EXPERIMENTAL AND COMPUTATIONAL STUDIES ON PROTEIN FOLDING, MISFOLDING AND STABILITY

A Dissertation

by

YUN WEI

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2009

Major Subject: Biochemistry

EXPERIMENTAL AND COMPUTATIONAL STUDIES ON PROTEIN FOLDING, MISFOLDING AND STABILITY

A Dissertation

by

YUN WEI

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	J. Martin Scholtz
Committee Members,	C. Nick Pace
	Mary Bryk
	Jerry Tsai
Head of Department,	Gregory D. Reinhart

May 2009

Major Subject: Biochemistry

ABSTRACT

Experimental and Computational Studies on

Protein Folding, Misfolding and Stability. (May 2009)

Yun Wei, B.S., Wuhan University

Chair of Advisory Committee: Dr. J. Martin Scholtz

Proteins need fold to perform their biological function. Thus, understanding how proteins fold could be the key to understanding life. In the first study, the stability and structure of several β-hairpin peptide variants derived from the C-terminus of the B1 domain of protein G (PGB1) were investigated by a number of experimental and computational techniques. Our analysis shows that the structure and stability of this hairpin can be greatly affected by one or a few simple mutations. For example, removing an unfavorable charge near the N-terminus of the peptide (Glu42 to Gln or Thr) or optimization of the N-terminal charge-charge interactions (Gly41 to Lys) both stabilize the peptide, even in water. Furthermore, a simple replacement of a charged residue in the turn (Asp47 to Ala) changes the β-turn conformation. Our results indicate that the structure and stability of this β–hairpin peptide can be modulated in numerous ways and thus contributes towards a more complete understanding of this important model β-hairpin as well as to the folding and stability of larger peptides and proteins.

The second study revealed that PGB1 and its variants can form amyloid fibrils *in vitro* under certain conditions and these fibrils resemble those from other proteins that

have been implicated in diseases. To gain a further understanding of molecular mechanism of PGB1 amyloid formation, we designed a set of variants with mutations that change the local secondary structure propensity in PGB1, but have similar global conformational stability. The kinetics of amyloid formation of all these variants have been studied and compared. Our results show that different locations of even a single mutation can have a dramatic effect on PGB1 amyloid formation, which is in sharp contrast with a previous report. Our results also suggest that the α-helix in PGB1 plays an important role in the amyloid formation process of PGB1.

In the final study, we investigate the forces that contribute to protein stability in a very general manner. Based on what we have learned about the major forces that contribute to the stability of globular proteins, protein stability should increase as the size of the protein increases. This is not observed: the conformational stability of globular proteins is independent of protein size. In an effort to understand why large proteins are not more stable than small proteins, twenty single-domain globular proteins ranging in size from 35 to 470 residues have been analyzed. Our study shows that nature buries more charged groups and more non-hydrogen-bonded polar groups to destabilize large proteins.

DEDICATION

This dissertation is dedicated to my mother, Manxiang Tang.

Her constant love and support have sustained me throughout my life.

ACKNOWLEDGEMENTS

I would like to gratefully and sincerely thank my advisor, Dr. J. Martin Scholtz for his guidance, understanding, patience, consistent encouragement and considerable freedom on the direction of my research. I would like to thank Dr. Nick Pace for his enormously valuable advice on my research. His tenacity and passion for research has certainly inspired me. I would like to thank Dr. Jerry Tsai for his guidance and great support on the computational part of the β-hairpin project. I also thank Dr. Mary Bryk for her advice and assistance throughout the course of my research and Dr. Sarah E. Bondos for her time and help. I thank Dr. Beatrice Huyghues-Despointes, Dr. Abbas Razvi, Dr. Saul Trevino, Dr. Xiangming Kong, Dr. Lisa M. Perez and Xiaotao Qu for their generosity with their time and expertise. I thank all the members, former and current, of the Pace-Scholtz lab for their friendship and support. I also thank the staff of the Molecular and Cellular Medicine Department and Biochemistry Department for their great help over the years.

Finally, thanks to my parents, my elder sister, Qin, and my friends for their love and support over the years.

TABLE OF CONTENTS

Page

ABSTRACT iii

DEDICATION v

ACKNOWLEDGEMENTS vi

TABLE OF CONTENTS vii

LIST OF FIGURES ix

LIST OF TABLES xii

CHAPTER

I INTRODUCTION 1

II NMR STUDY AND MOLECULAR DYNAMICS SIMULATIONS OF OPTIMIZED β-HAIRPIN FRAGMENTS OF PROTEIN G 3

- Introduction 3
- Materials and Methods 8
- Results and Discussion 11
- Conclusions 34
- Supplemental Materials 34

III A KINETIC STUDY OF AMYLOID FORMATION IN PROTEIN PGB1 39

- Introduction 39
- Materials and Methods 45
- Results and Discussion 49

IV WHY LARGE PROTEINS ARE NOT MORE STABLE THAN SMALL PROTEINS 63

- Introduction 63
- Materials and Methods 65
- Results and Discussion 66

CHAPTER Page

Conclusions 96
Supplemental Materials 99

V SUMMARY 102

REFERENCES 105

VITA 132

LIST OF FIGURES

FIGURE		Page
1	Thermal Unfolding Curves Based on the Chemical Shifts Change of the Tyr45-Hδ with Temperature at pH 7 for Each Peptide: G41(wt) (Circles), K41 (wt*)(Squares), K41+A47 (Diamonds), K41+T42 (Crosses), K41+Q42+A47 (Triangles), K41+T42+A47 (Pluses) and K41+T42+A47 With 20% Methanol (Filled Circles)	15
2	Deviations in CαH Chemical Shifts from Random Coil Values for Each Residue in G41, K41, K41+Q42+A47 and in the C-Terminal Region of PGB1 in D_2O at pH 7 and 4ºC	18
3	Deviations in CαH and H_N Random Coil Values for the Turn Residues In G41, K41, K41+Q42+A47, and the Native PGB1	19
4	A Summary of the NOE Crosspeaks Found in K41+Q42+A47 in 5 mM Sodium Phosphate 10% D_2O Solutions at pH 7 and 4 ºC	21
5	A Summary of NOE Interactions Observed in the K41+Q42+A47 Peptide but Not in K41 or G41 (Thin Arrows), and the Enhanced NOE Interactions Relative to Those Observed in K41 or G41 (Thick Arrows)	22
6	Structural Depictions of the Native β-Turn of 4:4 Hairpin Conformation of G-Hairpin in PGB1 (A) and the β-Turn of the Nonnative 3:5 Hairpin (B), Generated from MD Simulations.	27
7	Representative of β-Hairpin Conformations of K41+Q42+A47 Selected from MD-Generated Structures Based on Their Similarity with the NOE Data.	27
8	Comparison of Probability Distributions of the Normalized CαRMSD in the MD-Gererated Ensembles in the β-Turn Region (Residue 47-50)..	31
9	The Ramachandran Plot Showing the Main-chain ϕ, ψ Angle Distributions for the Five-Residue Loop (Residue 47 to 51) in K41+Q42+A47 Generated from MD Simulations	33

FIGURE Page

10 Portions of the 500 MHz NOESY Spectrum of K41+Q42+A47 Peptide Recorded in 100% D_2O at 5 mM Sodium Phosphate Buffer pH 7 and 4ºC (t_m = 200 ms). 36

11 Amide and Aromatic Protons Region of the 500 MHz NOESY Spectrum of K41+Q42+A47 Peptide Recorded in Water (10% D_2O) at 5 mM Sodium Phosphate Buffer pH 7 and 4 ºC..... 38

12 A Ribbon Diagram of PGB1. 43

13 Illustration of the Time Course for Amyloid Formation in PGB1. 48

14 Influence of Seeding and Ionic Strength on PGB1 Amyloid Formation Monitored by UV 280 Sedimentation Method..... 52

15 Influence of pH on PGB1 Amyloid Formation Monitored by UV 280 Sedimentation Method..... 53

16 Fibril Formation of A34F in 50mM Sodium Citrate, 27% TFE, pH 3.75 at 37ºC with Agitation Displays a Two-state Curve. 56

17 Effect of Adding the β-Hairpin Peptide K41T42 on Kinetics of D40M PGB1 Amyloid Formation. 61

18 Electron Micrographs of a Representative Set of PGB1 Mutant Fibrils.... 62

19 The Predicted Stability for the 20 Proteins Compared to Experimentally Measured Stability 80

20 The Number of Buried Non-hydrogen-bonded Polar Groups Increases with the Increase of Protein Size..... 86

21 Comparison of the Predicted Stability after Energetic Cost Corrections with the Measured Stability for the 20 Proteins..... 87

22 Calculated van der Waals and Electrostatic Interaction Energy. 91

23 Comparison of the Trend of the Increase of Packing Densities (Calculated By NSC Method) with Protein Size with the Trend of the Increase of vdW Interactions Per Residue..... 100

FIGURE Page

24 The Relationship Between Experimental and Predicted m-Values as a Function of Protein Size 101

LIST OF TABLES

TABLE		Page
1	Thermal Unfolding and Sequences of G-Hairpin Peptide Variants	13
2	Hydrophobic Contact Surface Area Between β-Hairpin Strands	28
3	Terminal Ion Pair (IP) Interactions in the Ensembles of MD-Generated Structures	29
4	Chemical Shift Assignments for K41+Q42+A47 Peptide	35
5	Lag Time and Growth Rate Constant for D40M PGB1 and Variants	60
6	Summary of Studied Proteins	68
7	Estimated Energetic Contribution to the Conformational Stability for Each Factor.	72
8	Estimated Conformational Stability of All Proteins by Using Different Combinations of Methods	81
9	Groups Burials for the 20 Proteins	82
10	Burial of Charged Groups for All the Proteins	83
11	Estimated Conformational Stability Before and After Corrections for the Destabilizing Forces	89
12	Calculated Vacuum Enthalpy of Unfolding, ΔH^{U}_{N}(vac), and Its Decomposition on Caloric/Gram Basis	90
13	Calculated Packing Densities for All the Proteins	99

CHAPTER I

INTRODUCTION

Globular proteins need to fold into a unique, often intricate and close-packed, three-dimensional structure to perform their biological functions. These functions include the control and regulation of essentially every biological process on which our lives depend. A fundamental biological process is thus the protein folding itself, the process through which disordered polypeptides convert into their native folded structures to function. Although the importance of this folding process has been recognized for many years, the underlying mechanism by which the complex process takes place is still not completely understood. As a result of the application of a wide range of biophysical and computational methods used in combination with biochemical and protein engineering techniques, much has been learned about protein folding in the past few decades. Of note is the fact that much has been learned by studying peptide and protein model systems. Short peptides that can fold into the two regular minimal secondary structures, α-helix and β-hairpin, are used to study the specific noncovalent interactions that have been identified crucial to protein folding, including hydrophobic interactions, hydrogen bonding and electrostatic interactions. The first chapter of this dissertation focuses on examining the folding and stability of a naturally occurring β-hairpin with a combination of experimental and theoretical approaches. The third chapter of this

---

This dissertation follows the style of *Journal of Molecular Biology*.

dissertation provides a theoretical study that tries to unravel the mystery of why the stability of large proteins is similar to that of small proteins.

Proteins fold to function, but lack of function is not always the worst scenario. Over the last few years, we have learned that a misfolded protein is potentially worse than proteins that do not perform their biological functions because a misfolded protein can poison the cells. Recent discoveries show that some apparently unrelated devastating diseases all result from protein misfolding, including Alzheimer's, Parkinson's disease, Mad Cow disease and Cystic fibrosis. Amyloid fibrils and related protein aggregations have therefore been a critical subject of recent studies. Because the formation of amyloid fibrils is now generally believed to be a generic structural property of polypeptides, it is of great interest to investigate protein misfolding in model systems that have been extensively studied with various biochemical and biophysical approaches. In the second chapter of this dissertation, a kinetic study of amyloid formation in the well-studied model system Protein G B1 domain (PGB1) has been conducted in order to gain further insights into the molecular mechanism of amyloid fibril formation.

CHAPTER II

NMR STUDY AND MOLECULAR DYNAMICS SIMULATIONS OF OPTIMIZED β-HAIRPIN FRAGMENTS OF PROTEIN G*

INTRODUCTION

Some short peptide fragments excised from proteins can fold autonomously in aqueous solution into native-like conformations in the absence of interactions with other structural elements [1; 2]. These polypeptide fragments may thus play important roles in initiating protein folding by narrowing the conformational search path to one where local effects dominate the interactions during the early stages of protein folding [3; 4; 5]. To further understand early events in protein folding, short peptides have often been used as simple model systems to identify specific non-covalent interactions that are crucial to secondary structure formation [6; 7; 8; 9; 10; 11]. Many features of protein folding have also been observed in the folding of peptide models, including the general hydrophobic collapse, close packing of side chains and the formation of intramolecular hydrogen bonds [9; 10; 11; 12; 13]. For example, over the last couple of decades, model peptides have been used quite successfully to identify local interactions and determine their contributions to the formation and stability of the α-helix [14; 15; 16]. Similar studies on the

---

* Reprinted from *Proteins*, **69**, Wei Y., Huyghues-Despointes B.M.P., Tsai J. and Scholtz J.M. 'NMR study and molecular dynamics simulations of optimized β-hairpin fragments of protein G', pp 285-296, 2007, with the permission of Wiley-Liss, Inc. a subsidiary of John Wiley & Sons, Inc.

formation of simple β-hairpins, the minimal unit of the β-sheet structure, have lagged behind and have been hampered by the marginal stability of the structure and the strong tendency of the peptides to aggregate.

Fortunately, the last decade has seen considerable progress in our understanding of β-sheet and β-hairpin formation with the developments of well-folded 'designed β-hairpin' systems, often employing amino acid mimics as key building blocks; a recent review summarizes some of the important findings [17]. There are several key interactions and features of the peptides that appear to be critical to the folding and stability of β-sheets and β-hairpins. Briefly, the specific nature and propensity of residues in the β-turn itself has been shown to be important in defining the conformation and stability of model β-hairpin peptides. Griffiths-Jones *et al.* [18] have shown that significant β-turn conformation is still populated in a truncated variant of their model β-hairpin peptide in the absence of interstrand hydrophobic contacts, suggesting the intrinsic turn propensity plays an important role in nucleating β-hairpin folding. In a separate study with a different model peptide, de Alba *et al*. (1997) demonstrated that a peptide can fold in aqueous solution to two β-hairpin structures that can co-exist in equilibrium and single-residue substitutions in the β-turn region can dramatically alter the preferences between the β-turn conformations [19; 20; 21]. These results demonstrate the important role of the β-turn sequence in determining both the stability and conformational preferences of the β-hairpin [7; 22].

In addition to studies on designed, *de novo* β-sheet and β-hairpin peptides, there have been a few studies on peptides derived directly from proteins and thus comprised of

native residues with interactions that are probably more biologically relevant. To our knowledge, the first reported β-hairpin peptide that was demonstrated to autonomously fold into a native-like structure was the 17-residue peptide from the N-terminus of ubiquitin, however this peptide only adopts ~20% β-hairpin conformation in aqueous solution [1; 23]. A major breakthrough in the study of β-hairpin formation came when Serrano and his colleagues showed that the 16-residue peptide corresponding to the C-terminal hairpin (residues 41-56, known here as G-hairpin, proposed first by Kobayashi *et al.* [24]) of B1 domain of the *streptococcal* protein G (pdb: 1PGB) populates up to ~40% of a monomeric, native-like β-hairpin structure in aqueous solution [2; 7].

Since this original discovery, the G-hairpin has been the paradigm for β-hairpin formation using experimental (NMR, CD spectroscopy) [2; 8; 25; 26; 27; 28; 29] and computational [30; 31; 32; 33; 34; 35] approaches. These studies have led to the suggestions that there are four main factors that contribute to the folding and stability of the G-hairpin: 1) the intrinsic β-turn propensities of the turn residues [7; 36], 2) hydrophobic interactions between the side chains across the β-strands [12; 28; 37], 3) interstrand hydrogen bonds that help define and maintain the architecture of the hairpin [33], and 4) polar side-chain to side-chain (sc-sc) interactions, including electrostatic interactions and salt bridges [6; 8; 19; 25]. Among these, the interstrand hydrophobic interactions between the aromatic groups of Trp43, Tyr45, Phe52 and the nonpolar residue Val54 have been suggested to provide the largest contribution to stability and have even been implicated as the nucleation site for folding [30; 32; 35; 38].

Most previous studies using the G-hairpin as the reference model for β-hairpin structure have focused on either the interstrand hydrophobic interactions or on optimization of the whole loop sequence to increase the stability. The hydrophobic cluster of G-hairpin is formed between a valine (Val54) side chain and the rings of three aromatic residues (Trp43, Tyr45 and Phe52). In one elegant study to explore hydrophobic interactions, Cochran *et al.* have shown replacements of the other three residues in the hydrophobic cluster with tryptophan significantly stabilize the G-hairpin, thus demonstrating that these interactions could be transferred from another model system that they called the tryptophan zipper [28]. Kobayashi *et al.* showed single mutations to alanine of either Trp43, Tyr45 or Phe52 in the hydrophobic core can dramatically destabilize the G-hairpin [8].

In intact PGB1, the second β-turn region consists of a six-residue loop (DDATKT) that forms a 4:4 hairpin with a type IV turn in the protein structure according to the nomenclature of Thornton [39; 40]. The loop is closed by two hydrogen bonds formed between Asp46 and Thr51 in the protein, whereas only one hydrogen bond between Asp46 NH and Thr51 carboxyl is possible in the peptide. The β-turn of G-hairpin is a long loop and presumably more flexible and entropically unfavorable compared to short and rigid β-turns [29; 36]. Fesinmeyer *et al*. has shown that replacing the native turn with NPATGK, a sequence selected by comparisons to known turns, can substantially enhance the stability of the resulting G-hairpin variant [29]. In fact, replacement of the six-residue loop with the shorter turn-favoring sequence Val-$_D$Pro-Gly-Lys in the context of hydrophobic cluster from the G-hairpin is stabilizing [36; 41].

McCallister *et al.* have examined the effects of replacing the turn residues Asp46, Asp47, Thr49 or Thr51 with alanine in the protein G and the D47A mutation is the only one found to stabilize the protein [42]. Also, the replacement of Asp47 at turn position *i* with Ala is the only single loop mutation reported that stabilizes the G-hairpin [8; 42]. The fact that the D47A mutation stabilizes the G-hairpin peptide is quite unexpected, because the D47A mutation removes not only a salt bridge with Lys50 across the turn, but also a hydrogen bond between Asp47 and Tyr45. However, no explanation for the increased in the stability of peptide has been given.

Single amino acid substitutions that reverse or remove the charges on solvent-exposed side chains have been shown to sometimes increase protein stability without altering the structure [43; 44]. Similarly, the pH and salt-dependence of β-hairpin formation in a number of model systems suggest that long- and short-range electrostatic interactions across β-strands can also contribute to stability [25; 45; 46; 47]. In both designed peptides and G-hairpin, enhanced electrostatic interactions at the terminal positions of the peptide can contribute to the stability of the β-hairpin [25; 48]. The introduction of cross-strand Lys-Glu ion pairs near the terminal regions of designed β-hairpins also show increased stability [45; 48]. However, these mutations might also create additional hydrophobic interactions due to the long hydrophobic side chain of Lys and it is difficult to separate these contributions to stability from the ion-pair interactions.

Our goal in this study is to combine known mutations with a few new ones to generate a more stable G-hairpin and determine the energetic contributions of the different types of interactions. To gain further insights into how long-range coulombic

interactions and hydrophobic interactions stabilize the β-hairpin, we investigated several replacements (E42T/E42Q/D47A) that remove the unfavorable interactions without introducing new hydrophobic interactions in the G-hairpin model system. Since the turn has such an important role in β-hairpin folding, we also provide structural studies of the peptides using 2D NMR spectroscopy. Surprisingly, we find that single mutations can have rather large effects on not only stability (not so unexpected), but also on structure. This observation appears to be more pronounced in β-strands than in α-helical peptide models or in proteins, where single amino acid changes almost never result in large or even detectable structural changes. Molecular dynamics modeling for each peptide variant was also performed to give us insights into structural features and to support and augment the NMR studies.

MATERIALS AND METHODS

Peptides Synthesis and Purification. The peptides were synthesized using Fmoc chemistry with HBTU activation and standard solid-phase methods on Wang Resin. Peptides were cleaved from the resin and deprotected by using a mixture of 90% TFA, 5% thioanisole, 3% ethanedithiol and 2% anisole. The peptides were purified using reversed-phase FPLC on a Pharmacia Source 15RPC column with two gradient runs of acetonitrile and 5 mM sodium phosphate or 10 mM ammonium bicarbonate[25]. The peptide identities were confirmed by MALDI mass spectrometry.

NMR Spectroscopy. The NMR samples (~1 mM) were prepared as described previously [25] in a volume of 0.75 mL either in a mixture of H_2O/D_2O (9:1 ratio by volume) or 99.8% D_2O buffered with 5 mM sodium phosphate. The pH was adjusted to 7 with a small addition of DCl or NaOD and measured by a PHM-220 pH meter with a combined glass electrode. The pH was not corrected for solvent isotope effects. All NMR experiments were acquired with 500 MHz Varian Inova spectrometers and processed with NMRPipe [49]. The temperature of the NMR probe was calibrated using the temperature dependent chemical shifts (difference between the methyl and hydroxyl resonances) of a 100% (v/v) methanol standard [50]. Sodium 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) was used as an internal reference and set to 0 ppm for all temperatures and co-solvent concentrations. The thermal dependences of ^1H chemical shifts of the peptides in D_2O at pH 7 were determine from 1D NMR spectra obtained by accumulating 128 scans with 4096 complex points and spectral width of 6000. Measurements were taken with increasing temperature from 2 to 27ºC at ~2ºC intervals and from 30 to 80ºC at ~5ºC intervals. The thermal dependence of ^1H chemical shifts of the K41+T42+A47 peptide in D_2O in 20% methanol was obtained at a wider temperature range (-2 to 80ºC). A combination of TOCSY and NOESY 2D-NMR spectra was used to assign all resonances. A mixing time of 200 ms was used for NOESY spectra in both D_2O and 9:1 mixture of H_2O/D_2O solutions. The TOCSY experiments used a mixing time of 80 ms.

Thermal Melting Curves by 1D NMR. The chemical shifts of Tyr45 Cδ protons were chosen in this study as the probe to determine the relative stabilities of G-hairpin

variants. We chose Tyr45 Cδ proton because a large signal change was observed with temperature, in agreement with what has been observed previously [25; 26]. The changes in chemical shift with temperature were analyzed assuming a two-state folding transition between folded and unfolded β-hairpins as previously described [25]. The cosolvent methanol, known to stabilize secondary structure in peptides [6], was used to help define the folded state pre-transition baseline (5.98 ppm) by following thermal transition in 20% methanol for the most stable peptide (K41+T42+A47) from temperatures below 0ºC (~ -2ºC). The unfolded baseline value (6.86 ppm) of Tyr45 Cδ proton was determined by fitting the chemical shift values of Tyr45 Cδ proton of an unfolded G-hairpin variant at high temperatures in a previous study [25; 26].

Molecular Dynamics Simulation. Using the ENCAD program and potential energies with the F3C water model [51; 52], MD simulations were performed on all peptide variants as described previously [25]. The coordinates of the starting structure used those for the β-hairpin in the crystal structure of PGB1 (pdb: 1PGB). The peptide was placed in a periodic box of water without boundaries [51; 52] and the box of water was then trimmed so that the edges were at least 8 Å away from the closest peptide atom. All waters within 1.67 Å of the peptide were then removed and the box size was adjusted to match the density of water (0.997 g/ml) at 298K [53; 54]. The system was then relaxed by performing 3,000 conjugate gradient energy minimization steps. At least ten 10 ns simulations with different random seed numbers were performed for each peptide (12 simulations for K41+T42 and 10 for other peptides). Each 10 ns simulation generated 10,000 structures and only structures generated in the last 9 ns were used to analyze the

structure and properties of the β-hairpin. The hydrophobic contact surface area (HCSA) was calculated using Voronoi polyhedra method [55; 56]. The polyhedra surrounding each atom were uniquely defined by using the Delauney tessellation [57] and two carbon atoms sharing a common polyhedron face are defined as being in contact [56]. The hydrophobic clusters were defined by the largest side-chain to side-chain contact network as previously described [25]. Clusters are defined as the set of all the molecules are in contact with at least one other molecule in the cluster.

A hydrogen bond is considered to be present if the distance between the donor hydrogen and the acceptor oxygen is 2.6 Å or less and the bond angle formed by the acceptor oxygen, hydrogen and the donor atom is greater than 120º. An ion pair is defined using 3.5 Å as the distance cutoff between the charged nitrogen of the amino group and the oxygen of the carboxyl group. A salt-bridge interaction is considered to be present if it satisfies both the hydrogen bond and ion-pair definition [25]. The Ramachandran plot showing the turn main-chain ϕ, ψ angle distributions and the turn CαRMSD distributions are represented using the R package [58].

RESULTS AND DISCUSSION

Peptide Design. As described above, G-hairpin is a good model system to investigate many aspects of β-hairpin folding and stability. Here, we extend these studies to investigate the effects of long-range coulombic interactions on β-hairpin stability. G-hairpin has several electrostatic interactions that are predicted to be unfavorable at pH 7,

so we constructed a set of four 16-residue G-hairpin variants (Table 1) that have the unfavorable electrostatic interactions removed or altered. In summary, our design strategy had the following considerations: 1) All four variants include the replacement of Gly41 with Lys (G41K). This replacement was shown to introduce a salt bridge between the side chain of Lys41 and either the C-termini or side-chain carboxyl of Glu56, thus greatly enhancing the stability of the G-hairpin [25]. 2) Glu42 is replaced with either Gln or Thr, residues with polar but uncharged side chains and relatively high β-sheet propensities [59; 60]. These substitutions were made to examine the effect of removing repulsive electrostatic interactions between Glu42 and Glu56 while still maintaining high overall β-sheet propensity. 3) A replacement of Asp47 with Ala at the first position of the β-turn, shown previously to stabilize the G-hairpin [8], is used in three out of four variants. We seek to determine how this replacement stabilizes the β-hairpin because it not only removes the unfavorable interactions between the side-chain carboxyl groups of Asp47 and Asp46, but also removes favorable interactions between the carboxyl group of Asp47 and the amino group of Lys50.

Table 1: Thermal unfolding and sequences of G-hairpin peptide variants.

Name	Sequence[a]			Net Charge at pH 7[b]	T_m(ºC)[c]
	41	47	56		
G41 (wt)	G-E-W-T-Y-D-D-A-T-K-T-F-T-V-T-E			-3	1.3[d]
K41 (wt*)	K-E-W-T-Y-D-D-A-T-K-T-F-T-V-T-E			-2	11.4[d]
K41+A47	K-E-W-T-Y-D-A-A-T-K-T-F-T-V-T-E			-1	18.3 (0.2)
K41+T42	K-T-W-T-Y-D-D-A-T-K-T-F-T-V-T-E			-1	26.4 (0.2)
K41+Q42+A47	K-Q-W-T-Y-D-A-A-T-K-T-F-T-V-T-E			0	33.0 (0.3)
K41+T42+A47	K-T-W-T-Y-D-A-A-T-K-T-F-T-V-T-E			0	35.5 (0.3)
					45.2 (0.3) (+20% methanol)

[a]The residue numbering corresponds to that for the intact protein G B1 domain.
[b]Calculated using standard pK values for the ionizable groups [61].
[c]The T_m values were determined by measuring chemical shifts of Tyr45-Hδ in 5 mM sodium phosphate in D_2O at pH 7. Values in parentheses show the standard errors in the fitting calculations.
[d]The T_m values for G41(wt) and K41(wt*) were determined under identical conditions by Huyghues-Despointes *et al.* [25]

Mutations Can Alter Thermal Stability. The folding rate of G-hairpin has been shown to be faster than the NMR time-scale (>10^3 s^{-1}) [26; 62]; therefore, the proton chemical shifts represent the populations of the folded and unfolded β-hairpin conformations at equilibrium [25]. The thermodynamics of G-hairpin [26] and several variants [8] have been determined by monitoring the temperature dependence of proton chemical shifts in 1D-NMR measurements. Importantly, Honda *et al.* have shown that the thermodynamic properties determined from the NMR probes on both the main-chain and the side-chains are similar and are identical to those determined from independent calorimetric experiments [26]. These results suggest that the thermal unfolding of G-hairpin can be adequately described by a cooperative two-state folding transition

between folded β-hairpins and the unfolded state, consistent with results found for other G-hairpin variants [28; 29]. These results provide enough evidence that the thermodynamic properties of our G-hairpin variants of equal length can be determined by this NMR method.

Here we use the chemical shift changes of the Tyr45 Cδ protons with temperature as the probe to determine the relative thermal stabilities of the G-hairpin variants as previously described [8; 25]. 1D proton NMR spectra for all the G-hairpin variants (Figure 1) were measured at pH 7 from 2 to 80ºC. As controls, the melting curves for the wild-type G-hairpin and the K41 peptide (wt*) from a previous study are also shown [25].

