

University of Massachusetts Medical School

eScholarship@UMMS

GSBS Dissertations and Theses

Graduate School of Biomedical Sciences

2009-7


Long-Range Side Chain-Main Chain Hydrogen Bonds: A Molecular Signature of the TIM Barrel Architecture: A Dissertation

Xiaoyan Yang

University of Massachusetts Medical School

Let us know how access to this document benefits you.

Follow this and additional works at: https://escholarship.umassmed.edu/gsbs_diss

 Part of the [Amino Acids, Peptides, and Proteins Commons](#), [Enzymes and Coenzymes Commons](#), and the [Inorganic Chemicals Commons](#)

Repository Citation

Yang X. (2009). Long-Range Side Chain-Main Chain Hydrogen Bonds: A Molecular Signature of the TIM Barrel Architecture: A Dissertation. GSBS Dissertations and Theses. <https://doi.org/10.13028/d9qz-4k61>. Retrieved from https://escholarship.umassmed.edu/gsbs_diss/430

This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in GSBS Dissertations and Theses by an authorized administrator of eScholarship@UMMS. For more information, please contact Lisa.Palmer@umassmed.edu.

**LONG-RANGE SIDE CHAIN-MAIN CHAIN
HYDROGEN BONDS:
A MOLECULAR SIGNATURE OF THE
TIM BARREL ARCHITECTURE**

A Dissertation Presented

By

XIAOYAN YANG

**Submitted to the faculty of the University of Massachusetts Medical School in partial
fulfillment of the requirements for the degree of**

DOCTOR OF PHILOSOPHY

IN BIOCHEMISTRY AND MOLECULAR PHARMACOLOGY

July 2009

**LONG-RANGE SIDE CHAIN-MAIN CHAIN HYDROGEN BONDS:
A MOLECULAR SIGNATURE OF THE TIM BARREL ARCHITECTURE**

**A Dissertation Presented
By
XIAOYAN YANG**

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

C. Robert Matthews, Ph.D., Thesis Advisor

Dan Bolon, Ph.D., Member of Committee

David Lambright, Ph.D., Member of Committee

William Royer, Ph.D., Member of Committee

Doug Barrick, 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**

Mary Munson, 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
Program in Biochemistry and Molecular**

**Program in Biochemistry and Molecular Pharmacology
July, 2009**

DEDICATION

I dedicate this thesis in loving memory of my grandparents, Jikang Li and Daqin Feng. And to my great-grandmother, Yunzhu Cui, this is the gift for your 100th birthday!

ACKNOWLEDGEMENTS

I would like to thank all the people who have helped me during my graduate study. First of all, I would like to thank my thesis advisor, Bob Matthews, for his continuous support, encouragement and patience. His intelligence and dedication to science always inspired me. What I learned from him, not only how to be a good scientist, but more importantly how to be a good person, will benefit me for my whole life. I would also thank Bob for leading a wonderful group of people to make the lab such a pleasant place to work, to communicate, and to have fun. I want to thank all the former and current members from Matthews' lab. They are the best colleagues and friends. I would like to thank Dr. Ying Wu for teaching me all the techniques and how to work efficiently. I thank Dr. Ramakrishna Vadrevu for initiating my exciting project. I thank Sagar Kathuria for excellent collaboration and all insightful discussions. I thank Dr. Zhenyu Gu for technical support, great advice and friendship. I thank Dr. Jill Zitzewitz for all the help on so many ways, especially scientific thinking and writing. I thank Dr. Osman Bilsel for all the patience on technical help and explanations on data analysis. I thank Can Kayatekin, Anna-Karin Svensson, Amanda Fitzgerald Noel, and Agnita Kundu for sharing my happiness and tears. I thank Shaun Lavalley, Jennifer Duong, Meme Tran, Paul Nobrega, and Gangadhara for bringing so much fun and energy to the lab. I thank Karen Welch for all the scheduling and paperwork. I love your sweet smile and voice.

I appreciate my committee members, Dan Bolon, Bill Royer, David Lambright, and committee chair, Mary Munson for their thoughtful suggestions and guidance through all my thesis research. I thank Doug Barrick for serving as my external

committee. I also would like to thank everyone in the BMP department, especially Yufeng Cai, Madhavi Kolli, Melonnie Furgason, Nicholas Willis, Naveen Kommajosyula, and John Pagano, for all the friendship.

I want to thank all my friends here in UMMS, and from high school and college. I want to thank my boyfriend Fan Xiao, who has shared everything with me along this journey. Finally I would like to send special appreciation to my family, especially my father, Yihua Yang, and my mother, Jun Li, who are the best parents. Without their love and support, I could never have done this. I would thank my grandmother Daqin Feng, who just passed away after my defense and my grandfather Jikang Li, who left me four years ago. They gave me so much love and taught me so many important things in life. You are always in my heart.

ABSTRACT

The hydrophobic effect and hydrogen bonding interactions have long been considered to be the dominant forces in protein folding. However, the contribution of hydrogen bonds to stabilizing proteins has been difficult to clarify. As the intramolecular hydrogen bonds are formed in place of hydrogen bonds with solvent during folding, measures of stability fail to give a significant change in free energy. Previous studies on hydrogen bonding interactions have shown that they are only marginally important.

Three long-range side chain-main chain hydrogen bonds have been found in the alpha subunit of tryptophan synthase (α TS), a $(\beta\alpha)_8$ TIM barrel protein. These long-range non-covalent interactions connect either the N-terminus of one β -strand with the C-terminus of the succeeding and anti-parallel α -helix (F19-D46 and I97-D124) or the N-terminus of an α -helix with the C-terminus of the succeeding β -strand (A103-D130). By analogy, these interactions are designated as $\beta\alpha$ - or $\alpha\beta$ -hairpin clamps. Surprisingly, the removal of any one of these clamp interactions, by replacement of the aspartic acid with alanine, results in significantly decreased thermodynamic stability for the native state and a substantial loss of secondary structure. When compared to several other side chain-side chain and short-range side chain-main chain interactions in α TS, these hairpin clamps clearly play a unique role in the structure and stability of α TS.

The generality of these observations for $\beta\alpha$ -hairpin clamps in TIM barrel proteins was tested by experimental analysis of the clamps in a pair of homologous indole-3-glycerol phosphate synthase (IGPS) TIM barrels of low sequence identity. The results suggest that

only the subset of conserved $\beta\alpha$ -hairpin clamps with hydrogen bond length less than 2.80 Å make substantive contributions to stability and/or structure. Those clamps with longer hydrogen bonds make modest contributions to stability and structure, similar to other types of side chain-main chain or side chain-side chain hydrogen bonds. The role of these clamps in defining the structures of the super-family of TIM barrel proteins was examined by a survey of 71 TIM barrel proteins from the structural database. Conserved features of $\beta\alpha$ -hairpin clamps are consistent with a 4-fold symmetry, with a predominance of main chain amide hydrogen bond donors near the N-terminus of the odd-number β -strands and side chain hydrogen bond acceptors in the loops between the subsequent α -helices and even-numbered β -strands. In this configuration, the clamps provide an N-terminal cap to odd-number β -strands in the β -barrel.

Taken together, these findings suggest that $\beta\alpha$ -hairpin clamps are a vestigial signature of the fundamental $\beta\alpha\beta$ building block for the $(\beta\alpha)_8$ motif and an integral part of the basic TIM barrel architecture. The relative paucity of $\beta\alpha$ -hairpin clamps remaining in TIM barrel structures and their variable contributions to stability imply that other determinants for structure and stability of the barrel have evolved to render a subset of the clamp interactions redundant. Distinct sequence preferences for the partners in the $\beta\alpha$ -hairpin clamps and the neighboring segments may be useful in enhancing algorithms for structure prediction and for engineering stability in TIM barrel proteins.

TABLE OF CONTENTS

TITLE PAGE.....	i
APPROVAL PAGE.....	ii
DEDICATION.....	iii
ACKNOWLEDGEMENTS.....	iv
ABSTRACT	vi
TABLE OF CONTENTS	viii
LIST OF TABLES.....	x
LIST OF FIGURES	xi
ABBREVIATIONS	xiii
PREFACE.....	xiv
CHAPTER I INTRODUCTION.....	1
Overview.....	2
Protein stability.....	2
Protein H-bonds	3
Side chain-main chain H-bonds.....	6
TIM barrel proteins.....	9
Structure.....	9
Stability.....	12
Targets for study	14
Scope of the thesis	22
CHAPTER II Long-range Side chain-main chain Interactions Play Crucial Roles in Stabilizing the $(\beta\alpha)_8$ Barrel Motif of the Alpha Subunit of Tryptophan Synthase.....	25
ABSTRACT.....	26
INTRODUCTION	27
RESULTS	31
Alanine replacements at D46, D124 or D130 significantly perturb the secondary and tertiary structure of α TS	31
Equilibrium folding properties of the clamp-deletion variants.....	35
Kinetic folding properties of the clamp-deletion variants	37
Thermodynamic parameters extracted from the kinetic experiments.....	43
Uniqueness of the clamp interactions	46
Uniqueness of aspartic acid as an H-bond acceptor in the clamp interaction.....	47
DISCUSSION.....	53
The role of clamp interactions in the structure and stability of α TS	53
The role of the clamp interactions in the folding mechanism of α TS	54

Structural implication of strong protection against HX for F19 NH and I97 NH	57
Clamp interactions define the boundaries of $\beta\alpha$ and $\alpha\beta$ hairpins	57
CHAPTER III $\beta\alpha$ -hairpin Clamps Highlight the $\beta\alpha\beta$ Module as a Fundamental Building Block of TIM Barrel Proteins	59
ABSTRACT	60
INTRODUCTION	61
RESULTS	64
Experimental analysis of $\beta\alpha$ -hairpin clamp H-bonds in two TIM barrel proteins ...	64
Survey of $\beta\alpha$ -hairpin clamps in the TIM barrel proteins.....	75
DISCUSSION	82
Recognition of potent $\beta\alpha$ -hairpin clamps in TIM barrel proteins	82
Survey analysis of $\beta\alpha$ -hairpin clamps in TIM barrel proteins.....	83
CHAPTER IV DISCUSSION AND FUTURE DIRECTIONS	87
Implications for TIM barrel protein structure and folding	88
Broader involvement of clamp interactions.....	90
Applications to protein structure prediction and protein engineering	93
CHAPTER V MATERIALS AND METHODS	95
Site-directed mutagenesis	96
Protein expression and purification	97
Circular dichroism	99
Equilibrium experiments.....	99
Kinetic experiments	100
Data analysis	101
Survey of TIM barrel structures.....	101
Definitions of $\beta\alpha$ -hairpin clamp H-bonds	102
APPENDIX.....	103
Verification of the intermediate state for eIGPS $\beta7\alpha7$ clamp-deletion variant N231A by kinetic jump experiments.....	103
One nonpotent Class II clamp H-bond in sIGPS	105
One clamp variant in eIGPS may enhance protein stability	109
References.....	114

LIST OF TABLES

- Table 2.1 Thermodynamic parameters for the urea-induced unfolding of WT α TS and eight variants at 25 °C
- Table 3.1 Thermodynamic parameters for the urea-induced unfolding of sIGPS, eIGPS, α TS and nine clamp-deletion variants at 25 °C
- Table 3.2 Donor/acceptor sequence preferences of Class I $\beta\alpha$ -hairpin clamps in 71 TIM barrel proteins
- Table 4.1 Potential potent clamp H-bonds predicted from the database of 71 TIM barrel proteins
- Table A.1 Thermodynamic parameters for the urea-induced unfolding of sIGPS clamp-deletion variants at 25 °C
- Table A.2 Thermodynamic parameters for the urea-induced unfolding of WT and S82D eIGPS at 25 °C

LIST OF FIGURES

- Figure 1.1 Classification of long-range side chain-main chain H-bonds (clamp H-bonds)
- Figure 1.2 Representation of the TIM barrel protein motif
- Figure 1.3 Schematic model of the TIM barrel fold
- Figure 1.4 Equilibrium and kinetic folding models of α TS
- Figure 1.5 ^1H - ^{15}N HSQC spectrum of α TS
- Figure 1.6 Equilibrium and kinetic folding models of σ IGPS
- Figure 1.7 Proposed folding model of ϵ IGPS
- Figure 2.1 Long-range side chain-main chain interactions in α TS
- Figure 2.2 Far-UV and near-UV CD spectra of WT α TS and three α TS variants
- Figure 2.3 Urea denaturation equilibrium unfolding curves of D46A, D124A, and D130A α TS
- Figure 2.4 Chevron plots for WT, D46A, D124A, and D130A α TS
- Figure 2.5 The dependence of the amplitude for the major unfolding phase for WT and D46A α TS on the initial urea concentration
- Figure 2.6 Far-UV and near-UV CD spectra of three clamp-deletion α TS variants
- Figure 2.7 Urea denaturation equilibrium unfolding curves of D38A, D112A, S33A, D46A, D46N, and D46E α TS
- Figure 3.1 Long-range side chain-main chain interactions in σ IGPS
- Figure 3.2 Long-range side chain-main chain interactions in ϵ IGPS

- Figure 3.3 Far-UV and near-UV CD spectra of the WT and IGPS clamp-deletion variants
- Figure 3.4 Equilibrium unfolding properties of IGPS variants
- Figure 3.5 Positional preference of Class 1 $\beta\alpha$ -hairpin clamps in 71 TIM barrels
- Figure 3.6 Architectural principles of the TIM barrel fold
- Figure A.1 The dependence of the amplitude for the unfolding phase for WT and N231A in eIGPS on the initial urea concentration
- Figure A.2 Far-UV and near-UV CD spectra of WT, T129A, and D128AT129A sIGPS
- Figure A.3 Urea denaturation equilibrium unfolding curves of WT, T129A, and D128AT129A sIGPS
- Figure A.4 Far-UV and near-UV CD spectra of the WT and S82D eIGPS
- Figure A.5 Urea denaturation equilibrium unfolding curves of WT and S82D eIGPS

ABBREVIATIONS

H-bond	hydrogen bond
TIM	triose phosphate isomerase
α TS	α -subunit of tryptophan synthase from <i>E.coli</i>
sIGPS	indole-3-glycerol phosphate synthase from <i>S. solfataricus</i>
eIGPS	indole-3-glycerol phosphate synthase from <i>E.coli</i>
CD	circular dichroism
FL	fluorescence
MRE	mean residue ellipticity
N	native state
U	unfolded state
I	intermediate state
I _{BP}	off-pathway burst-phase intermediate
WT	wild-type
HX	hydrogen-deuterium exchange
HX-MS	hydrogen-deuterium exchange mass spectrometry
HSQC	heteronuclear single quantum coherence
ILV	isoleucine, leucine, and valine
β ME	β -mercaptoethanol
DTE	dithioerythritol

PREFACE

References to publications that represent the work contained within this thesis:

Chapter II:

Yang X., Vadrevu R., Wu Y., Matthews C.R. 2007. “Long-range side-chain-main-chain interactions play crucial roles in stabilizing the $(\beta\alpha)_8$ barrel motif of the alpha subunit of tryptophan synthase.” *Protein Sci.* July; **16**:1398-1409.

Contributions to this work are as follows:

Xiaoyan Yang collected and analyzed all of the data; Ramakrishna Vadrevu found the distinct downfield cross peaks in his NMR work on α TS and initiated the project;

Ying Wu provided instructions on experimental work on α TS and insightful discussions regarding the project; Bob Matthews served as the principal investigator on this project and provided insightful analysis and discussions of the studies.

Chapter III:

Yang X., Kathuria S., Vadrevu R., Matthews C.R. “ $\beta\alpha$ -hairpin clamps brace $\beta\alpha\beta$ modules and can make substantive contributions to the stability of TIM barrel proteins.” *In preparation*

Contributions to this work are as follows:

Xiaoyan Yang performed the mutagenesis and characterization of the proteins; Sagar Kathuria performed the survey of 71 TIM barrel proteins; Ramakrishna Vadrevu helped design the queries for the survey analysis and provided insightful discussions

on the project; Bob Matthews served as the principal investigator on this project and provided insightful analysis and discussions of the studies.

Other publications:

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 alpha subunit of tryptophan synthase, a TIM barrel protein." *J. Mol. Biol.* Mar 9; **366**:1624-38.

Xiaoyan Yang participated in the protein purification and characterization of the α TS variants L50A, L127A, and L176A.

Wu Y., Vadrevu R., **Yang X.**, Matthews C.R. 2005. "Specific structure appears at the N terminus in the sub-millisecond folding intermediate of the alpha subunit of tryptophan synthase, a TIM barrel protein." *J. Mol. Biol.* Aug 19; **351**:445-52.

Xiaoyan Yang participated in the protein purification and characterization of the α TS variants L25A, I151A, and I232A.

CHAPTER I
INTRODUCTION

Overview

One of the foremost problems in biology is understanding how newly-synthesized proteins fold so rapidly and efficiently into the well-defined three-dimensional structures that support their essential biological functions. The failure of this process can lead to aggregation and numerous diseases, including Huntington's disease, Parkinson's disease, amyotrophic lateral sclerosis, Alzheimer's disease, and even cancers (Joerger and Fersht 2008; Soto and Estrada 2008; Winklhofer et al. 2008). Over the years, many investigators have worked on solutions for the protein-folding problem, both *in vitro* and *in vivo* (Barrick 2009; Bartlett and Radford 2009; Hartl and Hayer-Hartl 2009), attempting to reveal its basic principles and to discover how the amino acid sequence encodes the native functional conformation of a protein (Wiederstein and Sippl 2005). Because the native conformation is thought to be the most stable structure of a protein (Anfinsen 1973), a major focus of protein folding studies has been the factors that stabilize the native conformation.

Protein stability

Proteins require minimum stability to perform their biological functions. Stability arises from the complex interplay between the hydrophobic effect, hydrogen bonds (H-bonds), disulfide bonds, electrostatic interactions, and Van der Waals forces (Murphy 2001). Among these interactions, the two major factors affecting protein stability are thought to be the hydrophobic effect and hydrogen bonding (Pace et al. 1996). The hydrophobic effect has long been considered the driving force for protein folding

(Kauzmann 1959), but the role of H-bonds has been less consistently appreciated. For example, early in the twentieth century, H-bonds were viewed as highly important in protein structure, but not in protein stability (Mirsky and Pauling 1936). However, despite their evident structural role, thermodynamic evidence that “hydrogen bonding or van der Waals interactions among polar amino acids may be important, ...” was lacking (Dill 1990). Later, the importance of H-bonds was demonstrated in a series of mutational experiments that showed “hydrogen bonding and the hydrophobic effect make large but comparable contributions to the stability of globular proteins” (Myers and Pace 1996). The significant role of the peptide H-bonds to protein stability has been shown in the osmolyte-induced folding process (Bolen and Rose 2008).

Protein H-bonds

H-bonds are one of the most prevalent intramolecular and intermolecular non-covalent interactions in protein structures. H-bonds stabilize the backbone structural building blocks such as α -helices or β -sheets (Murphy 2001). In an α -helix, H-bonds occur regularly between positions i and $i + 4$ in amino acid sequences. A 3_{10} helix is formed when H-bonds occur between positions i and $i + 3$, and β -sheets are stabilized by H-bonding networks between neighboring β -strands in either parallel or anti-parallel alignments. H-bonds also participate in the tertiary structure of proteins through interactions between polar side chains, and side-chain H-bonding interactions are usually related to molecular recognition and enzymatic functions (Liu et al. 2000). On the

protein surface, unsatisfied H-bond donors or acceptors, in side chains or the backbone usually form H-bonds with water molecules.

The apparent conservation of H-bonds during folding, where H-bonds to water are replaced with intramolecular H-bonds, led to the initial proposal that H-bonds did not contribute significantly to protein stability (Mirsky and Pauling 1936). In the unfolded state of a protein, all potential H-bond donors and acceptors are satisfied with the solvent. After a protein folds, only 1.3% of backbone amino groups and 1.8% of carbonyl groups fail to form at least one H-bond. However, 80% of main chain carbonyls fail to form a second H-bond (McDonald and Thornton 1994). Thus, the H-bonding network appears to be destabilized after the protein folded. At the same time, the rupture in H-bonding to water would increase the entropy of the solvent. If the weights of these two factors are comparable, H-bonds would not make a significant contribution to protein stability.

This view was challenged by accumulating evidence that H-bonds make a positive contribution to protein stability (Myers and Pace, 1996). By removing the side chain of a H-bond donor or acceptor with mutagenesis, the contribution of the H-bonds was then estimated by comparing the stability of the variant to that of the corresponding wild-type (WT) protein. Site-directed mutagenesis studies on single side chain-side chain or side chain-main chain H-bonds indicate their contribution to protein stability is relatively small, $\sim 1\text{-}2 \text{ kcal mol}^{-1}$ (Horovitz et al. 1990; Serrano et al. 1992; Myers and Pace 1996; Pace et al. 2009). For example, the stability of the homodimeric coiled-coil peptide, GCN4-p1, was reduced less than 1 kcal mol^{-1} , when several salt bridges were disrupted by a single mutation (Ibarra-Molero et al. 2004). In another example, the

stability of staphylococcal nuclease was reduced after mutating eight polar residues individually, resulting in the loss of H-bonds and ion pairs (Schwehm et al. 2003). For example, the reduction in protein stability by 3 kcal mol^{-1} for the D77A protein was attributed to the loss of two H-bonds with Y91 and T120, respectively. The remaining seven variants had only minor perturbations in protein stability (Schwehm et al. 2003). Other examples include ribonuclease T1, in which twelve single mutations individually removed H-bonds. An average contribution of $1.3 \text{ kcal mol}^{-1}$ for each H-bond was observed (Shirley et al. 1992). For barnase, the Y78A mutation removed the two H-bonds between G81 and Y78 and resulted in a net destabilization of $1.4 \text{ kcal mol}^{-1}$ (Serrano et al. 1992). The variant D33A in ribonuclease Sa disrupted three intramolecular H-bonds, resulting in the loss of 6 kcal mol^{-1} in stability (Trevino et al. 2005). So the average observed magnitude of the strength of one H-bond is $1\text{-}2 \text{ kcal mol}^{-1}$ (Gao et al. 2009).

Despite these results, site-directed mutagenesis has several limitations. First, mutagenesis may change the hydrophobicity of the protein. For example, the side chain of Y24 in barnase forms a H-bond with the backbone carbonyl of D75. The Y24F mutation, which breaks this H-bond, did not change the overall protein stability. The loss of the energy due to breaking the H-bond was compensated by the increase in hydrophobicity by adding a Phe (Serrano et al. 1992). Second, the replacement of one amino acid may change the side chain entropy. For instance, the alanine mutation may increase the helical propensity by lowering the side chain entropy, and enhance protein stability (Chellgren and Creamer 2006). Third, the freed H-bond donor or acceptor may

make H-bonds with other partners or water molecules, especially if the replacement residue and the original residue are not comparable in size. These limitations may lead to ambiguities in estimating the contribution of H-bonds to protein stability.

Side chain-main chain H-bonds

Of all the intramolecular H-bonds, 68.1% are H-bonds formed between the backbone carbonyl oxygen and amide hydrogen. 10.6% are side chain-side chain H-bonds, which are usually long range in nature, and the remaining 21.3% are H-bond interactions between main chain and side chain residues (Myers and Pace 1996). Half of these side chain-main chain H-bonds are between main chain amide hydrogens and another side chain, and the other half are between main chain carbonyl oxygens and the side chain of another residue. A majority of these side chain-main chain interactions are local, involving side chains within about six residues from either side of the main chain peptide bond (Baker and Hubbard 1984; Stickle et al. 1992). For instance, side chain-main chain H-bonds play an important role in stabilizing α -helices as helix-capping motifs (Jimenez et al. 1994; Aurora and Rose 1998).

Surprisingly, in studies of the alpha subunit of tryptophan synthase (α TS), a TIM barrel protein, using native-state hydrogen exchange (HX), the slow exchange of the amide hydrogen of F19 (several days) and the extremely slow exchange of the amide hydrogen of I97 (several months) in $^2\text{H}_2\text{O}$ indicated that these residues are highly protected in the native state (Vadrevu et al. 2008). Examination of the protein structure revealed that the backbone amide hydrogens of F19 and I97 are involved in H-bonding

interactions with the side chains of D46 and D124, respectively (Vadrevu et al. 2003). As the donor and acceptor of each H-bond are 27 amino acids apart, these H-bonds are long-range interactions. Structural analysis of α TS shows that these long-range non-covalent interactions connect the N-terminus of one β -strand with the C-terminus of the succeeding and anti-parallel α -helix (F19-D46 and I97-D124). By analogy, these interactions are designated as $\beta\alpha$ -hairpin clamps. The strong protection against HX suggested that these long-range side chain-main chain H-bonds between F19-D46 and I97-D124 might be crucial to the structure and stability of both native and partially folded states in α TS. The main purpose of this dissertation is to study the role of the clamp H-bonds in protein structure and stability.

Clamp H-bonding interactions partition into two classes (Figure 1.1). For Class I clamp H-bonds, the main chain amide hydrogen is the H-bond donor and the acceptor is a polar side chain such as Asn, Asp, Gln, Glu, Ser, Thr, and Tyr; this class is denoted as $MC_{NH} \rightarrow SC$. For Class II clamp H-bonds, the backbone carbonyl oxygen atom serves as the acceptor while the H-bond donor is provided by polar side chains such as Arg, Asn, Gln, His, Lys, Ser, Thr, Trp, and Tyr; this class is denoted as $SC \rightarrow MC_{C=O}$.

Structurally, each of the two $\beta\alpha$ -hairpin H-bonds in α TS brackets exactly one β -strand and one α -helix. These clamps brace the $\beta\alpha\beta$ module, the fundamental building block of the TIM barrel structure (see below) (Branden and Tooze 1999). Similar clamp H-bonds that bracket a consecutive pair of secondary-structure elements ($\beta\alpha$ -hairpin clamps) are observed in other TIM barrel proteins, such as the *ioI* protein from *B. subtilis*, and the cyclase moiety of imidazole-3-glycerol phosphate synthase (HisF) from

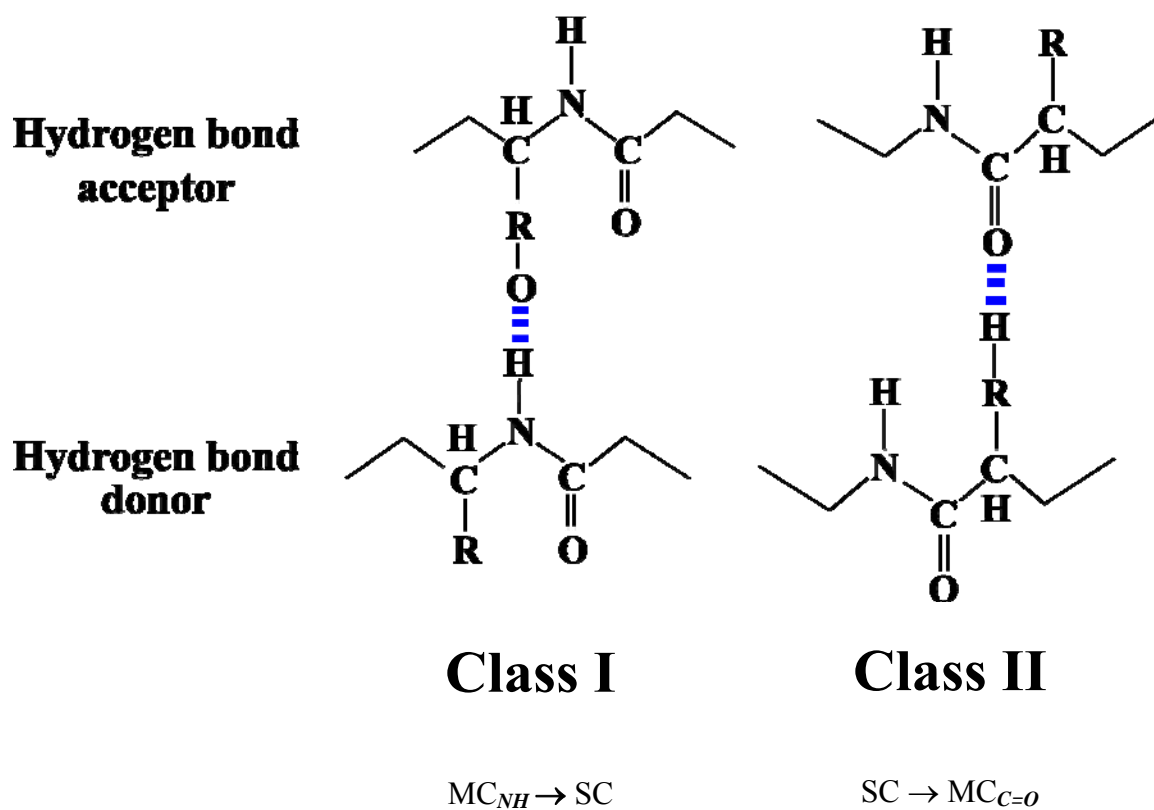


Figure 1.1. Classification of long-range side chain-main chain H-bonds (clamp H-bonds).

Class I clamp H-bonds are formed between the main chain amide hydrogen and another side chain. Class II clamp H-bonds are formed between the backbone carbonyl oxygen atom and another side chain.

