UNDERSTANDING FORCES THAT CONTRIBUTE TO PROTEIN STABILITY: APPLICATION FOR INCREASING PROTEIN STABILITY

A Dissertation

by

HAILONG FU

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

UNDERSTANDING FORCES THAT CONTRIBUTE TO PROTEIN STABILITY: APPLICATION FOR INCREASING PROTEIN STABILITY

A Dissertation

by

HAILONG FU

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:

Co-Chairs of Committee,	J. Martin Scholtz
	C. Nick Pace
Committee Members,	Gary R. Kunkel
	Gregory D. Reinhart
Head of Department,	Gregory D. Reinhart

May 2009

Major Subject: Biochemistry

ABSTRACT

Understanding Forces That Contribute to Protein Stability: Application for Increasing Protein Stability. (May 2009)

Hailong Fu, B.S., University of Science and Technology of China

Co-Chairs of Advisory Committee: Dr. J. Martin Scholtz
Dr. C. Nick Pace

The aim of this study is to further our understanding of the forces that contribute to protein stability and to investigate how site-directed mutagenesis might be used for increasing protein stability. Eleven proteins ranging from 36 to 370 residues have been studied here. A 36-residue VHP and a 337-residue VlsE were used as model systems for studying the contribution of the hydrophobic effect on protein stability. Mutations were made in both proteins which replaced bulky hydrophobic side chains with smaller ones. All variants were less stable than their wild-type proteins. For VHP, the destabilizing effects of mutations were smaller when compared with similar mutations reported in the literature. For VlsE, a similarity was observed. This different behavior was investigated and reconciled by the difference in hydrophobicity and cavity modeling for both proteins. Therefore, the stabilizing mechanism of the hydrophobic effect appears to be similar for both proteins.

Eight proteins were used as model systems for studying the effects of mutating non-proline and non-glycine residues to statistically favored proline and glycine residues in β-turns. The results suggest that proline mutations generally increase protein stability,

provided that the replaced residues are solvent exposed. The glycine mutations, however, only have a stabilizing effect when the wild-type residues have ϕ, ψ angles in the Lα region of Ramachandran plot. Nevertheless, this strategy still proves to be a simple and efficient way for increasing protein stability.

Finally, using a combination of eight previously identified stabilizing mutations; we successfully designed two RNase Sa variants (7S, 8S) that have both much higher T_ms and conformational stabilities than wild-type protein over the entire pH range studied. Further studies of the heat capacity change upon unfolding (ΔC_ps) for both proteins and their variants suggest that residual structure may exist in the denatured state of the 8S variant. An analysis of stability curves for both variants suggests that they achieve their stabilization through different mechanisms, partly attributed to the different role of their denatured states. The 7S variants may have a more rigid denatured state and the 8S variant may have a compact denatured state in comparison with that of wild-type RNase Sa.

DEDICATION

To my parents and sisters for their love and support

ACKNOWLEDGEMENTS

I thank my advisors, Dr. Martin Scholtz and Dr. Nick Pace, for allowing me to pursue projects of my choosing and for their guidance, encouragement and invaluable advice, which helped tremendously. I greatly appreciate all the advice from my committee members Dr. Gary Kunkel, Dr. Gregory Reinhart and Dr. David Giedroc. I thank Dr. Gerald Grimsley for being a good friend and my guru in many ways. I also thank all members of Dr. Scholtz's lab and Dr. Pace's lab for their help, including past member Dr. Saul Trevino with whom I studied and worked closely. I thank my friend Dr. Changxia Sun for being a true friend. Finally, I thank my family for their support and patience.

TABLE OF CONTENTS

CHAPTER Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

CHAPTER I

INTRODUCTION TO PROTEIN STABILITY

Proteins are involved in most biological processes. Globular proteins must fold into a unique, compact three dimensional structure (native state) in order to carry out their functions. Since the ground breaking work of Anfinsen on Ribonuclease A,[1] it has been widely accepted that the folding of this native state is determined by the protein's amino acid sequence. In contrast, the unfolded form of the protein (denatured state) was first proposed to be solvent-exposed and behave as a random coil, with no residual structure.[2] However, more and more evidence suggests that there is considerable residual structure in the denatured state of proteins. [3-11] Under any set of solvent conditions, the native state (N) and denatured state (D) are in the following equilibrium,

$$N \leftrightarrow D \tag{1.1}$$

The conformational stability of a protein, ΔG_U, can be defined as the free energy change for this equilibrium reaction,

$$\Delta G_U = G_D - G_N \tag{1.2}$$

Therefore for a protein to fold, ΔG_U must be positive. Under physiological conditions, the stabilities of most globular proteins studied are only 5 to 15 kcal/mol.[12] Since studies with variant forms of proteins show that many can be made more stable, it is not clear why the wild-type forms of most proteins are only marginally stable. One possibility is

---

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

that low stability facilitates protein turnover and degradation.[13] Another possibility is that evolution has chosen a marginal stability to maximize function.[14]

The conformational stability of a protein can be determined using the equation

$$\Delta G_U = -RT\ln K_{eq} \quad (1.3)$$

where K_{eq} is the equilibrium constant for the denaturation reaction (Eq. 1.1). The K_{eq} is usually determined experimentally as a function of a denaturing agent, such as temperature or chemical denaturants (GuHCl or urea), then ΔG_U is determined indirectly.[15]

For a protein to remain folded, the native state must have a lower free energy than the denatured state. The denaturation of protein is therefore the consequence of breaking labile (non-covalent) bonds that maintain this lower energy of native state.[16; 17] The major non-covalent forces that contribute favorably to protein stability are the van der Waals interactions, hydrogen bonding and hydrophobic effect. The stabilizing effects of these interactions are largely opposed by the major destabilizing force, which is the conformational entropy loss upon protein folding. Other forces, such as electrostatic interactions, can be either favorable or unfavorable, depending on the context.

The non-covalent forces that contribute to protein stability

van der Waals interactions

Often referred as dispersion forces or London forces, van der Waals attractions are dipole-dipole interactions resulting from transient, random fluctuations of electrons in neutral molecules. These fluctuations in one molecule can give rise to a set of

instantaneous dipole moments in that molecule, which in turn distorts the electron shell of a neighboring atom, resulting in a dipole coupling that produces attractive forces. In addition, van der Waals interactions also include the short-range repulsive forces resulting from the unfavorable spatial overlap of electron orbitals.[18]

The strength of van der Waals interactions depends strongly on the distance, r, between the atoms. The attractive dispersion forces are counterbalanced by the repulsion of the electron shells as the interacting atoms approach one another. The van der Waals interactions are often estimated by the Lennard-Jones 6, 12 potential

$$E_{vdw} = -A/r^6 + B/r^{12} \quad (1.4)$$

which includes the inverse twelfth power repulsive term and the inverse sixth power attractive term. The parameters of A and B have been evaluated from experimental data.[19] The van der Waals radius is given by the distance at which the function above reaches a minimum.[20]

Van der Waals attractions, although transient and weak, can provide an important component of protein stability because of their number. The hydrophobic interior of folded protein is highly packed like a crystal,[20-22] implying that a high number of close contacts are made in the folded form of the protein. The total van der Waals interactions could therefore contribute hundreds of kcal/mol to the stability of a typical globular protein.[20] Amino acid mutations that alter the internal packing of a protein often lead to a decrease in protein stability, supporting the important role of packing.[23-25]

Hydrophobic effect

Since the seminal work of Kauzmann nearly 50 years ago, the hydrophobic effect has been thought to be the dominant force driving protein folding.[17; 26] Based on the analyses of thermodynamic data characterizing the transfer of hydrocarbons from water to non-polar solvents, Kauzmann predicted that the association of non-polar amino acid side chains would be the major contribution to protein stability.

When a protein folds, most of its non-polar side-chains are buried in the protein interior.[27; 28] This burial results from the preference of non-polar atoms for a non-polar environment. This behavior is often described as the hydrophobic effect. The quantitative estimates for the hydrophobic effect are usually made using solvent transfer experiments by measuring the ΔG_{tr} for transferring different amino acid side-chains from aqueous solvent to different non-polar solvents, detergent micelles, or the vapor phases. The results have led to many hydrophobic scales.[27; 29-33] Although significant differences exist among these scales, the ΔG_{tr} values for the non-polar side-chains from these solvents are almost all positive, suggesting that they prefer the non-polar environment. At room temperature, this unfavorable transfer energy results from the unfavorable change in entropy due to the increased ordering of water molecules around the non-polar side chains needed for solvation.[17]

For more than twenty years, the contribution of hydrophobic effect to protein stability has been studied extensively using site-directed mutagenesis. The substitution of larger hydrophobic residues with smaller ones generally destabilizes a protein.[34-43] Pace reviewed and analyzed a set of 72 hydrophobic variants of 4 different proteins and

estimated that the contribution of one -CH_2- group burial to protein stability is ≈ 1.3 kcal/mol.[44] In a survey of a larger set of hydrophobic variants by Zhou and Zhou, this contribution was found to be 1.1 kcal/mol.[45] These results have excellent correlations with the corrected octanol-to-water transfer free energy.[46]

The measured ΔΔG values for the hydrophobic variants, however, exhibit a large variance. Since the protein interior closely resembles the packing density for crystals, the stronger van der Waals interactions in the folded protein can make a large contribution to the hydrophobic effect due to this tight packing.[36; 38; 47-52] The direct correlation of this ΔΔG values with pure van der Waals interaction parameters is only partially successful in a limited protein set,[36; 47; 49; 53] suggesting there is another contributing factor to the hydrophobic effect that varies among different variants. Eriksson et al.[39] crystallized 7 Leu to Ala and Phe to Ala variants in T4 lysozyme and found that the ΔΔG values could be attributed to a constant term and a variable term. The constant term correlates well with the transfer energy from water to n-octonal and the variable term was found to have linear dependence on the volumes of the cavities created by the mutations. Similar conclusions were also reached from the theoretical calculations and experimental results in other studies.[41; 54; 55]

Hydrogen bonding

A hydrogen bond is a weak electrostatic interaction which forms between covalently bonded hydrogen atom and a strongly electronegative atom with a lone pair of electrons,

$$\text{—}D^{\delta-}\text{—}H^{\delta+}\cdots A^{\delta-}\text{—} \qquad (1.5)$$

The hydrogen atom carries a partial positive charge induced by its covalently bonded donor atom (D) and interacts favorably with the acceptor atom (A) that carries a partial negative charge.

The important role of hydrogen bonds in globular structures was first demonstrated by the classic work of Pauling where he proposed the two major secondary structure elements, strands and helices, were stabilized by hydrogen bonds.[56; 57] This importance is further strengthened by the fact that almost every polar group capable of forming a hydrogen bond in a folded protein does so.[58-60] On average, proteins form 1.1 intromolecular hydrogen bonds per residue.[61] Nevertheless, the contribution of hydrogen bonds to protein stability has been debated for years. Theoretical studies suggest that hydrogen bonds make little contribution to protein stability.[62-64] Kauzmann also declared that "hydrogen bonds give a marginal stability to ordered structure" based on the low free energy for the formation of hydrogen bonds between urea molecules.[17] Similar conclusions were reached when other small model compounds were used.[26; 65-68] However, the interpretation of the results from these experiments is hindered by the difficulty in estimating the association state of the model compounds in solution; therefore the estimates of the free energy of hydrogen bonding are usually not reliable.

Site-directed mutagenesis on proteins has provided a method to assess the contribution of hydrogen bonds to protein stability. Most studies show that hydrogen bonds contribute favorably to protein stability.[69-72] In agreement with these results from proteins, studies of the helix-coil transition suggested that hydrogen bonds contribute about 1 kcal/mol/hydrogen-bond to peptide stability.[73; 74] Hydrogen bonding is now

generally accepted as a major contributor to protein stability comparable to hydrophobic interactions. On average, hydrogen bonds contribute 1 to 2 kcal/mol/hydrogen-bond to protein stability,[71; 72; 75-78] whether they are from side chains or peptide groups.

Electrostatic interactions

Electrostatic interactions between charges are important in most aspects of the physics, chemistry and biology. Coulomb's law states that the interaction energy of two charges i and j can be calculated as,

$$E = q_i q_j / \varepsilon r_{ij} \quad (1.6)$$

where q_i and q_j are the appropriate electronic charges, ε is the dielectric constant and r_{ij} is the distance between the charges. Quantification of the contribution of individual electrostatic interaction to protein stability is difficult, because the charges on the side chains of Arg, Lys, Asp, Glu, His and the N and C termini are not point charges, but are distributed over groups of atoms. In addition, the heterogeneity of the protein medium and the boundary between the protein and solvent makes it difficult to define the dielectric constant around the charges. Values of ε in folded proteins have been experimentally determined to range between 10 and 39.[79]

Due to the shorter r_{ij} and smaller ε values, the electrostatic interactions of charged groups in the native state are expected to be much stronger than that in the denatured state.[80] The charged groups are distributed in the native protein in a way that the total interactions between these charges are generally favorable.[81] However, the contribution of electrostatic interactions to protein stability is at most 10 kcal/mol as shown by studies of the pH dependence of protein stability.[82]

Charges that are closer (<4Å) in distance in proteins are called ion pairs. A salt bridge is a special type of ion pair in which hydrogen bonding interactions are also present between opposite charged groups. A single buried ion pair or salt bridge can contribute 3 – 5 kcal/mol to protein stability.[83; 84] In contrast, the contribution to stability of surface salt bridges is considerably less.[85; 86] Whether ion pairs contribute to the protein stability is still debated, since evidence suggests that the number of ion pairs in protein is usually small and these ion pairs are usually not conserved in a protein family.[87]

Theoretical studies also suggest that the electrostatic interactions of ion pairing are not enough to compensate for the desolvation energy for burying charged groups.[88] Indeed, the burial of a charged residue triad made an unfavorable contribution to the stability of arc repressor, in contrast to burial of hydrophobic residues.[89] Completely buried charged groups usually form hydrogen bonds, and thus contribute favorably to protein stability. Depending on the context, a buried charged group can contribute favorably or unfavorably to protein stability. For example, the buried Asp 33 of RNase Sa forms 3 good hydrogen bonds and contribute favorably to protein stability. On the other hand, the buried non-hydrogen bonded Asp 79 in RNase Sa destabilizes the protein by ≈ 3 kcal/mol.[90]

Taken together, the contribution of electrostatic interactions to protein stability is smaller in comparison with the two major stabilizing forces, the hydrophobic effect and hydrogen bonding.

Conformational entropy

An unfolded protein has an enormous number of conformations that it can access through rotations around the Cα-N(ϕ), Cα-C(ψ) and side chain (χ) bonds. However, in a folded protein such rotations are largely constrained, and the protein is generally restricted to a small number of conformations. Therefore, the folding of a protein results in a large decrease in conformational entropy ($\Delta S_{conf}<0$). This large negative ΔS_{conf} is the major force opposing protein folding.

The contribution of any given residue to the conformational entropy of a protein can be estimated by a statistical approach,

$$\Delta S_{conf} = R \ln z \qquad (1.7)$$

where R is the gas constant and z is the number of the conformations of equal energy available to the residue in the unfolded state.[91] This contribution was estimated to be 4.1 cal/mol/K/res by Kauzmann.[17] Later, a similar value of 4.2 cal/mol/K/res was obtained by Privalov based on calorimetric data of several small proteins.[92] More recently, using thermodynamic data from model compounds and protein unfolding studies, Spolar and Record have estimated the chain conformational entropy ΔS_{conf} to be 5.6 cal/mol/K/res (1.7 kcal/mol/res at 25 ºC).[93]

The chain conformational entropy can be further divided into the side chain conformational entropy and the backbone conformational entropy. Both components have been examined in detail using both experimental and computational approaches.[94-96] The combination of the contribution of both components are in good agreement with

the measured values for the chain conformational entropy determined by Spolar and Record,[93] suggesting that such estimates are reasonably accurate.

Disulfide bonds contribute substantially to protein stability by lowering the conformational entropy of denatured state. For proteins in which disulfide bonds have been removed by mutation or chemical modification, the contribution of the disulfide bonds to the conformational stability can be estimated using[97]

$$\Delta S_{conf} = -2.1\text{-}3/2R \ln n \quad (1.8)$$

where n is the number of residues in the loop formed by the disulfide bond. This equation predicts that the conformational stability of a protein will be raised ≈ 3.0 kcal/mol by adding one disulfide bond linking 15 residues.

Model systems for studying the hydrophobic effect

Our model systems for studying the contribution of hydrophobic effect to protein stability involves a small protein and a large protein. They are the C terminal subdomain of villin "head piece" (VHP) and the *Borrelia burgdorferi* outer surface lipoprotein Variable major protein-Like Sequence, Expressed (VlsE).

VHP

The small protein studied is the 36-residue Villin C terminal subdomain (VHP). Villin is a protein that crosslinks f-actin filaments in the cytoskeleton. The Villin C-terminal domain is also called the Villin "head piece". VHP is the C-terminal subdomain of this 8kDa "head piece".[98]

VHP folds into a three helices cluster as shown by its crystal structure (Fig. 1a).[99] It has a compact core consisting of three phenylalanine residues and several other hydrophobic residues. The Phe residues are completely buried and play an important role in stabilizing the protein.[100]

VHP is a good model for studying the stability of a small protein. The folding kinetics and structure of VHP36 have been extensively studied both experimentally and computationally.[101; 102] This protein might be the smallest protein studied so far. It has no disulfide bond and does not require cofactors for folding.