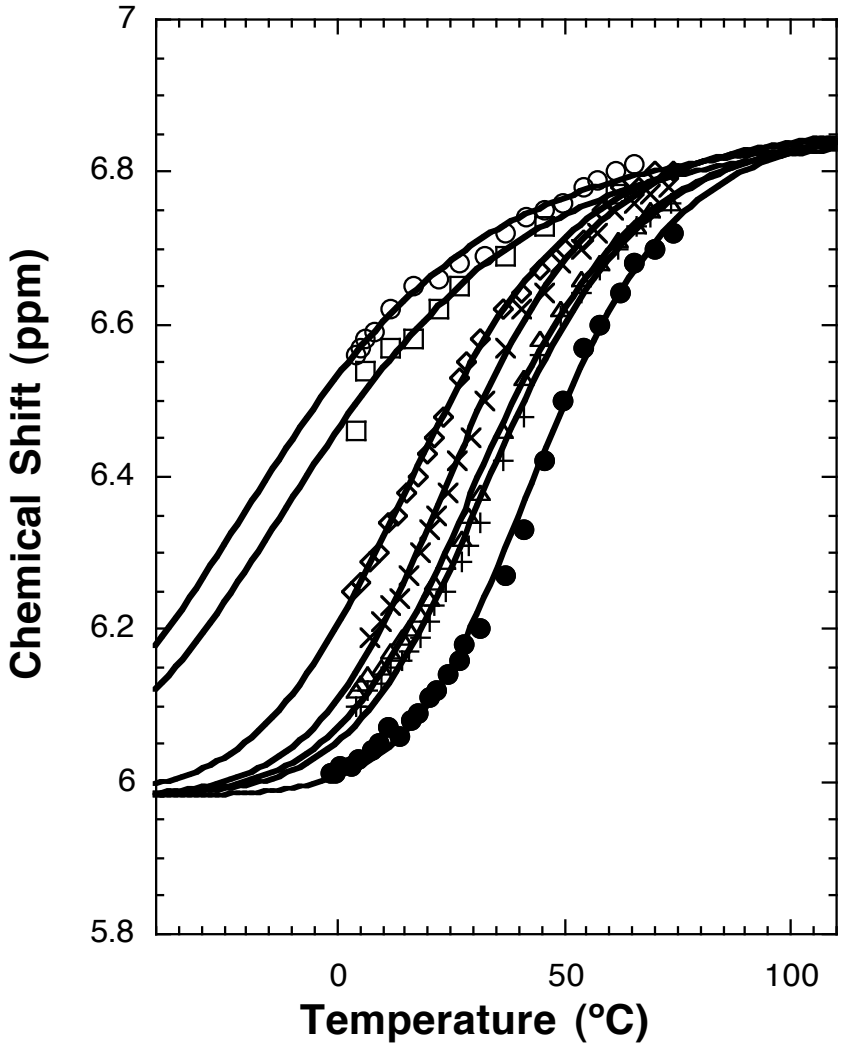

Figure 1: Thermal unfolding curves based on the chemical shifts change of the Tyr45-Hδ with temperature at pH 7 for each peptide: G41 (wt) (circles), K41 (wt*) (squares), K41+A47 (diamonds), K41+T42 (crosses), K41+Q42+A47 (triangles), K41+T42+A47 (pluses) and K41+T42+A47 with 20% methanol (filled circles).

All the melting curves show simple sigmoidal shapes, suggesting a two-state transition from folded to unfolded conformations with increasing temperature. The analysis of the thermal unfolding curves provides estimates for the melting temperatures, the apparent T_m (Table 1). Peptides with higher apparent T_m values populate more β-hairpin structure and the stability of the peptides can be ranked according to their apparent T_m values: K41+T42+A47 ≈ K41+Q42+A47 > K41+T42 > K42+A47> K41 (wt*) > wt. Single mutation D47A and G41K have been shown to stabilize the G-hairpin by ~7 ºC [8] and ~10 ºC [25] separately. For the double mutant K41+A47, an additive effect on β-hairpin stability was observed, providing an increase in apparent T_m from the wild-type G-hairpin of 17 °C [25]. A single replacement of Glu42 with Gln shows an substantial increase in apparent T_m of ~15 ºC as does the replacement of Glu42 with Thr, suggesting that a negatively charged residue is not favored at position 42.

Mutations Can Alter β-Hairpin Structure. The chemical shift dispersion in the 1D ^{1}H NMR spectra of these peptides is an excellent indicator of secondary structure formation. In particular, the deviations of the CαH chemical shifts from random coil values ($\Delta\delta_{C\alpha H}$) are useful parameters to characterize both α-helix and β-sheet structure [19; 63; 64; 65; 66]. Generally, upfield chemical shifts of CαH protons from random coil values (negative $\Delta\delta_{C\alpha H}$ values) are observed in helical and β-turn structures in proteins and peptides, whereas downfield shifts (positive $\Delta\delta_{C\alpha H}$ values) are observed in β-sheet conformations. The patterns we observe for $\Delta\delta_{C\alpha H}$ values for the peptides are consistent with two β-strands separated by a β-turn, as expected for the β-hairpin conformation [65]. In Figure 2, we compare the deviations of CαH chemical shift from random coil values

for each residue in K41+Q42+A47, G-hairpin (G41, wt) [2], K41 (wt*) [25] and the C-terminal region of the native intact PGB1 protein for comparison [67]. All the peptides have similar $\Delta\delta_{C_\alpha H}$ values for each residue to those observed in the protein, but with a reduced magnitude. As expected, the patterns of $\Delta\delta_{C_\alpha H}$ are in the same rank order in stability determined by the NMR thermal stability study (native β-hairpin in PGB1 > K41+Q42+A47 > K41 > G41). The $\Delta\delta_{C_\alpha H}$ values for residues Asp46, Ala47 and Ala48 in the β-turn (Figure 3a) show also an order that agrees with the stability rank (less negative values, less stable). The $\Delta\delta_{C_\alpha H}$ values for residues Thr49 and Lys50, however, deviate substantially from those in the native protein (Figure 3a) [67]: $\Delta\delta_{C_\alpha H}$ for Thr49 is slightly positive in the protein but negative in all the peptides. This result suggests that the G-hairpin peptides adopt a different β-turn structure from that in the protein. Additionally, a larger change in $\Delta\delta_{C_\alpha H}$ for Lys50 is observed in K41+Q42+A47 than in K41 or G41, and the $\Delta\delta_{C_\alpha H}$ patterns of G41 and K41 in the β-turn region are quite different from $\Delta\delta_{C_\alpha H}$ pattern of K41+Q42+A47 (shown in Figure 3a). To confirm the β-turn conformation change from 4:4 to 3:5 hairpin, we also examined the deviations of the backbone amide (H_N) chemical shifts from random coil values. As shown in Figure 3b, the $\Delta\delta_{NH}$ of turn region in the triple mutant K41+Q42+A47 has a similar 3:5 hairpin chemical shifts deviation pattern to other 3:5 hairpins shown by Fesinmeyer *et al.* [68]. Ala47 has a notable downfield shifted H_N comparing to that in K41 and in the native protein. In addition, the Thr49 and Thr51 have an enhanced upfield shifted H_N. All these results suggest that the D47A mutation has changed the conformations of the β-turn region of the G-hairpin variants.

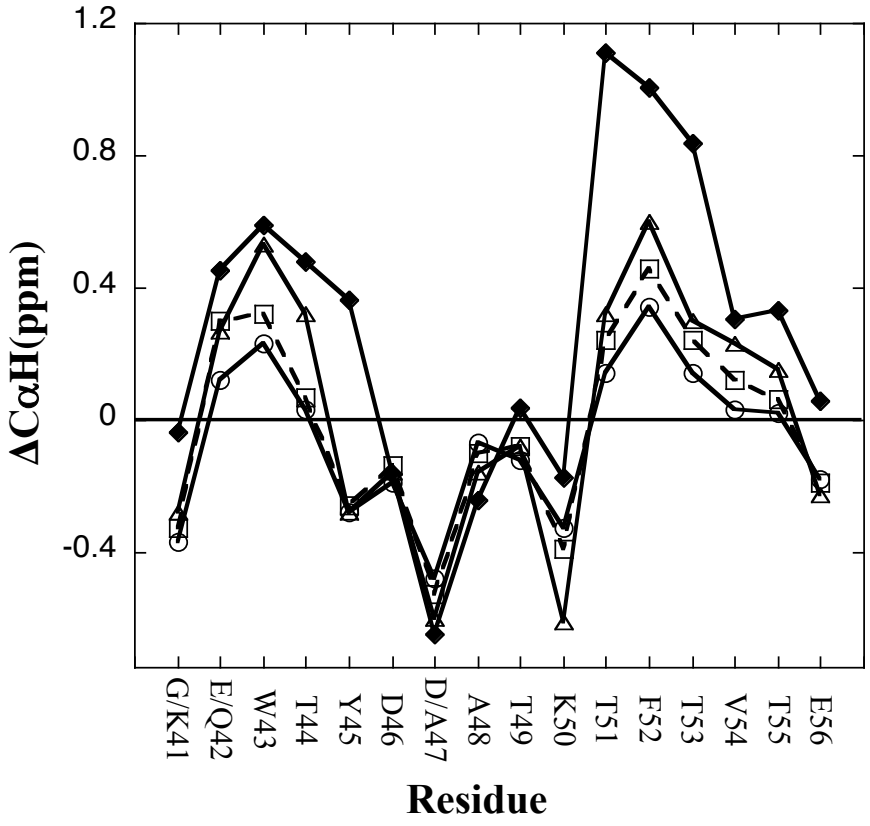

Figure 2: Deviations in CαH chemical shifts from random coil values for each residue in G41(wt) (circles), K41(wt*) (squares), K41+Q42+A47 (triangles) and in the C-terminal region of PGB1 (filled diamonds) [25] in D_2O at pH 7 and 4 ºC.

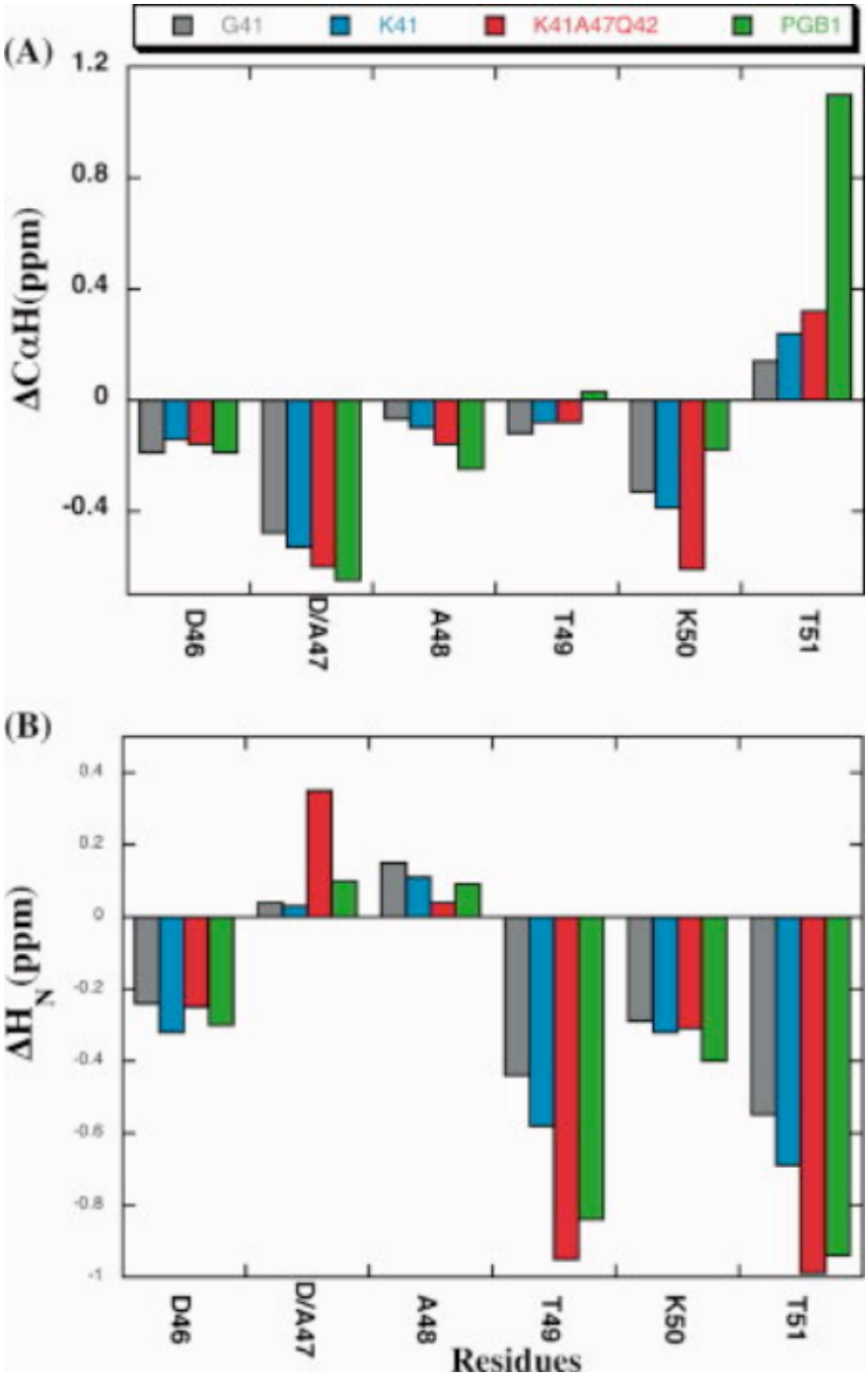

Figure 3: Deviations in $C_\alpha H$ (A) and H_N (B) random coil values for the turn residues in G41, K41, K41+Q42+A47, and the native PGB1.

A comparison of the short-range NOE contacts in K41+Q42+A47 with those in G41 and K41 confirms that the β-turn in the triple mutant has an altered conformation [2; 25]. Figure 4 shows a summary of the relevant NOE data, and Figure 4 illustrates the NOE crosspeaks that are observed in K41+Q42+A47 but not in K41 or G41. According to the nomenclature of Thornton [69] and Richardson [70], the β-hairpin in the protein adopts 4:4 hairpin conformation with a four-residue type IV turn, where the side-chain carboxyl group of Asp46 forms hydrogen bonds with the backbone NH groups of Thr51 and Lys50. In contrast, the NOE data for K41+Q42+A47 suggest the carboxyl of Asp46 forms hydrogen bonds with side chain of Lys50 and the backbone NH group of Thr49, consistent with a 3:5 β-hairpin conformation. In addition, the pattern of $d_{NN}(i,i+1)$ NOE crosspeaks observed in K41+Q42+A47 are from Ala47 to Thr51 rather than from Asp46 to Thr51 as seen for K41 or G41, which suggests K41+Q42+A47 has a five-residue β-turn instead of a six-residue β-turn as in G41 peptide [2]. We observed intensity enhanced $d_{NN}(i, i+1)$ NOE interactions between Ala48 and Thr49, Thr49 and Lys50 and a more intense $d_{NN}(i, i+1)$ NOEs between Lys50 and Thr51. Moreover, a new $d_{NN}(i, i+2)$ NOE crosspeak is observed between Ala48 and Lys50, suggesting the turn is tighter in K41+Q42+A47. Taken together, these results suggest that the "AATKT" sequence in the triple mutant adopts a 3:5 β-hairpin, where the side chain of Lys50 rotates clockwise with respect to its position in the protein, extends in an upward direction perpendicular to the plane of the β-hairpin and packs against the side chain of Ala47. Consistent with this structure, we observe long-range NOE contacts between side-chain protons of Tyr45 and the CαH of Lys50 (Y45εH↔K50αH), CβH of Ala47 (Y45εH↔A47βH)

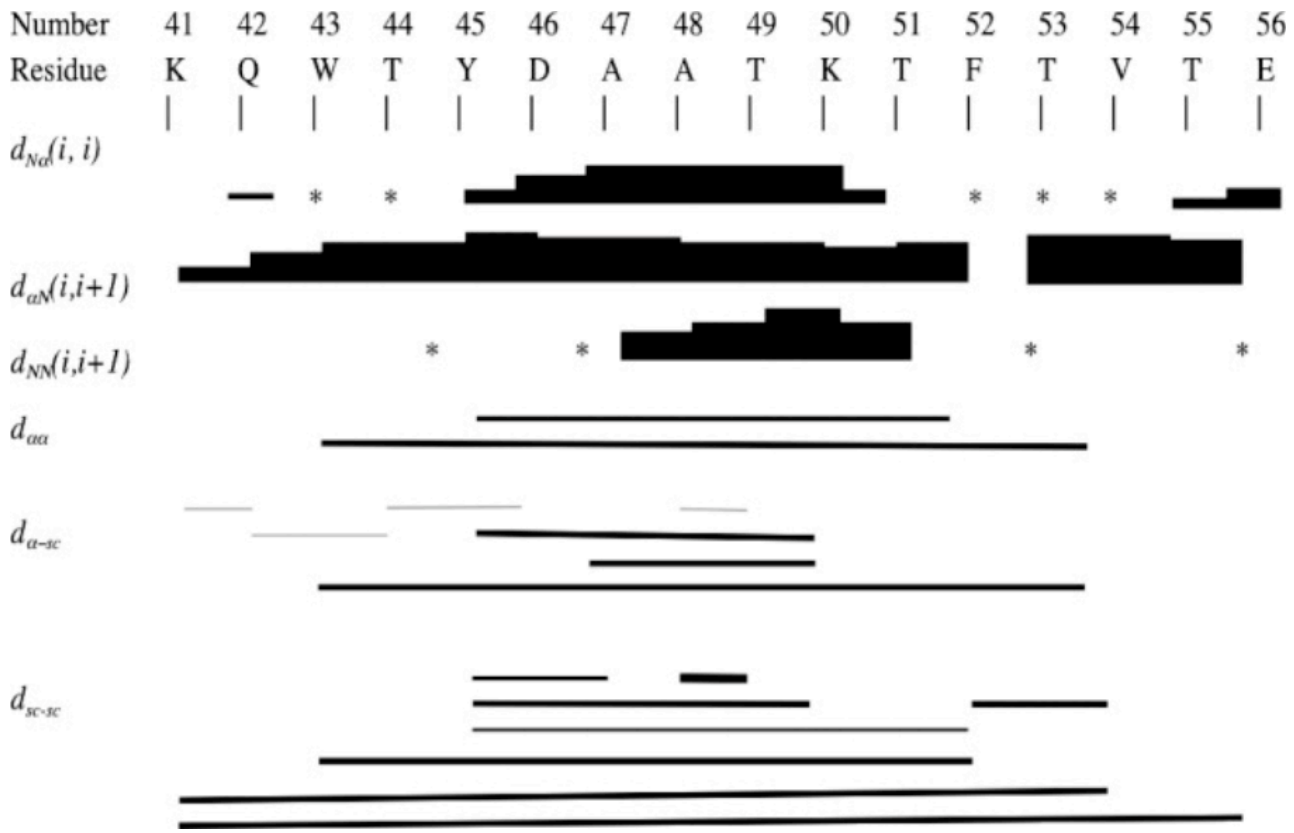

Figure 4: A summary of the NOE crosspeaks found in K41+Q42+A47 in 5 m*M* sodium phosphate 10% D_2O and 100% D_2O solutions at pH 7 and 4°C. The NOE intensities are proportional to the line thickness. Lines represent connectivities between two residues.

and between the CαH of Ala47 and the side chain of Lys50 (A47αH↔K50γH) in the β-turn region (Figure 5).

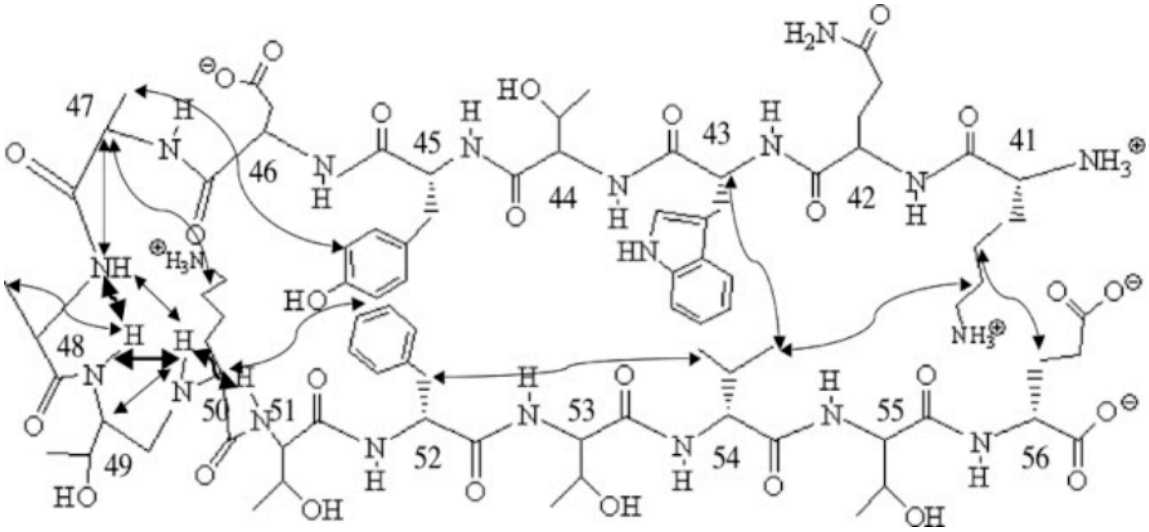

Figure 5: A summary of NOE interactions observed in the K41+Q42+A47 peptide but not in K41 or G41 (thin arrows), and the enhanced NOE interactions relative to those observed in K41 or G41 (thick arrows).

Three residue loops, like those in the 3:5 hairpin, are more prevalent in folded proteins than four residue loops [1; 19], which suggests the former may be more stable. This result is consistent with what we have observed here: replacement of Asp47 with the shorter side chain Ala47 leads to a 3:5 loop at the β-turn region and thus increases the

stability of the β-hairpin. However, Asp has been shown to have much higher β-turn potential than Ala at the first turn position (*i*) in a statistical study of protein structures [71]. One obvious explanation for this discrepancy is that the statistical study was based on protein structures, not isolated peptides. In proteins, interactions of tertiary structures may stabilize certain β-turn conformations even energetically unfavorable ones, whereas in the absence of other interactions, β-turns can form more entropically favorable conformations. This may indicate that the contributions to β-turns in the proteins are not completely dependent on turn sequence, but could be context-dependent. We have observed here that the side chain of Lys50 orients differently from that in the protein due to the absence of interactions with other structural elements in the intact protein. Although the D47A replacement removes the favorable Asp47-Lys50 charge pair in the original sequence of G-hairpin, it also removes the steric clash between Lys50 and Asp47 in the β-turn. Thus, the D47A mutation leads to a new orientation for the side chain of Lys50, making a 3:5 β-hairpin and larger burial of hydrophobic surface in the G-hairpin.

K41+Q42+A47 Retains the Native-like G-hairpin Register. It might be expected that the change of a 'doubly bulged' turn (four residue 4:4 loop) to a three-residue (3:5) turn might lead to a one-residue frameshift in the alignment of β-strand register. However, the sharper bend of the β-turn toward the hydrophilic side in K41+Q42+A47 might compensate for this one-residue shift, thus maintaining the same register of the β-hairpin. In our case, this is supported by strong $d_{NN}(i,i+1)$ and weak $d_{\alpha N}(i,i+1)$ interactions in the β-turn residues and strong $d_{\alpha N}(i,i+1)$ in the β-strand residues (Figure

4). Additionally, intense long-range interstrand αH↔αH NOE interactions between Trp43 and Val54 (Figure 4), and between Tyr45 and Phe52 are still maintained, consistent with the proposed structure.

The β-Hairpin Variants Have a More Stable Hydrophobic Cluster. We observe more long-range NOE interactions of all types between main-chain and side-chain atoms as well as between different side chains in K41+Q42+A47 than in K41 or G41 (Figure 5). In addition to the NOE contacts between the side chains of Phe52 and Trp43 observed in G41, we also observe strong NOE interactions between the side chains of Phe52 and Val54 (F52βH↔V54γH). In the K41+Q42+A47 peptide, new strong NOE contacts (Y45εH↔K50αH, A47αH↔K50γH) develop from the replacement of the hydrophilic residue Asp47 with a small hydrophobic residue Ala (Figure 5 and see Supplement Materials). Furthermore, a NOE crosspeak is observed between the CαH of Trp43 and the side chain of Val54 (W43αH↔V54γH) and new NOE contacts between the side chains of Lys41 and Val54 (K41βH↔V54γH) and Glu56 (K41γH↔E56βH) are also observed (Figure 5 and Supplement materials). Together these results suggest K41+Q42+A47 has a more defined "pleated" β-hairpin structure with a larger hydrophobic cluster extending from the four-residue core (Tyr45, Trp43, Phe52 and Val54) to both the β-turn and the termini. (Note: The "pleated" appearance of β strands arises from tetrahedral chemical bonding at the Cα atom: if a side chain points straight up, then the bond to the must point slightly downwards, since its bond angle is approximately 109.5ºC.) The formation of this larger hydrophobic cluster formation is partly due to the new turn in 3:5 hairpin conformation, but also because of the more

pleated hairpin structure in the variants. The more pronounced pleated structure not only helps to compensate for the one-residue shift of the new β-turn while maintaining a regular network of interstrand hydrogen bonding, it also helps to pack the hydrophobic side chains together better.

Removal of Unfavorable Electrostatic Interactions Stabilizes the β-Hairpin. We observe long-range NOE interactions between Lys41 and the side chains of Val54 and Glu56 in the K41+Q42+A47 peptide. These contacts suggest that the ion-pair interactions at the termini are present as was observed previously for the K41 peptide [25]. The NOE contacts between the side chains of Lys41 and Val54 are only observed in K41+Q42+A47, but not in K41. Larger absolute $\Delta\delta_{C_\alpha H}$ values for residue Val54, Thr55, Glu56 are also observed in K41+Q42+A47 than those in K41 and the triple mutant is more stable than the single mutant. Therefore, the results suggest that removal of the charge-charge repulsion between Glu56 and Glu42 helps prevent the terminal residues from fraying, thus increasing the hydrophobic packing and stabilizing the β-hairpin.

Molecular Dynamics Study of the β-Hairpin Variants. In an effort to gain insight into the structural properties of the β-hairpins in solution and to visualize the non-native 3:5 β-hairpin conformation, we have performed molecular dynamics simulations in explicit water for each peptide variant. All simulations were performed for 10 ns and several independent trajectories of β-hairpin behaviors for each peptide were generated.

A well-folded and stable β-hairpin should have a large amount of buried surface area from the interstrand interactions. To estimate the packing contributions from the mutations, we calculated the average total amount of buried hydrophobic contact surface

area (HCSA) and the HCSA in the major cluster of buried surface area. The side-chain packing in the peptides buries from 91 to 102 Å of HCSA (Table 2) and the variants show increased side-chain packing with most of the increase in HCSA originating from the largest hydrophobic cluster (see Materials and Methods). The results show that the average amount of HCSA is largest in the triple mutants, consistent with the NOE data and suggest that the triple mutants (K41+T42+A47 and K41+Q42+A47) have more compact β-hairpin structures and thus more stability than the wild-type G-hairpin or the single or double mutants. Thus, we agree with previous studies that emphasized the important role of the hydrophobic cluster in the stability of the β-hairpin [36; 48; 72; 73; 74].

Next, we used the NMR data as a guide to find the best representative structures from the MD simulations (Figure 6). The best structures have the β-turn of a 3:5 hairpin with the hydrogen bonds between Asp46 carboxyl and Thr49 NH and Lys50 NH. The β-turn bends at residue Thr51 allowing for the hydrogen bond between Asp46 NH and Thr51 carboxyl and maintaining similar β-strand register as G-hairpin. In these structures, the β-hairpin has a compact hydrophobic cluster that extends from the central region to the β-turn and the termini, consistent with our experimental data.

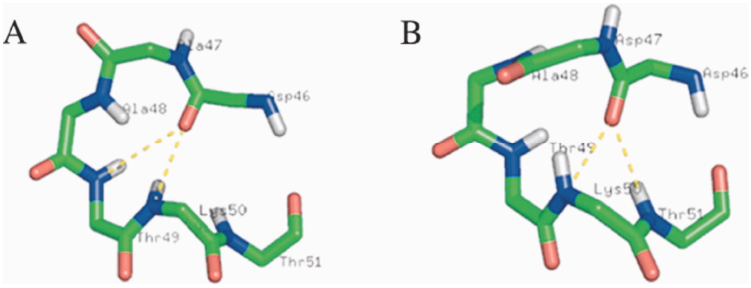

Figure 6: Structural depictions of the native β-turn of 4:4 hairpin conformation of G-hairpin in PGB1 (A) and the β-turn of the nonnative 3:5 hairpin (B), generated from MD simulations. The hydrogen bonds between Asp46 carboxyl and Thr49 NH and Lys50 NH are shown.

