks are labeled.

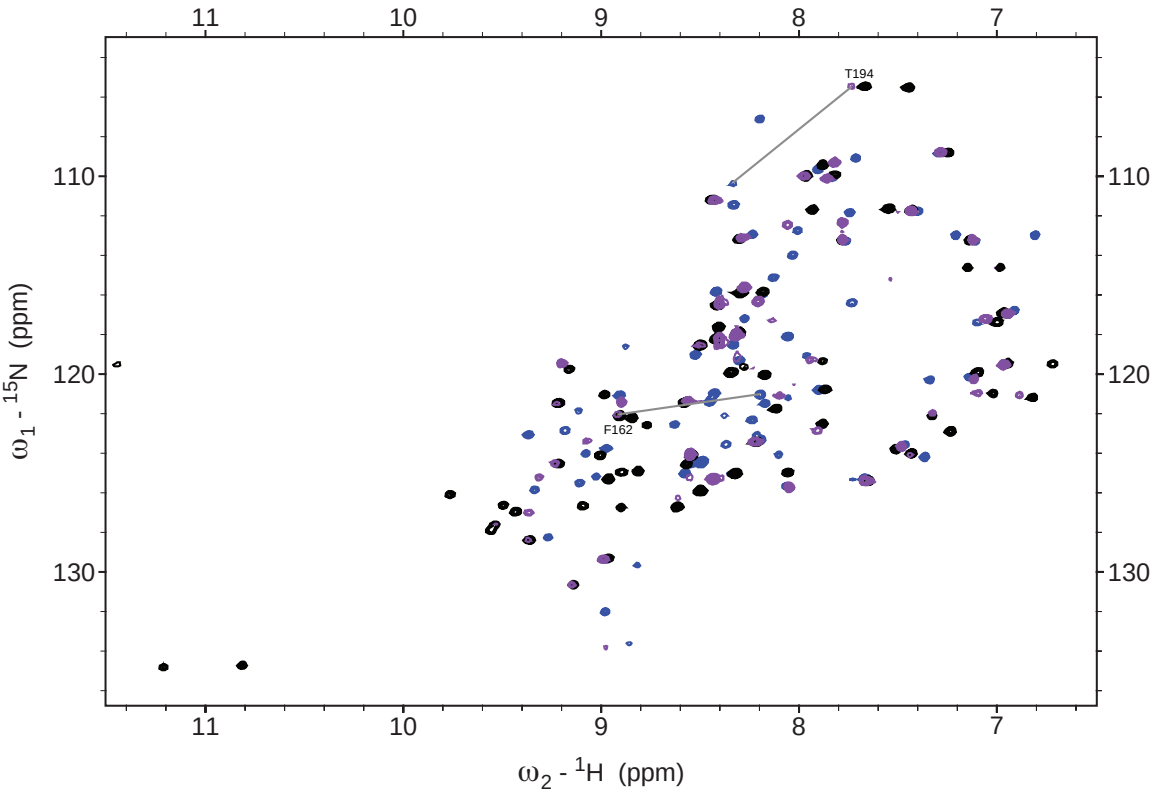


Figure 3.7

The chemical shifts of ZF2 of TIS11d-5′-UUUUUUUUAUUUU-3′ are nearly identical to the chemical shifts of TIS11d-ARE_{13}, **Figure 3.6 (C)**, except for residue Q175. In the NMR structure of TIS11d-ARE_{13}, the backbone amide of Q175 forms a hydrogen bond to U10. This data suggests that ZF1 binds to 5′-UUAU-3′ without forming a hydrogen bond between Q175 and U10 or the nature of this interaction at this location is altered. The backbone amide of residue R160 hydrogen bonds to A9 and undergoes one of the largest chemical shift changes upon binding ARE_{13}, this peak cannot be assigned and does not overlap with the cross-peak location of free TIS11d or TIS11d-$ARE_{13.}$ This suggests that this hydrogen bond is perturbed or absent in the TIS11d-5′-UUUUUUUUAUUUU-3′ complex.

Although the chemical shifts of ZF1 are most similar to the chemical shifts of TIS11d-ARE_{13}, Q175 and R160 are sensitive to the particular RNA sequence bound to ZF2 and may be important in determining the different binding affinity of TIS11d for the ARE_{13} and 5′-UUUUUUUUAUUUU-3′ sequences.

Chemical shift changes in TIS11d associated with high affinity binding to $FM4_{13}$

Upon binding 5′-UUUUAUUUUUUUU-3′, TIS11d undergoes minor chemical shift changes throughout ZF1, ZF2 and the linker, **Figure 3.8 (A)**. The largest chemical shift difference is at residue Y208. In the TIS11d-ARE_{13} NMR structure, Y208 forms an intramolecular stabilizing hydrogen bond with the sulfur of C206. The chemical shift differences of TIS11d- 5′-UUUUAUUUUUUUU-3′ are nearly identical and mostly overlap with TIS11d-U_{13}, **Figure 3.8 (B)** and **Figure 3.9 (B).** There are no large chemical shift differences between the two complexes; this suggests that TIS11d-5′-UUUUAUUUUUUUU-3′ adopts a similar conformation as TIS11d-U_{13}. Like TIS11d-U_{13}, TIS11d-5′-UUUUAUUUUUUUU-3′ is substantially different from TIS11d-ARE_{13}, **Figure 3.8 (C)** and **Figure 3.9 (A)**. Surprisingly, even though ZF2 binds to the sequence element 5′-UAUU-3′ the preceding KYKTEL sequence and ZF2 are more similar to TIS11d-U_{13} than to TIS11d-ARE_{13}. Residues in this region are missing from the TIS11d-5′-UUUUAUUUUUUUU-3′ $^{1}H^{15}N$-HSQC spectrum; these peaks are present in the TIS11d-U_{13} spectrum, **Figure 3.9 (B)**. This suggests that this finger may be flexible and is undergoing a chemical exchange process.

The backbone chemical shifts observed for TIS11d-5′-UUUUAUUUUUUUU-3′ suggests that although an adenine is present at ZF2, TIS11d does not bind it like TIS11d-$ARE_{13.}$ ZF2 of TIS11d binds adenine as in the high affinity recognition of ARE_{13} only if an adenine is also present at ZF1 suggesting the existence of communication between ZF1 and ZF2.

Figure 3.8: Amide backbone chemical shift difference analysis of TIS11d-5′-UUUUAUUUUUUUU-3′. Atop each plot is a schematic representation of the TZD of TIS11d; the black horizontal bars represent ZF1 and ZF2, the blue horizontal bars are the preceding conserved RYKTEL and KYKTEL sequences, white circles represent the CCCH residues of ZF1 and ZF2, the red circles represent residues whose side chains form base stacks with RNA bases in the TIS11d-ARE_{13} complex, black dots denote a residue that hydrogen bonds to a RNA base in the TIS11d-ARE_{13} complex. For each plot, black bars designate the magnitude of amide backbone chemical shift differences of TIS11d between states. A) Chemical shift differences between free TIS11d and TIS11d in complex with the high affinity sequence $FM4_3$, (Δδ =TIS11d – TIS11d-5′-UUUUAUUUUUUUU-3′). B) The chemical shift difference between TIS11d-5′-UUUUAUUUUUUUU-3′ and TIS11d-U_{13}, (Δδ =|TIS11d-5′-UUUUAUUUUUUUU-3′ – TIS11d-U_{13}|). This highlights the TIS11d chemical shift differences between the high affinity binding to 5′-UUUUAUUUUUUUU-3′ and the low affinity binding to U_{13}. C) The chemical shift differences of TIS11d-5′-UUUUAUUUUUUUU-3′ and TIS11d bound to ARE_{13}, (Δδ =|TIS11d-5′-UUUUAUUUUUUUU-3′ – TIS11d-ARE_{13}|). This figure highlights the residues of TIS11d in complex with 5′-UUUUAUUUUUUUU-3′ that are different from the highest affinity binding to the cognate ARE_{13}.

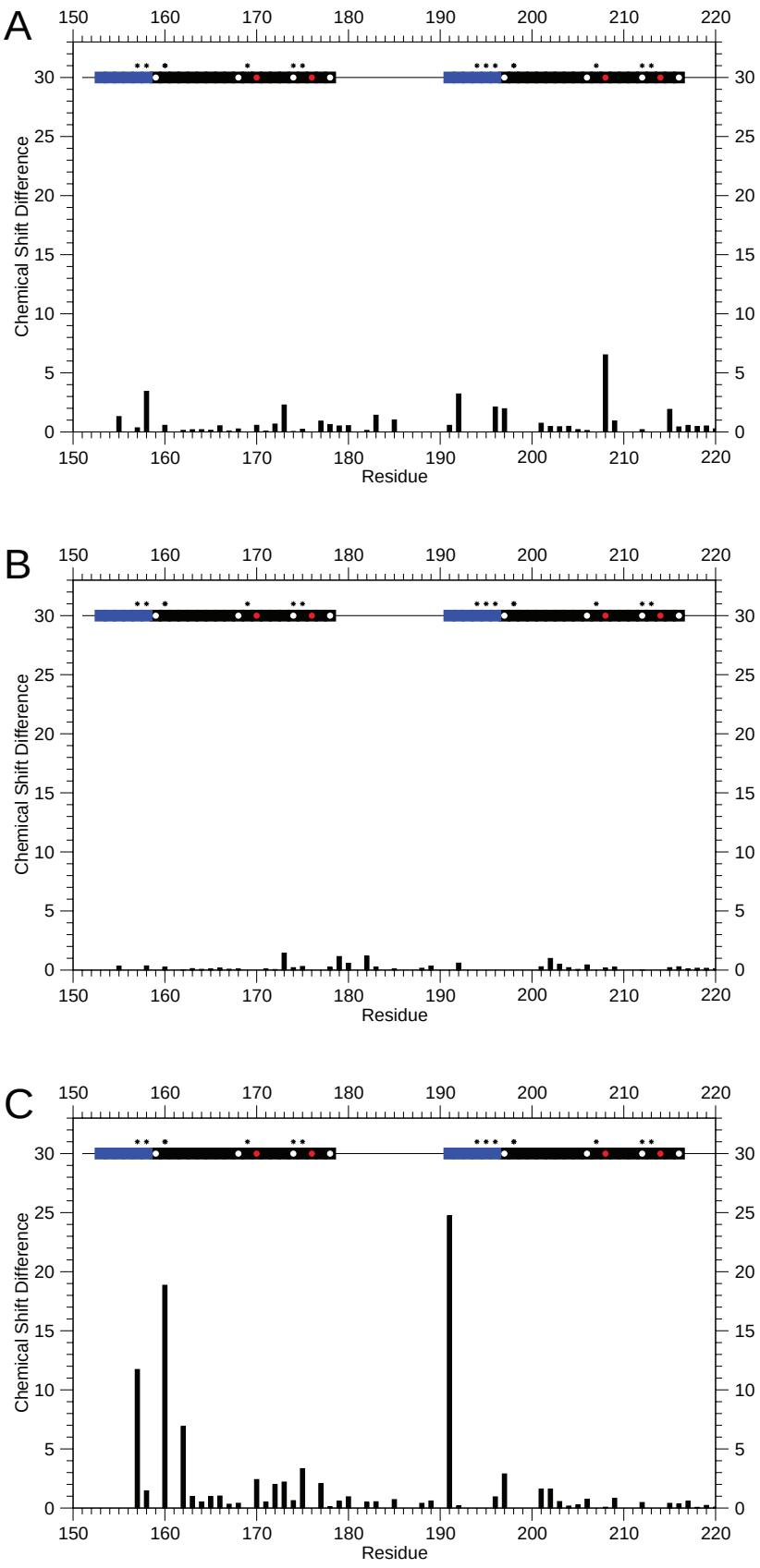

A
B
C
Chemical Shift Difference
Residue
150
160
170
180
190
200
210
220
0
5
10
15
20
25
30

Figure 3.8

Figure 3.9: Amide backbone chemical shifts of TIS11d bound to 5′-UUUUAUUUUUUUU-3′. A) The ^{1}H-^{15}N HSQC spectrum of TIS11d-5′-UUUUUUUUAUUUU-3′ is green, free TIS11d is blue, and TIS11d-ARE_{13} is black. B) The ^{1}H-^{15}N HSQC spectrum of TIS11d-5′-UUUUAUUUUUUUU-3′ is green, and TIS11d-U_{13} is red.

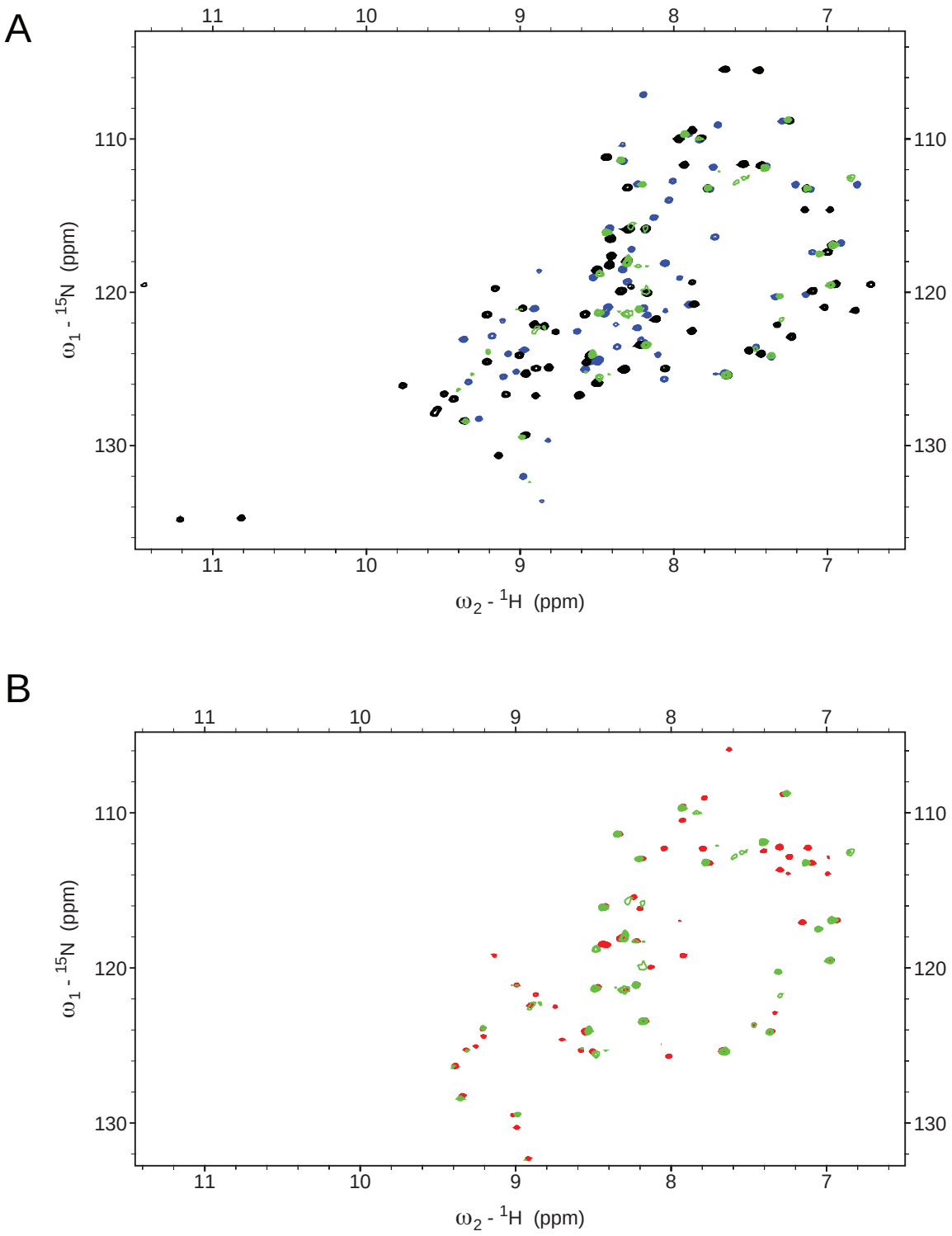

A
11
10
9
8
7
110
120
130
ω1 - 15N (ppm)
ω2 - 1H (ppm)
B
11
10
9
8
7
110
120
130
ω1 - 15N (ppm)
ω2 - 1H (ppm)

Figure 3.9

TIS11d-5′-UUUUAUUUUUUUU-3′ and U_{13}-TIS11d have similar chemical shifts, however, TIS11d has a high affinity for 5′-UUUUAUUUUUUUU-3′ but not for U_{13}. TIS11d binds 5′-UUUUAUUUUUUUU-3′ as tightly as 5′-UUUUUUUUAUUUU-3′ regardless of the favorable interaction associated with binding to an adenine at ZF1. Changes in the conformational entropy of TIS11d could drive the high affinity binding to 5′-UUUUAUUUUUUUU-3′. The backbone chemical shifts of TIS11d bound to ARE_{13}, 5′-UUUUUUUUAUUUU-3′ and 5′-UUUUAUUUUUUUU-3′ are different. Although we cannot identify the particular structural differences that result in the observed differences in backbone resonance frequencies for these TIS11d complexes, we can conclude that TIS11d binds these RNA sequences in different conformations.

Tis11d recognizes RNAs through different mechanisms

TIS11d requires an intermediate to bind ARE_{13} but not polyuridine. Binding to ARE_{13} is slow on the chemical shift timescale. At intermediate concentrations of RNA amide cross-peaks exist for both free and bound states of TIS11d, **Figure 3.10 (A)**. At substoichiometric concentrations of TIS11d (2% - 17% RNA : TIS11d), amide peaks of TIS11d migrates by fast exchange to an intermediate state. During this process cross-peaks do not shift towards the cross-peak position of TIS11d-ARE_{13}; this suggests that TIS11d access an intermediate encounter complex to bind ARE_{13}. Example shifts of cross-peaks to an intermediate are magnified within **Figure 3.10 (B)**.

Binding to U_{13} is slow on the chemical shift timescale. At intermediate concentrations of RNA amide cross-peaks exist for both states of TIS11d. TIS11d does not access an intermediate upon binding $U_{13.}$ The only exception in this case is for residue E182. E182 undergoes fast exchange at substoichiometric concentrations of U_{13} : TIS11d (0%-25%). E182 is part of the linker of TIS11d, and it is not clear what the role of this residue is in TIS11d specificity.

Titrations of TIS11d with 5′-UUUUUUUUAUUUU-3′ and 5′-UUUUAUUUUUUUU-3′ show that TIS11d recognizes RNA by intermediate exchange with respects to chemical shift and transitions to an intermediate are not observed in either case, **Table 3.2**. However, the TIS11d-ARE_{13} intermediate is an important step in binding $ARE_{13.}$

Figure 3.10: TIS11d accesses an intermediate to bind ARE_{13}. A) TIS11d is titrated with ARE_{13} (5′-UUUUAUUUAUUUU-3′), chemical shifts are determined by ^{1}H-^{15}N HSQCs. TIS11d binds ARE_{13} by accessing an intermediate and then transitioning to the high affinity complex. Fast chemical exchange of TIS11d on the chemical shift time-scale provide evidence of an intermediate. Cross-peaks of substoichiometric concentrations of TIS11d:ARE_{13} are colored as follows: 0% red, 3% orange, 7% yellow, 10% green and 13% blue. The final position of TIS11d-ARE_{13} is black. The trajectory of cross-peaks of TIS11d in fast exchange, are not in-line with the final positions of TIS11d-ARE_{13} cross-peaks. Example cross-peaks undergoing fast exchange are labeled in A and enlarged in B. For all other subsaturating stoichiometric concentrations of RNA_{13} cross-peaks of both free and bound TIS11d are present within the spectrum; this is typical of a slow exchange process with respects to chemical shift, and reflect a second transition. The chemical shift analysis if TIS11d with ARE_{13} suggests that TIS11d forms an intermediate encounter complex that is required to form the final TIS11d-ARE_{13} conformation.

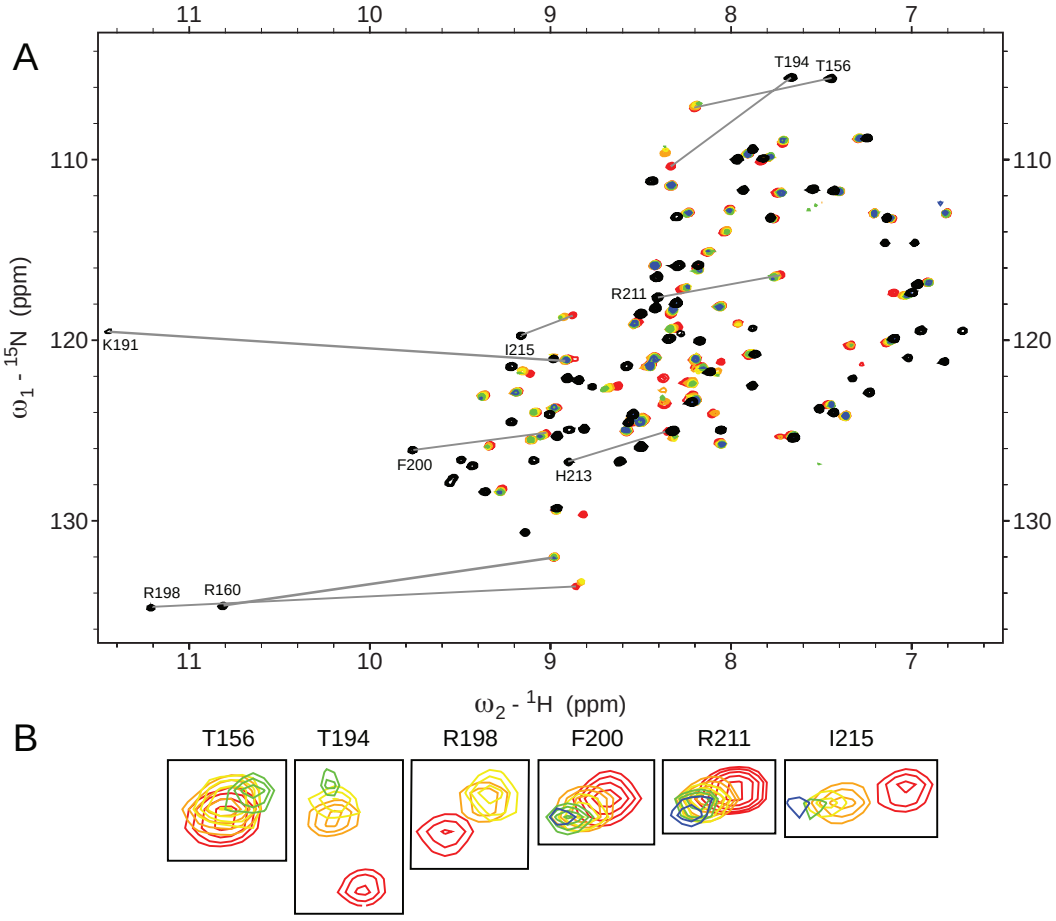

A
B
11
10
9
8
7
110
120
130
T194
T156
R211
K191
I215
F200
H213
R198
R160
ω1 - 15N (ppm)
ω2 - 1H (ppm)
T156
T194
R198
F200
R211
I215

Figure 3.10

Table 3.2: The exchange of ARE-like sequences. In this table RNA sequences are listed 5′ to 3′. The exchange processes are defined with respect to chemical shift.

Table 3.2: The exchange of ARE-like sequences.

RNA	Sequence	Exchange
ARE_{13}	UUU**UAUU**UAUUUU	Fast, Slow
$FM3_{13}$	UUU**UUUU**UAUUUU	Intermediate
$FM4_{13}$	UUU**UAUU**UUUUUU	Intermediate
ARE_{7}	UUU**UAUU**_______	
U_{13}	UUU**UUUU**UUUUUU	Slow

DISCUSSION

Mutations in the TIS11d gene have recently been linked to patients with acute myeloid leukemia.(*2*) However, the activity of TIS11d in the cell is unknown. In this study, gel mobility shift assays reveal that TIS11d is most selective for the ARE_{13} sequence. However, TIS11d also binds to the 5′-UUUUUUUUAUUUU-3′ and 5′-UUUUAUUUUUUUU-3′ sequences with high affinity. Given the high affinity of TIS11d binding to ARE_{13}, 5′-UUUUUUUUAUUUU-3′ and 5′-UUUUAUUUUUUUU-3′, each of these RNA sequences may be biologically important.

The backbone chemical shifts of TIS11d bound to U_{13} are indicative of non-specific binding of RNA. Low affinity recognition of polyuridine may be useful for scanning mRNA sequences for high affinity targets. The ability to bind non-specifically to mRNA will enhance the speed that TIS11d may locate a relevant mRNA target.

TIS11d is highly selective for UAUU sequence within a polyuridine stretch. However, ZF1 and ZF2 of TIS11d bind UAUU sequences directly but both ZF1 and ZF2 are sensitive to the RNA sequence bound by the other zinc finger. Q175 of ZF1 only undergoes a large chemical shift change upon binding $ARE_{13.}$ ZF2 only undergoes large chemical shifts in the context of full-length ARE_{13}. The full-length RNA sequence sensitivity of TIS11d may not be present in all of the TTP-like proteins. If ZF1 and ZF2 of TTP or TIS11b bind UAUU sequences independently with high affinity then these proteins could recognize shorter

sequences in vivo. This could also differentiate the specificity of each TTP-like protein for ARE sequences in mRNAs.

Amide chemical shifts of TIS11d in complex with each of these RNAs suggest that the conformation of TIS11d is dependent on the sequence it binds. If the ARE_{13} sequence is the only biologically significant target sequence, then adaptor proteins may recognize and select for that unique conformation. The conformation of TIS11d for each of these high affinity targets may be an indication of the mRNA sequence and could direct down stream effects.
The promiscuous RNA binding proteins human PUMILIO1 and PUMILIO2 bind different RNA sequences in three distinct conformations with high RNA binding affinity.(*23*) The conformation of protein-RNA complexes have an effect *in vivo*, especially if adaptor proteins that bind the protein-RNA complex select for a particular conformational mode.(*24*)

This study and the conformational studies on PUMILIO proteins highlight the importance of conformational analysis of promiscuous proteins. Studies of the DNA binding by CCHH type zinc fingers have allowed for zinc finger base selective engineering. In these studies molecular recognition is driven by the individual amino acids in direct contact with each base.(*25*) These studies suggest that CCHH type zinc fingers are modular, and the binding specificity of zinc finger chimeras exemplify CCHH type zinc finger modularity.(*25*) Our studies of TIS11d indicate that the CCCH zinc fingers of TIS11d are not modular as the RNA sequence recognized by one finger affects the RNA binding activity of the other zinc finger.

We hypothesize that the high affinity binding of TIS11d to 5′-UUUUAUUUUUUUU-3′ is entropically driven. Determination of the structure and flexibility of TIS11d in each of these complexes may provide insight into the mechanism of RNA recognition by TIS11d.

ACKNOWLEDGEMENTS

I would like to thank Sean Ryder and Brian Farley for helpful discussions.

REFERENCES:

1. Stumpo, D. J., Broxmeyer, H. E., Ward, T., Cooper, S., Hangoc, G., Chung, Y. J., Shelley, W. C., Richfield, E. K., Ray, M. K., Yoder, M. C., Aplan, P. D., and Blackshear, P. J. (2009) Targeted disruption of Zfp36l2, encoding a CCCH tandem zinc finger RNA-binding protein, results in defective hematopoiesis, *Blood 114*, 2401-2410.
2. Iwanaga, E., Nanri, T., Mitsuya, H., and Asou, N. (2011) Mutation in the RNA binding protein TIS11D/ZFP36L2 is associated with the pathogenesis of acute leukemia, *Int J Oncol 38*, 25-31.
3. Hodson, D. J., Janas, M. L., Galloway, A., Bell, S. E., Andrews, S., Li, C. M., Pannell, R., Siebel, C. W., MacDonald, H. R., De Keersmaecker, K., Ferrando, A. A., Grutz, G., and Turner, M. (2010) Deletion of the RNA-binding proteins ZFP36L1 and ZFP36L2 leads to perturbed thymic development and T lymphoblastic leukemia, *Nat Immunol 11*, 717-724.
4. Falini, B., Martelli, M. P., Bolli, N., Sportoletti, P., Liso, A., Tiacci, E., and Haferlach, T. (2011) Acute myeloid leukemia with mutated nucleophosmin (NPM1): is it a distinct entity?, *Blood 117*, 1109-1120.
5. Soria, G., and Gottifredi, V. (2010) PCNA-coupled p21 degradation after DNA damage: The exception that confirms the rule?, *DNA Repair (Amst) 9*, 358-364.
6. Johnson, B. A., and Blackwell, T. K. (2002) Multiple tristetraprolin sequence domains required to induce apoptosis and modulate responses to TNFalpha through distinct pathways, *Oncogene 21*, 4237-4246.
7. Lai, W. S., Carballo, E., Thorn, J. M., Kennington, E. A., and Blackshear, P. J. (2000) Interactions of CCCH zinc finger proteins with mRNA. Binding of tristetraprolin-related zinc finger proteins to Au-rich elements and destabilization of mRNA, *J Biol Chem 275*, 17827-17837.
8. Johnson, B. A., Geha, M., and Blackwell, T. K. (2000) Similar but distinct effects of the tristetraprolin/TIS11 immediate-early proteins on cell survival, *Oncogene 19*, 1657-1664.
9. Lin, N. Y., Lin, T. Y., Yang, W. H., Wang, S. C., Wang, K. T., Su, Y. L., Jiang, Y. W., Chang, G. D., and Chang, C. J. (2012) Differential Expression and Functional Analysis of the Tristetraprolin Family during Early Differentiation of 3T3-L1 Preadipocytes, *Int J Biol Sci 8*, 761-777.
10. Ramos, S. B. (2012) Characterization of DeltaN-Zfp36l2 mutant associated with arrest of early embryonic development and female infertility, *J Biol Chem 287*, 13116-13127.
11. Deleault, K. M., Skinner, S. J., and Brooks, S. A. (2008) Tristetraprolin regulates TNF TNF-alpha mRNA stability via a proteasome dependent mechanism involving the combined action of the ERK and p38 pathways, *Mol Immunol 45*, 13-24.
12. Sandler, H., and Stoecklin, G. (2008) Control of mRNA decay by phosphorylation of tristetraprolin, *Biochem Soc Trans 36*, 491-496.