The folding process for VHP is completely reversible in urea (Fig. 2a). At pH 7 it has a T_m of 74ºC and a conformational stability of 2.8 kcal/mol at 25 °C.

VlsE

VlsE is an outer surface lipoprotein of *Borrelia burgdorferi*. VlsE undergoes antigenic variation through an elaborate gene conversion mechanism and is thought to play a major role in the immune response to the Lyme disease *Borrelia*. The suggested function of VlsE is to present a variable region to the host immune system.[103]

The crystal structure of VlsE has been solved[104] and is shown in Fig. 1b. There are 11 helices and 4 short strands. Like VHP, this protein has no disulfide bonds or cofactors that are required for folding. VlsE is the largest single domain protein that folds in a two-state process. At pH 7, the unfolding in urea is nearly 100% reversible (Fig. 2b), and the conformational stability is 4.5 kcal/mol at 25 °C.[105]

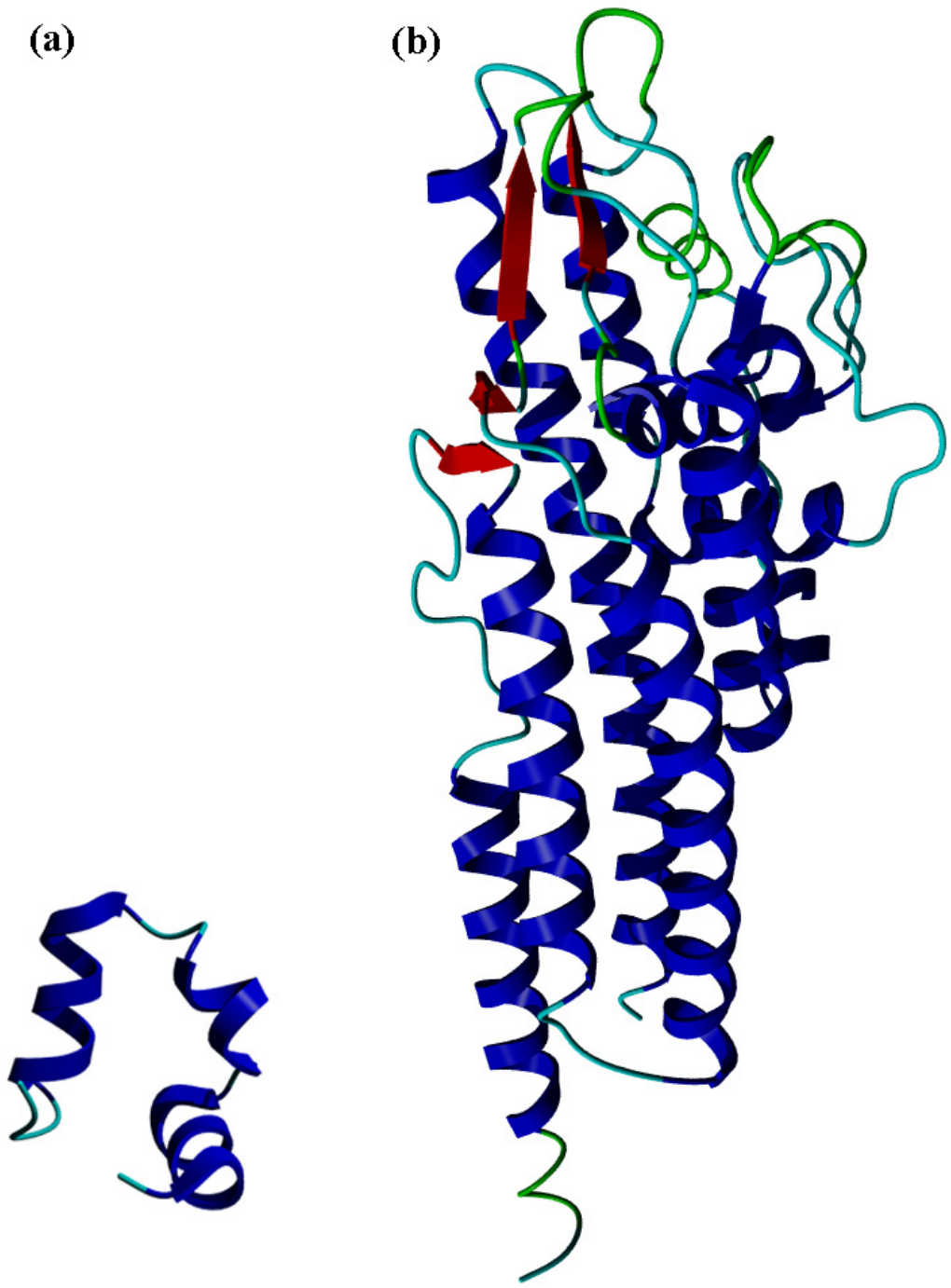

Fig. 1. Ribbon diagram showing the structure of VHP and VlsE. (a) VHP, the pdb file used is 1YRI. [99] (b)VlsE, the pdb file used is 1L8W.[104] Figures were generated using program YASARA.[106]

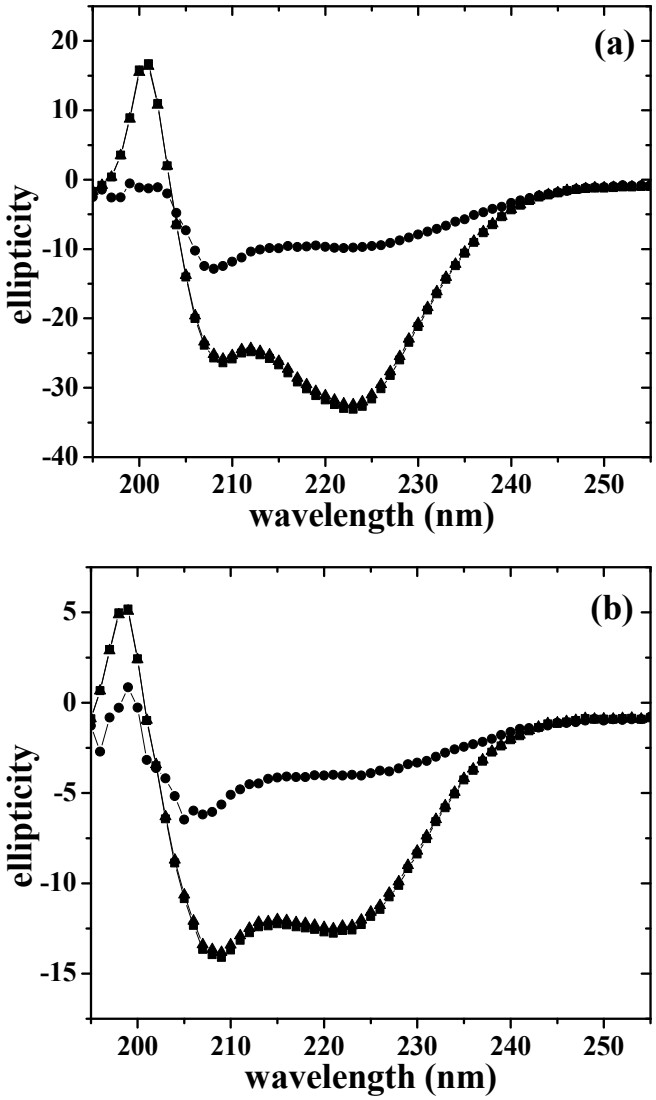

Fig. 2. Far-UV CD spectra for VHP and VlsE. (a) VHP spectra showing Native (■), Denatured (●) and Renatured (▲). (b) VlsE spectra showing Native (■), Denatured (●) and Renatured (▲).

Methods for increasing protein stability

There has always been a great interest in increasing protein stability for both basic research and industrial applications, since the conformational stability of a typical protein is only 5 to 15 kcal/mol.[12]. *In vitro*, protein stability can be increased by adding small molecules that bind to the native state,[107; 108] or by adding osmolytes.[109; 110] Alternately, protein stability can be increased intrinsically by changing protein amino acid sequence. Although *in vitro* selection[111; 112] and the consensus approach[113; 114] have been used to a limited extent to increase protein stability, most approaches involve site-directed mutagenesis in order to best optimize the various forces that contribute to protein stability.

As illustrated in Fig. 3, protein stability can be increased by increasing the free energy difference between the native state and the denatured state. This can be done by either decreasing the free energy of native state (I) or increasing the free energy of denatured state (II) or both.

Many strategies can be used to decrease the free energy of the native state either by the introduction of favorable interactions or by the removal of unfavorable interactions. Favorable interactions can be introduced by improving the packing interactions in the protein hydrophobic core, [115-117] adding hydrogen bonds or salt bridges,[90; 118; 119] optimizing the secondary structure propensity,[120-124] and optimizing the surface charge-charge interactions.[125-130] An example of removing unfavorable interactions would be to remove buried charged amino acid from the hydrophobic core.[90; 131-134]

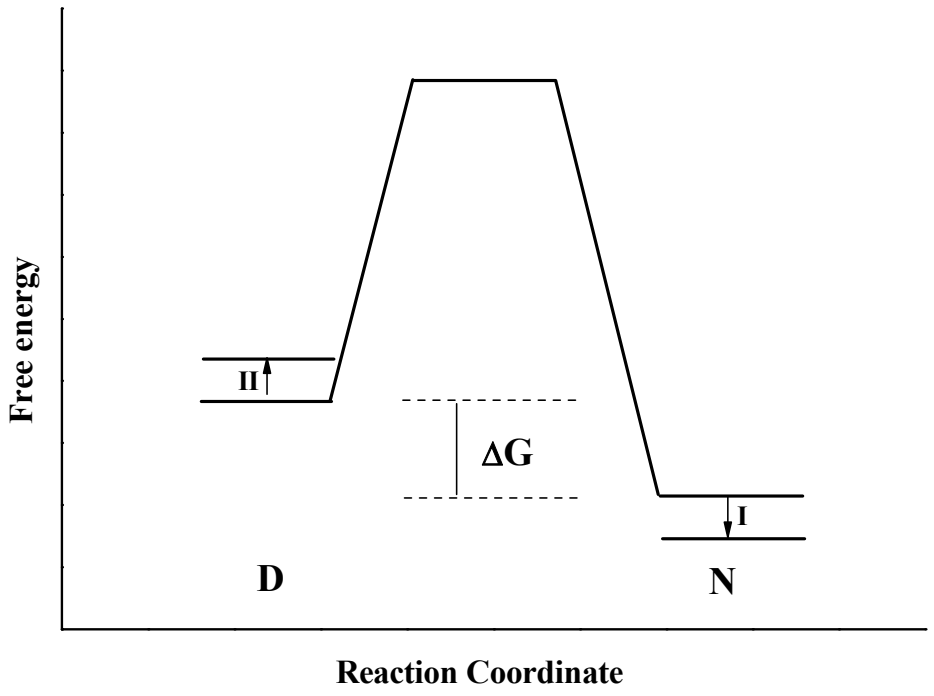

Fig. 3. Energy diagrams for protein folding reaction. Arrows indicate two methods to increase protein stability: by decreasing the free energy of native state energy (I) or increase the free energy of denatured state energy (II).

The strategies for increasing the free energy of the denatured state usually involve reducing the conformational entropy, Consequently, the denatured state is destabilized by restricting the conformational freedom. This can be done by the introduction of disulfide bonds,[135-139] or the introduction of rigid residue such as Pro,[140-144] or the substitution of a Gly with residues with a β-carbon.[141]

The strategies discussed above can result in large increase in protein stability, however, this can be problematic and difficult to achieve. For example, the introduction of disulfide bonds and Pro residues, as well as repacking a protein interior can cause steric strain.[24; 139; 145; 146] The optimization of secondary structures and removal of buried charges are limited to the low availability of such candidates,[131] and the stabilization effects of salt bridges may be only marginal in some cases.[147] The optimization of long-range charge-charge interactions appears promising,[127] although the accurate predictions for the stabilizing effects are difficult, because such interactions also influence the denatured state.[148; 149]

Previously, a successful strategy to increase protein stability was developed in our lab. By making mutations to Pro and Gly in solvent exposed turns, we successfully increased the conformational stability of RNase Sa by ≈ 1 kcal/mol for each mutation.[140] Here we extended this strategy by using 8 different proteins. In addition, we combine other strategies[90; 130] to achieve the most stability possible using using the model protein RNase Sa.

Research objectives

The major objectives of this work were: (1) To measure the stability difference between the variants of VHP and VlsE in order to gain a better understanding of the contribution of hydrophobic effect to the conformational stability of globular proteins. (2) To apply a previously developed strategy by making variants of 8 proteins that have mutations at sites in solvent exposed turns, to increasing protein stability. (3) To rationally design thermostable RNase Sa variants by the combination of different methods and study the stabilizing mechanism.

CHAPTER II

MATERIALS AND METHODS

Reagents for protein purification were of reagent grade and were purchased from various vendors including Sigma-Aldrich (St. Louis, MO), Fisher Scientific (Pittsburgh, PA) and EMD Chemicals Inc (Gibbstown NJ). Reagents and buffer solution for spectroscopic experiments were of best grade available also from the above vendors. Urea was from Amresco (Solon, OH) and Guanidine HCl was from MP Biomedicals; both were of proteomics grade and used without further purification.

Over-expression of RNase T1, Sa2, Sa3 were performed using plasmids derived from pEH100.[150] Plasmids for over-expression of all other proteins were all derived from pET vectors (Novagen).[151] Site-directed mutagenesis was performed using QuikChange™ Kit from Stratagene (La Jolla, CA). DNA oligonucleotides for cloning and mutagenesis were from IDT Integrated DNA Technology (Corallville, IA) and used without further purification. Restriction enzymes and polymerases used in DNA manipulations were from New England Biolabs (Ipswich, VA). Purification kit for plasmids and PCR products were from Qiagen (Valencia, CA). DNA sequencing was done in the Laboratory for Plant Genome Technology, Texas A&M University.

Protein overexpression was performed in different *E. coli* strains for different proteins. For VHP, Cold Shock Protein and Maltose Binding Protein, *E. coli* strain BL21 (DE3) (Novagen) was used. For VlsE and RNase Sa, *E. coli* strain C41(DE3)[152] was used. For RNase HI *E. coli* strain pLys BL21 (DE3) was used. For RNase T1, RNase

Sa2 and RNase Sa3 *E. coli* strain RY1988 (MQ) was used. For cloning and mutagenesis and plasmid miniprep procedure, *E. coli* strain XL-1 blue was used. Tryptone and Yeast Extract for bacterial cultures were obtained from Difco (Becton Dickinson , Sparks, MD). Media like ZYM505, ZYM 5052 were prepared as described.[153] Anion ion exchange medium DEAE sepharose and Q-Sepharase were from Amersham Biosciences (Uppsala Sweden). Sephadex C-25 cation ion exchange resin, G-50 and G-75 size exclusion gel filtration resins were from Sigma-Aldrich (St. Louis, MO).

All stability data were analyzed using the Origin 7 data analysis program from OriginLab Corporation. Visualization, structure manipulation and other calculations were performed using program packages SwissPDB viewer[154] or YASARA[106]. Accessible surface area and hydrogen bonding calculations were performed using the in-house program p*fis*.[75] Vector NTI advance 10 (Invitrogen) was used to generates plasmid maps of pHFVHP and pHFSa, and ClustalX[155] was used for protein sequence alignment.

Cloning and mutagenesis

Cloning of VHP

The VHP gene was cloned into a pET-22b(+) vector using standard PCR and restriction enzyme cutting and insertion methods. A diagrammatic representation of the procedure and the final construct is shown in Fig. 4.