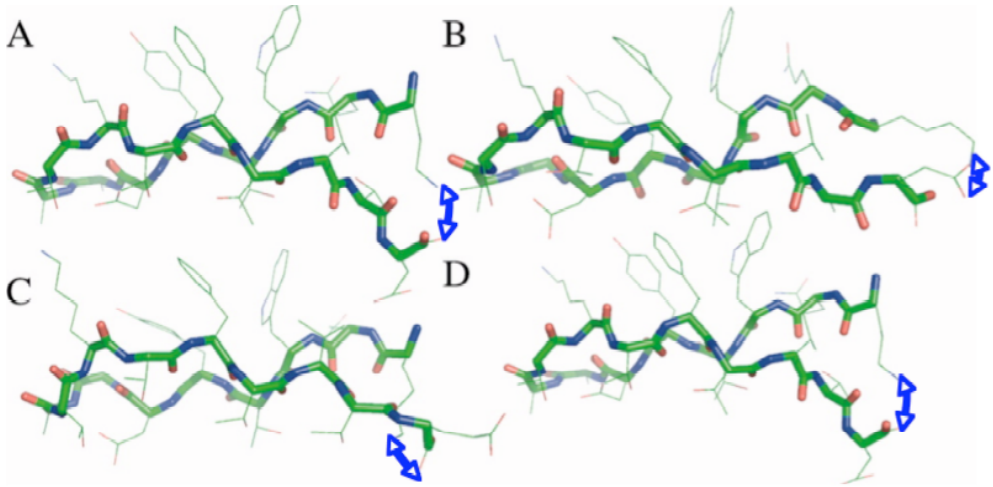

Figure 7: Representative β-hairpin conformations of K41+A47+Q42 selected from MD-generated structures based on their similarity with the NOE data. The ion pairs at the termini are marked with blue double arrow lines.

Table 2: Hydrophobic contact surface area between β-hairpin strands.

Property	Category	G41	K41	K41+A47	K41+T42	K41+A47+T42	K41+A47+Q42
Strand HCSA[a] (Å^2)	*Total*[b]	91±14	93±18	93±19	100±15	104±18	102±13
	Cluster[c]	72±15	75±19	70±19	78±18	83±20	81±16

[a]The hydrophobic contact surface area (HCSA) between β-strands. Standard deviations are shown.
[b]The total is the average over a sum of all side-chain hydrophobic contacts.
[c]The cluster averages over only the largest cluster (see Materials and Methods).

Table 3: Terminal ion pair (IP) interactions in the ensembles of MD-generated structures.

Property	Category	G41[e]	K41[e]	K41+A47	K41+T42	K41+Q42	K41+A47+T42	K41+A47+Q42
	Possible IP[a]	2	4	4	4	4	4	4
Terminal IP	*IP*[b]	0.6	1.1	1.3	1.3	1.8	1.2	1.2
	SB (%)[c]	92	78	100	100	100	97	100
	Nm to Os	0.2	0.0	0.0	0.1	0.3	0.1	0.0
Type of Terminal IP[d]	*Nm to Om*	0.4	0.1	0.2	0.1	0.2	0.0	0.1
	Ns to Os	-	0.0	0.1	0.2	0.4	0.1	0.1
	Ns to Om	-	1.0	1.0	1.0	0.9	1.0	1.0

[a]The total number of potential ion pairs involving the N- and C-terminal residues.
[b]The average number of ion pairs in the ensemble structures.
[c]The percent occurrence that salt bridges form as well as ion pairs, defined by distance and angle restraints for hydrogen bonds.
[d]The type of potential terminal ion pairs include N-terminal (Nm) or Lys41 side chain amino group (Ns) to C-terminal (Om) or Glu56 side-chain carboxyl (Os). The dashes indicate not significantly populated.
[e]The data for G41 and K41 are from Huyghues-Despointes *et al.* [25]

Since some of the replacements we introduced are at the termini of the peptides, we also analyzed the generated ensembles to check the effects of varying the electrostatic interactions at the termini on the conformation of β-hairpins and compared the results to those on K41 and G41 from the previous study [25]. Table 3 shows the frequencies and types of the termini ion-pair interactions formed in our ensemble of structures. More than one type of ion-pair contacts (occupancy>1.0) is observed in all the variants with the G41K replacement. The results show that almost all the double and triple mutants form ion pairs all of the time (100% salt bridges). It is also found that removing the unfavorable charge-charge interaction between Glu42 and Glu56 seems to increase the occupancy of the ion pair between the side-chain amino group of Lys41 and the side-chain carboxyl of Glu56. From the selected MD-simulated structures for K41+Q42+A47 generated based on the NOEs data (Figure 7), the β-hairpin tends to adopt a twisted and β-pleated structure to achieve the large hydrophobic cluster while keeping ion-pairs interactions at the termini (Figure 7).

To investigate the effect of replacement of Asp47 with Ala in the turn region, we calculated the CαRMSD values for all the ensembles of each variant. The crystal structure of G-hairpin in PGB1 was used as the starting structure for each MD trajectory and the reference for the CαRMSD calculation. The shape of the CαRMSD distribution represents how similar the structures are to the native structure of the G-hairpin in PGB1 and to each other. A single and narrow distribution peak indicates structures with less variation from the native G-hairpin, while a wide peak or more than one peak suggests differences from the native structure [25]. Figure 8 shows comparison of the distribution of

turn region CαRMSD values of peptides with or without the D47A replacement. The results show that the variants with the D47A replacement have a broader peak and often show more than one peak in the distribution of turn region CαRMSD values, while the variants with the wild-type sequence in the turn region have a narrow and single peak. The larger CαRMSD values observed in the D47A variants suggest that this replacement changes the conformation of the β-turn in the hairpins, consistent with our experimental data.

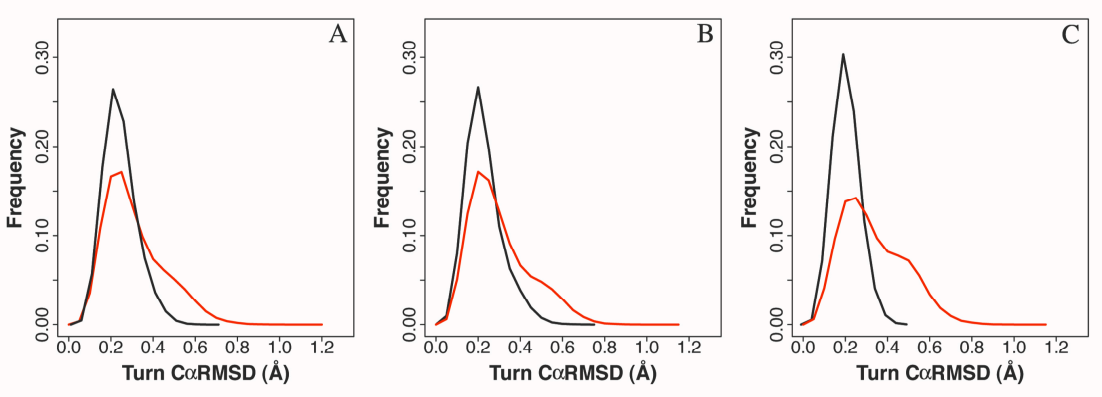

Figure 8: Comparison of probability distributions of the normalized CαRMSD in the MD-generated ensembles in the β-turn region (residue 47-50). Shown in the above figure from left to right are A) K41 [75] and K41+A47 (red); B) K41+T42 [75] and K41+A47+T42 (red); C) K41+Q42 [75] and K41+A47+Q42 (red).

Furthermore, according to the Ramachandran nomenclature for turns [76], the central residues (*i*+1, *i*+2) in the loop of residues 47-51 (from position *i* to *i*+4) of a 3:5 type I turn will be expected to have an $\alpha_R\alpha_R$ conformation in Ramachandran plot, which is different from the miscellaneous type IV turn in the 4:4 hairpin. Also, the α_L conformation is quite common at position *i*+3 in type I turns [71]. Therefore, we also determined the distribution of the main-chain ϕ, ψ for the loop of K41+Q42+A47 hairpin based on the generated structures from MD simulations. In Figure 9, a loop conformation $\alpha_R\alpha_R\alpha_R\alpha_L\beta$ is populated for the sequence Ala47-Ala48-Thr49-Lys50-Thr51 in the K41+Q42+A47 variant as expected for type I turn. Previously, the Thornton group has also shown that type I β-turn is the most abundant turn type in folded proteins [40; 71]. These results also suggest the type I turns are more favorable for β-hairpin formation.

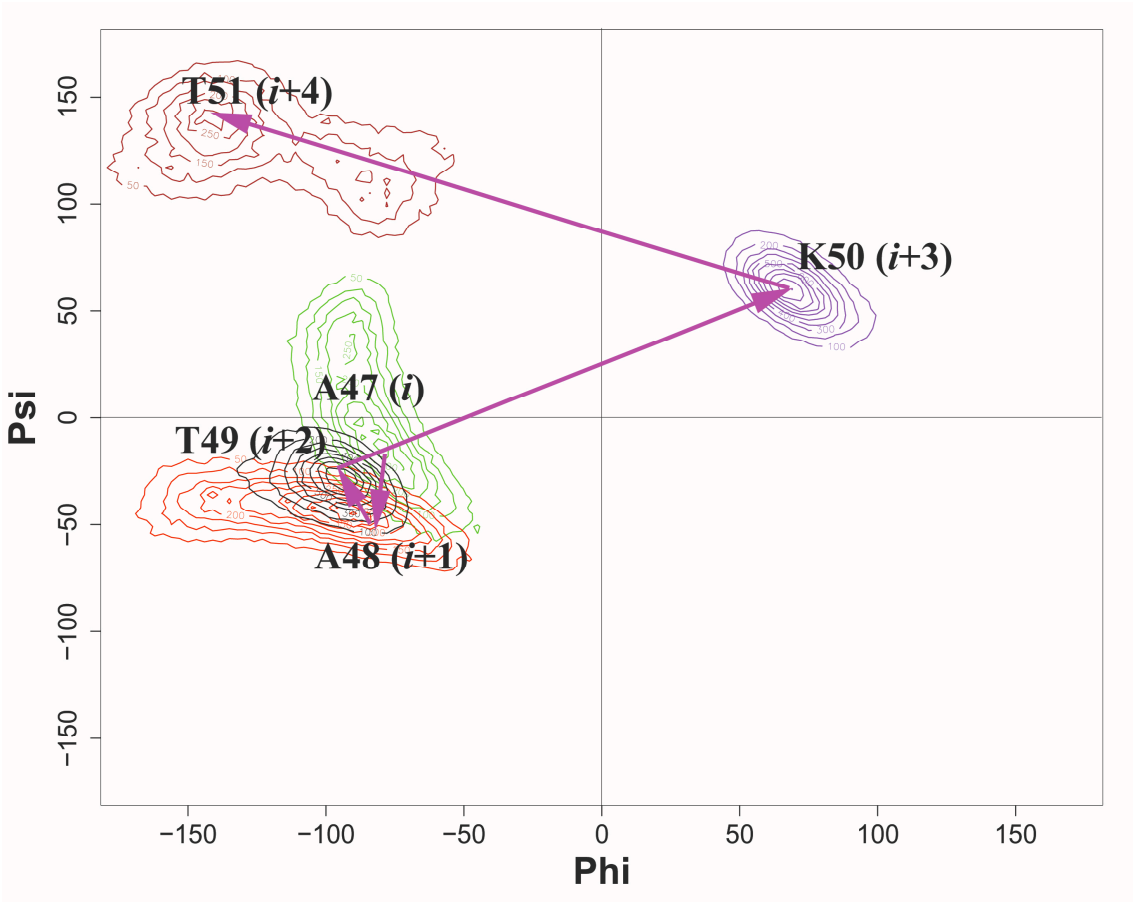

Figure 9: The Ramachandran plot showing the main-chain ϕ, ψ angle distributions for the five-residue loop (residue 47 to 51) in K41+Q42+A47 generated from MD simulations. The lines connect the most populated conformations (ϕ, ψ angles) for each residue in the turn.

CONCLUSIONS

Our data clearly show that removing the unfavorable charge-charge interactions can greatly increase the stability of β-hairpin. K41+T42+A47 is the most stable hairpin among all the studied variants and is ~ 96% folded in 20% methanol at 4ºC. Consistent with a folded structure, we observe more NOE interactions between side chains near the termini and a more pleated β-hairpin structure with larger hydrophobic cluster in K41+Q42+A47 than in the other G-hairpin peptides. We also demonstrate the replacement of Asp47 with alanine in the turn region changes the β-hairpin conformation to a 3:5 hairpin without altering the register and also increases the thermal stability of the β-hairpin. This suggests that the β-hairpin conformation is strongly affected by the turn sequence and the cross-strand side-chain interactions. Importantly, our results show that removing unfavorable electrostatic interactions, even without introducing extra hydrophobic interactions, can enhance the cross-strand side-chain packing and thus increase the stability of β-hairpin.

SUPPLEMENTAL MATERIALS

Chemical Shift Assignments for K41+Q42+A47 Peptide: We used 2D-NMR spectroscopy to determine how the mutations affect G-hairpin structure. The chemical shifts assignments of K41+Q42+A47 were determined by analysis of TOCSY spectra in D_2O and NOESY spectra in D_2O and 9:1 H_2O/D_2O mixture. The complete ^1H chemical

shift assignments for K41+Q42+A47 in 5mM sodium phosphate buffer at pH 7 and 4ºC are listed in Table 4.

Table 4: Chemical shift assignments for K41+Q42+A47 peptide.

Residues	δNH	$\delta C_\alpha H$	$\delta C_\beta H$	$\delta C_\gamma H$	$\delta C_\delta H$	δOther
Lys41	--	4.08	1.86 1.79	1.31	1.68	εH 2.92
Gln42	8.23	4.65	1.95 2.06	2.21 2.26		δNH_2 6.93/7.64
Trp43	8.79	5.24	3.14 3.15			2H 7.09; 4H 7.41; 5H 7.03; 6H 7.13; 7H 7.39; NH10.11;
Thr44	9.06	4.67	4.07	1.16		
Tyr45	8.95	4.29	2.69			2,6H 6.12; 3,5H 6.39;
Asp46	8.16	4.60	2.36 2.71			
Ala47	8.60	3.75	1.47			
Ala48	8.29	4.19	1.42 1.48			
Thr49	7.29	4.27	4.27	1.13		
Lys50	8.10	3.76	1.94 2.11	1.34	1.70	εH 3.02
Thr51	7.25	4.67	4.07	1.16		
Phe52	8.84	5.31	2.91 2.94			2,6H 7.18; 3,5H 7.40; 4H 7.20;
Thr53	9.00	4.65	4.07	1.17		
Val54	8.54	4.42	1.56	0.43 0.64		
Thr55	8.57	4.50	4.17	1.16		
Glu56	8.24	4.04	1.86 2.06	2.21		

∗ Assignment values (in ppm) are relative to DSS (0 ppm).

A.

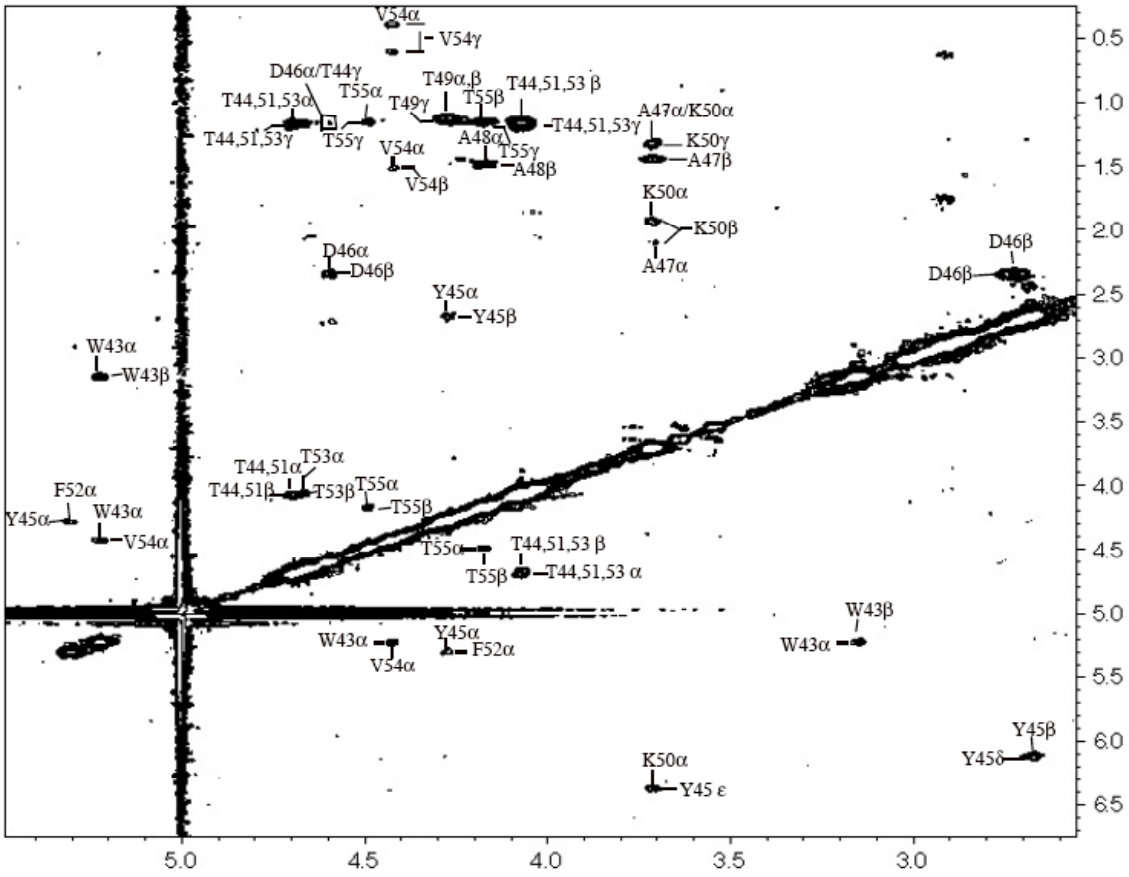

Figure 10: Portions of the 500 MHz NOESY spectrum of K41+Q42+A47 peptide recorded in 100% D_2O at 5 mM sodium phosphate buffer pH 7 and 4ºC (t_m = 200 ms). Panel A shows $C_\alpha H$-$C_\alpha H$ region of the NOESY spectrum; Panel B shows aromatic protons of the same spectrum.

B.

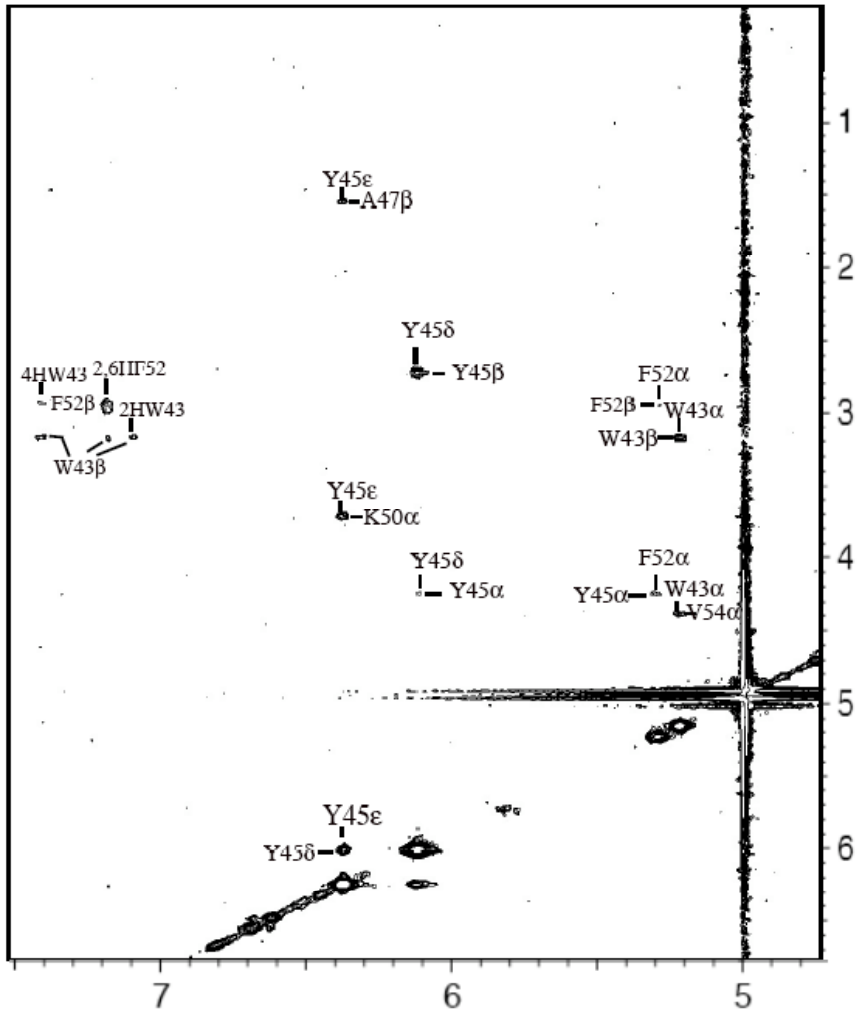

Figure 10: Continued

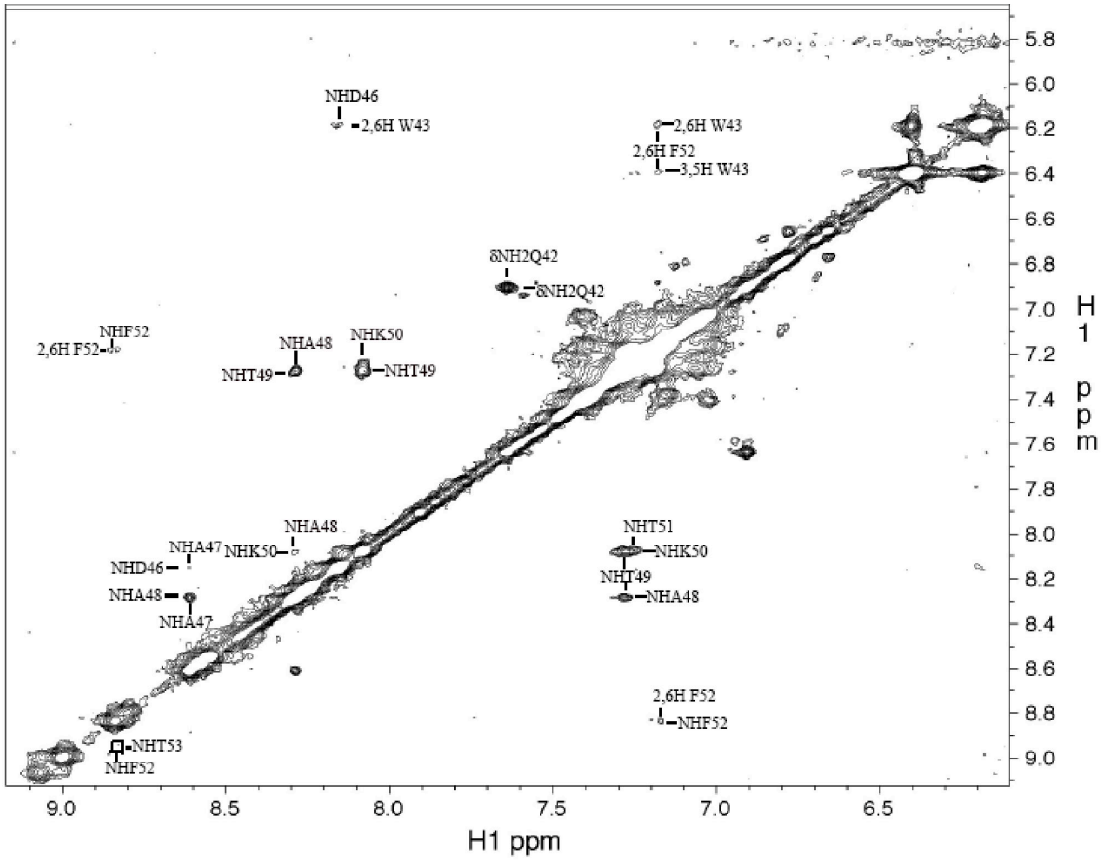

Figure 11: Amide and aromatic protons region of the 500 MHz NOESY spectrum of K41+Q42+A47 peptide recorded in water (10% D_2O) at 5 mM sodium phosphate buffer pH 7 and 4ºC.

CHAPTER III

A KINETIC STUDY OF AMYLOID FORMATION IN PROTEIN PGB1

INTRODUCTION

Amyloid fibrils and other non-native protein aggregates have been a critical subject of many studies because they are associated with a variety of devastating human diseases, including Alzheimer's, Creutzfeldt-Jakob, Huntington's and Parkinson's diseases [77; 78]. Many proteins, protein fragments and polypeptides that are not related to any known disease, such as the SH3 domain of phosphatidylinositol 3-kinase, human muscle acylphosphatase, a type II domain of fibronectin, insulin, Protein G B1 domain, have been also shown to undergo amyloid formation under certain *in vitro* conditions [79; 80; 81; 82]. The polypeptides known to form fibrils *in vitro* have no obvious sequence or structural similarities, which suggests that the formation of amyloid fibrils might be a generic property of polypeptides [83; 84].

In spite of the structural diversity of the proteins that form amyloid fibrils, all amyloid fibrils have a common core structure in which continuous β-sheets are formed with β-strands running perpendicular to the long axis of the fibrils and exhibit a cross-β diffraction pattern [85]. Also, mutants with the same native-like structure in their globular protein form could form morphologically different types of fibrils, as reported for Aβ, insulin, and β-microglobulin [86; 87; 88]. The amyloidosis mechanism that accounts for multiple fibril formations at a molecular level remains unclear.

For many proteins, it has been shown that a range of experimental conditions, such as elevated temperature, low or high pH, presence of organic solvents, bivalent cations, destabilizing mutations, low to moderate concentration of denaturant, can promote the formation of amyloid fibrils [80; 89; 90]. Also, even single mutations that reduce the stability of the native conformation of the protein enhance the ability of a protein to form amyloid [91; 92]. It is generally accepted that conformational stability of protein is a major factor that determines the ability of proteins to form amyloid. In an effort to explain the molecular basis of amyloid fibril formation, it has thus been proposed that the process of fibrillation usually starts from a partially folded conformation or intermediates under conditions in which specific intermolecular interactions, including electrostatic attraction, hydrogen bonding and hydrophobic contacts, are still favorable so that they can still stabilize the intermolecular β-sheets structure that is the basis of the highly ordered structure of amyloid fibrils [77; 89; 93; 94]. However, considerable debate exists as to the mechanism of fibril formation. It is still unclear what structures of the amyloidogenic precursors are sufficient to form the amyloid structures. For instance, is a high population of a specific conformation in denatured state is required to initiate forming the amyloid fibrils or do the amyloidogenic precursors assemble from existing β-sheet structures or association of unfolded polypeptide segments [86; 88]? Some studies show that aggregation can start from native-like states while others show they are initiated with disordered polypeptide segments, while still others have proposed both native folded states and partially folded states are involved [86; 95; 96]. Among these studies,

some proteins are populated with α-helical structures while some are populated with β-sheet structures under conditions in which fibrillation occurs.

Although the ability to form amyloid-like structures may be a generic property of the polypeptide chain itself, it is clear that some amino acid sequences have a higher propensity for fibril formation than others. The propensity of polypeptides to form amyloid fibrils or aggregates is influenced by the nature of the amino acid side chains [97; 98; 99]. Mutational studies for a range of model systems have mapped out several factors that affect the rate of amyloid formation or aggregation, such as intrinsic properties like β-sheet and α-helical propensity, hydrophobicity, charge and aromaticity along with other physicochemical environmental factors, such as pH, ionic strength and temperature [97; 98; 100; 101; 102; 103; 104]. Serrano and coworkers have shown that the aggregation propensity of peptides and proteins can be predicted by taking into account several terms: hydrophobicity, β-sheet propensity, electrostatic interactions and hydrogen bonding [105].

Based on experimental observations, the formation of amyloid fibrils has been described as a process that normally consists of three steps: nucleation (formation of stable nuclei), elongation (growth of nucleus representing the polymerization reactions), and an equilibrium phase [106; 107; 108; 109]. This process characterizes a sigmoidal curve defined by an initial lag phase, where no amyloid fibril is observed, a subsequent growth phase in which amyloid fibrils increase, and a final equilibrium phase, where the amyloid fibril contents reach a plateau [110]. Two parameters are introduced to study the kinetics of fibril formation: the lag time (t_{lag}), the time required for the nucleation phase

to take place and the growth rate constant (k_{app}) which measures the efficiency of the growth [108]. There is also some evidence that suggest the k_{app} and the lag time (t_{lag}) are positively correlated properties [98].

The development of an *in vitro* model system to investigate the physicochemical basis of amyloid fibril formation is very important. Many amyloid proteins have low solubility and lack of a known high-resolution three dimensional structure, which has led to experimental limitations for amyloid studies and hampered a comprehensive understanding of amyloid formation or fibrillation on a molecular level [110]. Recently, the Regan group has developed a series of variants of the immunoglobulin-binding domain B1 of the streptococcal protein G (PGB1) that can form amyloid *in vitro* under certain conditions [111]. The protein PGB1 has been used as a model system for studying the biophysical basis of protein folding, structural plasticity and dynamics using a variety of biophysical techniques [112; 113; 114; 115; 116; 117; 118; 119; 120; 121; 122; 123]. Thus, it is an excellent model system for examining the relation between protein folding and amyloid formation. The B1 domain belongs to the ubiquitin superfamily and comprises two β-strands, β1 and β2, connected by an α-helix as shown in Figure 12 [116].