13. Desroches-Castan, A., Cherradi, N., Feige, J. J., and Ciais, D. (2011) A novel function of Tis11b/BRF1 as a regulator of Dll4 mRNA 3'-end processing, *Mol Biol Cell 22*, 3625-3633.
14. Ciais, D., Cherradi, N., Bailly, S., Grenier, E., Berra, E., Pouyssegur, J., Lamarre, J., and Feige, J. J. (2004) Destabilization of vascular endothelial growth factor mRNA by the zinc-finger protein TIS11b, *Oncogene 23*, 8673-8680.
15. Blackshear, P. J., Lai, W. S., Kennington, E. A., Brewer, G., Wilson, G. M., Guan, X., and Zhou, P. (2003) Characteristics of the interaction of a synthetic human tristetraprolin tandem zinc finger peptide with AU-rich element-containing RNA substrates, *J Biol Chem 278*, 19947-19955.
16. Farley, B. M., Pagano, J. M., and Ryder, S. P. (2008) RNA target specificity of the embryonic cell fate determinant POS-1, *RNA 14*, 2685-2697.
17. Hudson, B. P., Martinez-Yamout, M. A., Dyson, H. J., and Wright, P. E. (2004) Recognition of the mRNA AU-rich element by the zinc finger domain of TIS11d, *Nat Struct Mol Biol 11*, 257-264.
18. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes, *J Biomol NMR 6*, 277-293.
19. Goddard, T. D., and D.G. Kneller. . SPARKY 3, University of California, San Francisco, CA.
20. DeLano, W. L. (2008) The PyMOL molecular graphics system., (LLC, D. S., Ed.), San Carlos, CA.
21. Stambulchik, E. (1996-2007) Grace Development Team.
22. Brewer, B. Y., Malicka, J., Blackshear, P. J., and Wilson, G. M. (2004) RNA sequence elements required for high affinity binding by the zinc finger domain of tristetraprolin: conformational changes coupled to the bipartite nature of Au-rich MRNA-destabilizing motifs, *J Biol Chem 279*, 27870-27877.
23. Lu, G., and Hall, T. M. (2011) Alternate modes of cognate RNA recognition by human PUMILIO proteins, *Structure 19*, 361-367.
24. Ryder, S. P. (2012) Pumilio RNA recognition: the consequence of promiscuity, *Structure 19*, 277-279.
25. Wolfe, S. A., Nekludova, L., and Pabo, C. O. (2000) DNA recognition by Cys2His2 zinc finger proteins, *Annu Rev Biophys Biomol Struct 29*, 183-212.

CHAPTER IV

DISCUSSION

THESIS IMPLICATIONS AND FUTURE DIRECTIONS

Protein flexibility contributes to protein function

Proteins are flexible molecules that often undergo conformational changes to perform their biological functions including catalysis, signal transduction, metabolism, transcription and translation. NMR spectroscopy is a powerful technique that can monitor protein motions over a broad range of time scales.

This thesis focuses on understanding the role of protein dynamics on allostery and molecular recognition, two essential functions of proteins that permit fine-tuning of a variety of essential processes in the cell. In particular, the fast time scale (ps-ns) dynamics of HbI were measured. This study showed that the flexibility of HbI is important in both the intersubunit communication as well as in the thermodynamics of cooperativity. Our NMR spectroscopic studies of TIS11d free and bound to four different RNA sequences suggest that structure and dynamics of TIS11d contribute to TIS11d binding specificity.

Fast protein motions and conformational entropy contribute to protein function

In ligand binding processes, changes in flexibility, due to a change in the number of states accessible to a protein (change in conformational entropy) can make a large contribution to the free energy of binding. Historically it has been difficult to estimate protein conformational entropy. The development of NMR relaxation experiments and their recent applications to study fast (ps-ns) protein dynamics, however, have highlighted the importance of changes in protein flexibility in ligand binding processes.(*1-3*) Therefore, studies of protein dynamics

in the ps-ns time scale are a critical component to understand cooperativity. In this study CO-HbI is shown to be more flexible than unliganded HbI while the opposite is truefor the less cooperative mutant (F97Y); CO-F97Y is more rigid than unliganded F97Y. These results are strong evidence that HbI modulates its flexibility to bind ligand cooperatively. The cooperativity of HbI has been shown to be entropically driven.(*4*) This study highlights the importance of evaluating the protein conformational entropy contribution to cooperativity and to ligand binding processes.

This is the first study of a cooperative protein that has significant changes in flexibility at the inter-subunit interface that is necessary for cooperativity, suggesting that changes in fast time scale motions are involved in the cooperative communication between monomers. These motions may be coupled to changes in water dynamics at the interface; therefore, this system is an ideal model for studying protein-water communication. Characterizing the fast time scale motions of mutants of HbI with altered cooperativity which either mimic the low affinity state or disrupt the interfacial water cluster (F97L, T72V and I114F), will reveal the balance between structural and dynamical changes in cooperativity. Previous studies have shown that the dynamic behavior of the side chains can be more complex and heterogeneous than that of the backbone.(*5*) Therefore, the potential entropic contribution of HbI side chains to cooperative binding is expected to be large.(*5*) Because our studies have used only the backbone dynamics to characterize the protein conformational entropy changes, determination of methyl side-chain dynamics of HbI and HbI mutants will provide

a more complete understanding of the dynamic contributions to HbI cooperativity. Molecular dynamic simulations will be used to understand the details of the coupling between changes in protein and water structures and dynamics in HbI cooperativity. Future molecular dynamic simulations can be used to measure changes in the total solvent accessible surface area (SASA) and estimate the change in entropy of the solvent.(*6*) By comparing simulations of CO bound and unliganded HbI changes in SASA estimates of the entropic contribution of water to cooperative binding can be obtained.

Studies of the fast time scale motions of TIS11d will be important to estimate the contribution of the change in conformational entropy to RNA binding, and these studies will provide insights into the combined role of protein structure and dynamics in controlling RNA binding affinity and specificity. In this study, TIS11d has been shown to bind three different RNAs with high affinity, and our data suggest that TIS11d may bind them in three different conformations. The different contribution of conformational entropy to binding each RNA sequence could explain the observed differences in structures and binding affinities. Changes in the conformational entropy of the protein, solvent or RNA likely contribute favorably to the free energy of the system and could drive high affinity recognition of RNA. Given that the affinities of TIS11d to ARE, FM3 and FM4 are similar, I hypothesize that favorable changes in conformational entropy compensate for unfavorable differences in the conformations of TIS11d-RNA.

Mechanism of long range-communications

Long-range communications within a protein and between proteins are essential in allosteric regulation and signal transduction.(*7*) The mechanism of long-range communications can be described according to a network of linked structural changes.(*8, 9*) However, local structural rearrangements are not the only source of long-range communication. Studies of transmembrane domains of signaling proteins demonstrate that long-range communications may be achieved through alterations in protein flexibility.(*10*) Although changes in flexibility have been shown to mediate long-range communication and allosteric regulation, methods to account for protein flexibility in communicative networks have not been developed.(*7*)

This thesis identifies several interfacial residues of helix E whose dynamics is greatly affected by ligand binding. A structure based network of residues which links the interface E and F helices to ligand binding at the heme has been proposed by Royer et al.(*9*) The integration of the results of the structural studies of Royer et al. with the results of future dynamics studies of different HbI mutants with altered cooperativity, described above, will reveal the coupling between structure and dynamical changes in the inter-subunit communication essential for cooperativity.

For each mutant, the entropic contribution to cooperativity can be measured using a Van't Hoff analysis. Backbone and side chain contributions to conformational entropy can be estimated from changes in order parameters. The

change in entropy associated with the water can be estimated using measurements of SASA from MD simulations of bound and unliganded mutants. Ideally, the change in entropy associated with HbI and mutant cooperativity will be linearly related to the sum of the change in conformational entropy of the protein and the change in entropy of the water. If the dynamics or the solvent associated with the intermediate are disrupted by a point mutation in HbI but the fully bound or unliganded states are not disrupted, then the measured change in entropy of cooperative binding may not be linearly related to the estimate of the change in entropy of the system. The relationship between the location of HbI point mutations and the change in the entropy of the protein or the water may reveal the connectivity between structural changes and dynamic changes of the system associated with HbI cooperativity.

Slow protein motions contribute to protein function

NMR relaxation experiments can monitor protein motions occurring on a slow (μs-ms) time scale.(*1, 11-14*) The slow timescale dynamics of TIS11d has not been characterized. This work shows that TIS11d undergoes a conformational transition to a binding competent state in order to recognize RNA targets. In future studies, NMR relaxation dispersion experiments will be used to measure slow time scale motions of TIS11d in the free state and at different concentrations of RNA to determine if a minor population of the binding-competent state is present in the absence of RNA. If, in the absence of RNA, TTP is at the equilibrium with its RNA-binding competent state, we will conclude that binding occurs through a selected-fit model of molecular recognition. The

absence of relaxation dispersion in the free state of TIS11d will indicate that RNA is necessary to promote the transition as predicted by the induced-fit model. In the induced-fit model, ligand induces a change in the protein conformation upon binding. In the selected-fit model, ligand selects and stabilizes a complementary protein conformation from a pre-existing equilibrium of states.

Impact on the study of biological systems

HbI has evolved to utilize conformational entropy to accomplish the necessary affinity and cooperativity to bind and release oxygen as needed. This study demonstrates that structural insights, even in a system that is as well understood as hemoglobin, may be insufficient to fully characterize protein function. A full analysis of the dynamics of the system is important to understand how binding, allosteric regulation and cooperativity occur.

TIS11d is a human protein that has been implicated in AML and ALL pathogenesis and hematopoiesis.(*15, 16*) TIS11d binds and targets mRNAs for degradation; however, the *in vivo* mRNA targets of TIS11d are unknown. The discovery of *in vivo* mRNA targets will provide insight into the role of TIS11d in hematopoiesis and the origin of TIS11d related pathogenesis. This study demonstrates that TIS11d can bind three different high affinity targets; each of these sequences may mediate TIS11d-mRNA recognition *in vivo*. Determination of the TTP, TIS11d and TIS11b consensus sequences will reveal the identity of potential mRNA targets of each protein that could be validated using mutagenesis. If the consensus sequences of TTP, TIS11d and TIS11b are

degenerate then the proteins may be capable of binding many *in vivo* targets or acquire functional specificity through conformational selection.

In addition, our study of TIS11d suggests that this protein binds different RNAs with different conformations that might be important for regulation. Structural, dynamical, and biochemical studies of full-length TIS11d in complex with its target RNAs and with other proteins of the RNA degradation machinery will elucidate the biological functions of the different conformations of TIS11d when bound to different RNAs.

Therapeutic interventions with TTP-like proteins.

A single intratumoral injection of TIS11b fused to R9, the cell penetrating poly-arginine peptide, suppresses tumor growth, enhances inflammatory cytokines and stops tumor vascularization.(*17*) The TIS11b-R9 therapy is a novel and potent approach to antitumor therapy and demonstrates the therapeutic power of modifying mRNA stability.(*17*) However, given that the biological function of TIS11d in the cell is still not well understood the side effects of these therapies cannot be predicted. Further studies on the binding specificity of TIS11d could illuminate potential side effects in humans and may provide insights into protein engineering strategies to reduce potential side effects.

CONCLUDING REMARKS

Fundamentally this work highlights the involvement of protein flexibility in cooperativity and binding specificity. The cooperativity of the homodimeric hemoglobin HbI of *scapharca inaequivalvis* is entropically driven and this study demonstrates that backbone conformational entropy and changes in interface dynamics favor HbI cooperativity. Protein dynamics are an essential element of protein function. This research demonstrates that the origins of protein function is intimately linked to protein flexibility and it calls for the investigation of backbone conformational entropy in all entropically driven processes.

The EF interface of HBI is conserved in invertebrate cooperative hemoglobins and has been observed in annelids, mollusks and the deuterostome phylum of echinoderms.(*18*) However, individual residues are not conserved within these hemoglobins, which suggests that the EF dimer assemblage has arisen through convergent evolution.(*18*) Therefore the EF interface has been independently selected to efficiently regulate oxygen transport in invertebrates. Although the thermodynamics of other EF assemblages is unknown, if future studies prove that HbI cooperativity is driven by changes in conformational entropy then perhaps all EF assemblages utilize changes in conformational entropy to bind oxygen cooperatively. If this is the case, then the EF interface may contain some innate feature that promotes cooperativity through dynamic and structural changes. This would be a prime example of the natural selection of changes in conformational entropy to generate cooperative binding.

TIS11d recognizes three similar RNA sequences with high affinity. I anticipate that fusions of TTP like proteins to R9 will act as potent antitumoral agents in humans. This research and future work on the binding specificity of TTP-like proteins will allow for intelligent design of TZFs that will limit potential side-effects of this novel treatment in humans.

REFERENCES:

1. Popovych, N., Sun, S., Ebright, R. H., and Kalodimos, C. G. (2006) Dynamically driven protein allostery, *Nat Struct Mol Biol 13*, 831-838.
2. Tzeng, S. R., and Kalodimos, C. G. (2009) Dynamic activation of an allosteric regulatory protein, *Nature 462*, 368-372.
3. Petit, C. M., Zhang, J., Sapienza, P. J., Fuentes, E. J., and Lee, A. L. (2009) Hidden dynamic allostery in a PDZ domain, *Proc Natl Acad Sci U S A 106*, 18249-18254.
4. Ikeda-Saito, M., Yonetani, T., Chiancone, E., Ascoli, F., Verzili, D., and Antonini, E. (1983) Thermodynamic properties of oxygen equilibria of dimeric and tetrameric hemoglobins from Scapharca inaequivalvis, *J Mol Biol 170*, 1009-1018.
5. Wand, A. J. (2001) Dynamic activation of protein function: a view emerging from NMR spectroscopy, *Nat Struct Biol 8*, 926-931.
6. Marlow, M. S., Dogan, J., Frederick, K. K., Valentine, K. G., and Wand, A. J. (2010) The role of conformational entropy in molecular recognition by calmodulin, *Nat Chem Biol 6*, 352-358.
7. Whitley, M. J., and Lee, A. L. (2009) Frameworks for understanding long-range intra-protein communication, *Curr Protein Pept Sci 10*, 116-127.
8. Gandhi, P. S., Chen, Z., Mathews, F. S., and Di Cera, E. (2008) Structural identification of the pathway of long-range communication in an allosteric enzyme, *Proc Natl Acad Sci U S A 105*, 1832-1837.
9. Ren, Z., Srajer, V., Knapp, J. E., and Royer, W. E., Jr. (2012) Cooperative macromolecular device revealed by meta-analysis of static and time-resolved structures, *Proc Natl Acad Sci U S A 109*, 107-112.
10. Langosch, D., and Arkin, I. T. (2009) Interaction and conformational dynamics of membrane-spanning protein helices, *Protein Sci 18*, 1343-1358.
11. Lange, O. F., Lakomek, N. A., Fares, C., Schroder, G. F., Walter, K. F., Becker, S., Meiler, J., Grubmuller, H., Griesinger, C., and de Groot, B. L. (2008) Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution, *Science 320*, 1471-1475.
12. Eisenmesser, E. Z., Bosco, D. A., Akke, M., and Kern, D. (2002) Enzyme dynamics during catalysis, *Science 295*, 1520-1523.
13. Cole, R., and Loria, J. P. (2002) Evidence for flexibility in the function of ribonuclease A, *Biochemistry 41*, 6072-6081.
14. Beach, H., Cole, R., Gill, M. L., and Loria, J. P. (2005) Conservation of mus-ms enzyme motions in the apo- and substrate-mimicked state, *J Am Chem Soc 127*, 9167-9176.
15. Stumpo, D. J., Broxmeyer, H. E., Ward, T., Cooper, S., Hangoc, G., Chung, Y. J., Shelley, W. C., Richfield, E. K., Ray, M. K., Yoder, M. C., Aplan, P. D., and Blackshear, P. J. (2009) Targeted disruption of Zfp36l2, encoding a CCCH tandem zinc finger RNA-binding protein, results in defective hematopoiesis, *Blood 114*, 2401-2410.

16. Iwanaga, E., Nanri, T., Mitsuya, H., and Asou, N. (2011) Mutation in the RNA binding protein TIS11D/ZFP36L2 is associated with the pathogenesis of acute leukemia, *Int J Oncol 38*, 25-31.
17. Planel, S., Salomon, A., Jalinot, P., Feige, J. J., and Cherradi, N. (2010) A novel concept in antiangiogenic and antitumoral therapy: multitarget destabilization of short-lived mRNAs by the zinc finger protein ZFP36L1, *Oncogene 29*, 5989-6003.
18. Royer, W. E., Jr., Zhu, H., Gorr, T. A., Flores, J. F., and Knapp, J. E. (2005) Allosteric hemoglobin assembly: diversity and similarity, *J Biol Chem 280*, 27477-27480.

APPENDIX A

MOLECULAR MECHANISMS OF VIRAL AND HOST-CELL SUBSTRATE RECOGNITION BY HCV NS3/4A PROTEASE

AUTHOR CONTRIBUTIONS

This chapter is a collaborative project by Keith Romano, myself, Laura Deveau, Hong Cao, Francesca Massi and Celia Schiffer. Keith Romano did all crystallization and structural analysis. Laura Deveau designed and cloned each of the protease mutants. Hong Cao and Keith Romano conducted and analyzed the fluorescence polarization experiments. Celia Schiffer and Keith Romano wrote the paper that includes this work. I performed the NMR experiments and both Francesca Massi and I analyzed and processed the NMR data for publication.

This work is published:

Romano, K. P., J. M. Laine, L. M. Deveau, H. Cao, F. Massi, and C. A. Schiffer. "Molecular Mechanisms of Viral and Host Cell Substrate Recognition by Hepatitis C Virus Ns3/4a Protease." *J Virol* 85, no. 13: 6106-16.

ABSTRACT

Hepatitis C NS3/4A protease is a prime therapeutic target responsible for cleaving the viral polyprotein at junctions 3-4A, 4A4B, 4B5A and 5A5B, and two host-cell adapter proteins of the innate immune response, TRIF and MAVS. In this study, NS3/4A crystal structures of both host-cell cleavage sites are determined and compared to the crystal structures of viral substrates. Two distinct protease conformations are observed and correlate with substrate specificity: (1) 3-4A, 4A4B, 5A5B and MAVS, which are processed more efficiently by the protease, form extensive electrostatic networks when in complex with the protease and (2) TRIF and 4B5A, which contain polyproline motifs in their full-length sequences, do not form electrostatic networks in their crystal complexes. These findings provide mechanistic insights into NS3/4A substrate recognition, which may assist in a more rational approach to inhibitor design in the face of the rapid acquisition of resistance.

INTRODUCTION

Hepatitis C virus (HCV) is a genetically diverse *Hepacivirus* of the *Flaviviridae* family infecting over 180 million people worldwide (33). HCV contains a positive, single-stranded RNA genome that is translated as a single polyprotein along the endoplasmic reticulum by host cell machinery. The viral polyprotein is subsequently processed by host-cell and viral proteases into structural (C, E1, E2) and non-structural components (p7, NS2, NS3, NS4A, NS4B, NS5A, NS5B) (23). The NS3/4A protein, a bifunctional protease/helicase enzyme formed by the non-covalent association of NS3 and NS4A, hydrolyzes four known sites along the viral polyprotein, thereby liberating non-structural proteins essential for viral replication. Previous kinetic data suggest that the first cleavage event at junction 3-4A occurs *in cis* as a unimolecular process, while processing of the remaining junctions 4A4B, 4B5A and 5A5B occur bimolecularly *in trans* (2, 22). Interestingly, these data demonstrated that the NS4A sequence is essential for the cleavage of junction 4B5A. These viral substrates share little sequence similarity, except for an acid at P6, cysteine or threonine at P1 and serine or alanine at P1' (**Table A.1**). Previous work by our group has revealed that the diverse set of NS3/4A substrate sequences are recognized in a conserved three-dimensional shape, defining a consensus van der Waals volume, or substrate envelope (29). This conserved mode of substrate recognition regulates polyprotein processing and thus the biology of HCV replication.

Table A.1 Primary cleavage sequences of NS3/4A substrates (genotype 1a).										
Substrate	**P6**	**P5**	**P4**	**P3**	**P2**	**P1**	**P1'**	**P2'**	**P3'**	**P4'**
TRIF	$P_{(8)}$	S	S	T	P	C	S	A	H	L
MAVS	E	R	E	V	P	C	H	R	P	S
3-4A	D	L	E	V	V	T	S	T	W	V
4A4B	D	E	M	E	E	C	S	Q	H	L
4B5A	E	C	T	T	P	C	S	G	S	W
5A5B	E	D	V	V	C	C	S	M	S	Y

In addition to its essential role in processing the viral polyprotein, NS3/4A protease also confounds the innate immune response to viral infection by disrupting activation of the transcription factors interferon regulatory factor 3 (IRF-3) and nuclear factor κB (NF-κB) (13, 21). Upon the detection of viral RNA in host cells, these transcription factors are induced through two distinct pathways involving signaling through Toll-like receptor 3 (TLR3) or retinoic acid-inducible gene I (RIG-I) (1, 37). NS3/4A protease disrupts the TLR3 and RIG-I cascades by cleaving the essential adaptor proteins Toll-interleukin-1 receptor domain-containing adaptor-inducing interferon-β (TRIF) and mitochondrial antiviral signaling protein (MAVS), respectively (12, 21). The TRIF and MAVS cleavage sites share little sequence homology with other viral substrates: TRIF contains cysteine at P1 and serine at P1', while MAVS contains glutamate at P6 and cysteine at P1 (**Table A.1**). Notably, in place of an acid at position P6, TRIF consists of a track of eight proline residues spanning P13–P6, which has been previously implicated as part of the substrate recognition motif for NS3/4A (11). The NS3-mediated processing of both viral and host-cell targets is central to the interplay between HCV replication and the innate immune response, thus highlighting the importance of better elucidating the mechanisms of substrate recognition.

Despite great efforts devoted to the development of NS3/4A protease inhibitors, the rapid rise of drug resistance in human clinical trials has limited the efficacy of the most promising drug candidates. Drug resistance mutations in the protease emerge as molecular changes that prevent the binding of drugs, but still

permitting the recognition and cleavage of substrates. A more detailed understanding of the molecular details underlying substrate recognition is therefore critical for explaining patterns of drug resistance and for designing novel drugs that are less susceptible to resistance. Here we analyze crystal structures of NS3/4A protease in complex with N-terminal products of viral substrates 3-4A, 4A4B, 4B5A and 5A5B. TRIF and MAVS crystal complexes further reveal that these host-cell products bind to the protease active site in a conserved three-dimensional manner similar to that of the viral products.

Notably, extensive electrostatic networks involving protease residues D81, R155, D168 and R123 form in product complexes 3-4A, 4A4B, 5A5B and MAVS while these networks are absent in the 4B5A and TRIF product complexes. Short peptides corresponding to the immediate cleavage sequences of TRIF and 4B5A have significantly weaker affinities for NS3/4, which correlate with their inability to form of such electrostatic networks with the protease. Taken together, our findings support previous biochemical studies implicating the role of polyproline II helices in TRIF cleavage by NS3/4A (11), and provide a structural basis for future studies aimed at better elucidating the detailed mechanism NS3/4A substrate recognition and cleavage.

MATERIALS AND METHODS

Mutagenesis and gene information

The HCV genotype 1a NS3/4A protease gene described in a Bristol-Meyers Squibb patent (34) was synthesized by GenScript and cloned into the pET28a expression vector (Novagen). The gene encodes a highly soluble form of the NS3/4A protease domain as a single-chain, with 11 core amino acids of NS4A located at the N-terminus. The inactive S139A protease variant was subsequently constructed using the QuikChange Site-Directed Mutagenesis Kit from Stratagene and sequenced by Davis Sequencing for confirmation.

Viral substrate and peptide product purchase and storage

Thirty milligrams of each substrate peptide and the corresponding N-terminal cleavage product (TRIF and MAVS) were purchased from 21st Century Biochemicals (Marlboro, MA). The TRIF and MAVS peptides were synthesized as a 13mer (P13–P1) and 7mer (P7–P1), respectively. The N-termini of all peptides were acetylated, while the C-termini of substrate peptides were blocked with amide groups. All peptides were stored as solids at -20ºC and dissolved in DMF to a final concentration of 50–100mM for crystallization trials.

Expression and purification of NS3/4A protease constructs

NS3/4A expression and purification were carried out as described previously (14, 34). Briefly, transformed BL21 (DE3) *E. coli* cells were grown at 37ºC and induced at an optical density of 0.6 by the addition of 1mM IPTG. Cells were harvested after 5 hours of expression, pelleted and frozen at −80ºC for

storage. Cell pellets were thawed, resuspended in 5mL/g of resuspension buffer (50mM phosphate buffer at pH 7.5, 500mM NaCl, 10% glycerol, 2mM β-ME) and lysed with a cell disruptor. The soluble fraction was retained, applied to a nickel column (Qiagen), washed with resuspension buffer, and eluted with resuspension buffer supplemented with 200mM imidazole. The eluant was dialyzed overnight (MWCO 10kD) to remove the imidazole and the his-tag was simultaneously removed with thrombin treatment. The nickel-purified protein was then flash-frozen and stored at −80ºC for up to six months.

Crystallization of product complexes and apo enzyme

For crystallization, the protein solution was thawed, concentrated to ~3mg/mL and loaded on a HiLoad Superdex75 16/60 column equilibrated with gel filtration buffer (25mM MES at pH 6.5, 500mM NaCl, 10% glycerol, 30μM zinc chloride, and 2mM DTT). The protease fractions were pooled and concentrated to 20–25mg/mL with an Amicon Ultra-15 10kD device (Millipore). The concentrated samples were either used for crystallization of apo structure or incubated for one hour with 2–20 molar excess of substrate product TRIF or MAVS. Diffraction-quality crystals were obtained overnight by mixing equal volume of concentrated protein solution with precipitant solution (20–26% PEG-3350, 0.1M sodium MES buffer at pH 6.5, and 4% ammonium sulfate) in 24-well VDX hanging drop trays.

Data collection and structure solution

Crystals large enough for data collection were flash-frozen in liquid nitrogen for storage. The TRIF, MAVS and apo crystals were mounted under

constant cryostream and x-ray diffraction data were collected at Advanced Photon Source LS-CAT 21-ID-F, BioCARS 14-BMC and our in-house RAXIS IV x-ray system, respectively. Diffraction intensities of product complexes were indexed, integrated and scaled using the program HKL2000 (27). All structure solutions were generated using simple isomorphous molecular replacement with PHASER (24). The B chain model of viral substrate product 4A4B (3M5M) (29) was used as the starting model for all structure solutions. Initial refinement was carried out in the absence of modeled ligand, which was subsequently built in during later stages of refinement. Upon obtaining the correct molecular replacement solutions, the phases were improved by building solvent molecules using ARP/wARP (25). Subsequent crystallographic refinement was carried out within the CCP4 program suite with iterative rounds of TLS and restrained refinement until convergence was achieved (6). The final structures were evaluated with MolProbity (7) prior to deposition in the protein data bank. 5% of the data was reserved for the free R-value calculation to limit the possibility of model bias throughout the refinement process (4). Interactive model building and electron density viewing was carried out using the program COOT (10).

Double-difference plots and global analysis

Double-difference plots were computed as described previously (28). Briefly, the atomic distances were calculated between each Cα of a given protease molecule and every other Cα in the same molecule. The differences of these Cα-Cα distances were then calculated between each pair of protease molecules and plotted as a contour graph for visualization. These analyses

allowed for effective structural comparisons without the biases associated with superimpositions and space group differences. Double-difference plots were used to determine the structurally invariant regions of the protease, consisting of residues 32–36, 42–47, 50–54, 84–86 and 140–143. Structural superimpositions were carried out in PyMOL (9) using the Cα atoms of these residues for all protease molecules. The apo structure was used as the reference structure for the alignments of the TRIF and MAVS product complexes. The Cα RMSD for each residue was subsequently calculated to access the degree of structural variation throughout the protein. The B-factor column of a representative structure was replaced with these values and used to generate the rainbow color spectrum to visualize these variations.

The viral substrate envelope

All active site alignments were performed with PyMOL using the Cα atoms of protease residues 137–139 and 154–160. For each alignment, the B chain of complex 4A4B was used as the reference structure. The NS3/4A viral substrate envelope, representing the consensus van der Waals volume shared by any three of the four viral substrate products, was computed as described previously using the full-length NS3/4A structure (1CU1) (35) and product complexes 4A4B (3M5M), 4B5A (3M5N) and 5A5B (3M5O) (29).

Active site comparisons of the four NS3/4A viral product complexes were performed by superposition the Cα atoms of residues 137–139 and 154–160, revealing that both the active site residues and substrate products spanning P6–P1 align closely with an average Cα rmsd of 0.24 Å and 0.35 Å, respectively.

The consensus van der Waals volume shared by any three of the four cleavage products was then calculated to generate the NS3/4A substrate envelope.

NMR spectroscopy and data processing

NMR data were recorded using 650µL of 390µM [U-^{15}N] NS3/4A protease (95% H_2O/5% D_2O, 25mM sodium phosphate at pH 7.2, 150mM KCl, 5µM zinc chloride, and 1mM TCEP). Backbone ^{1}H and ^{15}N resonance assignments of NS3/4A protease were kindly provided by Herbert Klei of Bristol-Myers Squibb, and confirmed using nuclear Overhauser enhancement spectroscopy (NOESY) experiments. Backbone ^{1}H and ^{15}N resonance assignments of NS3/4A protease bound to the N-terminal cleavage product of substrate 4A4B were obtained from the assignment of the free protein by following the chemical shift changes upon titration of the ligand. All experiments were performed at 298K using a Varian Inova spectrometer operating at 600 MHz (14.1 T). Spectra were processed using nmrPipe (8) and Sparky (15).

Binding of the unlabeled peptide corresponding to the N-terminal cleavage product of substrate 4A4B to [U-^{15}N] NS3/4A protease was monitored using a series of two-dimensional ^{15}N-^{1}H HSQC spectra collected at increasing concentrations of the peptide to a final concentration of 2.5 mM. The change of the cross-peak positions for NS3/4A residues was recorded as a function of titrated peptide concentration. The normalized change in the chemical shift was calculated for each protease residue using the following equation:

$$\Delta = \sqrt{\Delta\delta_H^2 + \left(\Delta\delta_N \gamma_N / \gamma_H\right)^2}$$

where $\Delta\delta$ is the chemical shift change observed between the free and bound states, γ is the gyromagnetic ratio, and the subscript indicates either ^{1}H or ^{15}N nuclei. The b-factors for each residue of the apo enzyme crystal structure were then replaced by the maximal chemical shift changes from the titration data. The crystal structure was colored in PyMOL (9) according to the chemical shift magnitudes to graphically depict the locations of the shifting residues.