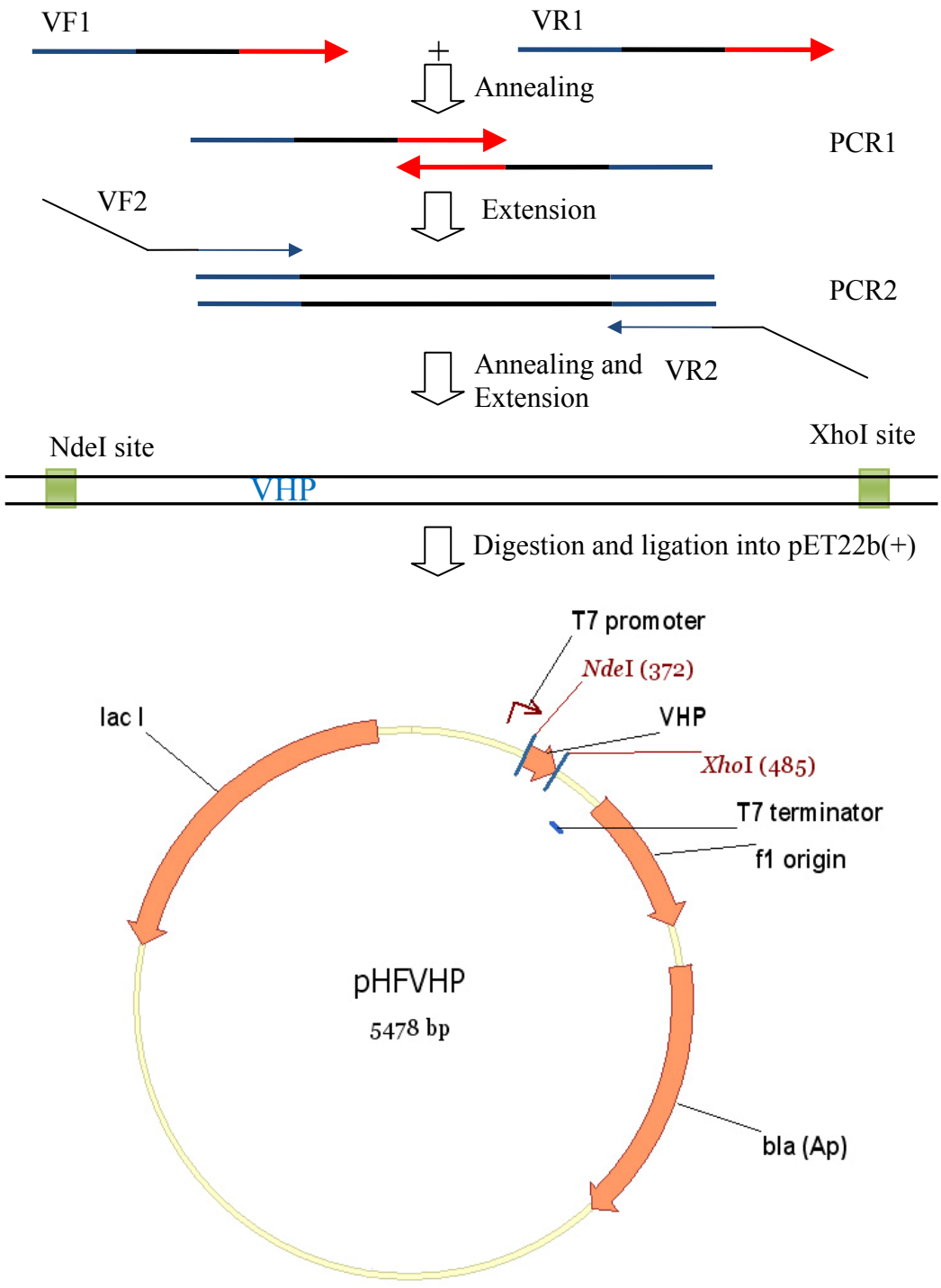

Fig. 4. Construction of VHP expression vector pHFVHP. The vector map was generated using Invitrogen Vector NTI 10 software.

The gene coding for VHP was constructed by standard PCR methods from four DNA oligonucleotides that span the whole coding region plus restriction enzyme sites that were necessary for cloning into vectors. The primers used are shown in Table 1.

Table 1. Primers used to construct and clone VHP into pET-22b(+)

PCR-1
VF1 *TCTGACGAAGACTTCAAAG*CTGTTTTTGGCATGACCCGCTCTGCCTTTGCCAACCTGCCG
VR1 *GAACAGACCTTTCTCTTT*TTTCAGGTTCTGCTGTTTCCACAGCGGCAGGTTGGCAAAGGC
PCR-2
VF2 GATTAGCAA**CATATG**CTG*TCTGACGAAGACTTCAAAG*
VR2 GAATCAGTAT**CTCGAG**TTA*GAACAGACCTTTCTCTTT*
Underlined sequences are complementary to each other. Bold sequences are restriction enzyme sites.

For both construction and amplification of the VHP gene, Turbo^{TM} polymerase was used with 50 cycles of amplification, which consists of a 40-second extension time, a 30-second annealing time and a 30-second melting time at 95^{o}C, with the amplification preceded by a 1 minute hold at 95^{o}C. A typical amplification schedule is shown in Fig. 5.

The annealing temperature was set for 50 ºC for both PCR reactions to assure the initial priming. The extension temperature was suggested by the vendor and the reaction mixture was prepared according to the manufacturer's manual.

95 ºC	95 ºC	50 ºC	72 ºC	72 ºC	4 ºC
1 min	30 sec	30 sec	40 sec	1 min	∞
Melt	Melt	Annealing	Extension	Extension	Hold
		50 cycles			

Fig. 5. A representative PCR amplification schedule. This schedule was used for the construction of VHP gene. The temperatures and cycles time shown here are typical for amplification using the Turbo^{TM} polymerase and for short procedures.

After the first PCR for the middle part of the gene was done, the amplification was confirmed using 2% agrose Gel. The product was purified using PCR purification kit to remove the enzymes and excessive primers. The product was used as a template in a second PCR reaction to add the extra VHP gene and appropriate restriction enzyme sites. The second product was also confirmed using 2% agrose gel and purified before cutting with NdeI and XhoI. The cloning vector pET-22b(+) was also cut with the same enzymes. The purified PCR product and vector were mixed and ligated using T4 ligase according the manufacturer's suggestion.

The ligation reaction mixture was directly transformed into chemically competent XL-1 blue *E. coli* cells without inactivation of the ligase as suggested by the manufacturer. Briefly, 2 µl of the reaction mixture was transformed into 50 µl of XL-1 blue cells and incubated on ice for 20 minutes before heat shock for 45 seconds at 42 ºC. After that, the cells were incubated on ice for extra 20 minutes and the entire

transformation mixture was plated onto a LB agar plate supplemented with 50 μg/ml of ampicillin. The plate was then incubated over night in a 37 ºC incubator after which the colonies were tested for insertion by colony PCR. For the colony PCR, 20 μl PCR reactions were setup and the T7 promoter primer and T7 terminator primer were used. The template DNA was either a pET-22b(+) plasmid which was used as a negative control or the plasmid that was provided by the bacterial colony that was picked from the selective plate. Taq polymerase was used for these reactions and the procedure was similar to that used above. The PCR reactions were analyzed on 2% agrose gel and positive clones were selected based on the presence of appropriate bands on the gel which were migrating differently from the negative control. Those colonies with positive clones were cultured in 2 ml ZYM505 media overnight and plasmid DNA was extracted using a QIAprep Spin Miniprep kit then sequenced to confirm the presence and accuracy of the VHP insertion.

Cloning of RNase Sa

Traditionally we used plasmid pEH100[150] to express RNase Sa. A ribonuclease inhibitor barstar is coexpressed from this plasmid to suppress the toxicity of RNase Sa to the *E. coli* host cells. However, some stabilizing RNase Sa variants became highly toxic to *E. coli* cells even at the presence of barstar, thus killing the cells. Therefore, the RNase Sa gene and the barstar gene were both cloned from the pEH100 into a pET-duet vector so that we could use a different *E. coli* strain C41(DE3), which was shown to tolerate toxic proteins.[156] By cloning the gene into a pET vector, we could also use the T7 expression system and take advantage of the newly developed auto induction

method.[153] The primers used for cloning the RNase Sa gene and the barstar gene into pET-duet vector are shown in Table 2.

Table 2. Primers used to clone RNase Sa and barstar into pET-duet

Barstar gene amplification
NcoIBarF (sense primer)
AGAAAGGAGCCG**CCATGG**AAAAAGCAG
BamHIBarR (anti-sense primer)
GCGTTCATCTCCCATT**GGATCC**TATTAAGAAAG
RNase Sa gene amplification
NdeIPhoSa (sense primer)
CAGG**CATATG**AAACAAAGCACTATTGC
XhoISa (anti-sense primer)
GATCA**CTCGAG**GGAAAGTGAAATCGTTAG
Bold sequences indicate restriction enzyme sites.

Both the barstar and the RNase Sa Genes were cloned into pET-duet vector by the similar procedure as that of VHP. Briefly, the barstar gene was first amplified using PCR from plasmid pEH100. Purified PCR product was cut with the restriction enzymes NcoI and BamHI and ligated into pET-duet vector cut with the same enzymes. Ligation mixture was transformed into XL-1 blue *E. coli* cells and the positive clones were identified by colony PCR. The plasmid from the positive clones was extracted and sequenced to confirm the presence and accuracy of the barstar insertion. The insertion of the RNase Sa gene followed the same procedure except that the restriction enzymes used were NdeI and XhoI and the template for insertion was the plasmid obtained from the previous step.

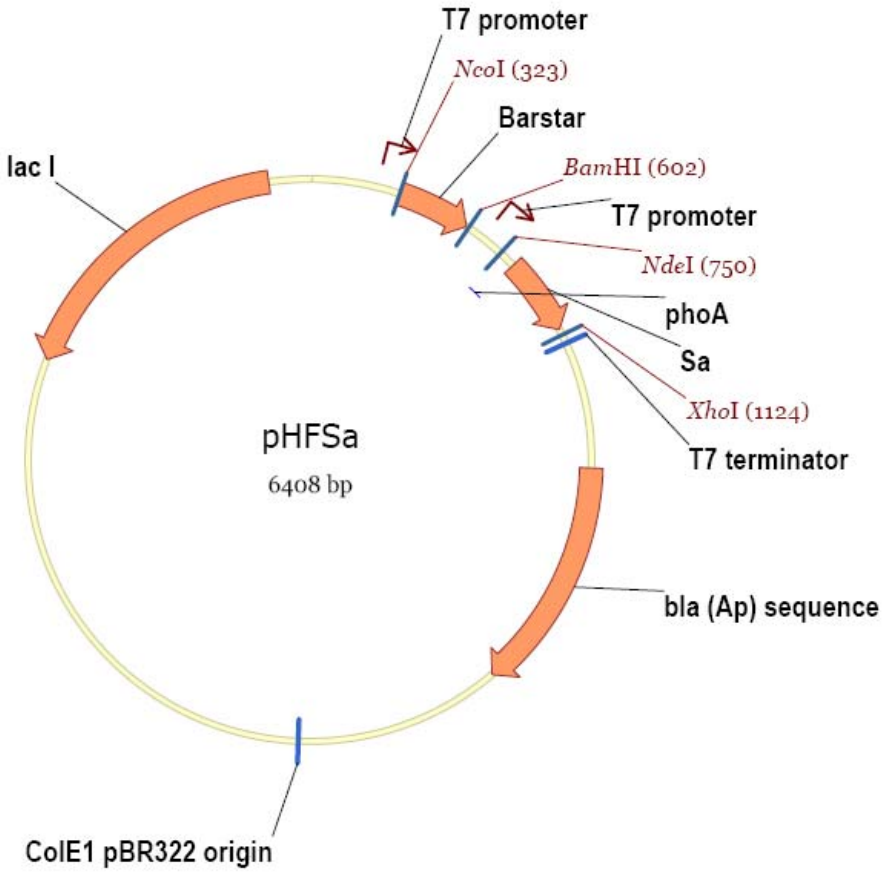

Fig. 6. Plasmid map for RNase Sa expression vector pHFSa.

The final construct was sequenced and the identity of both insertions was confirmed. The plasmid map for pHFSa is shown in Fig. 6.

Site-directed mutagenesis

Site-directed mutagenesis was done using the QuikChange™ kit from Stratagene. This kit uses the single strand amplification technique to generate a nicked double strand mutant plasmid DNA that can be taken up and repaired by bacterial *in vivo*. For single site mutation, two complementary DNA primers incorporating the desired mutated bases were designed according to general primer design guidelines using the Invitrogen Vector NTI software. The QuikChange reactions were setup according to the manufacturer's suggestions for the concentrations of the various reactants.

The amplification was verified on a 0.8% agrose gel. The product was digested with the restriction enzyme DPN1 that specifically cleaves methylated DNA. Since the reaction product does not contain any methylated DNA, the enzyme essentially removes the template plasmid DNA. This digestion product was applied to a regular transformation into chemically competent XL-1 blue *E. coli* cells and plated onto LB agar plates with appropriate antibiotics. The plates were incubated overnight; colonies were grown in ZYM505 media with appropriate antibiotics overnight before extraction of the plasmids. The plasmid DNA was sequenced to confirm the presence of the desired mutations.

Deletion of the original VlsE clone leading sequence

The original expression clone VlsE3 was provided by the Wittung-Stafshede lab.[105] This clone was derived from the T7 expression vector pCRT7NT (Invitrogen). Protein expressed from this clone has a His-tag and another leading sequence N-terminal to the VlsE sequence. The purification procedure was troublesome and not enough protein could be obtained. The leading sequence also introduced extra undesired sequence and the use of His-tag for purification prevented the using of EDTA. Therefore, this leading sequence was removed to keep only the original residues of VlsE. In this case, 102 bases were removed from the plasmid DNA. The QuikChange™ method discussed above is not efficient for this type of application. Instead, a SLIM method developed by Chiu et al[157] was successfully applied here to remove the leading sequence in one PCR reaction using four primers (Table 3).

Table 3. Primers used for deletion of VlsE leading sequence

VdFs (sense primer)
AAAAGCCAAGTTGCTGATAAGGACGAC
VdFt (sense primer)
GAGATATACATATG AAAAGCCAAGTTGCTGA
VdRt (anti-sense primer)
CATATGTATATCTC CTTCTTAAAGTTAAACAAAATTATTTC
VdRs (anti-sense primer)
CTTCTTAAAGTTAAACAAAATTATTTCTAGAGGG

The SLIM method is illustrated in Fig. 7. This approach uses the concept of linear plasmid DNA recirculization by annealing overhangs to generate nicked mutant double strand plasmid DNA. This method can be used to delete, insert or substitute large fragment of DNA bases from or into any position of a plasmid DNA.

The plasmid DNA was divided into three regions: the deletion region (grey), the overhang region (black ripple) and the body (white). The overhang region could be at either 5' or 3' of the deletion region. Primer VdFt and VdFs were sense primers. VdFs covered the 5 prime of the body while VDFt covered the overhang region and the 5' of the body. VdRt and VdRs were anti sense primers. Similarly, VdRs covered the 3 prime of the body and VdRt covered the overhang and the 3' of the body. Finally, VdFt and VdRt primers had overhangs that could anneal to each other.

The four primers and the plasmid were used in the usual QickChange^{TM} reaction except the concentration of each primer was reduced to half and the number of cycles was 30 instead of 18. There were four blunt end double strand products at the end of PCR reaction and after digestion with DPNI. The recombination of the 8 single strands after denaturation would potentially generate two hybrids that had both ends sticky and complementary. These sticky ends could anneal to each other when incubated at a low temperature for extended time. This led to two types of nicked circular DNA with identical sequence. These nicked plasmids were then inserted into the *E. coli* XL-1 blue cells by regular heat shock transformation and repaired *in vivo*. Positive clones from the selective plates were identified using colony PCR. Original VlsE3 plasmid was used

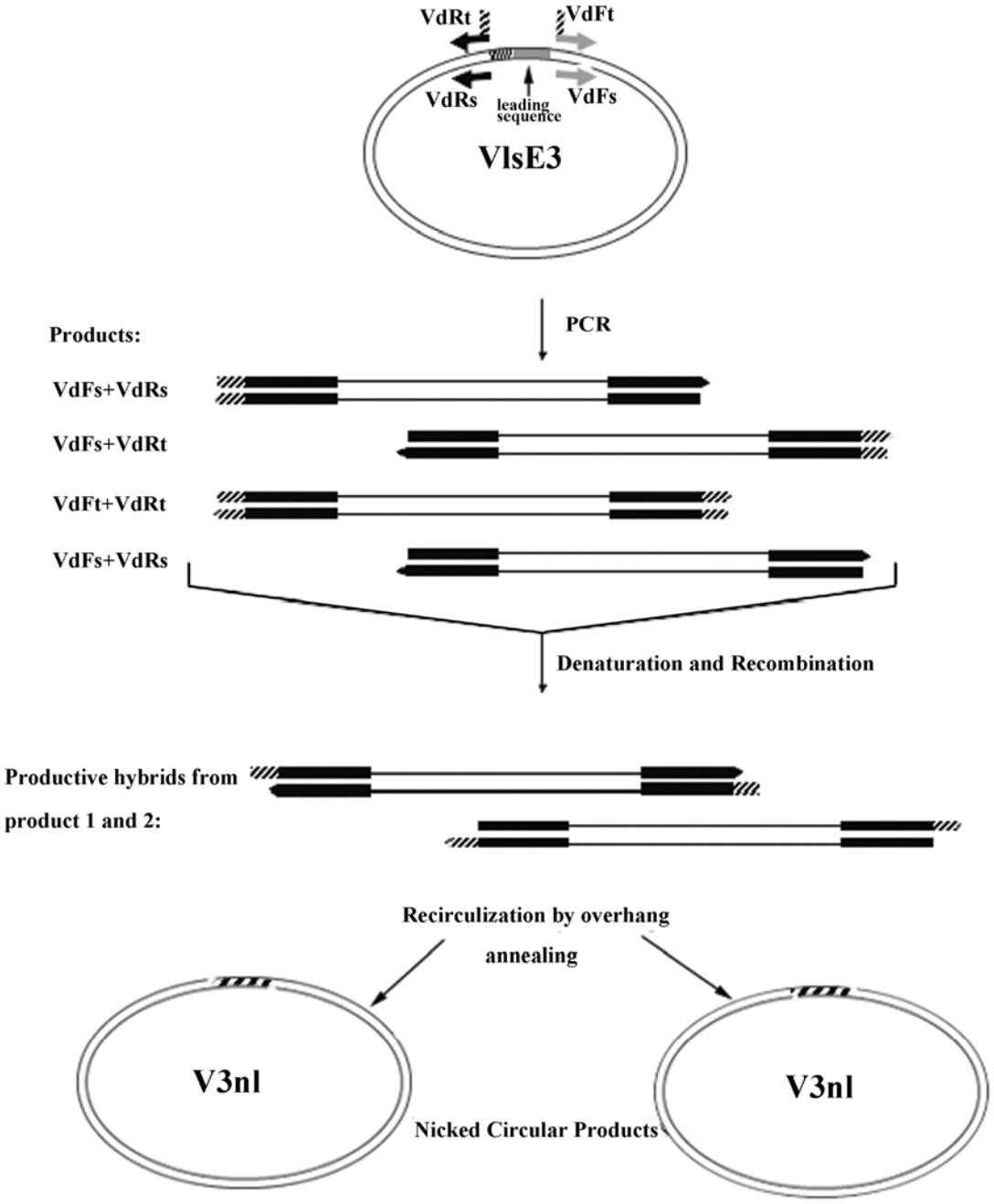

Fig. 7. A diagrammatic representation of the SLIM method. The figure illustrated the deletion of the VlsE leading sequence by one PCR reaction.

as a negative control. T7 promoter primer and T7 terminator primer were used for amplification. The plasmid from positive clones were then extracted and sequenced to confirm the deletion. The whole coding region including the T7 promoter and terminator regions were also checked for unintended mutations. This new clone was denoted as V3nl and had better expression and purification performance.