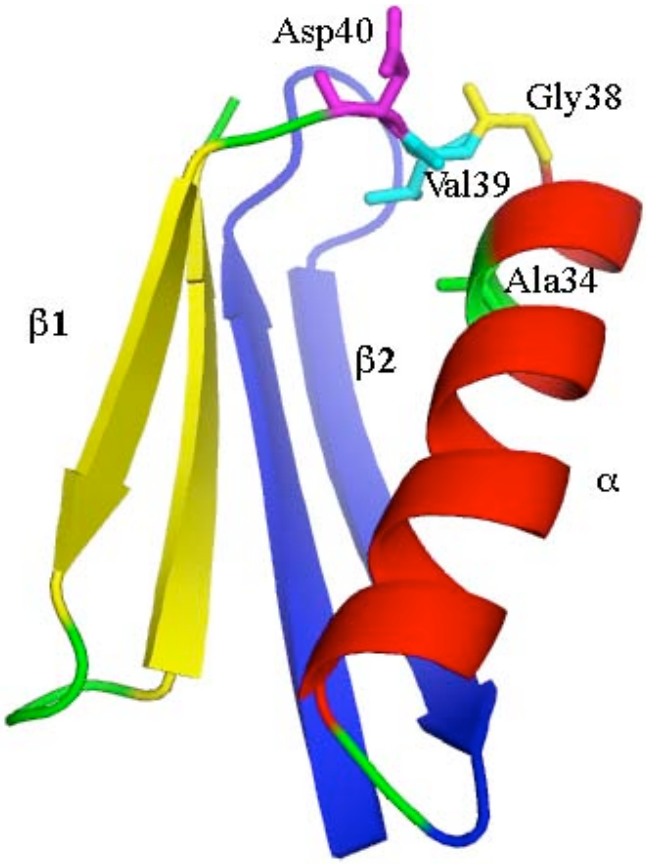

Figure 12: A ribbon diagram of PGB1. The PDB code for the structure used here is 1PGB [124]; indicated on the plot are Ala34 [125], Gly38 (yellow), Val39 (cyan) and the host-site Asp40 (magenta). The secondary structure elements (β1, β2 and α) are also labelled. This image was created using PyMol Molecular Viewer by Delano Scientific LLC.

Like many other proteins, amyloid formation in PGB1 variants can be induced under conditions in which there is a maximum population of intermediate conformations

present, such as near their melting temperatures [111]. Several variants of PGB1 in which amino acid substitutions were made either on the central or the edge of the β-strands are able to form amyloid fibrils near their melting temperature. This suggests that the location of a substitution destabilizing the protein is not a key factor to form the amyloid fibrils, but rather the global stability of the mutant is the most important factor [111].

Domain swapping, a specific means by which oligomeric proteins can be formed from their monomers, has been suggested as one of the mechanisms to accomplish the conformational changes that lead to amyloid formation [126; 127]. The domain-swapping model has been suggested for PGB1 amyloid fibril assembly in two separate studies [111; 128]. In both proposed domain-swapping models, the opening of the native structure of PGB1 is required and the β-hairpins are swapped to line up in a continuous manner to assemble into fibrils. Evidence that supports the model has also been found. Recently, the Gronenborn group has determined the structure of a PGB1 core mutant (L5V/F30V/Y33F/A34F), a domain-swapped dimer that exists in equilibrium with a partially folded monomeric species [129]. It has also been show that single cysteine disulfide cross-linking between residues in the outer edge β-strands of the domain-swapped dimer structure can lead to multimer formation prior or during the PGB1 amyloid formation [128]. In spite of this evidence, there are still many questions remaining unexplained about the domain-swapping model. For example, what is the role of the α-helix in the fibril assembly and does it need to partially unfold to undergo conformational rearrangement or remain as an intact secondary structural element in the dimerization? Furthermore, different morphologies have been observed in amyloid

fibrils formed by different PGB1 variants, suggesting that different models could be used or the atomic arrangement in the domain swap model could also depend on the sequence of monomer.

The purpose of this study is to investigate whether any correlation exists between the local secondary structure propensity and fibril formation tendency in PGB1. Since β-sheet structures are observed in the final structure of amyloid fibrils, will increase of the local β-hairpin propensity also increase the propensity of fibril formation in PGB1? Also, mutations in the α-helix were made to examine the role of the α-helix during the amyloid formation in PGB1. To exclude the global conformational stability of PGB1 as a factor that affects the propensity of amyloid formation, single mutations were made to change the local secondary structure propensity but maintain similar overall protein stability. Surprisingly, we observed significant changes in the propensity of amyloid formation in PGB1 by only single mutations located on the local α-helix propensity.

MATERIALS AND METHODS

Protein Expression and Purification. The pET15b plasmid containing the 6X-His-tagged PGB1 gene was generously provided by David Baker (University of Washington). Site-directed mutageneses were performed using the QuickChange™ kit from Stratagene. The primers are purchased from Integrated DNA technologies. Purified plasmids encoding mutant proteins were transformed into *E.coli* BL21(DE3) strain for protein overexpression. A single colony was used to inoculate a 75 mL Terrific Broth [130]

medium with 50 μg/mL ampicillin in a 250 mL baffled flask. The culture was incubated overnight with shaking (300 rpm) at 37ºC. An aliquot (5mL) of the preculture was transferred into 500 mL TB with 100 μg/mL ampcillin in a 2L baffled culture flask and grown to an OD_{600} value of approximately 0.8-1.0. The cells were then induced for 4 h at 37ºC with 1mM IPTG and harvested by centrifuging for 30 min at 4200 rpm and suspended in cold phosphate-buffered saline (PBS) with 1 mM PMSF. The cells were then lysed by passing through a French Press cell at 1200 psi and centrifuged again at 25000 g for 30 min to remove the cellular debris. The clear diluted cell lysate was loaded onto a 40 mL of Cu^{2+} charged His•Bind fractogel column (Novagen) equilibrated with 20 mM ammonium bicarbonate buffer and 9 mM imidazole. The His-tagged protein was eluted with 20 mM ammonium bicarbonate buffer and 200 mM imidazole. The elutant was then lyophilized and then loaded on a G-50 column with 50 mM ammonium bicarbonate buffer to remove the imidazole. The His tag for all variants was not cleaved. The purity of proteins was checked by polyacrylamide gel electrophoresis and electrospray mass spectrometry (Protein Chemistry Laboratory, Texas A&M University). The lyophilized pure proteins were kept in 4ºC for storage.

Transmission Electron Microscopy. Electron micrographs were recorded using a JOEL 1200 EX TEM transmission electron microscopy (Microscopy & Imaging Center, Texas A&M University) at 80 kV excitation voltages. Fibril samples were prepared by floating carbon-coated mica onto sample droplets followed by negatively staining with aqueous uranyl acetate (2% w/v).

Thioflavine T Fluorescence Assay for Fibrillation. The thioflavin T (ThT) fluorescence assay was performed on 96-well microplates using a SPECTRAmax GEMINI dual-scanning microplate spectrofluoromter. For each experiment, 100 μl aliquots of the sample were mixed with phosphate buffer, pH 6.0 and ThT stock solutions. The final concentration of dye was always 10 μM. The fluorescence was measured immediately after mixing using excitation and emission wavelengths of 450 nm and 482 nm respectively using a cut-off filter at 475 nm.

Thermal Denaturation Experiments. All the denaturation experiments were done at pH 3.75 in 50 mM sodium acetate on Aviv Circular Dichroism Spectrometer Model 62DS or 202. The thermal denaturation curves were analyzed as previously described [131].

Kinetics of Fibril Formation and UV 280 Sedimentation. The kinetics of PGB1 fibril formation are usually described as a sigmoidal curve defined by a lag time where no change in ThT fluorescence intensity was observed, a sigmoidal increase in ThT fluorescence denoting the growing of the fibrils and a plateau with a constant fluorescence intensity. Due to the high irreproducibility and sample manipulation with the ThT fluorescence measurement, here, we decided to use a UV 280 sedimentation method to track the formation of amyloid. Protein solution was centrifuged at 13000 rpm and the UV absorbance of the clarified supernatant at 280 nm was used to track the amount of protein that remained in solution or was not involved in the formation of amyloid. Plots of the extent of amyloid formation over time could be produced and fitted

to the sigmoidal curve described by the following equation described as previous study using Kaleidagragh [132]:

$$Y= (y_i + m_i x) + (y_f + m_f x) / [1 + \exp -(x - x_0)/\tau]$$

where Y is the fraction of protein that appears in the fibril formation, x is time, and x_0 is the time to 50% of maximal amyloid formation. Therefore, the apparent rate constant, k_{app} for the growth of fibrils is given by $1/\tau$; the lag time (t_{lag}) is given by $x_0 - 2\tau$ (Figure 13).

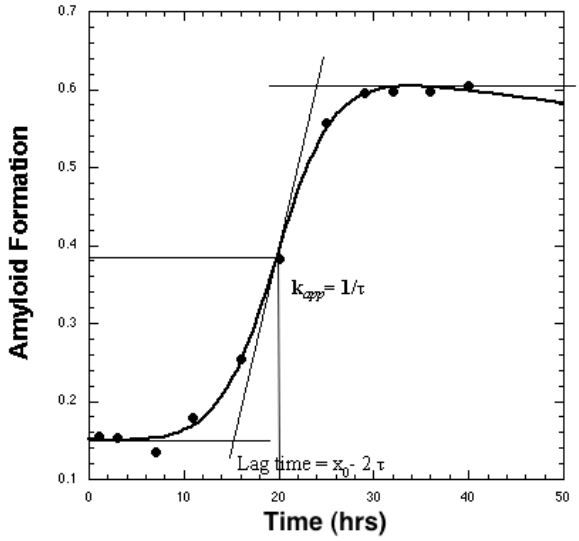

Figure 13: Illustration of the time course for amyloid formation in PGB1. The y-axis is normalized as the fraction of protein that appears in the fibril formation. The lag time (t_{lag}) is approximated by $x_0 - 2\tau$ and the apparent growth rate constant, k_{app}, is given by $1/\tau$.

Peptide Synthesis and FPLC Assay. The peptide K41T42 was synthesized and purified as previous described [133]. The peptide purity was confirmed by MALDI mass spectrometry. For studying the effect of adding the β-hairpin peptide K41T42 on kinetics of amyloid formation in PGB, 100 μM D40M PGB1 in 50 mM sodium citrate, pH 3.75 was incubated with agitation at 37 ºC with 50 μM K41T42 peptide added. Aliquots of samples over time were analyzed on FPLC column to determine the amount of protein and peptide.

Preparation of Seeds. The fibril stock solution containing 1000 μL 100 μM D40M PGB1 in 20 mM sodium citrate, pH 3.75 with 27% trifluoroethanol (TFE) was incubated with agitation at 37 ºC for 72 hours. Seeds were prepared by the sonication of 200 μL aliquots of the fibril stock solution using a Microson sonicator at intensity level 2.

RESULTS AND DISCUSSION

In an effort to investigate factors that affect the tendencies to form fibrils other than the conformational stability, the underlying strategy was to find protein mutants with similar conformational stabilities and conditions under which all mutants can form fibrils. A variety of conditions have been explored to find a possible condition under which variants can produce fibrils, including elevated temperatures, salt concentrations and co-solvent trifluoroethanol (TFE) additions. Our results show that the D40M PGB1 and its variants tend to form amyloid fibrils at acidic pH conditions (ranging from 2.5 to

4.5) at elevated temperature (tested from 37ºC to 60 ºC) in the presence of moderate concentrations of TFE (tested from 15% to 40%). No fibrils were formed without agitation, under neutral pH, basic pH or in the presence of the chemical denaturant guanidine chloride. No fibrils were formed at room temperature, even in the presence of fibril seeds. Similar to insulin and human calcitonin fibril formation [108], the kinetics of PGB1 amyloid formation also shows a dependence on the initial protein concentration, seeding and ionic strength. An increase in ionic strength and seeding can result in decreases in lag time (Figure 14). pH has also been found have an effect on the kinetics of PGB1 fibrillation (Figure 15). At 37ºC, more acidic solution can shorten the lag time to form fibrils. The pH effects on D40M PGB1 amyloid formation are also dependent on incubation temperature. The protein can form amyloid fibrils at 60ºC with a pH up to 4.5 while no amyloid fibrils were detected at 37ºC with a pH higher than 4. Both pH and temperature have great effects on the conformational stability of a protein. The effect of pH and temperature on kinetics of fibrillation could be a result of combined effects on the conformational stability.

Here, for the further study we use the following "standard" conditions to induce amyloid fibrils: incubating 100 μM D40M PGB1 variants in 50 mM sodium citrate, pH 3.75 and 27% TFE at 37ºC with agitation.

The propensity of PGB1 amyloid formation has been shown to correlate with the global stability of the protein [91; 111]. Here, a host-site mutation "D40M" was made to destabilize the wild type PGB1 in all of the PGB1 variants. A series of single amino acid substitutions that are located in different regions of PGB1 are introduced as guest

mutations that either increase or decrease the local secondary structure propensity (shown as in Table 5). The T18A mutation, which is located in β1, decreases the stability of the local secondary structure. The T25A, K28G, K28A, V29A, A34F, N35A and N37A mutations are located on the α-helix (residue 23-37). Previous studies show that alanine has the highest helix propensity while glycine has the lowest propensity [134]. The replacements by either Ala or Gly were thus designed to change the local helix propensity. The helix propensity of these sequences has been checked using the AGADIR program [135] and the predicted α-helix content for each sequence is shown in Table 5. The variant A34F was one of the variants that has been shown to stabilize the domain-swapped dimer by connecting the two helices with aromatic ring-ring interactions [129]. G38V and V39A are located on the turn connecting the α-helix and β2. Replacement of Gly38 with Val will result in disruption of the original "αGβ" turn motif, a populated αβ turn motif suggested to play a structural role [136]. Replacement of the hydrophobic residue Val39 with Ala is designed to remove the hydrophobic contact between Val39 and Ala34 that is important in maintaining the "close-up" packing of the α-helix against the β-sheets in the PGB1 native structure. The E42T, E42Q, D47A and T49A variants are located in β2. Our previous studies have shown that E42T, E42Q and D47A can stabilize the second β-hairpin [133]. Replacement of T49A was shown to destablize this β-hairpin [26].

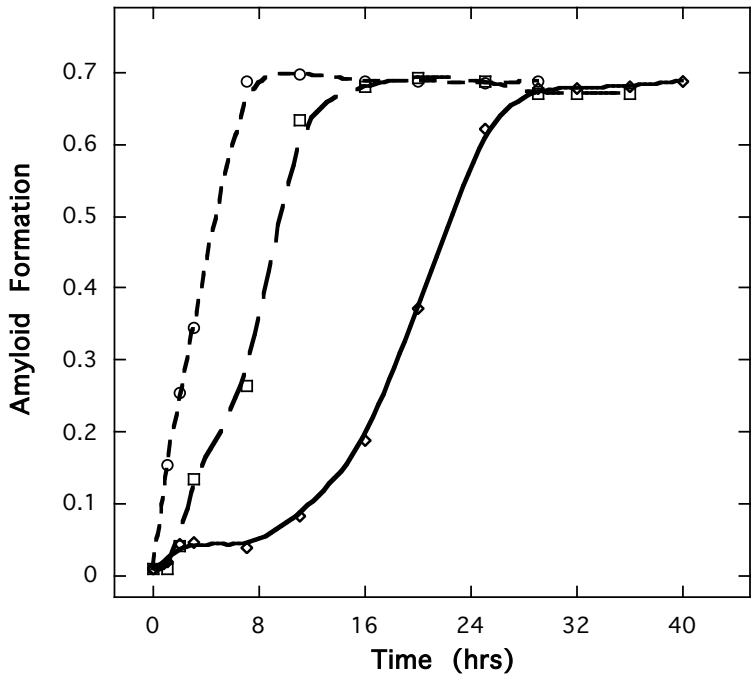

Figure 14: Influence of seeding and ionic strength on PGB1 amyloid formation monitored by UV 280 sedimentation method. The curves represent the change of amyloid formation as a function of incubation time. Fibrils were formed by 100 μM D40M PGB1 at 37ºC with agitation with 27% TFE at pH 3.75 in 20 mM sodium citrate without seeding (diamonds), 200 mM sodium citrate without seeding (squares) and 20 mM sodium citrate with 1% (v/v) seeding added (circles).

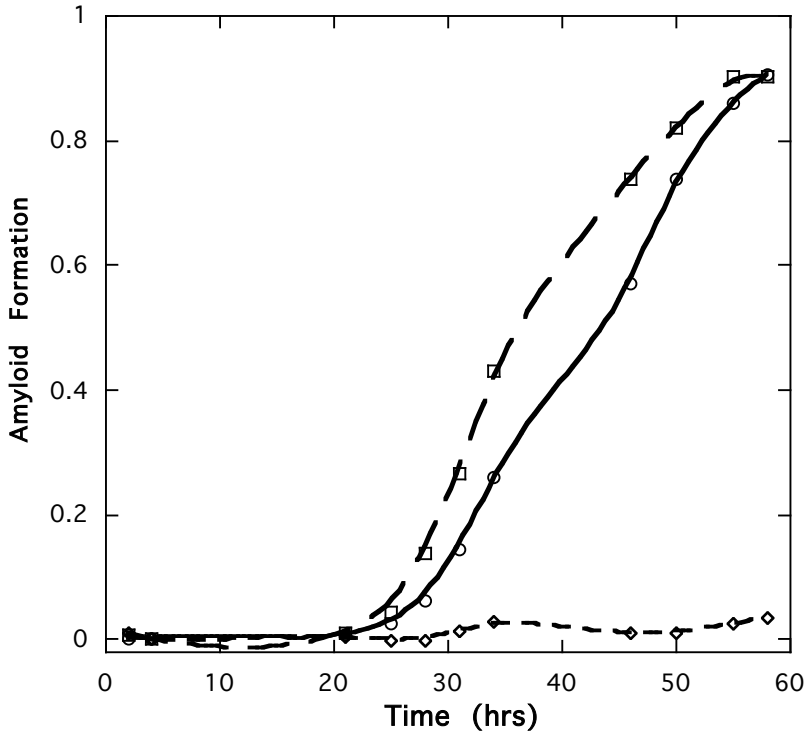

Figure 15: Influence of pH on PGB1 amyloid formation monitored by UV 280 sedimentation method. The curves represent the change of amyloid formation as a function of incubation time. The fibrils were formed by 100 μM E42Q at 37ºC with agitation and 30% TFE at pH 3.5 (squares), 3.75 (circles) and 4.0 (diamonds) in 20 mM sodium acetate.

The conformational stability of all of the mutants has been measured by thermal denaturation analysis. All of the mutants have similar stability to the D40M PGB1 (wt*) with difference ranging from -1.6 to 0.5 kcal/mol. Among these mutants, A34F, G34V and V39A destabilize D40M PGB1 the most ~ 1.6 kcal/mol. Our results show no correlation between the global protein conformational stability and the kinetics of amyloid formation for these variants (Table 5).

All of the mutants can be induced to form fibrils under the described conditions and have been tested for their propensities to form fibrils under these same conditions. The kinetics of all of the mutants show the described sigmoidal curve process except A34F, which displays a "two-state" amyloid formation process (Figure 16). During lag phase, no significant changes in ThT fluorescence or UV absorbance were observed. As shown in Table 5, the lag times and growth rates are independent of their conformational stability, suggesting there are factors other than global conformational stability affecting the propensity of amyloid fibrillation within the range of difference in stability of our mutants. It is surprising that the mutants with the most significantly increased lag times are all located in the α-helix instead of β-sheets given that β-sheet structures are found in the final amyloid fibrils. K28A, V29A, N35A and N37A are the four mutants that delay the onset of amyloid formation for the longest time. Except for N37A, all three mutants have increased local α-helix propensity. Substitutions in the β-strands show slight changes in the lag times but all have increased growth rates. The G38V and A34F mutants have no lag times and the V39A mutant also shows a longer lag time than the D40M PGB1.

If the initial conformation that associates on the fibrillation pathway is a partially folded intermediate that is in equilibrium with the monomer, the lag time would reflect both the population of the intermediate and its rate of formation from the monomer and its self-affinity. In contrast, the rate of elongation would reflect the rate that monomers can assemble onto the nucleus or ends of the fibrils. Our results show that stabilizing the α-helix in PGB1 will decrease the population of the intermediate or its rate of formation from the monomer. In other words, the α-helix of PGB1 needs to be unfolded in the partially (un)folded intermediate that is necessary for amyloid formation. Charged residues are suggested to play an important role in packing of protofilaments in fibrils. This could possibly explain why K28G slightly destabilizes the α-helix but still has a longer lag phase in amyloid formation.

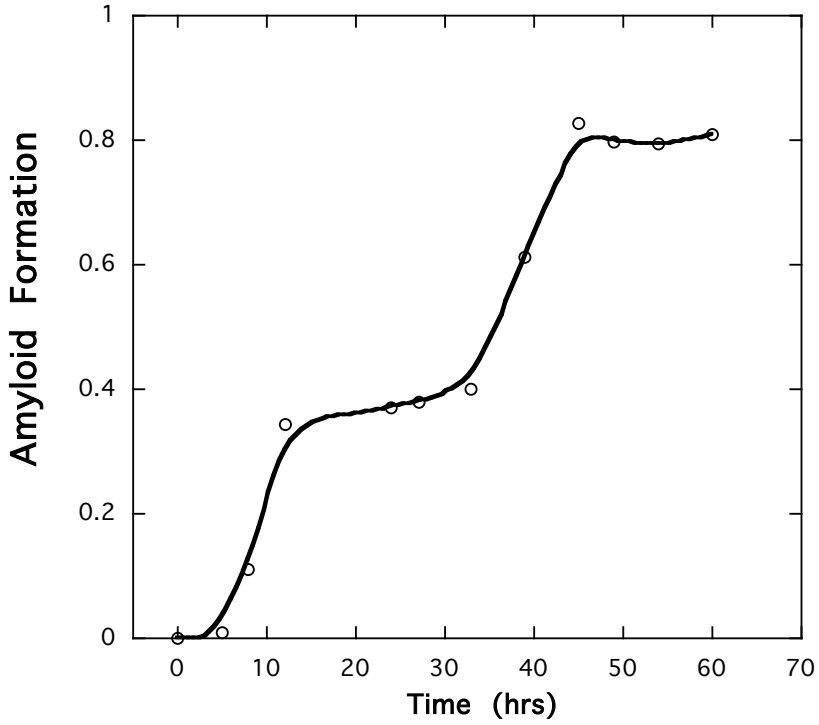

Figure 16: Fibril formation of A34F in 50 mM Na Citrate, 27% TFE, pH 3.75 at 37ºC with agitation displays a two-state curve. The amyloid formation was normalized on a scale from zero to one, which shows the fraction of protein that goes to the fibrils.

Previous studies suggest that the β2 is formed in the PGB1 folding/unfolding transition while the α-helix and β1 are not [119]. Our CD analysis of D40M PGB1 also shows β-sheet structures significantly populated under the amyloid formation conditions

(37ºC with 27% TFE). According to the proposed domain-swapping model for PGB1 amyloid formation [91; 128], both β1 and β2 contribute to the β-sheet assemble in PGB1 amyloid formation. Our results show that stabilizing the β-sheets has only small effects on the lag phase but almost all of the substitutions that stabilize the β-sheets increase the growth rate, which suggests that the side-by-side interactions between the two β-strands are involved in the monomer assembly to the fibrils. Ala34 has been suggested as a "gatekeeper" residue that blocks the formation of unwanted misfolded or aggregated states of PGB1 [128] and replacement to a larger side-chain residue in this position could promote the formation of the fibrillation specific interface. Our results with A34F, which can immediately start forming amyloid, agree with this suggestion. A34F is also the only mutant that displays the two-state fibrillation process. It is possible that there is more than one pathway that this mutant use to form fibrils and different conformations might be needed to initial the aggregation. More work would be needed to understand fibril formation in this variant.

G38V also speeds up the formation of the required unfolded conformation from monomer, which could be explained if the mutation opens up the native structure by disrupting the interface that the α-helix packs against the β-sheets. Another mutant V39A also removes the hydrophobic contact between Val39 and Ala34 in the monomer that connects the α-helix with β-sheets. However, a much longer lag phase was observed for this mutant. Possible explanation is that the hydrophobic residue Val39 could be also directly involved in the amyloid nucleus formation or the formation of the unfolded conformation that was required to initial the fibril formations. In the domain-swapping

model, Val39 is directly involved in the dimerization of PGB1 by hydrophobic interactions with Asn35 in the other monomer. Replacements of Val39 to alanine could also disrupt the favorable hydrophobic contacts that are needed for dimerization or other possible conformation required for amyloid fibrillation.

We also tested the effect of adding the β-hairpin peptide K41T42, which is derived from the 2nd β-hairpin of PGB1 [133], on the kinetics of PGB1 amyloid formation. The presence of peptide K41T42 leads to a shorter lag time (Figure 17), but causes no clear change on the fibril growth rate. The concentration of peptide K41T42 remains unchanged before and after the fibril formation, suggesting that the peptide itself is not incorporated into the amyloid fibrillation. The peptide K41T42 could help to form or maintain the partially unfolded precursor conformation by interacting its hydrophobic side with the helix thus precluding the formation of native structure.

Figure 18 shows images from electron microscopy corresponding to different mutant fibrils. Most of the mutants have overall similar morphology except V29A and K28G (Fig 18E and 18C). K28G shows longer and thinner fibrils compared with other mutants. The morphology of V29A fibrils is very different from other variants with very straight fibrils that do not contain a twist. This suggests that residue Val29 might be involved in the packing of protofilament into fibrils and the replacement of Val29 with Ala could result in a different assembly.

In conclusion, while it has been suggested that the location of mutation does not play an important role in determining the propensity of PGB1 amloid formation [91], our results show that different locations of even single mutations have dramatic different

effects on PGB1 amyloid formation. Our results suggest that the α-helix in PGB1 is involved in the fibril assembly of PGB1 and stabilizing the local helical structure could lead to a longer lag time for PGB1 amyloid formation.

Table 5**:** Lag time and growth rate constant for D40M PGB1 and variants.[a]

Protein	$\Delta\Delta G^o$ (kcal·mol^{-1})[b]	Location of mutated residue in native structure	Effect on loca l secondary structure	t_{lag} 98	k_{app}(h^{-1})
WT*	-	-	0.34[c]	14.4	0.37
T18A	- 0.5	β1	-[d]	12.4	0.50
T25A	0.2	α-helix	0.25[c]	15.5	0.25
K28G	- 0.1	α-helix	0.27[c]	30.9	0.24
K28A	- 0.2	α-helix	0.58[c]	172.7	0.36
V29A*	- 0.7	α-helix	0.72[c]	456	0.26
A34F	- 1.6	α-helix	0.26[c]	0	N.A.[e]
N35A	0.5	α-helix	0.48[c]	>30days	0.26
N37A	0.2	α-helix	0.32[c]	>30days	0.25
G38V	- 1.5	turn	-[d]	0	0.61
V39A	- 1.6	turn	-[d]	95.4	0.11
E42T	0.1	β2	+[d]	98.4	0.42
E42Q	0.1	β2	+[d]	26.0	0.33
D47A	0.5	β2	+[d]	16.3	0.59
T49A	- 0.7	β2	-[d]	7.1	0.83

[a]All amyloid tests were performed by incubating 100 μM protein in 50 mM sodium citrate, pH 3.75 with 27% TFE at 37ºC with agitation.
[b]$\Delta\Delta G^o$= 0.16 (the ΔS_m for WT* in kcal·mol^{-1})* ΔT_m.
[c]Values are α-helix propensity calculated by AGADIR from Serrano group. The number shows the predicted helical content for helix sequence part at pH 4.0, 310K and 0.1M ionic strength.
[d]The symbol,+/-, means the mutation increase/decrease the local secondary structures propensity.
[e]N.A., Not applicable, since this is a two-state process.

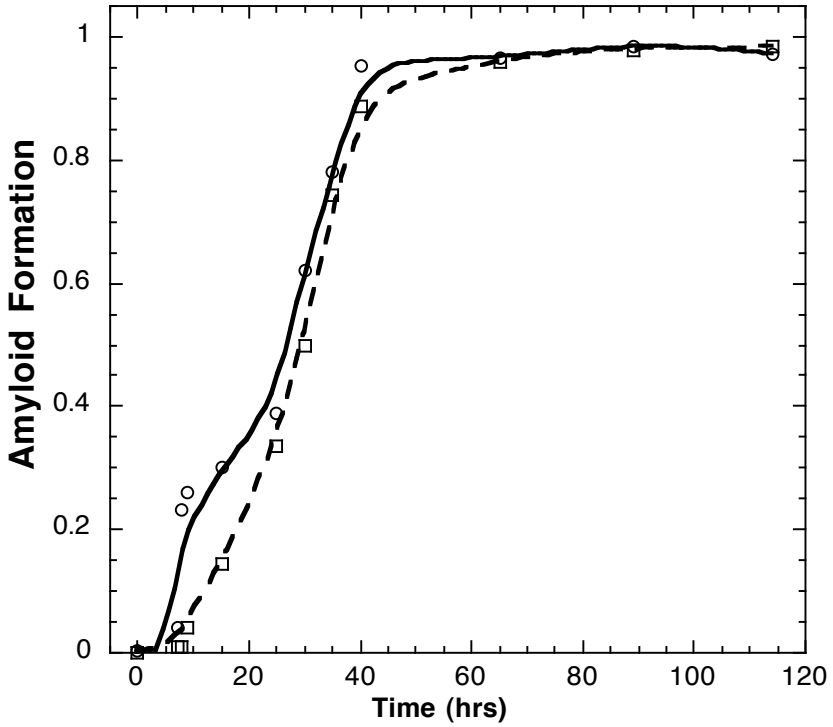

Figure 17: Effect of adding the β-hairpin peptide K41T42 [133] on kinetics of D40M PGB1 amyloid formation. Experiments were performed by incubating 100 μM D40M PGB1 in 50 mM sodium citrate, pH3.7 at 37ºC with agitation: with 50 μM K41T42 peptide added (circles), without peptide added (squares).