T. maritima. These findings suggest that the $\beta\alpha$ -hairpin clamp H-bonds in TIM barrel proteins might have a general role in defining the TIM barrel architecture.

TIM barrel proteins

One of the frequent targets for protein folding studies is the $(\beta\alpha)_8$ barrel family, one of the most common structural motifs in biology. Because the first protein structure solved with the $(\beta\alpha)_8$ motif was triosephosphate isomerase (TIM) (Sterner and Hocker 2005), this family is also referred to as the TIM barrel family. Many TIM barrels are enzymes with a variety of catalytic functions, and include oxidoreductases, transferases, lyases, hydrolases, and isomerases (Nagano et al. 2002).

Structure

The canonical TIM barrel fold consists of a central hydrophobic β -barrel made up of eight β -strands surrounded by a shell of eight amphipathic α -helices. Sequentially, a β -strand is followed by an α -helix that lies anti-parallel to the β -strand on which it docks. There are eight such $\beta\alpha$ units, which serve as repeat units for the TIM barrel motif (Figure 1.2a). The cylindrical core of the eight parallel β -strands is stabilized by H-bonds between the β -strands. The axes of the β -strands are tilted about 36° to the axis of the barrel, which reflects the staggered arrangement of the side chain in the barrel. The individual elements of secondary structure are connected by loops. Longer, more flexible loops between β -strands and α -helices ($\beta\alpha$ -loops) invariably define the active site (the catalytic face), whereas the shorter, more rigid loops between α -helices and β -strands

Figure 1.2. Representation of the TIM barrel protein motif. (a) Schematic model of the eight $\beta\alpha$ -repeats for the alpha subunit of tryptophan synthase (α TS). Secondary structural elements are indicated. (b) Ribbon diagram of α TS, an example of the TIM barrel protein family. The blue arrows highlight the β -strands. The red and yellow spiral ribbons highlight the α -helices and the gray segments represent the loops that connect the β -strands and the α -helices. As displayed in this figure, the top of the barrel is the catalytic face and the bottom of the barrel is the stability face.

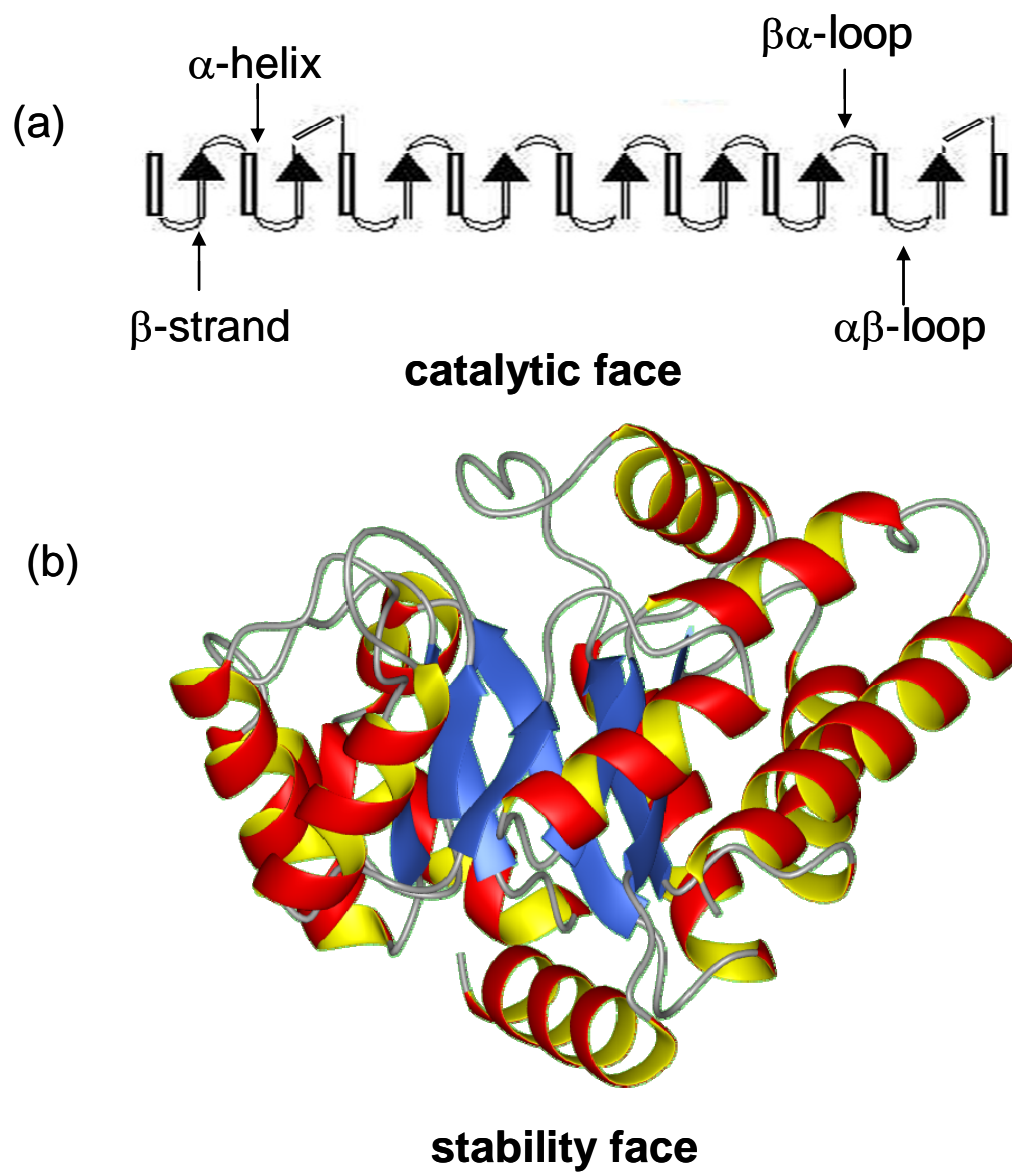


Figure 1.2

($\alpha\beta$ -loops) are thought to stabilize the barrel (the stability face) (Figure 1.2) (Urfer and Kirschner 1992; Sterner and Hocker 2005). Consistent with this view, $\beta\alpha$ -loops have been shown by knowledge-based potential analysis to have a large fraction of stable variants, but most $\alpha\beta$ -loops have almost no tolerance for mutation (Wiederstein and Sippl 2005).

In general, barrel structures can be classified based on geometric parameters such as the strand number and the shear number (Nagano et al. 1999). Using this classification, the TIM barrel structure has strand number $n=8$ and shear number $S=8$ (Figure 1.3). The side chains of β -strands form several layers inside the barrel. Each layer has four residues from four odd- or even-numbered strands. These layers reflect the alternating inside-outside positioning of the side chains in each β -strand (Figure 1.3).

Stability

For TIM barrel proteins, the central β -barrel core is the major source of the stability, with the $\alpha\beta$ -loops also playing an important role (Urfer and Kirschner 1992). Computational analysis of 71 TIM barrel proteins identified 957 residues that were predicted to play a crucial role in stability (Gromiha et al. 2004). Most of these stabilizing residues (80%) were in the β -strands, with far fewer found in α -helices. These residues were identified based on three major factors: long-range interactions, hydrophobicity of the environment, and sequence conservation (Gromiha et al. 2004).

The contribution of clusters of aliphatic side chains to protein stability has been proposed based upon correlations of large clusters of isoleucine, leucine, and valine (ILV)

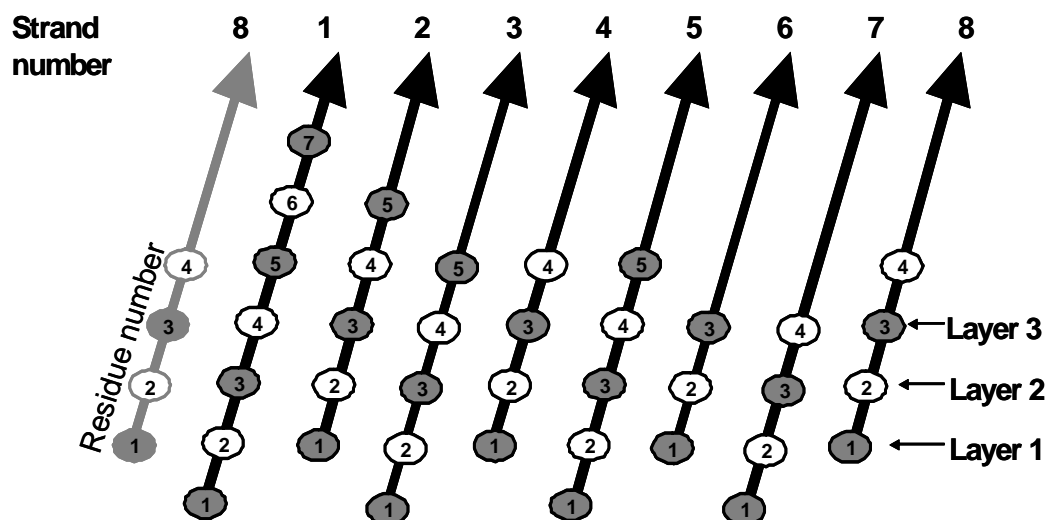


Figure 1.3. Schematic model of the TIM barrel fold. The eight β -strands of the TIM barrel architecture are represented by thick arrows. The β -strand number is indicated at the C-terminus of each β -strand. To convey the closed barrel architecture, $\beta 8$ is shown adjacent to $\beta 1$ (thick gray arrow) as well as adjacent to $\beta 7$. Each residue is represented as an ellipse and the position on the β -strand is indicated. The filled ellipses represent residues that point inside the barrel. The open ellipses represent residues that point outside the barrel. Layers formed by side chains are indicated.

side chains with the cores of stability in partially-folded states in two TIM barrel proteins, α TS and indole-3-glycerol phosphate synthase (IGPS) (Gu et al. 2007b; Wu et al. 2007). The unfavorable partitioning of small molecule analogs of the ILV side chains into water from the vapor phase (Radzicka and Wolfenden 1988) was involved to explain the strong resistance to HX for the underlying main chain amide hydrogens. The strengthening of the H-bonding network in the β -barrel by the hydrophobic cluster implies a linkage between the tertiary structure and the secondary structure (Trevino et al. 2008).

The **goal of this thesis** is to investigate the contribution of a not well recognized but as will be shown, rather common type of side chain-main chain H-bond in TIM barrel proteins. The results suggest that a subset of the $\beta\alpha$ -hairpin clamp interactions can make a significant contribution to structure and stability. As such, the clamps offer opportunities to engineer enhanced stability in TIM barrel proteins.

Targets for study

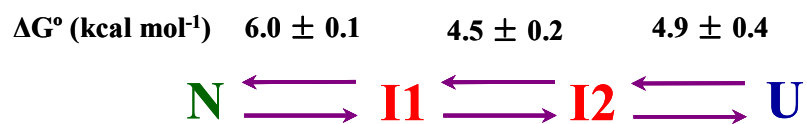
To study the $\beta\alpha$ -hairpin clamp H-bonds in TIM barrel proteins, I chose three target proteins: α TS from the mesophilic bacterium *E. coli*, IGPS from the thermoacidophilic archaeobacterium *S. solfataricus* (sIGPS), and IGPS from *E. coli* (eIGPS). Structural information is available and the folding mechanisms have been determined for these proteins (Hyde et al. 1988; Wilmanns et al. 1992; Sanchez del Pino and Fersht 1997; Bilsel et al. 1999b; Gualfetti et al. 1999; Forsyth and Matthews 2002; Nishio et al. 2005; Schneider et al. 2005), providing the information necessary to evaluate the roles of the $\beta\alpha$ -hairpin clamps in stabilizing their respective TIM barrel structures.

The alpha subunit of tryptophan synthase (α TS)

The α subunit, together with the β subunit, comprise the $\alpha_2\beta_2$ enzyme complex, which catalyzes the condensation of indole-3-glycerol phosphate and serine into tryptophan (Xie et al. 2002). α TS from *E. coli* has eight β -strands and eleven α -helices. One of the non-canonical α -helices precedes the first β -strand, and two others are short α -helices that are inserted into the loops preceding α_2 and α_8 . Because the protein does not have any disulfide bonds or prosthetic groups, it is a good model system for protein folding studies.

The folding mechanism of α TS is well characterized (Bilsel et al. 1999b; Gualfetti et al. 1999). In urea-denaturation equilibrium unfolding studies, α TS shows two stable intermediates, I1 and I2, which are populated at 3 and 5 M urea, respectively (Bilsel et al. 1999b; Gualfetti et al. 1999) (Figure 1.4a). The I1 intermediate, which can be detected by CD and fluorescence anisotropy, was originally thought to have a structured amino terminal region and an unfolded C-terminal region (Bilsel et al. 1999a). More recent native-state hydrogen-exchange NMR analysis suggests an intact β -barrel with a loosely-structured α -helical shell (Vadrevu et al. 2008). In contrast, the I2 intermediate has a structure that, by optical methods, resembles the unfolded state and is thought to be stabilized by the hydrophobic effect (Saab-Rincon et al. 1993).

(a)



(b)

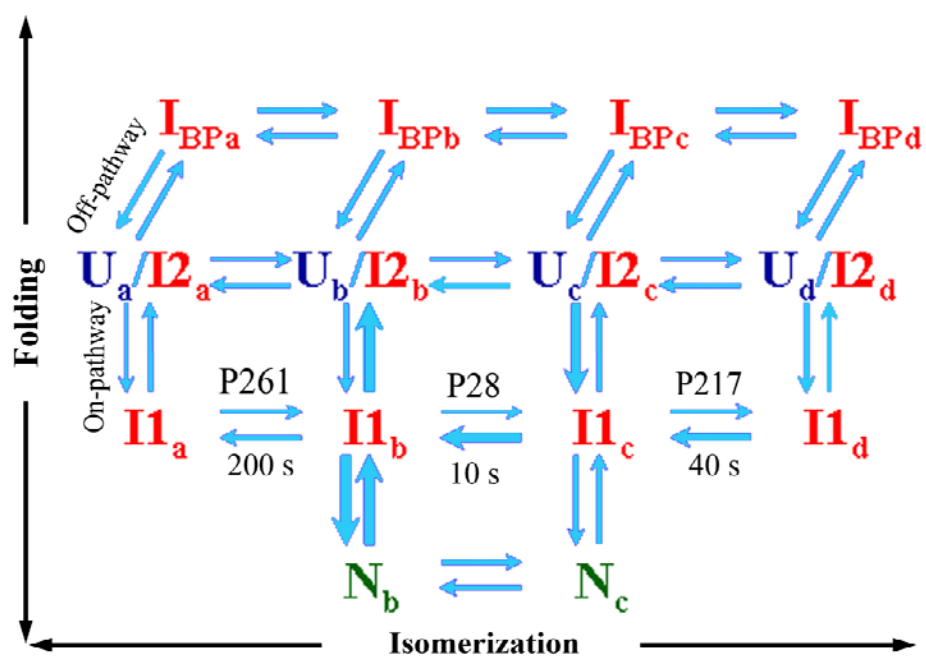


Figure 1.4. (a) Equilibrium (Gualfetti et al. 1999) and (b) kinetic (Bilsel et al. 1999b) folding models of α TS.

The kinetic folding mechanism of α TS is complex. The protein folds through four parallel channels (Figure 1.4b), each with on- and off-pathway intermediates. Upon folding, the first detectable step involves a rapid collapse (<5 ms) to form an off-pathway burst-phase species (I_{BP}) in each channel, whose unfolding controls the formation of an on-pathway intermediate in each channel (Bilsel et al. 1999b). Although the principal native species forms within hundreds of milliseconds, three proline isomerization reactions, which involve P28, P217, and P261, are responsible for the slow interconversion between these on-pathway intermediates (I1a-I1d) and serve as the rate-limiting steps to the native states (Wu and Matthews 2002b; Wu and Matthews 2002a; Wu and Matthews 2003).

$\beta\alpha$ - and $\alpha\beta$ -hairpin clamps in α TS

Examination of the ^1H - ^{15}N heteronuclear single quantum coherence (HSQC) spectrum of α TS revealed three distinct downfield cross peaks at ~ 10 ppm (Figure 1.5) (Vadrevu et al. 2003). These peaks were found to result from H-bond interactions between the backbone amide hydrogens of F19, I97, and A103 with the anionic side chains of D46, D124, and D130, respectively (Vadrevu et al. 2003). The F19-D46 H-bond clamps the secondary structure elements $\beta 1$ and $\alpha 1$, and the I97-D124 H-bond clamps $\beta 3$ and $\alpha 3$. The A103-D130 H-bond clamps $\alpha 3$ and $\beta 4$, and it represents an $\alpha\beta$ -hairpin clamp. All three side chain-main chain interactions are Class I clamp H-bonds (Figure 1.1), with the main chain amide nitrogen serving as the H-bond donor and the carboxyl side chain of aspartic acid as the H-bond acceptor.

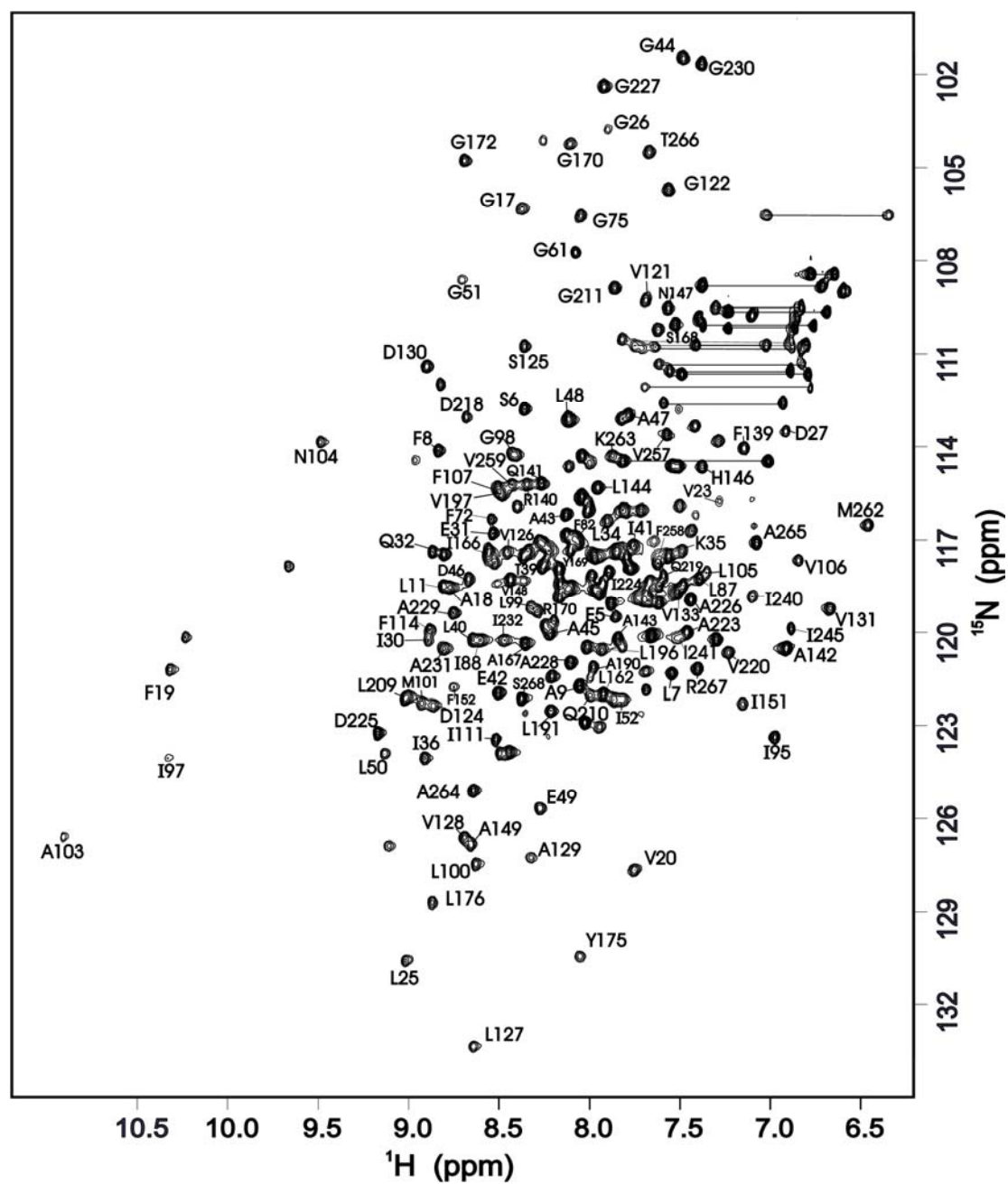


Figure 1.5. ^1H - ^{15}N HSQC spectrum of αTS . The three downfield (~ 10 ppm) cross peaks have been assigned to F19, I97, and A103 (Vadrevu et al. 2003).

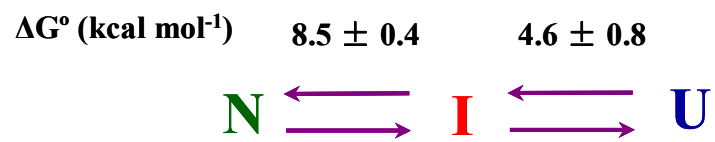
Indole-3-glycerolphosphate synthase (IGPS)

IGPS is involved in the fourth step of tryptophan synthesis, where it catalyzes the conversion of 1-(2-carboxyphenylamino)-1-deoxyribulose 5-phosphate (CdRP) to indole-3-glycerol phosphate (IGP) (Andreotti et al. 1994). Although the *S. solfataricus* IGPS (sIGPS) is a single-domain TIM barrel fold, the *E. coli* IGPS (eIGPS) chain is the N-terminal domain of a bifunctional enzyme with another TIM barrel domain N-(5'-phosphoribosyl anthranilate isomerase (PRAI) at the C-terminus (Altamirano et al. 2000). In the present study, the eIGPS domain was excised from the bifunctional enzyme and expressed as a single domain (Altamirano et al. 2000). Also the N-terminal 26 residues of sIGPS has been truncated to improve the solubility of the protein (Schneider et al. 2005). The structures of both forms of IGPS are available, and both folding mechanisms have been studied extensively (Sanchez del Pino and Fersht 1997; Forsyth and Matthews 2002; Gu et al. 2007a; Gu et al. 2007b).

***S. solfataricus* IGPS (sIGPS)**

sIGPS has been shown in urea-denaturation CD experiments to populate a stable intermediate at about 4.5 M urea. This intermediate retains about half of the ellipticity compared to the native state (Figure 1.6a) (Forsyth and Matthews 2002) and has a stability of 4.6 kcal mol⁻¹ relative to the unfolded state. The unfolding and refolding kinetics of sIGPS, as monitored by manual-mixing CD, stopped-flow CD and fluorescence (FL) methods, showed both a marginally stable off-pathway burst-phase (<5 ms) intermediate (I_{BP}) and a pair of on-pathway intermediates (Figure 1.6b)

(a)



(b)

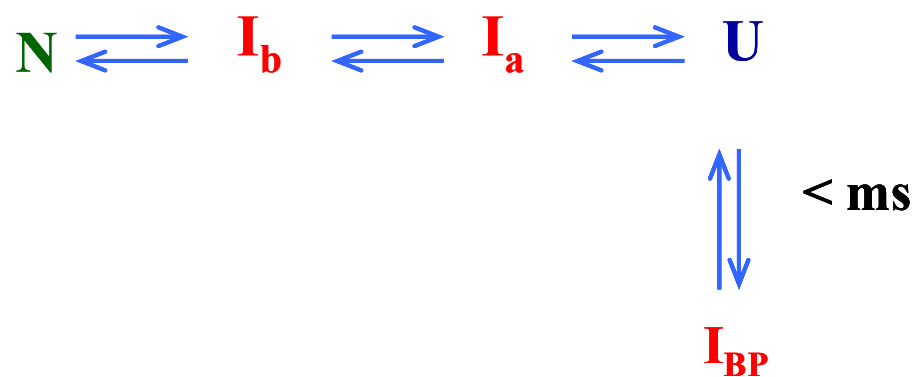


Figure 1.6. (a) Equilibrium (Forsyth and Matthews 2002) and (b) kinetic (Gu et al. 2007a) folding models of sIGPS.

(Forsyth and Matthews 2002; Gu et al. 2007a; Gu et al. 2007b), similar to α TS.

Comparison of the equilibrium and kinetic mechanisms has showed that the on-pathway intermediate is composed of two distinct species, I_a and I_b (Figure 1.6b).

The structures of these two on-pathway intermediates, as well as the off-pathway I_{BP} were probed by hydrogen-exchange mass spectroscopy (HX-MS). Peptide analysis from rapid pulse-quench experiments showed that the $(\beta\alpha)_4$ region is the most protected core in the first few milliseconds of folding. The $(\beta\alpha)_{1-3}$ and $(\beta\alpha)_5\beta_6$ segments are less well protected, and the remaining N- and C-terminal segments are not significantly protected (Gu et al. 2007a; Gu et al. 2007b). The initial protection in the $(\beta\alpha)_4$ segment was attributed to a high density of ILV residues and the rapid formation of a stable hydrophobic cluster. Strong protection in the $(\beta\alpha)_{2-5}\beta_6$ region after 5 s of folding demonstrates a stable core representing the on-pathway intermediate I_a . I_b has a more native-like structure, covering β_1 to β_7 . The increase in the size of the stable secondary structure as I_a matures to I_b and their similar estimated stability (I_a is ~ 0.7 kcal/mol more stable than I_b) provides a rationale for the observation of a single species in the equilibrium unfolding reaction. The protection against HX observed in the on-pathway intermediates was attributed to a pair of large ILV clusters that span the protected segments of the chain (Gu et al. 2007a; Gu et al. 2007b).

***E. coli* IGPS (eIGPS)**

eIGPS is the N-terminal domain (residues 1-255) of a bifunctional enzyme whose C-terminal domain is PRAI, which catalyzes the next step in the tryptophan synthesis

pathway (Ivens et al. 2002). eIGPS has only one tryptophan residue located at the N-terminal α -helix, α_0 , preceding the canonical TIM fold. eIGPS and sIGPS share 30% sequence identity, with sIGPS having a larger number of salt bridges and side chain-main chain H-bonds than eIGPS (Ivens et al. 2002). These additional non-covalent interactions reflect the environmental challenges experienced by thermophilic organisms.

The folding mechanism of eIGPS was proposed to follow a nonsequential unfolding process (Sanchez del Pino and Fersht 1997). Two different folding intermediates, appearing along two parallel unfolding pathways, were detected in guanidinium chloride-denaturation experiments monitored by CD, FL, and gel filtration chromatography. One intermediate (I_{NL}) was very native-like and kinetically trapped in the unfolding reaction. The other intermediate (I) was suggested to be a higher-order form, because the only tryptophan was shielded in this conformation (Figure 1.7).

Scope of the thesis

The goal of this thesis was to test the role of side chain-main chain clamp H-bonds in the stability and structure of several TIM barrel proteins. Clamp H-bonds were removed by mutating the side chain H-bond acceptors to alanine. The clamp-deletion variants were characterized by traditional protein folding techniques, e.g., urea-denaturation equilibrium and kinetic experiments monitored by CD. The contribution of the clamp H-bonds to structure and stability were determined by comparison with the properties of the corresponding WT protein.

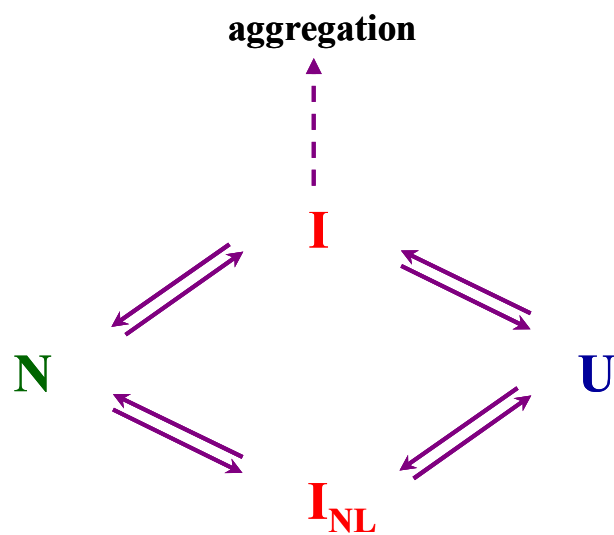


Figure 1.7. Proposed folding model of eIGPS (Sanchez del Pino and Fersht 1997).

In Chapter II, I report on a study in which α TS from *E. coli* served as the candidate (Yang et al. 2007). To complement equilibrium studies on three Class I clamp-deletion variants, I also performed kinetic studies to examine the effects of removing clamp H-bonds on a pair of rate-limiting transition states in the folding reaction. The unfolding and refolding reactions were probed with manual-mixing and stopped-flow CD and FL techniques. One local side chain-main chain H-bond and two salt bridges were tested as controls to show the uniqueness of the clamp interactions. Asparagine and glutamic acid replacements at one clamp H-bond were compared with the alanine mutation to test for the uniqueness of the aspartic acid as a clamp H-bond acceptor (Yang et al. 2007).