Protein over-expression and purification

Expression and purification of the variants of RNase T1, Sa2, Sa3, Cold shock protein, Tryptophan Synthetase alpha subunit, RNase HI and Maltose binding protein were described previously, [75; 150; 158-162] with only minor modifications.

Expression and purification of RNase Sa

The expression and purification of RNase Sa was following the procedure described previously with slight modifications.[150] All RNase Sa variants were expressed according to the Studier auto-induction procedure [153]. Briefly, C41 *E. coli* cells were transformed with pHFSa and plated on LB agar plate containing 50 µg/ml ampicillin. A single colony was grown in 2 ml LB media till saturation. The larger culture (2 L) of ZYM 5052 medium was inoculated from this culture and grown over night at 30°C. The purification procedure was the same as previously described.[150]

Expression and purification of VHP

Chemical competent *E. coli* BL21(DE3) cells were transformed with the pHFVHP plasmid. A single colony was picked from a selective agar plate to start a 2 ml LB culture containing 50 µg/ml of ampicillin. This 2 ml culture was grown utill

saturation and used to inoculate 2 L of ZYM5052 media containing 50 mg/L ampicillin. This large culture was grown overnight with vigorous shaking at 30ºC. Cells were harvested next morning by centrifugation at 4 ºC. The cell pellets were resuspended in ≈ 20 ml of TE buffer (15 mM Tris/3mM EDTA pH7.4) and passed through a French press twice at 12000 psi. The cell lysate was centrifuged down in a Beckman SS34 rotor at 16000 rpm, 4^{o}C for 50 minutes; the supernatant was collected and the volume was adjusted to 10% of the initial culture volume with TE buffer. Ammonia sulfate was added slowly into this suspension to 35% saturation on ice. This suspension was centrifuged at 12000 rpm in a SS34 rotor for 30 minutes and the pellet was discarded. The supernatant was then dialyzed against 4 litter NE buffer (50 mM NaOAc, 2 mM EDTA, pH 4.5) with one buffer change. The dialyzed suspension was centrifuged down at 10000 rpm in a GSA rotor for 20 minutes and the supernatant was loaded onto a 30 ml Sephadex C25 column equilibrated with NE buffer. After washing with ten column volume NE buffer, elution was performed with a 0 – 0.6 M NaCl gradient in NE buffer. The eluted fractions were collected and analyzed with SDS PAGE. Fractions containing the protein were pooled and freeze dried. Freeze dried protein was redissolved in minimal volume of ABC buffer (Ammonia bicarbonate, 50 mM, pH 8) and loaded on a 2 L G-50 size exclusion column equilibrated with ABC buffer. Fractions containing VHP were pooled after analysis with SDS PAGE and freeze dried. Dried protein powder was then stored in -20^{o}C freezer for future use. Typical yield for wild-type VHP was 30 mg per liter of culture. The purity and identity of the protein was confirmed by SDS PAGE and MALDI TOF mass spectrometry.

Expression and purification of VlsE

Plasmid V3nl was transformed into *E. coli* C41(DE3) cells using a heat shock procedure and plated onto LB agar plate containing 50 µg/ml ampicillin. The procedure used for VHP was then followed until the ammonia sulfate precipitation step. At that point, the volume of cell lysate was adjusted to 75 ml and ammonia sulfate was added slowly on ice until 90% saturation. This suspension was held on ice for 30 minutes and centrifuged with SS34 rotor at 12000 rpm for 30 minutes. The supernatant was discarded and the pellet was redissolved in a minimal volume of TE buffer and dialyzed against 2 L of the same buffer with two buffer changes. The dialyzed solution was then passed through a 10 ml Q-sepharose column equilibrated with TE buffer. The flow through and 2 column volume washes were pooled. The volume of this solution was adjusted to 200 ml by adding NE5 buffer (50 mM NaOAc / 2 mM EDTA / pH 5) and pH was adjusted to 5. The suspension was then centrifuged in a GSA rotor at 10000 rpm for 20 minutes. The supernatant was loaded onto a 100 ml Sephadex C-25 column equilibrated with NE5 buffer. Protein was eluted by a 0 – 0.6 salt gradient in NE5 buffer after washing with10 column volumes. The fractions containing VlsE were pooled and diluted 10 time with PE buffer (20 mM phosphate / 2 mM EDTA /pH 11.3) before loading onto a 100 ml Sepharose Q column equilibrated with PE buffer. Protein was then eluted by 0 – 0.7 M salt gradient in PE buffer after 10 column volume wash. Fractions containing VlsE were collected, desalted, concentrated and frozen at -20^{o}C.

Circular dichroism measurements

Far-UV CD measurements for VHP and VlsE

Far-UV CD spectra were collected for VHP and VlsE variants on an Aviv 62DS spectropolarimeter (Aviv Instruments, Lakewood, NJ) equipped with a temperature controlled stirring unit. All scans were done at 4ºC in 50 mM phosphate (for VHP) or 5 mM phosphate (for VlsE), pH 7.0 in a 1 cm pathlength cuvette. Protein concentrations were about 7 μM for VHP variants and 1 μM for VlsE variants determined as described.[163]

Thermal denaturation of proteins monitored by CD

Thermal denaturation of all protein variants were followed by circular dichroism spectrophotometry using an Aviv 62DS spectropolarimeter (Aviv Instruments, Lakewood, NJ) equipped with a temperature control, stirring unit. The general approach has been discussed previously.[164] The unfolding was followed by CD 234 nm for RNase Sa, Sa2, Sa3, 240 nm for RNase T1 and 222 nm for all other proteins. The denaturation for CspB-Bs, RNase Sa2, RNase T1, TSα, MBP variants was followed at pH 7.0 as described. [69; 158; 160; 161; 165] For RNase Sa3 we used 30 mM NaOAc (pH 5.0) and for RNase HI we used 30 mM NaOAc with 1 M GuHCl (pH 5.5).[162] For RNase Sa, the buffers were 30 mM Glycine (pH 1.64 – 3.25, 9 – 10), 30 mM NaOAc (pH 3.5 – 5.5), 30 mM MES (pH 5.5 –6.5), 30 mM MOPS (pH 7 –7.5) and 30 mM Tris (pH 8 – 8.5) and for VHP the buffer used was 50 mM Sodium Phosphate (pH 7.0). CD measurements were made at a 1 degree interval over temperature that ranged from 2 to 99^{o}C, at a

heating rate of 5-10^{o}C per hour with a 3 minute equilibration time between each measurement, a 1 nm bandwidth and a averaging time of 30 sec per measurement.

Thermal denaturation data were plotted as CD signal as a function of temperature. An example of RNase Sa unfolding curve is shown in Fig. 8.

Analysis of the denaturation curves was performed using the two-state model and the following equation[164]

$$y = \frac{(y_f + m_f \cdot T) + (y_u + m_u \cdot T) \cdot \exp\left(- \Delta H_m \cdot \frac{\frac{1}{T + 273.15} - \frac{1}{T_m + 273.15}}{R} \right)}{1 + \exp\left(- \Delta H_m \cdot \frac{\frac{1}{T + 273.15} - \frac{1}{T_m + 273.15}}{R} \right)} \quad (2.1)$$

where m_f is slope of the pre-transition baseline; y_f is y-intercept of the pretransition baseline; m_u is slope of the post-transition baseline; y_u is y-intercept of the post-transition baseline. T is the temperature at each point of measurement, T_m is the midpoint of the thermally induced denaturation of the protein. ΔH_m is the van't Hoff enthalpy of unfolding at the T_m.

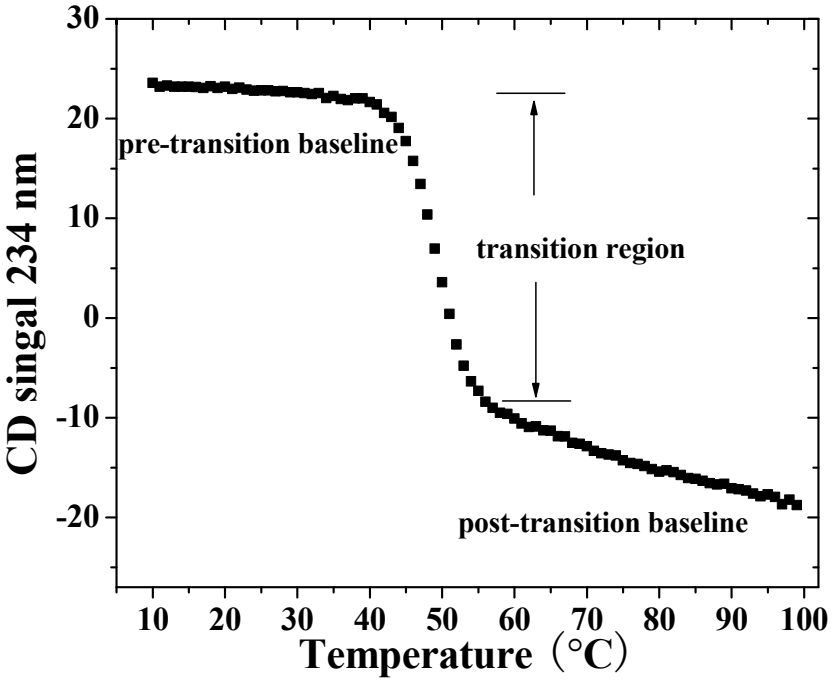

Fig. 8. Thermal denaturation curve for wild-type RNase Sa. Denaturation was monitored by circular dichroism at 234nm. The transition region of unfolding, together with the pre- and post- transition baselines are shown. The protein concentration was ~ 0.1 mg/ml in 30 mM MOPs, pH 7 in a 1cm cuvette.

Chemical denaturation of proteins monitored by CD

Chemical denaturation curves were determined using circular dichroism measurements at 222 nm for VHP and ubiquitin, 220 nm for VlsE and 234 nm for RNase Sa to follow unfolding. The general approach has been previously described.[164] The instrument was either an Aviv 62DS or 202SF spectropolarimeter (Aviv Instruments, Lakewood, NJ) equipped with a temperature control, stirring unit and automatic titrator. We typically used quartz cuvettes with a 1 cm path length. Protein working concentration was 7 μM for VHP, 2 μM for ubiquitin, 1 μM for VlsE and 10 μM for RNase Sa. Protein concentration was determined by measuring UV absorbance at 280 nm for VHP, RNase Sa and 205 nm for VlsE as previously described. [163]

Denaturant unfolding data were plotted as CD signal as a function of denaturant concentration. An example of RNase Sa unfolding curve is shown in Fig. 9.

Analysis of the denaturation curves was performed using the two-state model and the linear extrapolation method[166] using the following equation[164]

$$y = \frac{(y_f + m_f \cdot [D]) + (y_u + m_u \cdot [D]) \cdot \exp(m \cdot (\frac{[D] - D_{1/2}}{RT}))}{1 + \exp(m \cdot (\frac{[D] - D_{1/2}}{RT}))} \quad (2.2)$$

where m_f is slope of the pre-transition baseline; y_f is y-intercept of the pretransition baseline; m_u is slope of the post-transition baseline; y_u is y-intercept of the post-transition baseline; *m* is the dependence of ΔG on denaturant; $D_{1/2}$ is midpoint of the unfolding transition curve; and D is concentration of denaturant.

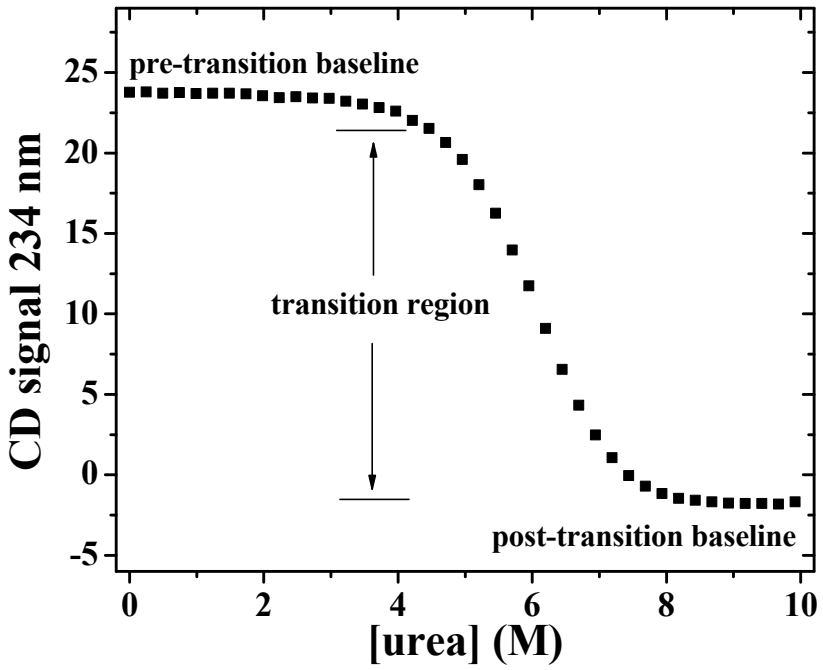

Fig. 9. Urea denaturation curve for wild-type RNase Sa. The transition region of unfolding, together with the pre- and post- transition baselines are shown. Experiments was done using a titrator to increase urea concentration while keeping the protein concentration constant. Protein concentration was ~ 10 μM in 30 mM MOPS, pH 7 in a 1 cm cuvette. CD signal was collected at 234 nm. Equilibration time for each point was 3 minutes and averaging time was 30 seconds.

CHAPTER III

CONTRIBUTION OF THE HYDROPHOBIC EFFECT TO THE CONFORMATIONAL STABILITY OF A SMALL PROTEIN AND A LARGE PROTEIN

Introduction

Globular proteins are essential for the existence of all biological systems. The hydrophobic effect is one of the major forces that stabilize the folded state of proteins, through the tight packing of non-polar atoms in the protein interior.[17; 22; 26] This packing has been extensively studied using site-directed mutagenesis,[35; 37; 39; 47; 167] and it has been generally shown that the substitution of larger hydrophobic groups with smaller ones generally destabilizes the protein. Based on these studies, Pace proposed that the burial of one -CH_2- group contributes approximately 1.3 kcal/mol to protein stability.[44] However, most of the proteins that have been studied are 'medium" sized, typically 100-150 residues in length. It is not clear whether Pace's finding will apply to proteins of all sizes, especially very small or very large ones.