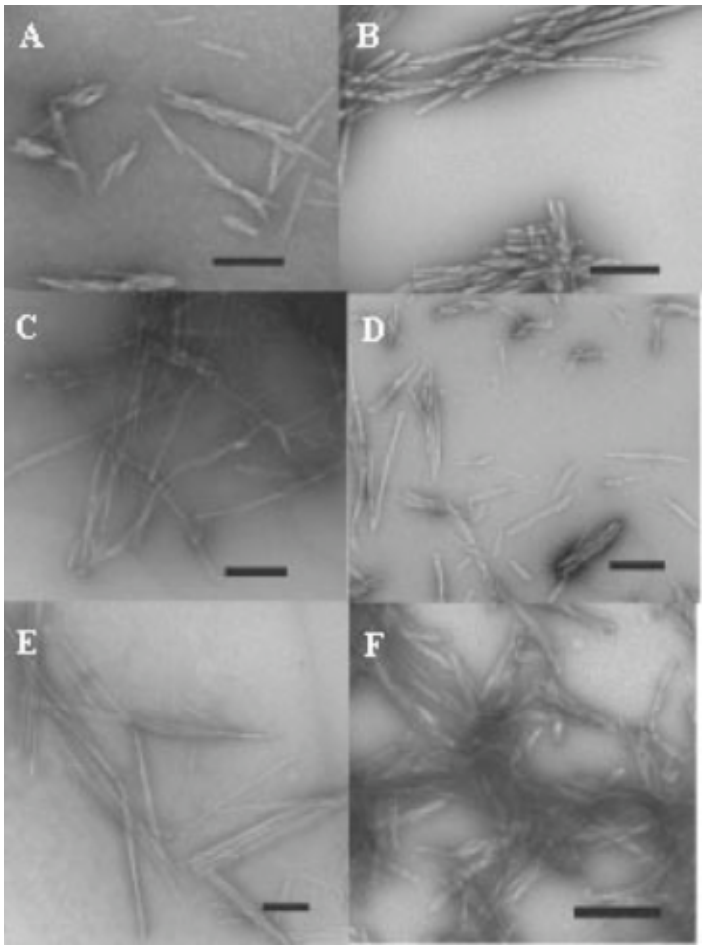

Figure 18: Electron micrographs of a representative set of PGB1 mutant fibrils. A. D40M PGB1; B. T18A; C. K28G; D. N37A; E. V29A; F. D47A. The black scale bars represent 200 nm.

CHAPTER IV

WHY LARGE PROTEINS ARE NOT MORE STABLE THAN SMALL PROTEINS

INTRODUCTION

In all living systems, globular proteins need to fold into their unique, compact three-dimensional conformations to carry out their biological functions. Understanding the energetics of protein folding helps us better predict protein stability from its structure. Among all the theoretical methods of prediction of protein stability, one traditional approach focuses on summing up all the major factors that contribute to the conformational stability based purely on experimental data. Based on what we have learned about protein folding energetics, the major factors are: the hydrophobic effect, hydrogen bonding, and conformational entropy. Covalent crosslinks also stabilize protein, if present. So, the conformational stability of a protein can be estimated by the following equation:

$$\Delta G = \Delta G_{CE} + \Delta G_{H\phi} + \Delta G_{HB} + \Delta G_{-s-s} \qquad (1)$$

In a study of the conformational stabilities of members of the ribonuclease family of proteins, Pace and his colleagues [137] showed that a remarkably good prediction of protein conformation stability can be obtained by just summing up the contributions deduced from rough estimations of a number of these terms. However, the predicted

conformational stabilities are much larger than the measured values when the same method was applied to larger proteins. In small proteins, the energetic contributions from the favorable folding forces, such as hydrophobic effect and hydrogen bonding, are nearly offset by the entropic penalization of folding. In contrast, large proteins bury more nonpolar groups, have a larger fraction of hydrophobic core residues [138] and form more hydrogen bonds than small proteins, which suggests that large proteins should be more stable than small proteins. But based on the large body of experimental data on conformational stability in the past 20 years, large proteins and small proteins have similar marginal stabilities. Thus, this observation raises a question: why are large proteins NOT more stable than small proteins? Or, what forces are used to destabilize large protein? How could we improve our traditional prediction method to make it also applicable for large proteins?

In an attempt to answer these questions in this study, we will re-examine several available prediction methods for each factor responsible for the conformational stability of proteins. We will also examine those structural characteristics that change with protein size and could possibly contribute to destablize the large proteins. These prediction methods will be used to compare 20 proteins with different sizes. Our studies suggest that burying non-hydrogen-bonded charged and polar groups are the strategies that nature uses to fine tune protein stability.

MATERIALS AND METHODS

Accessibility Analyses and Hydrogen Bonding. Solvent-accessible surface area (ASA), accessibility analyses and hydrogen bonding interactions are analyzed with the in-house program *pfis* - PDB file information software [139]. The atomic radii of Richards [140], a water probe radius of 1.4 Å and a slice width of 0.25 Å were used for the calculations. Hydrogen atoms were added to the PDB files using the Biopolymer module in the Insight II package. All water molecules were removed from the coordinate sets during the calculations. Intra- and intermolecular hydrogen bonds in the coordinated sets were identified with criteria that the hydrogen atom to acceptor atom distance <2.6 Å or the donor atom to acceptor atom distance was <3.6 Å. The unfolded protein structures were modeled with β-sheet extended dihedral angles using Insight II Biosym package.

Protein Unfolding Energies in Vacuum. To be consistent with Lazaridis et al. [141], we followed the same calculation procedures as theirs: Neutral side chains for all amino acid were employed. The topology file in PARAM19 was also modified in the same way as described [141] to yield a reasonable charge distribution for the neutral side chains of Asp, Glu, Arg, and Lys. Polar hydrogen positions were added to the native protein with the HBUILD program of CHARMm in the Insight II 2005 software package (Accelrys). The structures were minimized for 300 steepest descent steps. The extended structures were generated with β-sheet extended dihedral angles (ϕ = -140º and ψ = 135º) and also minimized for 300 steepest descent steps. Resulting deformations by the prolines were

repaired by local rearrangement during the minimizations. A cutoff of 999 was used in the calculations.

Packing Density. The protein volumes are calculated with the software VOIDOO package with van der Waals radii by Richards [140] and the NSC program as described by Eisenhaber and colleagues [142; 143].

RESULTS AND DISCUSSION

Protein Selection Criteria and Structural Characteristics. The 20 proteins used here have sizes ranging from 35 to 471 amino acids and were selected from a large number of proteins for which the conformational stabilities have been experimentally measured and x-ray crystal structures determined. In addition to protein size, the main criteria in choosing these proteins are summarized below.

A) Only monomeric single-domain globular proteins were used in the analysis. A protein is considered to be a single-domain protein if consensus domain assignment results are obtained from SCOP, CATH and 3Dee Server. B) All of the proteins used here approach a two-state folding mechanism and the conformational stabilities have been measured. C) Proteins with high-resolution three-dimensional x-ray crystal structures were preferred. This is important since the method used here investigates the correlation between the conformational stability and the structure, including the change in water accessible surface area on protein folding (ΔASA) of various groups, number of hydrogen bonds present in the native structures, etc. D) Since we focus mainly on non-

covalent intramolecular interactions, proteins without disulfide bonds or proteins with disulfide bonds but having available experimental data on energetic contribution from disulfide bonds were preferred. For large proteins over 200 residues, it is difficult to find qualified proteins without disulfide bonds, so three proteins that have disulfide bonds are included. E) Since this study aims to understand the correlation between protein stability and protein size, we only include proteins from ambient environments. Proteins from extreme environments such as thermophiles or haloalkaliphilic organisms are thus excluded.

Table 6 lists the 20 globular proteins studied here and summarizes some of their properties. The number of residues, N_r, refers to the number of residues present in the x-ray structures only. For some proteins, a few residues are missing in the x-ray structures. For example, the first five residues are not included in the x-ray structure of Staphylococcal nuclease (1SNO), and the last residue Lys is not included in the x-ray structure of pectate lyase C (2PEC). We assume these missing residues will not affect the result of our analysis. Most of the x-ray structures have a resolution better than 2 Å.

Estimation of the Contribution from Conformational Entropy. Conformational entropy is the major force that destabilizes proteins. The folding of a protein involves a large reduction of the number of conformations accessible both to the peptide backbones and the side chains of the protein; this reduction provides a strong entropic driving force

Table 6: Summary of studied proteins.

PDB	Protein & Source	R^a	N_r[a]	MW[a]	ΔG^b (H_2O)	N_{HB}[c]	N_{HB}/Res[c]	N_{s-s}[d]	ΔASA_{np}[e]	ΔASA_{np} /Res[e]	ΔASA_p[e]	ΔASA_p /Res[e]
1YRF	Villin subdomain HP-35, N68H *(Chicken)*	1.1	35	4082	3.3[f]	26	0.74	0	1659	47	574	16
1PGB	Protein G B1 domain *(Streptococcus sp.)*	1.9	56	6196	4.6[g]	57	1.02	0	2934	52	1150	21
1CSP	Cold shock protein *(B. subtilis)*	2.5	67	7365	3.5[h]	47	0.70	0	3699	55	1420	21
1UBQ	Ubiquitin *(Human)*	1.8	76	8565	6.7[i]	64	0.84	0	4530	60	1616	21
2HPR	HPr *(B. subtillis)*	1.8	87	9042	4.2[j]	60	0.95	0	5185	60	1844	21
1RGG	Ribonuclease *(S. aureofaciens)*	1.2	96	10576	6.4[k]	85	0.89	1	5355	56	2051	21
1A2P	Barnase *(B. amyloliquefaciens)*	1.5	108	12184	8.8[l]	107	0.99	0	5936	55	2783	26
3CHY	Chey (*E. coli*)	1.7	128	13966	5.7[m]	120	0.93	0	7757	61	3066	24
1LZ1	Lysozyme *(Human)*	1.5	130	14701	14.0[n]	137	1.05	4	8071	62	3284	25
1SNO	Staphylococcal Nuclease *(S. aureus)*	1.7	136	15482	6.1[o]	127	0.93	0	7373	54	3006	22
1IOB	Interleukin 1 beta (*Human*)	2.0	153	17377	9.1[p]	112	0.73	0	9763	64	3618	24
1RX4	Dihydrofolate reductase *(E. coli)*	2.2	159	18000	5.9[q]	137	0.86	0	9655	61	3497	22
2LZM	Lysozyme (*Bacteriophage T4*)	1.7	164	18635	15.0[r]	168	1.02	0	10624	65	4466	27
1BCX	Xylanase *(Bacillus circulans)*	1.8	185	20369	7.2[s]	186	1.01	0	11727	63	5169	28
3ADK	Cytosolic adenylate kinase *(Pig)*	2.1	194	21639	3.9[t]	162	0.84	0	11160	58	4502	23
1CAH	Carbonic anhydrase II (*Human*)	1.9	259	29115	9.3[u]	245	0.95	0	16751	65	6848	26
2ST1	Subtilisin (*B. amyloliquefaciens*)	1.8	275	27534	0.5[v]	268	0.97	0	17778	65	6933	25
2PEC	Pectate lyase C *(E. chrysanthemi)*	2.2	352	37569	12.1[w]	365	1.04	2	22430	64	10608	30
1DIL	Sialidase *(S. typhimurium)*	1.9	381	41942	4.2[x]	387	1.02	1	25315	66	12286	32
1GLM	Glucoamylase (*Aspergillus awamori)*	2.4	470	50423	6.7[y]	510	1.08	3	30512	65	12294	26

[a]*R*, resolution of the x-ray structures in Å. N_r , number of amino acid residues present in the analyzed x-ray structures. MW, molecular mass that is obtained by calculating amino acid sequence from the analyzed structures. [d]$\Delta G(H_2O)$, measured conformational stability of analyzed proteins in kcal/mol, All the $\Delta G(H_2O)$ were determined by chemical denaturation except for 2ST1, which was determined from thermal denaturation data. All the ΔG (H_2O) were measured at room temperature except 1LZ1(10ºC) and 1RX4(15ºC).

[e]N_{HB}-, number of hydrogen bond present in the analyzed x-ray structures, see methods for detail. N_{HB}/Res, number of HB per residue; the average for N_{HB}/Res is 0.93.
[d]N_{s-s}, number of disulfide bonds that present in the protein crystal structures.[e]ΔASA_{np}, $Å^2$, difference between ASA of non-polar groups for the native and unfolded states; ΔASA_p, difference between ASA of polar groups for the native and unfolded states; The average for ΔASA_{np}/Res is 60 $Å^2$; the average for ΔASA_p/Res is 24 $Å^2$.
[f]McKnight *et al.* 1996; [g]O'Neil *et al.* 1995;[h]Schindler *et al.* 1998;[i]Khorasanizadeh *et al.*1993;[j]Scholtz 1995;[k]Pace *et al.* 1998; [l]Serrano *et al.* 1992;[m]Filimonov *et al.* 1993; [n]Taniyama *et al.* 1992; [o]Schortle *et al.* 1986; [p]Chrunyk *et al.* 1993; [q]Perry *et al.* 1987; [r]Gray *et al.* 1996;Wetzel *et al.* 1988;[s]Davoodi *et al.* 2007; [t]Tian *et al.* 1988;[u]Aronsson *et al.* 1995; [v]Pantoliano *et al.* 1989;[w]Kamen *et al.* 2000; [x]Witarto *et al.* 2001;[y]Chen *et al.* 1996.

for protein unfolding. Several different approaches have been used to estimate the contribution of conformational entropy. In 1994, Spolar and Record [144] estimated the large unfavorable entropy contribution is observed constant for each globular protein ($\Delta S = -5.6 \pm 0.5$ e.u. per residue), which can be converted to ΔG_{CE} at 25ºC of 1.7 kcal/mol. In an earlier study with an independent method, Rashin [145] found a linearity of the increase of conformational entropy with molecular weight for proteins with size ranging from 100 up to 250 residues and an average value 1.8 kcal/mol per residue was obtained for conformational entropy loss upon protein folding. Rashin's value shows remarkable agreement with the value Spolar and Record proposed. Here, we will use 1.7 kcal/mol per residue as a rough method to estimate the conformational entropy.

Computational studies on conformational entropy suggest that a residue-specific method should be more accurate. The contribution of conformational entropy for individual residues can be dissected into two parts: the main-chain backbone and the side chain contribution. It is clear that the conformational entropy loss of peptide backbone will not be the same for every residue due to the influences of the side-chain of individual residues on the main chain dihedral angles [146]. Although many different methods have been used in computational studies, almost all the evaluation was based on the classical [147] definition of entropy converted from Boltzmann's formula:

$$S_{conf} = - R \Sigma p_i \ln p_i \qquad (2)$$

where p_i represents the fractional population of each conformational state i in the unfolded states. Studies on side-chain conformational entropy loss on protein folding suggest that the difference between different types of amino acids could be up to 2 kcal/mol [149]. A major breakthrough was made by the Freire group in a study where they obtained an estimation for each amino acid by using a combination of experimental and computational approaches [150]. The conformational entropy change, ΔS_{conf}, is composed of three different terms for each amino acid:

$$\Delta S_{conf} = \Delta S_{bb} + \Delta S_{ex->u} + \Delta S_{bu->ex} \qquad (3)$$

where ΔS_{bb} is the entropy loss of the backbone upon protein folding; $\Delta S_{ex->u}$ is the entropy loss of the surface-exposed side-chain upon protein folding; $\Delta S_{bu->ex}$ is the entropy loss when the surface-exposed side-chain is buried in the interior of the protein. Thus for a given amino acid residue on protein folding, the first two terms ($\Delta S_{bb} + \Delta S_{ex->u}$) will be constant in proteins with different sizes. The third term $\Delta S_{bu->ex}$ will reflect the difference of the residue burial for the proteins.

Although the Freire group approach shows large differences between each type of amino acid, the average value does not differ much from the value (1.7 kcal/mol per residue) in Spolar & Record method. We will consider whether it is necessary to calculate the conformational entropy with this residue-specific method. Large proteins bury a larger fraction of residues and thus have a smaller fraction of surface residues [138]. In a very recent study [151], it was shown that protein size affects both the amino acid

composition and the composition of the residue burial of proteins. The surface-to-volume ratio of protein decreases with the increase of protein size and the burial frequency for several hydrophilic residues such as Gln, Arg, Glu and Lys increases as the protein size increases. These hydrophilic residues are also characterized by a high side-chain conformational entropy loss upon burial. Therefore, it is of interest to also use a residue-specific method to calculate the conformational entropy in this study. For the third term, $\Delta S_{bu->ex}$, we have used a 50% burial cutoff for defining the "buried" side chain upon folding: a side chain is considered buried if its ASA shows larger than 50% buried and the $\Delta S_{bu->ex}$ term will be added to the ΔS_{conf} for this amino acid. The calculated energies from conformational entropy loss at 25ºC for each protein by both Spolar & Record method and Freire Group method with 50% cutoff are shown in Table 7.

Estimation of Energetic Contribution from the Hydrophobic Effect. Non-polar side chains have a low solubility in water and a high solubility in non-polar environments. Consequently, they tend to adhere to one another in aqueous environment, which is referred as the hydrophobic effect. Globular proteins contain hydrophobic cores. In 1959, Kauzmann proposed that the hydrophobic effect is the dominant driving force in protein folding [152]. Many experimental studies now support this widely believed idea. Kauzmann also proposed that the hydrophobic effects can be quantified by a solvent transfer model. He argued that the free energy change for burying a non-polar side-chain in the interior of a protein can be quantified by experiments in which a model compound is partitioned between water and a non-aqueous solvent. Effective application

Table 7: Estimated energetic contribution to the conformational stability for each factor (all values are in kcal/mol).

Protein	HΦ[a]				HB[a]	-S-S-[a]	CE[d]	
	Vajda[b]	Karplus[b]	RB (CH)[b]	RB (OT)[b]			Spolar & Record[d]	Freire Group[d]
1YRF	76	47	43	26	26	0	60	65
1PGB	139	58	65	35	57	0	95	103
1CSP	174	64	85	45	47	0	114	127
1UBQ	209	98	120	62	64	0	129	138
2HPR	239	106	126	60	83	0	148	147
1RGG	252	108	124	65	85	5	163	170
1A2P	296	141	151	81	107	0	184	204
3CHY	368	170	201	103	120	0	219	227
1LZ1	386	167	185	110	137	6	221	246
1SNO	353	158	194	98	127	0	231	249
1IOB	455	197	242	131	112	0	260	296
1RX4	447	195	220	122	137	0	270	280
2LZM	513	223	260	138	168	0	279	304
1BCX	574	224	248	140	186	0	315	358
3ADK	533	225	273	142	162	0	330	347
1CAH	802	374	385	209	245	0	440	483
2ST1	840	337	407	198	268	0	468	476
2PEC	1123	445	529	279	365	8	598	651
1DIL	1272	507	621	335	387	4	648	735
1GLM	1455	611	721	387	510	9	801	867

[a]HΦ, the contribution of hydrophobic effect; HB, the contribution of hydrogen bonding, which was calculated by multiplying the number of hydrogen bonds by 1.0 kcal/mol per hydrogen bond. –S-S-, the contribution of disulfide bonds, which was calculated using equation (5) at 25ºC [153b]Vajda and Karplus, both use ASA-based method. For Vajda method, the $\Delta G_{H\phi}$ was calculated by multiplying the water-surface-accessible-area change upon folding by 34 calmol^{-1} Å^{-2}. For Karplus method, by which all surfaces area associated with main- and side-chain carbon atoms were included except for amide, carboxylate and guanidion carbons. The aliphatic surface areas were then multiplied by 25 calmol^{-1} Å^{-2}. The aromatic surface areas were multiplied by 16 calmol^{-1} Å^{-2}. The sulfur atoms of Cys and Met are not accounted for in this method. RB, Residue-based method, by which $\Delta G_{H\phi}$ was calculated by multiplying the ΔG_{tr} values for cyclohexane (CH) or *n*-octanol (OT) by the number of hydrophobic groups buried estimated by pfis program. The hydrophobic groups included –CH_2- groups from all the polar and charged residues, plus the side-chains of the following amino acids: Ala, Val, Ile, Leu, Pro, Phe, Tyr, Trp, Met and Cys.[d]CE, conformational entropy; Spolar&Record, by which ΔG_{CE} was calculated using 1.7 kcal/mol per residue; Freire Group, the residue-specific method, by which ΔG_{CE} was calculated using D'Aquino's data for each residue. 50% cutoff was used to determine if the $\Delta_{Sbu->ex}$ term will be added or not.

of this model requires answering two questions. Which solutes can most accurately model the transfer free energies of non-polar protein side-chains? Which non-aqueous solvent can best model the interior of globular proteins? Amino acids, N-acetyl amino acids and short peptides have been used to model the side-chains, and various solvents have been used to model the hydrophobic core of proteins. These results have been summarized by Pace who compares the experimental data of contribution of burial of a $-CH_2-$ group from mutagenesis studies (*ca*. 1.2 kcal/mol) with the ΔG_{tr} for transfer a $-CH_2-$ from water to various solvents [154]. The ΔG_{tr} for transfer of a $-CH_2-$ from water to cyclohexane is the closest to experimental data with a value of *ca.* 1.0 kcal/mol while ΔG_{tr} for transfer of a $-CH_2-$ from water to wet *n*-octanol and other solvents are quite similar with a significantly smaller value of *ca.* 0.5 kcal/mol.

In this study, we will include both the ΔG_{tr} values from water to cyclohexane (CH) and from water to wet *n*-octanol (OT) to estimate the hydrophobic effects. An assumption is made that there is a linear correlation between the hydrophobic free energy of protein and solvent-accessible surface areas (ASA). ASA is calculated as the surface area accessible to a water probe with a radius of 1.4 Å that rolls over the surface of the solute [155]. Instead of using extended tripeptide Gly-X-Gly as a model for the unfolded state, here, for all the methods, we use the actual polypeptide chain in the extended β conformation to model the unfolded state. The crystal structures of the proteins were used to estimate the number of buried hydrophobic side-chains and the number of buried $-CH_2-$ groups from the polar and charged side-chains upon folding. These numbers are then multiplied by the corresponding ΔG_{tr} values based on

cyclohexane or *n*-octanol data. Since we are using the percent burial for each residue, this method is referred as residue-based method (RB) in this study. The estimated contributions of the hydrophobic effect for each protein are shown in Table 7.

In addition to the residue-based method, another way to quantify the hydrophobic effect is to estimate the total change in ASA between the folded and unfolded protein. This method treats the protein molecule as a whole instead of determining transfer energy for each type of amino acid. The energetic contribution of hydrophobic effect ($\Delta G_{H\phi}$) can thus be estimated by the following equation:

$$\Delta G_{H\phi} = k_h [(ASA)_N - (ASA)_U] \qquad (4)$$

where the k_h refers as a hydrophobicity coefficient [156; 157]. There have been several studies trying to find which value of k_h can better estimate the $\Delta G_{H\phi}$ [158; 159]. Different values have been proposed depending on which solvent is used in the solvent transfer model. For the value of cylcohexane-water transfer cofficient, several values have been suggested. Vajda and his colleagues suggested a value of 34 cal mol^{-1} $Å^{-2}$ that is later supported by Dill and Sancho and their colleagues in separate studies [160; 161; 162]. Karplus argued that the polar atoms should be ignored when estimating the hydrophobic effect because the transfer free energies for polar groups are highly dependent on burial environments and aromatic groups should have a smaller value than aliphatic groups [163]. He suggested that 25 cal mol^{-1} $Å^{-2}$ for aliphatic groups and 16 cal mol^{-1} $Å^{-2}$ for aromatic groups are better coefficients for roughly evaluating the hydrophobic effects. Here, we

also use his method to estimate the hydrophobic effect (see Table 7 for more details). The estimated contributions of the hydrophobic effect for each protein by ASA-based method with k_h values from both Vajda and Karplus are shown in Table 7.

By comparing the estimated hydrophobic effects by different methods shown in Table 7, we think the method with n-octanol transfer data underestimates the hydrophobic effect and the Vajda method overestimates the hydrophobic effect. The Karplus method and residue-based method with cyclohexane data give similar results, which suggests the hydrophobicity coefficients reported by Karplus are better than the larger value of 34 cal mol^{-1} $Å^{-2}$ used by Vajda *et al.* Since the residue-based method with cyclohexane transfer data is the closest to the experimentally measured values, we will use this method in the later calculations of this study.

Estimation of the Contribution from Hydrogen Bonding. On average, proteins form 1.1 intramolecular hydrogen bond per residue on folding [164]. Most of the hydrogen bonds are formed between the carbonyl oxygen and the amide hydrogen of the peptide bond. Many polar side chains are also buried and most of these groups also form hydrogen bonds with each other or with peptide groups [165; 166]. The contribution of hydrogen bonds to protein stability has been debated for a long time. Theoretical studies argue that hydrogen bonds are only responsible for maintaining the native structure of the protein and contribute little or unfavorably to protein stability [141]. However, mutational studies suggest hydrogen bonds make a favorable contribution to protein stability. Recently, Myers and Pace suggested that there is a substantial gain in stability from burying a polar group in the interior of protein and forming hydrogen bonds [167]. In

studies of 52 hydrogen-bonding mutants, they show that hydrogen bonds contribute 1.0 to 2.0 kcal/mol per intramolecular hydrogen bond to the stability of protein. In average, one hydrogen bond contributes ~1.2 kcal/mol to protein stability [154]. In this study, we'll make a conservative estimation that each hydrogen bond contributes to the conformational study by 1.0 kcal/mol. The number of hydrogen bonds and their energy contributions are shown in Table 7.

Estimation of the Contribution from Disulfide Bonds. Disulfide bonds covalently stabilize a protein by constraining the unfolded conformations of the protein and thereby decreasing their conformational entropy. This leads to an increase in the free energy of the denatured states of proteins. The decrease in conformational entropy is estimated by calculating the probability that the ends of a polymer chain will simultaneously occur in the same volume element, v_s [168]. The choice of the value of v_s has been a subject of discussion [153; 169]. Based on mutational studies, Pace and his coworkers determined 57.9 $Å^3$ for v_s based on the closest approach distance of two thiols. Using this value, the equation for estimating the effect of a cross-link on the conformational entropy becomes:

$$\Delta S = -2.1 - (3/2)\, R \ln n \qquad (5)$$

where R is the gas constant and n is the number of residues in the loop forming the cross-link [153]. Good agreement has been shown between the predictions of the contribution of cross-link on the conformational stability of protein by this equation and experimental data [137; 153]. Here, we also use this equation to estimate the contributions of

the disulfide bonds for the three proteins in the list for which no experimental data is available (Table 7).

Summary. By simply summing up these major forces contributing to protein stability that are estimated based only on experimental studies, we predicted the conformational stabilities of all the 20 proteins. Table 8 shows the predicted protein stabilities from different combinations of estimates of the hydrophobic effect and conformational entropy and comparisons with the measured stabilities are also given by the average standard deviations. Note the remarkably good agreement between predicted and measured stability when using *n*-octanol data for hydrophobic effect estimation with either conformational entropy estimation method. These are not used further because the *n*-octanol data is not in agreement with the experimental data of hydrophobic effect. The Vajda method will not be used further either because it overestimates the hydrophobic effect. So far, the combined method that is closest based on the experimental data gives an average standard deviation of 71 kcal/mol, for which the residue-based method with cyclohexane data were used for hydrophobic effect estimation and Freire group method with 50% cutoff for conformational entropy.

Using these values the predicted conformational stabilities plotted versus protein size (Figure 19) shows that the predicted stabilities increase almost linearly with the increase of protein size. The difference between the predicted stability and the measured stability for the biggest protein is 360 kcal/mol. Actually, all the results from these methods show linearly increased predicted stabilities with the increase of protein size. Thus, it appears that some forces that contribute to destabilizing the proteins, especially

for big proteins, are not taken into account. One question raised is what forces are not considered when equation (1) is used to estimate protein stability? The equation includes favorable contributions from buried hydrophobic residues, polar groups involved in hydrogen bonding and unfavorable conformational entropy. However, buried polar groups that do not form hydrogen bonds are not taken into account. Savage and Elliott concluded that the number of unfulfilled hydrogen bonds could be a destabilizing contribution and thus correlates with protein stability [170]. As shown in Table 9, proteins bury more polar groups and charged groups as well as more non-polar groups when the size of proteins increases. We will now take a closer look at these buried polar groups to find the forces that are possibly important in destabilizing large proteins.

Buried Charged Groups. Burying a charged groups generally makes a large unfavorable contribution to protein stability due to the Born self energy [171]. Exceptions are made only when very good hydrogen bonds or strong ion pairs are formed (Gilleto and Pace, 1999; Anderson et al. 1990). Both experimental and theoretical studies show that in most cases, replacements of buried charged residues yield more stable proteins [172; 173]. As shown in Table 9, there is a striking difference in the burial of ionizable groups between small proteins and large proteins: of the 20 ionizable groups in 1YRF 32% are buried while of the 155 ionizable groups in 1GLM 69% are buried. The fraction burial of charged groups is more than doubled in our largest proteins. Proteins tend to bury more charged groups as they get larger. This is in accord with the earlier observation of Kajander and his colleagues and suggests that nature may use charge burial as a strategy to reduce protein stability [173].