In Chapter III, I report on a study in which I tested three Class I $\beta\alpha$ -hairpin clamp H-bonds in two other TIM barrel proteins, sIGPS and eIGPS. To investigate the generality of these clamp H-bonds in the TIM barrel protein family, 71 TIM barrel proteins were surveyed for the frequencies, sequence preferences and locations of $\beta\alpha$ -hairpin clamp H-bonds. This analysis was performed in collaboration with another graduate student, Sagar Kathuria. The implications of the role played by clamp H-bonds in TIM barrel architecture are discussed. In this thesis, my goal was to deepen the understanding of a particular facet of TIM barrel structure. The insights obtained may improve structural predictions from TIM barrel sequences and provide a rational approach to engineer enhanced stability in this versatile enzymatic platform.

CHAPTER II

**LONG-RANGE SIDE CHAIN-MAIN CHAIN
INTERACTIONS PLAY CRUCIAL ROLES IN
STABILIZING THE $(\beta\alpha)_8$ BARREL MOTIF OF
THE ALPHA SUBUNIT OF TRYPTOPHAN SYNTHASE**

ABSTRACT

The role of hitherto unrecognized long-range hydrogen bonds between main chain amide hydrogens and polar side chains on the stability of a well-studied $(\beta\alpha)_8$, TIM barrel protein, the alpha subunit of tryptophan synthase (α TS), was probed by mutational analysis. The F19-D46 and I97-D124 hydrogen bonds link the N-terminus of a β -strand with the C-terminus of the succeeding anti-parallel α -helix, and the A103-D130 hydrogen bond links the N-terminus of an α -helix with the C-terminus of the succeeding anti-parallel β -strand, forming clamps for the respective $\beta\alpha$ or $\alpha\beta$ hairpins. The individual replacement of these aspartic acid side chains with alanine leads to what appear to be closely-related partially-folded structures with significantly reduced far-UV CD ellipticity and thermodynamic stability. Comparisons with the effects of eliminating another side chain-main chain hydrogen bond, G26-S33, and two electrostatic side chain-side chain hydrogen bonds, D38-H92 and D112-H146, all in the same N-terminal folding unit of α TS, demonstrated a unique role for the clamp interactions in stabilizing the native barrel conformation. Because neither the asparagine nor glutamic acid variant at position 46 can completely reproduce the spectroscopic, thermodynamic or kinetic folding properties of wild-type α TS with aspartic acid, both size and charge are crucial to its unique role in the clamp hydrogen bond. Kinetic studies suggest that the three clamp hydrogen bonds act in concert to stabilize the transition state leading to the fully-folded TIM barrel motif.

INTRODUCTION

Side chains stabilize the native, functional conformations of proteins by van der Waals interactions between nonpolar partners, hydrogen bonds (H-bonds) between polar donors and acceptors, and electrostatic interactions between opposing charged groups. Mutational analysis of buried nonpolar side chains has shown that replacements with alanine can lead to destabilizations of up to 7 kcal mol⁻¹ (Shortle et al. 1990; Eriksson et al. 1992; Serrano et al. 1992). The magnitudes of the effects correlate with the local packing density (Shortle et al. 1990) and the burial of nonpolar surface area (Jackson et al. 1993). With the exception of buried ionic interactions (Anderson et al. 1990; Vaughan et al. 2002), mutational analysis has also shown that the magnitude of an individual side chain-side chain or side chain-main chain H-bond or an electrostatic interaction is relatively small, ~1-2 kcal mol⁻¹ (Horovitz et al. 1990; Serrano et al. 1992; Myers and Pace 1996; Ibarra-Molero et al. 2004). The small contribution to stability for side chain-main chain interactions is underscored by the relatively rapid exchange of their amide hydrogens with solvent through local fluctuations within the native ensemble of conformers (Englander and Kallenbach 1983; Krishna et al. 2004). Thus, the results of a native-state hydrogen exchange (HX) experiment on the alpha subunit of tryptophan synthase (α TS) are quite striking.

α TS is a 29 kDa TIM barrel protein in which the eight parallel beta strands alternate in sequence with amphipathic anti-parallel helices that dock on the hydrophobic barrel H-bonds between β 1 and β 8 and van der Waals interactions between nonpolar side chains in β 1, α 1, β 8 and α 8'/ α 8 close the barrel (Figure 2.1a). Relevant to the present

Figure 2.1. Long-range side chain-main chain interactions in α TS. (a) Ribbon diagram of α TS highlighting the $\beta\alpha$ and $\alpha\beta$ clamps encompassing $\beta 1-\alpha 1$, F19 NH-D46 O δ 2; $\beta 3-\alpha 3$, I97 NH-D124 O δ 2, and $\alpha 3-\beta 4$, A103 NH-D130 O δ 1. The side chains involved in the H-bond clamp interactions are highlighted with the donor and acceptor shown in blue and red, respectively. (b) The three individual $\beta\alpha$ or $\alpha\beta$ clamps, illustrating the respective H-bond distances for each side chain-main chain interaction. The H-bonds and the corresponding distances are determined by using the program HBPLUS (McDonald and Thornton 1994). The structures were generated using PyMOL v 0.99 (Delano 2002) and PDB code: 1BKS (Hyde et al. 1988).

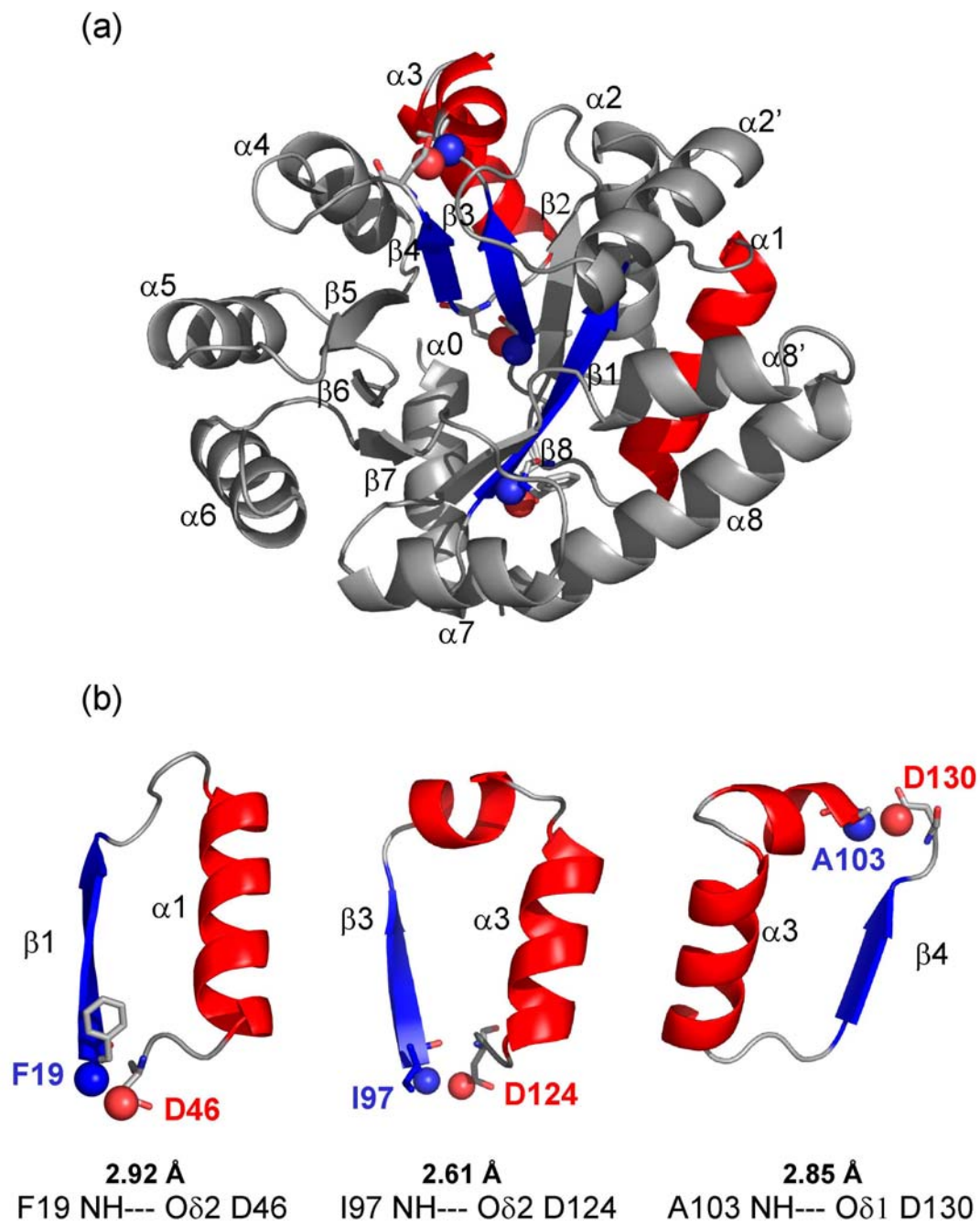


Figure 2.1

issue, the ^1H - ^{15}N heteronuclear single quantum coherence (HSQC) NMR spectrum of uniformly ^{15}N -labeled protein (Vadrevu et al. 2003) provided a vehicle for demonstrating that the main chain amide hydrogens of two residues, F19 and I97, are very resistant to exchange with solvent deuterium (Vadrevu et al. 2008). Inspection of the crystal structure of αTS (Hyde et al. 1988; Nishio et al. 2005) shows that both are involved in long-range side chain-main chain H-bonds, F19-D46 and I97-D124 (Figure 2.1b). The resistance of the amide hydrogen of F19 to exchange for several days and that of I97 for several months in $^2\text{H}_2\text{O}$ at pH 7.8 and 25 °C implies that their amide hydrogens become exposed to solvent in rare high energy states (Bai et al. 1995), not through local fluctuations in the native basin. By implication, the F19-D46 and I97-D124 H-bonds might play unusually important roles in the structure and stability of both native and partially-folded states in αTS .

This hypothesis was tested by subjecting the H-bond acceptor side chains, D46 and D124, to mutational analysis. To test the significance of these specific interactions, mutations were also introduced at another long-range main chain-aspartic acid side chain H-bond, A103-D130, and a short-range main chain-serine side chain H-bond, G26-S33, both of whose amide hydrogens exchange within a few hours. Comparisons were also made with mutations of a pair of surface salt bridges involving aspartic acids in the same N-terminal region, D38-H92 and D112-H146. The results demonstrated that each of the F19-D46, I97-D124 and A103-D130 side chain-main chain H-bonds makes a significant contribution to the structure and the stability of αTS that is distinct from the relatively minor perturbations for the other variants.

RESULTS

The X-ray structure of α TS (Figure 2.1a) shows that the F19-D46, I97-D124 and A103-D130 H-bonds each bracket a sequential pair of elements of secondary structure. Each of these three long-range H-bond interactions involves a main chain amide hydrogen donor and a side chain oxygen acceptor in aspartic acid. The F19-D46 H-bond links the N-terminus of β 1 with the C-terminus of α 1, the I97-D124 H-bond links the N-terminus of β 3 with the C-terminus of α 3 and the A103-D130 H-bond links the N-terminus of α 3 with the C-terminus of β 4 (Figure 2.1b). All three interactions span precisely 27 residues in sequence, and, by analogy, serve as structural clamps for $\beta\alpha$ hairpins, F19-D46 and I97-D124, or an $\alpha\beta$ hairpin, A103-D130, in α TS.

The uniqueness of these structural clamp interactions can be assessed by comparisons with other non-covalent interactions involving short-range or surface H-bonds. The G26-S33 H-bond braces a tight turn between β 1 and α 1 that contains the only cis prolyl peptide bond in α TS between D27 and P28. The D38-H92 and D112-H146 salt bridges on the surface of the barrel link the middle of α 1 to the loop between α 2 and β 2 and the N-terminus of α 3 with the loop between α 4 and β 5, respectively. (In the higher resolution structure of α TS from *S. typhimurium* (Hyde et al. 1988), D112 appears to form a second salt bridge with R145).

Alanine replacements at D46, D124 or D130 significantly perturb the secondary and tertiary structure of α TS

The effects of replacing D46, D124 or D130 with alanine on the secondary and tertiary structure of the α TS variants were determined by collecting far- and near-UV CD spectra (Figure 2.2). The far-UV CD spectra show a maximum mean residue ellipticity (MRE) at 195 nm and minima at 208 and 222 nm, demonstrating that all of the variants retain components of the α -helices and β -strands found in the wild-type (WT) protein. However, the significant decrease in the MRE by \sim 40% for the variants (Figure 2.2a-c) at all three wavelengths indicates that each individual replacement causes a substantial loss of secondary structure.

To probe the effects of the D46A, D124A and D130A variants on the packing of aromatic side chains, the near-UV CD spectra were collected (Figure 2.2e-g). The near-UV CD spectra of the D46A (Figure 2.2e) and D124A (Figure 2.2f) variants, which are very similar to each other and distinctly different from both D130A (Figure 2.2g) and WT α TS, indicate altered chiral environments for one or more of the 7 tyrosines and 12 phenylalanines (α TS does not contain tryptophan). The nearly featureless spectrum of D130A α TS between 270 and 290 nm suggests that the tyrosine side chains are mobile. Small peaks near 260 nm imply that one or more of the phenylalanines experience chiral packing environments.

These data show that the elimination of any one of the F19-D46, I97-D124 or A103-D130 H-bonds and/or the loss of the negative charge when aspartic acid is replaced by alanine have a substantial impact on the secondary and tertiary structure of α TS. The very similar response in the far-UV CD spectrum for all three variants (Figure 2.2d) and two of the three variants in the near-UV CD spectrum (Figure 2.2h) suggests that the loss

Figure 2.2. Far-UV (a)-(d) and near-UV (e)-(h) CD spectra of WT α TS (●) and three α TS variants: (a) and (e) D46A α TS (Δ), (b) and (f) D124A α TS (\diamond), (c) and (g) D130A α TS (\square). Panel (d) displays all the variant spectra in panels (a)-(c) for a direct comparison, and panel (h) displays all the variant spectra in panels (e)-(g). The protein concentration ranged from 5 to 7 μ M for the far-UV CD spectra and from 50 to 150 μ M for the near-UV CD spectra. The buffer contained 10 mM potassium phosphate, pH 7.8, 0.2 mM K_2EDTA , and 1 mM β ME at 25 °C.

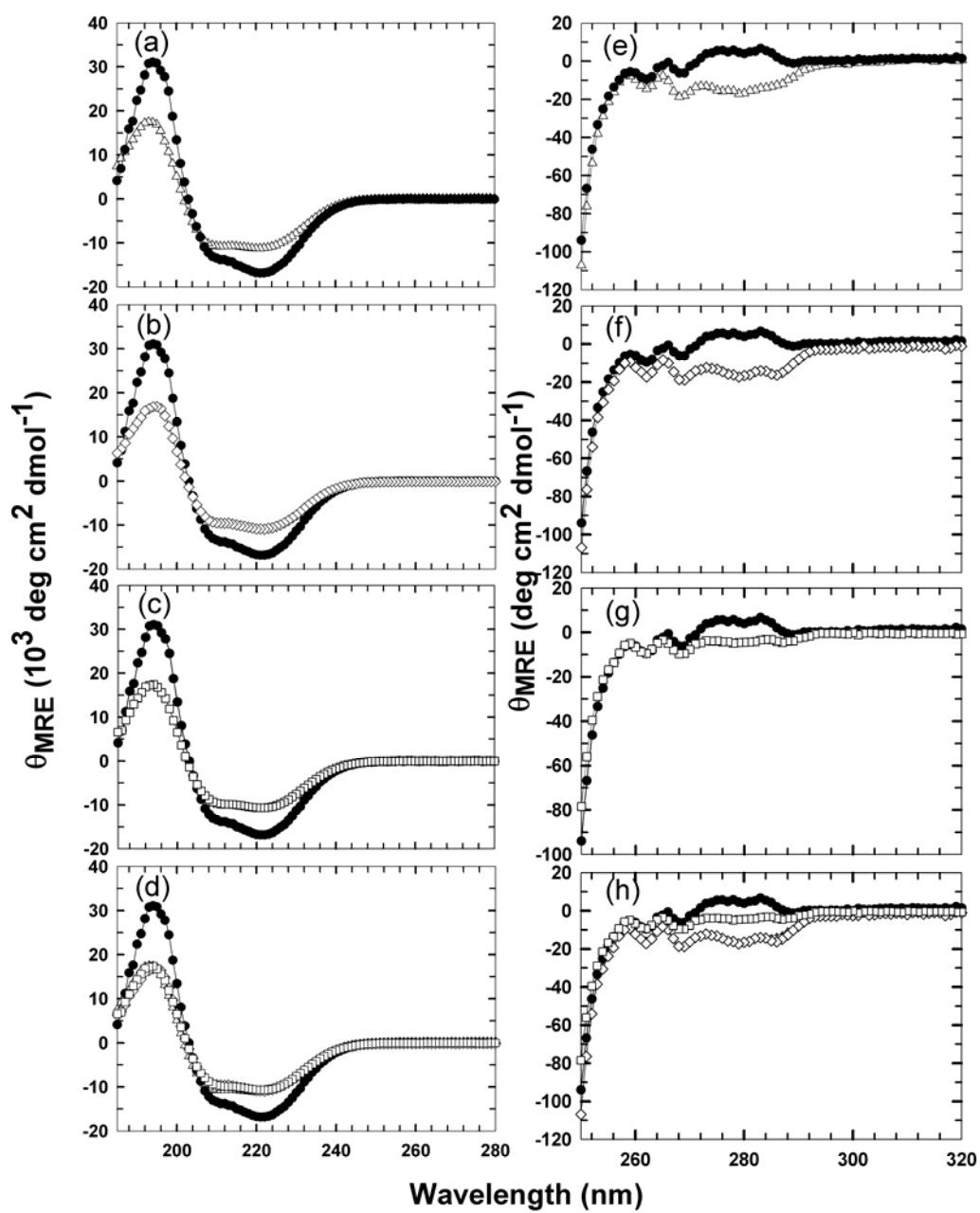


Figure 2.2

of either the F19-D46 or the I97-D124 non-covalent interactions lead to similar partially-folded structures. These folds retain a substantial fraction of the secondary structure and maintain tertiary packing around some of the aromatic side chains. The contrasting near-UV CD spectrum for D130A α TS shows that at least some aspect of the tertiary structure of this variant differs from those of the other two clamp variants.

Equilibrium folding properties of the clamp-deletion variants

Urea denaturation experiments were performed to test the contribution of the three main chain amide-aspartic acid side chain H-bonds to the thermodynamic properties of α TS. The dependence of the MRE at 222 nm on the urea concentration for each variant is compared to WT α TS in Figure 2.3. The equilibrium unfolding curve of the WT protein displays two overlapping sigmoidal transitions that have previously been assigned to interconversions between the native, N, and intermediate, I1, states and between the I1 and the unfolded-like, I2, and unfolded, U, states (Bilsel et al. 1999b). The far-UV CD spectra of I2 and U are identical, permitting the application of a three-state model, $N \rightleftharpoons I1 \rightleftharpoons I2/U$, to fit the equilibrium data. The I1 state for WT α TS reaches a maximal population of ~60% at 3 M urea (Gualfetti et al. 1999), and its presence is evident in the change in slope of the transition curve at that urea concentration (Figure 2.3).

The urea denaturation profile of the D130A clamp-deletion variant could be fit to the same three-state model used for WT α TS. This fitting procedure was justified by the appearance of a limited native baseline below 0.5 M urea and a small shoulder near 2 M

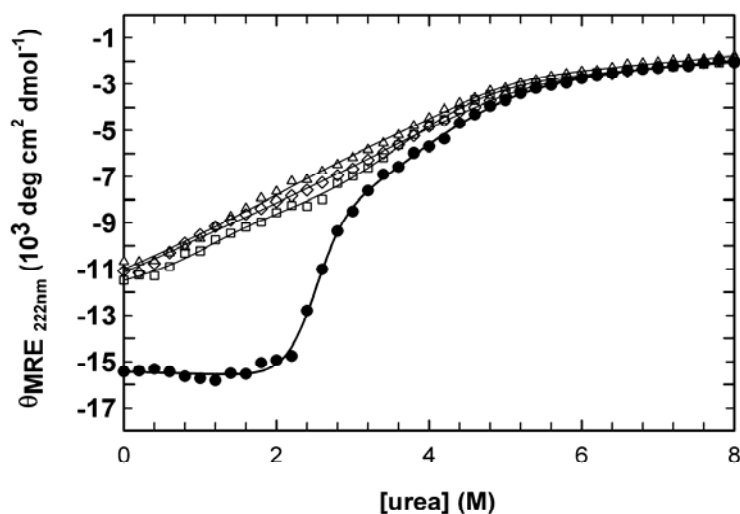


Figure 2.3. Urea denaturation equilibrium unfolding curves of D46A α TS (\triangle), D124A α TS (\diamond) and D130A α TS (\square). The data for WT α TS (\bullet) are shown for comparison in each panel. The continuous lines represent fits of the data for each α TS variant to a 3-state equilibrium folding model as described in the text. The protein concentration was $\sim 5 \mu\text{M}$, and the buffer contained 10 mM potassium phosphate, pH 7.8, 0.2 mM K_2EDTA , and 1 mM βME at 25 $^\circ\text{C}$.

urea. As shown in Table 2.1, the free energy difference between the N and I1 states is greatly reduced, as is the urea-dependence of this free energy difference, i.e., the m value, from the values for WT α TS. The $\Delta G^{\circ}_{N \rightleftharpoons I1}(\text{H}_2\text{O})$ parameters for D130A α TS and WT protein are 1.17 ± 0.39 and 7.19 ± 0.58 kcal mol⁻¹, respectively; the respective m values are 1.19 ± 0.21 and 2.85 ± 0.24 kcal mol⁻¹ M⁻¹. This three-state fit of the equilibrium unfolding data for D130A was confirmed by a complementary kinetic measurement of the stability of the native state (see below).

Unfortunately, both of the $\beta\alpha$ clamp-deletion variants, D46A and D124A α TS, display broad rather featureless transition curves (Figure 2.3) that could not be uniquely fit to a three-state model. In fact, the appearance of the unfolding profiles made it unclear whether the N, I1 and I2/U thermodynamic states were still being sampled in these two variants. As will be shown below, kinetic folding studies can be used to demonstrate the presence of stable folded states in the absence of denaturant and to estimate the stabilities of D46A and D124A α TS.

Kinetic folding properties of the clamp-deletion variants

To obtain insights into the role of the clamp interactions on the kinetic and thermodynamic properties of α TS, a comprehensive set of unfolding and refolding jumps at various final urea concentrations was performed on the clamp-deletion variants. A distinctive feature of the native state of WT α TS is its rate-limiting role in the kinetic unfolding reaction. WT α TS unfolds through two slow phases whose time constants decrease exponentially as the urea concentration is increased (Figure 2.4a). It has

Table 2.1. Thermodynamic parameters for the urea-induced unfolding of WT α TS and eight variants at 25 °C^a.

α TS variants	$\Delta G^{\circ}_{N \rightleftharpoons I1}(\text{H}_2\text{O})$ (kcal mol ⁻¹)	$-m_{N \rightleftharpoons I1}$ (kcal mol ⁻¹ M ⁻¹)	$\Delta G^{\circ}_{I1 \rightleftharpoons I2/U}(\text{H}_2\text{O})$ (kcal mol ⁻¹)	$-m_{I1 \rightleftharpoons I2/U}$ (kcal mol ⁻¹ M ⁻¹)	$\Delta \Delta G^{\circ}_{N \rightleftharpoons I1}$ (kcal mol ⁻¹) ^b
Wild type	7.19±0.58	2.85±0.24	3.04±0.85	0.81±0.17	-
D46A	1.98±0.45 ^c	0.78±0.17 ^c	4.97±1.96 ^d	1.07±0.39 ^d	-5.21±0.73
D124A	2.53±0.40 ^c	1.12±0.19 ^c	3.81±0.64 ^d	0.79±0.16 ^d	-4.66±0.70
D130A	1.17±0.39	1.19±0.21	2.75±0.24	0.80±0.05	-6.02±0.70
D38A	6.11±0.50	2.93±0.24	4.52±0.69	1.14±0.16	-1.08±0.77
D112A	4.90±0.28	2.34±0.14	3.43±0.41	0.89±0.09	-2.29±0.64
S33A	3.73±0.45	2.11±0.24	3.86±0.60	1.00±0.14	-3.46±0.73
D46N	5.04±0.57	3.23±0.38	2.01±0.32	0.67±0.08	-2.15±0.81
D46E	2.15±0.46 ^c	0.93±0.21 ^c	1.91±1.44 ^{d, e}	0.48±0.06 ^{d, e}	-5.04±0.74

a. Buffer conditions: 10 mM potassium phosphate, 0.2 mM K₂EDTA, and 1 mM β ME at pH 7.8.

b. Perturbation in stability for N \rightleftharpoons I1 reaction, calculated by $\Delta \Delta G^{\circ}_{N \rightleftharpoons I1} = \Delta G^{\circ}_{N \rightleftharpoons I1}(\text{H}_2\text{O}, \text{variant}) - \Delta G^{\circ}_{N \rightleftharpoons I1}(\text{H}_2\text{O}, \text{WT})$.

c. Determined by fitting the urea dependence of the major unfolding kinetic phase to a two-state model.

d. Determined by fitting the equilibrium unfolding data to a three-state model with parameters for the N \rightleftharpoons I1 transition fixed to the values determined as described in footnote c.

e. The native baseline was allowed to float to an optimal value and the unfolded baseline was fixed to the linear dependence on denaturant concentration observed at high urea concentration (Figure 2.7b).

Figure 2.4. Kinetic chevrons for (a) WT α TS, (b) D46A α TS, (c) D124A α TS and (d) D130A α TS, which show the observed refolding (\circ ◆ \triangle □) and unfolding (\blacktriangledown ▽■) relaxation times as a function of final urea concentrations monitored by manual-mixing CD (\circ ▽ \triangle ▽□) at 222 nm and stopped-flow FL (◆■). The solid lines in panels b-d represent the behavior of the WT protein. Refolding burst-phase amplitude at 222 nm (▲) by stopped-flow CD for (e) the WT α TS, (f) D46A α TS, (g) D124A α TS and (h) D130A α TS. The equilibrium unfolding curves (●) for each protein are shown for reference in panels e-h, and the continuous lines indicate the fit of each data set to a three-state model. The dashed and dotted lines in panels f-h represent the equilibrium unfolding transition and the dependence of the burst-phase amplitude on urea concentration, respectively, for the WT protein.

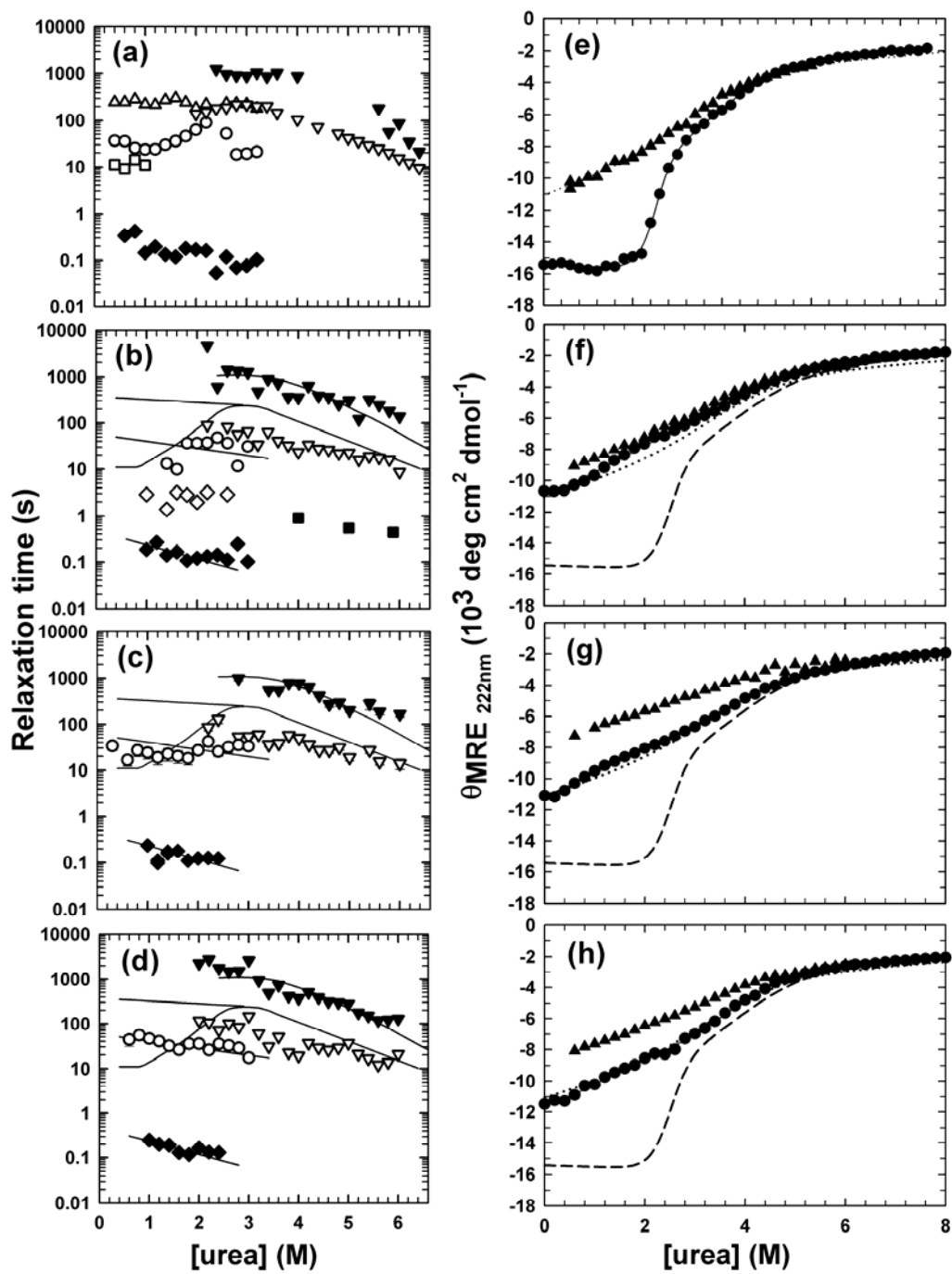


Figure 2.4

previously been shown that these two exponential phases correspond to the independent unfolding of a major and a minor native conformer (Bilsel et al. 1999b). These conformers reflect the native *cis*, 93%, and non-native *trans*, 7%, isomers for the D27-P28 peptide bond (Wu and Matthews 2002a). The subsequent unfolding reactions of the corresponding on-pathway I1 intermediates are far faster, rendering the pair of $N \rightleftharpoons I1$ reactions as the rate-limiting steps in unfolding for α TS. If the α TS variants, in the absence of denaturant, occupy a stable thermodynamic state related to the native state of the WT protein, they might be expected to display a vestige of the rate-limiting unfolding reactions.