The C terminal subdomain of the villin "head piece" (VHP) is a 36-residue protein with a high melting temperature (74 ºC at pH 7).[98] Because it is one of the smallest globular protein known to date, VHP has been extensively studied both experimentally and computationally.[101; 102; 168; 169] The conformational stability of VHP is

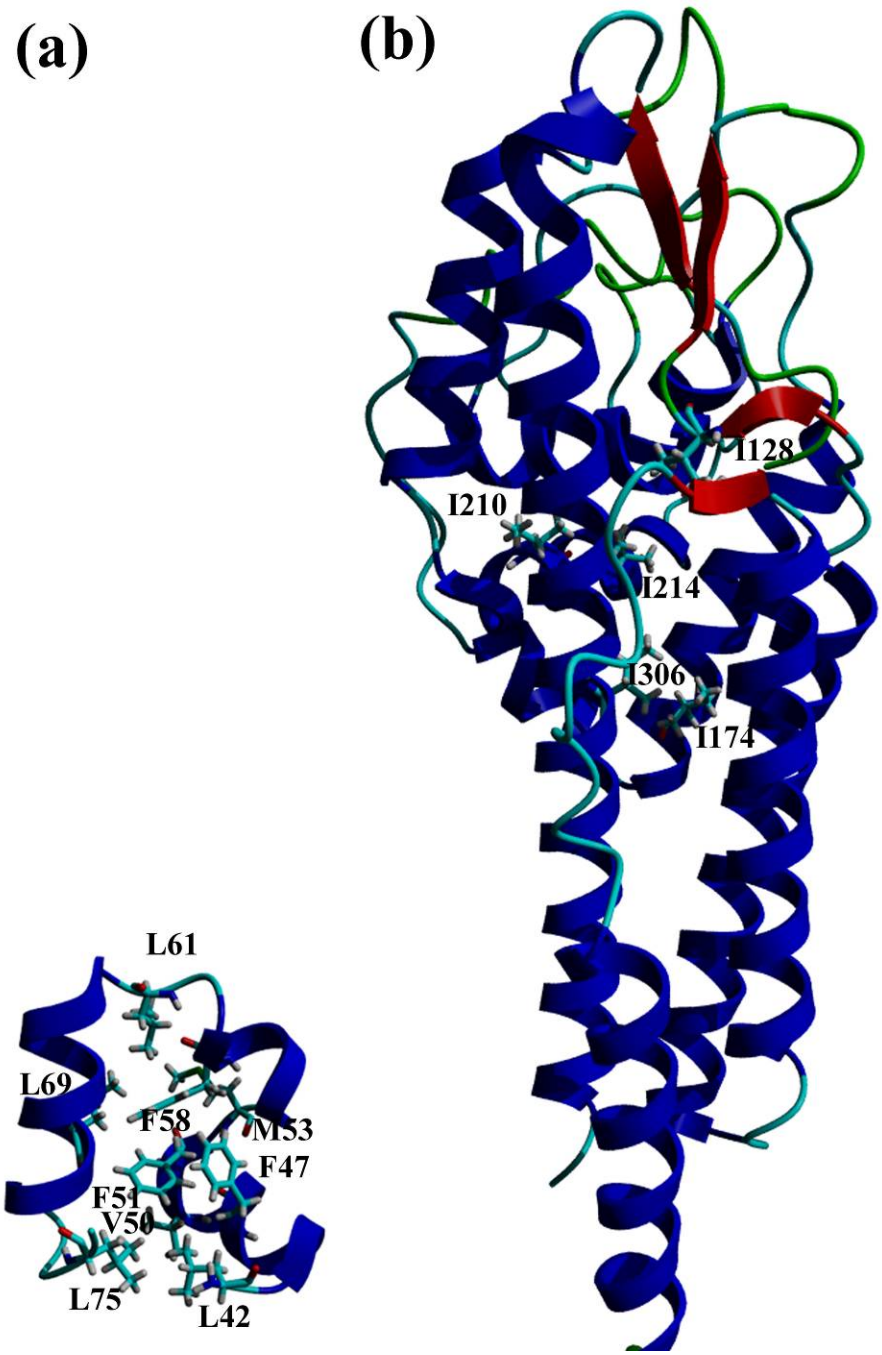

Fig. 10. Ribbon diagram of VHP and VlsE structures. The residues mutated are shown as sticks. The figure was generated using the YASARA program[106] and the X-ray structures for (a) VHP (PDB 1YRI) [16] and (b) VlsE (PDB 1L8W)[104].

~ 3 kcal/mol at pH 7, 25 ºC.[98] Both NMR and crystal structures of VHP have been determined. [99; 170] The structure of VHP consists of a three helix bundle held together by a core containing nine hydrophobic residues (Fig. 10a). A study of the 3 buried Phe residues suggests that they make a significant contribution to the stability of VHP.[100] However, this previous study employed Phe to Leu mutations, which is a non-standard mutation. In order to characterize the contribution of hydrophobic effects of these Phe residues, we made the Phe to Ala mutations in this study.

The *Borrelia burgdorferi* outer surface lipoprotein Variable major protein-Like Sequence, Expressed (VlsE), on the other hand, is a 337 residue protein and considered rather large protein for the protein folding study.[105] To date, VlsE is the largest single domain protein that folds by a two-state process.[103] At pH 7, the conformational stability of VlsE is only ~ 5 kcal/mol.[105] The crystal structure of VlsE shows its unique football shape, consisting of 11 helices and 4 short strands (Fig. 10b).[104] Interestingly, it has been recently suggested that a crowding agent may induce a shape change of VlsE to a more globular conformation.[171] The thermodynamics of the folding of VlsE has not been well studied.

VHP and VlsE have two interesting physico-chemical characteristics: (1) the stabilities of the two proteins are similar, despite huge differences in their size and folding rate and (2) the percentage of non-polar side-chain group burial for these two proteins is quite different, with 84% of non-polar surface buried in VlsE, and only 53% buried in VHP. Therefore, it may be possible that hydrophobic effect contributes to a different extent to protein stability in these two proteins.

Table 4. Structural characteristics of the mutated residues

residue	secondary structure	van der Waals contacts (<4Å)	% Side chain buried[a]
		VHP	
Val 50	helix	L42/F51/K73	66
Leu 42	loop	V50/L75	76
Leu 61	loop	F58/P62/L69	70
Leu 69	helix	F58/L61/K73	75
Leu 75	helix	L42/F51/K73	72
Met 53	loop	A57/F58	81
Phe 47	helix	F51/R55/F58	98
Phe 51	helix	F47/V50/F58/K70/K73/L75	100
Phe 58	helix	F47/F51/M53/F58/L61/Q66/L69	98
		VlsE	
Ile 128	loop	P45/V50/A272/V276	100
Ile 154	helix	L163/L212/V216/L307	100
Ile 210	other	V205/A260/V299	100
Ile 214	helix	I151/V261	100
Ile 306	helix	Y73/L233/I308	100

[a] Calculated using the in-house program *pfis*[75], based on the Lee and Richard method[172].

In this study, we focused on the hydrophobic residues buried in VHP and VlsE. To investigate the contribution of hydrophobic residues to the stability of these two proteins, 9 VHP variants and 5 VlsE variants were designed. The 9 VHP variants consist of single mutations to Ala: V50A, L42A, L61A, L69A, L75A, M53A, F47A, F51A, and F58A. The 5 VlsE variants consist of single mutations to Val: I128V, I154V, I210V, I214V,and I306V. The mutated residues of both proteins contain side chains that are mostly buried, as shown in Table 4 and Fig. 10. We have determined denaturation curves for these variants, using circular dichroism to follow unfolding. The results give additional insight into the general contribution of hydrophobic effects to protein stability.

Results

Efficient expression and purification of VHP variants

Efficient expression and purification of small proteins like VHP is often difficult because they are marginally stable and susceptible to degradation *in vivo*. Many approaches have been tried to produce VHP in large quantities for use in structural and biophysical studies. The VHP variants have been produced by: (1) intact gene expression in *E. coli*, which produces only small amounts of protein;[98] (2) fusion protein expression in *E. coli* [100; 173] which requires an additional cleavage step and (3) direct chemical synthesis, which is labor intensive.[99; 168] All the above methods have their disadvantages, so a new expression and purification system for VHP was developed.

Inspection of the original VHP gene revealed several rare codons, which may lead to the low expression level of the original expression system, pVHP42-76b, developed by McKnight et al.[98] Therefore, a redesigned VHP gene was constructed using standard molecular biology methods and cloned into a pET22B (+) vector. The primers used are shown in Table 1. They span the redesigned VHP gene sequence from the back-translation of VHP, and also provide restriction enzyme sites for cloning. The resulting expression system was used with the auto-induction method developed by Studier[153]. The protein expressed with a final yield of about 30 mg/L of culture. All VHP variants in this study were successfully expressed using this expression system.

CD spectra of wild-type proteins and variants

Both VHP and VlsE have a large fraction of α-helix content, which can be easily detected by far-UV CD spectroscopy. Therefore, far-UV CD spectroscopy was used to study the secondary structure features of wild-type VHP and VlsE as well as their variants. The VHP variants that have residues changed to Ala, as well as the Ile-Val variants of VlsE, have CD spectra that are very similar to that of the wild-type proteins, suggesting that single mutations here do not affect the secondary structure significantly (Fig. 11a, 11c). For the aromatic variants of VHP (Phe to Ala) the intensity of one variant (F58A) is slightly reduced compared to that of the wild-type VHP (Fig. 11b). This may be because there are more unfolded forms of this variant under these conditions or it may reflect that Phe 58 residue may contributed to the far-UV CD spectra. Overall, the scans suggest that the native structure of all variants are not significantly different for wild-type when the mutations are introduced.

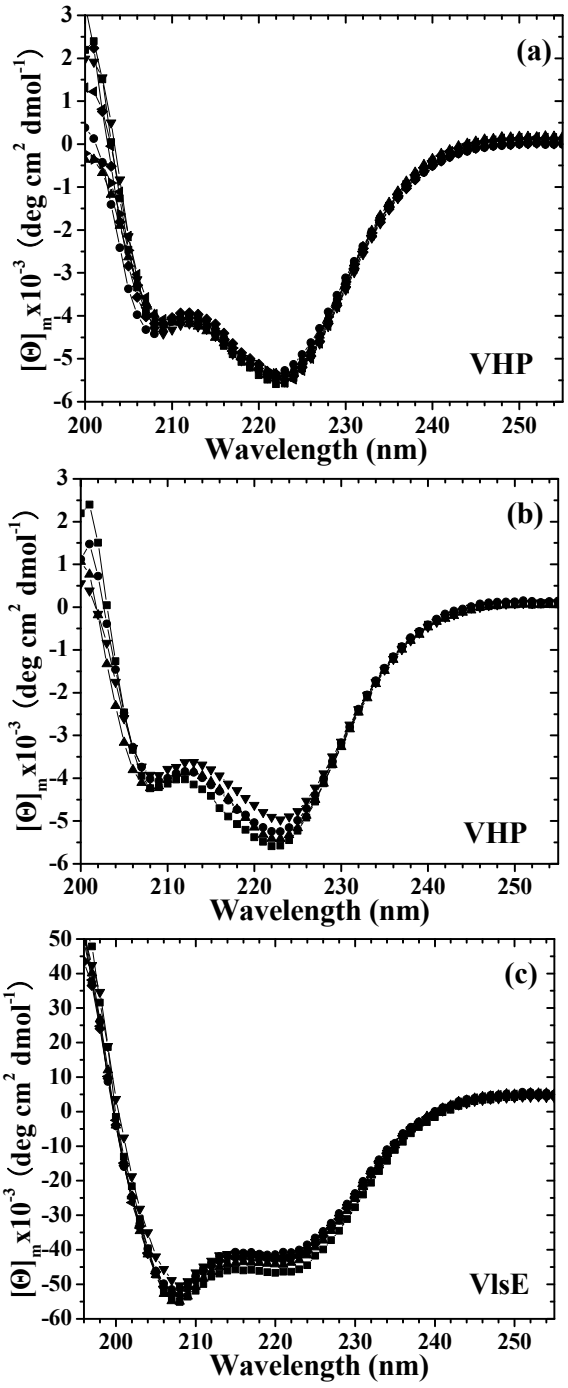

Fig. 11. Far-UV CD spectra of VHP and VlsE variants. (a) Wild-type VHP (■) and hydrophobic variants L42A (●), V50A (▲), L61A (▼), L69A (♦), L75A (◄) and M53A(►). (b) Wild-type VHP (■) and F47A (●), F51A (▲), F58A (▼). (c) Wild-type VlsE (■) and variants I128V (●), I154V (▲), I210V (▼), I214V (♦) and I306V (◄). Data were collected at 4°C in 50 mM or 5 mM sodium phosphate buffer (pH 7) for VHP and VlsE respectively.

Table 5. Parameters characterizing the thermal unfolding of the VHP variants

Variant	ΔH_m [a] (kcal/mol)	T_m [b] (ºC)	$\Delta\Delta T_m$ [c] (ºC)	$\Delta\Delta G^o$ [d] (kcal/mol)	$\Delta\Delta G^o$ [e] (kcal/mol)
VHP	31	74.4	-	-	-
V50A	28	67.0	-7.4	-0.6	-0.9
L42A	23	60.4	-14.0	-1.3	-1.7
L61A	22	45.3	-29.1	-1.8	-2.6
L69A	22	55.3	-19.1	-1.5	-1.9
L75A	29	67.2	-7.2	-0.4	-0.6
M53A	15	34.7	-39.7	-2.6	-3.2
F47A	18	41.7	-32.7	-2.2	-2.2
F51A	22	48.8	-25.6	-1.7	-1.7
F58A	16	28.3	-46.1	-2.9	-2.9

[a] ΔH_m, enthalpy of unfolding at T_m.The error is ±2 kcal/mol.
[b] $T_m = T$, where $\Delta G^o = 0$. The error is ±0.5 ºC.
[c] $\Delta T_m = T_m$ (variant) – T_m (VHP).
[d] $\Delta\Delta G^o = \Delta G^o$ (variant) - ΔG^o (VHP), calculated using Eq. 2.2. ΔG^o values were calculated at 25 ºC using the ΔC_p 0.374 kcal/mol/K determined for wild-type VHP in this study. Negative $\Delta\Delta G^o$ values here indicate decrease in stability. The mean error is ±0.1 kcal/mol.
[e] This column is the normalized $\Delta\Delta G$ for 100% side chain burial using the data in table 4 for each site.

Thermal denaturation of VHP variants

The stabilities of wild-type VHP and its variants were determined by thermal denaturation, using CD (222 nm) to follow unfolding. The denaturation curves were analyzed to determine the temperature at the midpoint of the denaturation curves, T_m and the enthalpy change at T_m, ΔH_m. All of the VHP variants were observed to unfold at lower temperatures than wild-type VHP indicating that they are all destabilized (Fig. 12 and Table 5). The T_m and ΔH_m values determined for wild-type VHP are in good agreement with previously reported values.[98] The six hydrophobic variants (Val/Leu/Met to Ala) have T_ms ranging from 67.2 to 34.7 ºC. These translate to a destabilization of -0.4 to -2.6 kcal/mol relative to wild-type VHP. For the three Phe to Ala variants, the decreases in stabilities are even bigger (Table 5): the T_ms are from 48.8 ºC (F51A) to 41.7 ºC (F47A) to 28.3 ºC (F58A); the corresponding $\Delta\Delta G$ values are from -1.7 (F51A) to -2.2 (F47A) to -2.9 (F58A).

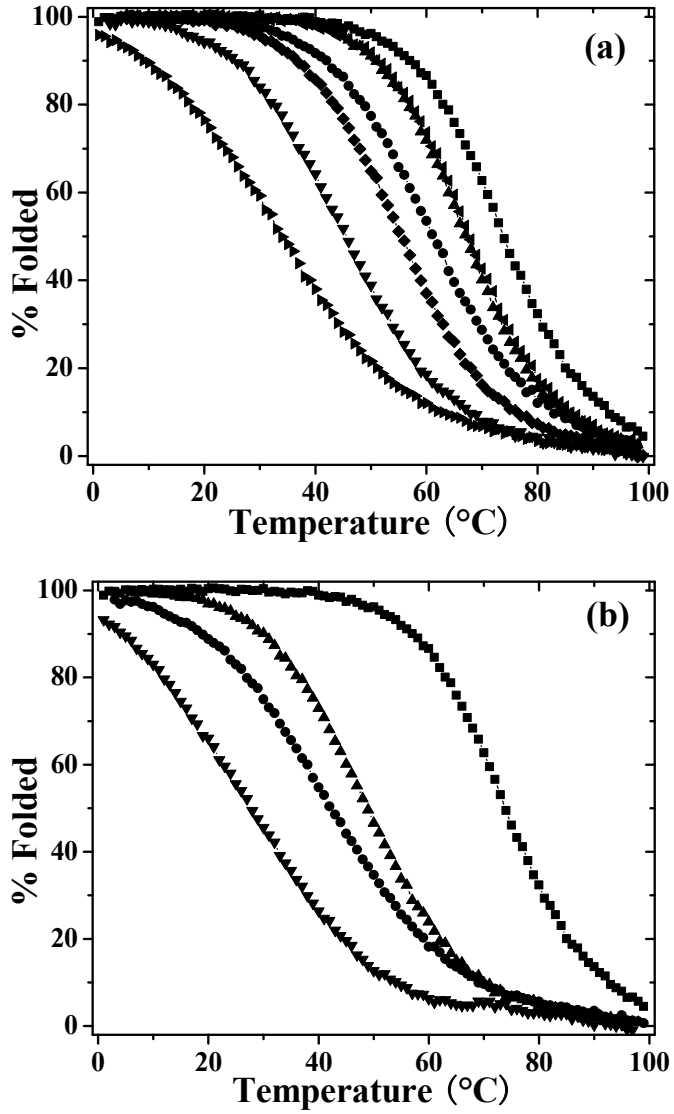

Fig. 12. Thermal unfolding curves for VHP and its variants. (a) Wild-type VHP (■) and hydrophobic variants L42A (●), V50A (▲), L61A (▼), L69A (♦), L75A (◄) and M53A(►). (b) Wild-type VHP (■) and F47A (●), F51A (▲), F58A (▼). Experiments were done in 50 mM sodium phosphate (pH 7).