Van der Waals contact energy

Van der Waals contact energies between protease residues and peptide products were computed using a simplified Lennard-Jones potential as described previously (26). Briefly, the Lennard-Jones potential (V_r) was calculated for each protease-product atom pair where r, ε and σ represent the interatomic distance, van der Waals well depth and atomic diameter, respectively.

$$V_r = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$$

V_r was computed for all possible protease-product atom pairs within 5Å, and potentials for non-bonded pairs separated by less than the distance at the minimum potential were equated to $-\varepsilon$. Using this simplified potential value for each non-bonded protease-product atom pair, the total van der Waals contact energy (ΣV_r) was computed for each peptide residue. For graphical convenience, Van der Waals energy indexes were then calculated by multiplying the raw values by a factor of -10.

Fluorescence polarization

For fluorescence polarization experiments, the NS3/4A protease domain was purified as described previously in purification buffer (25mM HEPES, pH 7.5, 150mM NaCl, 20% glycerol, 4mM DTT) and subsequently concentrated to 200–400μM. The concentrated stocks were then two-thirds serially diluted in 384-well plates (Corning) in reaction buffer using the Genesis Workstation (Tecan). An equal volume of substrate buffer (25mM HEPES, pH 7.5, 0mM NaCl, 20% glycerol, 4mM DTT) containing 10nM of fluorescein-tagged substrate product (4A4B, 4B5A or 5A5B) was added to each well to make a final well volume of 60μL. The final condition constituted 5nM fluorescein-tagged cleavage products in 25mM HEPES (pH 7.5), 75mM NaCl, 20% glycerol and 4mM DTT. The plates were incubated at room temperature for two hours and five fluorescence polarization measurements were taken for each well using the Victor-3 plate reader (Perkins Elmer). Five sets of binding data were collected for each substrate product and each trial was processed independently. The average and standard deviations were then calculated from the results of these five trials. Fluorescence polarization data (in milli-polarization units, *mP*) were fit to the hill equation, where E_T is the total NS3/4A concentration, K_d is the equilibrium binding constant, *b* is the baseline fluorescence polarization and *m* is the fluorescence polarization maximum.

$$mP = b + \left(\frac{m - b}{1 + K_d / E_T} \right)$$

RESULTS

Structure determination of apo NS3/4A and product complexes

The apo NS3/4A protease domain and host-cell product complexes TRIF and MAVS all crystallized in the space group $P2_12_12_1$ with one molecule in the asymmetric unit (**Table A.2**). For all structures, we utilized the highly soluble NS3/4A protease domain described previously, containing the essential residues of cofactor NS4A covalently linked at the N-terminus (34). This NS3/4A construct also contains the inactivating mutation S139A, designed to further enhance protein stability by minimizing autoproteolysis during the crystallization process. The partially inactivated variant still exhibits residual proteolytic activity, as observed for other serine proteases (5, 18), likely facilitated by the nucleophilic attack of water. Thus the complete characterization of full-length substrate peptides was not possible and short peptides corresponding to the N-terminal cleavage products of authentic substrates were used for all crystallization trials. Peptide products 4A4B, 4B5A, 5A5B and MAVS spanned P7–P1, while the TRIF sequence spanned P13–P1 with a track of eight proline residues at the N-terminus (**Table A.1**). The entire peptide sequence spanning P7–P1 could be modeled in each structure except for the TRIF complex, which revealed electron density for the residues spanning P6–P1 but not the polyproline track.

Tertiary structure analysis

Structural analyses of the NS3/4A apo enzyme were carried out in conjunction with: (1) the host-cell product complexes TRIF and MAVS, (2) the viral product complexes 4A4B, 4B5A and 5A5B (29) and (3) the full-length

Table A.2. Data collection and refinement statistics.

Dataset	TRIF	MAVS	APO
PDB ID	3RC4	3RC5	3RC6
Modeled Ligand	PSSTPC	QEREVPC	*(unliganded)*
Resolution (Å)	1.50 (1.50–1.55)	1.60 (1.60–1.66)	1.30 (1.30–1.35)
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Molecules in AU[†]	1	1	1
Cell dimensions:	--	--	--
a (Å)=	53.9	54.1	54.8
b (Å)=	58.1	58.2	58.6
c (Å)=	61.3	61.3	60.9
Completeness (%)	87.4 (93.1)	99.3 (99.1)	93.1 (97.1)
Total reflections	136929	97504	190364
Unique reflection	27770	26041	45595
Average I/σ	13.5 (4.4)	18.8 (5.4)	14.8 (4.1)
Redundancy	4.9 (4.7)	3.7 (3.7)	4.2 (4.0)
R_{sym} (%)[‡]	4.8 (30.1)	3.3 (19.3)	4.1 (29.3)
RMSD[∂] in:			
Bonds (Å)	0.009	0.009	0.009
Angles (°)	1.28	1.26	1.29
R_{factor} (%)[§]	18.7	17.3	16.2
R_{free} (%)[§]	21.5	19.3	19.1

[†] AU, asymmetric unit.

[‡] $R_{sym} = \Sigma|I - \langle I \rangle|/\Sigma I$, where I = observed intensity, <I> = average intensity over symmetry equivalent.

[∂] RMSD, root mean square deviation.

[§] $R_{work} = \Sigma||F_o| - |F_c||/\Sigma|F_o|$. R_{free} was calculated from 5% of reflections, chosen randomly, which were omitted from the refinement process.

() Denotes statistics for the highest resolution shell.

NS3/4A structure (35), in which the C-terminus represents the post-cleavage product 3-4A. The seven structures constitute a total of twelve NS3/4A protease monomers, which all adopt the same chymotrypsin-like tertiary fold defined by the labeling-scheme for trypsin (3). The N-terminal distorted β-barrel subdomain contains two α-helices (α0 and α1) and seven β-strands (A0–F1), while the C-terminal β-barrel subdomain comprises of two α-helices (α2 and α3) and six β-strands (A2–F2). The cofactor NS4A contributes a single β-strand to the N-terminal distorted β-barrel, which is essential for efficient catalytic function (22). The catalytic triad is located in the cleft between these subdomains, with the N-terminal β-barrel contributing residues H57 and D81 and the C-terminal β-barrel contributing the nucleophilic S139. The active site residues in all crystal structures share similar architecture, defined by the catalytic triad residues (S139A, H57, D81) and backbone nitrogens of the oxyanion hole (G137, S138, S139A).

Certain global differences are observed between these structures when superpositions are performed. Double-difference plots were therefore generated between each co-complex and the apo enzyme to determine the most invariant regions (**Figure A.1**). Product 3-4A varied most extensively from the apo state, which likely reflects differences in genotype, protein size and crystal packing of the full-length construct. The remaining product complexes derive from the same NS3/4A protease domain construct, and in general, vary less extensively from the apo enzyme. Notably, the host-cell product complexes are most similar to the apo enzyme, while viral product complexes 4A4B, 4B5A and 5A5B vary more

Figure A.1 Double-difference plots of product complexes.

Double-difference plots were computed for viral and host-cell product complexes relative to the apo enzyme structure. NS3 protease residues are numbered according to the conventional system, whereas negative numbers indicate residues of the cofactor NS4A. Red and blue contour lines represent positive differences of 1.0Å and 0.5Å, respectively, while black and green lines represent negative differences of the same magnitudes.

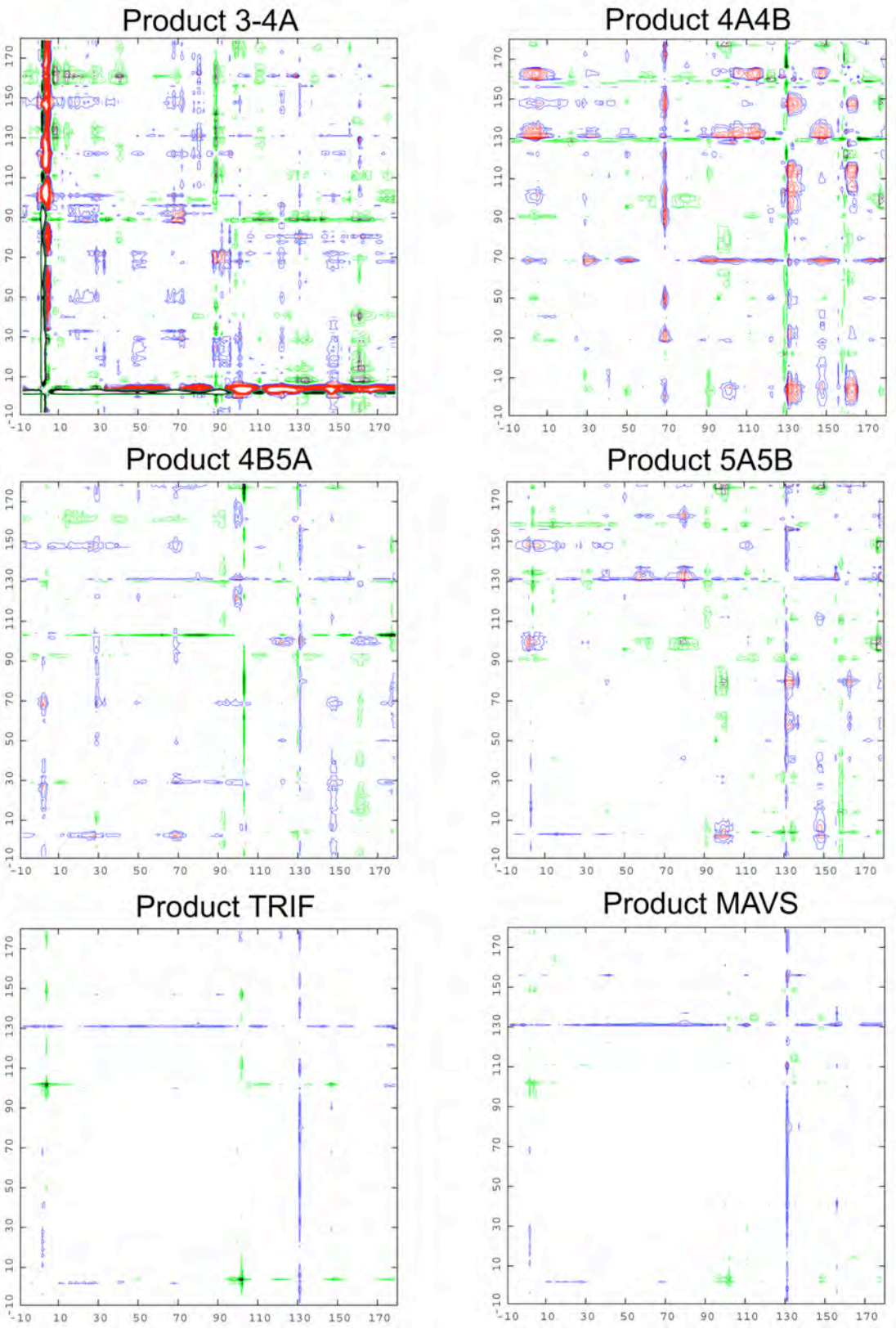

Product 3-4A
Product 4A4B
Product 4B5A
Product 5A5B
Product TRIF
Product MAVS

Figure A.1

extensively. Taken together, these findings suggest that structural differences likely reflect the inherent flexibility in certain regions of the protease.

The set of double difference plots were further analyzed and the most invariant regions of the protease were determined to contain residues 32–36, 42–47, 50–54, 84–86, and 140–143 (**Figure A.1**). Product complexes 4A4B, 4B5A, 5A5B, TRIF and MAVS consist of ten protease molecules, which were subsequently superposed onto the apo enzyme using the Cα atoms of the structurally invariant residues. The Cα average root mean square (RMS) deviation was calculated for each residue and the seven most variable regions of the protease were determined to be (**Figure A.2A)**: (1) the linker connecting cofactor 4A at the N-terminus, (2) the loop containing residues 65–70, (3) the zinc-binding site containing residues 95–105, (4) the 3_{10} helix region spanning residues 128–136, (5) the zinc-binding site containing residues 145–148, (6) the active site anti-parallel β-sheet containing residues 156–168 and (7) the C-terminal α3 helix. These regions are solvent-exposed and likely influenced by both crystal packing effects and inherent flexibility.

Extensive structural differences are observed near the active site as indicated by large RMS deviations for residues 156–168. These differences are most pronounced for the β-strands E2 and F2, which form the anti-parallel β-sheet constituting the majority of the active site. This region is least variable near the catalytic triad, while the average RMS deviations increase significantly toward the loop connecting these β-strands (**Figure A.2B**). Though the architecture of the protease catalytic triad is conserved, these observations suggest a potential

Figure A.2 Average RMS deviations of product complexes.

All protease molecules from product complexes 4A4B, 4B5A, 5A5B, TRIF and MAVS were superposed onto the apo enzyme structure using the most invariant core residues 32–36, 42–47, 50–54, 84–86 and 140–143. (A) The average Cα RMS deviations were plotted versus residue number and (B) mapped onto a representative protease molecule with the most variable regions depicted in red and the most invariant regions depicted in blue. The seven most variable regions of the protease are labeled in both panels.

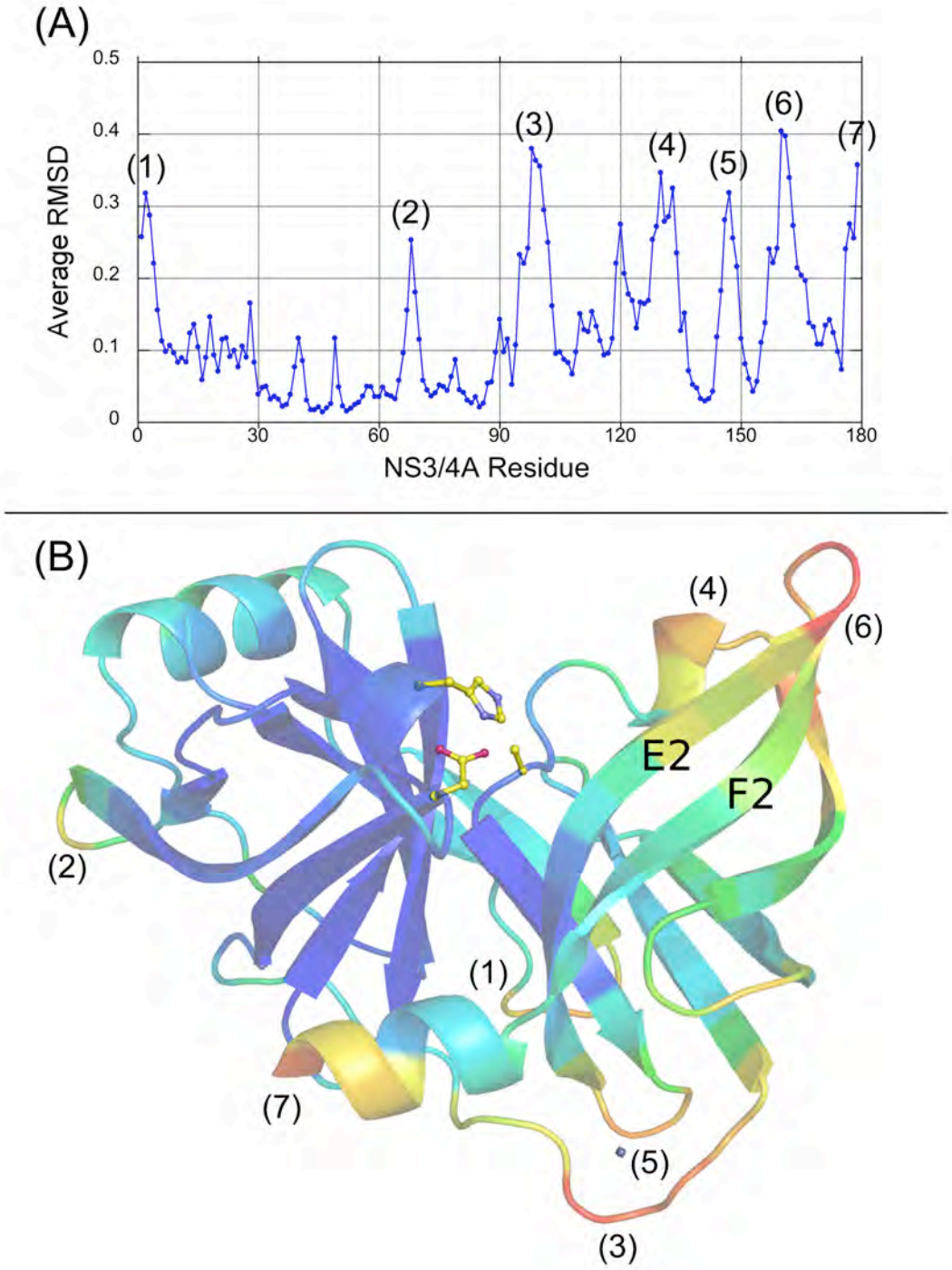

(A)
Average RMSD
0.5
0.4
0.3
0.2
0.1
0
0
30
60
90
120
150
180
NS3/4A Residue
(1)
(2)
(3)
(4)
(5)
(6)
(7)
(B)
(4)
(6)
E2
F2
(2)
(1)
(7)
(5)
(3)

Figure A.2

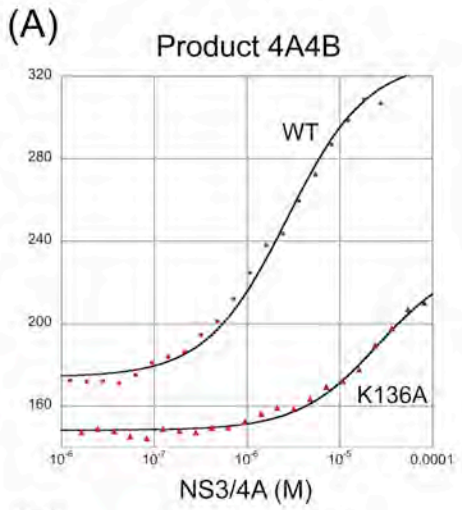
dynamic interaction between the anti-parallel β-sheet of the protease active site and substrate products. Further studies are necessary to probe the nature and extent of such dynamic interplay. Nevertheless, though the Cα atoms of active site residues shift relative to the protease core, these residues superpose well onto the APO enzyme with a RMS deviation range of 0.3–0.5Å. Moreover, the residue side chains adopt similar rotamer conformations and interact with the same surrounding residues, suggesting that potential flexibility in the protease active site would not disrupt its tertiary structure.

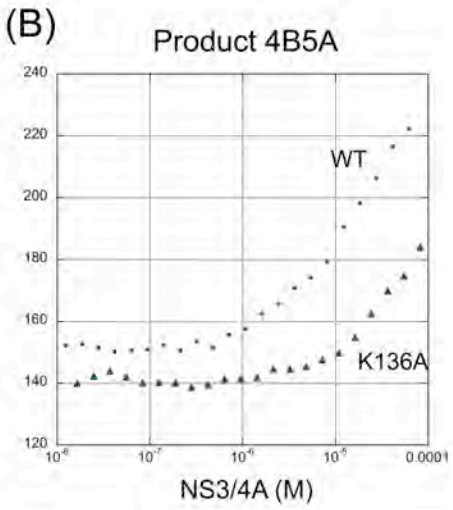
Analysis of viral product binding

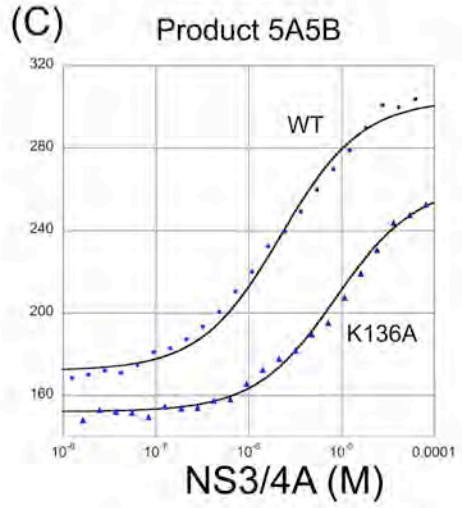
Fluorescence polarization experiments reveal that the viral products bind with different affinities to the protease, with the K_d for 4B5A over 10-fold weaker than those values for 4A4B and 5A5B (**Figure A.3**). Crystal structures reveal that these viral products bind in a conserved manner, forming an anti-parallel β-sheet with protease residues 154–160 and burying 500–600$Å^2$ of solvent accessible surface area (19). The peptide product backbone torsion angles are very similar, with positions P1–P4 being the most similar and residues P5–P7 deviating progressively toward the C-terminus. A constrained P2 ϕ torsion angle of about -60° is observed in product complexes 3-4A, 4A4B and 5A5B. Interestingly, these P2 residues could sterically tolerate the substitution of proline, which is found at the P2 position in substrates 4B5A, TRIF and MAVS. The ability of the P2 residue to adopt this constrained backbone torsion is a likely determinant in recognition process, allowing for the proper positioning of the P1 cysteine for catalysis in the active site.

Figure A.3 Fluorescence polarization of viral product binding.

Fluorescence polarization binding experiments were conducted with the inactive S139A protease (*circles*) and S139A/K136A variant (*triangles*) with fluorescein-tagged peptide products (A) 4A4B, (B) 4B5A and (C) 5A5B. For all conditions, trace amounts of peptide were incubated for two hours with increasing concentrations of NS3/4A protease in 25mM HEPES (pH 7.2), 75mM potassium chloride, 20% glycerol, 0.5µM zinc chloride and 4mM DTT. Each panel depicts a single representative trial, though each reaction condition was repeated five times. (D) Data for each trial were processed independently by least-squares regression fitting with the Hill Equation and the average and standard deviation of equilibrium dissociation constants (K_d) were calculated.







(D)

NS3/4A Cleavage Product Binding Affinities

	WT	*K136A*
Product 4A4B	***2.20μM ± 0.45***	***> 25μM***
Product 4B5A	***> 25μM***	***>50μM***
Product 5A5B	***2.60μM ± 0.32***	***9.80μM ± 2.80***

Figure A.3

There are many conserved features in viral product binding, involving both backbone and side chain interactions (**Figure A.4**). For example, eight hydrogen bonds between backbone amide and carbonyl groups are completely conserved in the product complexes, involving protease residues G137, S138, S139A, R155, A157 and S159. S159 (C159 in product complex 3-4A) and A157 each contribute two hydrogen bonds with bound products at positions P5 and P3, respectively. The P1 residue, cysteine in all substrates but 3-4A, interacts favorably with the π-system of electrons of F154. All P1 terminal carboxyl groups sit in the oxyanion hole, hydrogen bonding with the Nε nitrogen of H57 and the amide nitrogens of residues 137–139. Though the coordinates of the P1 terminal oxygen atom are not included in the full-length NS3/4A structure, geometric restraints would position it similarly to the other peptide products. Thus the same set of protease residues contacts all peptide sequences, although the precise nature of these interactions varies depending on the particular residue involved in each contact.

Despite these similarities, there are also unique features that likely underlie the particular specificity of NS3/4A for each substrate (**Figure A.4**). For example, the four acidic residues in product 4A4B lead to a highly charged peptide in solution. In the bound state, however, the atomic geometry suggests that the P5 glutamate is protonated and hydrogen bonding with the carboxyl group of the P3 glutamate, which itself forms an ionic interaction with the terminal nitrogen of K136. In fact, K136 interacts differently with all four viral products, forming: (1) a hydrogen bond with the P2 carbonyl oxygen of 3-4A, (2) a salt

Figure A.4 Stereo view of viral product binding to NS3/4A protease. N-terminal NS3/4A cleavage products (A) 3-4A, (B) 4A4B, (C) 4B5A, and (D) 5A5B are shown binding to the protease active site with the electrostatic interactions of backbone and sidechain atoms depicted in black and red, respectively.

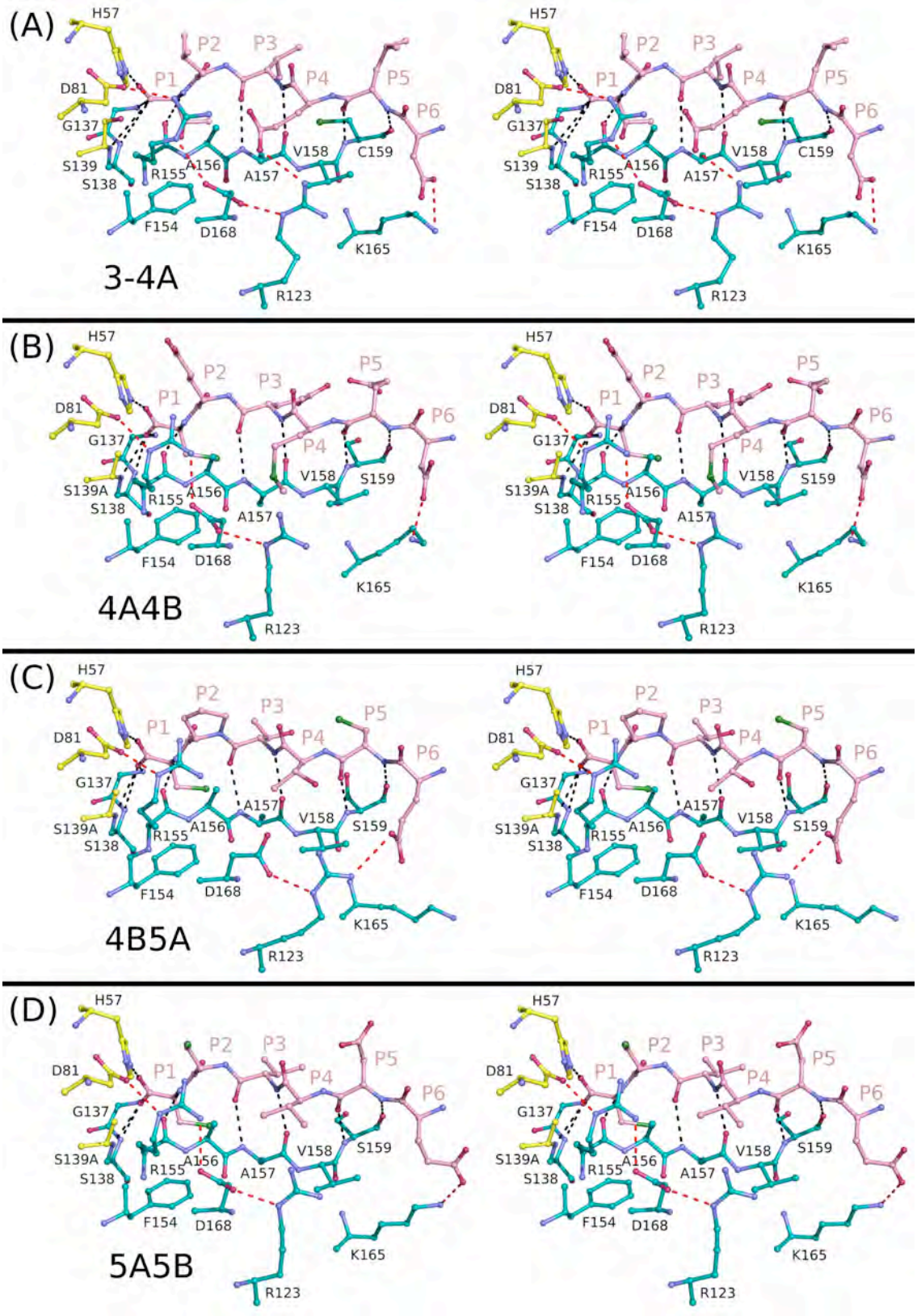

(A)
H57
P1
P2
P3
P4
P5
P6
D81
G137
S139
S138
R155
A156
A157
V158
C159
F154
D168
K165
R123
3-4A
(B)
S139A
S159
4A4B
(C)
4B5A
(D)
5A5B

Figure A.4

bridge with the P3 glutamate of 4A4B, and (3) an extended conformation that does not interact considerably in product complexes 4B5A and 5A5B. Fluorescence polarization data confirms a more significant loss in binding affinity of product 4A4B for the K136A protease variant compared to products 4B5A and 5A5B (**Figure A.3**). Thus the affinity of a substrate product likely arises from the side chain interactions unique to that particular product.

NMR solution studies of product 4A4B binding

The NS3/4A protease active site is located on the surface of the protein and thus highly solvent-exposed. The analysis of viral and host-cell product binding is therefore complicated by the proximity of symmetry-related molecules within the crystal lattice. To investigate the possibility of crystal packing effects confounding structural observations, we carried out NMR HSQC titration experiments using peptide product 4A4B spanning P7–P1 (**Figure A.5A**). HSQC chemical shift perturbations upon product titration are consistent with the molecular interactions observed in the crystal complex. The normalized chemical shift perturbations for each protease residue were compared to the buried surface area calculated directly from product complex 4A4B (**Figure A.5B**). These data indicate that the same set of protease residues with large chemical shift perturbations are also observed to interact extensively with product 4A4B in the crystal structure. The NMR solution studies recapitulate our structural observations and suggest that our crystal structure analyses are indeed representative of the interactions occurring in solution.

Figure A.5 NS3/4A H^1, N^{15} NMR HSQC titration data.

(A) NMR spectra reveal N^{15} labeled NS3/4A chemical shift perturbations upon titration of increasing concentrations of substrate 4A4B product peptide. Cross-peaks are colored according to the concentration of peptide product 4A4B, ranging from 0mM (blue) to 2.5mM (red). For clarity, the titration data for the highest peptide concentration (2.5mM) are only depicted in the insets. (B) The apo NS3/4A crystal structure was colored according to the chemical shift changes between the apo enzyme state and 2.5 mM product 4A4B. Protease residues undergoing chemical shifts are colored by a spectrum ranging from pink (small shifts) to red (large shifts). All unassigned residues are depicted in white.