Similar to the WT protein, the unfolding reactions for the D46A, D124A and D130A variants each exhibits a major fast and minor slow phase whose relaxation times accelerate exponentially with increasing urea concentration and are nearly coincident with those from the WT protein (Figure 2.4a-d). Thus, in the absence of denaturant, D46A, D124A and D130A α TS all occupy distinct thermodynamic states that retain the major *cis* and minor *trans* isomers at D27-P28 and the rate-limiting barriers to unfolding experienced by WT α TS.

The refolding of WT α TS detected by stopped-flow CD (SF-CD) and stopped-flow fluorescence (SF-FL) proceeds through a sub-millisecond burst-phase reaction followed by a hundreds of milliseconds reaction that accelerates as the urea concentration is increased (Figure 2.4a). This “back-tracking” phase reflects the unfolding of a misfolded off-pathway species and controls access to the subsequent on-pathway species, I1. *Cis-trans* isomerization at three key prolines, P28, P217 and P261, ultimately limits

folding to the major and minor native species evident in the unfolding reaction (Bilsel et al. 1999b; Wu and Matthews 2002a; Wu and Matthews 2002b). These rate-limiting isomerization reactions are the source of the three slow refolding phases in the 10 to 300 s time range for the WT protein under strongly folding conditions (Figure 2.4a).

By contrast, all three variants are missing two of the three slow urea-independent refolding phases observed in the WT protein (Figure 2.4a-d). Only two phases were detected under strongly folding conditions for D124A (Figure 2.4c) and D130A (Figure 2.4d) α TS. The relaxation time of the faster phase is weakly dependent on the final urea concentration and coincident with the back-tracking phase in the WT protein. The relaxation time of the slower refolding phase was nearly independent of the urea concentration and in the time range of phases previously assigned to *cis/trans* isomerization reactions among a set of four I1 intermediates in parallel folding channels (Wu and Matthews 2002a). In addition to these two refolding phases, D46A also displays a refolding phase in the seconds time range that connects smoothly with an additional unfolding reaction. This additional phase has also been detected for the L48A variant (Wu et al. 2007) and is thought to reflect the interconversion of the I1 and I2 states. Finally, the measurement of the amplitude of the sub-millisecond CD signal at 222 nm as a function of the final urea concentration confirmed the presence of a burst-phase intermediate for all three variants (Figure 2.4f-h).

Taken together, the kinetic refolding data demonstrate that all of the clamp-deletion variants retain the sub-millisecond off-pathway misfolded intermediate and one of the proline isomerization reactions between two I1-like intermediates. Because the

pair of unfolding phases for all three variants so closely agree in relaxation time and amplitude with those previously assigned to isomerization at P28, it is reasonable to suppose that the remaining isomerization reaction in refolding for all three clamp-deletion variants reflects the isomerization at P28. By implication, the P217 and P261 *cis/trans* isomerization reactions no longer limit access to the native conformation in the three clamp-deletion variants. The C-terminus of the clamp-deletion variants, including residues P217 and P261, does not appear to be well folded or at least not thermodynamically-coupled to the folded N-terminus in their respective native states.

Thermodynamic parameters extracted from the kinetic experiments

The kinetic unfolding experiment that verified the presence of stable thermodynamic states for the clamp variants can be modified to determine the stability of their folded states. Because the amplitudes of the unfolding reactions are directly proportional to the populations of the folded states from which these reactions initiate, the stability of the folded states can be measured by monitoring the amplitudes of the unfolding reactions (to constant urea concentration) as a function of the initial urea concentration. The amplitudes of the slow unfolding phases are expected to decrease in a sigmoidal fashion as the initial urea concentration is increased into the zone where the native states become depopulated.

This procedure was validated on WT α TS (Figure 2.5a), where a fit of the sigmoidal decrease in ellipticity at increasing initial urea concentrations to a two-state model yielded a stability of 7.06 ± 0.59 kcal mol⁻¹. This value is in excellent agreement

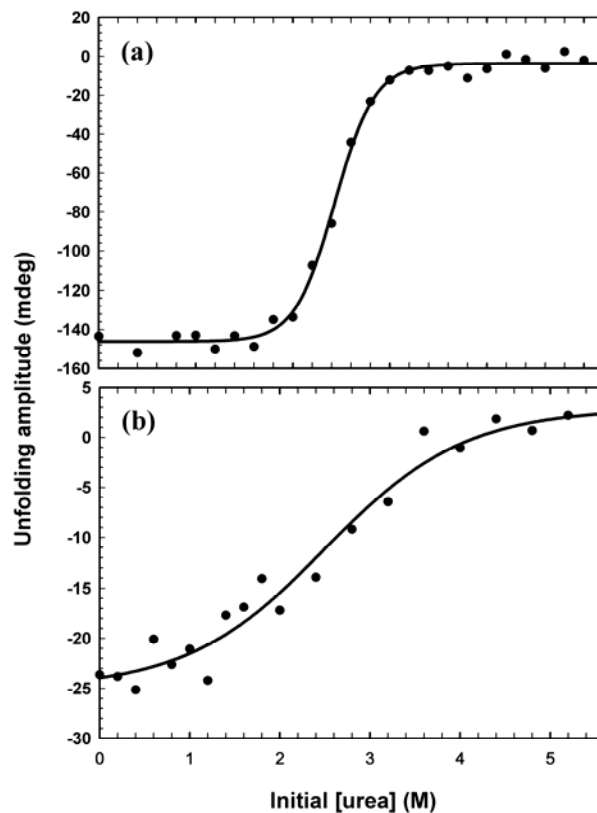


Figure 2.5. The dependence of the amplitude for the major unfolding phase for (a) WT α TS and (b) D46A α TS on the initial urea concentration; the final urea concentration in all cases was 6 M urea. The solid lines represent the fit of the data to a two-state model with $\Delta G^\circ = 7.06 \pm 0.59 \text{ kcal mol}^{-1}$ and $m = 2.89 \pm 0.24 \text{ kcal mol}^{-1} \text{ M}^{-1}$ for WT α TS and $\Delta G^\circ = 1.98 \pm 0.45 \text{ kcal mol}^{-1}$ and $m = 0.78 \pm 0.17 \text{ kcal mol}^{-1} \text{ M}^{-1}$ for D46A α TS. The protein concentration ranged from 5 to 6 μM .

with that for the $N \rightleftharpoons I1$ transition from the standard equilibrium titration experiment on the WT protein, $7.19 \pm 0.58 \text{ kcal mol}^{-1}$ (Figure 2.3 and Table 2.1). The urea dependence of the stability, the m value, is also in excellent agreement, 2.90 ± 0.24 vs $2.85 \pm 0.24 \text{ kcal mol}^{-1} \text{ M}^{-1}$ (Table 2.1). Confirmation of the kinetic measurement of stability was provided by the D130 α TS variant. The traditional equilibrium titration data (Figure 2.3) yielded a stability of $1.17 \pm 0.39 \text{ kcal mol}^{-1}$ and an m value $1.19 \pm 0.21 \text{ kcal mol}^{-1} \text{ M}^{-1}$ (Table 2.1). The kinetic measurement yielded a stability of $1.05 \pm 0.17 \text{ kcal mol}^{-1}$ and an m value $1.25 \pm 0.16 \text{ kcal mol}^{-1} \text{ M}^{-1}$ (Table 2.1).

The stabilities of the folded states of the D46A and D124A α TS clamp-deletion variants were then determined in a similar fashion. Figure 2.5b displays a representative example of the amplitude of the major unfolding phase for D46A α TS as a function of the initial urea concentration. The stability of the folded state for D46A α TS, $1.98 \pm 0.45 \text{ kcal mol}^{-1}$, is $5.21 \text{ kcal mol}^{-1}$ less than that of the N state for the WT protein; the m value is also significantly reduced, 0.78 ± 0.17 vs $2.85 \pm 0.24 \text{ kcal mol}^{-1} \text{ M}^{-1}$ (Table 2.1). The stability of the folded state for D124A α TS is decreased to $2.53 \pm 0.40 \text{ kcal mol}^{-1}$. The m value for the $N \rightleftharpoons I1$ transition of D124A α TS variant is also significantly decreased (Table 2.1).

Although the absence of a prominent shoulder in the MRE titration data for the D46A (Figure 2.4f) and D124A (Figure 2.4g) variants, leaves the existence of an $I1$ -like state ambiguous, the folded states for both proteins, as measured by the kinetic amplitude experiment, disappear, e.g., at 3-4 M urea for D46A α TS (Figure 2.5b), before the unfolded state is completely populated, e.g., at 5-6 M urea for D46A α TS (Figure 2.3).

This discrepancy in the ellipticities implies the presence of a stable intermediate between 3 and 6 M urea. The stabilities of the intermediates for D46A and D124A α TS can be estimated by fitting the unfolding data for these clamp-deletion variants (Figure 2.3) to a three-state $N \rightleftharpoons I1 \rightleftharpoons I2/U$ model, for which the parameters for the $N \rightleftharpoons I1$ reaction are fixed to those obtained from the kinetic unfolding experiments. The thermodynamic parameters obtained for the $I1 \rightleftharpoons I2/U$ transition are shown in Table 2.1.

The stabilities of the I1 states for all three clamp-deletion variants (relative to their I2/U states) and the m values for the $I1 \rightleftharpoons I2/U$ transition are within error of the values for the I1 state relative to the I2/U state for WT α TS. Because the removal of the clamp interaction by replacing aspartic acid with alanine has no significant effect on the stabilities of the intermediate states, it appears that each of the clamp interactions is broken in the intermediate of the WT protein.

Uniqueness of the clamp interactions

All three clamp interactions are located in the N-terminal half of α TS, which is known to serve as a core of stability in both the N and I1 states (Zitzewitz and Matthews 1999; Rojsajakul et al. 2004). To determine whether the surprisingly large perturbations of structure and stability are related to the participation of these clamp H-bonds in this stable core, the S33A, D38A, and D112A variants of α TS were constructed. S33 forms an H-bond with the main chain amide nitrogen of G26, D38 forms an electrostatic interaction with H92, and D112 forms an electrostatic interaction with H146. The heavy atom distances for these three interactions are 3.00, 2.75, and 2.61 Å, respectively,

comparable to 2.92, 2.61, and 2.85 Å for the F19-D46, I97-D124, and A103-D130 clamp H-bonds (Figure 2.1b). The solvent accessible surface areas for the G26-S33, D38-H92, and D112-H146 interactions are 16, 89, and 90 Å², also comparable to those for the F19-D46, I97-D124, and A103-D130 clamp H-bonds, 28, 19, and 80 Å², respectively.

The D38A and D112A variants display a similar reduction in the far-UV CD signal as the clamp-deletion variants (Figure 2.6a and 2.6b). However, both retain the distinctive near-UV CD spectrum for the WT protein (Figure 2.6d and 2.6e), which suggests that they retain the similar tertiary structure as WT protein. The reductions in the far-UV CD signal, therefore, likely reflect the local unfolding of the surface helices and loops linked by these salt bridges. The far-UV (Figure 2.6c) and near-UV CD (Figure 2.6f) spectra of the S33A variant are both very similar to those for WT αTS. The unfolding transition curves for all three variants reveal native baselines and the biphasic behavior characteristic of a three-state process (Figure 2.7a). Fits of these data to a three-state model show that the perturbations on stability for D38A, D112A, and S33A αTS are less than those of their clamp-deletion counterparts for the N \rightleftharpoons I1 transition and negligible for the I1 \rightleftharpoons I2/U transition (Table 2.1). Thus, the effects on the structure and/or stability of the native state accompanying the loss of the βα or αβ clamps in D46A, D124A, and D130A αTS are quantitatively different than those of other side chain-main chain or side chain-side chain electrostatic interactions in the N-terminal folding unit of αTS.

Uniqueness of aspartic acid as an H-bond acceptor in the clamp interaction

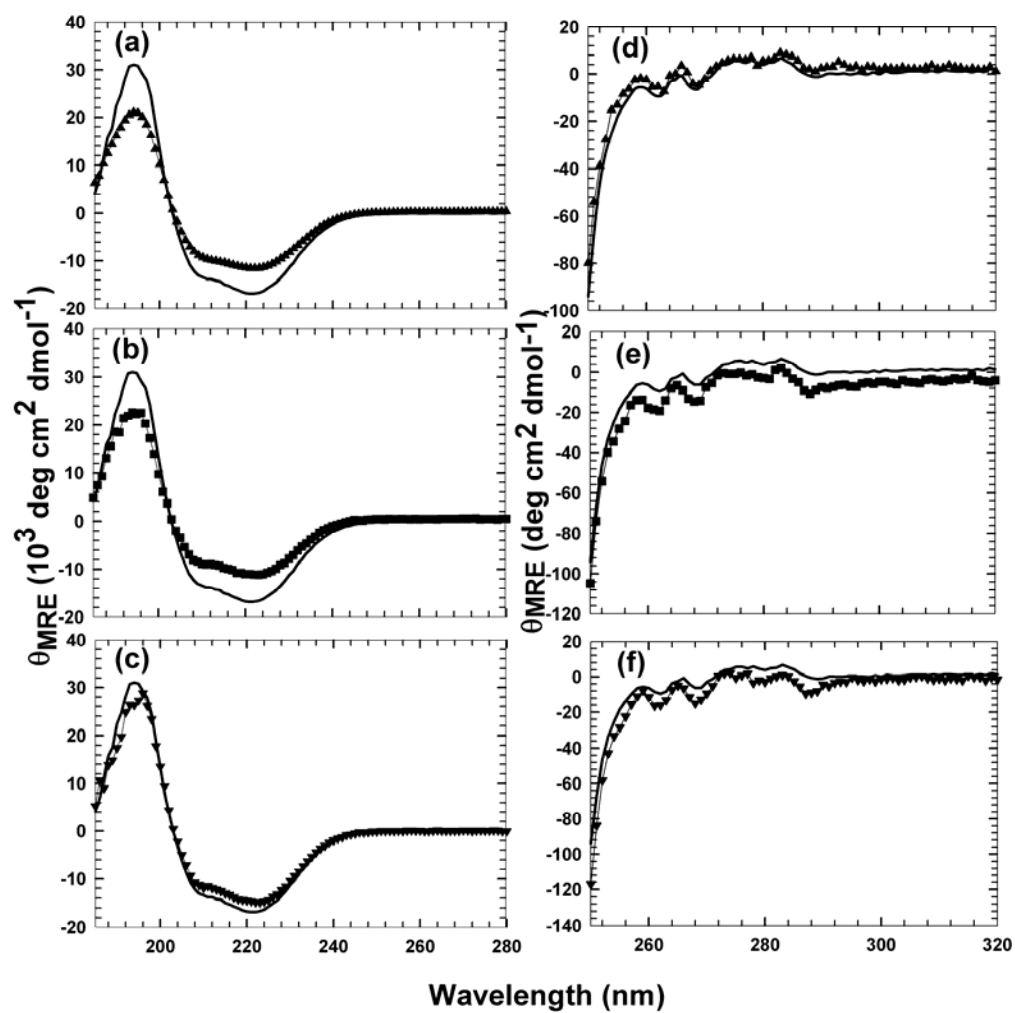


Figure 2.6. Far-UV (a)-(c) and near-UV (d)-(f) CD spectra of three clamp-deletion α TS variants: (a) and (d) D38A α TS (▲), (b) and (e) D112A α TS (■), and (c) and (f) S33A α TS (▼). The solid line in each panel represents the behavior of the WT α TS for reference.

Figure 2.7. Urea denaturation equilibrium unfolding curves of (a) D38A α TS (\blacktriangle), D112A α TS (\blacksquare) and S33A α TS (\blacktriangledown), and (b) D46A α TS (\triangle), D46N α TS (\blacklozenge) and D46E α TS (\circ). The data for WT α TS (\bullet) are shown for comparison in each panel. The continuous lines represent fits of the data for each α TS variant to a 3-state equilibrium folding model as described in the text. The protein concentration was $\sim 5 \mu\text{M}$, and the buffer contained 10 mM potassium phosphate, pH 7.8, 0.2 mM K_2EDTA , and 1 mM β ME at 25 $^\circ\text{C}$.

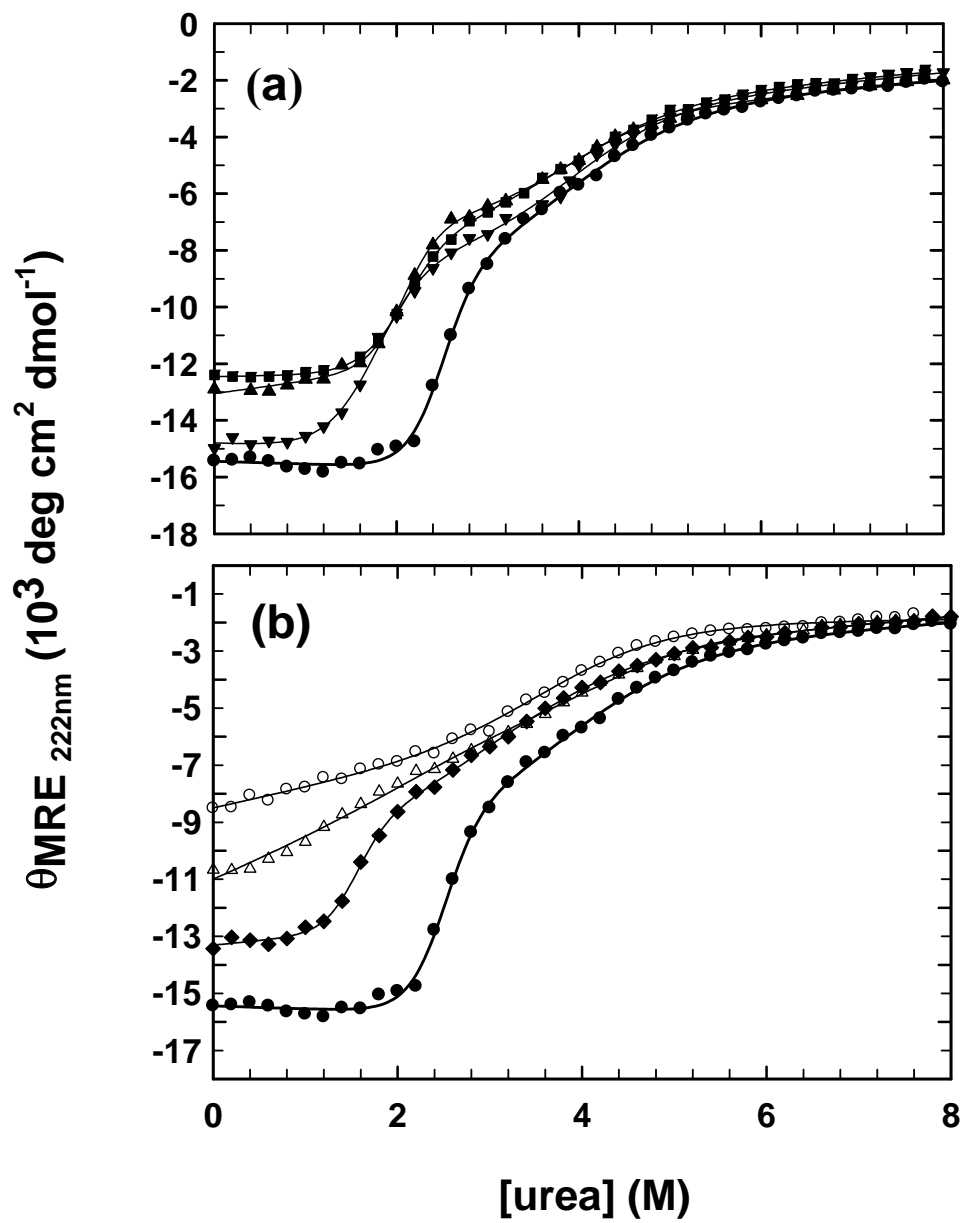


Figure 2.7

The significance of the formal negative charge and the size of the aspartic acid side chain for the clamp interaction were examined by replacing the aspartic acid at position 46 with asparagine or glutamic acid. The asparagine side chain has a very similar size but lacks the formal charge, and glutamic acid retains the charge but has an additional methylene in the side chain. The D46N α TS variant experiences a 20% reduction in the far-UV CD signal and retains the three-state equilibrium unfolding profile (Figure 2.7b). The free energy change for the N \rightleftharpoons I1 reaction is 5.04 ± 0.57 kcal mol⁻¹, only 2.15 kcal mol⁻¹ less than that of WT α TS (Table 2.1). The absence of a significant perturbation in the free energy change for the I1 \rightleftharpoons I2/U reaction is consistent with the results for the D46A variant and the disruption of the clamp H-bond in the I1 state for WT α TS. Kinetic analysis showed that D46N α TS retains the two unfolding phases observed for the WT protein; however, two of the three proline isomerization-limited refolding reactions are absent. Thus, the isosteric replacement of aspartic acid with asparagine at position 46 in the N-terminal domain appears to preclude the proper folding and/or docking of the C-terminal region containing P217 and P261.

The ellipticity at 222 nm for the D46E α TS variant is smaller than that for any other variant examined, including that for D46A α TS (Figure 2.7b). D46E α TS retains the pair of slow unfolding reactions characteristic of the native conformation of α TS, and the kinetic measurement of the free energy change for the N \rightleftharpoons I1 transition is also greatly destabilized, 2.15 ± 0.46 kcal mol⁻¹ (Table 2.1). Similar to D46A and D46N, the free energy difference for the I1 \rightleftharpoons I2/U reaction is within error of that for WT α TS. Neither

asparagine nor glutamic acid can completely recapitulate the spectroscopic, thermodynamic and kinetic folding properties of aspartic acid at position 46.

DISCUSSION

The role of clamp interactions in the structure and stability of α TS

Elimination of any one of the three $\alpha\beta$ or $\beta\alpha$ clamp interactions in α TS, F19-D46, I97-D124 or A103-D130, results in surprisingly large decreases in the stability of the native state relative to the equilibrium intermediate, I1. Equally surprising was the appearance of what appear to be closely-related partially-folded states with substantially diminished secondary structure and altered tertiary structure for all three clamp-deletion variants. This tentative conclusion is based on (1) virtually identical far-UV CD spectra for all three clamp-deletion variants, and identical near-UV CD spectra for the D46A and D124A α TS variants, (2) similar reductions in the stability of the N state, 4.66-6.02 kcal mol⁻¹, and (3) a pair of unfolding reactions with identical relaxation times and relative amplitudes for all three proteins. The contrasting near-UV CD spectrum for D130A α TS (Figure 2.2g) may reflect the peculiar and localized behavior of Y102, adjacent to the peptide bond containing the amide hydrogen of A103 with which D130 forms an H-bond. The observation of a common urea-independent slow refolding reaction, absent the prolyl isomerization reactions for P217 and P261, implies that the C-terminal region of α TS cannot properly fold and/or dock on the N-terminal domain in the native states of all three clamp-deletion variants.

Other variants in which either aspartic acid, which serves as the anionic partner in surface electrostatic interactions, D38-H92 and D112-H146, or serine, which serves as an H-bond acceptor for a nearby main chain amide nitrogen donor, G26-S33, is replaced with alanine show smaller decreases in stability and lesser effects on the secondary

and/or tertiary structure. The comparable range of H-bond lengths and solvent accessible surface areas with the clamp-deletion variants suggest that other factors, possibly the concerted disruption of all three clamp H-bonds (see below), are responsible for the dramatically different effects. The observation that neither asparagine nor glutamic acid can completely recover the structural, thermodynamic or kinetic folding properties of WT α TS implies, at least for this position, that both the formal charge and the precise positioning of that charge are crucial for full acquisition of the TIM barrel structure and stability.

The role of the clamp interactions in the folding mechanism of α TS

The retention of the burst-phase ellipticity and the hundreds of milliseconds unfolding reaction under refolding conditions, i.e., the back-tracking of the off-pathway burst-phase intermediate to the productive folding pathway, for all three clamp-deletion variants (Figure 2.4f-h) demonstrate that none of the clamps are essential for the misfolding reaction. The observation of a single urea-independent subsequent refolding reaction attributed to the *trans* \rightarrow *cis* isomerization of the peptide main chain at D27-P28, shows that the prolyl isomerizations at P217 and P261 are no longer rate limiting in folding for any of the clamp-deletion variants. By contrast, the retention of the *cis* isomer for P28 in the native state of all three variants implies that the tight turn between β 1 and α 1, responsible for stabilizing this higher energy form, is preserved in the absence of the clamps.

These results can be explained by examination of the properties of a stable C-terminal truncated version of α TS, 1-188 α TS (Zitzewitz and Matthews 1999). 1-188 α TS retains the burst-phase reaction, the early back-tracking reaction, a single slow folding reaction and two slow unfolding reactions. All but the unmeasured burst-phase reaction have relaxation times similar to full-length α TS. Although the secondary and tertiary structures of the clamp-deletion variants are somewhat disrupted compared to the WT 1-188 fragment (with the clamp H-bonds), the similarities of the kinetic folding properties of 1-188 α TS (Zitzewitz and Matthews 1999) and the clamp-deletion variants suggest that the C-terminal region of each of the clamp-deletion variants is not well folded under native-favoring conditions. Apparently, the loss of any of the clamp interactions in the N-terminus alters the structure of the N-terminus so as to preclude the folding and docking of the C-terminus. An earlier fragmentation analysis of α TS (Higgins et al. 1979a) showed that the C-terminal region, residues 189-268, is poorly folded in isolation. The spontaneous recovery of structure and function for α TS when these two fragments are complemented showed that the N-terminal region serves as the template for the folding and docking of the C-terminal region. If the template structure is altered, as appears to be the case for all three clamp-deletion variants, the C-terminal region cannot fold and dock on the N-terminal region.

It is also interesting to note that the two slow unfolding relaxation times exhibited by all three variants are within error of those for WT α TS at high denaturant concentration. Given the substantial loss in the stability of their respective native states, the similar destabilization of the unfolding transition states for all three variants means

that all three clamp interactions in the WT α TS must be maintained in both unfolding transition states. Because the $N \rightleftharpoons I1$ reaction is reversible, these clamps must, therefore, also be very native-like in the transition states for the refolding reactions. The absence of the clamps in the I1 state, as implied by the lack of perturbation in the nature of $\Delta G^\circ_{I1 \rightleftharpoons I2/U}$ (H_2O) when aspartic acid is replaced by alanine (Table 2.1), implies that the clamps first appear in the final, rate-limiting transition state for folding, after the $U \rightarrow I2$ and $I2 \rightarrow I1$ reactions are complete. By acting in concert to stabilize the transition state of WT α TS, the clamps direct the folding to the fully-formed TIM barrel structure rather than to the partially-folded states observed for the clamp variants. The stabilization of the N-terminus of the native state for WT α TS by the clamps presumably enhances the folding of the C-terminus by providing a properly-structured template on which the C-terminus can fold.

The closely-related partially folded structures implied by the nearly identical far-UV CD spectra (Figure 2.2d) and unfolding rates (Figure 2.4b-d) for the clamp-deletion variants suggest that the loss of a single clamp results in the rupture of the two remaining clamps. Thus, the unusually large decrease in stability observed for each clamp-deletion variant (Table 2.1) does not reflect the strength of a single side chain-main chain H-bond. This conjecture is supported by the substantial perturbations in the far- and near-UV CD spectra that imply the loss of numerous non-covalent interactions in these variants.