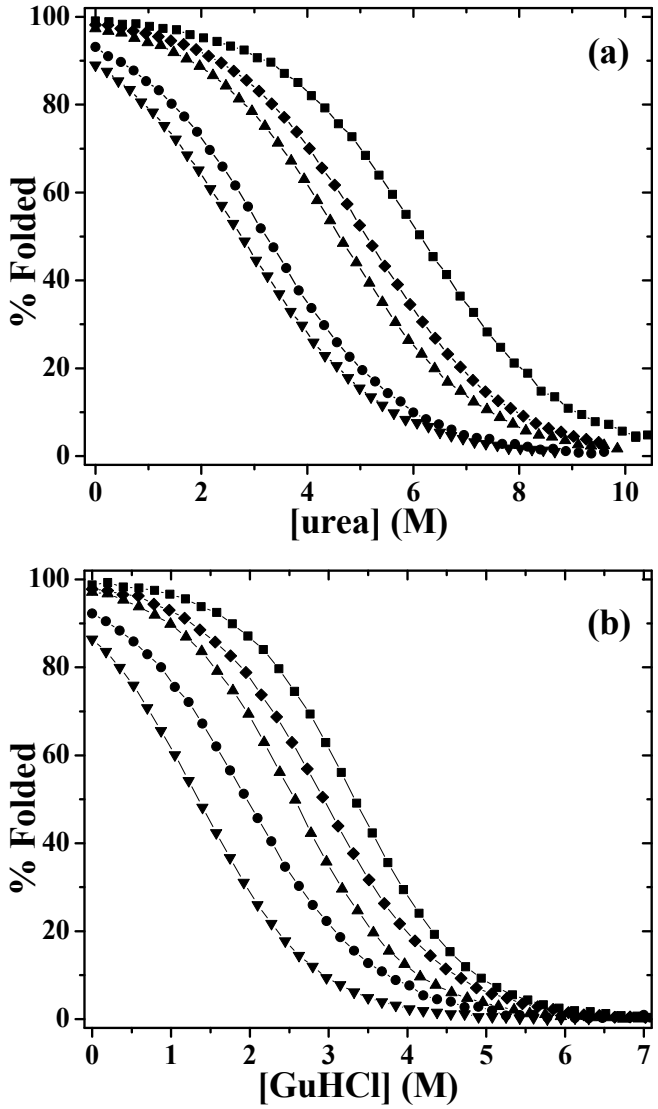

Fig. 13. Chemical denaturation curves for VHP and its variants. (a) Urea and (b) GuHCl unfolding curves for wild-type VHP (■) and variants L42A (●), V50A (▲),L69A (▼), L75A (♦).

Chemical denaturation of VHP and VlsE variants

The effect of mutations on the conformational stability of VHP was also determined using urea and GuHCl denaturation. Fig. 12 shows that all nine VHP variants are less stable than wild-type VHP, since they unfold at lower urea and GuHCl concentrations. However, only 4 of the 9 variants were stable enough to give reliable pre-transition base line for both denaturants. They are L42A, V50A, L69A and L75A (Fig. 13). The denaturation curves were analyzed assuming a two-state folding mechanism, using the linear extrapolation method (LEM)[166] to obtain the thermodynamic parameters for both urea and GuHCl unfolding (Table 6). For the four variants, the destabilization ranges from -0.3 to -1.6 kcal/mol (GuHCl) and -0.5 to -1.6 kcal/mol (urea) experiments (Table 6). The results determined by the two denaturants are in good agreement with each other.

Table 6. Parameters characterizing chemical unfolding of VHP variants

Variant	GuHCl denaturation			Urea denaturation			
	m (kcal /mol/M)	$D_{1/2}$ (M)	$\Delta\Delta G(H_2O)$[a] (kcal/mol)	*m* (kcal /mol/M)	$D_{1/2}$ (M)	$\Delta\Delta G(H_2O)$[a] (kcal/mol)	$\Delta\Delta G(H_2O)$[b] (kcal/mol)
VHP	0.84±0.01	3.33±0.04	-	0.45±0.02	6.17±0.04	-	
V50A	0.82±0.01	2.56±0.03	-0.6±0.1	0.46±0.01	4.61±0.02	-0.7±0.1	-1.1
L42A	0.74±0.03	1.95±0.03	-1.2±0.1	0.47±0.01	3.22±0.06	-1.4±0.1	-1.8
L69A	0.82±0.01	1.35±0.04	-1.6±0.2	0.45±0.01	2.76±0.07	-1.6±0.1	-2.1
L75A	0.78±0.01	2.94±0.03	-0.3±0.1	0.47±0.01	5.19±0.06	-0.5±0.1	-0.7

[a] $\Delta\Delta G(H_2O)=m_{av}[D_{1/2}(\text{variant})-D_{1/2}(\text{VHP})]$. m_{av} is the average *m* value for all variants. m_{av}(GuHCl) = 0.80±0.04 kcal/mol/M, m_{av}(urea) = 0.46±0.01 kcal/mol/M. Errors are propagated from the error of *m* and $D_{1/2}$.

[b] This column is urea $\Delta\Delta G(H_2O)$ normalized to 100% side-chain burial.

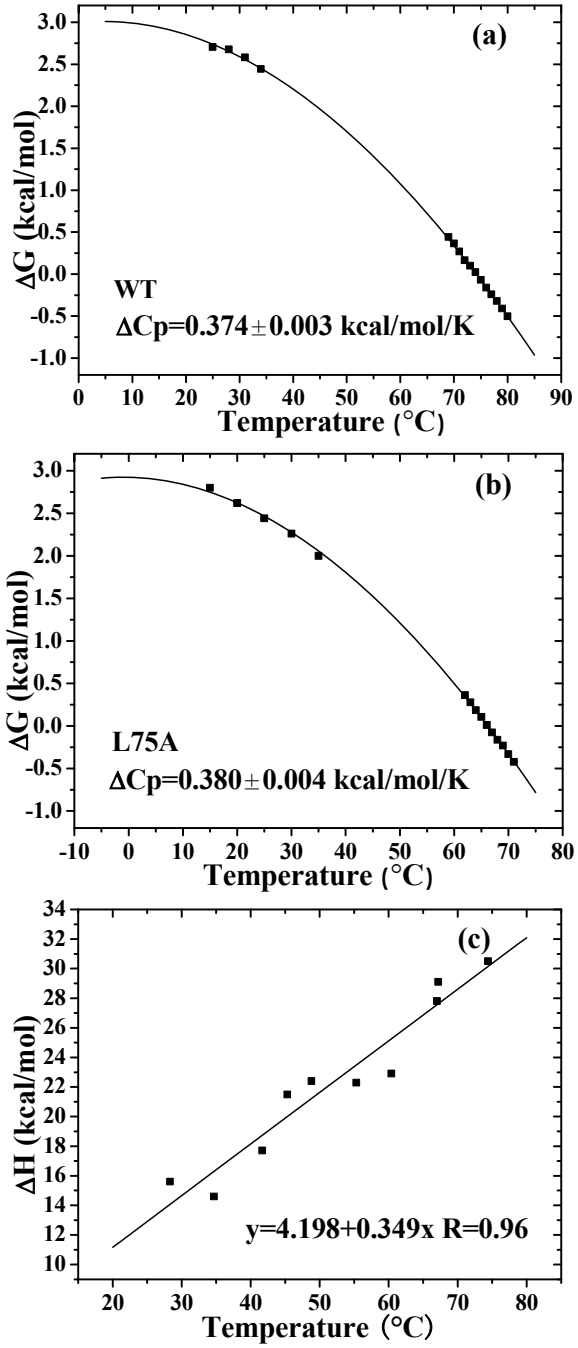

Fig. 14. Estimation for ΔC_p of unfolding of VHP. (a) Global fit for wild-type VHP. (b) Global fit for L75A VHP. (c) Kirchoff analysis for all the VHP variants.

The conformational stabilities of wild-type VHP at temperatures from 25 ºC to 35 ºC were also determined by urea denaturation. A global fit of these data using the modified Gibbs-Helmholtz equation,

$$\Delta G\,(T) = \Delta H_m\,(1\text{-}T/T_m) + \Delta C_p\,[(T\text{-}T_m) - T\ln(T/T_m)] \qquad (3.1)$$

gave a change of heat capacity of unfolding, ΔC_p, of 0.374±0.003 kcal/mol/K for wild-type VHP (Fig. 14a). This value was used to calculate the ΔΔG values in Table 3, which vary from -0.4 to -2.9 kcal/mol. They are in good agreement with the results from the chemical denaturation experiments described above. When the values are normalized for their side chain percentage burial, to give the ΔΔG values that would be expected if the side chains were fully buried, they range are from -0.6 to -3.2 kcal/mol.

Thermal denaturation of VlsE is not reversible at pH 7.0; therefore, the stabilities of VlsE variants were determined by urea denaturation (here the unfolding is completely reversible). The urea denaturation curves for wild-type VlsE and its variants were analyzed assuming a two-state folding mechanism. All five variants unfold at a lower urea concentration than that of wild-type VlsE (Fig. 15), indicating that all mutations are destabilizing. The urea denaturation curves for VlsE were analyzed using the LEM[166] method to determine the thermodynamic parameters $D_{1/2}$ (urea concentration at the midpoint of transition) and *m* (the slope of the denaturation transition curve) (Table 7). All variants have $D_{1/2}$ less than 1 M, yet they all have similar *m* values. The $\Delta\Delta G(H_2O)$ values were calculated by using the average *m* value. All variants are destabilized by 1.1 to 1.9 kcal/mol, relative to wild-type VlsE.

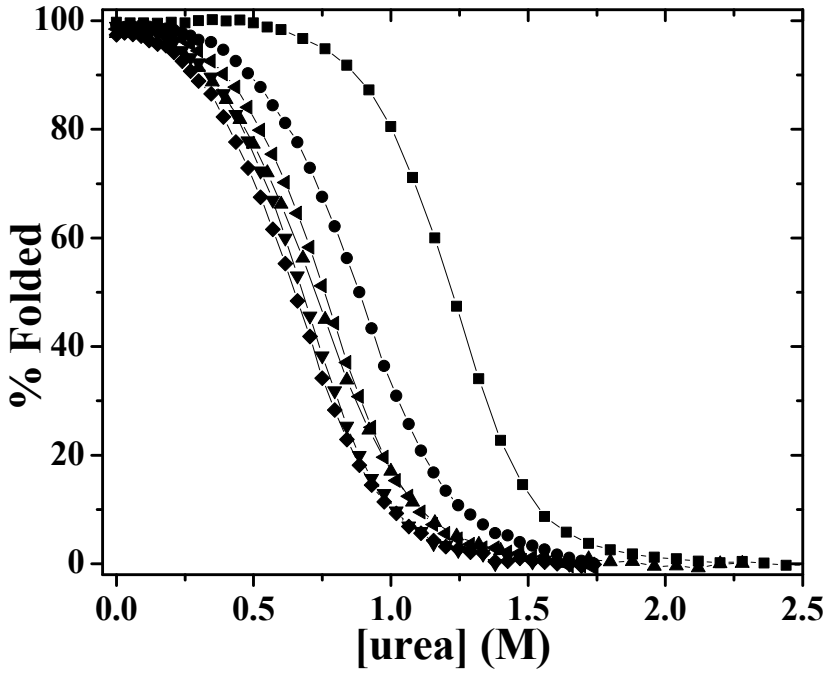

Fig. 15. Urea unfolding curves for VlsE and its variants. Shown in figure are wild-type VlsE (■), I128V (●), I154V (▲), I210V (▼), I214V (♦) and I306V (◄).

Table 7. Parameters characterizing urea unfolding of VlsE variants

Variant	*m* (kcal /mol/M)	$D_{1/2}$(M)	$\Delta\Delta G(H_2O)$[a] kcal/mol
VlsE	3.86±0.20	1.19±0.05	-
I128V	3.36±0.04	0.88±0.03	-1.1±0.2
I154V	3.35±0.05	0.70±0.02	-1.7±0.2
I210V	3.77±0.13	0.67±0.02	-1.8±0.2
I214V	3.48±0.21	0.65±0.01	-1.9±0.2
I306V	3.70±0.02	0.77±0.03	-1.5±0.2
Average	3.59±0.22	-	-1.6 ±0.3

[a] $\Delta\Delta G(H_2O) = m_{av}[D_{1/2}(\text{variant}) - D_{1/2}(\text{WT})]$. m_{av} is the average m value for all variants, =3.59 kcal/mol/M.

Discussion

ΔC_p of unfolding for VHP

In this study, it is essential to determine the ΔC_p value for VHP variants for two reasons: (1) the contribution of mutations to VHP stability determined by different denaturation methods can be directly compared by the $\Delta\Delta G$ values, and (2) some of the variants are not stable enough to characterize using chemical denaturation. Thus, a reliable ΔC_p value is required for calculating the ΔG values using Eq. 3.1. The traditional method for determining ΔC_p is to use differential scanning calorimetry (DSC) with the Kirchoff analysis,

$$d(\Delta H_m)/d(T_m) = \Delta C_p \qquad (3.2)$$

where a variation of pH in the acidic region is used to obtain the ΔH and T_m.[92] In most cases, ΔH is found to vary linearly with T_m. For VHP, however, it is difficult to apply this method because the VHP protein has small enthalpy of unfolding (ΔH_{cal}) with a very broad DSC transition. The small ΔH makes its measurement less reliable, while the broad transition results in incomplete pre or post-transition baselines, making the interpretation of DSC data extremely difficult.

Another approach that has been widely used was proposed by Pace and Laurents[174] in which the ΔG values determined at lower temperatures by chemical denaturation are constrained by the data at the thermal denaturation transition region, and globally fit with Eq. 3.1 to obtain the ΔC_p. Therefore, a series of urea denaturation curves were determined for wild-type VHP at temperatures from 25 to 34 ºC. These

data, and the T_m and ΔH_m values determined by thermal denaturation, were fit to Eq. 3.1 to give a ΔC_p value of 0.374±0.003 kcal/mol/K.

It is well known that ΔC_p is correlated with the change in accessible surface area upon unfolding (ΔASA). Several equations have been developed to predict ΔC_p from ΔASA.[175-177] After careful study of 49 proteins, Myers et al.[177] proposed the following equation:

$$\Delta C_p \text{ (in cal/mol/K)} = -119 + 0.20*(\Delta ASA) \quad (3.3)$$

Using Eq. 3.3, and the ΔASA for VHP (2465Å, calculated using the program *pfis*[75] and PDB file 1YRI[16]), the calculated ΔC_p for wild-type VHP is 374 cal/mol/K which agrees exactly with the experimental determined value. This agreement suggests that VHP is not unique in its thermodynamic behavior. However, one wonders if this ΔC_p can be safely used for all variants. Using Eq. 3.3, as well as equations from others, [175-177] a substitution of a smaller hydrophobic side chain may lead to a decrease in ΔC_p up to 35 cal/mol/K[178], because of the reduced non-polar surface area exposure upon unfolding. Therefore, the ΔC_p value for the L75A variant was also determined by global fit of the urea and thermal denaturation data (Fig. 14b). The resulting ΔC_p value (0.380±0.004 kcal/mol/K) differs little from that of the wild-type VHP. This agrees with what was observed with ubiquitin, where very small changes in ΔC_p were found for mutations from Leu/Val to Ala.[178] In addition, the effects of mutations on ΔC_p have also been studied in two cold shock proteins by Gribenko and Makhatadze.[179] Their results suggest that if variants of the same protein share a similar stabilization mechanism, they also share the similar enthalpy functions. The ΔH_m and T_m values determined for all the variants in this

study are plotted in Fig. 14c. The ΔH_m showed a good linear dependence on T_m, with a slope of 0.349±0.034 kcal/mol/K. This is within error of the ΔC_p values described above, indicating that the ΔC_p values for the different variants studied here are similar. The $\Delta\Delta G$ values for all variants are essentially the same when calculated with the different ΔC_p values (not shown). Therefore, the ΔC_p value of 0.374 kcal/mol/K determined for wild-type VHP was used for all variants.

Phe to Ala mutations significantly destabilize VHP

VHP has a hydrophobic core that contains three conserved Phe residues, Phe 47, Phe 51 and Phe 58 (Fig. 10). These Phe residues have been previously mutated to Leu, either individually or in combination.[100] Studies of those variants suggest that the Phe residues make a significant contribution to the stability of VHP. However, the mutation from Phe to Leu may introduce steric hindrance as well as a cavity[22], making the interpretation and quantification of the stabilizing effect difficult. Therefore, we changed Phe to Ala for the variants in this study. The ΔT_ms for the three Phe to Ala single variants are large, ranging from -25.6 ºC (F51A) to -46.1 ºC (F58A), following the same order of destabilization as previously reported.[100] Using Eq.3.1 with ΔC_p= 0.374 kcal/mol/K, gave ΔG(25 ºC) values of 1.3, 0.8 and 0.2 kcal/mol for F51A, F47A and F58A respectively. These values are in good agreement with the calculated ΔG(25 ºC) values from the transition region of thermal denaturation curves, suggesting that the ΔC_p used is reasonable for describing the folding of these variants. While all of the stabilities of these Phe to Ala variants are significantly lower than that of wild-type VHP, the degree of destabilization varies. The key residue that seems to define the VHP fold

appears to be Phe 58. The difference of destabilization observed for these 3 Phe to Ala variants may reflect the different environment of the mutation sites. Inspection of the crystal structure reveals that Phe 58 is the center of a larger hydrophobic cluster. Phe 58 has van der Waals interactions with five of the other eight hydrophobic residues studied here (Table 4). The other two Phe, however, have less van der Waals interactions. This implies that the Phe to Ala mutations are destabilizing mainly because of the loss of van der Waals interactions.