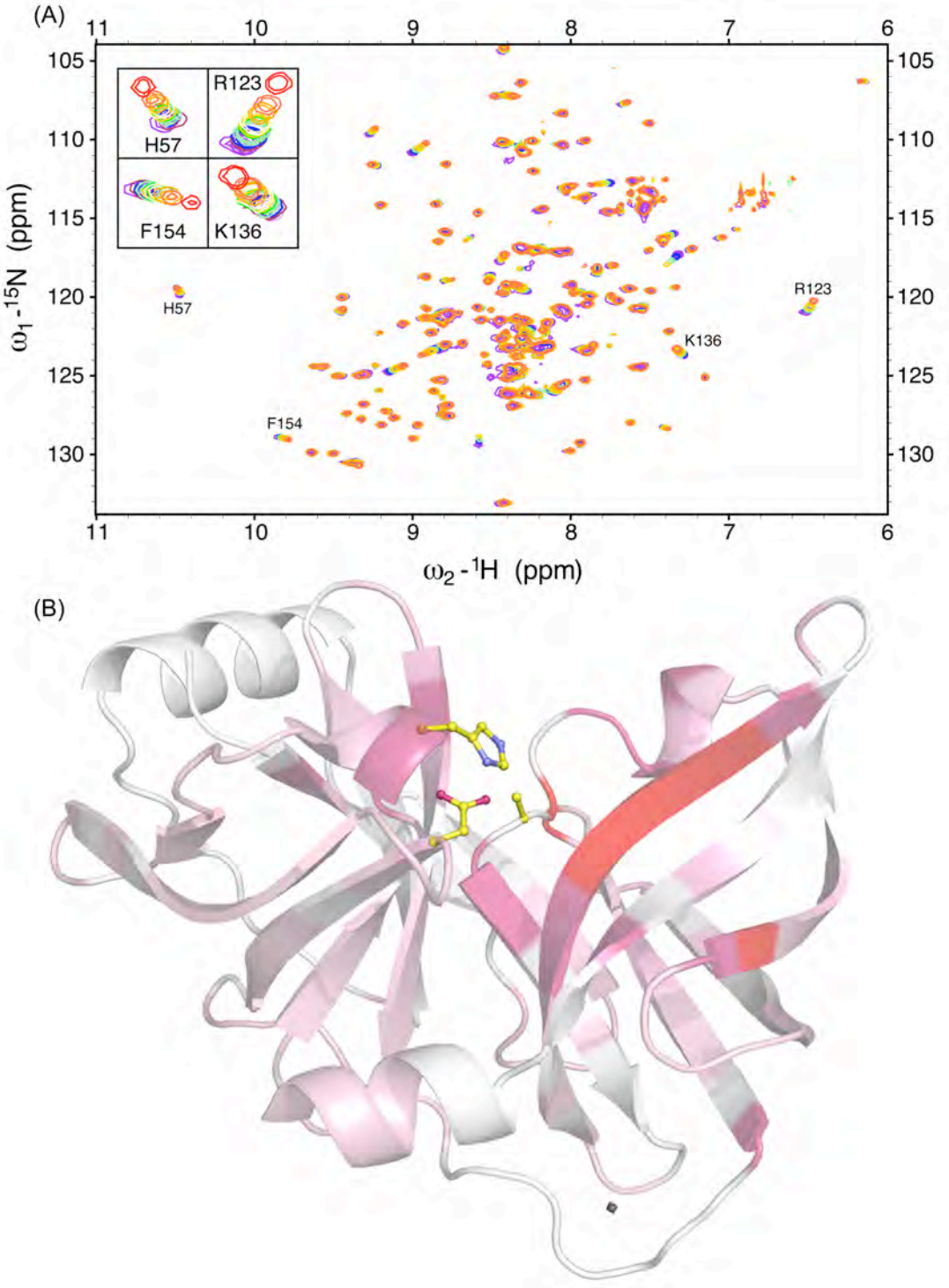

(A)
11
10
9
8
7
6
105
110
115
120
125
130
R123
H57
F154
K136
H57
R123
K136
F154
ω1 -15N (ppm)
ω2 -1H (ppm)
(B)

Figure A.5

Analysis of host-cell product complexes TRIF and MAVS

Active site superpositions (residues 137–139; 154–160) reveal that TRIF and MAVS peptide products both bind to the protease active site in a conserved three-dimensional shape (**Figure A.6**). Both substrate products form anti-parallel β-sheets with protease residues 154–160. There is no clear electron density in the TRIF complex for the proline residues spanning P13–P7. Nevertheless, the residues from P6–P1 overlap closely with the corresponding residues in the MAVS complex. The P1 cysteine residues of both substrate products interact with the aromatic ring of F154. Eight hydrogen bonds are observed in both structures, involving the amide nitrogens or carboxyl oxygens of residues G137, S138, S139A, R155, A157 and S159. A157 and S159 each contribute two hydrogen bonds with the P3 and P5 residues, respectively. In both structures, the carbonyl groups at position P1 interact with the protease oxyanion hole, defined by the backbone amide nitrogens of residues 137–139. Thus, the post-cleavage products of the cellular substrates TRIF and MAVS bind to the protease active site in a conserved manner despite their large variations in primary sequence.

There are also many differences in the binding of TRIF and MAVS involving mainly sidechain interactions with the protease (**Figure A.6**). For example, MAVS interacts closely with the protease electrostatic network formed by residues D81, R155, D168 and R123. The P4 glutamate of MAVS interacts with R155 and R123 in this network, while the P6 glutamate forms a salt bridge with R123. The TRIF peptide product, however, lacks such extended residues

Figure A.6 Stereo view of host-cell product binding to NS3/4A protease. (A) The apo enzyme structure and N-terminal cleavage products (B) TRIF and (C) MAVS are shown binding to the protease active site with the electrostatic interactions of backbone and sidechain atoms depicted in black and red, respectively.

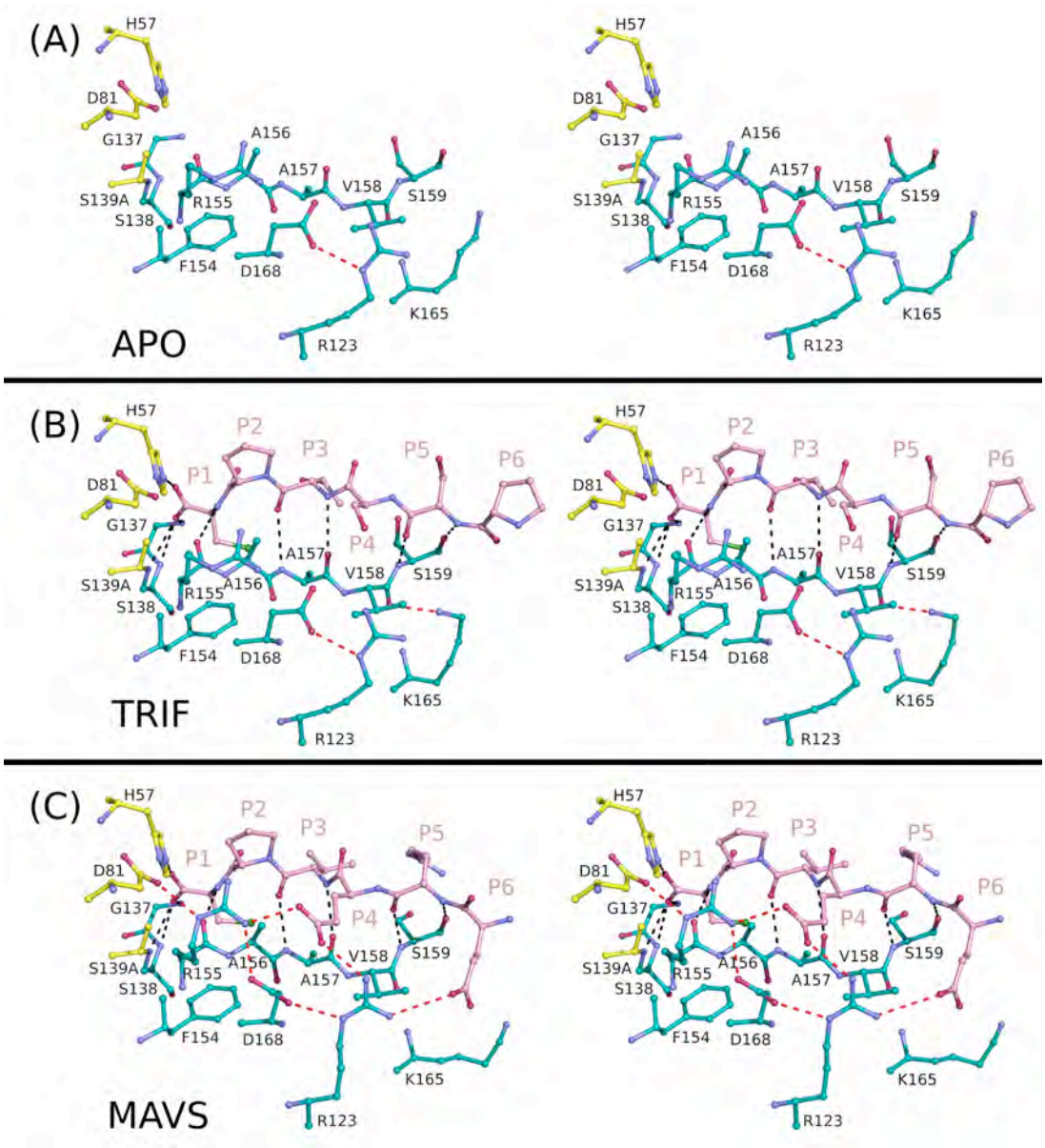

(A)
H57
D81
G137
A156
A157
V158
S159
S139A
R155
S138
F154
D168
K165
R123
APO
(B)
P1
P2
P3
P4
P5
P6
TRIF
(C)
MAVS

Figure A.6

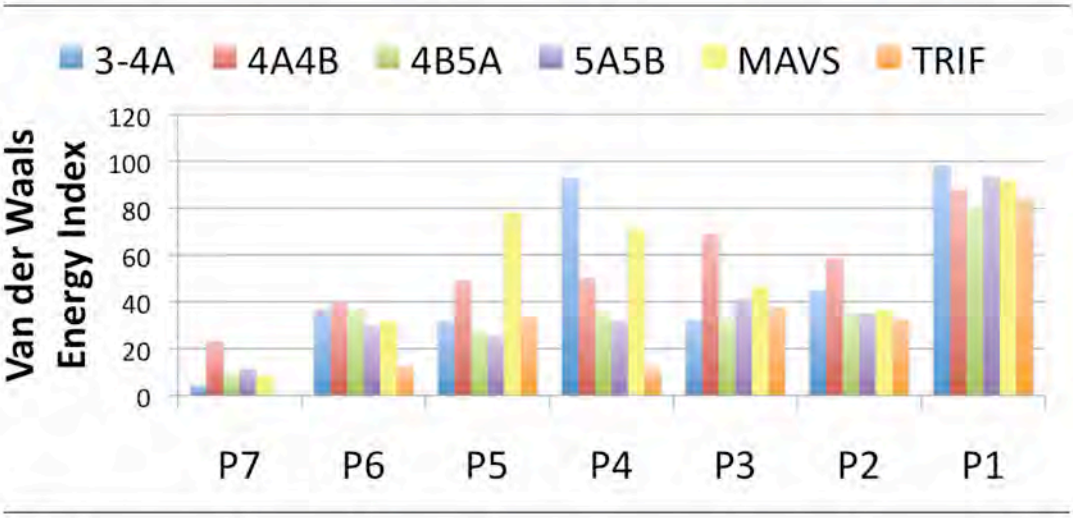
on this surface of the molecule, and the electrostatic network is notably absent with the participating residues adopting conformations observed in the apo structure. Previous studies demonstrate that a large fraction of full-length TRIF exists as polyproline II helices, and that the interaction of NS3/4A with a polyproline II helix facilitates TRIF cleavage (11). The absence of clear electron density for the proline residues suggests that full-length TRIF may be necessary to stabilize the polyproline track in a conformation capable of specific interaction with NS3/4A.

Comparison of host-cell and viral product binding

Host-cell product binding was analyzed on a broader basis by comparison with the binding of viral substrates, previously reported by our group (29). Both viral (3-4A, 4A4B, 4B5A, 5A5B) and host-cell (TRIF, MAVS) substrate products bind to the protease active site in a conserved three-dimensional shape. The peptide backbone torsional angles are very similar, being most conserved at position P1 and deviating slightly toward position P6. All peptide products adopt constrained P2 Ψ torsion angles, even those containing non-proline residues at this site. However, van der Waals analyses of substrate products indicate large variations in sidechain interactions with the protease. All of the NS3/4A substrates contain either cysteine or threonine at position P1, while five of the six contain an acid at position P6. The P1 and P6 substrate residues each contribute the same amount of van der Waals energies in all product complexes (**Figure A.7**). The amino acid makeup of viral cleavage sequences is much more diverse at positions P5–P2, and in general, the van der Waals energies at

Figure A.7 Van der Waals energies of viral substrate binding.

The van der Waals binding energies for subsites P7–P1 of each NS3/4A cleavage product (3-4A, 4A4B, 4B5A, 5A5B, TRIF and MAVS) are graphically depicted with their primary amino acid sequences tabulated below.



3-4A	A	D	L	E	V	V	T
4A4B	F	D	E	M	E	E	C
4B5A	S	E	C	T	T	P	C
5A5B	T	E	D	V	V	C	C
MAVS	Q	E	R	E	V	P	C
TRIF	P	P	S	S	T	P	C

Figure A.7

each position correlate with amino acid size. For example, the P4 glutamates in substrates 3-4A and MAVS, and to a lesser extent the P4 methionine in substrate 4A4B, are associated with larger van der Waals energies relative to the substrates with smaller P4 residues. Likewise, the larger glutamate residues at both P3 and P2 of substrate 4A4B also correlate with greater contact energies compared to the other substrates, which contain smaller amino acids at these positions.

Though similar in shape, protease substrate binding can be further categorized into two groups: (1) product complexes 3-4A, 4A4B, 5A5B and MAVS bind with an intact electrostatic network involving residues D81, R155, D168 and R123, while (2) product complexes 4B5A and TRIF bind without this network such that R155, D168 and R123 maintain conformations observed in the apo enzyme. Notably, the 4B5A and TRIF cleavage sites contain fewer charged residues compared to the other substrates, which may underlie their inability to form the electrostatic network. Binding studies demonstrate that both 4B5A and TRIF have relatively weaker affinities for NS3/4A compared to the other substrates (11, 31). However, most biochemical studies have been conducted on small peptides corresponding to the immediate cleavage sequences of TRIF and 4B5A. Indeed, kinetic studies revealed that full-length TRIF is processed more efficiently than peptides corresponding to the cleavage sequence (11). Additional molecular interactions by the full-length proteins or by adaptor proteins in the authentic cellular environment may better facilitate substrate binding. Nevertheless, the current structural analyses suggest that NS3/4A substrates

vary in specificity by their ability to form and stabilize the protease electrostatic network; the sequence specificity is particularly influenced by amino acid variation at positions P6, P4 and P2.

Analysis using the viral substrate envelope

The binding of NS3/4A cellular substrates was analyzed in terms of the viral substrate envelope (**Figure A.8**), which was previously defined as the van der Waals volume shared by any three of four viral products (29). This shape could not be predicted by the primary sequences alone and highlights the conserved mode of viral substrate recognition despite their high sequence diversity. The backbone chains of both TRIF and MAVS fit entirely within the substrate envelope, as well as the sidechains of TRIF spanning P5–P1. The sidechains of MAVS are also mostly confined within the substrate envelope, except for the longer side chains of the P6 glutamate, P5 arginine and P4 glutamate. The carboxylic acids of these glutamate residues interact extensively with the protease electrostatic network, while the P5 arginine packs against loop residues 159–162. As these interactions occur outside the viral substrate envelope, we speculate that mutations that disrupt the electrostatic network, such as R155K and D168A, would preferentially reduce the proteolytic processing of MAVS compared to TRIF.

Figure A.8 Host-cell product binding and the NS3/4A substrate envelope. (A) The NS3/4A substrate envelope is calculated from the consensus van der Waals volume shared by any three of the four viral products. (B) The TRIF cleavage product is largely confined within the substrate envelope, (C) while MAVS is mostly located within the substrate envelope except for the sidechain atoms spanning P6–P4.

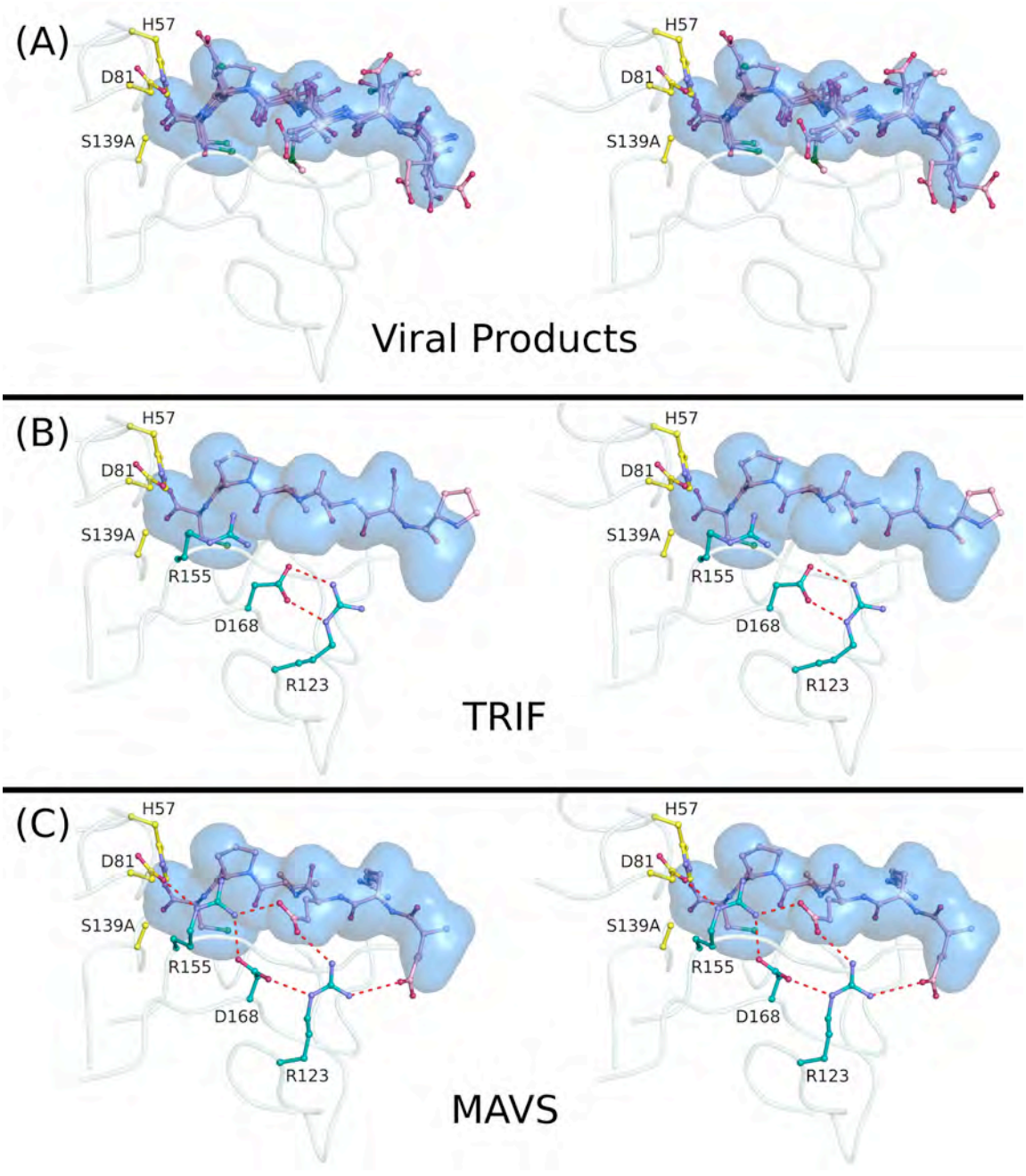


Figure A.8

DISCUSSION

The recognition and proteolysis of the viral polyprotein and host-cell adaptor proteins by NS3/4A protease play an integral role in the ability of HCV to replicate and evade the innate immune response to viral infection (12, 21). In this study, crystal structures of the NS3/4A protease domain reveal that viral and host-cell products bind to the protease active site in a similar three-dimensional shape, defined by the viral substrate envelope reported previously (29). The MAVS product complex reveals the formation of an extensive electrostatic network involving protease residues D81, R155, D168 and R123, which also form in viral product complexes 3-4A, 4A4B and 5A5B. No such networks form in the TRIF and 4B5A complexes, and residues in this region of the protease adopt the same conformations observed in the apo state. The absence or presence of electrostatic networks also correlate with the affinities of product binding, with the K_d of 4B5A being 10-fold weaker compared to products 4A4B and 5A5B. The greater catalytic efficacies of NS3/4A for substrates 4A4B and 5A5B relative to 4B5A (31) may also derive from the formation of electrostatic networks. However, short peptide may only partially mimic how the viral cleavage sequences are processed along the viral polyprotein in the natural cellular environment. Additional molecular features may further modulate the binding of TRIF and 4B5A, perhaps facilitated by the proline-rich regions contained in both proteins (11). Thus, the specificity of substrate processing by NS3/4A protease seems to arise from at least two distinct molecular interaction

patterns, which likely influence the order and kinetics of polyprotein processing during the HCV lifecycle.

In fact, these structural observations can be further linked to the known biology of NS3/4A processing during viral replication. Previous NS3-mediated cleavage assays of HCV polyprotein substrates revealed that NS4A is essential for the *trans* cleavage of junction 4B5A, but not required for the processing of junctions 4A4B and 5A5B (2, 22). Our structural analyses provide further insight into the molecular interactions underlying these previous findings. The NS4A cofactor likely stabilizes the tertiary protease fold required for the binding of NS3/4A substrates, and the binding of substrate 4B5A may absolutely depend on this particular protease conformation. Substrates 4A4B and 5A5B, however, may be able to induce these conformational changes through charge interactions, even in the absence of cofactor NS4A. Thus our findings support the previous published data for NS3 HCV polyprotein processing, and future research is warranted to better ascertain the dynamic mechanisms of substrate recognition.

The ability for HCV to establish chronic human infections is highly dependent on the viruses' ability to effectively replicate while simultaneously evading the host-cell immune response. The virally encoded NS3/4A protein plays an integral role in this process by mediating the cleavage of essential viral proteins and antiviral host cell adaptors. NS3/4A protease is thus a prime therapeutic target and great efforts have been devoted in the development of protease inhibitors, which have demonstrated efficacy in late phases of human clinical trials. Nevertheless, the high rate and error-prone nature of HCV

replication has led to the emergence of drug resistance against the most promising protease inhibitors to date, such as boceprevir, telaprevir and ITMN-191 (16, 17, 20, 30, 32, 36). Inhibitor potency often derives from molecular interactions that are not essential for substrate recognition and cleavage. Mutations in these regions of the protease can selectively prevent drug binding while still allowing for the recognition and cleavage of viral and host-cell substrates. Thus identification of the protease residues that are important for substrate binding is crucial and will ultimately facilitate the design of drugs that target these particular residues. A more detailed understanding of the mechanisms underlying viral and host-cell substrate recognition is therefore essential in facilitating a more rationale approach to the design of more robust NS3/4A protease inhibitors.

ACKNOWLEDGEMENTS

NMR chemical assignments were kindly provided by Herbert Klei of Bristol-Myers Squibb. We thank David Smith of the LS-CAT beamline at Argonne National Laboratory for data collection of the apo enzyme; we also thank Shivender Shandilya and Vukica Šrajer for data collection of the TRIF complex at BioCARS; Madhavi Nalam and Rajintha Bandaranayake for assistance with structural refinement; and Aysegul Ozen, Seema Mittal and Madhavi Kolli and for their computational support. Use of the Advanced Photon Source was supported by the U. S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. DE-AC02-06CH11357. Use of the BioCARS Sector 14 was supported by the National Institutes of Health, National Center for Research Resources, under grant number RR007707. Use of the LS-CAT Sector 21 was supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor for the support of this research program (Grant 085P1000817). The National Institutes of Health grant 2R01-GM4347 supported this work.

REFERENCES

1. Alexopoulou, L., A. C. Holt, R. Medzhitov, and R. A. Flavell. 2001. Recognition of double-stranded RNA and activation of NF-kappaB by Toll-like receptor 3. Nature 413:732-8.
2. Bartenschlager, R., L. Ahlborn-Laake, J. Mous, and H. Jacobsen. 1994. Kinetic and structural analyses of hepatitis C virus polyprotein processing. J Virol 68:5045-55.
3. Bode, W., and R. Huber. 1978. Crystal structure analysis and refinement of two variants of trigonal trypsinogen: trigonal trypsin and PEG (polyethylene glycol) trypsinogen and their comparison with orthorhombic trypsin and trigonal trypsinogen. FEBS Lett 90:265-9.
4. Brunger, A. T. 1992. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. Nature 355:472-5.
5. Carter, P., and J. A. Wells. 1988. Dissecting the catalytic triad of a serine protease. Nature 332:564-8.
6. COLLABORATIVE COMPUTATIONAL PROJECT, N. 1994. The CCP4 Suite: Programs for Protein Crystallography. Acta Crystallographica 50:760-763.
7. Davis, I. W., A. Leaver-Fay, V. B. Chen, J. N. Block, G. J. Kapral, X. Wang, L. W. Murray, W. B. Arendall, 3rd, J. Snoeyink, J. S. Richardson, and D. C. Richardson. 2007. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res 35:W375-83.
8. Delaglio, F., S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer, and A. Bax. 1995. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 6:277-93.
9. DeLano, W. L. 2008. The PyMOL Molecular Graphics System. DeLano Scientific LLC, San Carlos, CA.
10. Emsley, P., and K. Cowtan. 2004. Coot: model-building tools for molecular graphics. Acta Crystallogr D Biol Crystallogr 60:2126-32.
11. Ferreon, J. C., A. C. Ferreon, K. Li, and S. M. Lemon. 2005. Molecular determinants of TRIF proteolysis mediated by the hepatitis C virus NS3/4A protease. J Biol Chem 280:20483-92.
12. Foy, E., K. Li, R. Sumpter, Jr., Y. M. Loo, C. L. Johnson, C. Wang, P. M. Fish, M. Yoneyama, T. Fujita, S. M. Lemon, and M. Gale, Jr. 2005. Control of antiviral defenses through hepatitis C virus disruption of retinoic acid-inducible gene-I signaling. Proc Natl Acad Sci U S A 102:2986-91.
13. Foy, E., K. Li, C. Wang, R. Sumpter, Jr., M. Ikeda, S. M. Lemon, and M. Gale, Jr. 2003. Regulation of interferon regulatory factor-3 by the hepatitis C virus serine protease. Science 300:1145-8.
14. Gallinari, P., D. Brennan, C. Nardi, M. Brunetti, L. Tomei, C. Steinkuhler, and R. De Francesco. 1998. Multiple enzymatic activities associated with recombinant NS3 protein of hepatitis C virus. J Virol 72:6758-69.
15. Goddard, T. D., and D. G. Kneller. SPARKY 3. University of California, San Francisco.

16. He, Y., M. S. King, D. J. Kempf, L. Lu, H. B. Lim, P. Krishnan, W. Kati, T. Middleton, and A. Molla. 2008. Relative replication capacity and selective advantage profiles of protease inhibitor-resistant hepatitis C virus (HCV) NS3 protease mutants in the HCV genotype 1b replicon system. Antimicrob Agents Chemother 52:1101-10.
17. Kieffer, T. L., C. Sarrazin, J. S. Miller, M. W. Welker, N. Forestier, H. W. Reesink, A. D. Kwong, and S. Zeuzem. 2007. Telaprevir and pegylated interferon-alpha-2a inhibit wild-type and resistant genotype 1 hepatitis C virus replication in patients. Hepatology 46:631-9.
18. Krishnan, R., J. E. Sadler, and A. Tulinsky. 2000. Structure of the Ser195Ala mutant of human alpha--thrombin complexed with fibrinopeptide A(7--16): evidence for residual catalytic activity. Acta Crystallogr D Biol Crystallogr 56:406-10.
19. Krissinel, E., and K. Henrick. 2007. Inference of macromolecular assemblies from crystalline state. J Mol Biol 372:774-97.
20. Lenz, O., T. Verbinnen, T. I. Lin, L. Vijgen, M. D. Cummings, J. Lindberg, J. M. Berke, P. Dehertogh, E. Fransen, A. Scholliers, K. Vermeiren, T. Ivens, P. Raboisson, M. Edlund, S. Storm, L. Vrang, H. de Kock, G. C. Fanning, and K. A. Simmen. 2010. In vitro resistance profile of the HCV NS3/4A protease inhibitor TMC435. Antimicrob Agents Chemother.
21. Li, K., E. Foy, J. C. Ferreon, M. Nakamura, A. C. Ferreon, M. Ikeda, S. C. Ray, M. Gale, Jr., and S. M. Lemon. 2005. Immune evasion by hepatitis C virus NS3/4A protease-mediated cleavage of the Toll-like receptor 3 adaptor protein TRIF. Proc Natl Acad Sci U S A 102:2992-7.
22. Lin, C., B. M. Pragai, A. Grakoui, J. Xu, and C. M. Rice. 1994. Hepatitis C virus NS3 serine proteinase: trans-cleavage requirements and processing kinetics. J Virol 68:8147-57.
23. Major, M. E., and S. M. Feinstone. 1997. The molecular virology of hepatitis C. Hepatology 25:1527-38.
24. McCoy, A. J., R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, and R. J. Read. 2007. Phaser crystallographic software. J Appl Crystallogr 40:658-674.
25. Morris, R. J., A. Perrakis, and V. S. Lamzin. 2002. ARP/wARP's model-building algorithms. I. The main chain. Acta Crystallogr D Biol Crystallogr 58:968-75.
26. Nalam, M. N., A. Ali, M. D. Altman, G. S. Reddy, S. Chellappan, V. Kairys, A. Ozen, H. Cao, M. K. Gilson, B. Tidor, T. M. Rana, and C. A. Schiffer. 2010. Evaluating the substrate-envelope hypothesis: structural analysis of novel HIV-1 protease inhibitors designed to be robust against drug resistance. J Virol 84:5368-78.
27. Otwinowski, Z., and W. Minor. 1997. Processing of X-ray Diffraction Data Collected in Oscillation Mode, p. Volume 276: Macromolecular Crystallography, part A, p.307-326. In J. R. M. S. C.W. Carter, Eds. (ed.). Methods in Enzymology.
28. Prabu-Jeyabalan, M., E. A. Nalivaika, K. Romano, and C. A. Schiffer. 2006. Mechanism of substrate recognition by drug-resistant human

immunodeficiency virus type 1 protease variants revealed by a novel structural intermediate. J Virol 80:3607-16.
29. Romano, K. P., A. Ali, W. E. Royer, and C. A. Schiffer. 2010. Drug resistance against HCV NS3/4A inhibitors is defined by the balance of substrate recognition versus inhibitor binding. Proc Natl Acad Sci U S A.
30. Sarrazin, C., R. Rouzier, F. Wagner, N. Forestier, D. Larrey, S. K. Gupta, M. Hussain, A. Shah, D. Cutler, J. Zhang, and S. Zeuzem. 2007. SCH 503034, a novel hepatitis C virus protease inhibitor, plus pegylated interferon alpha-2b for genotype 1 nonresponders. Gastroenterology 132:1270-8.
31. Steinkuhler, C., G. Biasiol, M. Brunetti, A. Urbani, U. Koch, R. Cortese, A. Pessi, and R. De Francesco. 1998. Product inhibition of the hepatitis C virus NS3 protease. Biochemistry 37:8899-905.
32. Tong, X., S. Bogen, R. Chase, V. Girijavallabhan, Z. Guo, F. G. Njoroge, A. Prongay, A. Saksena, A. Skelton, E. Xia, and R. Ralston. 2008. Characterization of resistance mutations against HCV ketoamide protease inhibitors. Antiviral Res 77:177-85.
33. WHO 02/08/2010 2010, posting date. Vaccine Research: Hepatitis C. [Online.]
34. Wittekind, M., S. Weinheirner, Y. Zhang, and V. Goldfarb. August 8, 2002 2002. Modified forms of hepatitis C NS3 protease for facilitating inhibitor screening and structural studies of protease:inhibitor complexes. United States of America patent US Patent 6333186.
35. Yao, N., P. Reichert, S. S. Taremi, W. W. Prosise, and P. C. Weber. 1999. Molecular views of viral polyprotein processing revealed by the crystal structure of the hepatitis C virus bifunctional protease-helicase. Structure 7:1353-63.
36. Yi, M., Y. Ma, J. Yates, and S. M. Lemon. 2009. Trans-complementation of an NS2 defect in a late step in hepatitis C virus (HCV) particle assembly and maturation. PLoS Pathog 5:e1000403.
37. Yoneyama, M., M. Kikuchi, K. Matsumoto, T. Imaizumi, M. Miyagishi, K. Taira, E. Foy, Y. M. Loo, M. Gale, Jr., S. Akira, S. Yonehara, A. Kato, and T. Fujita. 2005. Shared and unique functions of the DExD/H-box helicases RIG-I, MDA5, and LGP2 in antiviral innate immunity. J Immunol 175:2851-8.