Structural implication of strong protection against HX for F19 NH and I97 NH

The absence of the F19-D46 and I97-D124 clamp interactions in the stable I1 folding intermediate for α TS (Table 2.1) begs the question as to the source of the unusually strong protection of these amide hydrogens against HX with solvent. If I1 were that source, one would have expected exchange at both of these positions in a few hours at 25 °C and pH 7.8 via an EX1 mechanism (Englander and Kallenbach 1983; Bilsel et al. 1999b). The implication of protection sufficient to retard exchange for days and months is either that these amide hydrogens have an alternative H-bond acceptor partner and/or are buried in non-native nonpolar environments that inhibit exchange with solvent in the I1 state. Exchange must occur through other, higher energy states on the folding free energy surface, e.g., the I2 state that lacks discernable secondary structure. Either implied explanation means that the N-termini of at least two β -strands, β 1 and β 3, are involved in a very stable but non-native structure(s) that differentiate the I1 and the N state. It is interesting to speculate that both aspartic acid H-bond donors are replaced by carbonyl oxygens in adjacent strands, i.e., β 2 and β 4, whose lengths are extended to enable these additional stabilizing interactions. Whatever the source of these non-native stabilizing interactions, they must be disrupted to enable formation of the clamp H-bonds required to access the transition state leading to the fully-folded TIM barrel.

Clamp interactions define the boundaries of $\beta\alpha$ and $\alpha\beta$ hairpins

The clamp interactions probed in this analysis are analogous to helix-capping interactions previously identified by Rose (Presta and Rose 1988). H-bond acceptor side

chains near the N-termini of helices can form H-bonds with unsatisfied main chain amide hydrogen donors within but near the N-termini of the helices. These H-bonds are typically sheltered from solvent by a pair of nonpolar side chains that presumably serve to increase the strength of the H-bond. Inspection of the local environments of each of the $\beta\alpha$ and $\alpha\beta$ clamp interactions in α TS shows hydrophobic clusters in those regions: A103 has van der Waals contacts with V128 and V131, I97 has contacts with L85 and I88, and F19 has contacts with V259 and M262. As with the helix-capping interactions, screening of the solvent by hydrophobic clusters would be expected to strengthen the clamp H-bond.

In addition to enhancing the stability of the native conformation, helix-capping interactions are thought to limit the length of the helices and, thereby, promote the formation of loops and turns that link the helices to neighboring elements of secondary structure. The clamp interactions may serve a similar role for $\beta\alpha$ and $\alpha\beta$ hairpins by delineating and reinforcing the boundaries of its constituent β strand and α helix. The specificity conferred by the formation of a clamp interaction would ensure the proper register of the strand and helix forming the hairpin. The identification of these long-range clamp interactions in the amino acid sequence of TIM barrel proteins of unknown structure could, therefore, assist in the prediction of their structures from the sequence. Nonpolar side chains at the N-termini of predicted β strands or α helices that precede polar H-bond acceptor side chains at the C-termini of subsequent predicted helices or strands would be candidates for the clamp H-bonds.

CHAPTER III

$\beta\alpha$ -HAIRPIN CLAMPS HIGHLIGHT THE $\beta\alpha\beta$ MODULE AS A FUNDAMENTAL BUILDING BLOCK OF TIM BARREL PROTEINS

ABSTRACT

Long-range hydrogen bonds between main chain amide hydrogen atoms and side chain acceptors that bracket consecutive $\beta\alpha$ or $\alpha\beta$ elements of secondary structure in the $(\beta\alpha)_8$, TIM barrel protein, α TS, have previously been found to make substantial contributions to structure and stability in the native conformation. Experimental analysis of $\beta\alpha$ -hairpin clamps in a homologous pair of TIM barrel proteins of low sequence identity, indole-3-glycerol phosphate synthase from *S. solfataricus* and *E. coli*, reveals that this phenomenon is not unique to α TS. A survey of 71 TIM barrel proteins demonstrates a pseudo 4-fold symmetry for the placement of $\beta\alpha$ -hairpin clamps, bracing the fundamental $\beta\alpha\beta$ building block and reinforcing the shear for the $(\beta\alpha)_8$ motif. The observed sequence preferences and locations of $\beta\alpha$ -hairpin clamps will enhance structure prediction algorithms and provide a strategy for engineering stability in TIM barrel proteins.

INTRODUCTION

The $\beta\alpha$ -repeat class of proteins includes the $(\beta\alpha)_8$, TIM, barrel structure and the $(\beta\alpha)_n$ Rossmann-fold. These two super-families are among the most common folds in all three super-kingdoms of life, undoubtedly reflecting gene duplication of the elemental $\beta\alpha$ repeat building block as higher-order structure and, eventually, numerous functions evolved. Experimental (Zitzewitz et al. 1999) and bioinformatics (Gerstein 1997; Nagano et al. 2002) analyses of the TIM barrel architecture have led to the conclusion that a pair of adjacent parallel β -strands and the intervening anti-parallel α -helix, i.e., the $\beta\alpha\beta$ module, serves as the minimal unit of stability in the TIM barrel. The interactions stabilizing this super-secondary structure include: (1) main chain-main chain H-bonds between the parallel β -strands, (2) intra-helical main chain-main chain H-bonds, (3) hydrophobic interactions between the side chains protruding from the β -strands and the α -helix, (4) side chain-side chain H-bonds and salt bridges, (5) dipole-dipole interaction between the α -helix and the two β -strands on which it docks (FarzadFard et al. 2008) and (6) side chain-main chain H-bonds (Yang et al. 2007). The role of a fascinating subset of the latter interactions in stabilizing TIM barrel proteins (Yang et al. 2007) is the subject of this communication.

A majority of side chain-main chain interactions in proteins are local in sequence, usually within six residues (Baker and Hubbard 1984; Stickle et al. 1992), and can cap both the N- and the C-termini of α -helices (Presta and Rose 1988; Aurora and Rose 1998). Mutational analysis has shown that these non-covalent interactions usually

contribute modestly to the stability of their resident proteins, typically in the range of 1-2 kcal mol⁻¹ (Horovitz et al. 1990; Serrano et al. 1992; Myers and Pace 1996; Ibarra-Molero et al. 2004). Surprisingly, three long-range side chain-main chain H-bonds have been found to be critical to the formation of the fully-folded alpha subunit of tryptophan synthase (α TS) TIM barrel and each contribute 3-5 kcal mol⁻¹ to its stability. An intriguing feature of these interactions is that the main chain amide H-bond donor and the side chain H-bond acceptor for each of the three H-bond pairs is precisely 27 residues apart in sequence and connects the N-terminus of one element of secondary structure, either β -strand or α -helix, to the C-terminus of the subsequent element of structure, either α -helix or β -strand, respectively. These structures have been designated as $\beta\alpha$ -hairpin clamps and $\alpha\beta$ -hairpin clamps (Yang et al. 2007). While this sequence separation may be expected from the average length of a $\beta\alpha$ -repeat element in TIM barrel proteins (Frenkel and Trifonov 2005), the observations that these side chain-main chain H-bonds span consecutive elements of secondary structure and can make significant contributions to structure and stability raise the possibility that these $\beta\alpha$ - and $\alpha\beta$ -hairpin clamps are an important general feature of the TIM barrel architecture.

A two-pronged approach was taken to probe the significance of $\beta\alpha$ -hairpin clamps to TIM barrels. First, mutational analysis of two other representative TIM barrel proteins, indole-3-glycerol phosphate synthase (IGPS) from thermo-acidophilic archaeobacterium *S. solfataricus* (sIGPS) and the mesophilic *E. coli* (eIGPS) assessed the roles of several conserved $\beta\alpha$ -hairpin clamps in structure and stability. Second, a survey of 71 TIM barrel proteins explored the frequency, location and sequence preferences of

all $\beta\alpha$ -hairpin clamps. The results show that side chain-main chain H-bonds are extremely common in TIM barrel proteins and that a subset of $\beta\alpha$ -hairpin clamps may be crucial to protein structure and stability. In combination, the experimental and survey results imply that $\beta\alpha$ -hairpin clamps might have been a critical stabilizing feature of the elemental $\beta\alpha\beta$ module from which TIM barrels evolved.

RESULTS

Experimental analysis of $\beta\alpha$ -hairpin clamp H-bonds in two TIM barrel proteins

The generality of the potent hairpin clamps found in the α TS TIM barrel was tested by a mutational analysis of $\beta\alpha$ -hairpin clamps in two homologous TIM barrel proteins with low sequence identity (<30%) to the α TS TIM barrel and to each other. sIGPS and eIGPS, each contain three $\beta\alpha$ -hairpin clamps, some of which are conserved with those in α TS and some of which are conserved with each other. Figure 3.1 and 3.2 display the elements of secondary structure framed by the clamps, the side chains corresponding to the amide H-bond donors and the side chain H-bond acceptors observed in sIGPS and eIGPS. The $\beta 1\alpha 1$ clamp is conserved in α TS and eIGPS, the $\beta 2\alpha 2$ clamp only appears in sIGPS, the $\beta 3\alpha 3$ clamp is conserved in all three proteins, and the $\beta 7\alpha 7$ clamp is conserved in sIGPS and eIGPS.

Perturbation of the secondary and tertiary structure by clamp deletion in sIGPS and eIGPS.

The contribution of each side chain-main chain H-bond to the structure of the sIGPS and eIGPS TIM barrels was assessed by replacing the H-bond accepting side chain with alanine and monitoring the effects on the secondary and tertiary structure by far-UV and near-UV circular dichroism (CD) spectroscopy. The far-UV CD spectra for the wild-type (WT) and clamp-deletion variants of sIGPS and eIGPS are shown in Figure 3.3a and

Figure 3.1. Long-range side chain-main chain interactions in sIGPS. (a) Ribbon diagram of sIGPS highlighting the $\beta\alpha$ -hairpin clamps. (b) The intervening elements of secondary structures (ribbon) between the residues forming the clamps: $\beta 2\alpha 2$, S104 NH-E74 O ϵ 1; $\beta 3\alpha 3$, I107 NH-D128 O δ 1; and $\beta 7\alpha 7$, K207 NH-N228 O δ 1. The side chains involved in the H-bond clamp are highlighted with the donor and acceptor shown in blue and red, respectively. The H-bonds and the corresponding distances were determined by using the program HBPLUS (McDonald and Thornton 1994). The structures were generated using PyMOL v 0.99 (Delano 2002) and PDB code is 2C3Z (Schneider et al. 2005) for sIGPS.

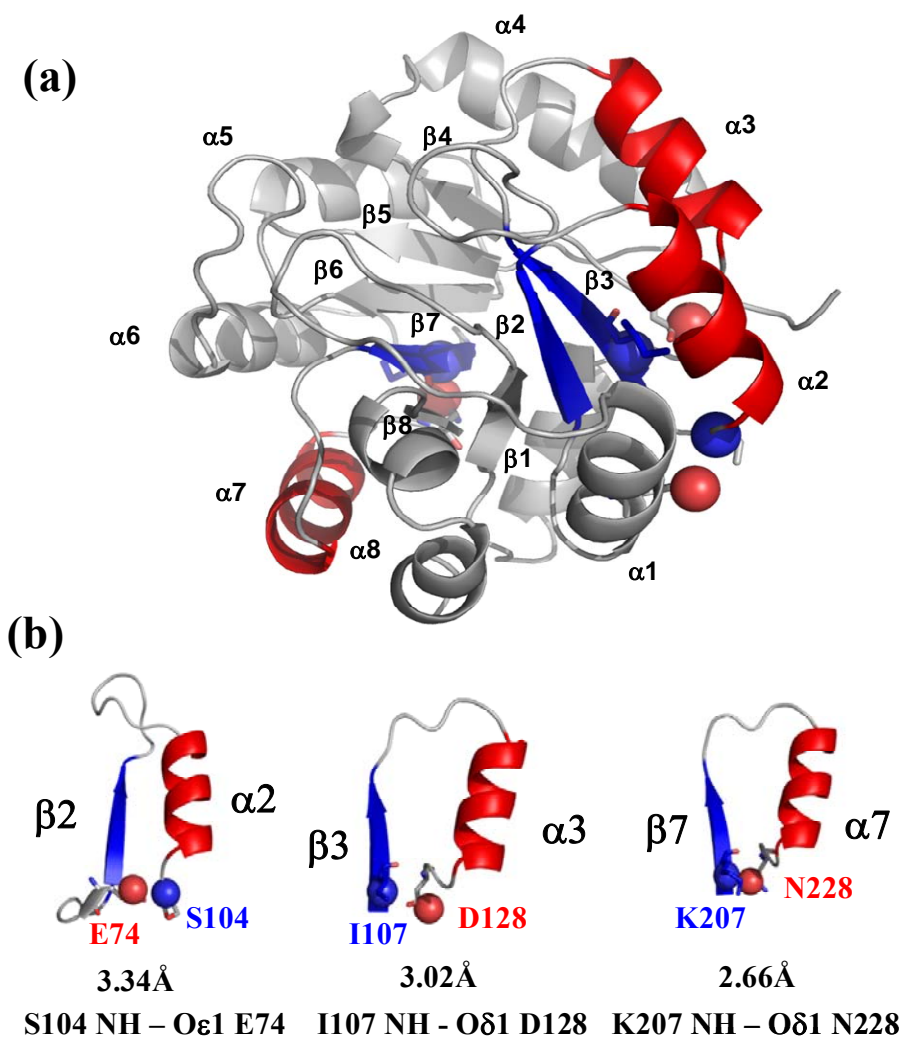


Figure 3.1

Figure 3.2. Long-range side chain-main chain interactions in eIGPS. (a) Ribbon diagram of eIGPS highlighting the $\beta\alpha$ -hairpin clamps. (b) The intervening elements of secondary structures (ribbon) between the residues forming the clamps: $\beta 1\alpha 1$, F50 NH-S82 O γ ; $\beta 3\alpha 3$, I111 NH-D132 O $\delta 2$; and $\beta 7\alpha 7$, V211 NH-N231 O $\delta 1$. The side chains involved in the H-bond clamp are highlighted with the donor and acceptor shown in blue and red, respectively. The H-bonds and the corresponding distances were determined by using the program HBPLUS (McDonald and Thornton 1994). The structures were generated using PyMOL v 0.99 (Delano 2002) and PDB code is 1PII (Wilmanns et al. 1992) for eIGPS.

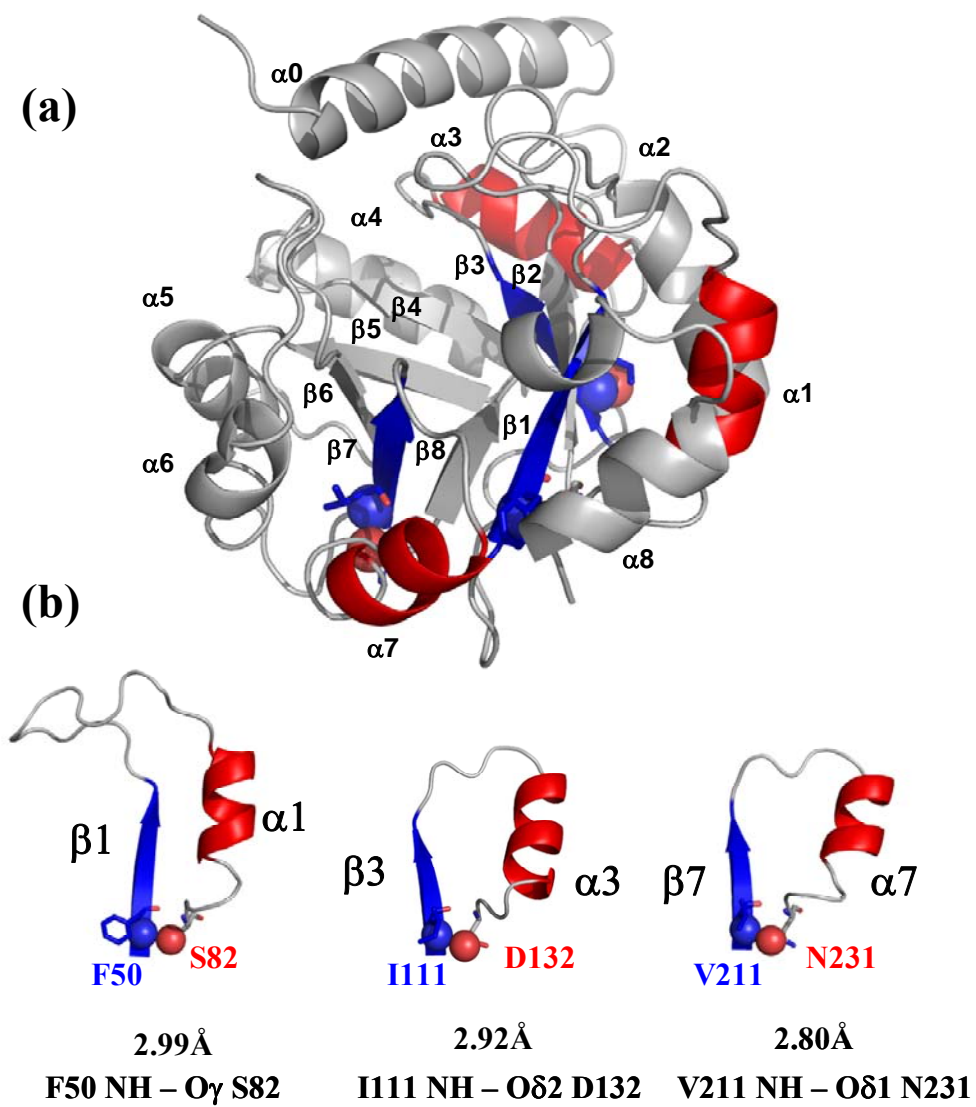


Figure 3.2

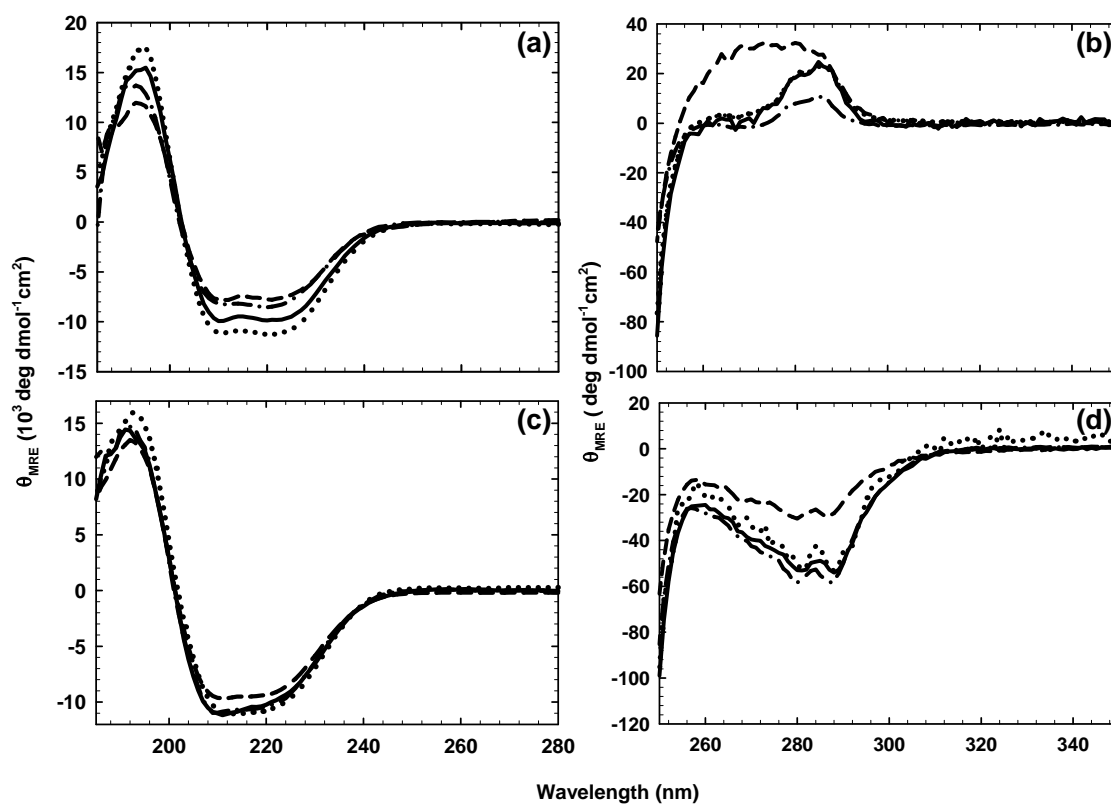


Figure 3.3. Far-UV (a, c) and near-UV (b, d) CD spectra of the WT (solid line) and IGPS clamp-deletion variants. sIGPS (a) and (b): E74A (dashed-dotted line), D128A (dotted line), and N228A (dashed line). eIGPS (c) and (d): S82A (dashed-dotted line), D132A (dotted line), and N231A (dashed line). The buffer contained 10 mM potassium phosphate, 0.2 mM K_2EDTA , and 1 mM βME at pH 7.8 for sIGPS and pH 7.0 for eIGPS, at 25 °C.

3.3c, and the near-UV CD spectra are shown in Figure 3.3b and 3.3d. For sIGPS, the elimination of the $\beta 2\alpha 2$ and $\beta 7\alpha 7$ clamps with the E74A and N228A mutations, respectively, reduces the ellipticity at 222 nm by $\sim 15\%$, while the elimination of the $\beta 3\alpha 3$ clamp with the D128A mutation actually increases the ellipticity by $\sim 15\%$ (Figure 3.3a). The perturbations of the aromatic side chain packing observed in the near-UV CD spectrum (Figure 3.3b) are diverse, with the dichroic bands similar in shape but reduced in magnitude for the elimination of the $\beta 2\alpha 2$ clamp, similar to WT for the elimination of the $\beta 3\alpha 3$ clamp and substantially increased over the wavelength range from 260 to 280 nm for the elimination of the $\beta 7\alpha 7$ clamp. The perturbations of the secondary and tertiary structure for the clamp-deletion variants in eIGPS, S82A ($\beta 1\alpha 1$ clamp deletion), and D132A ($\beta 3\alpha 3$ clamp deletion) are minimal, with the exception of the N231A variant ($\beta 7\alpha 7$ clamp deletion). The decreases in the magnitude of the bands at 222 nm and 284 nm for the N231A variant imply a fraying of the secondary structure and a loosening of the packing near aromatic side chains when the $\beta 7\alpha 7$ clamp is eliminated.

Perturbation of stability by clamp deletion in sIGPS and eIGPS

The effect of $\beta\alpha$ -hairpin clamp deletion on the stability of sIGPS and eIGPS was determined by urea denaturation. As for α TS (Matthews and Crisanti 1981), both sIGPS (Forsyth and Matthews 2002) and eIGPS (Sanchez del Pino and Fersht 1997) unfold via a highly populated intermediate, and their unfolding titration curves are well-fit by a three-state model, $N \rightleftharpoons I \rightleftharpoons U$. With the exception of the N231A variant for eIGPS, the urea-

induced unfolding transition for each of the remaining five clamp-deletion variants was also well-described by this three-state model (Figure 3.4). Because a distinct transition between the native state (N) and the intermediate state (I) was not observed during the urea induced denaturation of eIGPS N231A (Figure 3.4c), kinetic unfolding experiments were performed to verify the existence of I and measure the free energy difference between the N and I states (Yang et al. 2007).

The presence of I in eIGPS N231A was verified by the observation of a slow kinetic unfolding phase where relaxation times decrease with increasing final denaturant concentration (Bilsel et al. 1999b), when eIGPS was subjected to a jump from 0 to 3 M urea where the I state would be expected to be highly populated. Because the amplitude of the unfolding phase is proportional to the population of the N state from which the reaction initiates, the decrease in the amplitude as a function of increasing initial urea concentrations (Figure A.1) can be fit to a two-state model, $N \rightleftharpoons I$, to extract the stability, ΔG°_{NI} , and the urea dependence of stability, m_{NI} . These parameters were used to fit the CD unfolding transition for the eIGPS N231A (Figure 3.4c) and extract ΔG°_{IU} and m_{IU} .

The perturbations of the stabilities of the N and I states for the clamp-deletion variants vs. the parent sequences are illustrated graphically in Figure 3.4b and 3.4d for sIGPS and eIGPS, respectively. The free energy differences between the N and I states, ΔG°_{NI} , and between the I and U states, ΔG°_{IU} , as well as the urea-dependences of these free energies, the m-values, are tabulated in Table 3.1. The elimination of the $\beta 2\alpha 2$ clamp in sIGPS, E74A, only reduces the stability of the N state by 1.08 kcal mol⁻¹, and

Figure 3.4. Equilibrium unfolding properties of IGPS variants. Panels (a) and (c) display urea denaturation unfolding curves of WT (●, solid line) and IGPS clamp-deletion variants. The lines represent fits of the data for each variant to a 3-state equilibrium folding model as described in the text. (a) sIGPS: E74A (▲, dashed-dotted line), D128A (◆, dotted line), and N228A (■, dashed line). (c) eIGPS: S82A (▲, dashed-dotted line), D132A (◆, dotted line), and N231A (■, dashed line). Panels (b) and (d) are bar graphs representing the free energy differences for the N to I step in unfolding, ΔG°_{NI} , (black bars) and the I to U step, ΔG°_{IU} (gray bars) for WT and the variants of (b) sIGPS and (d) eIGPS (Table 3.1).

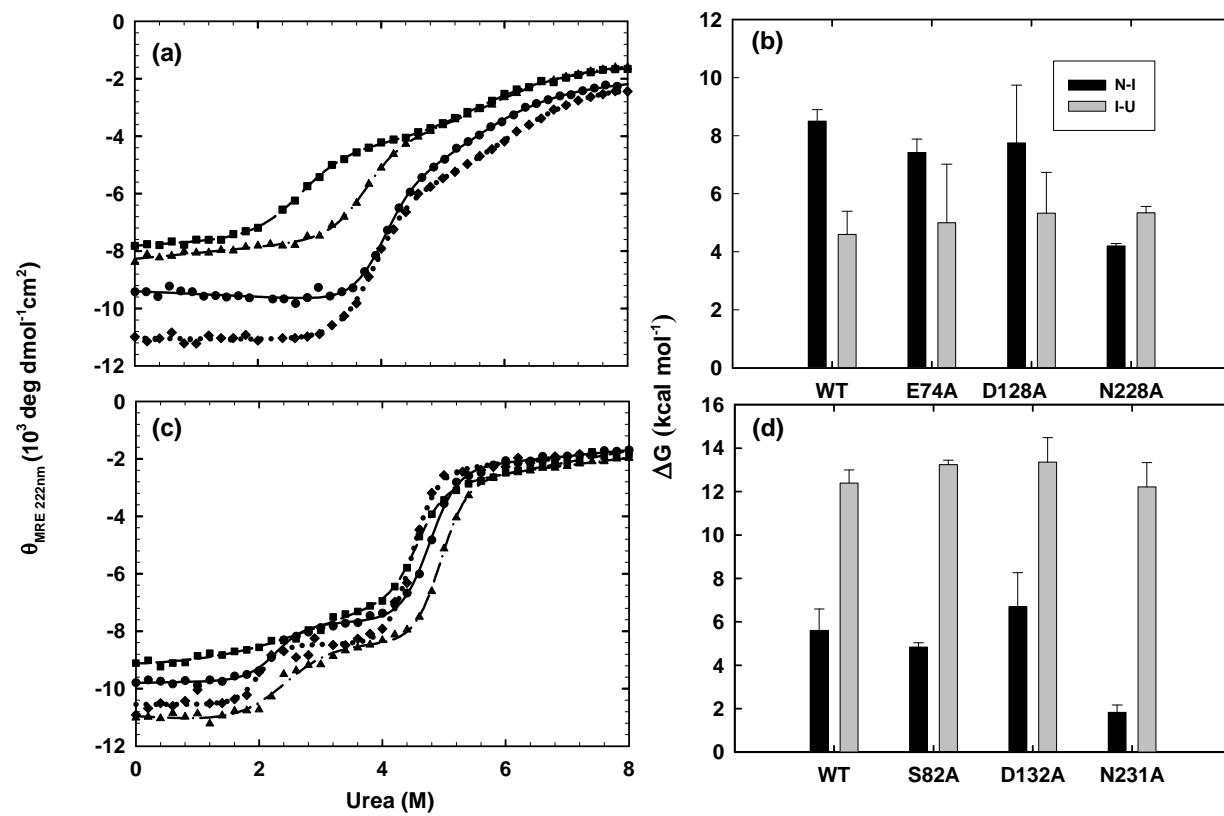


Figure 3.4

Table 3.1. Thermodynamic parameters for the urea-induced unfolding of sIGPS, eIGPS, α TS and nine clamp-deletion variants at 25 °C^a.