Mutations from larger hydrophobic residues to smaller ones have destabilizing effects on both VHP and VlsE

Mutations that change larger hydrophobic side chains in the protein interior to smaller ones are generally destabilizing, while in the reverse case, substitutions of larger hydrophobic groups that fill cavities tend to be stabilizing.[37; 44; 116] These types of mutations when applied to proteins of extreme size (very small or very large) may provide unique information on the origin and strength of the hydrophobic effects.

The three Phe residues in VHP make substantial contributions to its stability. Their stabilizing effects may also arise from π-π interactions between Phe residues[99] and a cation-π[180] interaction between Phe 47 and Arg 55. The six buried hydrophobic residues and the three Phe residues studied here define the hydrophobic region of this small protein. Six of these nine residues (Phe 47, Phe 51, Phe 58, Met 53, Leu 61 and Leu 69) surround Phe 58 and bridge helices 1, 2 and 3; the other three residues Leu 42, Leu 75 and Val 50 lie at the open end of the VHP molecule. These data suggest that the hydrophobic core consists of these six residues, with Phe 58 in the center. Presumably,

Phe 58 would be the most important residue in the core as far as stability is concerned, while Phe 47 and Phe 51 would play smaller. Indeed, the double variant F47L/F51L has been shown to retain the VHP fold.[100] In addition, mutations of these six "core" residues should be more deleterious than mutations of the other three. The results from thermal and chemical denaturation experiments are in good agreement (Table 5, 6), and both results suggest that all six hydrophobic mutations are destabilizing, depending on the position of the residue side chains. The destabilization for the six hydrophobic residues ranges from -0.5 to -2.6 kcal/mol (Table 5,6). The average ΔΔG(25ºC) value for the three "core" variants (M53A, L61A and L69A) is -2.0 kcal/mol, similar to that of the three Phe variants (-2.2 kcal/mol). In contrast, the average ΔΔG(25ºC) value for V50A, L42A and L75A is only -0.7 kcal/mol. Moreover, M53A is almost as destabilizing as F58A. These studies suggest that the VHP hydrophobic core consists of the six residues surrounding Phe 58, and this defines the VHP fold. Mutations in this core region destabilize VHP by an average of 2.1 kcal/mol/residue. Note that the ΔG of wild-type VHP is only ~2.9 kcal/mol, making these destabilizations very significant.

One of the ways to accurately assess the contribution of the hydrophobic effect is to make mutations that change Ile to Val. [36; 44] This mutation will not introduce steric hinderence, and only eliminates one methylene group, thus minimizing cavity formation. Five Ile to Val variants (I128V, I154V, I210V, I214V and I306V) were designed in VlsE where the wild-type Ile residues are fully buried. All VlsE variants studied here are less stable than the wild-type protein, as shown by their lower ΔG. This means that the buried Ile residues contribute more to the stabilization of VlsE than Val. The *m*-values of

all variants are similar to that of wild-type VlsE, suggesting that urea unfolding of all these variants can be assumed to be two-state process. The resulting $\Delta\Delta G(H_2O)$ values for the five variants range from -1.1 kcal/mol to -1.9 kcal/mol (average = -1.6 kcal/mol), which shows accurately the destabilizing effect of removing one methylene group from the interior of VlsE.

Structural interpretation of the hydrophobic effect in VHP and VlsE

In both VHP and VlsE proteins, mutations that change a larger hydrophobic side chain to a smaller one lead to a decrease in protein stability. Many attempts have tried to correlate structural parameters to such destabilization. $\Delta\Delta G$s were shown to correlate well with different structural parameters, such as the number of Cα groups within a sphere of 10Å radius, centered in the Cα of the mutated residue (for SNase, r=0.67)[37], the change in side-chain solvent accessible surface areas upon mutation (for Barnase, r=0.83)[47] ,or the number of methylene groups within 6Å of the groups deleted (for Barnase, r=0.91)[47]. However, each of these analyses was originally based on variants of one medium sized protein (~150 amino acid residues) and they did not perform well in a larger set of variants of two or proteins varying in size.[53] The application of these same analyses on variants of both VHP and VlsE did not perform well either, as suggested by the poor linear relationships (Fig. 16). Since these types of analyses reflect the van der Waals interactions that may be different in wild-type proteins and variants, results here suggest that the contribution of van der Waals interactions may be protein specific, or there may be other contributing factors.

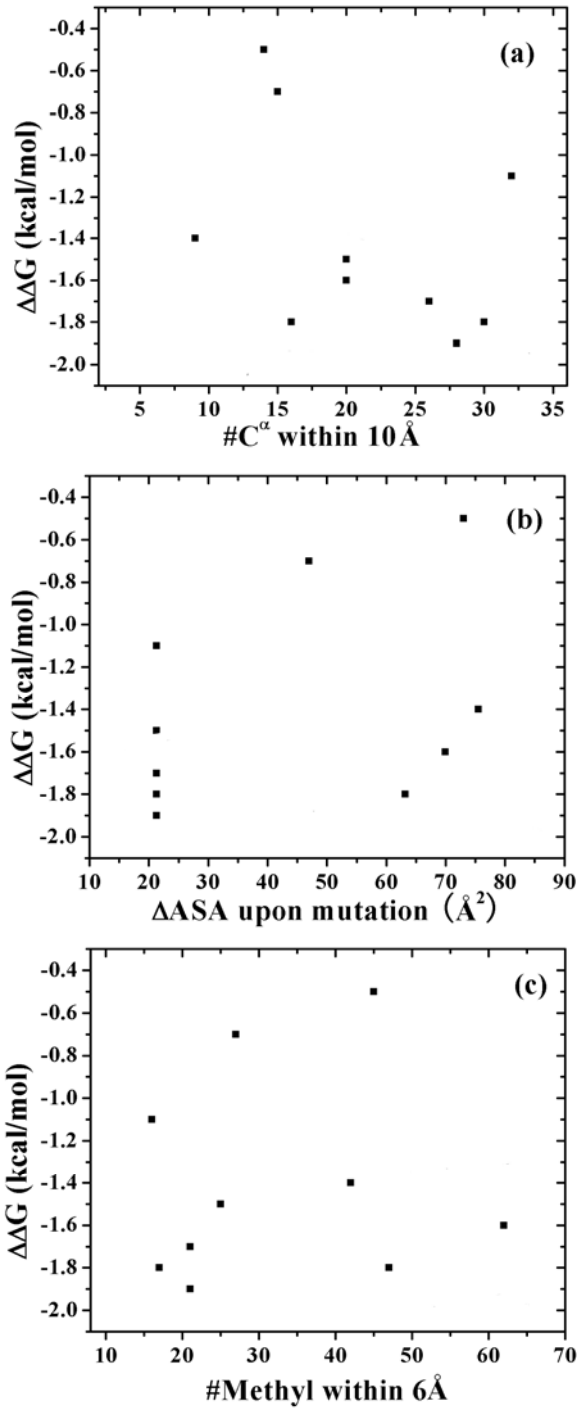

Fig. 16. Plots of ΔΔG vs. structural parameters describing the environments of the changed residues. (A) The number of Cα groups within a sphere of 10Å radius centered in the Cα of the mutated residue. (B) Difference in side-chain solvent accessible surface areas upon mutation. (C) Number of methylene groups within 6Å of the groups deleted.

Table 8. Comparison of ΔΔG values for mutations from larger hydrophobic residues to smaller ones with solvent transfer values

Mutation	$\Delta\Delta G^{a}_{Pace}$ (kcal/mol)	$\Delta\Delta G^{b}_{Survey}$ (kcal/mol)	$\Delta\Delta G^{c}_{VHP}$ (kcal/mol)	$\Delta\Delta G^{d}_{VlsE}$ (kcal/mol)	$\Delta\Delta G^{e}_{transfer}$ (kcal/mol)
Ile to Val	-1.3 ± 0.4	-1.2± 0.9	-	-1.6(-1.0)	-0.8
Val to Ala	-2.5 ± 0.9	-2.3± 1.2	-1.0	-	-1.2
Leu to Ala	-3.5 ± 1.1	-2.8 ±1.4	-1.8	-	-1.9
Met to Ala	-3.0 ± 0.9	-3.4 ± 2.4	-3.2	-	-1.2
Phe to Ala	-3.8 ± 0.3	-3.8± 2.5	-2.3	-	-2.0
Per $-CH_2-$	-1.3	-1.2	-0.6	-1.6	

[a] From Pace (1992).[44]
[b] Data from ProTherm[181] database, where the criteria are those mutations with <30% ASA for the wild-type residue.
[c] From Table 5, 6.
[d] The value in parentheses is the intercept value from Fig. 17.
[e] Difference in side chain transfer energy from water to n-octanol.[32]

Pace conducted a survey on four proteins and reported the average destabilizing effects of mutations from larger hydrophobic residues to smaller ones[44] (Table 8). One of his key findings is that the cost for the loss of one single -CH_2- group is 1.3 kcal/mol. Later, Zhou and Zhou surveyed 1083 hydrophobic variants from various proteins and reached a similar result.[182] Here we also looked at stability data for these same types of mutations as reported in the ProTherm[181] database, and the results were very similar to that of Pace[44], but with larger standard errors (Table 8). Besides, a variation in $\Delta\Delta G$ values among variants for same mutations (i.e. Leu to Ala) is generally observed as that in VHP and VlsE.[37; 183] However, when normalized to 100% surface burial, the average stability decrease from mutation for VHP variants is especially low in comparison with reported values for the same mutation (Table 8). The cost of the loss of one single -CH_2- group is only 0.6 kcal/mol, half of that expected (1.3 kcal/mol/-CH_2-). Such a contribution appears to be observed in VlsE, although a slightly higher average value is observed (1.6 kcal/mol vs. 1.3 kcal/mol). One of the reasons for this may be because the van der Waals interactions are different in both proteins. Since the residues in a larger protein like VlsE are more buried, there are more neighboring interacting residues. Hence, van der Waals interactions may contribute more in larger proteins. This could explain the slightly larger value of 1.6 kcal/mol discussed above, as well as the lower value for VHP. However, the contribution from van der Waals interactions alone cannot account for such big difference in stability observed in both proteins, as shown by the low correlations discussed above. This suggests that the difference in stability among the variants may arise from the environment of the mutation site and the changes brought by

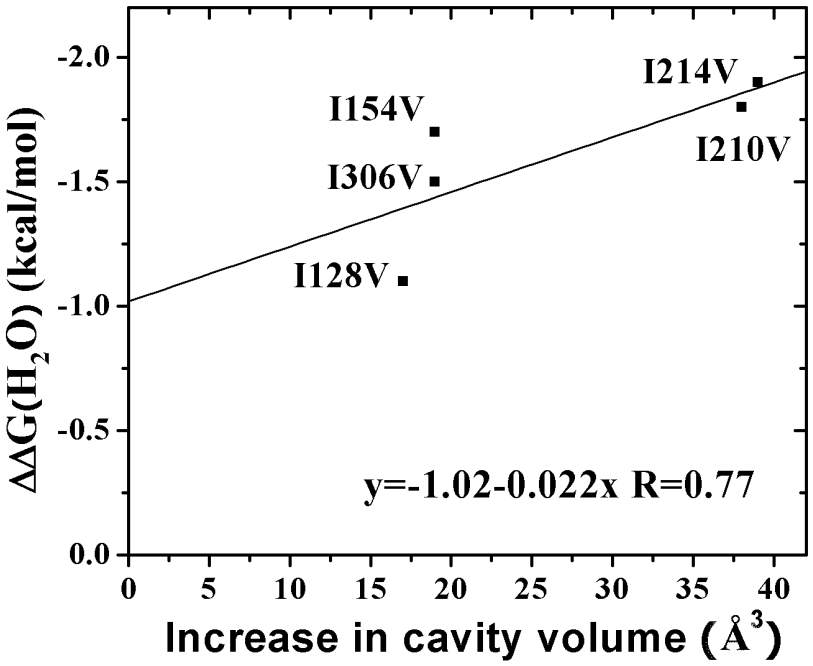

Fig. 17. Plot of ΔΔG vs. the modeled cavity volume for VlsE variants.

the mutations. For example, the size of created cavity, [39] water penetration[184] or secondary structure propensity[42] could all contribute to the stabilities reported here.

It has been proposed that the energetic effect of mutating larger hydrophobic residues to smaller ones can be attributed to a constant term and a variable term.[39; 54] The constant term is related to the water/octanol transfer energy (ΔG_{tr}) of side chain groups,[32] while the variable term correlates with the size of the introduced cavities. Studies of several proteins have found that a slope of ~22 cal/mol/Å^3 best represents this correlation. [39; 41; 53; 54] To investigate if this approach can be applied here, a similar analysis was attempted on both VHP and VlsE variants. Structural analysis of VlsE revealed three non-polar cavities of 174, 116 and 20 Å^3. Such big cavities can greatly destabilize the protein, and this might account for the low stability of VlsE. The cavities in five Ile to Val variants were modeled,[185] and the increase in cavity volumes was found to range from 17 to 39 Å^3. The plot of $\Delta\Delta G$ values on the dependence of the cavity volumes gave a slope of 22 cal/mol/Å^3, in good agreement with other reported values. [39; 41; 53; 54] The intercept of -1 kcal/mol is also in reasonable agreement with the reported transfer energy difference of 0.8 kcal/mol for an Ile to Val (Fig. 17).[32] In contrast, for VHP, since the size of the protein is so small, the "buried" hydrophobic residues are actually much closer to the protein surface. Therefore, the mutations from larger hydrophobic residues to smaller ones do not create a cavity at all. A mutation in this case would only leave a groove on the protein surface. For the most buried residue, Phe 58, the mutation creates a solvent accessible hole across the whole molecule, instead of a cavity. This implies that for VHP, the "variable term" caused by cavity creation may be

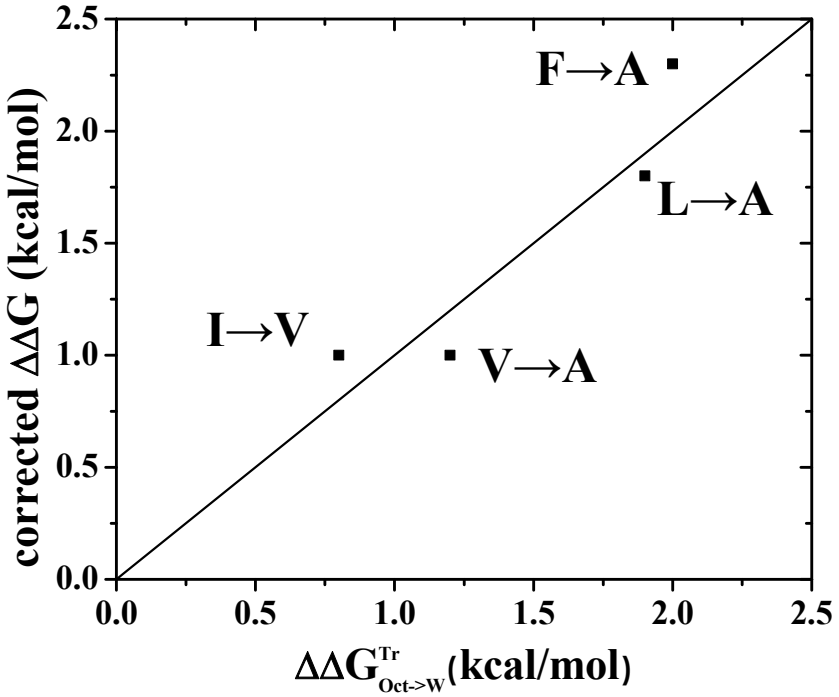

Fig. 18. Correlation between the octanol to water transfer energy difference and the corrected ΔΔG values. The corrected ΔΔG values are the average normalized ΔΔG values (Leu/Val/Met to Ala) for VHP and the intercept from ΔΔG (Ile to Val) vs. cavity volume of VlsE. The solid line shows a unity fit.

negligible. The energetic contribution of mutations in VHP could be described solely by the "constant" term, which may be modeled as the difference in water/octanol transfer energy of side chain groups ($\Delta\Delta G_{tr}$).[32] Indeed, the normalized $\Delta\Delta G$ values for Leu/Val/Phe to Ala from the VHP mutations are in qualitative agreement with the corresponding transfer energy differences (Table 8, Fig. 18). As for the Met to Ala mutation, a cavity appears to exist in the modeled variant structure that could lead to its higher destabilizing effect. Taken together, these studies on both VHP and VlsE support the practice of using $\Delta\Delta G_{tr}$ values from water/octanol to estimate the stability gained by burying the hydrophobic side chains in folded proteins.[186]

Conclusion

We have made 9 variants of a small protein (VHP) and 5 variants of a large protein (VlsE), where in each case a larger hydrophobic side-chain was changed to a smaller one. These mutations did not introduce any obvious steric hindrance or disruption of tertiary structure. All the mutations lead to a decrease in stability. The mutations in the 6 residues that consist of the hydrophobic core of VHP cause substantial decreases in stability; the average $\Delta\Delta G$ value for these 6 variants is -2.1 kcal/mol. However, the protein still folds, suggesting that such a mutation is well tolerated by the very small protein. An analysis of the decrease in stabilities reveals that the cost for the loss of one single $-CH_2-$ group in VHP is 0.6 kcal/mol, only half of the expected value (1.3 kcal/mol)[44]. In contrast, the same cost in VlsE is 1.6 kcal/mol, which is very similar to the expected 1.3 kcal/mol value. Several structure-energy correlations were considered to explain such a difference. The analysis indicates that the different

destabilization magnitudes of hydrophobic mutations for VHP and VlsE can be well explained by the difference in hydrophobicity and modeled cavities, as was previous proposed for T4 lysozyme variants. [39] The results suggest that the hydrophobic effect of small protein like VHP and large protein like VlsE can be described using a similar energy function, hence, it provides a general protein stabilization mechanism. Our results will be useful for both experimental and computational studies on protein folding and stability.