APPENDIX B

CLUSTERS OF BRANCHED ALIPHATIC SIDE CHAINS SERVE AS CORES OF STABILITY IN THE NATIVE STATE OF THE HISF TIM BARREL PROTEIN

AUTHOR CONTRIBUTIONS

This chapter is a collaborative project between Gangadhara N. Basavanapura, myself, Sagar V. Kathuria, Francesca Massi and C. Robert Matthews. Gangadhara N. Basavanapura expressed and purified HisF, conducted the kinetics experiments and analyzed and assigned the hydrogen deuterium exchange data. Francesca Massi and I acquired and processed the hydrogen deuterium exchange and assignment experiments. Francesca Massi verified the assignments of HisF. Sagar V. Kathuria did the ILV cluster and kinetic analysis of HisF. C. Robert Matthews and Gangadhara N. Basavanapura wrote the paper that includes this work.

This work is currently under review,

Gangadhara N. Basavanapura, Jennifer M. Laine, Sagar V. Kathuria, Francesca Massi and C. Robert Matthews. "Clusters of Branched Aliphatic Side Chains Serve as Cores of Stability in the Native State of the HisF TIM Barrel Protein." *Journal of Molecular Biology*

ABSTRACT

Clusters of branched aliphatic side chains, isoleucines, leucines and valines (ILV), have been found to stabilize partially-folded intermediates populated during the folding of TIM barrel proteins. The evidence supporting this conclusion derives from native-state amide hydrogen exchange (NS-HX) or pulse-quench HX analysis, which finds that amide hydrogens associated with ILV clusters are preferentially resistant to exchange with solvent deuterium in these intermediate states. The unusually long lifetime of the native state of the HisF TIM barrel from a thermophilic bacteria, T. maritima, enabled an NS-HX test of the role of ILV clusters in stabilizing its native conformation. Higher energy states offering protection against exchange were found to be associated with a pair of large ILV clusters. The clusters, however, do not segregate with the N- and C- terminal halves of the β-barrel, as might be expected from the $(ba)_4$ + $(ba)_4$ gene duplication thought to give rise to HisF. Sequence variation following duplication gave rise to ILV clusters that offer very strong protection against exchange in the β1-β2 and β4-β7 segments of the $(\beta\alpha)_8$ barrel. The strength of the protection against exchange through local fluctuations varies inversely with the mean squared fluctuations predicted from a simple normal mode analysis. The correlations between the ILV clusters, protection against hydrogen exchange and the vibrational properties demonstrate a key role for these clusters in determining the thermodynamic, dynamic and structural properties of HisF, and, by inference, other representatives of the most common platform for catalysis in biology.

INTRODUCTION

The $(ba)_8$, TIM barrel, motif one of the most common folds in biology,[1,2] has been the subject of numerous studies of its structure,[2,3-4] function,[5-8] folding,[9-12] design,[13-21] and evolution.[2,22-26] Evolution presumably recapitulated the ba module through gene duplication until its 8-fold manifestation resulted in a stable closed barrel with b1 hydrogen bonded to b8 and the 8 intervening a-helices forming a continuous amphipathic shell around the hydrophobic b-barrel.[24] The short loops/turns between the α-helices and the β-strands at one end of the barrel are crucial for stability.[27,28] The longer loops between the β-strands and the subsequent α-helices at the opposite end of the barrel invariably form the active site for a host of enzymes from all three superkingdoms.[29,30]

Although the hydrophobic effect, hydrogen bonding and electrostatic interactions all contribute to stabilizing the native fold of TIM barrels, one might ask whether there are sequence-specific contributions to stability that differentiate one barrel from another. Mutational analysis used to assess the contributions of individual side chains to stability invariably finds that most if not all buried side chains are involved in defining the free energy of the native state relative to the unfolded state.[31-33] However, these classical thermodynamic studies do not address the question of whether rare high energy microstates within the native manifold are stabilized in a sequence-dependent fashion.

We have approached this problem using protection against exchange of amide hydrogens for solvent deuterium under conditions favoring the native state of proteins.[34] Less stable segments readily expose their main chain amide hydrogens to solvent and are rapidly replaced by deuterium. More stable segments only slowly exchange with

deuterium. 2D NMR experiments allow site-specific measurement of the protection against exchange and, thereby, a high resolution assessment of the contributions of sequence to stability.

Application of the NS-HX NMR method to the alpha subunit of Trp synthase from *E. coli* (αTS) revealed patterns of protection that were found to reflect the structures of partially-folded states on the folding free energy surface.[28] The protection patterns vary with the sequence in a manner dictated by the location of a large cluster of isoleucine, leucine and valine side chains at the N-terminus of the barrel. We had previously hypothesized that ILV clusters, which are preferentially resistant to the penetration of water and/or hydroxide required for HX,[35] serve as cores of stability in partially-folded states of αTS.[36] Unfortunately, the time constraints of the NMR experiment, ~ 30 minutes for data collection, and the comparable lifetime of the native state at neutral pH and room temperature precluded an analysis of the protection in the fully-folded αTS TIM barrel and, thereby, its determinants of stability.

The *T. maritima* imidazole-3-glycerol phosphate synthase TIM barrel, the product of the HisF gene, proves to be an especially favorable candidate to explore sequence-specific contributions to the stability of the native state. Sterner and his colleagues,[22,37] have examined the evolution of this very stable HisF barrel and concluded that the $(\beta\alpha)_8$ motif arose from a $(\beta\alpha)_4 + (\beta\alpha)_4$ gene duplication event. Subsequent studies revealed differential stability of the N- and C- terminal $(\beta\alpha)_4$ modules of the HisF barrel,[38] supporting this view and suggesting that differential HX patterns might persist in these modules in the native state.

As will be shown, the unusual thermodynamic and kinetic folding properties of the HisF barrel permit a site-specific HX examination of the essential hydrogen-bonding network in the native thermodynamic state. Similar to their roles in stabilizing folding intermediates, ILV clusters preferentially persist in high energy microstates that define the essential features of the TIM barrel motif for HisF. The sequences, however, have evolved since the gene duplication event in such a way that the clusters are not simply associated with either the N- and/or C-terminal halves of HisF.

RESULTS

As a critical prelude to the NS-HX analysis of HisF, it was necessary to identify the species that appear during the folding reaction and the barriers that separate them along the reaction coordinate. Such data were essential in associating the HX protection patterns with on- and off-pathway intermediates in the αTS[28] and sIGPS TIM barrels.[39] By contrast, the unusual thermodynamic and kinetic folding properties of HisF enabled unique access to the exchange properties of its native state.

Equilibrium studies of HisF denaturation

Gdn-HCl induced equilibrium denaturation experiments were performed on HisF, and the disruption of secondary structure was monitored by far UV-CD spectroscopy. The reversibility was tested by comparing the transition curves generated from incubating the native state with increasing concentrations of denaturant with those generated by incubating the denatured state with decreasing concentrations of denaturant. Anticipating strong protection against hydrogen exchange in a thermophilic TIM barrel, we performed the folding study under the conditions selected for the NS-HX experiment, 40 ℃ and pH 7.2. Surprisingly, the CD-detected unfolding reaction required ~18 days at pH 7.2 and 40 ℃ to equilibrate (**Figure B.1(a)**). By contrast, the refolding reaction appeared to equilibrate within 24 hours (**Figure B.1(b)**). Although all of the data are consistent with a cooperative two-state unfolding transition, the contrasting responses in the approach to equilibrium imply the presence of an additional rapidly-accessible and native-like state on the unfolding side of the major barrier that typically separates the native state from partially folded states in folding reactions. After 18 days at pH 7.2 and 40 ℃, the estimated apparent free energy of folding is 10.81 kcal $mol^{-1}$for

Figure B.1. Chemical denaturation curves of HisF monitored by mean residue ellipticity at 222 nm and plotted as a function of Gdn-HCl concentration in a buffer containing 10 mM KPi, 1 mM K_2EDTA and 0.5 mM DTT at pH 7.2 and 40 ℃. (a) Unfolding equilibrium: 1 day (●), 2 days (○), 4 days (■), 10 days (▲), 13 days (▼) and 18 days (◆). (b) Refolding equilibrium: 1 day (●), 2 days (○), 4 days (■), 10 days (▲) and 18 days (◆). Solid lines indicate fits to a two-state model.

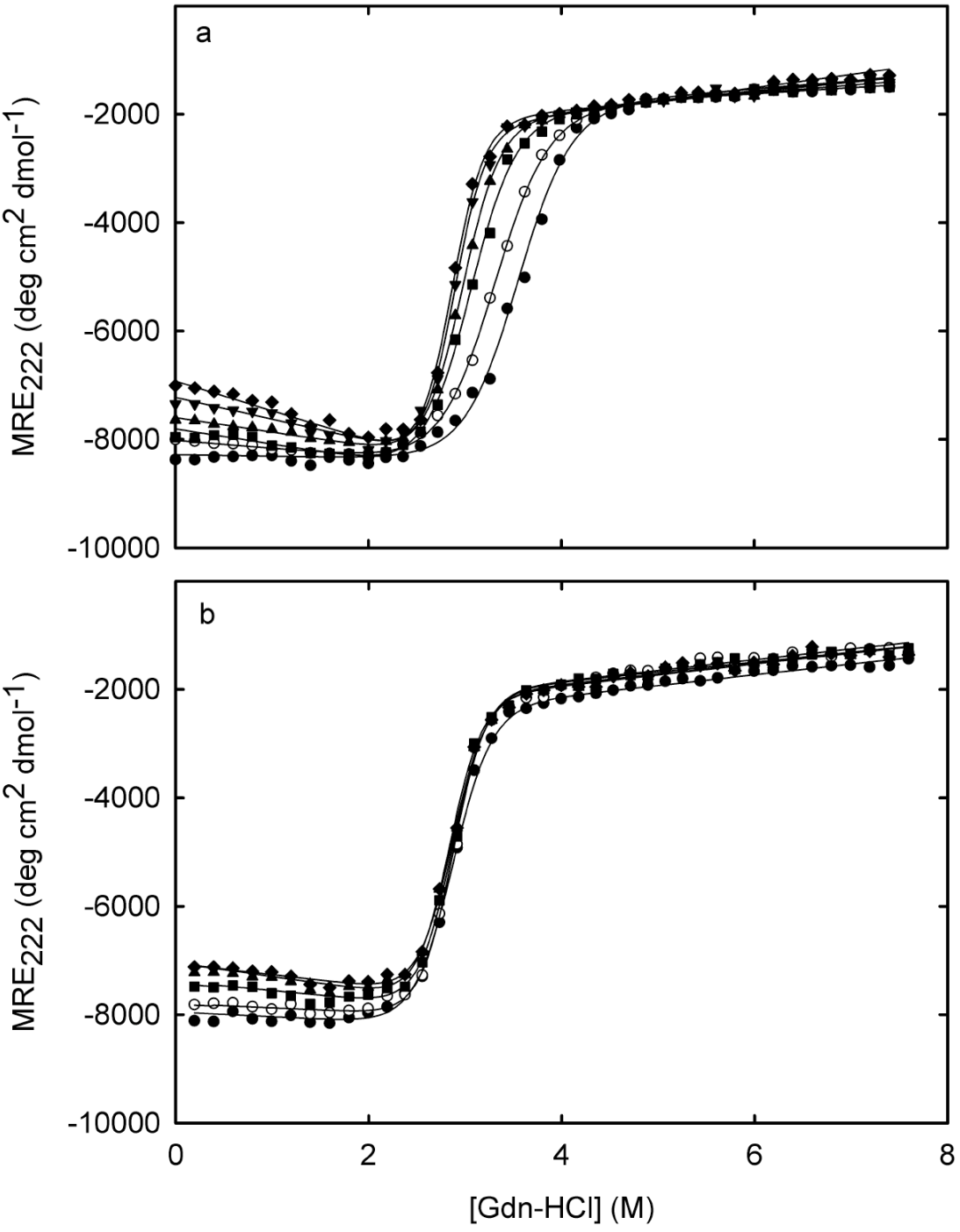

a
b
-2000
-4000
-6000
-8000
-10000
MRE222 (deg cm2 dmol-1)
0
2
4
6
8
[Gdn-HCl] (M)

Figure B.1

the unfolding equilibrium reaction and 10.10 kcal mol^{-1} for the refolding equilibrium reaction (**Table B.S1**). The small but progressive increase in the denaturant dependence of the ellipticity in the native baseline region over the time course of the experiment (**Figures B.1(a) & B.1(b)**) suggest that prolonged incubation in Gdn-HCl may lead to the slow degradation or aggregation of the protein. Although the equilibrium study did not provide reliable insights, kinetic folding studies proved to be a more fruitful approach towards elucidating the folding mechanism of HisF (see below).

Kinetic studies of the folding mechanism of HisF

Unfolding and refolding kinetic reactions of HisF were monitored by manual-mixing and stopped-flow CD. Unfortunately, optical studies of the refolding reaction were plagued by aggregation reactions below 1.5 M Gdn-HCl at 40 ℃ and pH 7.2. To enable access to strongly refolding conditions, the kinetic experiments were performed at 20 ℃ and pH 7.2 where self-association was not observed. The traces were fit to one or a sum of exponentials and a semi-log plot of the relaxation times is shown as a function of the denaturant concentration (**Figure B.2(a)**). A satisfactory fit of the unfolding traces between 4.4 and 5.6 M Gdn-HCl revealed a major slow phase that accelerates exponentially with increasing concentrations of Gdn-HCl and a minor faster phase, <5%, in the hundreds of seconds time range that is weakly dependent on the denaturant concentration. Above 5.6 M Gdn-HCl, only a single, denaturant-dependent phase was observed. Refolding was more complex, with a sub-millisecond burst-phase in ellipticity (**Figure B.2(b)**), a small amplitude phase, ~10 %, in the 1-10 s time range that accelerates at higher Gdn-HCl concentration and a major slow phase that accelerates exponentially between 2.6 and ~1.6 M Gdn-HCl before rolling over to a

Incubation Time (days)	$\Delta G_{app}°$ (kcal mol^{-1})		$-m_{app}$ (kcal mol^{-1} M^{-1})		C_m (M)	
	Unf.	Ref.	Unf.	Ref.	Unf.	Ref.
1	7.98 ± 0.21	8.88 ± 0.17	2.25 ± 0.05	3.10 ± 0.05	3.54	2.86
2	7.95 ± 0.15	10.51 ± 0.16	2.41 ± 0.05	3.65 ± 0.07	3.29	2.87
10	9.52 ± 0.10	10.09 ± 0.17	3.20 ± 0.04	3.52 ± 0.12	2.97	2.86
18	10.81 ± 0.16	10.10 ± 0.33	3.78 ± 0.02	3.53 ± 0.15	2.86	2.86

Table B.S1. Apparent thermodynamic parameters derived from 2-state fits of the equilibrium unfolding and refolding curves of the Gdn-HCl denaturation of HisF at pH 7.2 and 40 °C monitored by far UV-CD spectroscopy.

Figure B.2. (a) A chevron plot of the observed refolding (closed symbols) and unfolding (open symbols) relaxation times of HisF, monitored by manual-mixing CD (●) and stopped flow CD (SF-CD), (▲) at 222 nm, extracted from exponential fits and plotted as a function of final denaturant concentrations at 20 ℃. The assignments of the kinetic phases to steps in the proposed folding mechanism (Scheme 1) are indicated. The buffer was 10 mm KPi, 1 mM K_2EDTA and 0.5 mM DTT at pH 7.2. (b) The 4 day equilibrium refolding curve for HisF monitored by CD at 222 nm (●) at 20 ℃. The refolding burst-phase ellipticity obtained from SF-CD (▲) under the same conditions is also shown. The dashed line indicates a linear extrapolation of the ellipticity of the unfolded state.

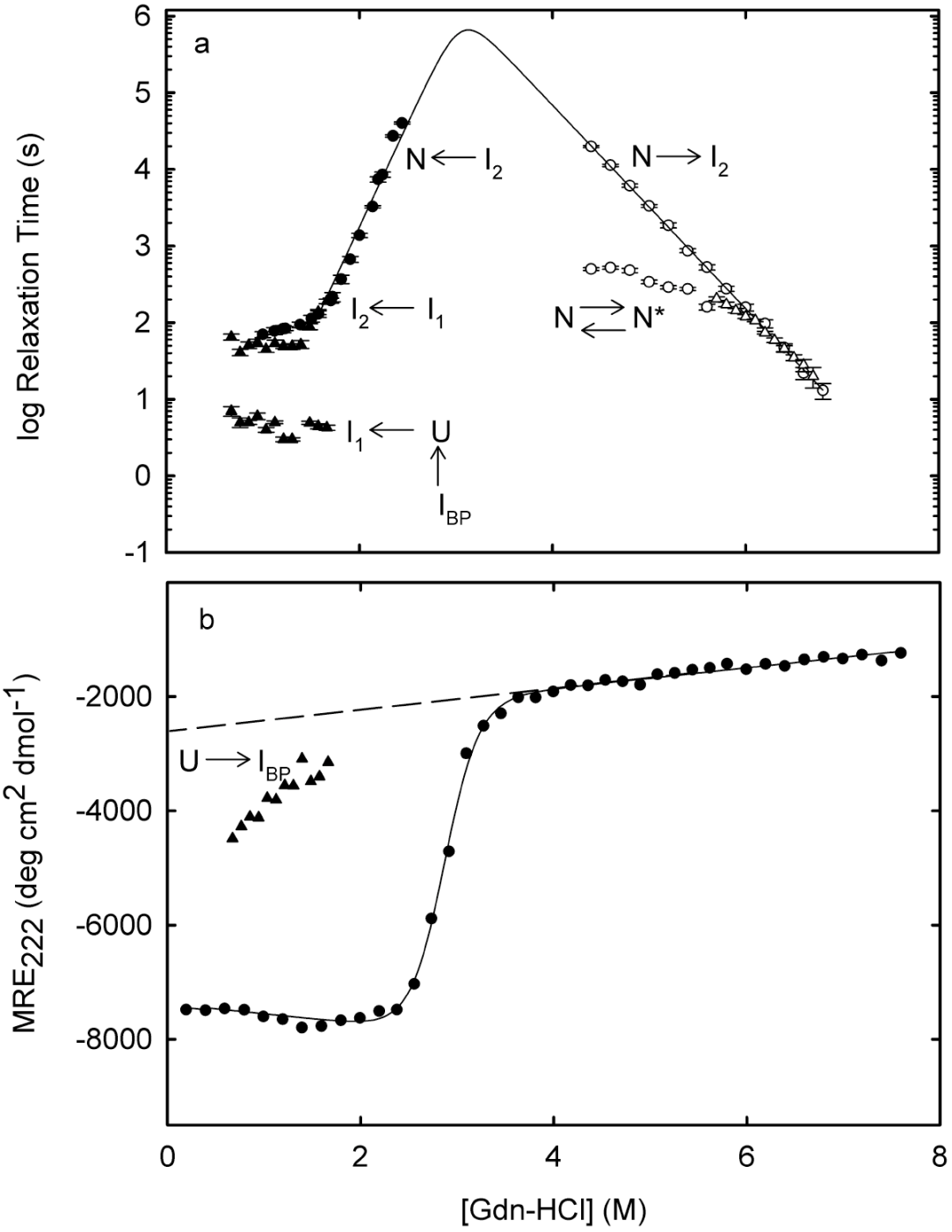

a
log Relaxation Time (s)
N ← I2
N → I2
I2 ← I1
N → N*
I1 ← U
IBP
b
MRE222 (deg cm2 dmol-1)
U → IBP
[Gdn-HCl] (M)

Figure B.2

denaturant-independent phase whose relaxation time is ~60 s (**Figure B.2(a)**). These data are consistent with a five state folding model involving an off-pathway early folding intermediate and two sequential on-pathway intermediates, as observed previously for sIGPS.[40] The small amplitude unfolding phase in the hundreds of seconds time range, however, requires an additional species on the native side of the major barrier to unfolding, N*. The small amplitude of this phase and its modest denaturant dependence implies a subtle change in the secondary structure compared to the native state.

$$\begin{array}{l} [N \leftrightarrows N^*] \leftrightarrows I_2 \leftrightarrows I_1 \leftrightarrows U \\ \qquad\qquad\qquad\quad \downarrow\uparrow \\ \qquad\qquad\qquad\quad I_{BP} \end{array} \qquad \text{Scheme I}$$

The assignments of the steps in the mechanism to the refolding and unfolding phases are shown in **Figure B.2(a) & B.2(b)**. A detailed analysis of the kinetic folding mechanism of HisF will be presented elsewhere.

The denaturant dependence of the relaxation times for the major unfolding and refolding phases in HisF, an inverted V shape or chevron shape, is typical of reversible protein folding reactions. By linear extrapolations, the chevron can be used to estimate the rate constants for the rate-limiting unfolding and refolding reactions in the absence of denaturant.[41] The inverse of the relaxation time, t, equals the rate constant, k. Although the predicted refolding rate constant from the refolding leg of the chevron, k_f = 144 s^{-1}, is in the range observed previously for another TIM barrel,[42] the unfolding rate constant, $k_u = 7.12 \times 10^{-11}$ s^{-1}, is responsible for the substantial stability of the HisF native state relative to the I_2 state; $DG_{N/I2} = -RT \ln (k_u/k_f) = 13.85$ kcal mol^{-1}. Pertinent to the NS-HX study, the lifetime of the N state in the absence of Gdn-HCl is an astounding

~445 years at pH 7.2 and 20 ℃. The unfolding leg of the chevron at 40 ℃ and pH 7.2, the conditions for the NS-HX study, reveals a lifetime of the native state of ~116 years (k_u = 2.72 × 10^{-10} s^{-1}, & $DG_{N/I2}$ = 12.44 kcal mol^{-1}. **Figure (B.S1)**). The denaturant dependence of the unfolding relaxation time is such that the lifetime of the native state exceeds 115 days up to 2.0 M Gdn-HCl at pH 7.2 and 40 ℃. These extraordinary lifetimes enable a unique assessment of the HX properties of the native state of a TIM barrel protein.

Native-state HX-NMR experiments

The availability of the main chain amide ^{1}H and ^{15}N resonance assignments for HisF,[43] provided the starting point for the NS-HX experiments. Examination of the TROSY 2D ^{15}N -^{1}H correlation spectrum (**Figure B.S2(a))** revealed that 212 out of 240 crosspeaks were sufficiently well resolved at 40 ℃ to enable accurate measurements of their intensities for this study. Exchange of the amide hydrogen atoms for deuterium was monitored by periodically recording the TROSY 2D ^{15}N -^{1}H correlation spectrum as a function of time after dissolving lyophilized protein into $^{2}H_2O$ buffer at pH 6.8 (meter reading) at 40 ℃ (**Figures B.S2(b) & B.S2(c)**). Note that the relationship between pH and pD, pD = pH (meter reading) + 0.4,[44] means that the actual acidity in solution was equivalent to pH 7.2. The precision of the measurements was enhanced by leaving the sample in the NMR probe for the entire time course of the experiment, 10 days.

The 90 Class I NHs exchanged within the time required to dissolve the sample, shim the magnet and collect a useful 2D spectrum, 30 min. The 70 Class II NHs underwent > 95% exchange over 7 days, and the 52 Class III NHs did not undergo detectable exchange over the full 10 day course of the experiment. The exchange rate

Figure B.S1. Chevron plot of unfolding (open circles, open triangles) and refolding (closed circles, closed triangles) phases of HisF at pH 7.2 and 40 °C. Data were collected by manual-mixing and stopped-flow CD spectroscopy and fit to exponentials as described in Materials and Methods. Assignments of the phases to specific steps in the folding mechanism shown in Scheme 1 are indicated. The buffer used was 10 mm KPi, 1 mM K_2EDTA and 0.5 mM DTT at pH 7.2 and 40 °C.

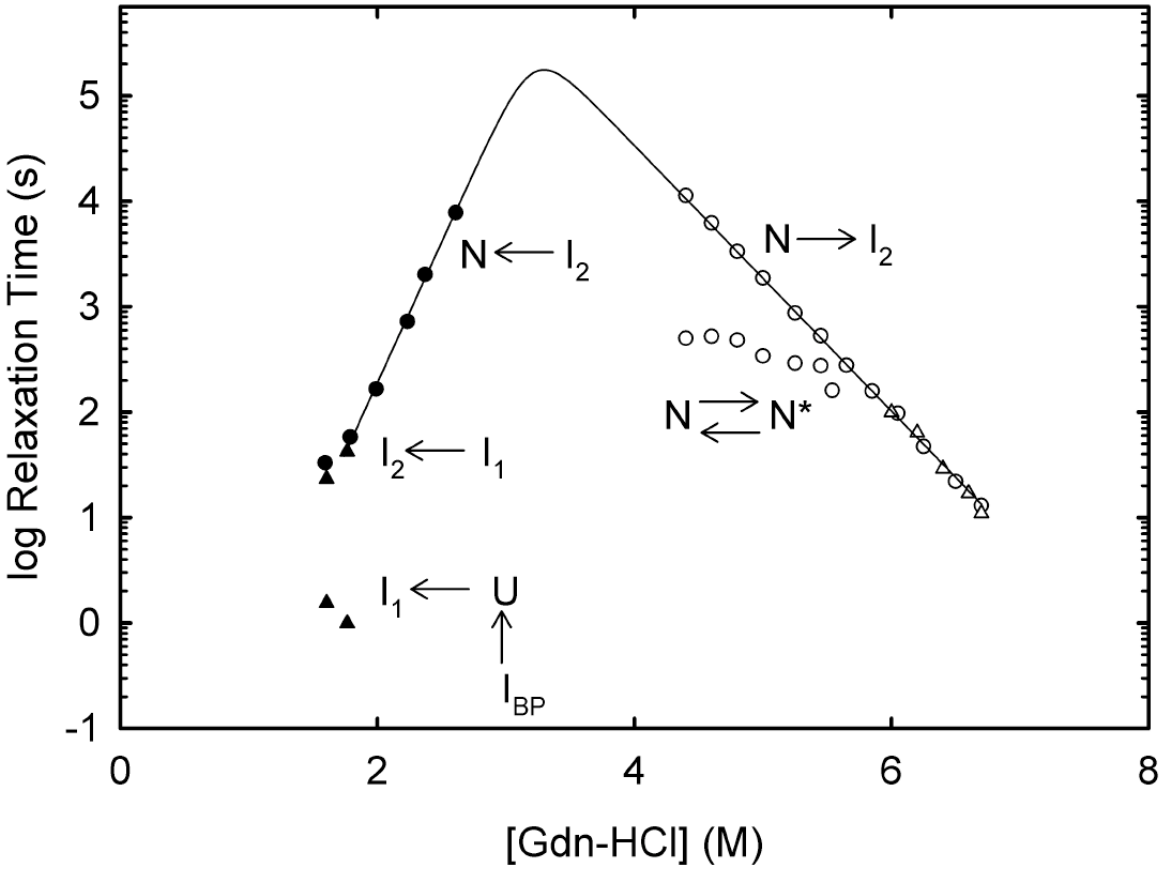


Figure B.S1

Figure B.S2. ^{15}N-^{1}H TROSY 2D correlation spectra of HisF in (a) aqueous buffer; (b) after 30 minutes in $^{2}H_2O$ buffer; (c) after 10 days in $^{2}H^{2}O$ buffer at pH 7.2 and 40 °C. The buffer contained 10 mM KPi, 50 mM KCl and 1 mM K_2EDTA.