	Donor and acceptor pairs	Donor and acceptor distance (Å)	strains	$\Delta G^{\circ}_{\text{NI}}(\text{H}_2\text{O})$ (kcal mol ⁻¹)	$-m_{\text{NI}}$ (kcal mol ⁻¹ M ⁻¹)	$\Delta G^{\circ}_{\text{IU}}(\text{H}_2\text{O})$ (kcal mol ⁻¹)	$-m_{\text{IU}}$ (kcal mol ⁻¹ M ⁻¹)	$\Delta\Delta G^{\circ}_{\text{NI}}$ (kcal mol ⁻¹) ^b
sIGPS			WT	8.50 ± 0.40	2.10 ± 0.10	4.60 ± 0.80	0.86 ± 0.13	-
	S104-E74	3.34	E74A	7.42 ± 0.46	1.97 ± 0.13	5.00 ± 2.02	0.86 ± 0.36	-1.08 ± 0.61
	I107-D128	3.02	D128A	7.75 ± 1.99	1.99 ± 0.12	5.33 ± 1.40	0.89 ± 0.23	-0.75 ± 2.03
	K207-N228	2.66	N228A	4.20 ± 0.08	1.56 ± 0.03	5.34 ± 0.22	0.97 ± 0.04	-4.30 ± 0.41
eIGPS			WT	5.60 ± 0.99	2.46 ± 0.42	12.39 ± 0.60	2.60 ± 0.13	-
	F50-S82	2.99	S82A	4.84 ± 0.19	2.05 ± 0.08	13.24 ± 0.20	2.68 ± 0.04	-0.76 ± 1.01
	I111-D132	2.92	D132A	6.69 ± 1.57	3.34 ± 0.77	13.36 ± 1.12	2.99 ± 0.26	1.09 ± 1.86
	V211-N231	2.80	N231A	1.28 ± 0.15 ^c	0.89 ± 0.11 ^c	11.84 ± 1.45 ^d	2.62 ± 0.31 ^d	-4.32 ± 1.00
α TS ^e			WT	7.19 ± 0.58	2.85 ± 0.24	3.04 ± 0.85	0.81 ± 0.17	-
	F19-D46	2.79	D46A	1.98 ± 0.45	0.78 ± 0.17	4.97 ± 1.96	1.07 ± 0.39	-5.21 ± 0.73
	I97-D124	2.57	D124A	2.53 ± 0.40	1.12 ± 0.19	3.81 ± 0.64	0.79 ± 0.16	-4.66 ± 0.70
	A103-D130	2.75	D130A	1.17 ± 0.39	1.19 ± 0.21	2.75 ± 0.24	0.80 ± 0.05	-6.02 ± 0.70

a. Buffer conditions: 10 mM potassium phosphate, 0.2 mM K₂EDTA, 1 mM β ME at pH 7.8 for sIPGS and pH 7.0 for eIGPS.

b. Perturbation in stability for N to I reaction, calculated by $\Delta\Delta G^{\circ}_{\text{NI}} = \Delta G^{\circ}_{\text{NI}}(\text{H}_2\text{O}, \text{variant}) - \Delta G^{\circ}_{\text{NI}}(\text{H}_2\text{O}, \text{WT})$.

c. Determined by fitting the urea dependence of the amplitude of the unfolding kinetic phase to a two-state model.

d. Determined by fitting the equilibrium unfolding data to a three-state model with parameters for the N to I transition fixed to the values determined as described in footnote c.

e. Values are from Yang *et al* (Yang *et al.* 2007).

the elimination of the $\beta 3\alpha 3$ clamp, D128A, has no significant effect on its stability. In striking contrast, the elimination of the $\beta 7\alpha 7$ clamp, N228A, reduces the stability of the N state by 4.30 kcal mol⁻¹. Consistent with the disruption of the clamps in the $N \rightleftharpoons I$ reaction previously observed for α TS (Yang et al. 2007), the free energy differences between the I and U states for the clamp-deletion variants are very comparable to the value for WT sIGPS (Figure 3.4b). Similar results were obtained for eIGPS (Figure 3.4d). Only the elimination of the $\beta 7\alpha 7$ clamp, N231A, significantly decreased the stability of the N state by 4.32 kcal mol⁻¹.

Deletion of the $\beta 1\alpha 1$ clamp, S82A, and the $\beta 3\alpha 3$ clamp, D132A, in eIGPS had no significant effect on the stability of the N state, and none of the clamp deletions perturbed the stability of the I state relative to the U state. Thus, only the $\beta 7\alpha 7$ clamps in both sIGPS and eIGPS significantly perturb the structure and the stability of their resident TIM barrels. The $\beta 1\alpha 1$, $\beta 2\alpha 2$ and $\beta 3\alpha 3$ clamps have lesser effects on structure and/or stability of the native states.

Survey of $\beta\alpha$ -hairpin clamps in the TIM barrel proteins

The observation that $\beta\alpha$ - and $\alpha\beta$ -hairpin clamps can have a significant effect on structure and stability in three TIM barrel proteins motivated a survey of the prevalence of long-range side chain-main chain clamps in the super-family. The analysis was carried out for a structural database of 71 TIM barrel domains (Gromiha et al. 2004), selected as a non-redundant representation of the TIM barrel fold from the SCOP (Murzin et al. 1995) and HOMSTRAD (Mizuguchi et al. 1998) databases, with a pair-wise sequence

homology of < 25%. The highest resolution structure for each domain was chosen from the Protein Data Bank (Berman et al. 2000). The 35% homology set of proteins from the CATH database (Cuff et al. 2009), CATH-35, was used as a comparative data set.

The long-range side chain-main chain H-bonding interactions were determined using HBPlus program (McDonald and Thornton 1994). A minimum sequence separation of 15 residues was chosen to define the long-range H-bonds in this analysis, thereby eliminating shorter-range helix-capping interactions that have been considered previously (Aurora and Rose 1998). On average, there are 10 Class I H-bonds per protein (2.9/100 residues) and 20 Class II H-bonds (6.8/100 residues) for this group of TIM barrels. For comparison, the proteins in the CATH-35 database contain an average of 3 Class I H-bonds per protein (1.6/100 residues) and 6 Class II H-bonds (3.5/100 residues). It is evident that the TIM barrel motif is an especially favorable target for these non-covalent interactions between the main chains and polar side chains. The sequence separation between the donor and acceptor residues follows the distribution previously observed for the length of closed loops in TIM barrel proteins (Frenkel and Trifonov 2005).

A further classification of the $MC_{NH} \rightarrow SC$ and $SC \rightarrow MC_{C=O}$ H-bonds defines interactions that can clamp consecutive elements of secondary structure as either $\beta\alpha$ - or $\alpha\beta$ -hairpins. Because some of the H-bonds occur between residues that belong to domains other than the TIM architecture or encompass more or less than a pair of consecutive secondary structure elements, only H-bonds between residues that spanned exactly one β -strand and one α -helix, were selected. The additional requirement that

both secondary structure elements be a part of the TIM architecture was also imposed. Each interaction was classified into one of four categories based on the order of the secondary structure elements and the class of the H-bond. In the 71 TIM barrel proteins examined, there are 131 Class I $\beta\alpha$ -hairpin clamps, 163 Class 1 $\alpha\beta$ -hairpin clamps, 286 Class II $\beta\alpha$ -hairpin clamps and 232 Class II $\alpha\beta$ -hairpin clamps. There are also 1,427 long-range side chain-main chain H-bonds that do not fall into these four categories.

Preferences in residue composition and location of Class I $\beta\alpha$ -hairpin clamps

Having established a high frequency of Class I and Class II $\alpha\beta$ and $\beta\alpha$ -hairpin clamps in TIM barrel proteins, the survey then focused on the pair-wise amino acid preferences for the main chain H-bond donors and side chain acceptors for the 131 Class I H-bonds that clamp $\beta\alpha$ -hairpins for comparison with the experimental results. There is a very significant preference, >40% of the clamps, for aspartic acid side chains forming H-bonds with amide hydrogens of isoleucine, leucine and valine residues (Table 3.2). These branched aliphatic side chains are among the most frequently buried amino acids (Chothia 1976; Pace et al. 2004) and, along with alanine and glycine, are the only amino acids that do not partition favorably from the vapor phase to water (Radzicka and Wolfenden 1988). As such, they would be especially effective at excluding water from side chain-main chain H-bonds represented by the Class I $\beta\alpha$ -hairpin clamps. The exclusion of water would be expected to strengthen these H-bonds and make them more resistant to exchange with solvent (Vadrevu et al. 2008). The locations of the entire

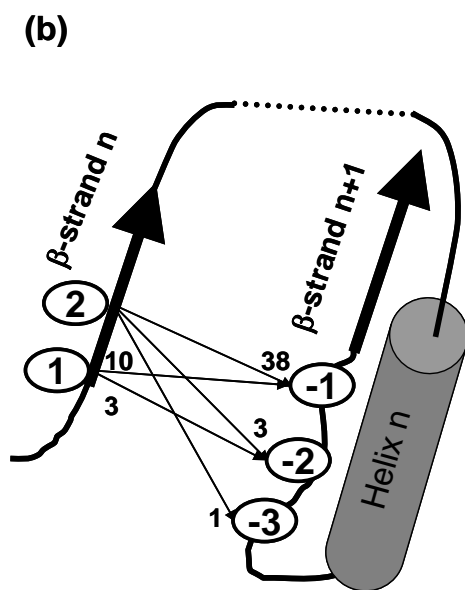
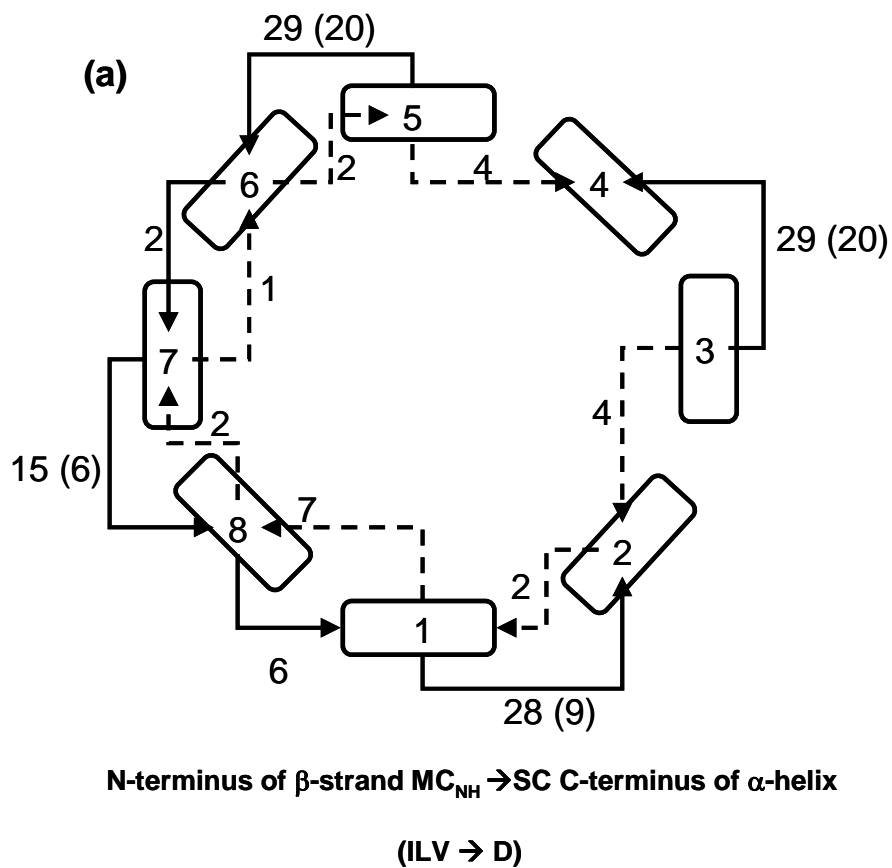
Table 3.2. Donor/acceptor sequence preferences of Class I $\beta\alpha$ -hairpin clamps in 71 TIM barrel proteins. The numbers of Class I $\beta\alpha$ -hairpin clamps between different main chain amide donors (horizontal list) and the side chain acceptors (vertical list) are shown.

Donor \ Acceptor	ALA	ARG	ASP	CYS	GLN	GLY	HIS	ILE	LEU	LYS	MET	PHE	SER	THR	TYR	VAL
ASN	1	1	1	1		2		4	2	2					1	2
ASP	2	2		3			1	18	14	2	2	5		2	1	23
GLN	1		1					1	1			1		2		1
GLU		2			2			2	1	1	1	1	2			1
HIS								1	1		1					
SER						2		1	2							2
THR			1			1		1				1		1		2
TYR									1							

group of 131 Class I $\beta\alpha$ -hairpins, displayed in Figure 3.5a, show a very strong preference for $\beta1\alpha1$, $\beta3\alpha3$, $\beta5\alpha5$ and $\beta7\alpha7$ clamps (77% of the clamps) in which the side chain acceptor is C-terminal to the main chain H-bond donor. With the exception of a half-dozen $\beta8\alpha8$ clamps, the paucity of clamps between $\beta2\alpha2$, $\beta4\alpha4$ and $\beta6\alpha6$ is distinctly different from their odd strand counterparts. With exception of the $\beta8\alpha8$ clamp, the number of $\beta\alpha$ -hairpin clamps in which the side chain acceptor is N-terminal to the main chain H-bond donor is much lower than the C-terminal counterparts. The additional number of clamps for the $\beta8\beta1$ interface could reflect the necessity for securing the N- and C-terminal strands.

Inspection of the locations of the 55 members of the Ile, Leu and Val (ILV)→Asp (D) subset of Class I $\beta\alpha$ -hairpin clamps reveals a startling observation. The ILV residues are always located in the odd-numbered strands, $\beta1$, $\beta3$, $\beta5$ or $\beta7$, and the D side chains are always located before the succeeding even-numbered strands, $\beta2$, $\beta4$, $\beta6$ or $\beta8$. Further, there is a strong preference for the ILV residue to occupy the 2nd position in the odd-numbered strand and for the D residue to occupy the position immediately preceding the even-numbered strand (Figure 3.5b). The 4-fold symmetry preference for the $\beta\alpha$ -hairpin clamps and their preferred placement in or adjacent to consecutive odd and even numbered β -strands have important implications for the evolution of TIM barrel architecture.

Figure 3.5. Positional preference of Class 1 $\beta\alpha$ -hairpin clamps in 71 TIM barrels. (a) The TIM barrel architecture is represented by a cross-sectional view of the eight β -strands. The number of $MC_{NH} \rightarrow SC$ $\beta\alpha$ -hairpin clamp interactions connecting adjacent β -strands with side chain H-bond acceptors C-terminal to the MC_{NH} donors (—), and with side chain H-bond acceptors N-terminal to the MC_{NH} donors (---) are indicated. The number of $\beta\alpha$ -hairpin clamps with (ILV) $MC_{NH} \rightarrow SC$ (D) is represented in parenthesis. (b) The positional preference of (ILV) $MC_{NH} \rightarrow SC$ (D) relative to the β -strands. The donor prefers either the first or second position of the β -strand and the acceptor prefers to be in the loop immediately preceding the subsequent β -strand. The number of times each pair of interactions occurs in the 71 TIM barrel database is indicated.



MC_{NH} (ILV) \rightarrow SC (D)

Figure 3.5

DISCUSSION

Recognition of potent $\beta\alpha$ -hairpin clamps in TIM barrel proteins

Experimental analysis of $\beta\alpha$ -hairpin clamps between main chain H-bond donors and side chain H-bond acceptors in three TIM barrels, α TS, sIGPS and eIGPS, has shown that a subset of these non-covalent interactions make substantial contributions to structure and stability. Comparison of the potency of the Class I $\beta\alpha$ -hairpin clamps in these three TIM barrels of low sequence identity shows no correlation between the contributions of these clamps to stability and either the location of the clamps in the structure or their conservation. What is apparent from an examination of the crystal structures of the three proteins, however, is that the length of the H-bond differentiates between the clamps that make major or minor contributions to stability (Table 3.1). The two potent $\beta\alpha$ -hairpin clamps in α TS, β 1 α 1 and β 3 α 3, one in sIGPS, β 7 α 7, one in eIGPS, β 7 α 7, and the single potent α 3 β 4 $\alpha\beta$ -hairpin clamp in α TS all have H-bond lengths of less than 2.80 Å. The lengths of the four $\beta\alpha$ -hairpin clamps whose elimination has minimal effects on structure and stability in these three TIM barrels all exceed 2.92 Å. The shorter H-bond distances for potent clamps, their unusually large contribution to stability (4-6 kcal mol⁻¹ vs. 1-2 kcal mol⁻¹ for typical side chain-main chain H-bonds)(Horovitz et al. 1990; Serrano et al. 1992; Myers and Pace 1996; Ibarra-Molero et al. 2004) and their role in defining secondary and tertiary structure must reflect the complex coupling of the H-bonds with their environments. Although the nominal resolutions of the crystal structures of these TIM barrel proteins dictate that the correlation between H-bond length and TIM barrel

stability be viewed as tentative, the apparent correlation does provide a testable and logical hypothesis for future experiments on $\beta\alpha$ -hairpin clamps in other TIM barrel proteins.

Survey analysis of $\beta\alpha$ -hairpin clamps in TIM barrel proteins

The observed non-equivalence of odd- and even- numbered strands in the TIM barrel architecture (Nagano et al. 2002) has its origins in the shear number (Wierenga 2001), $S = 8$, associated with the $(\beta\alpha)_8$ repeat. With $S = 8$, the β -sheet H-bond ladder connects the first residue of β_8 with the eighth residue of β_1 (Wierenga 2001). The tilt of the β -strands with respect to the central axis of the barrel required to achieve this shear permits a favorable orientation for the H-bonding network between adjacent parallel β -strands. This staggered arrangement of β -strands also permits a layered packing of side chains inside the barrel (Branden and Tooze 1999). One layer is composed of four residues oriented towards the inside of the barrel from all four even-numbered β -strands. The next layer is composed of side chains from the four odd-numbered β -strands, point outside the barrel (Figure 3.6) (Branden and Tooze 1999). The staggered arrangement of pairs of β -strands also provides an opportunity for an N-terminal cap of the β -sheet ladder between the main chain of the second residue on an odd-numbered β -strand and the side chain of a residue on the loop connecting the subsequent helix and even-numbered β -strand, i.e., the $\beta\alpha$ -hairpin clamp (Figure 3.6).

Figure 3.6. Architectural principles of the TIM barrel fold. The eight β -strands of the TIM architecture are represented as thick arrows. The β -strand number is indicated at the C-terminus of each β -strand. To convey the closed barrel architecture, $\beta 8$ is shown adjacent to $\beta 1$ (thick gray arrow) as well as adjacent to $\beta 7$. Each residue is represented as an ellipse and the position on the β -strand is indicated. The filled ellipses represent residues that point inside the barrel. Residues in the loop preceding the β -strand show the one letter code for the most common amino acid at that position (>15%) in the 71 TIM barrel database. The dotted arrows represent the H-bond network for the β -barrel. The unbroken lines represent the $\beta\alpha$ -hairpin clamp interactions between the second residue of an odd-numbered β -strand and the side chain of the residue immediately preceding the subsequent β -strand. These interactions are “double braced” with main chain-main chain interactions between the same two residues.

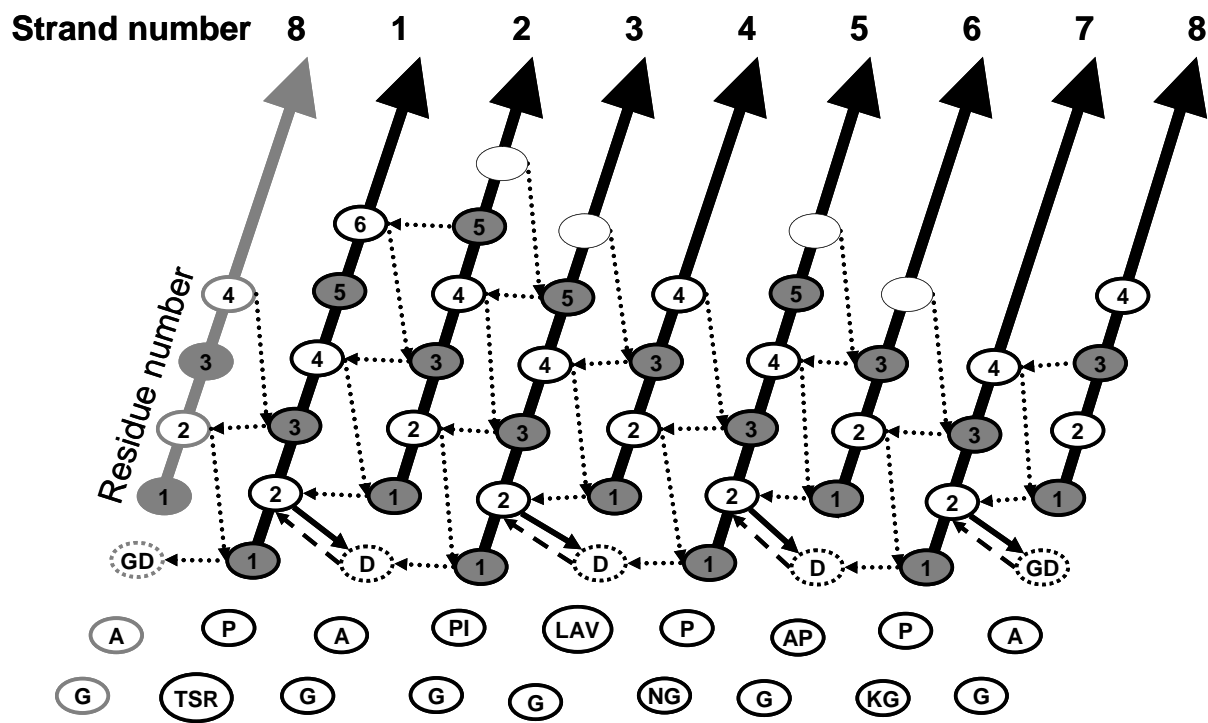


Figure 3.6

The chemical origin for the asymmetry between odd- and even-numbered β -strands is apparent from an inspection of the residue preference (>15%) at positions encompassing the N-terminus of each β -strand (Figure 3.6). The conserved proline at the -1 position of odd-numbered β -strands is well suited to provide a kink in the backbone that marks the beginning of a β -strand (FarzadFard et al. 2008). The preferred sequence pattern of the tight turn connecting the α -helix and the subsequent even-numbered β -strand, GAD, has been observed previously (Nagano et al. 2002). The glycine breaks the α -helix, the minimal size of the alanine side chain enables the tight turn and the aspartic acid just prior to the beginning of even-numbered β -strands forms the $\beta\alpha$ -hairpin clamp and braces the $\beta\alpha\beta$ module. The N-terminal cap for the odd-numbered β -strand is very often reinforced by a main chain-main chain H-bond between the two residues, with the amide group of the aspartic acid acting as the donor to the main chain carbonyl oxygen of the partner residue (Figure 3.6). While other side chain acceptors are observed, the length of the aspartic acid side chain appears to be ideal for the reinforcement of the side chain-main chain H-bond with the main chain-main chain H-bond, providing a plausible explanation for its higher frequency in $\beta\alpha$ -hairpin clamps.

The common occurrence of the $\beta\alpha$ -hairpin clamp H-bonds in TIM barrel proteins implicates their importance in TIM barrel structure. The residue preferences, locations of $\beta\alpha$ -hairpin clamp H-bonds, and their contributions to the stability of TIM barrel proteins, may be useful for protein structure prediction and protein engineering (see Discussion chapter).

CHAPTER IV

DISCUSSION AND FUTURE DIRECTIONS

Implications for TIM barrel protein structure and folding

From the experimental results on three TIM barrel proteins, α TS, sIGPS and eIGPS, and the analysis of the survey of 71 TIM barrel proteins, we conclude that the $\beta\alpha$ -hairpin clamp H-bond is a common feature for the TIM barrel protein family. $\beta\alpha$ -hairpin clamps can also make substantive contributions to the stability of TIM barrel proteins.

The 4-fold symmetry of the preferred $\beta\alpha$ -hairpin clamps is mirrored in the packing of the side chains in the interior of the β -barrel. A residue oriented towards the inside of the β -barrel from an even-numbered β -strand is at the same level as corresponding residues from the three remaining even-numbered β -strands. The next layer is composed of the four side chains from the odd-numbered β -strands (Figure 3.6); the third, and usually final layer, is composed again of side chains from the odd-numbered β -strands (Branden and Tooze 1999). Both the preferred clamp locations and the layering of side chains inside the β -barrel have their origins in the tilt of the β -strands with respect to the central axis of the barrel. The $S = 8$ shear provides a favorable orientation for the H-bonding network between adjacent parallel β -strands, opportunities for side chain-main chain clamp interactions and the layering of the side chains inside the barrel. Together, these non-covalent interactions stabilize the $(\beta\alpha)_8$, TIM barrel structure (Figure 3.6) (Wierenga 2001).

The $\beta\alpha\beta$ module appears to be the minimum, self-contained folding unit that could have been an evolutionary precursor of the TIM barrel architecture (Branden and Tooze 1999; Zitzewitz et al. 1999; Zitzewitz and Matthews 1999; Nagano et al. 2002;

Frenkel and Trifonov 2005). The consistent positioning of the ILV to D, Class I, $\beta\alpha$ -hairpin clamps within the $\beta\alpha\beta$ modules leads to plausible speculation on its importance in the origin of TIM barrel and other $\beta\alpha$ -repeat structures. For the primordial $\beta\alpha\beta$ module and during its subsequent evolution, the clamps may have played an important role in defining the register between the intra-module β -strands and in stabilizing the fold. Different packing arrangements of two or more modules can be envisioned to achieve different protein folds including the P-loop hydrolase fold, the flavodoxin fold, the Rossmann fold and the thiamin binding fold (Gerstein 1997). Several possible gene duplication or gene fusion events could have converged to the final $(\beta\alpha)_8$ barrel proteins, resulting in the observed 4-fold (2+2+2+2) (Gerstein 1997; Branden and Tooze 1999; Zitzewitz et al. 1999; Nagano et al. 2002; Frenkel and Trifonov 2005) and 2-fold (4+4) symmetry (Hocker et al. 2002) or even the possibility of a 6+2 barrel (Higgins et al. 1979b).

In the primordial TIM barrel, each clamp interaction may have been important for maintaining stability of the protein. However, as the sequences diverged, the development of other sources of stability could bridge the $\beta\alpha\beta$ modules and lead to global cooperativity. This is clearly evident from the existence of hydrophobic clusters of isoleucine, leucine and valine residues in TIM barrels that play an important role in maintaining their structure and stability (Gu et al. 2007b; Wu et al. 2007). While the locations of these ILV clusters are also not conserved, clusters containing 10 or more side chains and burying more than 500 \AA^2 of surface area are crucial for the formation of transient and stable folding intermediates in these proteins (Gu et al. 2007a; Gu et al.

2007b; Wu et al. 2007; Vadrevu et al. 2008). These ILV clusters always span more than one $\beta\alpha\beta$ module, offering, in addition to the H-bonds between the β -strands in adjacent modules, a thermodynamic linkage between the modules. This evolving and enhanced stabilization of the structure with clusters of ILV residues may have ultimately rendered some of the clamp interactions dispensable.

The absence of the $\beta\alpha$ -hairpin clamps in the I states of all three TIM barrels demonstrates that the potent effects of these clamps only appear as the N state appears. Kinetic folding studies on α TS (Yang et al. 2007) revealed further that each clamp is crucial for accessing the transition state ensemble required to reach the properly-folded TIM barrel structure. Although the local connectivity of the $\beta\alpha\beta$ module might have been expected to enable the clamp to play a significant role in the early stages of the folding reaction, the primary role of the potent set of clamps is to drive the final stage of the reaction to completion and fully develop the global cooperativity. The coupling between the formation of the side chain-main chain H-bond with substantial changes in secondary and tertiary structure provides a rationale for the unexpectedly large contributions to the free energy of folding by these clamp.

Broader involvement of clamp interactions

Similar $\beta\alpha$ -hairpin clamps exist in other TIM barrel proteins, including the *ioII* protein from *B. subtilis* (L106-N72), the cyclase moiety of imidazole-3-glycerol phosphate synthase (HisF) from *T. maritima* (F77-D98), N-(5'-phosphoribosyl)

anthranilate isomerase (PRAI) from *E.coli* (I355-D374), and the founding member of the motif, triosephosphate isomerase (TIM) from chicken (I206-D227).

HisF from the hyperthermophile *T. maritime* could be the next target to study the role of the clamp H-bonds, as the folding mechanism of HisF is one ongoing project in the Matthews lab. HisF is one half of the heterodimeric protein complex with imidazole glycerol phosphate synthase which is responsible for the fifth step of histidine biosynthesis (Beismann-Driemeyer and Sterner 2001). In HisF, three Class I $\beta\alpha$ -hairpin clamp H-bonds were observed, including I6-D45 ($\beta1\alpha1$), F77-D98 ($\beta3\alpha3$) and I198-D219 ($\beta7\alpha7$). All three clamp H-bonds share the Asp residue as the H-bond acceptor, two of them have Ile as the H-bond donor, and the other has Phe as the H-bond donor. The H-bond lengths for these three H-bonds are near 2.80 Å (2.77 Å, 2.85 Å and 2.83 Å), suggesting that all three might make significant contributions to structure and stability.