As shown in Table 10, most of buried charges are hydrogen bonded. On average, buried (≥70% buried) charged groups forms 1.4 hydrogen bonds. However, not all of the buried charged groups are hydrogen bonded and the number with no hydrogen bonds tends to increase when proteins get bigger (Table 10). There are 35 buried charged groups without any hydrogen bonds in the large protein 1GLM while only 1 in the small protein 1YRF.

Among these buried, non-hydrogen-bonded, charged groups, most are not involved in favorable salt bridge interactions. The contribution of salt bridges to protein stability remains disputable. Hendsch and Tidor show that the net contribution of most salt bridges to protein stability is unfavorable due to the large unfavorable desolvation energy that is not fully compensated by favorable electrostatic interactions in the folded protein [172]. Whereas, the Nussinov group found that salt bridges are likely to stabilize the proteins [174].

Now, if we correct the predicted stability by taking these buried charged groups into consideration, the best results (assessed by the lowest average S.D.) we get is 24±20 kcal/mol (Table 11). However, this requires 11 kcal/mol for the energetic cost of burying one non-hydrogen-bonded charged group. Based on what we have learned from mutational studies on buried charged groups, buried charged groups destablize proteins by 2-7 kcal/mol [175; 176]. The value of 11 kcal/mol would obviously be an overestimation of the energy penalty for burying of a charged residue. This suggests that there are other strategies that nature uses to destablize large proteins besides burying charged residues.

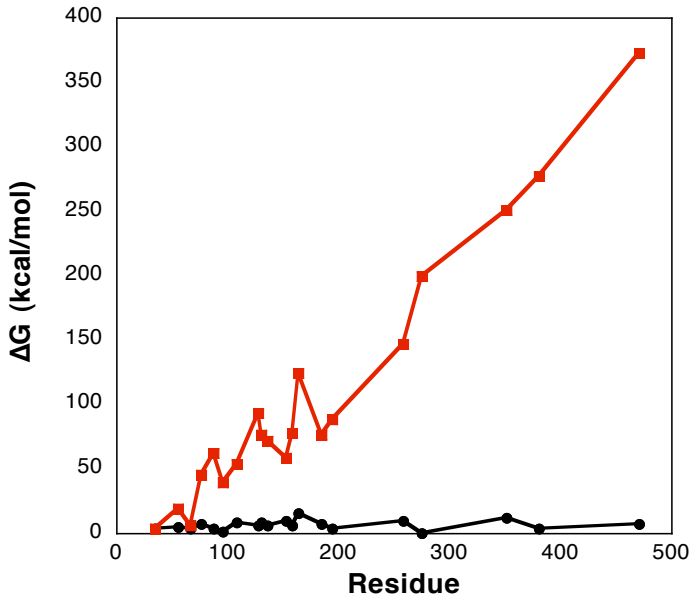

Figure 19: The predicted stability for the 20 proteins (red filled squares) compared to experimentally measured stability (black filled circles). The stability was predicted by equation 1 with using residue-based method based cyclohexane transfer data for hydrophobic effect estimations and Freire group method for the conformational entropy estimations.

Table 8: Estimated conformational stability (ΔG) of all proteins by using different combinations of methods. (kcal/mol)[a].

Protein	ΔG_m [b]	Vajda[c]		Karplus[c]		RB (CH)[c]		RB (OT)[c]	
		Spolar&Record	Freire Group	Spolar&Record	Freire Group	Spolar&Record	Freire Group	Spolar&Record	Freire Group
1YRF	3.3	42	37	13	8	9	4	-9	-13
1PGB	4.6	101	93	20	12	27	19	-3	-11
1CSP	3.5	107	94	-3	-15	18	6	-22	-34
1UBQ	6.7	144	135	33	24	55	46	-3	-12
2HPR	4.2	174	175	41	42	61	62	-5	-4
1RGG	1.8	174	167	30	23	46	39	-13	-20
1A2P	8.8	219	199	64	43	74	54	4	-16
3CHY	5.7	269	261	71	63	102	93	4	-4
1LZ1	8.2	302	277	83	58	101	76	26	1
1SNO	6.1	249	231	54	37	90	72	-6	-24
1IOB	9.1	307	271	49	13	94	58	-17	-53
1RX4	5.9	314	304	62	52	87	78	-11	-20
2LZM	15.0	402	377	112	87	149	124	27	2
1BCX	7.2	445	402	95	53	119	76	11	-32
3ADK	3.9	365	348	57	41	105	88	-26	-43
1CAH	9.3	607	564	179	136	190	147	14	-29
2ST1	0.5	640	632	137	129	207	200	-2	-10
2PEC	12.1	898	845	220	167	304	251	54	1
1DIL	4.2	1015	928	250	163	364	277	79	-9
1GLM	6.7	1173	1107	329	263	439	373	105	39
Avg. SD	-	277±223	240±201	63±59	46±46	89±79	71±67	15±17	17±10

[a]All the estimated conformational stabilities were calculated using equation (1). The estimated contributions from hydrogen bonding and disulfide bonds are the same for each.
[b]ΔG_m, measured protein stability, For 1RGG and 1LZ1, the contribution from disulfide bonds have been subtracted.
[c]Vajda, Karplus, RB(CH) and RB(OT) refer as the methods used for the hydrophobic effects estimations; Spolar&Record and Freire Group, the two methods used for the conformational entropy estimations.

Table 9: Groups burials for the 20 proteins.

PDB ID	N_r[a]	Buried SC[b] Non-polar		Buried SC Polar[c]		Buried Peptide Polar[d]		Total Buried Polar[e]		Buried Ionizable groups[f]	
		#	%	#	%	#	%	#	%	#	%
1YRF	35	10.0	62.5	4.4	48.9	48.5	68.4	52.9	66.1	6.4	32.1
1PGB	56	15.5	73.7	11.0	49.8	84.2	74.5	95.2	70.5	9.8	33.7
1CSP	67	20.7	76.6	11.2	58.9	95.4	70.7	106.6	69.2	11.3	31.3
1UBQ	76	23.6	81.4	14.3	53.0	115.3	75.4	129.6	72.0	14.0	33.4
2HPR	87	31.5	82.9	12.2	41.9	141.7	81.0	153.9	75.4	37.2	37.2
1RGG	96	30.2	71.8	23.5	55.8	139.7	72.4	163.2	69.4	21.5	52.5
1A2P	108	34.4	79.9	26.7	62.0	162.6	75.0	189.3	72.8	29.2	57.3
3CHY	128	51.5	79.2	13.6	43.9	213.1	82.9	226.7	78.7	28.4	47.3
1LZ1	130	50.7	83.1	30.1	61.4	211.1	80.9	241.2	77.8	29.3	48.9
1SNO	136	47.5	80.5	22.4	58.9	215.3	78.8	237.7	76.4	42.1	56.1
1IOB	153	53.4	80.9	36.8	55.7	244.0	79.5	280.8	75.3	33.3	52.0
1RX4	159	63.8	80.8	23.0	62.2	248.6	77.9	271.6	76.3	37.5	42.2
2LZM	164	61.3	82.8	30.8	54.0	275.5	83.7	306.3	79.3	43.9	54.8
1BCX	185	60.2	83.6	58.4	56.2	297.7	80.2	356.1	75.0	27.8	66.3
3ADK	194	64.6	82.9	29.3	61.1	322.8	83.0	352.1	80.6	45.6	41.4
1CAH	259	94.1	84.8	52.6	66.6	425.2	82.2	477.8	80.2	71.4	55.4
2ST1	275	110.5	85.0	65.0	56.0	469.0	85.1	534.0	80.1	41.8	69.7
2PEC	352	126.2	84.7	103.9	67.0	604.1	85.8	708.0	82.3	73.2	62.0
1DIL	381	133.9	88.7	115.8	72.0	650.2	85.2	766.0	82.9	103.2	66.6
1GLM	470	190.4	89.4	124.6	63.3	808.2	85.7	932.8	81.8	106.7	68.9

[a]Number of residues.
[b]Buried SC Non-polar, side chain non-polar groups include side chains of Ala, Val, Ile, Leu, Phe, Tyr (without –OH), Trp, Cys, Met and all the $–CH_2$ groups in polar and charged side chains.
[c]Buried SC Polar, side chain polar groups include the O and $–NH_2$ groups of Asn and Gln and the –OH groups in Tyr, Thr and Ser.
[d]Buried Peptide BB, only the carbonyl O and -NH in the peptide backbone.
[e]Buried Polar, sum of the side chain and peptide backbone polar groups.
[f]Buried Ionizable groups, besides the ionizable groups in Asp, Glu, Lys, Arg and His, N- and C-termini were also taken into account.

Table 10: Burial of charged groups for all the proteins.[a]

PDB ID	N_r	H-Bonded Charged Groups (≥70% buried)	# HB formed	#H-bonds per Buried charged groups	Non-Hydrogen-Bonded Charged Groups (≥70% buried)	Energetic cost (3 kcal/mol per group)
1YRF	35	2	2	1.0	1	3
1PGB	56	6	7	1.2	0	0
1CSP	67	2	3	1.5	2	6
1UBQ	76	6	6	1.0	1	3
2HPR	87	6	8	1.3	1	3
1RGG	96	10	17	1.7	5	15
1A2P	108	20	30	1.5	3	9
3CHY	128	14	19	1.4	3	9
1LZ1	130	14	19	1.4	3	9
1SNO	136	22	31	1.4	5	15
1IOB	153	12	21	1.8	12	36
1RX4	159	17	24	1.4	4	12
2LZM	164	18	22	1.2	12	36
1BCX	185	19	30	1.6	6	18
3ADK	194	12	12	1.0	13	39
1CAH	259	38	54	1.4	11	33
2ST1	275	29	43	1.5	7	21
2PEC	352	39	61	1.6	19	57
1DIL	381	60	89	1.5	20	60
1GLM	470	57	81	1.4	35	105

[a]Charged groups refer as both side chain charged groups in Asp, Glu, Lys, Arg, His and N-, C-termini.

Buried Unsatisfied Polar Groups. As proteins get larger, they bury both more peptide polar groups and side chain polar groups (-OH groups of Ser, Thr and Tyr; O and –NH2 groups of Asn and Gln) as shown in Table 9. Proteins bury more than 70% of their peptide groups and 50% of their side chain polar groups upon folding (Table 9). Most of the peptide bonds and side chain polar groups are involved in hydrogen bonding. On average, proteins form 1.1 hydrogen bonds per residue when they fold, 1.0 of which are by the peptide groups [164]. For buried side chain polar groups, most of them form more than one hydrogen bond. However, proteins also bury unsatisfied polar groups. For instance, there are 27 buried (≥70%) peptide bonds that form no hydrogen bond and 78 that form only one hydrogen bond in our largest protein 1GLM.

The contribution of hydrogen bonding by peptide groups to the conformational stability has been a controversial issue. The peptide backbones have been suggested to be the most polar of the uncharged polar groups and to have a preference for hydrogen bonding to water, rather than to other peptides [152; 177; 178]. It has been thus argued that burying the peptide backbone that contains polar NH and CO groups will involve a significant energetic penalty [179]. In a more recent study, Bolen and Rose propose that the peptide backbone hydrogen bonding play a more important role in protein folding than hydrophobic effects [180]. Based on recent mutational studies, Pace and his co-workers suggest that a buried hydrogen bonded peptide group can overcome the large dehydration penalty and still make a favorable contribution to the stability [154; 167; 181; 182]. While these arguments continue, it seems likely that burying peptide groups without satisfying the hydrogen bond donors or acceptors in the interior of protein will

destabilize the protein. Fleming and Rose suggest that it is so energetically expensive that the likelihood of finding unsatisfied peptide bond in a protein is insignificant [183].

The net favorable contribution of a buried hydrogen bond may come from the combined effect of hydrogen bonding and van der Waals interactions of peptide groups in the interior of a folded protein [139]. For polar groups burial, mutation studies show that polar groups contribute favorably to protein stability when they form hydrogen bonds [182]. In a more recent study, it was shown that the contribution of polar group burial to protein stability is strongly context dependent: burial of a –OH group could be favorable even if it is not hydrogen bonded [184].

Generally speaking, burial of polar groups without forming any hydrogen bond will destabilize the protein due to the desolvation penalty. Here, to see if large proteins bury more non-hydrogen-bonded polar groups as the size of the protein increases, we analyzed the crystal structures of our 20 proteins. The buried non-hydrogen-bonded carbonyl oxygen and amide group from peptide groups and side chain polar groups are treated the same way. As shown in Figure 20, both the buried unsatisfied carbonyl oxygen and amide groups increase, as the proteins get bigger. It is possible that nature uses this approach to destabilize large proteins.

This observation suggests that we need to include these buried unsatisfied polar groups in the stability estimation. As a first approximation, we will apply the same energetic penalty estimations for the carbonyl oxygen and the amide group. All the buried non-hydrogen-bonded carbonyl oxygen and amide groups are listed in Table 11. If we correct the predicted stability by only taking the buried unsatisfied polar groups

into consideration, we get is 21±20 kcal/mol with an estimation of 1.6 kcal/mol for the energetic cost of burying one non-hydrogen-bonded polar group (Table 11).

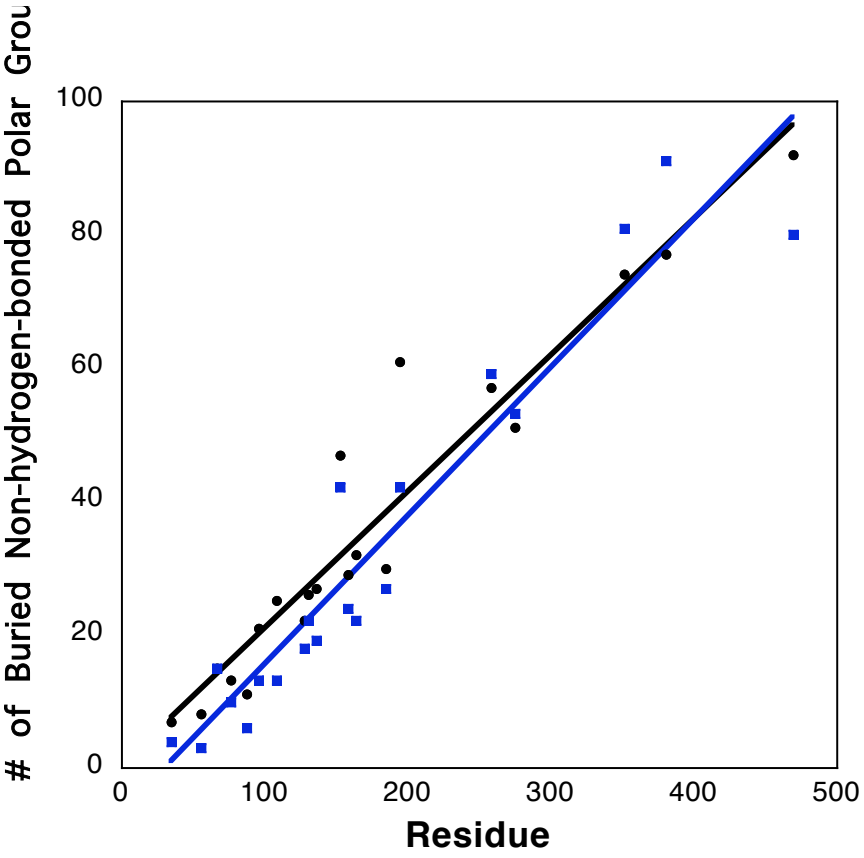

Figure 20: The number of buried non-hydrogen-bonded polar groups increases with the increase of protein size. Shown in the figure are buried non-hydrogen-bonded carbonyl oxygens (black filled circles) and buried non-hydrogen-bonded amide nitrogens (blue filled squares).

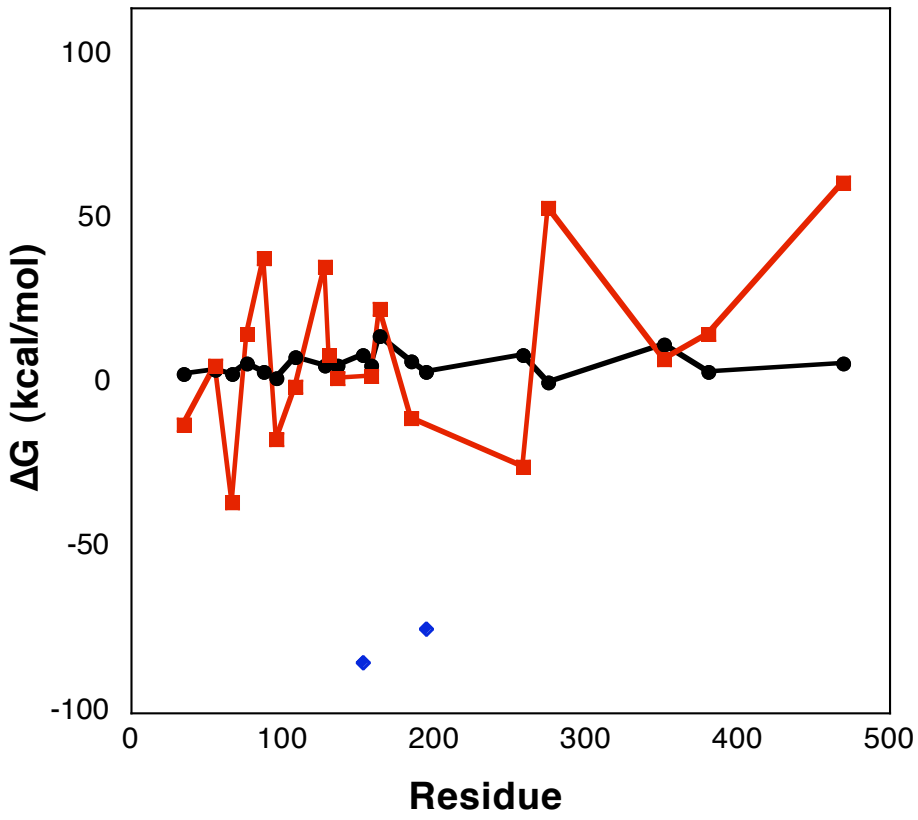

Figure 21: Comparison of the predicted stability after energetic cost corrections with the measured stability for the 20 proteins. Corrections were made for the energetic costs for burying non-hydrogen-bonded charged groups and polar groups (red filled squares) and compared to the measured stability (black filled circles). The outliers (1IOB and 3ADK) are shown in blue filled diamonds.

Packing (van der Waals) Interactions. It is difficult to separate the hydrophobic effects from van der Waals or packing interactions, so it was estimated together as hydrophobic effects here. Therefore, it will also be interesting to know the correlation of van der Waals contribution with protein size. Currently, no experimental data is available to quantify van der Waals interactions in proteins. In a theoretical analysis, Lazaridis, Archontis and Karplus (abbr. as LAK) simulated the unfolding enthalpies in vacuum of four proteins by using the molecular force field CHARMM [141] and dissected into two terms: the van der Waals interaction energy and the electrostatic (chiefly hydrogen bonds) interaction energy. The calculated unfolding energy on a per gram basis for their four proteins are similar, therefore they concluded that all energy components are similar for different proteins and the mechanism of protein stability is essentially independent of protein. By using the same method, we calculated the unfolding energies for the 20 proteins with different sizes shown in Table 12. Our results show that the van der Waals energy per gram/residue increases when proteins get larger (Figure 22A), suggesting that more contribution from van der Waals interactions in large proteins than in small proteins. The reason that LAK get similar values for each of the four proteins is because those four proteins have similar sizes (all medium size proteins with 103 to 153 residues). Their values are also in agreement with our results for different proteins but within same range of sizes (shown as bold region of Table 12). Another term, the electrostatic (chiefly hydrogen bonds) interaction energy per residue, doesn't show obvious size dependence (Figure 22B). Therefore, if these calculations are practical, our estimations for hydrophobic effects might be even lower for large proteins.

Table 11: Estimated conformation stability before and after corrections for the destabilizing forces.

PDB ID	N_r	ΔG_m	Buried[e] Unsatisfied		Energetic Costs (1.2 kcal/mol)	Before Corr.[a]	Corr. after non-hydrogen-bonded buried charged groups[b]	Corr. after non-hydrogen-bonded buried polar groups[c]	Corr. after non-hydrogen-bonded buried charged and polar groups[d]
			N	O		RB (CH) + Freire Group			
1YRF	35	3.3	7	4	13.2	4	-7	-14	-12
1PGB	56	4.6	8	3	13.2	19	19	1	6
1CSP	67	3.5	15	15	36	6	-16	-42	-36
1UBQ	76	6.7	13	10	27.6	46	35	9	15
2HPR	87	4.2	11	6	20.4	62	51	35	39
1RGG	96	1.8	21	13	40.8	39	-16	-15	-17
1A2P	108	8.8	25	13	45.6	54	21	-7	-1
3CHY	128	5.7	22	18	48	93	60	29	36
1LZ1	130	8.2	26	22	57.6	76	43	-1	9
1SNO	136	6.1	27	19	55.2	72	17	-2	2
1IOB	153	9.1	47	42	106.8	58	-74	-84	-85*
1RX4	159	5.9	29	24	63.6	78	34	-7	2
2LZM	164	15.0	32	22	64.8	124	-8	38	23
1BCX	185	7.2	30	27	68.4	76	10	-15	-10
3ADK	194	3.9	61	42	123.6	88	-55	-77	-75*
1CAH	259	9.3	57	59	139.2	147	26	-39	-25
2ST1	275	0.5	51	53	124.8	200	123	34	54
2PEC	352	12.1	74	81	186	251	42	3	8
1DIL	381	4.2	77	91	201.6	277	57	8	15
1GLM	470	6.7	92	80	206.4	373	-12	98	62
Avg. SD	-	-	-	-	-	**71±67**	**24±20**	**21±20**	**18±18***

[a]The RB(CH) method was used for hydrophobic effect estimations and Freire Group residue-specific method for the CE estimations.
[b]The correction was made by subtracting the energetic costs of 11 kcal/mol for each buried non-hydrogen-bonded charged group.
[c]The correction was made by subtracting the energetic costs of 1.6 kcal/mol for each buried non-hydrogen-bonded polar group.
[d]The correction was made by subtracting sum of the energetic costs of 3 kcal/mol for buried charged group and 1.2 kcal/mol for polar group.
[e]All the buried were defined as ≥70% buried.
*Average standard deviation drops to 14±12 after removing these two outliers.

Table 12: Calculated vacuum enthalpy of unfolding, ΔH^{U}_{N}(vac), and its decomposition on caloric/gram basis. (kcal/mol)

PDB ID	**N_r**	vdW	Elec	Δe^{U}_{N}(vdW) (cal/g)	Δe^{U}_{N}(Elec) (cal/g)	Δe^{U}_{N}(vdW) (kcal/residue)	Δe^{U}_{N}(Elec) (kcal/residue)
1YRF	35	160	60	39	15	4.6	1.7
1PGB	56	301	193	49	31	5.4	3.5
1CSP	67	334	188	45	25	5.0	2.8
1UBQ	76	431	206	50	24	5.7	2.7
2HPR	87	469	243	52	27	5.4	2.8
1RGG	96	508	304	48	29	5.3	3.2
1A2P	**108**	**653**	**291**	**54**	**24**	**6.0**	**2.7**
3CHY	**128**	**773**	**337**	**55**	**24**	**6.0**	**2.6**
1LZ1	**130**	**786**	**311**	**53**	**21**	**6.0**	**2.4**
1SNO	**136**	**849**	**292**	**55**	**19**	**6.2**	**2.1**
1IOB	**153**	**901**	**348**	**52**	**20**	**5.9**	**2.3**
1RX4	159	945	340	53	19	5.9	2.1
2LZM	164	1019	369	55	20	6.2	2.2
1BCX	185	1246	788	61	39	6.7	4.3
3ADK	194	1110	395	51	18	5.7	2.0
1CAH	259	1752	681	60	23	6.8	2.6
2ST1	275	1692	778	61	28	6.2	2.8
2PEC	352	2402	1111	64	30	6.8	3.2
1DIL	381	2659	1164	63	28	7.0	3.1
1GLM	470	3236	1498	64	30	6.9	3.2
Avg.	-	-	-	**54±7**	**25±6**	6.0±0.6	2.7±0.7

*Bold font shows the protein size region of LAK's proteins.

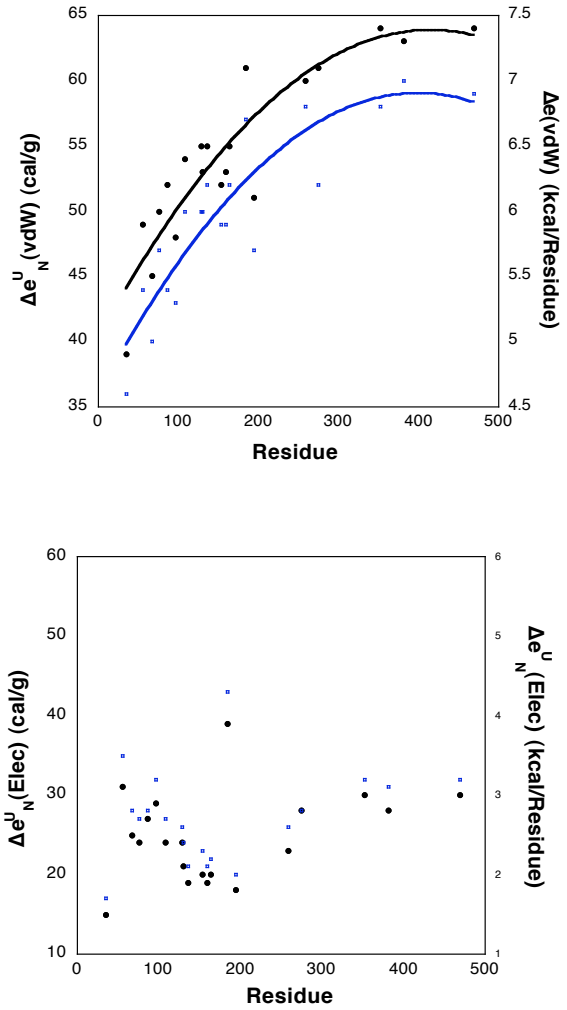

Figure 22: Calculated van der Waals and electrostatic interaction energy. A) The calculated van der Waals energy on a per gram Δe^{U}_{N}(vdW) (black filled circles) and per residue (blue squares) basis as a function of protein size. B). The calculated electrostatic (chiefly hydrogen bond) interaction energy on a per gram Δe^{U}_{N}(Elec) (black filled circles) and per residue (blue filled squares) basis as a function of protein size.

Mutation studies have shown that creation of one or more cavities inside the hydrophobic core of protein can destablize the protein [185; 186], thus it intuitively raises the question of whether large proteins pack more loosely than small proteins. Here we checked the packing densities of all the proteins with two different calculation methods, Voidoo and NSC programs [142; 143]. Different from the results that reported by Liang and Dill [187], both of our results show that the packing densities slightly increase with increase of protein sizes (see Figure 23 in Supplemental Materials). It is noticeable that the trend of the increase of packing densities with protein size is very similar to the trend of the increase of van der Waals interactions with a high correlation coefficient of 0.90 (Figure 24).

More Compact Denatured States for Large Proteins? One possible explanation for large proteins not being more stable than small proteins is that large proteins might have more compact denatured states. It is quite plausible that the unfolded states tend to have more residual hydrophobic cores or residual structures, in other words, more compact denatured state, when the polypeptide chains get longer. The m-value upon urea denaturation has been shown to be proportional to the surface area newly exposed on denaturation. By comparing predicted m-values with two modeled denatured states with experimental measured m-values, Auton and his colleagues [188] show that the denatured states of most proteins are halfway between random coils and compact denatured states and independent of protein size (see Supplemental Material for more). Therefore, we exclude the differences in denatured states as an explanation for the instability of large proteins.

Other Possible Contributions. Other possible explanations for the marginal stability of large proteins have also been taken into consideration. For instance, electrostatic interactions are not included here. Compared to hydrogen bonding and hydrophobic interactions, experimental studies show that the electrostatic interactions among charged groups make only minor contribution (no more than 10 kcal/mol) to protein stability and are highly context dependent [189; 190; 191]. Recently, several studies show that electrostatic interactions not only stabilize the folded state of proteins, but also play an important role in the denatured state [192; 193; 194]. Thus it is suggested the net contributions from electrostatic interactions are relatively small. Hydrogen bonds are stronger when the distance is shorter or the geometry is optimal. Our analysis of the average hydrogen bond length in different sized proteins suggests that the hydrogen bond lengths are also independent of protein sizes.

Summary. If we take both the unsatisfied polar groups and the buried charged groups into the stability prediction, the best results improves to18±18 kcal/mol as shown in Table 11 with an estimation of 1.2 kcal/mol as energetic cost for burying one unsatisfied polar group and 3 kcal/mol as energetic cost for burying one non-hydrogen-bonded charged group. As shown in Figure 22, after the correction made for the destabilizing forces, the predicted stabilities for large proteins show much better agreement with the measured stabilities.

There are two outliers for which the predicted stabilities are much lower than the measured stabilities: 1IOB and 3ADK. Removing the two outliers, the average agreement and standard deviation become much better with a value of 14±12 kcal/mol.