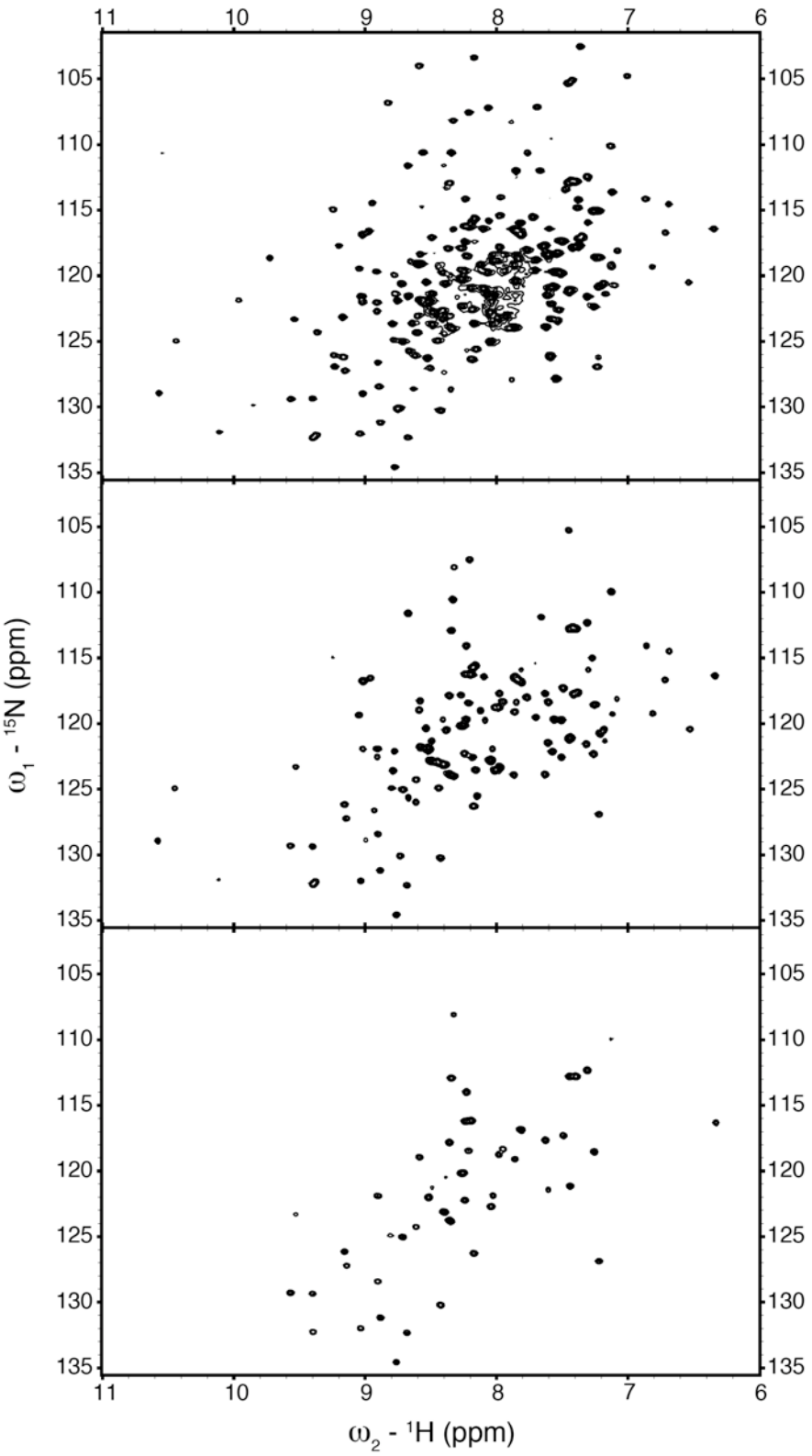


Figure B.S2

constants for the Class II NHs were obtained by fitting the fractional occupancy of hydrogen as a function of time to a single exponential function. Traces and fits at 0 M Gdn-HCl for representative set of NHs are shown in **Figures B.3(a)** and **B.3(b**) for two different time regimes, 30 - 1,000 minutes and 30 - 10,000 minutes, spanned by these experiments. From these measurements, protection factors (PFs) were calculated as the ratio of the exchange rate constant of the amino acid in an unstructured peptide, $k_{int,}$ to the observed exchange rate constant for a particular amino acid in the folded protein $k_{obs,}$ PF = k_{int}/k_{obs}. The k_{int} was taken as the exchange rate for amino acids in random coil conformation calculated using the program Sphere (www.fccc.edu/research/labs/roder/sphere/). The protection factors range from 3.59×10^3 for the more rapidly exchanging NHs to 1.09×10^7 for slowly exchanging NHs in the absence of Gdn-HCl (Table S2). Although there was no evidence in the NMR spectra for the time-dependent changes observed in the native baseline region of the CD data (**Figures B.1(a) and B.1(b)**), the fits of the hydrogen occupancies were only considered over the first 7 days of exchange. Less than 10% of the native ellipticity was lost in this time frame.

EX1 vs. EX2 limit: Characterizing the exchange regime

To interpret the protection factors for HisF, one must determine whether exchange is limited by the opening rate constant for breaking the hydrogen bond, the EX1 limit, or by the fraction of the open state required for exchange, the EX2 limit.[45] This determination can be made by comparing the k_{obs} values at two different pH values.[45,46-48] In the EX1 limit, a linear dependence in the log-log plot is expected to intersect the y-axis at the same log unit as the x-axis and have a gradient of one. In the EX2 limit, one

Residue	Structure[b]	k_{obs} (s^{-1})	k_{int} (s^{-1})	Protection Factor[c]	$\Delta G°_{HX}$ (kcal mol^{-1})[d]
V12	β1	2.829×10^{-5}	2.17	7.67×10^{4}	6.99 ± 0.01
R16	β turn	6.148×10^{-5}	29.1	4.73×10^{5}	8.12 ± 0.01
S29	loop	2.769×10^{-4}	25.5	9.21×10^{4}	7.12 ± 0.03
V33	α1	4.888×10^{-4}	1.88	3.84×10^{3}	5.13 ± 0.04
E34	α1	4.307×10^{-4}	3.68	8.54×10^{3}	5.62 ± 0.08
L35	α1	5.747×10^{-4}	3.06	5.34×10^{3}	5.33 ± 0.03
G36	α1	1.097×10^{-4}	18.8	1.71×10^{5}	7.49 ± 0.04
Y39	α1	3.305×10^{-5}	10.1	3.05×10^{5}	7.85 ± 0.02
S40	α1	7.614×10^{-5}	43.1	5.66×10^{5}	8.23 ± 0.03
E41	α1	1.881×10^{-4}	10.1	5.36×10^{4}	6.77 ± 0.02
I42	loop	2.629×10^{-5}	2.17	8.25×10^{4}	7.04 ± 0.02
I44	loop	1.194×10^{-5}	4.51	3.77×10^{5}	7.98 ± 0.02
D45	loop	2.033×10^{-5}	4.86	2.39×10^{5}	7.70 ± 0.03
I52	loop	1.753×10^{-6}	2.02	1.15×10^{6}	8.67 ± 0.06
S55	loop	8.681×10^{-4}	38.4	4.42×10^{4}	6.65 ± 0.05
E67	α2	8.503×10^{-5}	3.68	4.32×10^{4}	6.63 ± 0.02
A70	α2	7.672×10^{-6}	11.9	1.55×10^{6}	8.86 ± 0.02
Q72	loop	9.557×10^{-5}	13.4	1.40×10^{5}	7.36 ± 0.02
I73	loop	6.337×10^{-4}	4.84	7.63×10^{3}	5.56 ± 0.03
I75	loop	4.946×10^{-4}	2.02	4.08×10^{3}	5.17 ± 0.05
F77	β3	8.188×10^{-6}	5.43	6.63×10^{5}	8.33 ± 0.04
T78	β3	1.740×10^{-5}	16	9.19×10^{5}	8.53 ± 0.03
G80	β3	3.676×10^{-5}	22.1	6.01×10^{5}	8.27 ± 0.02
A89	α3	1.261×10^{-5}	26	2.06×10^{6}	9.04 ± 0.03
S90	α3	1.211×10^{-5}	38.4	3.17×10^{6}	9.30 ± 0.04
E91	α3	1.756×10^{-5}	10.1	5.75×10^{5}	8.24 ± 0.03
L94	α3	4.044×10^{-6}	2.54	6.28×10^{5}	8.30 ± 0.05
G96	loop	2.070×10^{-4}	50.6	2.44×10^{5}	7.71 ± 0.06
A97	loop	5.214×10^{-6}	24.2	4.64×10^{6}	9.54 ± 0.06
D98	loop	5.720×10^{-5}	8.25	1.44×10^{5}	7.38 ± 0.04
E108	loop	1.022×10^{-3}	3.68	3.59×10^{3}	5.09 ± 0.07
I113	α4	3.712×10^{-4}	1.88	5.06×10^{3}	5.30 ± 0.04
T114	α4	7.457×10^{-5}	8.21	1.10×10^{5}	7.21 ± 0.04
Q115	α4	1.044×10^{-4}	29.8	2.85×10^{5}	7.81 ± 0.03
Q118	α4	3.281×10^{-6}	18.8	5.73×10^{6}	9.67 ± 0.02
F120	α4	2.050×10^{-4}	14.9	7.26×10^{4}	6.96 ± 0.03
G121	loop	1.724×10^{-4}	35	2.03×10^{5}	7.60 ± 0.02
A124	loop	1.070×10^{-4}	26	2.42×10^{5}	7.71 ± 0.04
D130	β5	1.941×10^{-4}	4.86	2.50×10^{4}	6.29 ± 0.05
F141	β turn	5.805×10^{-6}	6.83	1.17×10^{6}	8.69 ± 0.07
T142	β turn	1.173×10^{-3}	16	1.36×10^{4}	5.92 ± 0.03
K147	β turn	6.918×10^{-6}	19.7	2.84×10^{6}	9.24 ± 0.04
I151	β turn	8.456×10^{-5}	4.51	5.33×10^{4}	6.76 ± 0.05
L152	β turn	5.631×10^{-4}	2.54	4.51×10^{3}	5.23 ± 0.05
L153	loop	5.923×10^{-6}	2.66	4.49×10^{5}	8.09 ± 0.03
D155	α5	1.507×10^{-5}	13.7	9.08×10^{5}	8.53 ± 0.04
W156	α5	9.324×10^{-6}	4.23	4.53×10^{5}	8.10 ± 0.02
E159	α5	7.429×10^{-6}	3.68	4.95×10^{5}	8.15 ± 0.03
E161	α5	1.462×10^{-4}	3.68	2.51×10^{4}	6.30 ± 0.06

R163	loop	1.085×10^{-5}	26	2.39×10^{6}	9.13 ± 0.07
S172	β6	1.111×10^{-4}	60.9	5.48×10^{5}	8.21 ± 0.05
M186	α6	1.773×10^{-4}	11.4	6.42×10^{4}	6.88 ± 0.05
F189	α6	4.507×10^{-6}	15.6	3.46×10^{6}	9.36 ± 0.07
L193	α6	6.190×10^{-6}	2.48	4.06×10^{5}	8.02 ± 0.04
T194	loop	1.633×10^{-6}	8.6	5.26×10^{6}	9.62 ± 0.06
S201	β7	2.094×10^{-5}	38.4	1.83×10^{6}	8.96 ± 0.05
H209	α7	2.501×10^{-5}	15.9	6.35×10^{5}	8.30 ± 0.03
L211	α7	3.451×10^{-5}	4.95	1.43×10^{5}	7.38 ± 0.04
E212	α7	4.789×10^{-4}	3.13	6.53×10^{3}	5.46 ± 0.02
L215	α7	9.099×10^{-6}	4.95	5.44×10^{5}	8.21 ± 0.02
A218	loop	1.376×10^{-4}	24.2	1.75×10^{5}	7.51 ± 0.03
D219	loop	1.631×10^{-4}	8.25	5.05×10^{4}	6.73 ± 0.05
A221	β8	1.082×10^{-5}	16.4	1.51×10^{6}	8.85 ± 0.04
A223	β8	2.068×10^{-5}	10.1	4.88×10^{5}	8.14 ± 0.04
A224	loop	1.502×10^{-6}	16.4	1.09×10^{7}	10.0 ± 0.09
E239	α8	6.391×10^{-5}	6.7	1.04×10^{5}	7.18 ± 0.03
K243	α8	1.207×10^{-3}	19.7	1.63×10^{4}	6.03 ± 0.04
V246	loop	3.767×10^{-6}	4.84	1.28×10^{6}	8.74 ± 0.03

[a]The buffer contained 10 mM potassium phosphate, 50 mM KCl, 1 mM K_2EDTA at pH 7.2.
[b]The element of secondary structure containing the residue.
[c]Protection Factor = (k_{int}/k_{obs})
[d]$\Delta G°_{HX} = -RT \ln(k_{obs}/k_{int})$

Table B.S2. Protection factors and ΔG°HX for main chain NHs in HisF in the absence of Gdn-HCl at pH 7.2 and 40 °Ca.

Figure B.3. Normalized intensity of ^{15}N-^{1}H cross peak vs log time plots for a representative set of NHs that exchange over the time frame (a) 30 - 1,000 min (b) 30 - 10,000 min at 40 ℃ and pH 7.2 in deuterated 10 mM KPi, 50 mM KCl and 1 mM K_2EDTA. The presence of 50 mM KCl, included to mimic the solvent for the resonance assignments,[43] had no detectable effect on the rate constants for the major unfolding and refolding phases corresponding to the [N ⇆ N*] ⇆ I_2 reaction (data not shown) (Scheme 1) and, therefore, the lifetime of the native state of HisF.

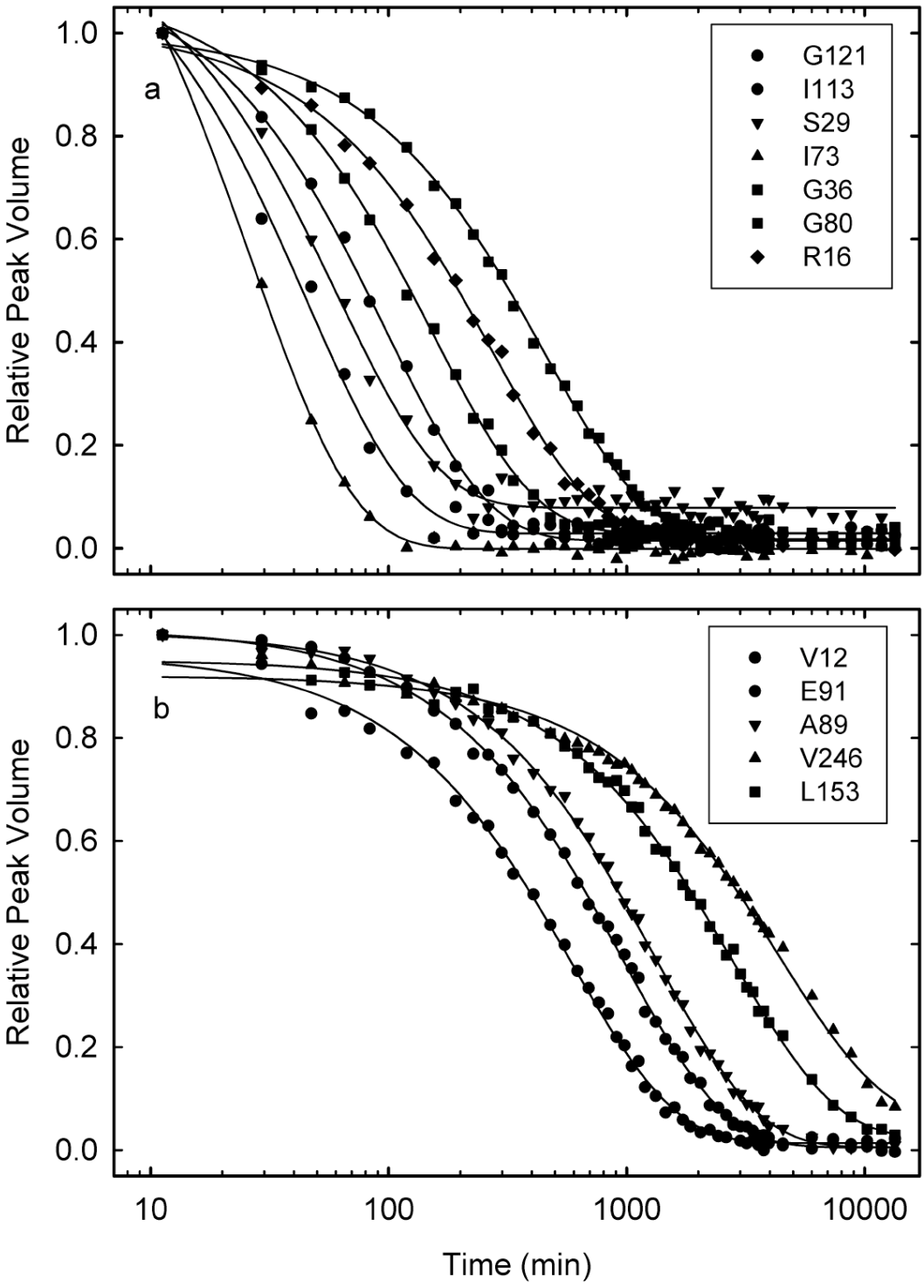

a
G121
I113
S29
I73
G36
G80
R16
Relative Peak Volume
1.0
0.8
0.6
0.4
0.2
0.0
b
V12
E91
A89
V246
L153
Relative Peak Volume
1.0
0.8
0.6
0.4
0.2
0.0
10
100
1000
10000
Time (min)

Figure B.3

expects a linear dependence intersecting the y-axis at one log unit higher than the x-axis and with a gradient of one. The shift along the y-axis reflects the known acceleration of the rate-limiting exchange reaction with increasing pH above ~2.3.[49] Although the raw data do not follow the expectation of either the EX1 or EX2 regime (**Figure B.S3(a)**), when corrected for the increase in the stability of the native state of HisF with pH (**Figure B.S3(b)**), the exchange process for the Class II NHs was found to obey the EX2 limit (**Figure B.4**). In this limit, the protection factors reflect the free energy difference between the closed and open forms of the H-bond forming/breaking reactions, $DG°_{HX}$.[50] $\Delta G°_{HX}$ is defined as $\Delta G°_{HX} = -RT\ ln(k_{obs}/k_{int})$. The $DG°_{HX}$ values for the Class II NHs in the absence of denaturant range from 5 to 11 kcal mol^{-1} and are shown in **Table B.S2**. Given the exceedingly long lifetime of the native state of HisF under these conditions, ~116 years in the absence of denaturant, the observed exchange over a period of 10 days must occur within the native manifold of states rather than from partially-folded states on the unfolded side of the rate-limiting barrier. Although the exchange behavior of the Class I and Class III NHs could not be measured, if both exchange by an EX2 mechanism, their $\Delta G°_{HX}$ value would be less than 5 kcal mol^{-1} and greater than 11 kcal mol^{-1} respectively.

HX protection patterns in HisF

The protection patterns for the Class I, Class II and Class III NHs are shown in **Figures B.5** and **B.6**. The rapidly exchanging Class I NHs preferentially appear in the active-site loops at the C-termini of the β-strands, which for β4, β5 and β8, are either

Figure B.S3. (a) The log-log plot of the observed exchange rate constants measured at pH 7.2 and pH 6.2 at 40 °C for amide hydrogens of HisF, before correction for the effect of pH on stability. The continuous line represents EX2 limit, and the dotted line represents EX1 limit. (b) The chevron plots for the major unfolding and refolding phases in HisF at pH 6.2 (●) and pH 7.2 (▲) at 40 °C, obtained from manual mixing CD experiments at 222 nm. The buffer was 10 mm KPi, 1 mM K2EDTA and 0.5 mM DTT. The increase in the pH from 6.2 to 7.2 slows the major unfolding reaction by 5-fold while leaving the major refolding reaction unchanged. Correcting the kobs for the 5-fold slowing of the unfolding reaction at pH 7.2 yields the correlation shown in **Figure B.4**.

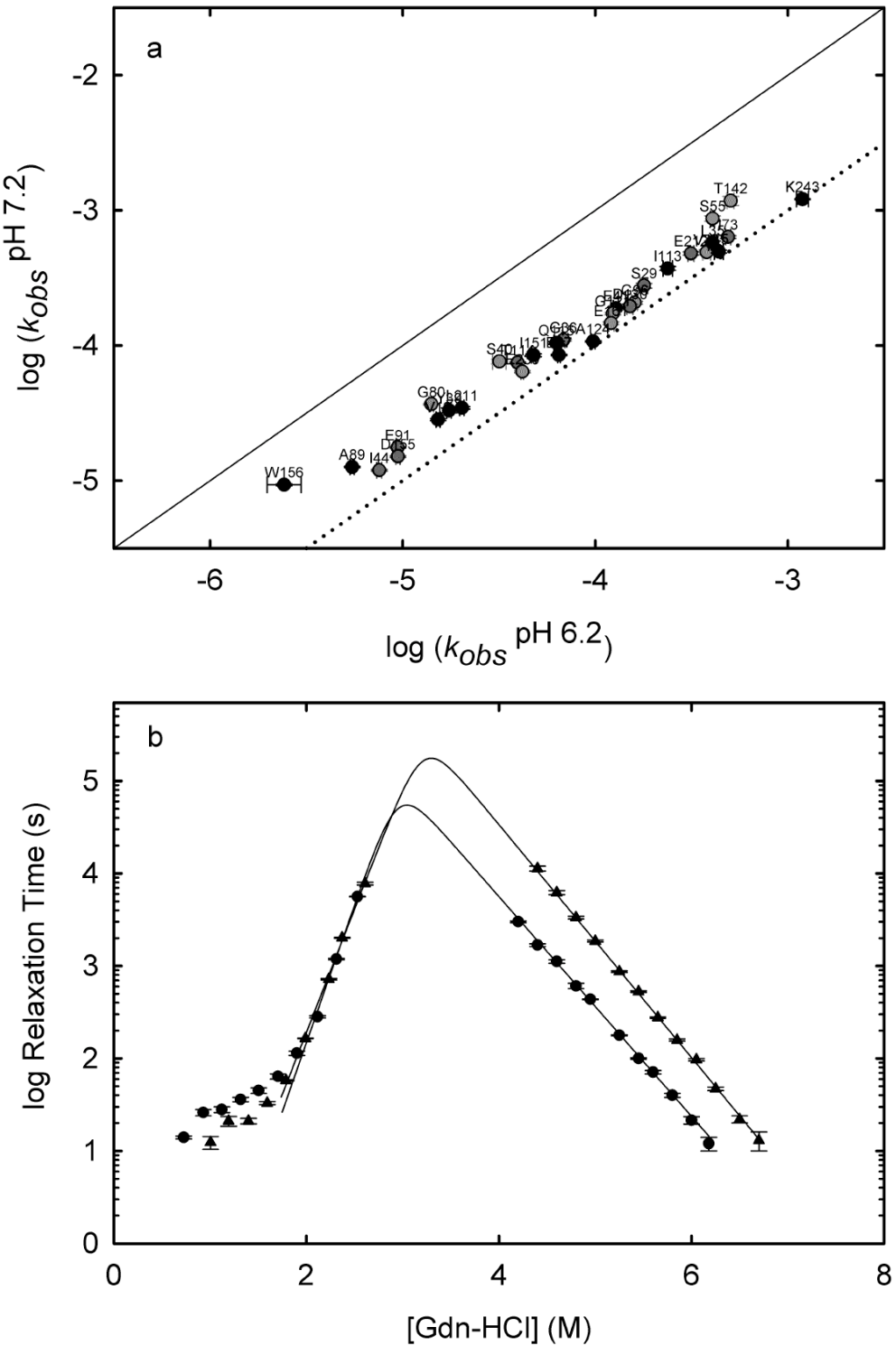

a
log (k_{obs} pH 7.2)
log (k_{obs} pH 6.2)
-2
-3
-4
-5
-6
-5
-4
-3
T142
K243
S55
I113
S29
A124
W156
A89
I44
E91
b
log Relaxation Time (s)
[Gdn-HCl] (M)
0
1
2
3
4
5
0
2
4
6
8

Figure B.S3

Figure B.4. (a) The log-log plot of the observed exchange rate constants measured at pH 7.2 and pH 6.2 at 40 ℃ for amide hydrogens of HisF, after correction for the effect of pH on stability (**Figure B.S3**(b)). The continuous line represents the EX2 limit, and the dotted line represents EX1 limit.

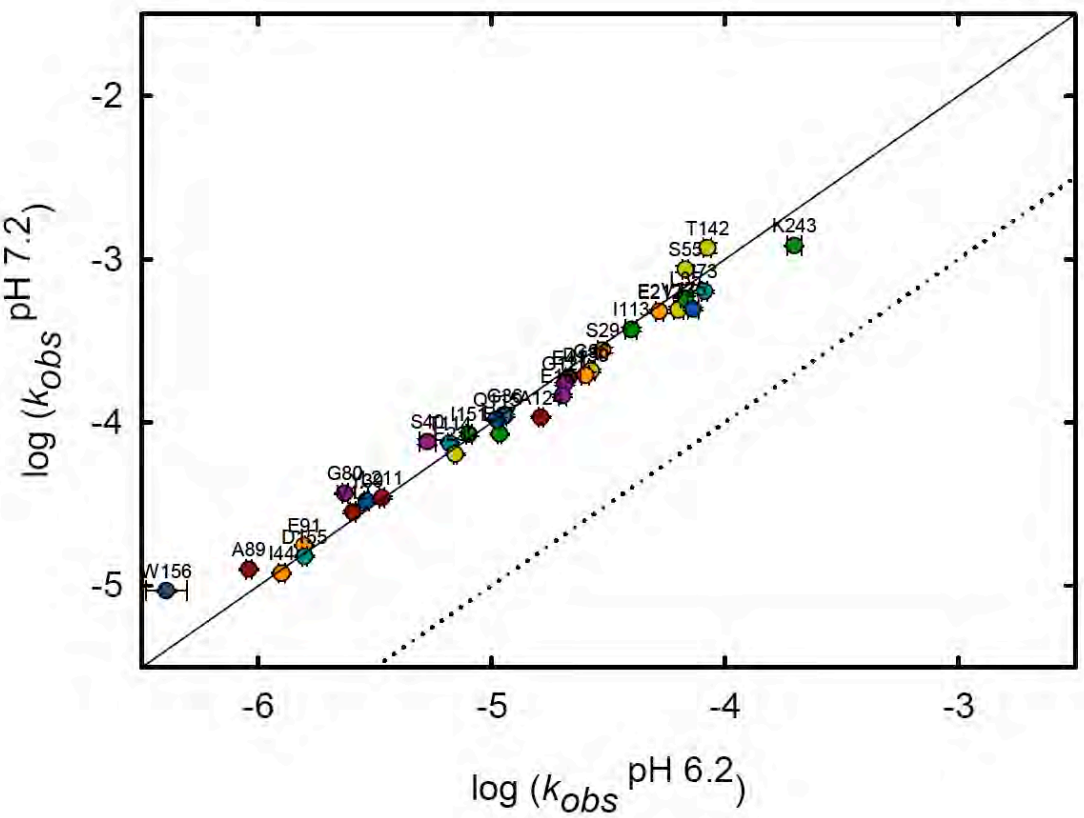

-2
-3
-4
-5
-6
-5
-4
-3
log (k_{obs} pH 7.2)
log (k_{obs} pH 6.2)
T142
K243
S55
I113
S29
A89
I44
W156

Figure B.4

Figure B.5 A three-dimensional representation of the location of the three classes of protection for HisF at pH 7.2 and 40 °C and the location of the five ILV clusters. (a) Class I, red, $\Delta G°_{HX} < 5$ kcal mol^{-1}; (b) Class II, yellow, $5 < \Delta G°_{HX} < 7$ kcal mol^{-1}, green, $7 < \Delta G°_{HX} < 9$ kcal mol^{-1}, blue, $9 < \Delta G°_{HX} < 11$ kcal mol^{-1}; (c) Class III, purple, $\Delta G°_{HX} > 11$ kcal mol^{-1}; (d) Five ILV clusters: Cluster 1 (green), Cluster 2 (red), Cluster 3 (cyan), Cluster 4 (purple), and Cluster 5 (black). The structures were generated using PyMol and Protein Data Bank entry 1THF.

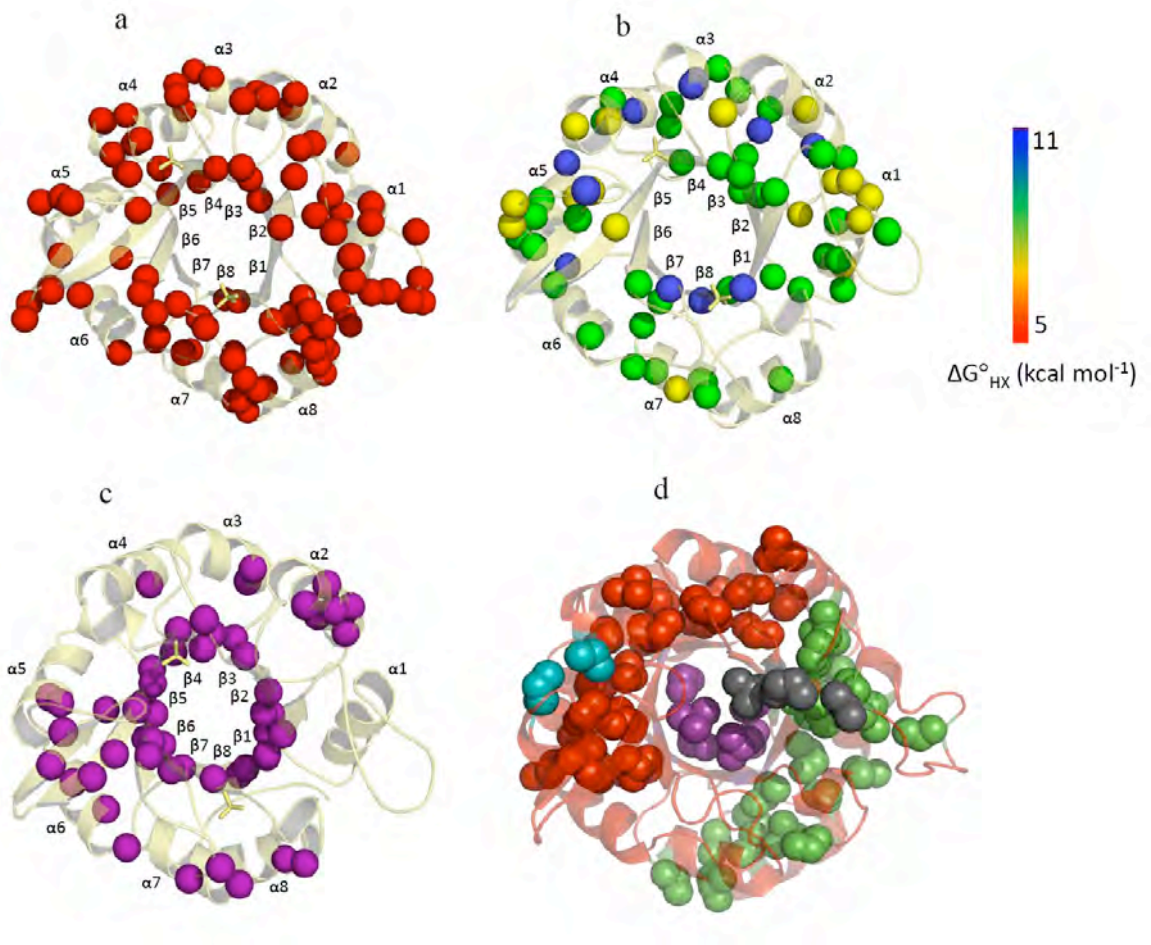

a
b
c
d
α1
α2
α3
α4
α5
α6
α7
α8
β1
β2
β3
β4
β5
β6
β7
β8
11
5
ΔG°HX (kcal mol-1)

Figure B.5

Figure B.6. 2D representation of protection against exchange in HisF at pH 7.2 and 40 °C. Class I, red, $\Delta G°_{HX} < 5$ kcal mol^{-1}; Class II: yellow, $\Delta G°_{HX}$ 5 - 7 kcal mol^{-1}, green, $\Delta G°_{HX}$ 7 - 9 kcal mol^{-1} and blue, $\Delta G°_{HX}$ 9 - 11 kcal mol^{-1}; Class III, purple, $\Delta G°_{HX} > 11$ kcal mol^{-1}. Open circles indicate either the absence of the NMR assignment or the inability to obtain accurate fits due to spectral overlap and (◌) indicates a proline residue.