Because it appears that the length of the clamp H-bond differentiates between the clamps that make major or minor contributions to stability, the bond length may be one criterion to predict potent clamp H-bonds. As summarized from the three TIM barrel proteins I tested, the cutoff for the bond length is 2.80 Å. Table 4.1 lists seven proteins from the database of 71 TIM barrel proteins, which contain the Class I ILV \rightarrow D $\beta\alpha$ -hairpin clamp H-bonds with bond length shorter than 2.80 Å. These clamp H-bonds are predicted to be crucial to protein structure and stability and would be interesting candidates for further study.

Expansion of the search of the $\beta\alpha$ -hairpin clamp H-bonds to other types of $(\beta\alpha)_n$ structures, e.g., the $\alpha/\beta/\alpha$ sandwich flavodoxin-fold, the Rossmann fold, and the leucine-

Table 4.1. Potential potent clamp H-bonds predicted from the database 71 TIM barrel proteins.

PDB	Main Chain	Donor	Side Chain	Acceptor	DA Distance (Å)	Protein (Organism)
1d3g	252	VAL	278	ASP	2.80	Dihydroorotate Dehydrogenase (Human)
1eom	82	VAL	121	ASP	2.75	Endo-Beta-N-Acetylglucosaminidase F3 (Flavobacterium sp)
1gvf	131	VAL	170	ASP	2.78	Tagatose-1,6-bisphosphate Aldolase (E.coli)
1huv	228	LEU	248	ASP	2.56	Mandelate Dehydrogenase (Pseudomonas sp)
1km0	67	ILE	93	ASP	2.73	Orotidine Monophosphate Decarboxylase (Methanobacterium sp)
2ebn	12	LEU	40	ASP	2.72	Endo-Beta-N-Acetylglucosaminidase F3 (Chryseobacterium sp)
	87	VAL	125	ASP	2.72	
2tmd	219	ILE	252	ASP	2.80	Trimethylamine dehydrogenase (Methylophilus sp)

rich repeat fold revealed the presence of one or more clamps in selected members of these motifs as well. Further experimental and bioinformatic analysis is required to determine if the enhancement of stability gained by linking consecutive elements of secondary structure in α TS and IGPS through these types of long-range non-covalent interactions is a general property of the other $(\beta\alpha)_n$ classes of proteins.

Applications to protein structure prediction and protein engineering

The occurrence of the $\beta\alpha\beta$ motif in a large number of protein families (Gerstein 1997; Branden and Tooze 1999) suggests that the N-terminal capping of β -strands by $\beta\alpha$ -hairpin clamps (FarzadFard et al. 2008), akin to the analogous N-capping of α -helices (Presta and Rose 1988; Aurora and Rose 1998), may be a useful property for the refinement of protein fold prediction and for engineering stability in $\beta\alpha$ -repeat proteins.

By screening for $\beta\alpha$ -hairpin clamps between the main chain amide hydrogens at favored positions near the N-terminus of a β -strand and H-bond acceptor side chains in the loop before the subsequent β -strand (~25 residues apart in sequence), the 3D structural prediction could be improved. These clamps would establish the register of the pair of β -strands, and, with the very short loop linking the intervening α -helix to the second β -strand. It might be possible to establish the register of the α -helix on the β -strand pair in the $\beta\alpha\beta$ module. Although TIM barrels typically contain only a few $\beta\alpha$ -hairpin clamps, defining the spatial relationships of the components of those $\beta\alpha\beta$ modules might increase the probability of predicting the packing of adjacent $\beta\alpha$ -repeats

in the structures. The effect of accurately predicting the structure of one $\beta\alpha\beta$ module might, therefore, propagate throughout the TIM barrel.

For protein engineering, the TIM barrel architecture provides a scaffold that is capable of a very diverse set of enzymatic functions (Nagano et al. 2002). TIM barrel enzymes have been re-engineered to accommodate alternative substrates (Vick and Gerlt 2007; Alahuhta et al. 2008; Gerlt and Babbitt 2009) and even to catalyze non-biological reactions (Rothlisberger et al. 2008). Because the active sites of TIM barrel enzymes are invariably composed of the loops protruding from the C-termini of the β -strands, the stability of engineered constructs might be enhanced without compromising activity by introducing $\beta\alpha$ -hairpin clamps at the other pole of the TIM barrel (Williams et al. 1999).

For instance, the $\beta 1\alpha 1$ clamp H-bond F50-S82 in eIGPS has no significant effect on protein stability, while its counterpart in α TS (F19-D46) is crucial to protein structure and stability. This comparison implies that Asp might make a stronger H-bond than Ser. Preliminary data for the eIGPS clamp variant S82D (see Appendix) suggest that the replacement of Ser with Asp as the clamp H-bond acceptor may increase the stability of the protein. If the stability of the native state of eIGPS is increased for the S82D variant, a structural analysis could be performed to demonstrate the presence of the clamp H-bond.

In conclusion, this thesis adds to an appreciation for a common structural feature of the TIM barrel structure. The long-range side chain-main chain H-bonds, $\beta\alpha$ -hairpin clamp H-bonds, can play a surprising and significant role in the stability and the structure of one of the most common motifs in biology. Continuing studies will increase our understanding of these clamp H-bonds and improve structure prediction and engineering.

CHAPTER V

MATERIALS AND METHODS

Site-directed mutagenesis

Oligonucleotides for mutagenesis were purchased from Eurofins MWG Operon (Huntsville, AL), and the QuickchangeTM site-directed mutagenesis kit was obtained from Stratagene (La Jolla, CA). The method involved one round of PCR amplification using two complementary primers, and the variant was amplified by PCR in the plasmid containing WT protein. The endonuclease, Dpn1, was used to selectively digest the methylated parent plasmid, and the product was transformed into XL1-Blue competent cells from Stratagene (La Jolla, CA).

The site-directed mutations were confirmed by DNA sequence analysis ((UC Davis Sequencing Facility, Davis, CA and Genewiz Inc, NJ). The pXH plasmid containing α TS was available in Matthews' lab (Wu and Matthews 2002b). The plasmid encoding a truncated version of sIGPS in which the non-canonical additional α -helix (α 00) at the N-terminus, residues 1-26, was deleted to eliminate aggregation during folding, pTNI4, was obtained from Dr. K. Kirschner at the University of Basel, Switzerland (Forsyth and Matthews 2002). The plasmid coding for eIGPS, pJB122, was obtained from Dr. J.M. Blackburn at the University of the Western Cape, South Africa (Altamirano et al. 2000). The eIGPS, with an additional Ala residue after the start codon and a C-terminal FLAG peptide sequence (GSDYKDDDDK), is fully folded and catalytically active (Altamirano et al. 2000).

Protein expression and purification

α TS variant proteins were expressed in CB149 cells, and the cells were grown at 37°C for about 12 to 16 hours. The cells were harvested, and the cell pellet was sonicated in 100 mM potassium phosphate, 5 mM K₂EDTA, 2 mM dithioerythritol (DTE), and 0.2 mM Phenylmethylsulfonyl fluoride (PMSF) at pH 7.8 with lysozyme (1mg/g cell). The α TS variants were largely found in the insoluble fractions of the cell lysate. The insoluble fraction of the cell lysate was solubilized in 6 M urea, 10 mM potassium phosphate, 2 mM K₂EDTA, and 2 mM DTE. Then protein was refolded by adding the solution drop-wise to 15X volume of refolding buffer (10 mM potassium phosphate, 2 mM K₂EDTA, and 2 mM DTE at pH 7.8) or dialyzed in refolding buffer to remove the urea. The refolded protein was then loaded onto a DE52 resin column preequilibrated with refolding buffer. After washing the column with one column volume of refolding buffer, the protein was eluted with linear gradient-elution buffer (low-salt buffer: 10 mM potassium phosphate, 4 mM K₂EDTA, and 2 mM DTE at pH 7.8; high-salt buffer: 100 mM potassium phosphate, 4 mM K₂EDTA, and 2 mM DTE at pH 7.8). The peak fractions were pooled and the purity was demonstrated by the appearance of a single band by SDS-PAGE. The protein was concentrated in an Amicon concentrator fit with a YM-10 membrane. The protein was precipitated by 80% (w/v) ammonium sulfate and stored at 4 °C. The protein was redissolved in the standard buffer (10 mM potassium phosphate, 0.2 mM K₂EDTA, 1 mM β -mercaptoethanol (β ME) at pH 7.8) and dialyzed to remove ammonium sulfate. The protein concentration was

determined from the measurement of the UV absorbance at 278 nm, using an extinction coefficient of $12,600 \text{ M}^{-1} \text{ cm}^{-1}$.

The sIGPS variant proteins were expressed in BL21/DE3 cells, and the cells were grown at 37°C to an OD of 0.6 at 600 nm. Expression was induced by the addition of IPTG (1 mM) for twelve hours, after which the cells were harvested. The cell pellet was suspended in the lysis buffer (100 mM potassium phosphate, 2 mM K_2EDTA , 1 mM DTE, 0.3 mM PMSF at pH 7.8 and lysozyme (1mg/g cell)). After sonication, the protein was found in the soluble fraction of the cell lysate. The supernatant was dialyzed against running buffer (10 mM potassium phosphate, 2 mM K_2EDTA at pH 7.8) and then loaded onto a DEAE-Sepharose Fast Flow (Amersham Pharmacia Biotech, Piscataway, NJ) column preequilibrated with running buffer. The column was washed with one column volume of running buffer and eluted with a linear gradient from low-salt buffer (10 mM potassium phosphate, 2 mM K_2EDTA , and 2 mM DTE at pH 7.8) to high-salt buffer (1.0 M potassium phosphate, 2 mM K_2EDTA , and 2 mM DTE at pH 7.8). Protein fractions (20 ml) were collected and analyzed by SDS-PAGE. Fractions containing the desired protein were concentrated and dialyzed against running buffer before loading on a Sephacryl S-200 (Amersham Pharmacia Biotech, Piscataway, NJ) column. Fractions were collected and analyzed by SDS-PAGE. The peak fractions were pooled and the purity (>95%) was confirmed using electrospray mass spectrometry at the Proteomics Facility at University of Massachusetts Medical School (Worcester, MA). The

concentration of sIGPS was determined from its extinction coefficient at 278 nm, 10,240 $M^{-1} \text{ cm}^{-1}$.

The expression and purification of eIGPS and its variants followed the same protocol as that of sIGPS, with the exception that pH for all of buffers was 7.0. The extinction coefficient of eIGPS is 20,130 $M^{-1} \text{ cm}^{-1}$ at 278 nm.

Circular dichroism

CD spectroscopy was employed to monitor the secondary structure and the tertiary structure near aromatic side chains. Spectra were obtained on a Jasco Model J-810 spectropolarimeter equipped with a thermoelectric cell holder. Far-UV CD data were collected from 280 nm to 185 nm at a scan rate of 50 nm/min and at 1 nm intervals using a 0.1 cm pathlength cell, a bandwidth of 2.5 nm, and an averaging time of 8 s. Three replicate spectra were collected and averaged. The protein concentration was approximately 5 μM . Near-UV CD data were collected from 350 nm to 250 nm at 5 nm/min using a 0.5 cm path length cell, and the protein concentration was 50-150 μM . The temperature was maintained at 25 °C with a computer-controlled Peltier system.

Equilibrium experiments

The stability of the variants was measured by a urea-denaturation equilibrium unfolding experiment. A Hamilton 540B automatic titrator was used to prepare the samples in solution with from 0 to 8 M urea at concentration increments of 0.2 M urea to

enhance the precision of the measurements. The samples were incubated overnight at 25 °C to ensure equilibration. Far-UV CD data were collected from 265 nm to 215 nm for each sample.

Kinetic experiments

CD manual-mixing kinetic experiments were performed on a Jasco Model J-810 spectropolarimeter equipped with a thermoelectric cell holder using a 1 cm pathlength cell, a bandwidth of 2.5 nm, and an averaging time of 1 s. The dead-time of the experiments was 3 s and the instrument response time was about 5 s. The change in ellipticity as a function of time was monitored at 222 nm. Unfolding jumps were from 0 M urea to various final concentrations of urea (2 M–6 M), and refolding jumps were performed from 6 M to various final urea concentrations (0.6 M–3 M). The final concentration of protein was 3–5 μ M. Stopped-flow CD experiments were performed with an Aviv model 202SF stopped-flow CD spectrometer. The cell path length was 1.0 mm, and the dead time for mixing was 5 ms. Stopped-flow fluorescence experiments were performed on an Applied Photophysics SX-18MV stopped-flow fluorometer. The excitation wavelength was 280 nm, and the emission above 300 nm was detected through a cut-off filter.

Kinetic unfolding experiments to determine the stability of the folded states of the clamp-deletion variants were performed by jumping from different initial urea concentration (0–5.6 M for α TS and 0–2.8 M for eIGPS) to a final urea concentration (6M for α TS and 3 M for eIGPS). Protein samples were first equilibrated in the initial urea

concentration overnight and then jumped to the desired urea-buffer solutions by a 1:10 dilution. The final protein concentration was around 5 μM .

Data analysis

Equilibrium CD data at 222 nm were fit to a three-state model, $\text{N} \rightleftharpoons \text{I} \rightleftharpoons \text{U}$ (Bilsel et al. 1999b). The kinetic data were fit to a sum of exponentials and a constant, using equation $A(t) = A(\infty) + \sum A_i \exp(-t/\tau_i)$. Each τ_i was plotted as a function of urea concentration to provide a chevron analysis of the folding reaction (Matthews 1987). All thermodynamic and kinetic folding data were fit using Savuka version 6.2, an in-house, non-linear, least-squares program.

Survey of TIM barrel structures

The CATH-35 control data set was downloaded from <http://release.cathdb.info/v3.2.0/CathDomainPdb.S35.v3.2.0.tgz>. A database of 71 TIM barrel proteins had been previously developed and made available at <http://www.cbrc.jp/~gromiha/tim/proteinlist.html> (Gromiha et al. 2004). The secondary structure and solvent accessibility were calculated using DSSP program (Kabsch and Sander 1983), and the H-bond interaction parameters were calculated using default settings of the HBPLUS program (McDonald and Thornton 1994).

Definitions of $\beta\alpha$ -hairpin clamp H-bonds

For each protein, the eight canonical β -strands and α -helices in the context of the TIM barrel architecture were identified and numbered sequentially. The $\beta\alpha$ -hairpin clamp H-bonds were subjected to the following filters: 1) the H-bonds must be between main chain and side chain, 2) the amino acid chain length between the donor and acceptor must be ≥ 15 residues, and 3) the chain must include exactly one β -strand and one α -helix identified in the context of the TIM barrel architecture. For the case of the $\beta 8\alpha 8$ clamps, the residues prior to $\beta 1$ were included. The H-bonds that passed each stage of the filtering process were exported to a PyMOL script in color-coded fashion for manual confirmation.

APPENDIX

Verification of the intermediate state for eIGPS $\beta 7\alpha 7$ clamp-deletion variant N231A by kinetic jump experiments

To verify the presence of the intermediate state, I, for the clamp-deletion variant N231A in eIGPS, and determine the free energy difference between the N and I states, the kinetic unfolding experiment was employed. Because the amplitude of the unfolding reaction is directly proportional to the population of the folded state from which the reaction is initiated, the stability of the folded state can be measured by monitoring the amplitude of the unfolding reaction (to constant urea concentration) as a function of the initial urea concentration. The amplitude of the unfolding phase, proportional to the population of the N state, is expected to decrease in a sigmoidal fashion as the initial urea concentration is increased into the zone where the native state becomes depopulated. This procedure was validated previously on α TS (Yang et al. 2007).

A fit of the sigmoidal decrease in ellipticity at increasing initial urea concentrations to a two-state model for WT eIGPS yielded a stability of 5.29 ± 1.71 kcal mol⁻¹ (Figure A.1). This value is in excellent agreement with that for the N to I transition from the standard equilibrium titration experiment on the WT protein, 5.60 ± 0.99 kcal mol⁻¹ (Figure 3.4c and Table 3.1). The urea dependence of the stability, the m-value, is also in excellent agreement, 2.34 ± 0.74 vs 2.46 ± 0.42 kcal mol⁻¹ M⁻¹ (Table 3.1).

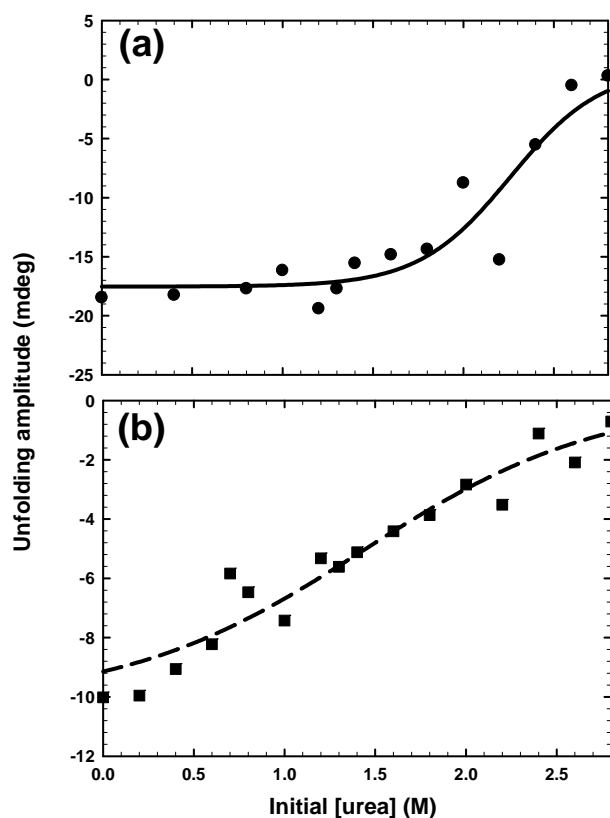


Figure A.1. The dependence of the amplitude for the unfolding phase for (a) WT eIGPS (●, solid line) and (b) eIGPS N231A (■, dashed line) on the initial urea concentration; the final urea concentration in all cases was 3 M urea. The lines represent the fit of the data to a two-state model with $\Delta G^\circ = 5.29 \pm 1.71 \text{ kcal mol}^{-1}$ and $m = 2.34 \pm 0.74 \text{ kcal mol}^{-1} \text{ M}^{-1}$ for WT eIGPS and $\Delta G^\circ = 1.28 \pm 0.15 \text{ kcal mol}^{-1}$ and $m = 0.89 \pm 0.11 \text{ kcal mol}^{-1} \text{ M}^{-1}$ for eIGPS N231A.

Applying this method to eIGPS N231A (Figure A.1), the free energy difference for N to I transition extracted from the kinetic experiment is 1.28 ± 0.15 kcal mol⁻¹ and the m-value is 0.89 ± 0.11 kcal mol⁻¹ M⁻¹. The free energy difference for the I to U transition for eIGPS N231A can be estimated by fitting the unfolding titration data (Figure 3.4c) to a three-state model. In this fit, the parameters for the N to I reaction were fixed to those obtained from the kinetic unfolding experiments. The thermodynamic parameters obtained for the I to U transition are shown in Table 1 in main text.

One nonpotent Class II clamp H-bond in sIGPS

A back-to-back Class II clamp H-bond T129-E155 was found right next to the Class I $\beta 3\alpha 3$ clamp H-bond I107-D128 in sIGPS discussed in Chapter III. The side chain of T129, which is next to D128, makes a H-bond with the main chain of E155, clamping $\beta 4$ and $\alpha 4$. Both the single mutation T129A and the double mutation D128AT129A were tested to explore the role of this Class II clamp H-bond in sIGPS. The very similar response in far-UV and near-UV CD spectra for the clamp-deletion variant T129A and double mutant D128AT129A indicates that neither of them significantly perturbs the TIM barrel structure (Figure A.2). The urea denaturation unfolding curves for T129A and the double mutant D128AT129A are also similar to that of the WT sIGPS (Figure A.3). Both unfolding titration curves are well-fit by a three-state model, $N \rightleftharpoons I \rightleftharpoons U$. The thermodynamic parameters for these two sIGPS clamp-deletion variants and together with the variant D128A are listed in Table A.1. The

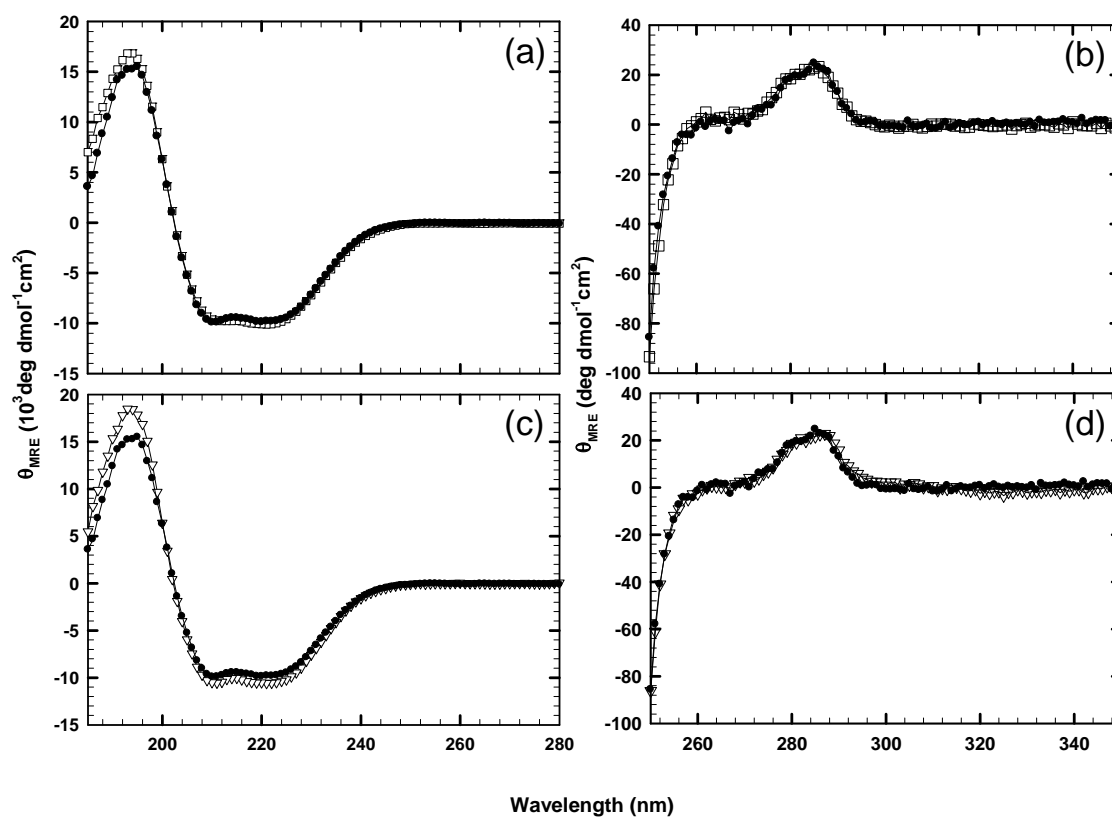


Figure A.2. Far-UV (a, c) and near-UV (b, d) CD spectra of WT (●), T129A (a, b) (□), and D128AT129A sIGPS (c, d) (▽). Buffer contains 10 mM potassium phosphate, 0.2 mM K_2EDTA , and 1 mM βME at pH 7.8, 25 °C.

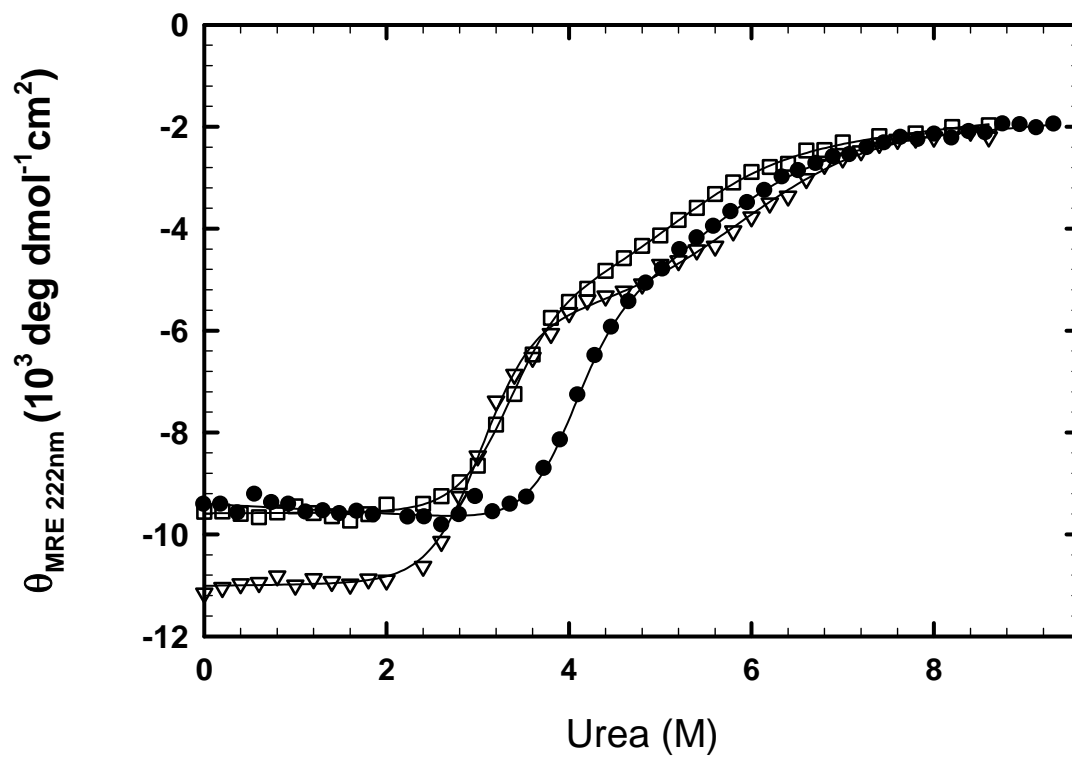


Figure A.3. Urea denaturation equilibrium unfolding curves of WT (●), D128A (□) and D128AT129A sIGPS (▽). The continuous lines represent fits of the data for each protein to a 3-state equilibrium folding model as described in the text.

Table A.1. Thermodynamic parameters for the urea-induced unfolding of sIGPS clamp-deletion variants at 25 °C^a.

sIGPS variants	$\Delta G^{\circ}_{\text{NI}}(\text{H}_2\text{O})$ (kcal mol ⁻¹)	$-m_{\text{NI}}$ (kcal mol ⁻¹ M ⁻¹)	$\Delta G^{\circ}_{\text{IU}}(\text{H}_2\text{O})$ (kcal mol ⁻¹)	$-m_{\text{IU}}$ (kcal mol ⁻¹ M ⁻¹)	$\Delta\Delta G^{\circ}_{\text{NI}}$ (kcal mol ⁻¹) ^b
WT	8.50±0.40	4.60±0.80	2.10±0.10	0.86±0.13	-
D128A	7.75±1.99	5.33±1.40	1.99±0.12	0.89±0.23	-0.75±1.44
T129A	6.74±2.02	5.21±1.22	2.02±0.13	1.01±0.22	-1.76±1.46
D128AT129A	6.20±2.04	4.53±0.52	2.04±0.15	0.77±0.08	-2.30±1.47

a. Buffer conditions: 10 mM potassium phosphate, 0.2 mM K₂EDTA, 1 mM βME at pH 7.8.

b. Perturbation in stability for N to I reaction, calculated by $\Delta\Delta G^{\circ}_{\text{NI}} = \Delta G^{\circ}_{\text{NI}}(\text{H}_2\text{O}, \text{variant}) - \Delta G^{\circ}_{\text{NI}}(\text{H}_2\text{O}, \text{WT})$.

$\Delta\Delta G^{\circ}_{\text{NI}}$ for either single mutation D128A or T129A is relatively small (-0.75 ± 1.44 kcal mol⁻¹ for D128A and -1.76 ± 1.46 kcal mol⁻¹ for T129A). The loss of the Class I $\beta 3\alpha 3$ clamp H-bond or Class II $\beta 4\alpha 4$ clamp H-bond has only minor effect on protein stability. As the $\Delta\Delta G^{\circ}_{\text{NI}}$ for the double mutant D128AT129A is -2.30 ± 1.47 kcal mol⁻¹, the net effect of the elimination of both clamp H-bonds is the sum of two single eliminations. Thus, the Class II clamp H-bond T129-E155 in sIGPS is not crucial to protein stability.

One clamp variant in eIGPS may enhance protein stability

As from previous results, the $\beta 1\alpha 1$ clamps which are conserved in α TS (F19-D46) and eIGPS (F50-S82) behave differently in two proteins. The $\beta 1\alpha 1$ clamp H-bond in α TS is potent to protein structure and stability, while that in eIGPS is not. The two clamps share the same H-bond donor, Phe, and have different acceptors, which are Asp for α TS and Ser for eIGPS. The strategy is to replace the acceptor S82 with Asp in eIGPS, and test the hypothesis that a new clamp H-bond F50-D82 is introduced and will increase the protein stability in eIGPS.

Similar experimental methods have been used. Characterization of the clamp H-bond variant, S82D is compared with WT eIGPS.

In Figure A.4, the far-UV and near-UV CD spectrum for S82D and WT eIGPS are very similar, except small differences in the 185-195 nm and 215-225 nm ranges. This suggests that the replacement of Ser with Asp may still retain the TIM barrel structure of eIGPS.