1IOB is an all β-sheet structure protein and has less hydrogen bonds present in the crystal structure: on average, protein forms 0.93 hydrogen bonds per residue in our protein dataset, while 1IOB form only 0.73 per residue. Hydrogen bonds in β-sheets usually have shorter distances than those in α-helices, which is also observed in 1IOB and might be an explanation for the high measured stability of 1IOB. For 3ADK, a possible explanation is that the crystal structure was determined with the present of contained sulfate ions while the hydrogen bonds and hydrophobic effects are calculated for the crystal structure without the presence of the sulfate ions. The hydrogen bonding present in the structure was suggested to be different with or without the presence of sulfate ions and thus relatively less hydrogen bonds are observed in the calculation (0.84 per residue) [148]. The ΔASA_{np} and ΔASA_{p} of 3ADK on a per residue basis are both smaller than that of the proteins with similar size (Table 6). Furthermore, it is noticeable that both 1IOB and 3ADK have nearly doubled the buried unsatisfied charged groups and polar groups compared to proteins with similar sizes. All these factors explain the underestimation of the stability of these two proteins. There is also one positive outlier for which the predicted stabilities are much higher than the measured stability: 2ST1. The relative less buried unsatisfied charged and polar groups could explain the overestimation.

Comparison with Experimental Data. From the above calculations, we found that the energetic cost for burying one non-hydrogen-bonded charged group is ~3 kcal/mol and the energetic cost for burying one non-hydrogen-bonded polar group is ~1.2 kcal/mol. Are these results in agreements with the results that previously found from

experimental mutant studies? In a study of 40 Val → Thr mutants in 11 proteins, Takano and his colleagues show that replacing a $-CH_3$ group with an –OH group will lead to an average decrease in stability of 1.8 kcal/mol. This replacement was unfavorable in 39 out of 40 mutants [182]. For buried charged group, experimental data have shown that replacing a naturally buried non-hydrogen-bonded Asp with Ala can increase the protein stability by 2.8 kcal/mol [176]. Karp and his colleagues showed that replacing a Val buried in the hydrophobic core of Staphylococcal nuclease with Asp can destabilize protein 7.0 kcal/mol at pH 7 [175]. Thus, we could expect an average of ~ 5.0 kcal/mol for buried a non-hydrogen-bonded charged group. The values we estimated from the calculations are slightly smaller than the experimental data because the residue based method with cyclohexane data underestimated the hydrophobic effect and made a conservative estimation for hydrogen bond contribution too. As we mentioned before, experimental studies yield an average value of 1.2 kcal/mol for burying a –CH2 group while the cyclohexane data gives only 1.0 kcal/mol; also, we have made a very conservative estimation (1 kcal/mol) for hydrogen bond (compared to the average contribution of 1.2 kcal/mol per hydrogen bond). If we do a correction by multiplying the hydrophobic effect by 1.2, the results we get are 5.1 kcal/mol for burying one non-hydrogen-bonded charged group and 1.7 kcal/mol for one non-hydrogen-bonded polar group. These values are in great agreement with the values from experimental studies.

CONCLUSIONS

Why are large proteins not more stable than small proteins? When proteins get bigger, the surface-to-volume ratio decreases, which suggests that large proteins have a larger hydrophobic core and tend to bury more residues. However, as for the amino acid composition, no increase was observed for the fraction of hydrophobic residues for large proteins, actually small proteins (25-50) have greater hydrophobic residue content [195]. Therefore large proteins bury more hydrophilic groups as well as more nonpolar groups, but the fractions of total buried polar and nonpolar surface areas remain constant in both large proteins and small proteins (Harpaz et al. 1994). Intuitively, buried nonpolar groups result in larger hydrophobic effect contribution and buried polar groups yields larger hydrogen bond contribution. What offsets these increased favorable contributions of protein stability with the increased protein size? Our calculations show that increased conformational entropy loss upon these increased burials are insufficient to offset the increases of the sum of hydrophobicity and hydrogen bonds of large proteins as protein sizes get bigger. Several possible destabilizing forces for large proteins have then been taken into consideration, including number of buried non-hydrogen-bonded charged and polar groups, different packing density, different denatured states and difference in hydrogen bond length. Our calculations also show that the packing density increases slightly with the increase of protein size, which might be also due to the increased burial of polar groups and hydrogen bonds [196]. The packing interactions, i.e. van der Waals, are also shown to contribute more to the stability in large proteins than small proteins.

Therefore, we conclude that the increased burial of non-hydrogen-bonded charged and polar groups are the major destabilizing forces in large proteins.

Proteins bury more polar groups and charged groups as the protein sizes increase. In the stability prediction by equation 1, the effects of increased buried polar groups are not completely included. By simply taking the energetic costs of burying non-hydrogen-bonded charged groups and polar groups into account, we show a much improved stability prediction that applies to both small proteins and big proteins:

$$\Delta G = \Delta G_{H\phi} + \Delta G_{HB} + \Delta G_{s\text{-}s} - \Delta G_{CE} - \Delta G_{\text{Buried NonHB Pol}} - \Delta G_{\text{Buried NonHB Charged}} \quad (6)$$

where the hydrophobic effects are estimated by using cyclohexane solvent based data and 50% cutoff was used for the residue-specific method for conformational entropy estimation. Although the hydrophobicity effect estimation is based on a model of unfolded state with an accessible area that is too high, which leads to an overestimation with our method, it is still possible that we have underestimated the contribution of hydrophobic effects given the contributions of van der Waals interactions, esp. for large proteins. Our calculation found that burying one non-hydrogen-bonded polar group will destabilize the protein by 1.2 kcal/mol and burying one non-hydrogen-bonded charged group will destabilize the protein by 3 kcal/mol, which show good agreement with the values that found in experimental data. Even a relatively rough estimation of conformational stability of proteins as this is, our study suggests that burying more non-

hydrogen-bonded polar and charged groups are strategies that Mother Nature uses to fine tune the stability of large proteins.

The reasons of burial of more polar and charged residues in large proteins could be two-fold. One is to keep large proteins marginally stable. Most folded proteins are only marginally stable and ~ 2-10 kcal/mol more stable than unfolded proteins [197; 198]. The biological function of proteins requires structural flexibility [199] and high stability could reduce the activity. Also, protein turnover in the cell may require moderation of stability [200]. The other reason could be that large proteins need to bury more polar groups to form structures. By comparing the partially folded α-Lactalbumin and better packed hen lysozyme, Demarest and his colleagues have suggested that burial of polar groups might be necessary for the protein capable of forming a folded structure instead of a molten globule [201]. Long-chain proteins need to bury more polar and charged groups to form a single domain structure. Thus, it is possible that burial of more polar and charged groups could be a result of both protein stability optimization and structure formation.

SUPPLEMENTAL MATERIALS

Table 14: Calculated packing densities (PD) for all the proteins

PDB ID	Molecular Weight (Da)	PD by VOIDOO[a] (Da/Å^3)	PD by NSC[b] (Da/Å^3)
1YRF	4082	1.06	0.50
1PGB	6196	1.09	0.54
1CSP	7365	1.09	0.54
1UBQ	8565	1.06	0.54
2HPR	9042	1.07	0.56
1RGG	10576	1.08	0.50
1A2P	12184	1.08	0.57
3CHY	13966	1.06	0.57
1LZ1	14701	1.09	0.59
1SNO	15482	1.07	0.56
1IOB	17377	1.09	0.58
1RX4	18000	1.09	0.58
2LZM	18635	1.07	0.58
1BCX	20369	1.10	0.61
3ADK	21639	1.08	0.56
1CAH	29115	1.08	0.59
2ST1	27534	1.08	0.61
2PEC	37569	1.09	0.62
1DIL	41942	1.09	0.62
1GLM	50423	1.09	0.63

[a]VOIDOO, program used to calculate the protein volumes written by Kleywegt and Jones [143].
[b]NSC, numerical surface calculation, protein volumes were calculated as described by Eisenhaber and colleagues [142].

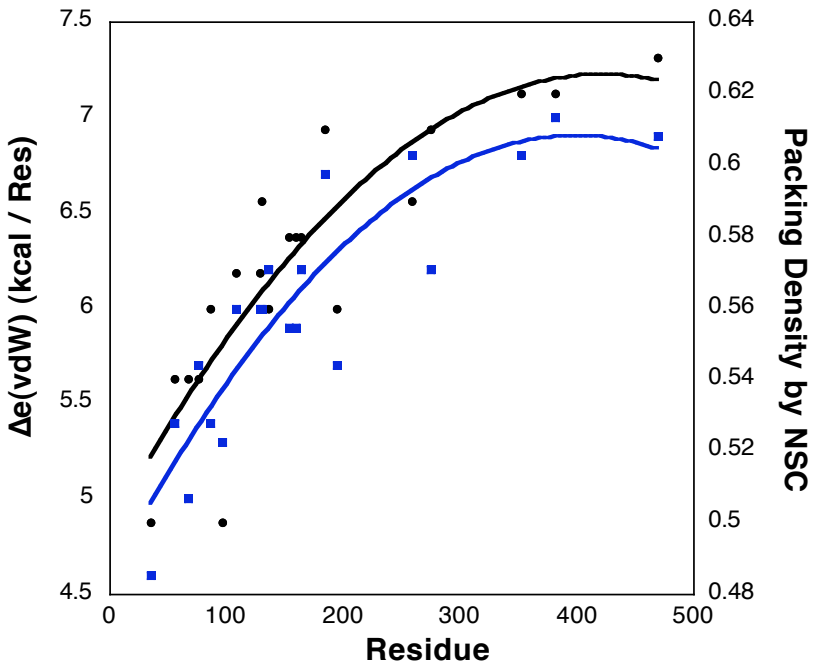

Figure 23: Comparison of the trend of the increase of packing densities (calculated by NSC method) with protein size with the trend of the increase of vdW interactions per residue.

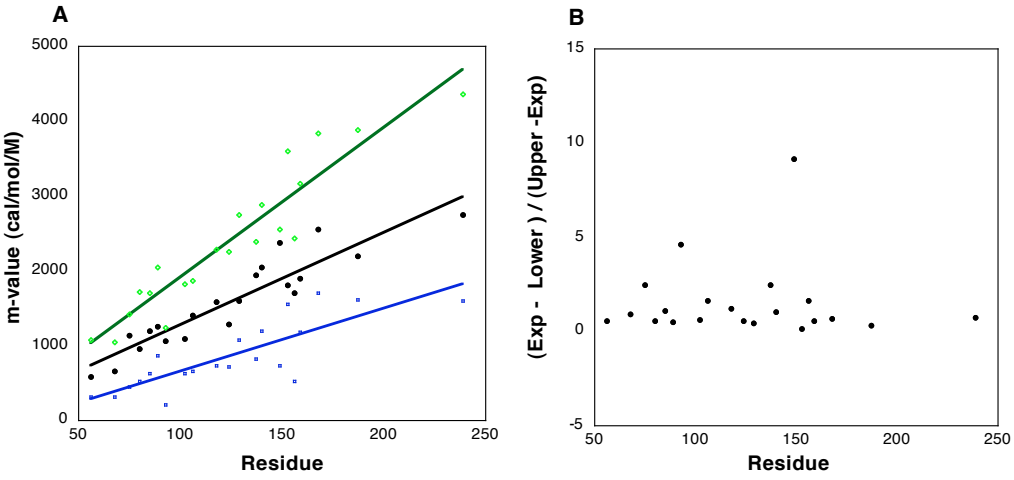

Figure 24: The relationship between experimental and predicted m-values as a function of protein size. Panel A: the predicted m-values: Lower-bound (blue squares), Upper-bound (green diamonds) and the measured m-values (black circles) as a function of protein size; Panel B: The ratio as a function of protein size. The ratio equals (experimental m value – lower-bound predicted m value) / (upper-bound predicted m value – experimental m-value). Large ratio suggests the denatured state is closer to random coil, while small ratio suggests the denatured state is closer to the compacted state.

CHAPTER V

SUMMARY

This dissertation consists of three different projects. The first was dedicated to a structural and thermodynamic study on a series of β-hairpin peptide variants derived from the C-terminus of PGB1 (G-hairpin). The G-hairpin peptide is one of the two reported naturally occurring peptides that can autonomously fold into β-hairpin structure in aqueous solution [1; 2]. Several important noncovalent interactions that contribute to the protein stability, such as hydrophobic interactions and hydrogen bonding, have also been identified in this peptide model system. Single mutations that optimize charge-charge interactions in proteins have been shown to sometimes increase the protein stability [202; 203; 204]. In contrast, our results show that even a single mutation (E42T or E42Q) that removes an unfavorable charge-charge interaction can greatly increase the stability of the G-hairpin peptide. We also examined the structure of a single replacement in the turn region, D47A, to understand why this replacement can stabilize the G-hairpin. Our data suggest that only single replacement of Asp47 with alanine can change the β-hairpin from a 4:4 conformation to a 3:5 hairpin conformation, which leads to a tighter packing of the side-chains of the β-hairpin. Our results prove that the β-turn propensity of the turn residue plays an important role in β-hairpin stability. The additivity of these effects has also been shown in our results, indicating the stabilizing interactions in this model system can be isolated and evaluated. The most stable β-hairpin peptide

(K41+T42+A47) in our study can populate ~96% folded structure in 20% methanol at 4ºC.

The second project correlates the secondary structure propensity and the ability to form amyloid fibrils using PGB1 as a model system. PGB1 is a small protein (56 residue) with a mixed α+β structure. The D40M mutation was incorporated as a context host site to destabilize the wild-type protein and thus to be able to form amyloid fibrils under relatively mild conditions. A number of mutations were introduced to change different local secondary structure propensity while maintaining the global conformational stability of PGB1. Our results show that single mutations that stabilize the α-helix dramatically delay the onset of amyloid formation in PGB1. By contrast, substitutions that localize on the β-sheets have only small effects on the amyloid formation in PGB1. Our results suggest that the α-helix plays an important role in the PGB1 amyloid formation. Our data show the A34F mutant can form amyloid fibrils without visible lag time and has a two-state amyloid formation process. The EM data show that V29A fibrils are morphologically different from other variants; further work is needed to explain these observations.

The third project was focused on answering the question: why are large proteins not more stable than small proteins? Despite the apparent larger hydrophobic interactions and large numbers of hydrogen bonds in large proteins, large proteins and small proteins have similar conformational stability. To answer the question, we selected twenty single-domain globular proteins ranging in size from 35 to 470 residues to analyze. Several major stabilizing forces as well as destabilizing forces have been

investigated for the 20 proteins. Our results suggest that the increased conformational entropy loss in large proteins is not enough to offset the increase in the sum of hydrophobicity and hydrogen bonds of large proteins. Proteins tend to bury more and more non-hydrogen-bonded charged and polar groups as proteins get bigger. With relative conservative estimations for hydrophobic effect and hydrogen bonds, our results found that burying one non-hydrogen-bonded polar group will destabilize the protein by 1.2 kcal/mol and burying one non-hydrogen-bonded charged group will destabilize the protein by 3 kcal/mol. Our results suggest that burying more non-hydrogen-bonded polar and charged groups might be strategies that Mother Nature uses to destabilize large proteins.

REFERENCES

1. Searle, M. S., Williams, D. H. & Packman, L. C. (1995). A short linear peptide derived from the N-terminal sequence of ubiquitin folds into a water-stable non-native beta-hairpin. *Nat. Struct. Biol.* **2**, 999-1006.
2. Blanco, F. J., Rivas, G. & Serrano, L. (1994). A short linear peptide that folds into a native stable beta-hairpin in aqueous solution. *Nat. Struct. Biol.* **1**, 584-590.
3. Dill, K. A. (1990). Dominant forces in protein folding. *Biochemistry* **29**, 7133-7155.
4. Kim, P. S. & Baldwin, R. L. (1990). Intermediates in the folding reactions of small proteins. *Annu. Rev. Biochem.* **59**, 631-660.
5. Karplus, M. & Weaver, D. L. (1994). Protein folding dynamics: the diffusion-collision model and experimental data. *Protein Sci.* **3**, 650-668.
6. Ciani, B., Jourdan, M. & Searle, M. S. (2003). Stabilization of beta-hairpin peptides by salt bridges: role of preorganization in the energetic contribution of weak interactions. *J. Am. Chem. Soc.* **125**, 9038-9047.
7. Ramirez-Alvarado, M., Kortemme, T., Blanco, F. J. & Serrano, L. (1999). Beta-hairpin and beta-sheet formation in designed linear peptides. *Bioorg. Med. Chem.* **7**, 93-103.

8. Kobayashi, N., Honda, S., Yoshii, H. & Munekata, E. (2000). Role of side-chains in the cooperative beta-hairpin folding of the short C-terminal fragment derived from streptococcal protein G. *Biochemistry* **39**, 6564-6571.

9. Shoemaker, K. R., Kim, P. S., Brems, D. N., Marqusee, S., York, E. J., Chaiken, I. M., Stewart, J. M. & Baldwin, R. L. (1985). Nature of the charged-group effect on the stability of the C-peptide helix. *Proc. Natl. Acad. Sci. USA* **82**, 2349-2353.

10. Bierzynski, A., Kim, P. S. & Baldwin, R. L. (1982). A salt bridge stabilizes the helix formed by isolated C-peptide of RNase-A. *Proc. Natl. Acad. Sci. USA* **79**, 2470-2474.

11. Bruch, M. D., Dhingra, M. M. & Gierasch, L. M. (1991). Side chain-backbone hydrogen bonding contributes to helix stability in peptides derived from an alpha-helical region of carboxypeptidase A. *Proteins* **10**, 130-139.

12. Espinosa, J. F., Munoz, V. & Gellman, S. H. (2001). Interplay between hydrophobic cluster and loop propensity in beta-hairpin formation. *J. Mol. Biol.* **306**, 397-402.

13. Santiveri, C. M., Santoro, J., Rico, M. & Jimenez, M. A. (2004). Factors involved in the stability of isolated beta-sheets: turn sequence, beta-sheet twisting, and hydrophobic surface burial. *Protein Sci.* **13**, 1134-1147.

14. Scholtz, J. M., Marqusee, S., Baldwin, R. L., York, E. J., Stewart, J. M., Santoro, M. & Bolen, D. W. (1991). Calorimetric determination of the enthalpy change for the alpha-helix to coil transition of an alanine peptide in water. *Proc. Natl. Acad. Sci. USA* **88**, 2854-2858.

15. Williams, S., Causgrove, T. P., Gilmanshin, R., Fang, K. S., Callender, R. H., Woodruff, W. H. & Dyer, R. B. (1996). Fast events in protein folding: helix melting and formation in a small peptide. *Biochemistry* **35**, 691-697.

16. Chakrabartty, A. & Baldwin, R. L. (1995). Stability of alpha-helices. *Adv. Protein. Chem.* **46**, 141-176.

17. Hughes, R. M. & Waters, M. L. (2006). Model systems for beta-hairpins and beta-sheets. *Curr. Opin. Struct. Biol.* **16**, 514-524.

18. Griffiths-Jones, S. R., Maynard, A. J. & Searle, M. S. (1999). Dissecting the stability of a beta-hairpin peptide that folds in water: NMR and molecular dynamics analysis of the beta-turn and beta-strand contributions to folding. *J. Mol. Biol.* **292**, 1051-1069.

19. de Alba, E., Rico, M. & Jimenez, M. A. (1997). Cross-strand side-chain interactions versus turn conformation in beta-hairpins. *Protein Sci.* **6**, 2548-2560.

20. de Alba, E., Jimenez, M. A., Rico, M. & Nieto, J. L. (1996). Conformational investigation of designed short linear peptides able to fold into beta-hairpin structures in aqueous solution. *Fold. Des.* **1**, 133-144.

21. de Alba, E., Jimenez, M. A. & Rico, M. (1997). Turn residue sequence determines beta-hairpin conformation in designed peptides. *J. Am. Chem. Soc.* **119**, 175-183.

22. Santiveri, C. M., Pantoja-Uceda, D., Rico, M. & Jimenez, M. A. (2005). beta-Hairpin formation in aqueous solution and in the presence of trifluoroethanol: a

(1)H and (13)C nuclear magnetic resonance conformational study of designed peptides. *Biopolymers* **79**, 150-162.

23. Cox, J. P., Evans, P. A., Packman, L. C., Williams, D. H. & Woolfson, D. N. (1993). Dissecting the structure of a partially folded protein: circular dichroism and nuclear magnetic resonance studies of peptides from ubiquitin. *J. Mol. Biol.* **234**, 483-492.

24. Kobayashi, N., Endo, S. & Munekata, E. (1993). Conformational study on the IgG binding domain of protein G. *Peptide Chemistry 1992*, 278-280.

25. Huyghues-Despointes, B. M., Qu, X., Tsai, J. & Scholtz, J. M. (2006). Terminal ion pairs stabilize the second beta-hairpin of the B1 domain of protein G. *Proteins* **63**, 1005-1017.

26. Honda, S., Kobayashi, N. & Munekata, E. (2000). Thermodynamics of a beta-hairpin structure: evidence for cooperative formation of folding nucleus. *J. Mol. Biol.* **295**, 269-278.

27. Andersen, N. H., Olsen, K. A., Fesinmeyer, R. M., Tan, X., Hudson, F. M., Eidenschink, L. A. & Farazi, S. R. (2006). Minimization and optimization of designed beta-hairpin folds. *J. Am. Chem. Soc.* **128**, 6101-6110.

28. Cochran, A. G., Skelton, N. J. & Starovasnik, M. A. (2001). Tryptophan zippers: stable, monomeric beta-hairpins. *Proc. Natl. Acad. Sci. USA* **98**, 5578-5583.

29. Fesinmeyer, R. M., Hudson, F. M. & Andersen, N. H. (2004). Enhanced hairpin stability through loop design: the case of the protein G B1 domain hairpin. *J. Am. Chem. Soc.* **126**, 7238-7243.

30. Zagrovic, B., Sorin, E. J. & Pande, V. (2001). Beta-hairpin folding simulations in atomistic detail using an implicit solvent model. *J. Mol. Biol.* **313**, 151-169.

31. Ma, B. & Nussinov, R. (2000). Molecular dynamics simulations of a beta-hairpin fragment of protein G: balance between side-chain and backbone forces. *J. Mol. Biol.* **296**, 1091-1104.

32. Pande, V. S. & Rokhsar, D. S. (1999). Molecular dynamics simulations of unfolding and refolding of a beta-hairpin fragment of protein G. *Proc. Natl. Acad. Sci. USA* **96**, 9062-9067.

33. Munoz, V., Henry, E. R., Hofrichter, J. & Eaton, W. A. (1998). A statistical mechanical model for beta-hairpin kinetics. *Proc. Natl. Acad. Sci. USA* **95**, 5872-5879.

34. Sheinerman, F. B. & Brooks, C. L., 3rd. (1997). A molecular dynamics simulation study of segment B1 of protein G. *Proteins* **29**, 193-202.

35. Tsai, J. & Levitt, M. (2002). Evidence of turn and salt bridge contributions to beta-hairpin stability: MD simulations of C-terminal fragment from the B1 domain of protein G. *Biophys. Chem.* **101-102**, 187-201.

36. Espinosa, J. F., Syud, F. A. & Gellman, S. H. (2002). Analysis of the factors that stabilize a designed two-stranded antiparallel beta-sheet. *Protein Sci.* **11**, 1492-1505.

37. Pastor, M. T., Lopez de la Paz, M., Lacroix, E., Serrano, L. & Perez-Paya, E. (2002). Combinatorial approaches: a new tool to search for highly structured beta-hairpin peptides. *Proc. Natl. Acad. Sci. USA* **99**, 614-619.

38. Dinner, A. R., Lazaridis, T. & Karplus, M. (1999). Understanding beta-hairpin formation. *Proc. Natl. Acad. Sci. USA* **96**, 9068-9073.

39. Sibanda, B. L., Blundell, T. L. & Thornton, J. M. (1989). Conformation of beta-hairpins in protein structures: a systematic classification with applications to modelling by homology, electron density fitting and protein engineering. *J. Mol. Biol.* **206**, 759-777.

40. Wilmot, C. M. & Thornton, J. M. (1988). Analysis and prediction of the different types of beta-turn in proteins. *J. Mol. Biol.* **203**, 221-232.

41. Haque, T. S. & Gellman, S. H. (1997). Insights on beta-hairpin stability in aqueous solution from peptides with enforced type I' and type II' beta-turns. *J. Am. Chem. Soc.* **119**, 2303-2304.

42. McCallister, E. L., Alm, E. & Baker, D. (2000). Critical role of beta-hairpin formation in protein G folding. *Nat. Struct. Biol.* **7**, 669-673.

43. Grimsley, G. R., Shaw, K. L., Fee, L. R., Alston, R. W., Huyghues-Despointes, B. M. P., Thurlkill, R. L., Scholtz, J. M. & Pace, C. N. (1999). Increasing protein stability by altering long-range coulombic interactions. *Protein Sci.* **8**, 1843-1849.

44. Perl, D., Mueller, U., Heinemann, U. & Schmid, F. X. (2000). Two exposed amino acid residues confer thermostability on a cold shock protein. *Nat. Struct. Biol.* **7**, 380-383.

45. Ramirez-Alvarado, M., Blanco, F. J. & Serrano, L. (2001). Elongation of the BH8 beta-hairpin peptide: electrostatic interactions in beta-hairpin formation and stability. *Protein Sci.* **10**, 1381-1392.

46. Olsen, K. A., Fesinmeyer, R. M., Stewart, J. M. & Andersen, N. H. (2005). Hairpin folding rates reflect mutations within and remote from the turn region. *Proc. Natl. Acad. Sci. USA* **102**, 15483-15487.

47. de Alba, E., Blanco, F. J., Jimenez, M. A., Rico, M. & Nieto, J. L. (1995). Interactions responsible for the pH-dependence of the beta-hairpin conformational population formed by a designed linear peptide. *Euro. J. Biochem.* **233**, 283-292.

48. Searle, M. S., Griffiths-Jones, S. R. & Skinner-Smith, H. (1999). Energetics of weak interactions in a beta-hairpin peptide: electrostatic and hydrophobic contributions to stability from lysine salt bridges. *J. Am. Chem. Soc.* **121**, 11615-11620.

49. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277-293.

50. Cavanagh, J., Fairbrother, W., Palmer, A. & Skelton, N. (1996). *Protein NMR spectroscopy: principles and practice*, Academic Press, San Diego.

51. Levitt, M. & Hirshberg, M. (1997). Calibration and testing of a water model for simulation of the molecular dynamics of proteins and nucleic acids in solution. *Comp. Phy. Comm.* **91**, 215-231.

52. Levitt, M., Hirshberg, M., Sharon, R. & Daggett, V. (1995). Potential energy function and parameters for simulation of the molecular dynamics of proteins and nucleic acids in solution. *Comp. Phy. Comm.* **91**, 215-231.

53. Grindley, T., Lind, J.E., Jr. (1971). PVT properties of water and mercury. *J. Chem. Phys.* **54**, 3983-3989.

54. Vedam, R., G. Holton. (1968). Specific volumes of water at high pressures, obtained from ultasonic-propagation measurements. *J. Acous. Soc. Am.* **43**, 108-116.

55. Gerstein, M., Tsai, J. & Levitt, M. (1995). The volume of atoms on the protein surface: calculated from simulation, using Voronoi polyhedra. *J. Mol. Biol.* **249**, 955-966.

56. Tsai, J., Gerstein, M. & Levitt, M. (1997). Simulating the minimum core for hydrophobic collapse in globular proteins. *Protein Sci.* **6**, 2606-2016.

57. Delauney, B. (1934). Sur la sphère vide. *Bull Acad. Sci. USSR(VII), Classe Sci. Mat. Nat.*, 783-800.

58. Becker, R. A., Chambers, J. M. & Wilks, A. R. (1988). *The New S Language: A Programming Environment for Data Analysis and Graphics.* Wadsworth & Brooks/Cole, Pacific Grove, USA.

59. Smith, C. K., Withka, J. M. & Regan, L. (1994). A thermodynamic scale for the beta-sheet forming tendencies of the amino acids. *Biochemistry* **33**, 5510-5517.

60. Griffiths-Jones, S. R., Sharman, G. J., Maynard, A. J. & Searle, M. S. (1998). Modulation of intrinsic phi, psi propensities of amino acids by neighbouring

residues in the coil regions of protein structures: NMR analysis and dissection of a beta-hairpin peptide. *J. Mol. Biol.* **284**, 1597-1609.

61. Thurlkill, R. L., Grimsley, G. R., Scholtz, J. M. & Pace, C. N. (2006). pK values of the ionizable groups of proteins. *Protein Sci.* **15**, 1214-1218.

62. Munoz, V., Thompson, P. A., Hofrichter, J. & Eaton, W. A. (1997). Folding dynamics and mechanism of beta-hairpin formation. *Nature* **390**, 196-199.

63. Case, D. A., Dyson, H. J. & Wright, P. E. (1994). Use of chemical shifts and coupling constants in nuclear magnetic resonance structural studies on peptides and proteins. *Methods Enzymol.* **239**, 392-416.

64. Wishart DS, S. B. (1994). Chemical shifts as a tool for structure determination. *Methods Enzymol.* **239**, 363-392.

65. Wishart, D. S., Sykes, B. D. & Richards, F. M. (1992). The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry* **31**, 1647-1651.

66. Wüthrich, K. (1986). *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, Inc, New York.

67. Blanco, F. J. & Serrano, L. (1995). Folding of protein-G B1 domain studied by the conformational characterization of fragments comprising its secondary structure elements. *Eur. J. Biochem.* **230**, 634-649.