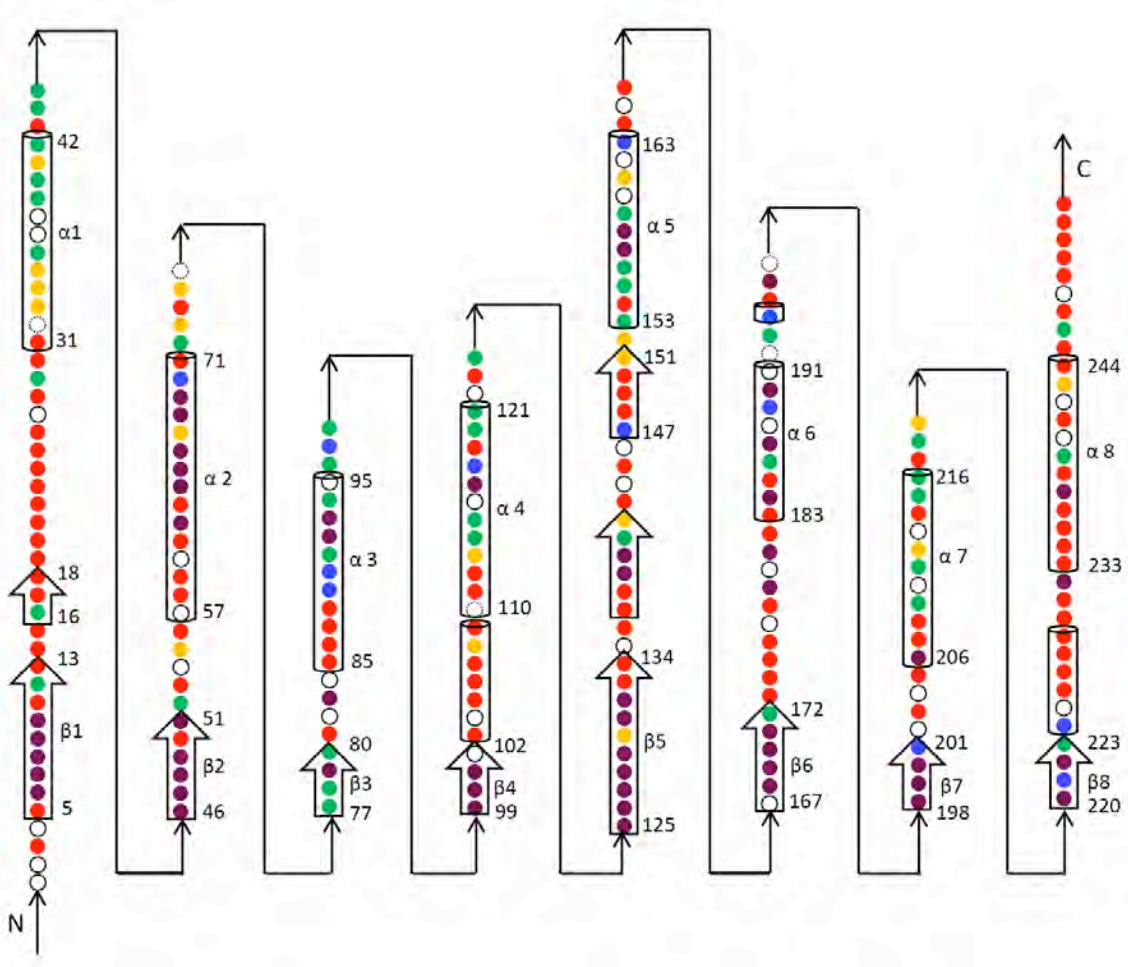

N
C
5
13
16
18
31
42
α1
β1
46
51
57
71
α 2
β2
77
80
85
95
α 3
β3
99
102
110
121
α 4
β4
125
134
147
151
153
163
α 5
β5
167
172
183
191
α 6
β6
198
201
206
216
α 7
β7
220
223
233
244
α 8
β8

Figure B.6

short α-helices (β4 and β8) or a β-hairpin (β5). Class I NHs also tend to be found on the hydrophilic faces of the α-helices and at the C-termini of β1, β2 and β5. Class II NHs are generally found on the nonpolar faces of the α-helices, at the C-termini of β1, β6, β7 and β8 and at interior positions in β3 and β8. They also can be found in the short C-terminal loops following the α-helices, which have previously been implicated in stabilizing the TIM barrel.[27,28] Non-exchanging Class III NHs are predominantly found in the β-strands, with one or more members of this class in all 8 β-strands. With the possible exception of α1 (3 NHs could not be resolved), at least one Class III NH is also found in the other 7 α-helices. Thus, almost the entire β-barrel and segments of most if not all of the α-helices in the external shell appear to be intact in higher energy states within the native manifold of conformers of HisF.

Structural dynamics of HX in the Class II NHs

Insights into the conformational changes accompanying the HX reactions at individual sites can be gained by performing HX in increasing concentrations of denaturant.[51] The denaturant dependence of the protection factors is correlated with the exposure of buried surface area in the opening reaction[45] and, thereby, is indicative of the magnitude of the conformational change. The Gdn-HCl dependence of $\Delta G°_{HX}$ for a representative set of Class II NHs is shown in **Figure B.7**. Within the uncertainty of the measurement, +/- 0.16 kcal mol^{-1}, the $\Delta G°_{HX}$ values are independent of the denaturant concentration over the range from 0 to 2.0 M Gdn-HCl. Exchange within the native manifold occurs through local fluctuations, not by a significant exposure of buried surface area in partially-unfolded states on the native side of the rate-limiting

Figure B.7. The dependence of ΔG°HX as a function of Gdn-HCl concentration for a representative set of amide hydrogens on HisF.

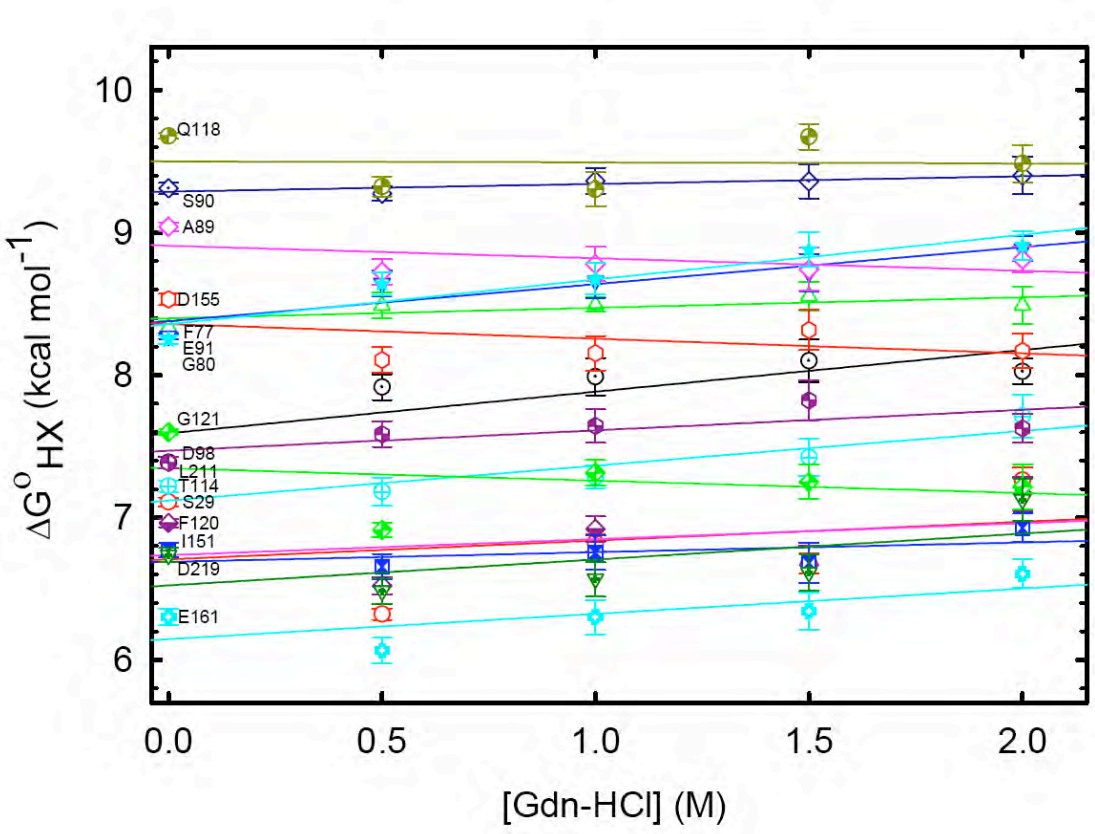

10
9
8
7
6
0.0
0.5
1.0
1.5
2.0
[Gdn-HCl] (M)
ΔG°HX (kcal mol-1)
Q118
S90
A89
D155
F77
E91
G80
G121
D98
L211
T114
S29
F120
I151
D219
E161

Figure B.7

barrier.[34,52-54] Estimating a denaturant dependence of ΔG°_{HX} of less than 0.2 kcal mol^{-1} M^{-1}, the surface area exposed in these exchange reactions must be < 200 $Å^2$.[55] Could small scale fluctuations of the sort experienced in normal vibrational modes be responsible for HX in the native conformation of HisF? This possibility was explored for HisF by calculating the mean squared fluctuations for a simplified C_α model of HisF with an on-line Gaussian Network Model (GNM) algorithm (http://ignm.ccbb.pitt.edu/). The GNM assumes that a protein in its folded state is equivalent to a fully elastic network, along the lines of the elasticity theory of random polymer networks.[56] The predicted equilibrium fluctuations of C_α atoms for several proteins are in close agreement with B-factors from x-ray crystallographic measurements,[57,58] supporting the validity of this approach. Pertinent to the present study, Bahar and her colleagues have previously observed an inverse correlation between B-factors predicted by the GNM and protection against exchange in 29 proteins.[59] Comparison with the ΔG°_{HX} values shows a strong correlation with the mean square fluctuations (**Figure B.8**). Segments with the smallest fluctuations, e.g., β-strands, tend to have high protection factors, the segments with the highest fluctuations, e.g., loops, tend to have low HX protection factors. The a-helices tend to display intermediate fluctuations, perhaps reflecting a mixture of all three classes of protection. The correlations imply that HX in the native conformation occurs through dynamic channels whose size and lifetimes are dictated by its global vibrational properties.[60]

Figure B.8. (a) Mean square fluctuations vs sequence, calculated for HisF from the GNM analysis. (b) ΔG°_{HX} values vs sequence for HisF at 40 ℃ and pH 7.2. The color code represents the β-strands (purple), α-helices (green), and loops (red).

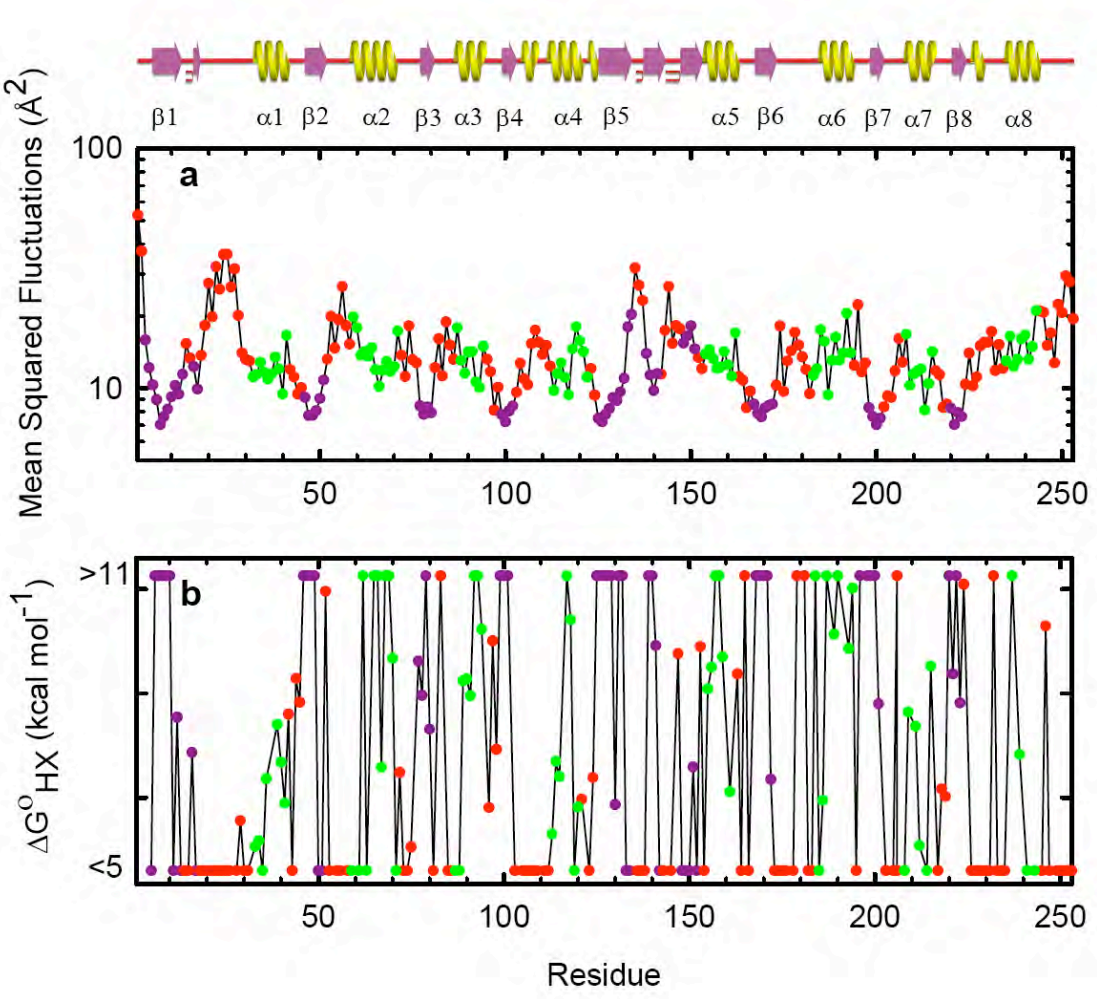

β1 α1 β2 α2 β3 α3 β4 α4 β5 α5 β6 α6 β7 α7 β8 α8
Mean Squared Fluctuations (Å2)
100
10
a
50
100
150
200
250
ΔGoHX (kcal mol-1)
>11
<5
b
50
100
150
200
250
Residue

Figure B.8

DISCUSSION

Native-state HX properties of HisF

The lifetime of the native state of HisF at pH 7.2 and 40 ℃, ~116 years in the absence of denaturant, far exceeds the time required to examine the HX properties of this unusually stable TIM barrel protein.

The protection of main chain NHs against exchange falls into three classes:

1. Class I NHs exchange with solvent within the time required to initiate the exchange reaction by dilution of lyophilized protein into deuterated buffer and collect a usable 2D NMR spectrum, ~30 minutes. In general, the Class I NHs are found in loops connecting the C-termini of β-strands with their subsequent α-helices, the solvent-exposed surface of the α-helical shell and the ends of α-helices and β-strands **(Figures B.5(a) & B.6**). Because the Class I NHs exchange far more rapidly than the rate of unfolding of HisF, they are most likely exchanging by an EX2 mechanism and have $\Delta G°_{HX}$ values less than 5 kcal mol^{-1}.
2. Class II NHs exchange over the time period from ~30 minutes to 10 days at pH 7.2 and 40 ℃ via an EX2 mechanism. The Class II NHs are primarily found at the interior, buried positions in the α-helices, in the short turns linking α-helices to the subsequent β-strands and preferentially in β3 and β8 (**Figures B.5(b) & B.6**). The $\Delta G°_{HX}$ values range from 5 to 11 kcal mol^{-1} for the Class II NHs (**Figure B.5(b) & Table B.S2**).
3. Class III NHs do not detectably exchange over 10 days at pH 7.2 and 40 ℃. Class III NHs most often appear in the 8 β-strands, but are also found in at least 7 of the 8 canonical helices (**Figures B.5(c) & B.6**). If they exchange by an EX2

mechanism, the $\Delta G°_{HX}$ values for Class III NHs are greater than 11 kcal mol^{-1}. If a subset of the Class III NHs exchange through intermediates on the unfolded side of the barrier, they must do so through species whose free energy are at least 12.44 kcal mol^{-1} above the native state.

Insights into the molecular underpinnings of the HisF TIM barrel

A closer examination of the protection patterns in the β-strands, α-helices, loops and turns in HisF (**Figure B.6**) hints at the essential structural features that define its native thermodynamic state. Most notably, the 8 canonical β-strands contain 58% of the non-exchanging NHs and 64% of the β-strand NHs fall into the Class III category (**Figure B.S4**). The interior of the HisF barrel is well described by 4 layers of side chains, as originally recognized by Thornton.[61] Each layer consists of 4 side chains from either the odd or even numbered strands, alternating as the layers proceed from the N- to the C-termini of the β-strands (**Figure B.S4**). Only the third layer contains 4 Class III NHs from the even numbered β-strands, V48 in β2, S101 in β4, L169 in β6 and L222 in β8, and 4 Class III NHs from the external side chains of the odd numbered β-strands, A8 in β1, V79 in β3, V127 in β5 and A200 in β7, that do not exchange over the period of 10 days. Examination of the structure further shows that Class III NHs in α-helices, L65 and V66 in α2, L92 and I93 in α3, A117 in α4, V157 and V158 in α5, I187 in α6 and L237 in α8, are closely associated with the side chains in layer 3 of the β-barrel (**Figure B.S5**). This collection of side chains linking the β-strands and α-helices may well represent one of the essential features defining the structures encompassed by the native thermodynamic state of the HisF barrel.

Figure B.S4. 2D representation of the hydrogen-bonding pattern of the β-sheet in HisF. The arrows (→) depict the hydrogen bonds, pointing from the backbone amide (NH) to the main chain carbonyl oxygen (C=O) with the exception of I6 in β1, F77 in β3 and I198 in β7 where their respective main chain amides are hydrogen-bonded to the side chain carboxyl of D45, D98 and D219, respectively. The (◌) with a cross indicate proline residues, and open circles indicates either the absence of the NMR assignment or the inability to obtain accurate fits due to spectral overlap. The (Δ) represents amino acids which are not a part of β-strands. Small circles represent main chain positions whose side chains point outside the barrel, large circles represent main chain positions whose side chains pointing inside the barrel. The ( ) with dashes indicates amide hydrogen not involved in hydrogen bonding in crystal structure. The NHs are color-coded as Class I, red, ΔG°HX $<$ 5 kcal mol -1; Class II: yellow, ΔG°HX 5 - 7 Kcal mol -1; green, ΔG°HX 7 - 9 kcal mol -1; blue, ΔG°HX 9 - 11 kcal mol -1; Class III, purple, ΔG°HX $>$ 11 kcal mol -1. The layers of side chains within the β- strands are indicated by dotted lines.

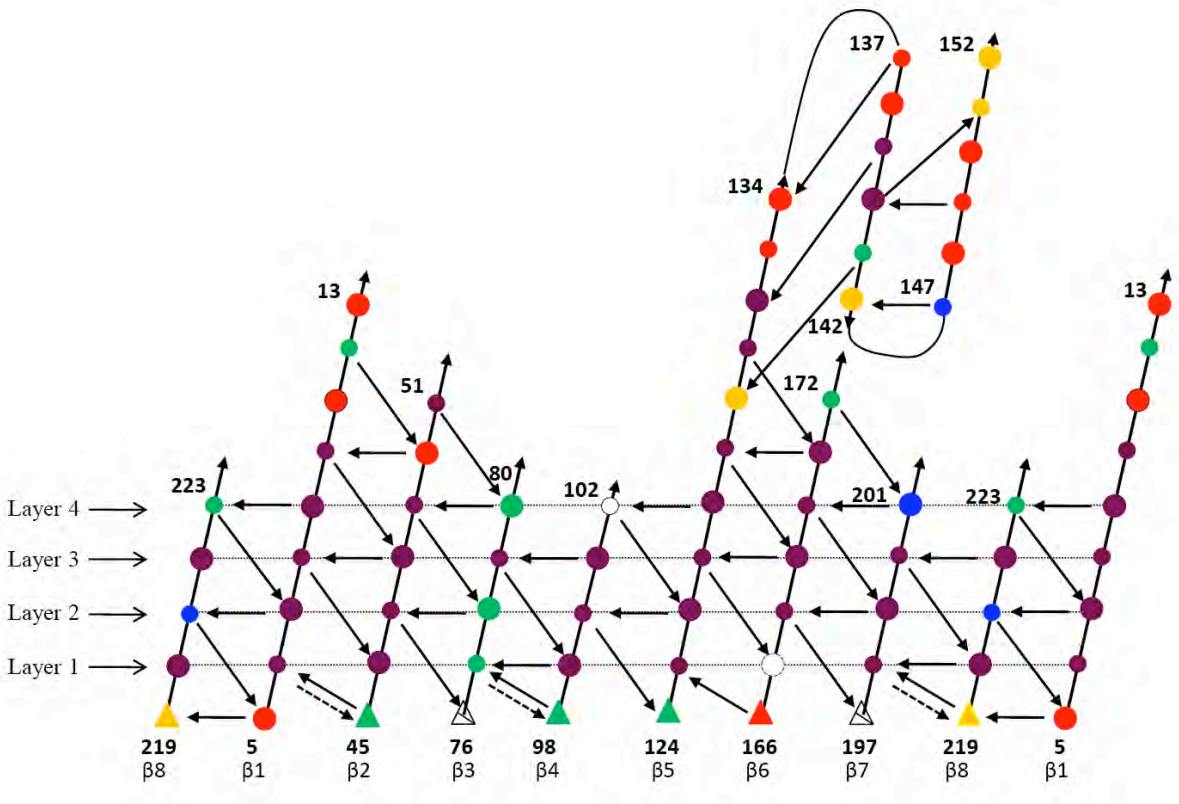

137
152
134
13
147
13
142
51
172
223
80
102
201
223
Layer 4
Layer 3
Layer 2
Layer 1
219
β8
5
β1
45
β2
76
β3
98
β4
124
β5
166
β6
197
β7
219
β8
5
β1

Figure B.S4

Figure B.S5. (a) & (b) A three-dimensional representation of the location of the protected 3rd layer of β-strands along with protected α-helices in different dimension. Class III NHs in helices are shown as green spheres and Class III NHs in β-strands are shown as purple spheres.

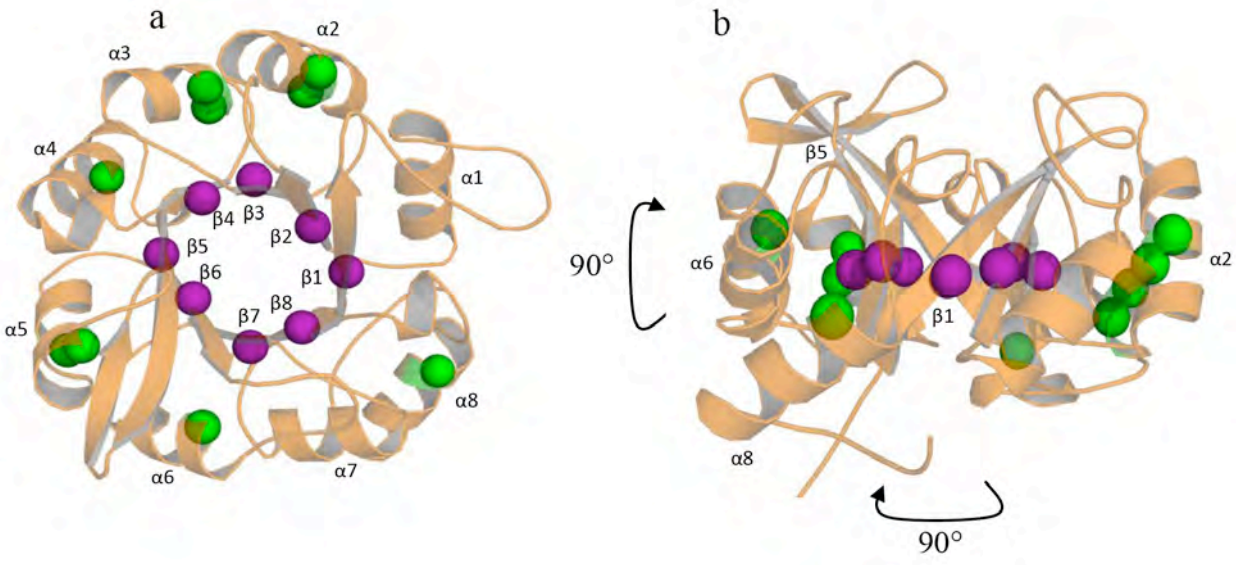


Figure B.S5.

The free energy difference between the native state and I_2 intermediate (Scheme 1) can be estimated from the chevron plot (**Figure B.2(a)**) and yields a value of ~12.44 kcal mol^{-1}, implying that the intervening transition state ensemble (TSE) must exceed that value. Access to the protection factors between 11 and at least 12.44 kcal mol^{-1}, i.e., those most crucial to the integrity of the native thermodynamic state was, unfortunately, precluded by the aggregation of the protein after 10 days at pH 7.2, at 40 ℃ and 2.0 M Gdn-HCl.

The BASiC hypothesis and HisF

The BASiC Hypothesis proposes that large clusters of ILV side chains play crucial roles in stabilizing partially-folded states in TIM barrel proteins.[36] Application of the BASiC Hypothesis to the αTS,[28] and sIGPS,[39,62] TIM barrels found a very good correlation between HX protection in their partially folded states and ILV clusters observed in their native conformations. The enhanced protection is ascribed to the preferentially resistance of such clusters to the penetration of the water and/or hydroxide ion required for exchange of underlying NHs and is consistent with the hydrophobicity scales of Wolfenden,[63] and Kyte & Doolittle.[64] Comparison of the protection patterns (**Figures B.5(b & c)**) with the locations of the 5 ILV clusters in HisF (**Figure B.5(d)**) shows that 90% of the Class II and Class III NHs lie within a surface defined by a 6$Å^2$ shell around these clusters. If one focuses on the very strongly protected, Class III, NHs, one finds that elements of one large ILV cluster stabilize helix-strand interactions in the β4/α4/β5/α5/β6/α6/β7 region and another stabilizes helix-strand interactions in the β1/α1/β2 region (**Figure B.5(d)**). Thus, the BASiC Hypothesis accounts for the cores of stability in the native conformation of the HisF TIM barrel

protein as well. Although almost all buried side chains are expected to contribute to the stability of the native state,[36] the dynamic behavior assessed by the HX experiment demonstrates a more heterogeneous situation for high energy microstates in the native basin.

Recently, the Rosetta design software package was employed to predict the most stable two-fold symmetric HisF TIM barrel,[65] motivated by the 4 + 4 gene duplication studies of Sterner and colleagues,[37] and an earlier conjecture of two-fold symmetric of many TIM barrels.[66] Exploring the full sequence of HisF, the cut points for the $(\beta\alpha)_4$ modules were moved along the sequence, the modules duplicated, fused and docked and the energies were minimized. When the cut points over-lapped the natural N- and C-termini, the termini were connected with a short linker peptide. The most stable symmetric barrels invariably contained a covalently-connected β4/α4/β5/α5/β6/α6/β7/α7 segment that agrees very well with the strongly-protected region spanned by the large ILV cluster (**Figures B.5(d) and B.S4**).

Implications for folding

The lack of dependence of the $\Delta G°_{HX}$ values on Gdn-HCl concentrations up to the unfolding transition zone (**Figure B.7**) shows that exchange occurs via local fluctuations in structure. The possible increase in $\Delta G°_{HX}$ with the concentration of Gdn-HCl for S90, E91 and D98 (**Figure B.7**), could reflect the binding of Gdn-HCl to these spatially-proximal amino acids. Dunbar and Farber,[67] have previously observed the binding of Gdn-HCl to several sites on the surface of ribonuclease A. The behavior of HisF contrasts with that of other proteins whose unfolding reactions near the transition zone are sufficiently rapid that they can traverse the rate-limiting barrier on the time

scale of the NS-HX experiment, typically a few hours. In those cases,[50,68-69] HX accelerates and the $\Delta G°_{HX}$ values decrease as high-energy, partially-unfolded species on both sides of the rate-limiting TSE contribute to the exchange process. By contrast, the HX results for HisF reflect the dynamic properties of native-like conformers that fluctuate around the lowest energy structure and are not simply related to those larger conformational changes that depend upon the denaturant concentration. Recognizing this limitation, however, the Class II protection against HX observed for half or more of the NHs in β3 and β8 (**Figures B.6 & B.S4**), $\Delta G°_{HX}$ ~ 9 kcal mol^{-1}, implies that these two weakened β-strands are the points at which the β-barrel initially fractures in the process of unfolding.

The 2 main chain Class II NHs in β3 form main chain-main chain H-bonds to β2 and the 2 Class II NHs in β8 form similar H-bonds to β1. The third Class II NH in β3, at F77, forms an H-bond to the side chain of D98 prior to β4 (**Figure B.S4**). Thus, one might predict that the two domains bordered by these two weakened β-stands interactions, β1/α1/β2/α2 and β4/α4/β5/α5/β6/α6/β7/α7, would comprise stable subdomains in one or both of the partially-folded states, I_1 and I_2, that appear during the folding of HisF (Scheme I). As described above, these subdomains are closely associated with clusters of ILV residues that are particularly resistant to exchange. Similar behavior has been previously observed in folding intermediates for sIGPS and αTS.[28,62] Direct assessment of the HX protection patterns in the HisF folding intermediates that appear during the folding reaction by pulse-quench HX-MS analysis is required to test this conjecture for HisF. These experiments are in progress.

Mechanism of HX

The absence of a denaturant dependence for $\Delta G°_{HX}$ for the measurable, Class II, NHs (**Figure B.7**) shows that the fluctuations enabling exchange do not involve large-scale changes in buried surface area. Thus, the mechanism of HX for HisF must involve the molecular events that are small in scale and sufficiently rapid to enable EX2 behavior. Lumry and Rosenberg,[70] proposed the “mobile defect” hypothesis that rests on packing imperfections in globular proteins.[71] These defects are not static, but rather move rapidly through the interior of a protein in response to thermal fluctuations. Consistent with this conjecture, Lakowicz and Weber,[72] observed the rapid diffusion of oxygen into carboxypeptidase A, presumably taking advantage of these mobile defects in the process of quenching the triplet state of a buried tryptophan. One might suppose that more tightly packed regions would be less receptive to defects because a higher free energy cost would be required to disrupt the more efficient packing. More loosely packed regions would be less energetically expensive to open and more easily able to accommodate a defect. If HX occurs via these mobile defects, resistance to exchange might reflect packing densities.