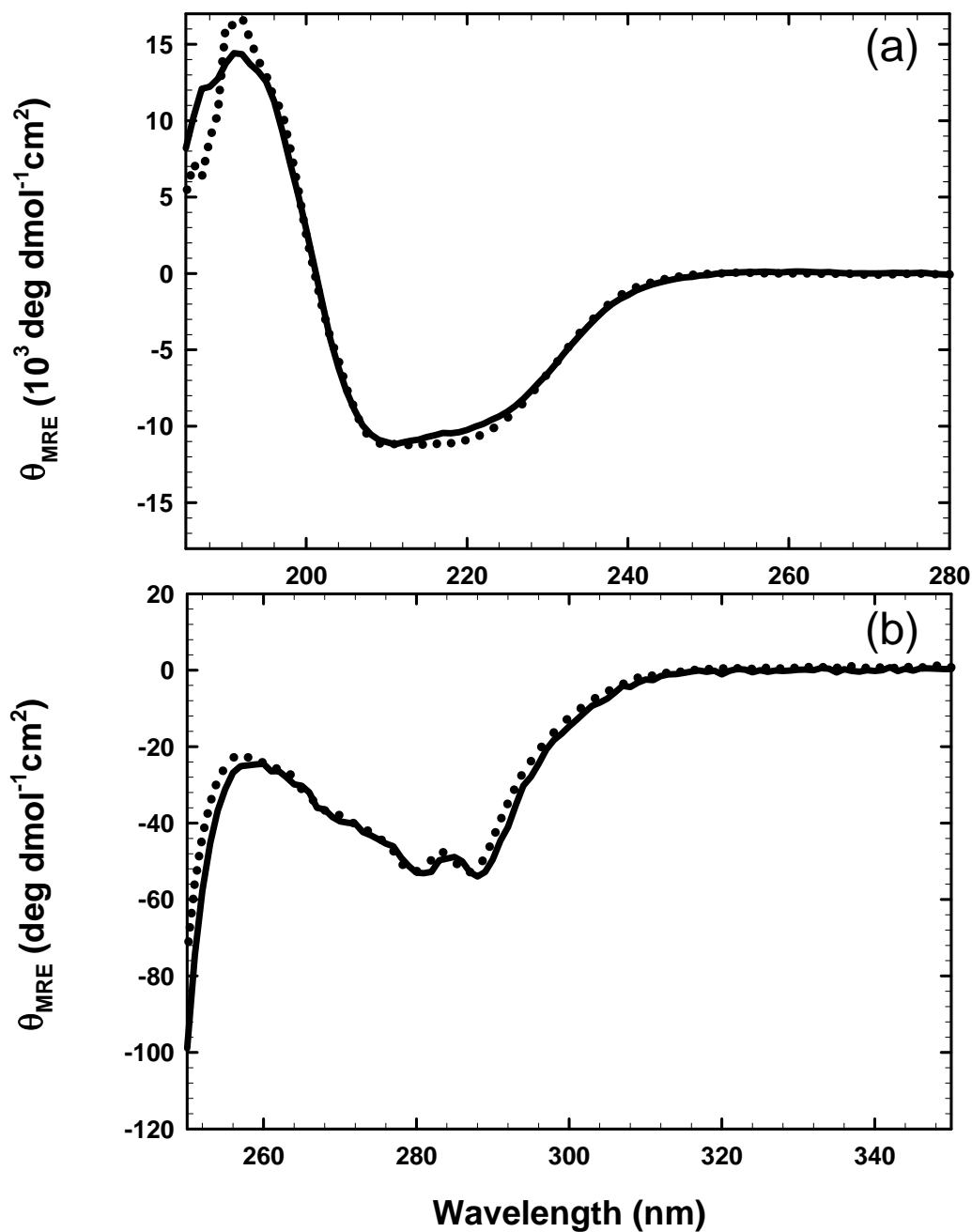


Figure A.4. (a) Far-UV and (b) near-UV CD spectra of WT (solid line) and S82D (dotted line) eIGPS. Buffer contains 10 mM potassium phosphate, 0.2 mM K_2EDTA , and 1 mM βME at pH 7.0, 25 °C.

The urea denaturation equilibrium unfolding experiments were performed by CD to probe the protein stability. From Figure A.5, the unfolding reaction of the clamp variant S82D is well-described by a three-state model, $N \rightleftharpoons I \rightleftharpoons U$. Compared to the WT eIGPS, both $N \rightleftharpoons I$ and $I \rightleftharpoons U$ transitions have changed. Table A.2 lists the thermodynamic parameters when the curves in Figure A.5 are fitted to a three-state model. For $N \rightleftharpoons I$ transition, the ΔG°_{NI} for eIGPS S82D is increased by about 3 kcal mol⁻¹ as compared to WT protein, while the m-value is increased by about 1 kcal mol⁻¹ M⁻¹. For $I \rightleftharpoons U$ transition, the ΔG°_{IU} is increased by about 2 kcal mol⁻¹ and the m-value is also increased by about 1 kcal mol⁻¹ M⁻¹.

Overall, the protein stability of eIGPS S82D has been increased by about 5 kcal mol⁻¹, and the I state has also been affected with this single amino acid replacement. Unlike other clamp interactions discussed before, which disappear at the I state, the new clamp H-bond F50-D82, if it was there, seems to be retained in the I state and help stabilizing the I state.

These data are only preliminary results. More thermodynamic experiments are needed to verify the increase on protein stability for eIGPS S82D. The HX-NMR technique could be applied to probe the protection of the hydrogen involved in this H-bond. The structural information will be a strong argument of the existence of the new clamp H-bond F50-D82. If the hypothesis is true, to build in the clamp H-bond will be a useful strategy to enhance TIM barrel stability.

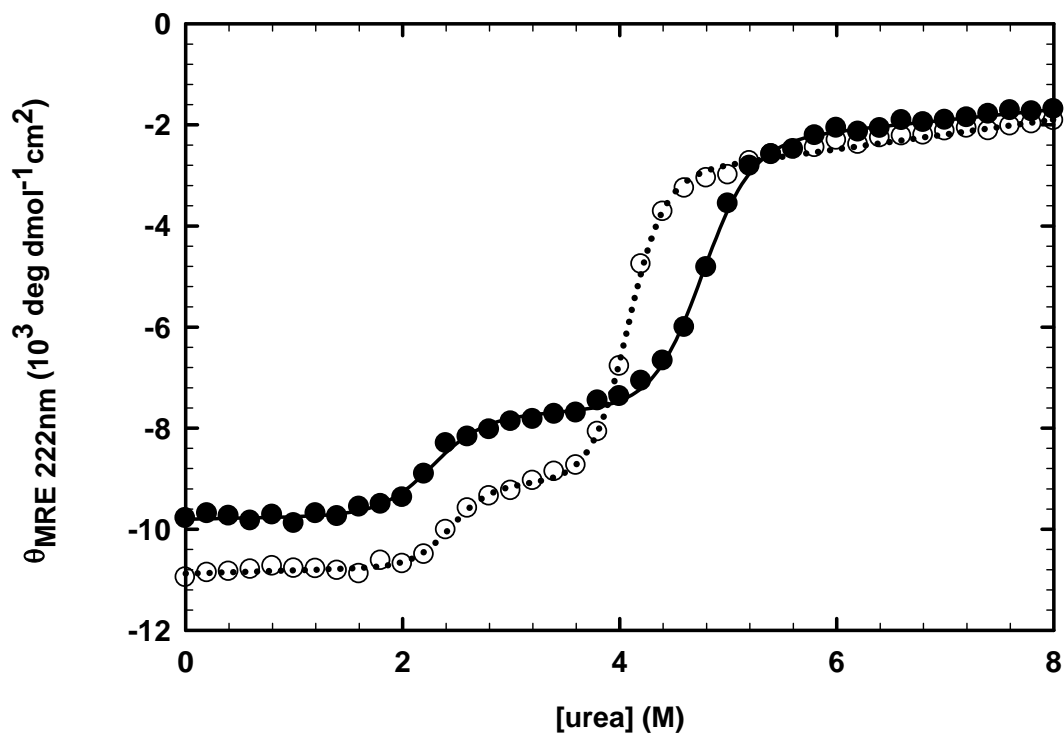


Figure A.5. Urea denaturation equilibrium unfolding curves of WT (●, solid line) and S82D (○, dotted line) eIGPS. The lines represent fits of the data for each protein to a 3-state equilibrium folding model as described in the text.

Table A.2. Thermodynamic parameters for the urea-induced unfolding of WT eIGPS and clamp variant S82D at 25 °C^a.

eIGPS variants	$\Delta G^{\circ}_{\text{NI}}(\text{H}_2\text{O})$ (kcal mol ⁻¹)	$-m_{\text{NI}}$ (kcal mol ⁻¹ M ⁻¹)	$\Delta G^{\circ}_{\text{IU}}(\text{H}_2\text{O})$ (kcal mol ⁻¹)	$-m_{\text{IU}}$ (kcal mol ⁻¹ M ⁻¹)
WT	5.60 ± 0.99	2.46 ± 0.42	12.39 ± 0.60	2.60 ± 0.13
S82D	8.46 ± 1.57	3.45 ± 0.71	14.51 ± 0.69	3.56 ± 0.17

a. Buffer conditions: 10 mM potassium phosphate, 0.2 mM K₂EDTA, 1 mM βME at pH 7.0.

REFERENCES

- Alahuhta, M., Salin, M., Casteleijn, M.G., Kemmer, C., El-Sayed, I., Augustyns, K., Neubauer, P., and Wierenga, R.K. 2008. Structure-based protein engineering efforts with a monomeric TIM variant: the importance of a single point mutation for generating an active site with suitable binding properties. *Protein Eng Des Sel* **21**(4): 257-266.
- Altamirano, M.M., Blackburn, J.M., Aguayo, C., and Fersht, A.R. 2000. Directed evolution of new catalytic activity using the alpha/beta-barrel scaffold. *Nature* **403**(6770): 617-622.
- Anderson, D.E., Bechtel, W.J., and Dahlquist, F.W. 1990. pH-induced denaturation of proteins: a single salt bridge contributes 3-5 kcal/mol to the free energy of folding of T4 lysozyme. *Biochemistry* **29**(9): 2403-2408.
- Andreotti, G., Tutino, M.L., Sannia, G., Marino, G., and Cubellis, M.V. 1994. Indole-3-glycerol-phosphate synthase from *Sulfolobus solfataricus* as a model for studying thermostable TIM-barrel enzymes. *Biochim Biophys Acta* **1208**(2): 310-315.
- Anfinsen, C.B. 1973. Principles that govern the folding of protein chains. *Science* **181**(96): 223-230.
- Aurora, R. and Rose, G.D. 1998. Helix capping. *Protein Sci* **7**(1): 21-38.
- Bai, Y., Sosnick, T.R., Mayne, L., and Englander, S.W. 1995. Protein folding intermediates: native-state hydrogen exchange. *Science* **269**(5221): 192-197.
- Baker, E.N. and Hubbard, R.E. 1984. Hydrogen bonding in globular proteins. *Prog Biophys Mol Biol* **44**(2): 97-179.
- Barrick, D. 2009. What have we learned from the studies of two-state folders, and what are the unanswered questions about two-state protein folding? *Phys Biol* **6**(1): 15001.
- Bartlett, A.I. and Radford, S.E. 2009. An expanding arsenal of experimental methods yields an explosion of insights into protein folding mechanisms. *Nat Struct Mol Biol* **16**(6): 582-588.
- Beismann-Driemeyer, S. and Sterner, R. 2001. Imidazole glycerol phosphate synthase from *Thermotoga maritima*. Quaternary structure, steady-state kinetics, and reaction mechanism of the hienzyme complex. *J Biol Chem* **276**(23): 20387-20396.
- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. 2000. The Protein Data Bank. *Nucleic Acids Res* **28**(1): 235-242.
- Bilsel, O., Yang, L., Zitzewitz, J.A., Beechem, J.M., and Matthews, C.R. 1999a. Time-resolved fluorescence anisotropy study of the refolding reaction of the alpha-subunit of tryptophan synthase reveals nonmonotonic behavior of the rotational correlation time. *Biochemistry* **38**(13): 4177-4187.
- Bilsel, O., Zitzewitz, J.A., Bowers, K.E., and Matthews, C.R. 1999b. Folding mechanism of the alpha-subunit of tryptophan synthase, an alpha/beta barrel protein: global

- analysis highlights the interconversion of multiple native, intermediate, and unfolded forms through parallel channels. *Biochemistry* **38**(3): 1018-1029.
- Bolen, D.W. and Rose, G.D. 2008. Structure and energetics of the hydrogen-bonded backbone in protein folding. *Annu Rev Biochem* **77**: 339-362.
- Branden, C. and Tooze, J. 1999. *Introduction to Protein Structure*. Garland Science Publishing, New York.
- Chellgren, B.W. and Creamer, T.P. 2006. Side-chain entropy effects on protein secondary structure formation. *Proteins* **62**(2): 411-420.
- Chothia, C. 1976. The nature of the accessible and buried surfaces in proteins. *J Mol Biol* **105**(1): 1-12.
- Cuff, A.L., Sillitoe, I., Lewis, T., Redfern, O.C., Garratt, R., Thornton, J., and Orengo, C.A. 2009. The CATH classification revisited--architectures reviewed and new ways to characterize structural divergence in superfamilies. *Nucleic Acids Res* **37**(Database issue): D310-314.
- Delano, W.L. 2002. *The PyMOL Molecular Graphics System*. DeLano Scientific.
- Dill, K.A. 1990. Dominant forces in protein folding. *Biochemistry* **29**(31): 7133-7155.
- Englander, S.W. and Kallenbach, N.R. 1983. Hydrogen exchange and structural dynamics of proteins and nucleic acids. *Q Rev Biophys* **16**(4): 521-655.
- Eriksson, A.E., Baase, W.A., Zhang, X.J., Heinz, D.W., Blaber, M., Baldwin, E.P., and Matthews, B.W. 1992. Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science* **255**(5041): 178-183.
- FarzadFard, F., Gharaei, N., Pezeshk, H., and Marashi, S.-A. 2008. [beta]-Sheet capping: Signals that initiate and terminate [beta]-sheet formation. *Journal of Structural Biology* **161**(1): 101-110.
- Forsyth, W.R. and 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**(5): 1119-1133.
- Frenkel, Z.M. and Trifonov, E.N. 2005. Closed loops of TIM barrel protein fold. *J Biomol Struct Dyn* **22**(6): 643-656.
- Gao, J., Bosco, D.A., Powers, E.T., and Kelly, J.W. 2009. Localized thermodynamic coupling between hydrogen bonding and microenvironment polarity substantially stabilizes proteins. *Nat Struct Mol Biol* **16**(7): 684-690.
- Gerlt, J.A. and Babbitt, P.C. 2009. Enzyme (re)design: lessons from natural evolution and computation. *Curr Opin Chem Biol*.
- Gerstein, M. 1997. A structural census of genomes: comparing bacterial, eukaryotic, and archaeal genomes in terms of protein structure. *J Mol Biol* **274**(4): 562-576.
- Gromiha, M.M., Pujadas, G., Magyar, C., Selvaraj, S., and Simon, I. 2004. Locating the stabilizing residues in (alpha/beta)₈ barrel proteins based on hydrophobicity, long-range interactions, and sequence conservation. *Proteins* **55**(2): 316-329.
- Gu, Z., Rao, M.K., Forsyth, W.R., Finke, J.M., and Matthews, C.R. 2007a. Structural analysis of kinetic folding intermediates for a TIM barrel protein, indole-3-glycerol phosphate synthase, by hydrogen exchange mass spectrometry and Go model simulation. *J Mol Biol* **374**(2): 528-546.

- Gu, Z., Zitzewitz, J.A., and Matthews, C.R. 2007b. 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 *Sulfolobus solfataricus*. *J Mol Biol* **368**(2): 582-594.
- Gualfetti, P.J., Bilsel, O., and Matthews, C.R. 1999. The progressive development of structure and stability during the equilibrium folding of the alpha subunit of tryptophan synthase from *Escherichia coli*. *Protein Sci* **8**(8): 1623-1635.
- Hartl, F.U. and Hayer-Hartl, M. 2009. Converging concepts of protein folding in vitro and in vivo. *Nat Struct Mol Biol* **16**(6): 574-581.
- Higgins, W., Fairwell, T., and Miles, E.W. 1979a. An active proteolytic derivative of the α subunit of tryptophan synthase. Identification of the site of cleavage and characterization of the fragments. *Biochemistry* **18**(22): 4827-4835.
- . 1979b. An active proteolytic derivative of the alpha subunit of tryptophan synthase. Identification of the site of cleavage and characterization of the fragments. *Biochemistry* **18**(22): 4827-4835.
- Hocker, B., Schmidt, S., and Sterner, R. 2002. A common evolutionary origin of two elementary enzyme folds. *FEBS Lett* **510**(3): 133-135.
- Horovitz, A., Serrano, L., Avron, B., Bycroft, M., and Fersht, A.R. 1990. Strength and co-operativity of contributions of surface salt bridges to protein stability. *J Mol Biol* **216**(4): 1031-1044.
- Hyde, C.C., Ahmed, S.A., Padlan, E.A., Miles, E.W., and Davies, D.R. 1988. Three-dimensional structure of the tryptophan synthase alpha 2 beta 2 multienzyme complex from *Salmonella typhimurium*. *J Biol Chem* **263**(33): 17857-17871.
- Ibarra-Molero, B., Zitzewitz, J.A., and Matthews, C.R. 2004. Salt-bridges can stabilize but do not accelerate the folding of the homodimeric coiled-coil peptide GCN4-p1. *J Mol Biol* **336**(5): 989-996.
- Ivens, A., Mayans, O., Szadkowski, H., Jurgens, C., Wilmanns, M., and Kirschner, K. 2002. Stabilization of a (betaalpha)₈-barrel protein by an engineered disulfide bridge. *Eur J Biochem* **269**(4): 1145-1153.
- Jackson, S.E., Moracci, M., elMasry, N., Johnson, C.M., and Fersht, A.R. 1993. Effect of cavity-creating mutations in the hydrophobic core of chymotrypsin inhibitor 2. *Biochemistry* **32**(42): 11259-11269.
- Jimenez, M.A., Munoz, V., Rico, M., and Serrano, L. 1994. Helix stop and start signals in peptides and proteins. The capping box does not necessarily prevent helix elongation. *J Mol Biol* **242**(4): 487-496.
- Joerger, A.C. and Fersht, A.R. 2008. Structural biology of the tumor suppressor p53. *Annu Rev Biochem* **77**: 557-582.
- Kabsch, W. and Sander, C. 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**(12): 2577-2637.
- Kauzmann, W. 1959. Some factors in the interpretation of protein denaturation. *Adv Protein Chem* **14**: 1-63.
- Krishna, M.M., Hoang, L., Lin, Y., and Englander, S.W. 2004. Hydrogen exchange methods to study protein folding. *Methods* **34**(1): 51-64.

- Liu, A., Hu, W., Majumdar, A., Rosen, M.K., and Patel, D.J. 2000. Detection of very weak side chain-main chain hydrogen bonding interactions in medium-size $^{13}\text{C}/^{15}\text{N}$ -labeled proteins by sensitivity-enhanced NMR spectroscopy. *J Biomol NMR* **17**(1): 79-82.
- Matthews, C.R. 1987. Effect of point mutations on the folding of globular proteins. *Methods Enzymol* **154**: 498-511.
- Matthews, C.R. and Crisanti, M.M. 1981. Urea-induced unfolding of the alpha subunit of tryptophan synthase: evidence for a multistate process. *Biochemistry* **20**(4): 784-792.
- McDonald, I.K. and Thornton, J.M. 1994. Satisfying hydrogen bonding potential in proteins. *J Mol Biol* **238**(5): 777-793.
- Mirsky, A.E. and Pauling, L. 1936. On the Structure of Native, Denatured, and Coagulated Proteins. *Proc Natl Acad Sci U S A* **22**(7): 439-447.
- Mizuguchi, K., Deane, C.M., Blundell, T.L., and Overington, J.P. 1998. HOMSTRAD: a database of protein structure alignments for homologous families. *Protein Sci* **7**(11): 2469-2471.
- Murphy, K.P. 2001. Stabilization of protein structure. *Methods Mol Biol* **168**: 1-16.
- Murzin, A.G., Brenner, S.E., Hubbard, T., and Chothia, C. 1995. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol* **247**(4): 536-540.
- Myers, J.K. and Pace, C.N. 1996. Hydrogen bonding stabilizes globular proteins. *Biophys J* **71**(4): 2033-2039.
- 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**(10): 2072-2084.
- Nagano, N., Orengo, C.A., and 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**(5): 741-765.
- Nishio, K., Morimoto, Y., Ishizuka, M., Ogasahara, K., Tsukihara, T., and Yutani, K. 2005. Conformational changes in the alpha-subunit coupled to binding of the beta 2-subunit of tryptophan synthase from *Escherichia coli*: crystal structure of the tryptophan synthase alpha-subunit alone. *Biochemistry* **44**(4): 1184-1192.
- Pace, C.N., Grimsley, G.R., and Scholtz, J.M. 2009. Protein ionizable groups: pK values and their contribution to protein stability and solubility. *J Biol Chem* **284**(20): 13285-13289.
- Pace, C.N., Shirley, B.A., McNutt, M., and Gajiwala, K. 1996. Forces contributing to the conformational stability of proteins. *FASEB J* **10**(1): 75-83.
- Pace, C.N., Trevino, S., Prabhakaran, E., and Scholtz, J.M. 2004. Protein structure, stability and solubility in water and other solvents. *Philos Trans R Soc Lond B Biol Sci* **359**(1448): 1225-1234; discussion 1234-1225.
- Presta, L.G. and Rose, G.D. 1988. Helix signals in proteins. *Science* **240**(4859): 1632-1641.

- Radzicka, A. and 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**(5): 1664-1670.
- Rojsajjakul, T., Wintrode, P., Vadrevu, R., Robert Matthews, C., and 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**(1): 241-253.
- Rothlisberger, D., Khersonsky, O., Wollacott, A.M., Jiang, L., DeChancie, J., Betker, J., Gallaher, J.L., Althoff, E.A., Zanghellini, A., Dym, O. et al. 2008. Kemp elimination catalysts by computational enzyme design. *Nature* **453**(7192): 190-195.
- Saab-Rincon, G., Froebe, C.L., and Matthews, C.R. 1993. Urea-induced unfolding of the alpha subunit of tryptophan synthase: one-dimensional proton NMR evidence for residual structure near histidine-92 at high denaturant concentration. *Biochemistry* **32**(50): 13981-13990.
- Sanchez del Pino, M.M. and Fersht, A.R. 1997. Nonsequential unfolding of the alpha/beta barrel protein indole-3-glycerol-phosphate synthase. *Biochemistry* **36**(18): 5560-5565.
- Schneider, B., Knochel, T., Darimont, B., Hennig, M., Dietrich, S., Babinger, K., Kirschner, K., and Sterner, R. 2005. Role of the N-terminal extension of the (betaalpha)8-barrel enzyme indole-3-glycerol phosphate synthase for its fold, stability, and catalytic activity. *Biochemistry* **44**(50): 16405-16412.
- Schwehm, J.M., Fitch, C.A., Dang, B.N., Garcia-Moreno, E.B., and Stites, W.E. 2003. Changes in stability upon charge reversal and neutralization substitution in staphylococcal nuclease are dominated by favorable electrostatic effects. *Biochemistry* **42**(4): 1118-1128.
- Serrano, L., Kellis, J.T., Jr., Cann, P., Matouschek, A., and Fersht, A.R. 1992. The folding of an enzyme. II. Substructure of barnase and the contribution of different interactions to protein stability. *J Mol Biol* **224**(3): 783-804.
- Shirley, B.A., Stanssens, P., Hahn, U., and Pace, C.N. 1992. Contribution of hydrogen bonding to the conformational stability of ribonuclease T1. *Biochemistry* **31**(3): 725-732.
- Shortle, D., Stites, W.E., and Meeker, A.K. 1990. Contributions of the large hydrophobic amino acids to the stability of staphylococcal nuclease. *Biochemistry* **29**(35): 8033-8041.
- Soto, C. and Estrada, L.D. 2008. Protein misfolding and neurodegeneration. *Arch Neurol* **65**(2): 184-189.
- Sterner, R. and Hocker, B. 2005. Catalytic versatility, stability, and evolution of the (betaalpha)8-barrel enzyme fold. *Chem Rev* **105**(11): 4038-4055.
- Stickle, D.F., Presta, L.G., Dill, K.A., and Rose, G.D. 1992. Hydrogen bonding in globular proteins. *J Mol Biol* **226**(4): 1143-1159.
- Trevino, S.R., Gokulan, K., Newsom, S., Thurlkill, R.L., Shaw, K.L., Mitkevich, V.A., Makarov, A.A., Sacchettini, J.C., Scholtz, J.M., and Pace, C.N. 2005. Asp79

- makes a large, unfavorable contribution to the stability of RNase Sa. *J Mol Biol* **354**(4): 967-978.
- Trevino, S.R., Scholtz, J.M., and Pace, C.N. 2008. Measuring and increasing protein solubility. *J Pharm Sci* **97**(10): 4155-4166.
- Urfer, R. and Kirschner, K. 1992. The importance of surface loops for stabilizing an eightfold beta alpha barrel protein. *Protein Sci* **1**(1): 31-45.
- Vadrevu, R., Falzone, C.J., and Matthews, C.R. 2003. Partial NMR assignments and secondary structure mapping of the isolated alpha subunit of Escherichia coli tryptophan synthase, a 29-kD TIM barrel protein. *Protein Sci* **12**(1): 185-191.
- Vadrevu, R., Wu, Y., and 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**(1): 294-306.
- Vaughan, C.K., Harryson, P., Buckle, A.M., and Fersht, A.R. 2002. A structural double-mutant cycle: estimating the strength of a buried salt bridge in barnase. *Acta Crystallogr D Biol Crystallogr* **58**(Pt 4): 591-600.
- Vick, J.E. and Gerlt, J.A. 2007. Evolutionary potential of (beta/alpha)₈-barrels: stepwise evolution of a "new" reaction in the enolase superfamily. *Biochemistry* **46**(50): 14589-14597.
- Wiederstein, M. and Sippl, M.J. 2005. Protein sequence randomization: efficient estimation of protein stability using knowledge-based potentials. *J Mol Biol* **345**(5): 1199-1212.
- Wierenga, R.K. 2001. The TIM-barrel fold: a versatile framework for efficient enzymes. *FEBS Lett* **492**(3): 193-198.
- Williams, J.C., Zeelen, J.P., Neubauer, G., Vriend, G., Backmann, J., Michels, P.A., Lambeir, A.M., and Wierenga, R.K. 1999. Structural and mutagenesis studies of leishmania triosephosphate isomerase: a point mutation can convert a mesophilic enzyme into a superstable enzyme without losing catalytic power. *Protein Eng* **12**(3): 243-250.
- Wilmanns, M., Priestle, J.P., Niermann, T., and Jansonius, J.N. 1992. Three-dimensional structure of the bifunctional enzyme phosphoribosylanthranilate isomerase: indoleglycerolphosphate synthase from Escherichia coli refined at 2.0 Å resolution. *J Mol Biol* **223**(2): 477-507.
- Winklhofer, K.F., Tatzelt, J., and Haass, C. 2008. The two faces of protein misfolding: gain- and loss-of-function in neurodegenerative diseases. *EMBO J* **27**(2): 336-349.
- Wu, Y. and Matthews, C.R. 2002a. A cis-prolyl peptide bond isomerization dominates the folding of the alpha subunit of Trp synthase, a TIM barrel protein. *J Mol Biol* **322**(1): 7-13.
- 2002b. 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**(2): 309-325.
- 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**(5): 1131-1144.

- Wu, Y., Vadrevu, R., Kathuria, S., Yang, X., and Matthews, C.R. 2007. A tightly packed hydrophobic cluster directs the formation of an off-pathway sub-millisecond folding intermediate in the alpha subunit of tryptophan synthase, a TIM barrel protein. *J Mol Biol* **366**(5): 1624-1638.
- Xie, G., Forst, C., Bonner, C., and Jensen, R.A. 2002. Significance of two distinct types of tryptophan synthase beta chain in Bacteria, Archaea and higher plants. *Genome Biol* **3**(1): RESEARCH0004.
- Yang, X., Vadrevu, R., Wu, Y., and Matthews, C.R. 2007. Long-range side-chain-main-chain interactions play crucial roles in stabilizing the (betaalpha)₈ barrel motif of the alpha subunit of tryptophan synthase. *Protein Sci* **16**(7): 1398-1409.
- Zitzewitz, J.A., Gualfetti, P.J., Perkons, I.A., Wasta, S.A., and Matthews, C.R. 1999. Identifying the structural boundaries of independent folding domains in the α subunit of tryptophan synthase, a β/α barrel protein. *Protein Sci* **8**(6): 1200-1209.
- Zitzewitz, J.A. and Matthews, C.R. 1999. Molecular dissection of the folding mechanism of the alpha subunit of tryptophan synthase: an amino-terminal autonomous folding unit controls several rate-limiting steps in the folding of a single domain protein. *Biochemistry* **38**(31): 10205-10214.