68. Fesinmeyer, R. M., Hudson, F. M., Olsen, K. A., White, G. W., Euser, A. & Andersen, N. H. (2005). Chemical shifts provide fold populations and register of beta hairpins and beta sheets. *J. Biomol. NMR* **33**, 213-231.

69. Sibanda, B. L. & Thornton, J. M. (1985). Beta-hairpin families in globular proteins. *Nature* **316**, 170-174.

70. Richardson, J. S. (1981). The anatomy and taxonomy of protein structure. *Adv. Protein Chem.* **34**, 167-339.

71. Hutchinson, E. G. & Thornton, J. M. (1994). A revised set of potentials for beta-turn formation in proteins. *Protein Sci.* **3**, 2207-2216.

72. Pastor, M. T., Gimenez-Giner, A. & Perez-Paya, E. (2005). The role of an aliphatic-aromatic interaction in the stabilization of a model beta-hairpin peptide. *Chembiochem.* **6**, 1753-1756.

73. Seshasayee, A. S., Raghunathan, K., Sivaraman, K. & Pennathur, G. (2005). Role of hydrophobic interactions and salt-bridges in beta-hairpin folding. *J. Mol. Model.* **12**, 197-204.

74. Kiehna, S. E. & Waters, M. L. (2003). Sequence dependence of beta-hairpin structure: comparison of a salt bridge and an aromatic interaction. *Protein Sci.* **12**, 2657-2667.

75. Bernado, P., Blackledge, M. & Sancho, J. (2006). Sequence-specific solvent accessibilities of protein residues in unfolded protein ensembles. *Biophys. J.* **91**, 4536-4543.

76. Wilmot, C. M. & Thornton, J. M. (1990). Beta-turns and their distortions: a proposed new nomenclature. *Protein Eng.* **3**, 479-493.

77. Dobson, C. M. (2001). The structural basis of protein folding and its links with human disease. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **356**, 133-145.

78. Sunde, M. & Blake, C. C. (1998). From the globular to the fibrous state: protein structure and structural conversion in amyloid formation. *Q. Rev. Biophys.* **31**, 1-39.

79. Fandrich, M., Fletcher, M. A. & Dobson, C. M. (2001). Amyloid fibrils from muscle myoglobin. *Nature* **410**, 165-166.

80. Guijarro, J. I., Sunde, M., Jones, J. A., Campbell, I. D. & Dobson, C. M. (1998). Amyloid fibril formation by an SH3 domain. *Proc. Natl. Acad. Sci. USA* **95**, 4224-4228.

81. Pertinhez, T. A., Bouchard, M., Tomlinson, E. J., Wain, R., Ferguson, S. J., Dobson, C. M. & Smith, L. J. (2001). Amyloid fibril formation by a helical cytochrome. *FEBS Lett.* **495**, 184-186.

82. Calamai, M., Chiti, F. & Dobson, C. M. (2005). Amyloid fibril formation can proceed from different conformations of a partially unfolded protein. *Biophys. J.* **89**, 4201-4210.

83. Fandrich, M. & Dobson, C. M. (2002). The behaviour of polyamino acids reveals an inverse side chain effect in amyloid structure formation. *EMBO J.* **21**, 5682-5690.

84. Uversky, V. N. & Fink, A. L. (2004). Conformational constraints for amyloid fibrillation: the importance of being unfolded. *Biochim. Biophys. Acta.* **1698**, 131-153.

85. Sunde, M. & Blake, C. (1997). The structure of amyloid fibrils by electron microscopy and X-ray diffraction. *Adv. Protein. Chem.* **50**, 123-159.

86. Fandrich, M., Forge, V., Buder, K., Kittler, M., Dobson, C. M. & Diekmann, S. (2003). Myoglobin forms amyloid fibrils by association of unfolded polypeptide segments. *Proc. Natl. Acad. Sci. USA* **100**, 15463-15468.

87. Grudzielanek, S., Jansen, R. & Winter, R. (2005). Solvational tuning of the unfolding, aggregation and amyloidogenesis of insulin. *J. Mol. Biol.* **351**, 879-894.

88. Smith, D. P., Jones, S., Serpell, L. C., Sunde, M. & Radford, S. E. (2003). A systematic investigation into the effect of protein destabilisation on beta 2-microglobulin amyloid formation. *J. Mol. Biol.* **330**, 943-954.

89. Chiti, F., Webster, P., Taddei, N., Clark, A., Stefani, M., Ramponi, G. & Dobson, C. M. (1999). Designing conditions for in vitro formation of amyloid protofilaments and fibrils. *Proc. Natl. Acad. Sci. USA* **96**, 3590-3594.

90. Litvinovich, S. V., Brew, S. A., Aota, S., Akiyama, S. K., Haudenschild, C. & Ingham, K. C. (1998). Formation of amyloid-like fibrils by self-association of a partially unfolded fibronectin type III module. *J. Mol. Biol.* **280**, 245-258.

91. Ramirez-Alvarado, M. & Regan, L. (2002). Does the location of a mutation determine the ability to form amyloid fibrils? *J. Mol. Biol.* **323**, 17-22.

92. Morel, B., Casares, S. & Conejero-Lara, F. (2006). A single mutation induces amyloid aggregation in the alpha-spectrin SH3 domain: analysis of the early stages of fibril formation. *J. Mol. Biol.* **356**, 453-468.

93. Lansbury, P. T., Jr. (1999). Evolution of amyloid: what normal protein folding may tell us about fibrillogenesis and disease. *Proc. Natl. Acad. Sci. USA* **96**, 3342-3344.

94. Zerovnik, E. (2002). Amyloid-fibril formation. Proposed mechanisms and relevance to conformational disease. *Eur. J. Biochem.* **269**, 3362-3371.

95. Bousset, L., Thomson, N. H., Radford, S. E. & Melki, R. (2002). The yeast prion Ure2p retains its native alpha-helical conformation upon assembly into protein fibrils in vitro. *EMBO J.* **21**, 2903-2911.

96. Plakoutsi, G., Taddei, N., Stefani, M. & Chiti, F. (2004). Aggregation of the Acylphosphatase from Sulfolobus solfataricus: the folded and partially unfolded states can both be precursors for amyloid formation. *J. Biol. Chem.* **279**, 14111-14119.

97. Chiti, F., Stefani, M., Taddei, N., Ramponi, G. & Dobson, C. M. (2003). Rationalization of the effects of mutations on peptide and protein aggregation rates. *Nature* **424**, 805-808.

98. Christopeit, T., Hortschansky, P., Schroeckh, V., Guhrs, K., Zandomeneghi, G. & Fandrich, M. (2005). Mutagenic analysis of the nucleation propensity of oxidized Alzheimer's beta-amyloid peptide. *Protein Sci.* **14**, 2125-2131.

99. DuBay, K. F., Pawar, A. P., Chiti, F., Zurdo, J., Dobson, C. M. & Vendruscolo, M. (2004). Prediction of the absolute aggregation rates of amyloidogenic polypeptide chains. *J. Mol. Biol.* **341**, 1317-1326.

100. Chiti, F., Calamai, M., Taddei, N., Stefani, M., Ramponi, G. & Dobson, C. M. (2002). Studies of the aggregation of mutant proteins in vitro provide insights into the genetics of amyloid diseases. *Proc. Natl. Acad. Sci. USA* **99 Suppl 4**, 16419-16426.

101. Pawar, A. P., Dubay, K. F., Zurdo, J., Chiti, F., Vendruscolo, M. & Dobson, C. M. (2005). Prediction of "aggregation-prone" and "aggregation-susceptible" regions in proteins associated with neurodegenerative diseases. *J. Mol. Biol.* **350**, 379-392.

102. Tartaglia, G. G., Cavalli, A., Pellarin, R. & Caflisch, A. (2004). The role of aromaticity, exposed surface, and dipole moment in determining protein aggregation rates. *Protein Sci.* **13**, 1939-1941.

103. Tartaglia, G. G., Cavalli, A., Pellarin, R. & Caflisch, A. (2005). Prediction of aggregation rate and aggregation-prone segments in polypeptide sequences. *Protein Sci.* **14**, 2723-2734.

104. Wurth, C., Guimard, N. K. & Hecht, M. H. (2002). Mutations that reduce aggregation of the Alzheimer's Abeta42 peptide: an unbiased search for the sequence determinants of Abeta amyloidogenesis. *J. Mol. Biol.* **319**, 1279-1290.

105. Fernandez-Escamilla, A. M., Rousseau, F., Schymkowitz, J. & Serrano, L. (2004). Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat. Biotechnol.* **22**, 1302-1306.

106. Jarrett, J. T. & Lansbury, P. T., Jr. (1993). Seeding "one-dimensional crystallization" of amyloid: a pathogenic mechanism in Alzheimer's disease and scrapie? *Cell* **73**, 1055-1058.

107. Lomakin, A., Chung, D. S., Benedek, G. B., Kirschner, D. A. & Teplow, D. B. (1996). On the nucleation and growth of amyloid beta-protein fibrils: detection of nuclei and quantitation of rate constants. *Proc. Natl. Acad. Sci. USA* **93**, 1125-1129.

108. Nielsen, L., Khurana, R., Coats, A., Frokjaer, S., Brange, J., Vyas, S., Uversky, V. N. & Fink, A. L. (2001). Effect of environmental factors on the kinetics of insulin fibril formation: elucidation of the molecular mechanism. *Biochemistry* **40**, 6036-6046.

109. Teplow, D. B. (1998). Structural and kinetic features of amyloid beta-protein fibrillogenesis. *Amyloid* **5**, 121-142.

110. Harper, J. D. & Lansbury, P. T., Jr. (1997). Models of amyloid seeding in Alzheimer's disease and scrapie: mechanistic truths and physiological consequences of the time-dependent solubility of amyloid proteins. *Annu. Rev. Biochem.* **66**, 385-407.

111. Ramirez-Alvarado, M., Merkel, J. S. & Regan, L. (2000). A systematic exploration of the influence of the protein stability on amyloid fibril formation in vitro. *Proc. Natl. Acad. Sci. USA* **97**, 8979-8984.

112. Alexander, P., Orban, J. & Bryan, P. (1992). Kinetic analysis of folding and unfolding the 56 amino acid IgG-binding domain of streptococcal protein G. *Biochemistry* **31**, 7243-7248.

113. Blanco, F. J., Ortiz, A. R. & Serrano, L. (1997). Role of a nonnative interaction in the folding of the protein G B1 domain as inferred from the conformational analysis of the alpha-helix fragment. *Fold. Des.* **2**, 123-133.

114. Choi, E. J. & Mayo, S. L. (2006). Generation and analysis of proline mutants in protein G. *Protein Eng. Des. Sel.* **19**, 285-289.

115. Goehlert, V. A., Krupinska, E., Regan, L. & Stone, M. J. (2004). Analysis of side chain mobility among protein G B1 domain mutants with widely varying stabilities. *Protein Sci.* **13**, 3322-3330.

116. Gronenborn, A. M., Filpula, D. R., Essig, N. Z., Achari, A., Whitlow, M., Wingfield, P. T. & Clore, G. M. (1991). A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein G. *Science* **253**, 657-661.

117. Kirsten Frank, M., Dyda, F., Dobrodumov, A. & Gronenborn, A. M. (2002). Core mutations switch monomeric protein GB1 into an intertwined tetramer. *Nat. Struct. Biol.* **9**, 877-885.

118. Lindman, S., Xue, W. F., Szczepankiewicz, O., Bauer, M. C., Nilsson, H. & Linse, S. (2006). Salting the charged surface: pH and salt dependence of protein G B1 stability. *Biophys. J.* **90**, 2911-2921.

119. McCallister, E. L., Alm, E. & Baker, D. (2000). Critical role of beta-hairpin formation in protein G folding. *Nat. Struct. Biol.* **7**, 669-673.

120. Meyer, S. C., Huerta, C. & Ghosh, I. (2005). Single-site mutations in a hyperthermophilic variant of the B1 domain of protein G result in self-assembled oligomers. *Biochemistry* **44**, 2360-2368.

121. Park, S. H., O'Neil, K. T. & Roder, H. (1997). An early intermediate in the folding reaction of the B1 domain of protein G contains a native-like core. *Biochemistry* **36**, 14277-14283.

122. Ramirez-Alvarado, M., Cocco, M. J. & Regan, L. (2003). Mutations in the B1 domain of protein G that delay the onset of amyloid fibril formation in vitro. *Protein Sci.* **12**, 567-576.

123. Strop, P., Marinescu, A. M. & Mayo, S. L. (2000). Structure of a protein G helix variant suggests the importance of helix propensity and helix dipole interactions in protein design. *Protein Sci*. **9**, 1391-1394.

124. Gallagher, T., Alexander, P., Bryan, P. & Gilliland, G. L. (1994). Two crystal structures of the B1 immunoglobulin-binding domain of streptococcal protein G and comparison with NMR. *Biochemistry* **33**, 4721-4729.

125. Greenfield, N. J. (1999). Applications of circular dichroism in protein and peptide analysis. *Trac-Trends in Analytical Chemistry* **18**, 236-244.

126. Bennett, M. J., Schlunegger, M. P. & Eisenberg, D. (1995). 3D domain swapping: a mechanism for oligomer assembly. *Protein Sci.* **4**, 2455-2468.

127. Liu, Y., Gotte, G., Libonati, M. & Eisenberg, D. (2001). A domain-swapped RNase A dimer with implications for amyloid formation. *Nat. Struct. Biol.* **8**, 211-214.

128. Louis, J. M., Byeon, I. J., Baxa, U. & Gronenborn, A. M. (2005). The GB1 amyloid fibril: recruitment of the peripheral beta-strands of the domain swapped dimer into the polymeric interface. *J. Mol. Biol.* **348**, 687-698.

129. Byeon, I. J., Louis, J. M. & Gronenborn, A. M. (2003). A protein contortionist: core mutations of GB1 that induce dimerization and domain swapping. *J. Mol. Biol.* **333**, 141-152.

130. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). The protein data bank. *Nucleic Acids Res.* **28**, 235-242.

131. Pace, C. N. & Scholtz, J. M. (1997). Measuring the conformational stability of a protein. In *Protein Structure: A Practical Approach* (Creighton, T. E., ed.), pp. 299-321, IRL Press, Oxford.

132. Nielsen, L., Frokjaer, S., Brange, J., Uversky, V. N. & Fink, A. L. (2001). Probing the mechanism of insulin fibril formation with insulin mutants. *Biochemistry* **40**, 8397-8409.

133. Wei, Y., Huyghues-Despointes, B. M., Tsai, J. & Scholtz, J. M. (2007). NMR study and molecular dynamics simulations of optimized beta-hairpin fragments of protein G. *Proteins* **69**, 285-296.

134. Pace, C. N. & Scholtz, J. M. (1998). A helix propensity scale based on experimental studies of peptides and proteins. *Biophys. J.* **75**, 422-427.

135. Munoz, V. & Serrano, L. (1997). Development of the multiple sequence approximation within the AGADIR model of alpha-helix formation: comparison with Zimm-Bragg and Lifson-Roig formalisms. *Biopolymers* **41**, 495-509.

136. Wintjens, R., Wodak, S. J. & Rooman, M. (1998). Typical interaction patterns in alpha/beta and beta/alpha turn motifs. *Protein Engineering* **11**, 505-522.

137. Pace, C. N., Hebert, E. J., Shaw, K. L., Schell, D., Both, V., Krajcikova, D., Sevcik, J., Wilson, K. S., Dauter, Z., Hartley, R. W. & Grimsley, G. R. (1998). Conformational stability and thermodynamics of folding of ribonucleases Sa, Sa2 and Sa3. *J. Mol. Biol.* **279**, 271-286.

138. Tokuriki, N., Stricher, F., Schymkowitz, J., Serrano, L. & Tawfik, D. S. (2007). The stability effects of protein mutations appear to be universally distributed. *J. Mol. Biol.* **369**, 1318-1332.

139. Hebert, E. J., Giletto, A., Sevcik, J., Urbanikova, L., Wilson, K. S., Dauter, Z. & Pace, C. N. (1998). Contribution of a conserved asparagine to the conformational stability of ribonucleases Sa, Ba, and T1. *Biochemistry* **37**, 16192-16200.

140. Richards, F. M. (1977). Areas, volumes, packing and protein structure. *Annu. Rev. Biophys. Bioeng.* **6**, 151-176.

141. Lazaridis, T., Archontis, G. & Karplus, M. (1995). Enthalpic contribution to protein stability: insights from atom-based calculations and statistical mechanics. *Adv. Protein Chem.* **47**, 231-306.

142. Eisenhaber, F. (1995). The double cubic lattice method: efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies. *J. Comput. Chem.* **16**, 273-284.

143. Kleywegt, G. J. & Jones, T. A. (1994). Detection, delineation, measurement and display of cavities in macromolecular structures. *Acta Crystallogr D Biol Crystallogr* **50**, 178-185.

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

145. Rashin, A. A. (1984). Buried surface area, conformational entropy, and protein stability. *Biopolymers* **23**, 1605-1620.

146. Pal, D. & Chakrabarti, P. (1999). Estimates of the loss of main-chain conformational entropy of different residues on protein folding. *Proteins* **36**, 332-339.

147. Hill, T. L. (1956). *Statistical Mechanics*, McGraw-Hill Book Co., New York.

148. Tian, G. C., Sanders, C. R., 2nd, Kishi, F., Nakazawa, A. & Tsai, M. D. (1988). Mechanism of adenylate kinase: histidine-36 is not directly involved in catalysis, but protects cysteine-25 and stabilizes the tertiary structure. *Biochemistry* **27**, 5544-5552.

149. Doig, A. J. & Sternberg, M. J. (1995). Side-chain conformational entropy in protein folding. *Protein Sci.* **4**, 2247-2251.

150. D'Aquino, J. A., Gomez, J., Hilser, V. J., Lee, K. H., Amzel, L. M. & Freire, E. (1996). The magnitude of the backbone conformational entropy change in protein folding. *Proteins* **25**, 143-156.

151. Shirota, M., Ishida, T. & Kinoshita, K. (2008). Effects of surface-to-volume ratio of proteins on hydrophilic residues: decrease in occurrence and increase in buried fraction. *Protein Sci.* **17**, 1596-1602.

152. Kauzmann, W. (1959). Some factors in the interpretation of protein denaturation. *Adv. Protein Chem.* **14**, 1-63.

153. Pace, C. N., Grimsley, G. R., Thomson, J. A. & Barnett, B. J. (1988). Conformational stability and activity of ribonuclease T1 with zero, one, and two intact disulfide bonds. *J. Biol. Chem.* **263**, 11820-11825.

154. Pace, C. N. (1995). Evaluating contribution of hydrogen bonding and hydrophobic bonding to protein folding. *Methods Enzymol.* **259**, 538-54.

155. Lee, B. & Richards, F. M. (1971). The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **55**, 379-400.

156. Sharp, K. A., Nicholls, A., Friedman, R. & Honig, B. (1991). Extracting hydrophobic free energies from experimental data: relationship to protein folding and theoretical models. *Biochemistry* **30**, 9686-9697.

157. Chothia, C. (1974). Hydrophobic bonding and accessible surface area in proteins. *Nature* **248**, 338-339.

158. Chothia, C. (1976). The nature of the accessible and buried surfaces in proteins. *J. Mol. Biol.* **105**, 1-12.

159. Sanchez-Ruiz, J. (1995). An estimate of shape-related contributions to hydrophobic Gibbs energies. *J. Phys. Chem.* **99**, 12076-12080.

160. Bueno, M., Camacho, C. J. & Sancho, J. (2007). SIMPLE estimate of the free energy change due to aliphatic mutations: superior predictions based on first principles. *Proteins* **68**, 850-862.

161. Chan, H. S. & Dill, K. A. (1997). Solvation: how to obtain microscopic energies from partitioning and solvation experiments. *Annu. Rev. Biophys. Biomol. Struct.* **26**, 425-459.

162. Vajda, S., Weng, Z. & DeLisi, C. (1995). Extracting hydrophobicity parameters from solute partition and protein mutation/unfolding experiments. *Protein Eng.* **8**, 1081-1092.

163. Karplus, P. A. (1997). Hydrophobicity regained. *Protein Sci.* **6**, 1302-1307.

164. Stickle, D. F., Presta, L. G., Dill, K. A. & Rose, G. D. (1992). Hydrogen bonding in globular proteins. *J. Mol. Biol.* **226**, 1143-1159.

165. Baker, E. N. & Hubbard, R. E. (1984). Hydrogen bonding in globular proteins. *Prog. Biophys. Mol. Biol.* **44**, 97-179.

166. McDonald, I. K. & Thornton, J. M. (1994). Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **238**, 777-793.

167. Myers, J. K. & Pace, C. N. (1996). Hydrogen bonding stabilizes globular proteins. *Biophys. J.* **71**, 2033-2039.

168. Flory, P. J. (1956). Theory of elastic mechanism in fibrous proteins. *J. Am. Chem. Soc.* **78**, 5222-5235.

169. Poland, D. C. & Scheraga, H. A. (1965). Statistical mechanics of noncovalent bonds in polyamino acids III interhelical hydrophobic bonds in short chains. *Biopolymers* **3**, 305-313.

170. Savage, H. J., Elliott, C. J., Freeman, C. M. & Finney, J. L. (1993). Lost hydrogen bonds and buried surface area: rationalising stability in globular proteins. *J. Chem. Soc. Faraday Trans.* **89**, 2609-2617.

171. Braun-Sand, S., Strajbl, M. & Warshel, A. (2004). Studies of proton translocations in biological systems: simulating proton transport in carbonic anhydrase by EVB-based models. *Biophys. J.* **87**, 2221-2239.

172. Hendsch, Z. S., Jonsson, T., Sauer, R. T. & Tidor, B. (1996). Protein stabilization by removal of unsatisfied polar groups: computational approaches and experimental tests. *Biochemistry* **35**, 7621-7625.

173. Kajander, T., Kahn, P. C., Passila, S. H., Cohen, D. C., Lehtio, L., Adolfsen, W., Warwicker, J., Schell, U. & Goldman, A. (2000). Buried charged surface in proteins. *Structure* **8**, 1203-1214.

174. Kumar, S. & Nussinov, R. (1999). Salt bridge stability in monomeric proteins. *J. Mol. Biol.* **293**, 1241-1255.

175. Karp, D. A., Gittis, A. G., Stahley, M. R., Fitch, C. A., Stites, W. E. & Garcia-Moreno, E. B. (2007). High apparent dielectric constant inside a protein reflects structural reorganization coupled to the ionization of an internal Asp. *Biophys. J.* **92**, 2041-2053.

176. 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. & Pace, C. N. (2005). Asp79 makes a large, unfavorable contribution to the stability of RNase Sa. *J. Mol. Biol.* **354**, 967-978.

177. Schellman, J. A. (1955). The stability of hydrogen-bonded peptide structures in aqueous solution. *C. R. Trav. Lab. Carlsberg [Chim]* **29**, 230-259.

178. Wolfenden, R. (1978). Interaction of the peptide bond with solvent water: a vapor phase analysis. *Biochemistry* **17**, 201-204.

179. Honig, B. & Cohen, F. E. (1996). Adding backbone to protein folding: why proteins are polypeptides. *Fold. Des.* **1**, R17-R20.

180. Bolen, D. W. & Rose, G. D. (2008). Structure and energetics of the hydrogen-bonded backbone in protein folding. *Annu. Rev. Biochem.* **77**, 339-362.

181. Pace, C. N., Horn, G., Hebert, E. J., Bechert, J., Shaw, K., Urbanikova, L., Scholtz, J. M. & Sevcik, J. (2001). Tyrosine hydrogen bonds make a large contribution to protein stability. *J. Mol. Biol.* **312**, 393-404.

182. Takano, K., Yamagata, Y. & Yutani, K. (2001). Contribution of polar groups in the interior of a protein to the conformational stability. *Biochemistry* **40**, 4853-4858.

183. Fleming, P. J. & Rose, G. D. (2005). Do all backbone polar groups in proteins form hydrogen bonds? *Protein Sci.* **14**, 1911-1917.

184. Takano, K., Scholtz, J. M., Sacchettini, J. C. & Pace, C. N. (2003). The contribution of polar group burial to protein stability is strongly context-dependent. *J. Biol. Chem.* **278**, 31790-31795.

185. Eriksson, A. E., Baase, W. A., Zhang, X. J., Heinz, D. W., Blaber, M., Baldwin, E. P. & Matthews, B. W. (1992). Response of a protein structure to cavity-creating mutations and its relation to the hydrophobic effect. *Science* **255**, 178-183.

186. Lee, B. (1993). Estimation of the maximum change in stability of globular proteins upon mutation of a hydrophobic residue to another of smaller size. *Protein Sci.* **2**, 733-738.

187. Liang, J. & Dill, K. A. (2001). Are proteins well-packed? *Biophys. J.* **81**, 751-766.

188. Auton, M., Holthauzen, L. M. & Bolen, D. W. (2007). Anatomy of energetic changes accompanying urea-induced protein denaturation. *Proc. Natl. Acad. Sci. USA* **104**, 15317-15322.

189. Pace, C. N., Laurents, D. V. & Thomson, J. A. (1990). pH dependence of the urea and guanidine hydrochloride denaturation of ribonuclease A and ribonuclease T1. *Biochemistry* **29**, 2564-2572.

190. Grimsley, G. R., Shaw, K. L., Fee, L. R., Alston, R. W., Huyghues-Despointes, B. M., Thurlkill, R. L., Scholtz, J. M. & Pace, C. N. (1999). Increasing protein stability by altering long-range coulombic interactions. *Protein Sci.* **8**, 1843-1849.

191. Whitten, S. T. & Garcia-Moreno, E. B. (2000). pH dependence of stability of staphylococcal nuclease: evidence of substantial electrostatic interactions in the denatured state. *Biochemistry* **39**, 14292-14304.

192. Cho, J. H. & Raleigh, D. P. (2005). Mutational analysis demonstrates that specific electrostatic interactions can play a key role in the denatured state ensemble of proteins. *J. Mol. Biol.* **353**, 174-185.

193. Makhatadze, G. I., Loladze, V. V., Ermolenko, D. N., Chen, X. & Thomas, S. T. (2003). Contribution of surface salt bridges to protein stability: guidelines for protein engineering. *J. Mol. Biol.* **327**, 1135-1148.

194. Pace, C. N., Alston, R. W. & Shaw, K. L. (2000). Charge-charge interactions influence the denatured state ensemble and contribute to protein stability. *Protein Sci.* **9**, 1395-1398.

195. Shen, M., Davis, F. P. & Sali, A. (2005). The optimal size of a globular protein domain: A simple sphere-packing model. *Chem. Phys. Let.* **405**, 224-228.

196. Schell, D., Tsai, J., Scholtz, J. M. & Pace, C. N. (2006). Hydrogen bonding increases packing density in the protein interior. *Proteins* **63**, 278-282.

197. Pace, C. N., Shirley, B. A., McNutt, M. & Gajiwala, K. (1996). Forces contributing to the conformational stability of proteins. *Faseb. J.* **10**, 75-83.

198. Pace, C. N., Trevino, S., Prabhakaran, E. & Scholtz, J. M. (2004). Protein structure, stability and solubility in water and other solvents. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **359**, 1225-1235.

199. Fields, P. A. (2001). Review: protein function at thermal extremes: balancing stability and flexibility. *Comp. Biochem. Physiol. A. Mol. Integr. Physiol.* **129**, 417-431.

200. Parsell, D. A. & Sauer, R. T. (1989). The structural stability of a protein is an important determinant of its proteolytic susceptibility in Escherichia coli. *J. Biol. Chem.* **264**, 7590-7595.

201. Demarest, S. J., Zhou, S. Q., Robblee, J., Fairman, R., Chu, B. & Raleigh, D. P. (2001). A comparative study of peptide models of the alpha-domain of alpha-lactalbumin, lysozyme, and alpha-lactalbumin/lysozyme chimeras allows the elucidation of critical factors that contribute to the ability to form stable partially folded states. *Biochemistry* **40**, 2138-2147.

202. Gribenko, A. V. & Makhatadze, G. I. (2007). Role of the charge-charge interactions in defining stability and halophilicity of the CspB proteins. *J. Mol. Biol.* **366**, 842-856.

203. Matousek, W. M., Ciani, B., Fitch, C. A., Garcia-Moreno, B., Kammerer, R. A. & Alexandrescu, A. T. (2007). Electrostatic contributions to the stability of the GCN4 leucine zipper structure. *J. Mol. Biol.* **374**, 206-219.

204. Jelesarov, I. & Karshikoff, A. (2008). Defining the role of salt bridges in protein stability. In *Protein Structures, Stability and Interactions* (Shriver, J. W., ed.), pp. 362-410, Humana Press, Totowa, NJ.

VITA

Name: Yun Wei

Address: 440 Reynolds Medical Building
1114 TAMU, College Station, TX 77843-1114

Email Address: yun_wei@tamu.edu

Education: B.S., Biotechnology, Wuhan University, 2002
Ph.D., Biochemistry, Texas A&M University, 2009

Publication:

Wei, Y., Huyghues-Despointes, B.M., Tsai, J., Scholtz, J.M. (2007) NMR study and molecular dynamics simulations of optimized β-hairpin fragments of protein G. *Proteins* **69**: 285-296.
Wei, Y., Pace, C.N., Scholtz, J.M. (2009) Why large proteins are not more stable than small proteins. *(In Process)*
Wei, Y., Scholtz, J. M. (2009) A kinetic study of amyloid formation on Protein G *(In Process)*