Another view, espoused by Wagner & Wuthrich,[73] supposes that exchange is governed by global processes that transiently expose main chain NHs to solvent. In a proposed model, they suggest that exchange is enhanced at polar interfaces between buried clusters of nonpolar side chains. If one instead considers the global process to be the normal vibrational modes of a protein, one finds a connection with the mobile defects view. A more tightly packed region would be expected to be involved in higher frequency, smaller displacement modes that would restrict the temporal or spatial access of water and hydroxide to the underlying main chain NH and vice versa. Either

the mobile defects or the global process view could be consistent with exchange via an EX2 mechanism, a wide range of protection factors and insensitivity to the denaturant concentration. The good correlation between the $\Delta G°_{HX}$ values of amides and the predicted mean-squared fluctuations of the respective C_α atoms by the GNM model (**Figures B.8(a) & (b**)), however, supports the global process mechanism of exchange via the normal vibrational modes of HisF. The correlation between protection against exchange and the location of clusters of ILV side chains probably reflects their propensity to be buried in the interior of the protein,[74,75] where the higher density of C_α atoms within 8 $Å^2$ would lead to smaller mean squared displacements and retarded HX. The unique aversion of aliphatic side chains to water[76] may also contribute to the enhanced protection seen in and near ILV clusters.

Insights into the evolution of the HisF barrel

The location of the stability cores in the native state of HisF does not mirror the 4 + 4 gene duplication event thought to have given rise to this TIM barrel protein.[2,22] Sequence variation following the duplication event has caused one core to span three over-lapping β/α/β modules, β4/α4/β5/α5/β6/α6/β7, and another core to stabilize the β1/α1/β2 module.

It is interesting, however, that the locations of a different stabilizing feature of TIM barrel proteins, the main chain-side chain H-bond clamp, are consistent with the 4 + 4 model. These clamps are normally formed between the NH of the amino acid in the second position from the N-terminus in an odd-numbered β-strand and, most often, the carboxylic moiety of an aspartic acid immediately preceding the subsequent even-numbered β-strand.[77] The side chains associated with the NHs in the odd-numbered β-

strand are large hydrophobes, often branched aliphatic side chains, that sequester the H-bond from solvent and significantly decrease the propensity of the NH to exchange with solvent. The Class II F77- D98 H-bond clamp is found between β3/α3/ β4, and the Class III I198-D219 clamp is found in the analogous position between β7/α7/ β8 (**Figure B.S4**). Although the clamp is conserved, the large hydrophobe is not: F77 in β3 shields the β3/α3/ β4 clamp and I198 shields the β7/α7/ β8 clamp. However, both clamps preserve a proline just prior to the first position of β3 and β7, and the sequence preceding both D98 and D219, GAD, is conserved.[77] The tripeptide is consistent with previous observations of a conserved GXD sequence prior to even-numbered b-strands in TIM barrel proteins.[2]

Interestingly, a third clamp between I6 in β1 and D45 prior to β2 is also found in HisF but its analog between β5 and β6 is not (**Figure B.S4**). The GID sequence prior to β2 is replaced with GAG prior to β6, eliminating the carboxylic acid H-bond acceptor required for the clamp with β5. The sequence of the β1/α1/β2 module has been predicted to most closely resemble that for the earliest progenitor of the fundamental β/α/β module,[37] suggesting that this clamp might have been a key stabilizing element in that module. The loss of the clamp as these modules diverged in sequence presumably reflects the development of other stabilizing components in the sequence, e.g., the large ILV cluster that spans the β4/α4/ β5/α5/β6/α6/ β7 region and, in particular, the β5/α5/β6 module.

Summary

The cores of stability in the native thermodynamic state for HisF, as assessed by protection of main chain amide hydrogens against exchange with solvent, correlate well

with the location of two clusters of isoleucine, leucine and valine side chains. The higher energy states exchange via local fluctuations that expose less than a few hundred square angstroms of buried surface area to solvent,[55] and can be understood as the transient voids created by the normal vibrational modes of HisF. When considered with the exchange properties of two other TIM proteins, αTS[28] and sIGPS,[39] whose ILV clusters offer protection against HX to folding intermediates, it is evident that ILV clusters serve as cores of stability all along the folding reaction coordinate. These clusters are not conserved in sequence or structure. Rather, the clusters move as the sequences evolve. What is conserved, however, is the existence of one or more ILV clusters that spans multiple adjacent β/α/β modules and plays a primary role in stabilizing folding intermediates and the native state of TIM barrel proteins.

MATERIALS AND METHODS

Reagents

Ultrapure Gdn-HCl was purchased from MP Biomedicals, LLC (Solon, OH), deuterium oxide (99.9%), ^{15}N ammonium chloride and, potassium EDTA were obtained from Sigma-Aldrich (St Louis, MO). Monobasic and dibasic potassium phosphates were purchased from J. T. Baker Inc. (Phillipsburg, NJ) and Fisher Scientific (Fair Lawn, NJ), respectively. All other chemicals were reagent grade.

Protein expression and purification

HisF was expressed from the pET11c plasmid containing the HisF gene (a gift from Dr. Reinhard Sterner) in *E. coli* BL21 (DE 3) cells and was purified according to a published protocol.[78] An overnight culture of freshly transformed cells in LB medium supplemented with 0.1 mg/ml ampicillin was used to inoculate a 10 lt culture stocks. The cells were harvested, lysed by sonication, centrifuged and the supernatant was heat shocked at 75 ℃ for 15 minutes to precipitate host proteins. The HisF in the supernatant was dialysed against 10 mM Kpi, containing 2 mM K_2EDTA, and 1 mM DTT buffer and purified by column chromatography with a DEAE-Sepharose fast-flow column followed by a Sephacryl S-200 column. The purity of the HisF was confirmed by SDS-PAGE and by measuring the molecular mass with liquid chromatography-electrospray ionization mass spectrometry.

For NMR studies, the transformed *E. coli* BL21 (DE 3) cells were grown on minimal media supplemented with ^{15}N ammonium chloride.[43] Cells were grown at 37 ℃ until the OD at 600 nm reached 0.9 at which point the temperature was reduced to 30 ℃, and the cells were allowed to grow for an additional 14 h. The purification followed

the above procedures, and the purity of the HisF was confirmed by SDS-PAGE and by mass spectrometry.

Equilibrium measurements

Circular Dichroism. All CD spectroscopy was performed on a JASCO-810 spectropolarimeter (Jasco Inc., Easton, MD) equipped with a water-cooled Peltier temperature control system. The CD spectra were obtained using a 5 mm path length quartz cuvette, a scan rate of 50 nm min^{-1}, and a response time of 2 s. The buffer contained 10 mM Kpi, pH 7.2, 1 mM K_2EDTA and 0.5 mM DTT. The Gdn-HCl induced unfolding was monitored from 215 to 250 nm, and the protein concentration was 5 µM. Samples for the unfolding titrations were prepared by adding appropriate volumes of 0-8 M Gdn-HCl in standard buffer to a stock of native HisF in buffer. Samples for the refolding titrations were prepared by adding appropriate volumes of buffer to unfolded HisF in 7.5 M Gdn-HCl. The samples were incubated at 40 ℃ for up to 18 days, and aliquots were periodically withdrawn and their CD spectra recorded to measure the progress of the unfolding and refolding reactions to equilibrium. The Gdn-HCl concentration was determined by refractive index on a Leica Mark II refractometer.[79] The data were fit to a two-state model as described previously.[42]

Kinetic measurements

Circular dichroism. The slow unfolding and refolding kinetics of HisF were initiated by manual mixing and monitored with a JASCO-810 spectropolarimeter (Jasco Inc., Easton, MD). Data were collected at 222 nm in 10 mm cuvette under continuous stirring with a solution volume of 2 ml (manual mixing dead time ~6 s). The fast refolding and unfolding kinetics were monitored at 222 nm using an AVIV-202 stopped-flow CD

spectrophotometer (dead time ~5 ms). Unfolding experiments were initiated from the native state in standard folding buffer, and refolding experiments were initiated from protein incubated in 7.5 M Gdn-HCl overnight. The kinetic data were fit to one or more exponentials,[42,80] using an in-house nonlinear least-squares fitting program, Savuka as described.[9] The logarithm of the observed relaxation time (s) was plotted as a function of final denaturant concentration to produce a chevron plot.[41]

Native-state hydrogen-deuterium exchange

Purified HisF was dialyzed overnight into a buffer containing 10 mM Kpi, pH 7.2, 50 mM KCl and 1 mM K_2EDTA. The sample was concentrated for the HX-NMR study with an Amicon Ultra-15 centrifugal filter unit with a 10 kDa membrane and then lyophilized. Exchange was initiated by dissolving the lyophilized protein in ^2H_2O buffer, which had been prepared at the required pH and buffer conditions (all the components in the buffers were pre-deuterated by repeated cycles of dissolution and lyophilization). All pH values reported are corrected meter readings. Upon addition of ^2H_2O buffer, the sample was immediately transferred to a 5 mm NMR tube (Wilmad LabGlass, Vineland, NJ) and placed in the spectrometer at 40 ℃; the HisF concentration was ~0.4 mM. The time between the initiation of exchange, the transfer to the NMR tube, placement in the spectrometer, tuning and shimming, and the beginning of data collection averaged 15 min. TROSY 2D ^{15}N-^{1}H correlation spectra were recorded over a period of hours to days, and the sample remained in the spectrometer for the entire course of the exchange reactions. All NMR experiments were recorded on a Varian 600-MHz spectrometer, and the spectra were processed in NMR Pipe,[81] and analyzed with Sparky.[82] The temperature was calibrated using a sample of 100% methanol. A list of

the ^{1}H and ^{15}N chemical shifts was obtained from the BioMagResBank (http://www.bmrb.wisc.edu/) under accession number BMRB-15741, and the TROSY spectrum assigned accordingly.

Analysis of H/D exchange data

Exchange rate constants for main chain NHs were obtained by fitting the decay of ^{15}N-^{1}H TROSY spectrum cross-peak intensities as a function of exchange time to a single exponential, $I = I_0 \exp(-k_{obs}t)$, with the initial intensity I_0 and the observed exchange rate k_{obs} as free variables in the fit. The exchange time was defined as the period of time from the dissolution of lyophilized HisF in $^{2}H_2O$ buffer to the end of each TROSY 2D experiment. The uncertainties for the k_{obs} values were taken as standard errors of the fits. Under the demonstrated EX2 exchange mechanism for the Class II NHs, $\Delta G°_{HX}$ was obtained from the equation: $\Delta G°_{HX} = -RT \ln(k_{obs}/k_{int})$, where k_{int} is the intrinsic exchange rate calculated for amide protons in unstructured peptides. The k_{int} values were obtained using the program SPHERE from model data.[83]

Analysis of ILV clusters from the crystal structure of HisF

Clusters of ILV side chains in the crystal structure of HisF, 1THF,[22] were identified using in-house software.[84] This software identifies networks of ILV side chains each of which bury at least 10 $Å^2$ and together bury more than 500 $Å^2$ surface areas using the CSU software.[85]

ACKNOWLEDGMENTS

We graciously thank Dr. Reinhard Sterner for generously providing the pET11c expression plasmid. Thanks also to Osman Bilsel, Jill Zitzewitz, Paul Nobrega, Ramakrishna Vadrevu and Can Kayatekin for helpful discussions. This work was supported by National Institutes of Health grant GM 23303 to C.R.M.

REFERENCES

1. Branden, C. I. (1991). The TIM barrel-the most frequently occurring folding motif in proteins. *Curr. Opin. Struct. Biol.* **1,** 978-983.
2. Nagano, N., Orengo, C.A. & Thornton, J.M. (2002). One fold with many functions: the evolutionary relationships between TIM barrel families based on their sequences, structures and functions. *J. Mol. Biol.* **321,** 741-765.
3. Farber, G. K. & Petsko, G. A. (1990). The evolution of α/β barrel enzymes. *Trends Biochem. Sci.* **15,** 228-234.
4. Beaucamp, N., Hofmann, A., Kellerer, B. & Jaenicke, R. (1997). Dissection of the gene of the bifunctional PGK-TIM fusion protein from the hyperthermophilic bacterium *Thermotoga maritima*: design and characterization of the separate triosephosphate isomerase. *Protein Sci.* **6,** 2159-2165.
5. Nickbarg, E. B. & Knowles, J. R. (1988). Triosephosphate isomerase: energetics of the reaction catalyzed by the yeast enzyme expressed in *Escherichia coli*. *Biochemistry,* **27,** 5939-5947.
6. Wilmanns, M., Hyde, C. C., Davies, D. R., Kirschner, K. & Jansonius, J. N. (1991). Structural conservation in parallel β/α-barrel enzymes that catalyze three sequential reactions in the pathway of tryptophan biosynthesis. *Biochemistry,* **30,** 9161-9169.
7. Knowles, J. R. (1991). Enzyme catalysis; not different, just better. *Nature,* **350,** 121-124.
8. Go, M. K., Amyes, T. L. & Richard, J. P. (2009). Hydron transfer catalyzed by triosephosphate isomerase. Products of the direct and phosphite-activated isomerization of [1-^{13}C] glycolaldehyde in D_2O. *Biochemistry,* **48,** 5769-5778.
9. Bilsel, O., Zitzewitz, J. A., Bowers, K. E. & Matthews, C. R. (1999). Folding mechanism of the α-subunit of tryptophan synthase, an α/β barrel protein: global analysis highlights the interconversion of multiple native, intermediate, and unfolded forms through parallel channels. *Biochemistry*, **38**, 1018-1029.
10. Rojsajjakul, T., Wintrode, P., Vadrevu, R., Matthews, C. R. & Smith, D. L. (2004). Multi-state unfolding of the alpha subunit of tryptophan synthase, a TIM barrel protein: insights into the secondary structure of the stable equilibrium intermediates by hydrogen exchange mass spectrometry. *J. Mol. Biol.* **341,** 241-253.
11. Finke, J. M. & Onuchic, J. N. (2005). Equilibrium and kinetic folding pathways of a TIM barrel with a funneled energy landscape. *Biophys. J.* **89,** 488-505.
12. Wu, Y., Kondrashkina, E., Kayatekin, C., Matthews. C. R. & Bilsel, O. (2008). Microsecond acquisition of heterogeneous structure in the folding of a TIM barrel protein. *Proc. Natl. Acad. Sci. USA,* **105,** 13367-13372.
13. Tanaka, T., Kimura, H., Hayashi, M., Fujiyoshi. Y., Fukuhara, K. I. & Nakamura, H. (1994). Characteristics of a de novo designed protein. *Protein Sci.* **3,** 419-427.
14. Joseph-McCarthy, D., Petsko, G. A. & Karplus, M. (1995). Use of a minimum perturbation approach to predict TIM mutant structures. *Protein Eng.* **8,** 1103-1115.
15. Babbitt, P. C., Hasson, M. S., Wedekind, J. E., Palmer, D. R. J., Barrett, W.C., Reed, G. H. *et al.* (1996). The enolase superfamily: a general strategy for

enzyme-catalyzed abstraction of the α-protons of carboxylic acids. *Biochemistry,* **35,** 16489-16501.
16. Babbitt, P. C. & Gerlt, J. A. (2001). New functions from old scaffolds: how nature reengineers enzymes for new functions. *Adv. Protein Chem.* **55,** 1-28.
17. Silverman, J. A., Balakrishnan, R. & Harbury, P. B. (2001). Reverse engineering the $(\beta/\alpha)_8$ barrel fold. *Proc. Natl. Acad. Sci. USA,* 98, 3092-3097.
18. Bonneau, R., Strauss, C. E. M., Rohl, C. A., Chivian, D., Bradley, P., Malmstrom, L. *et al.* (2002). De novo prediction of three-dimensional structures for major protein families. *J. Mol. Biol.* **322,** 65-78.
19. Shukla, A. & Guptasarma, P. (2004). Folding of β/α-unit scrambled forms of *S. cerevisiae* triosephosphate isomerase: evidence for autonomy of substructure formation and plasticity of hydrophobic and hydrogen bonding interactions in core of $(\beta/\alpha)_8$-barrel. *Proteins,* **55,** 548-557.
20. Rothlisberger, D., Khersonsky, O., Wollacott, A. M., Jiang. L., DeChancie, J., Betker, J. *et al.* (2008). Kemp elimination catalysts by computational enzyme design. *Nature,* **453,** 190-195.
21. Jiang, L., Althoff, E. A., Clemente, F. R., Doyle, L., Röthlisberger, D., Zanghellini, A. et al. (2008). De novo computational design of retro-aldol enzymes. *Science,* **319,** 1387-1391.
22. Lang, D., Thoma, R., Henn-Sax, M., Sterner, R. & Wilmanns, M. (2000). Structural evidence for evolution of the β/α barrel scaffold by gene duplication and fusion. *Science,* **289,** 1546-1550.
23. Gerlt, J. A. & Babbitt, P. C. (2001). Barrels in pieces? *Nat. Struct. Biol.* **8,** 5-7.
24. Höcker, B., Jürgens, C., Wilmanns, M. & Sterner, R. (2001). Stability, catalytic versatility and evolution of the $(\beta/\alpha)_8$-barrel fold. *Curr. Opin. Biotechnol.* **12,** 376-381.
25. Setiyaputra, S., Mackay, J. P. & Patrick, W. M. (2011). The structure of a truncated phosphoribosylanthranilate isomerase suggests a unified model for evolution of the $(\beta\alpha)_8$ barrel fold. *J. Mol. Biol.* **408,** 291-303.
26. Reardon, D. & Farber G. K. (1995). The structure and evolution of α/β barrel proteins. *FASEB J.* **9,** 497-503.
27. Niermann, T. & Kirschner, K. (1990). Improving the prediction of secondary structure of “TIM” barrel enzymes. *Protein Eng.* **4,** 137-147.
28. Vadrevu, R., Wu, Y. & Matthews, C. R. (2008). NMR Analysis of partially folded states and persistent structure in the alpha subunit of tryptophan synthase: implications for the equilibrium folding mechanism of a 29-kDa TIM barrel protein. *J. Mol. Biol.* **377,** 294-306.
29. Anantharaman, V., Arvind, L. & Koonin, E. V. (2003). Emergence of diverse biochemical activities in evolutionarily conserved structural scaffolds of proteins. *Curr. Opin. Chem. Biol,* **7,** 12-20.
30. Burroughs, A. M., Allen, K. N., Dunaway-Mariano, D. & Aravind, L. (2006). Evolutionary genomics of the HAD superfamily: understanding the structural adaptations and catalytic diversity in a superfamily of phosphoesterases and allied enzymes. *J. Mol. Biol.* **361,** 1003-1034.
31. Shortle, D. (1986). Guanidine hydrochloride denaturation studies of mutant forms of *Staphylococcal Nuclease*. *J. Cell. Biochem.* **30,** 281-289.

32. Shortle, D. & Meeker, A. K. (1986). Mutant forms of *Staphylococcal Nuclease* with altered patterns of guanidine hydrochloride and urea denaturation. *Proteins: Struct. Funct. Bioinformatics,* **1,** 81-89.
33. Lim, W. A., Farruggio, D. C. & Sauer, R. T. (1992). Structural and energetic consequences of disruptive mutations in a protein core. *Biochemistry,* **31,** 4324-4333.
34. Bai, Y., Sosnick, T. R., Mayne, L. & Englander, S. W. (1995). Protein folding intermediates: native-state hydrogen exchange. *Science,* **269,** 192-197.
35. Radzicka, A., Pedersen, L. & Wolfenden, R. (1988). Influences of solvent water on protein folding: free energies of salvation of Cis and Trans peptides are nearly identical. *Biochemistry,* **27,** 4538-4541.
36. Wu, Y., Vadrevu, R., Kathuria, S., Yang, X. & Matthews, C. R. (2007). A tightly packed hydrophobic cluster directs the formation of an off-pathway sub-millisecond folding intermediate in the α subunit of tryptophan synthase, a TIM barrel protein. *J. Mol. Biol.* **366,** 1624-1638.
37. Richter, M., Bosnali, M., Carstensen, L., Seitz, T., Durchschlag, H., Blanquart, S. *et al.* (2010). Computational and experimental evidence for the evolution of a $(\beta\alpha)_8$-barrel protein from an ancestral quarter-barrel stabilised by disulfide bonds. *J. Mol. Biol.* **398,** 763-773.
38. Höcker, B., Beismann-Driemeyer, S., Hettwer, S., Lustig, A. & Sterner, R. (2001). Dissection of a $(\beta\alpha)_8$-barrel enzyme into two folded halves. *Nature Struct. Biol.* **8,** 32-36.
39. Gu, Z., Zitzewitz, J. A. & Matthews, C. R. (2007). Mapping the structure of folding cores in TIM barrel proteins by hydrogen exchange mass spectrometry: the roles of motif and sequence for the indole-3-glycerol phosphate synthase from *S. solfataricus*. *J. Mol. Biol.* **368,** 582-594.
40. Forsyth, F. R., & Matthews, C. R. (2002). Folding mechanism of Indole-3-glycerol phosphate synthase from *Sulfolobus solfataricus*: a test of the conservation of folding mechanisms hypothesis in $(\beta\alpha)_8$ barrels. *J. Mol. Biol.* **320,** 1119-1133.
41. Matthews, C. R. (1987). Effect of point mutations on the folding of globular proteins. *Methods Enzymol.* **154,** 498-511.
42. Wu, Y. & Matthews, C. R. (2002). Parallel channels and rate-limiting steps in complex protein folding reactions: prolyl isomerization and the alpha subunit of trp synthase, a TIM barrel protein. *J. Mol. Biol.* **323,** 309-325.
43. Lipchock, J. M. & Loria, J. P. (2008). ^{1}H, ^{15}N and ^{13}C resonance assignment of imidazole glycerol phosphate (IGP) synthase protein HisF from *Thermotoga maritima*. *Biomol. NMR. Assign.* **2,** 219-221.
44. Covington, A. K., Paabo, M., Robinson, R. A. & Bates, R. G. (1968). Use of the glass electrode in deuterium oxide and the relation between the standardized pD (pa_D) scale and the operational pH in heavy water. *Anal. Chem.* **40,** 700-706.
45. Englander, S.W. (2000). Protein folding intermediates and pathways studied by hydrogen exchange. *Annu. Rev. Biophys. Biomol. Struct.* **29,** 213-238.
46. Kiefhaber, T. & Baldwin, R. L. (1995). Kinetics of hydrogen bond breakage in the process of unfolding of ribonuclease A measured by pulsed hydrogen exchange. *Proc. Natl Acad. Sci. USA,* **92,** 2657-2661.

47. Yan, S., Kennedy, S. D. & Koide, S. (2002). Thermodynamic and kinetic exploration of the energy landscape of Borrelia burgdorferi OspA by native-state hydrogen exchange. *J. Mol. Biol.* **323,** 363-375.
48. Huang, J. R., Craggs, T. D., Christodoulou, J. & Jackson, S. E. (2007). Stable intermediate states and high energy barriers in the unfolding of GFP. *J. Mol. Biol.* **370,** 356-371.
49. Matthew, J. B. & Richards, F. M. (1983). The pH dependence of hydrogen exchange in protein. *J. Biol. Chem.* **258,** 3039-3044.
50. Bai, Y., Milne, J. S., Mayne, L. & Englander, S.W. (1994). Protein stability parameters measured by hydrogen exchange. *Proteins: Struct. Funct. Genet.* **20,** 4-14.
51. Xu, Y., Mayne, L. & Englander, S.W. (1998). Evidence for an unfolding and refolding pathway in cytochrome *c*. *Nature Struct. Biol.* **5,** 774-778.
52. Chamberlain, A. K., Fischer, K. F., Reardon, D., Handel, T. M. & Marqusee, S. (1999). Folding of an isolated ribonuclease H core fragment. *Protein Sci.* **8,** 2251-2257.
53. Feng, H., Vu, N. D. & Bai, Y. (2004). Detection and structure determination of an equilibrium unfolding intermediate of Rd-apocytochrome b_{562}: native fold with non-native hydrophobic interactions. *J. Mol. Biol.* **343,** 1477-1485.
54. Bernstein, R., Schmidt, K. L., Harbury, P. B. & Marqusee, S. (2011). Structural and kinetic mapping of side-chain exposure onto the protein energy landscape. *Proc. Natl. Acad. Sci. USA,* **108,** 10532-10537.
55. Myers, J. K., Pace, C. N. & Scholtz, J. M. (1995). Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci.* **4,** 2138-2148.
56. Flory, P. J. (1976). Statistical thermodynamics of random networks. *Proc. R. Soc. Lond. Ser. A.* 351-380.
57. Bahar, I., Atilgan, A. R. & Erman, B. (1997). Direct evaluation of thermal fluctuations in proteins using a single parameter harmonic potential. *Folding Des.* **2,** 173-181.
58. Bahar, I., Erman, B., Jernigan, R. L., Atilgan, A. R. & Covell, D. G. (1999). Collective motions in HIV-1 reverse transcriptase: examination of flexibility and enzyme function. *J. Mol. Biol.* **285,** 1023-1037.
59. Rader, A. J. & Bahar, I. (2004). Folding core predictions from network models of proteins. *Polymer,* **45,** 659-668.
60. Yang, L. W., Liu, X., Jursa, C. J., Holliman, M., Rader, A. J., Karimi, H. A. *et al.* (2005). *i*GNM: a database of protein functional motions based on Gaussian Network Model. *Bioinformatics,* **21,** 2978-2987.
61. Nagano, N., Hutchinson, E. G. & and Thornton, J. M. (1999). Barrel structures in proteins: automatic identification and classification including a sequence analysis of TIM barrels. *Protein Sci*. **8,** 2072-2084.
62. Gu, Z., Rao, M. K., Forsyth, W. R., Finke, J. M. & Matthews, C. R. (2007). Structural analysis of kinetic folding intermediates for a TIM barrel protein, indole-3-glycerol phosphate synthase, by hydrogen exchange mass spectrometry and Gō model simulation. *J. Mol. Biol.* **374,** 528-546.

63. Wolfenden, R., Andersson, L., Cullis, P. M. & Southgate, C. C. B. (1981). Affinities of amino acid side chains for solvent water. *Biochemistry,* **20,** 849-855.
64. Kyte, J. & Doolite, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157,** 105-132.
65. Fortenberry, C., Bowman, E. A., Proffitt, W., Dorr, B., Combs, S. A., Harp, J. M., Mizoue, L. & Meiler, J. (2011). Exploring symmetry as an avenue to the computational design of large protein domains. *J. Am. Chem. Soc*. 133, 18026-18029.
66. Wierenga, R. K. (2001). The TIM-barrel fold: a versatile framework for efficient enzymes. *FEBS Lett.* **492,** 193-198.
67. Dunbar, J., Yennawar, H, P., Banerjee, S., Luo, J. & Farber, G. K. (1997). The effect of denaturants on protein structure. *Protein Sci*. **6,** 1727-1733.
68. Parker, M. J. & Marqusee, S. (2001). A kinetic folding intermediate probed by native state hydrogen exchange. *J. Mol. Biol.* **305,** 593-602.
69. Cortajarena, A. L., Mochrie, S. G. J. & Regan, L. (2008). Mapping the energy landscape of repeat proteins using NMR-detected hydrogen exchange. *J. Mol. Biol.* **379,** 617-626.
70. Lumry, R. & Rosenberg, A . (1975). The mobile defect hypothesis of protein function. *Coll. Int., C. N. R. S. I'Eau. Syst. Biol.* **246,** 53.
71. Richards, F.M. (1977). Areas, volumes, packing and protein structure. *Ann. Rev. Biophys. Bioeng.* **6,** 151-176.
72. Lakowicz, J. R. & Weber, G. (1973). Quenching of protein fluorescence by oxygen. detection of structural fluctuations in proteins on the nanosecond time scale. *Biochemistry,* **12,** 4171-4179.
73. Wagner, G. & Wuthrich, K. (1979). Structural interpretation of the amide proton exchange in the basic pancreatic trypsin inhibitor and related proteins. *J. Mol. Biol.* **134,** 75-94.
74. Pace, C. N., Shirley, B. A., McNutt, M. & Gajiwala, K (1996). Forces contributing to the conformational stability of proteins. *FASEB J.* **10,** 75-83.
75. Pace, C. N., Fu, H., Fryar, K. L., Landua, J., Trevino, S. R., Shirley, B. A. *et al.* (2011). Contributions of hydrophobic interactions to protein stability. *J. Mol. Biol.* **408,** 514-528.
76. Radzicka, A. & Wolfenden, R. (1988). Comparing the polarities of the amino acids: side-chain distribution coefficients between the vapor phase, cyclohexane, 1-octanol, and neutral aqueous solution. *Biochemistry,* **27,** 1664-1670.
77. Yang, X., Kathuria, S. V., Vadrevu, R. & Matthews, C. R. (2009). βα-Hairpin clamps brace βαβ modules and can make substantive contributions to the stability of TIM barrel proteins. PLoS One, 4, e7179.
78. Thoma, R., Obmolova, G., Lang, D. A., Schwander, M., Jeno, P., Sterner, R. *et al.* (1999). Efficient expression, purification and crystallisation of two hyperthermostable enzymes of histidine biosynthesis. *FEBS Lett*, **454,** 1-6.
79. Nozaki, Y. (1972). The preparation of guanidine hydrochloride. *Methods Enzymol. **27,*** 43-50.
80. Wu, Y. & Matthews, C. R. (2003). Proline replacements and the simplification of the complex, parallel channel folding mechanism for the alpha subunit of trp synthase, a TIM barrel protein. *J. Mol. Biol.* **330,** 1131-1144.

81. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995). NMR Pipe-a multidimensional spectral processing system based on unix pipes. *J. Biomol. NMR,* **6**, 277-293.
82. Kneller, D. G. & Kuntz, I. D. (1993). UCSf Sparky: an NMR display, annotation and assignment tool. *J. Cell. Biochem*, **53**, 254.
83. Bai, Y., Milne, J. S., Mayne, L. & Englander, S. W. (1993). Primary structure effects on peptide group hydrogen exchange. *Proteins: Struct. Funct. Genet.* **17**, 75-86.
84. Kathuria, S. V., Day, I. J., Wallace, L. A. & Matthews, C. R. (2008). Kinetic traps in the folding of βα-repeat proteins: CheY initially misfolds before accessing the native conformation. *J. Mol. Biol.* **382,** 467-484.
85. Sobolev, V., Sorokine, A., Prilusky, J., Abola, E. E. & Edelman, M. (1999). Automated analysis of interatomic contacts in proteins. *Bioinformatics,* **15,** 327-332.