CHAPTER IV

ENGINEERING PROTEIN STABILITY BY MODULATING ENTROPIC CONTRIBUTIONS IN BETA-TURNS

Introduction

Proteins are involved in most biological processes, and globular proteins must fold into a unique three dimensional structure in order to carry out their function. Because the folded conformation of a protein is only about 5 to 15 kcal/mol more stable than its unfolded forms,[15] increasing the conformational stability of many proteins has always been an important goal for both basic research and industrial applications. Although *in vitro* selection has been used to a limited extent to increase protein stability,[111; 112] most approaches involve site-directed mutagenesis in order to best optimize the various forces that contribute to protein stability .[90; 115; 119; 121; 130; 187]

Reverse turns (or β-turns) are the smallest component of protein secondary structure. These turns usually contain four residues designated as i, i+1, i+2 and i+3, and may be grouped into seven classes, depending on the values of their ϕ and ψ angles. The role of β-turns in determining protein stability is not well understood. Statistical analysis has identified the occupancy frequency for amino acids at β-turn positions, and this analysis has been used as one approach to increase protein stability by optimizing the turns according to their positional propensities .[120; 124; 188-190]

We have previously described such a strategy for increasing protein stability by optimizing β-turns.[140] This strategy was based on the fact that at certain positions of β-

turns, the backbone ϕand ψ angles are intrinsically suitable for proline or glycine. For example, the type I β-turn position i+1 assumes a ϕ angle of -60º, similar to that of proline, while the type II β-turn position i+2 assume a ϕ angle of 90º, which causes a steric conflict for all amino acid residue types except glycine. In the case of RNase Sa, at such positions, we replaced the wild-type residue with the statistically favored proline and glycine residues, resulting in an increased stability for five variants from 0.7 to 1.3 kcal/mol.[140]

Here we report our results when this strategy is applied to eight diverse proteins: (1) the cold shock protein from *Bacillus subtilis* (CspB-Bs), (2) ubiquitin , (3) RNase Sa2, (4) RNase Sa3, (5) RNase T1, (6) RNase HI, (7) tryptophan synthatase α-subunit (TSα) and (8) maltose binding protein (MBP). These proteins range in size from 66 to 370 residues and have well-defined x-ray structures. The thermodynamic properties of these proteins have been extensively studied by either chemical or thermal denaturation.[75; 158-162; 165]

Results

Design of mutations

The program PROMOTIF [191] was used to search the pdb files of the proteins of interest to identify β-turn sequences. The study was limited to type I, II and IV β-turns since they represent the majority of the nine different types of known β-turns. The statistical preferences of the turn residues were then compared to those from Guruprasad

Table 9. Positional preferences of wild-type residues and statistically favored proline or glycine residues

Protein	Residue	turn-type	turn-position	wild-type preference	favored preference [a]	increase in preference
CspB-Bs	Asn 55	I	i+1	Asn (0.71)	Pro (4.29)	3.58
	Gln 23	IV	i+3	Gln (0.81)	Gly (1.96)	1.15
ubiquitin	Ser 19	I	i+1	Ser (1.61)	Pro (4.29)	2.68
RNase Sa2	Asn 33	I	i+1	Asn (0.71)	Pro (4.29)	3.58
	Gly 50	II	i	Gly (0.97)	Pro (1.91)	0.94
	Asn 51	II	i+1	Asn (0.47)	Pro (4.92)	4.45
RNase Sa3	Ser 34	I	i+1	Ser (1.61)	Pro (4.29)	2.68
	Ser 51	II	i	Ser (0.65)	Pro (1.91)	1.26
	Thr 52	II	i+1	Thr (0.71)	Pro (4.92)	4.21
RNase T1	Ser 8	IV	i+3	Ser (1.03)	Gly (1.96)	0.93
	Ser 37	I	i+3	Ser (0.94)	Gly (2.57)	1.63
	Ser 63	I	i+1	Ser (1.61)	Pro (4.29)	2.68
RNase HI	Ala 93	I	i+1	Ala (1.07)	Pro (4.29)	3.22
	Gly 123	I	i+1	Gly (0.42)	Pro (4.29)	3.87
	Ala 125	I	i+3	Ala (0.90)	Gly (2.57)	1.67
TSα	Thr 24	II	i	Thr (0.94)	Pro (1.91)	0.97
	Ala 158	I	i+3	Ala (0.59)	Gly (2.57)	1.98
	Ala 189	IV	i+3	Ala (0.83)	Gly (1.96)	1.13
MBP	Gly 13	I	i+1	Gly (0.42)	Pro (4.29)	3.87
	Asn 205	IV	i+3	Asn (1.24)	Gly (1.96)	0.72
	Ala 206	I	i+1	Ala (1.07)	Pro (4.29)	3.22

Positional preference as determined by Guruprasad and Rajkumar

[a] Favored refers to glycine or proline preferences that are statistically significant as determined by Guruprasad and Rajkumar. [192]

and Rajkumar[192] in order to eliminate the already favored proline or glycine sites. We also applied some guidelines for replacing wild-type residues with proline or glycine that we learned from our previous study.[140] We did not target residues that (1) had charged side chains (2) were involved in hydrogen bonding or were near a disulfide bond. (3) for a proline substitution, if the backbone amide proton participated in a hydrogen bond or (4) if the residue had a bulky hydrophobic side chain. In total, 21 sites (14 proline and 7 glycine) were identified for the 8 proteins, and the characteristics of these sites are given in Table 9. In addition, their positions in the protein structures are shown in Fig. 19. From these 21 sites, 14 are proline sites and 7 are glycine sites. These are rather small numbers considering that 8 sites were identified previously for just one protein, RNase Sa. However, the number of sites seems to vary in different proteins. For example, we did not find any suitable sites in thioredoxin and NTL9 although in each protein more than 8 turn sequences were found. Most of the proteins in our set have no more than 3 sites. It is interesting that the most turn sequences (13) were found in the 370 residue protein MBP, yet there are only 3 sites suitable for mutation. This may be because the percentage of turn residues is smaller in a large protein like MBP. Another reason could be that the more surface burial in large protein leads to the formation of more hydrogen bonds for those polar turn residues.

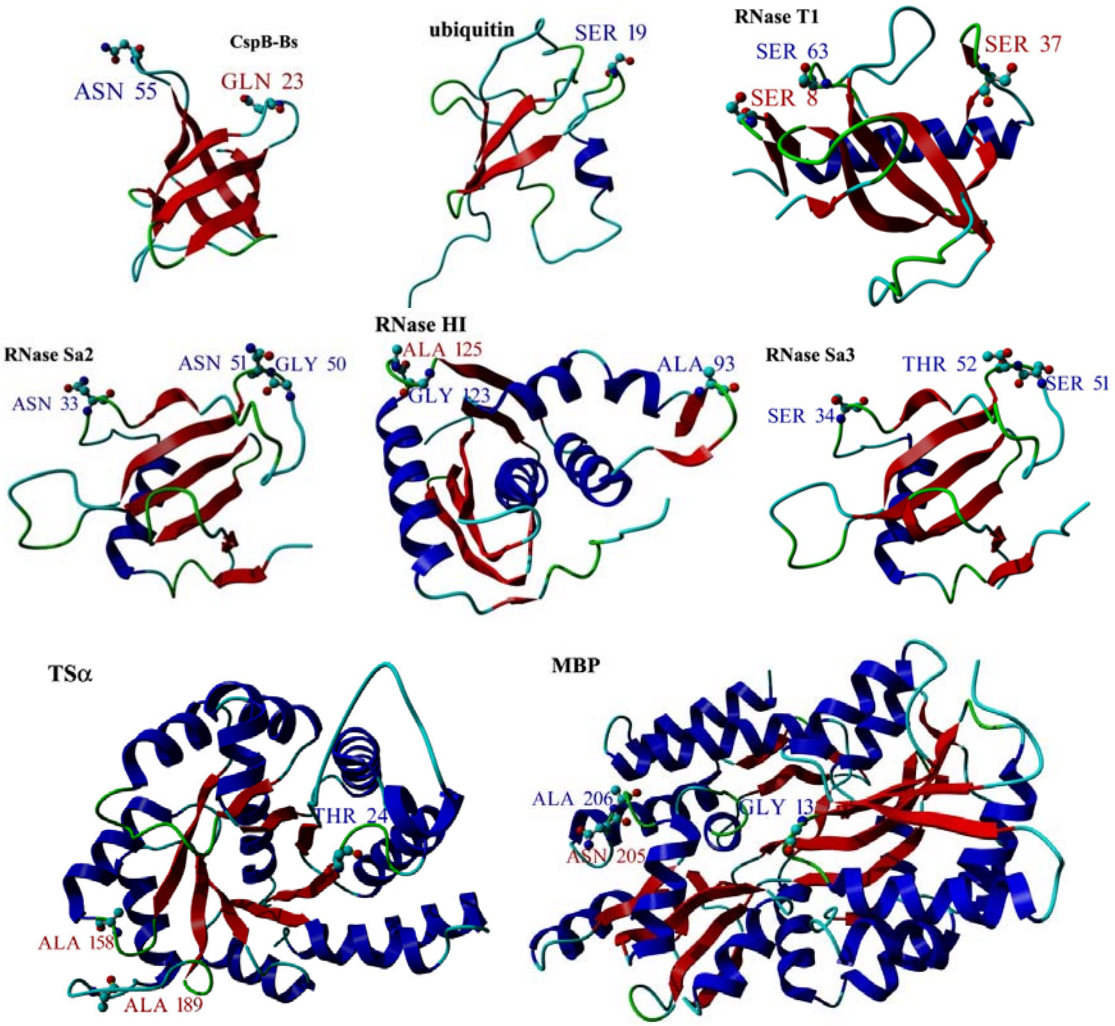

Fig. 19. Ribbon diagram showing the candidates for mutation in β-turn sequence in the 8 proteins studied. The following crystal structures were used: 1CSP (CspB-Bs),[193] 1ZW7 (ubiquitin),[194] 1PYL (RNase Sa2),[195] 1MGR (RNase Sa3), [196] 9RNT (RNase T1),[197] 2RN2 (RNase HI),[198] 1WQ5 (Tryptophan Synthase α subunit),[199] and 3MBP (Maltose Binding Protein).[200] The program YASARA was used to generate the figures.[106] Those residues labeled as blue were mutated to proline and those labeled as red were mutated to glycine.

Table 10. Parameters characterizing the thermal denaturation of β-turn protein variants

Protein	Variant	Turn -type	Position	ΔH_m [a]	$\Delta\Delta H_m$ [b]	ΔS_m [c]	$\Delta\Delta S_m$ [d]	T_m [e]	ΔT_m [f]	$\Delta\Delta G$ [g]
RNaseSa2	WT	N/A	-	69.1±0.5	-	220	-	41.4±0.1	-	-
	N33P	I	i+1	72.9±1.5	3.8	230	10	43.8±0.2	2.4	0.5
	G50P	II	i	73.2±0.3	4.2	229	9	46.7±0.1	5.3	1.2
	N51P	II	i+1	71.9±1.0	2.8	226	7	44.4±0.1	3.0	0.7
	P50/51	II	i/i+1	72.6±1.4	3.5	226	6	48.4±0.2	7.0	1.5
RNaseSa3	WT	-	-	122±3	-	366	-	59.2±0.1	-	-
	S34P	I	i+1	120±2	-1	359	-7	61.7±0.1	2.5	0.9
	S51P	II	i	121±2	-1	362	-3	61.7±0.1	2.5	0.9
	T52P	II	i+1	118±1	-4	353	-12	60.6±0.1	1.4	0.5
	P51/52	II	i/i+1	119±4	-3	365	-11	62.2±0.1	3.0	1.1
RNaseT1	WT	-	-	99.7±1.4	-	306	-	52.3±0.1	-	-
	S8G	IV	i+3	93.5±1.8	-6.2	289	-18	50.8±0.1	-1.5	-0.5
	S37G	I	i+3	93.2±0.3	-6.5	285	-21	53.5±0.2	1.2	0.4
	S63P	I	i+1	92.7±1.0	-7.0	282	-22	55.0±0.1	2.7	0.8
CspB-Bs	WT	-	-	44.9±1.9	-	137	-	55.0±0.1	-	-
	N55P	I	i+1	48.3±0.1	3.4	144	7	62.4±0.1	7.4	1.0
	Q23G	IV	i+3	40.0±0.5	-4.9	122	-15	54.7±0.1	-0.3	0.0
RNaseHI	WT	-	-	69.9±1.7	-	216	-	50.3±0.3	-	-
	A93P	I	i+1	68.8±5.1	-1.1	213	-3	49.8±0.2	-0.4	-0.1
	G123P	I	i+1	73.4±1.0	3.5	226	10	51.7±0.3	1.4	0.3
	A125G	I	i+3	70.0±2.5	0.1	217	1	49.9±0.2	-0.4	-0.1
TSα	WT	-	-	88.3±2.5	-	266	-	59.4±0.1	-	-
	T24P	II	i	73.3±3.0	-15.1	222	-44	56.9±0.2	-2.5	-0.7
	A158G	I	i+3	81.3±2.1	-7.0	247	-19	56.0±0.1	-3.4	-0.9
	A189G	IV	i+3	81.1±1.6	-7.2	244	-22	59.5±0.2	0.1	0.0
MBP	WT	-	-	199±3	-	593	-	62.5±0.1	-	-
	G13P	I	i+1	198±5	-1	589	-4	62.4±0.1	0.0	0.0
	N205G	IV	i+3	186±9	-14	555	-38	61.3±0.2	-1.2	-0.7
	A206P	I	i+1	198±11	-1	589	-4	62.4±0.2	-0.1	-0.1

[a] ΔH_m, enthalpy of unfolding at T_m (in kcal/mol).
[b] $\Delta\Delta H_m = \Delta H_m$(variant)$-\Delta H_m$(wild-type) (in kcal/mol).
[c] $\Delta S_m = \Delta H_m / T_m$ (in cal/mol/K).
[d] $\Delta\Delta S_m = \Delta S_m$(variant)$-\Delta S_m$(wild-type) (in cal/mol/K).
[e] T_m=T (in °C), where $\Delta G°=0$.
[f] $\Delta T_m = T_m$(variant)$-T_m$(wild-type) (in °C).
[g] $\Delta\Delta G = \Delta S_m$(wild-type)$*\Delta T_m$.(in kcal/mol) A positive value indicates an increase in stability.

Characterizing of wild-type proteins and variants

We do not expect single mutations on the surface to alter the three dimensional structure of a protein. Nevertheless, the structural properties of the wild-type proteins and their variants were characterized using circular dichroism spectroscopy. The far UV CD spectra were very similar for all wild-type proteins and their variants (data not shown). In addition, for RNase Sa2, RNase Sa3 and RNase T1, all variants retained their specific activity comparable to that of wild-type proteins (data not shown). These results suggest that the protein conformations are not changed significantly by the mutations made in this study.

A thermodynamic characterization of the wild-type proteins and their variants were done by thermal denaturantion or urea denaturation. The results are summarized below:

RNase Sa2/Sa3

RNase Sa2, RNase Sa3 and RNase Sa share 50% amino acid sequence identity. The sequence alignment identified 7 turns that exist in all three proteins (Fig. 20). In our previous study of RNase Sa, three residues were mutated and the effects of mutation were assessed. The S31P and S48P mutations increased the protein stability while Y49P had almost no effect on protein stability.[140] The latter may be attributed to the removal of favorable π-π interactions between the aromatic ring of this tyrosine and neighboring phenylalanines. To further test our strategy, we first designed three single mutation variants and one double mutation variant of RNase Sa2 and RNase Sa3 on the residues at the same turn positions as in RNase Sa (Fig. 20, in bold). All variants of RNase Sa2 and

RNase Sa3 were observed to unfold at higher temperature as expected (Fig. 21a, 21b). The ΔT_ms for RNase Sa2 variants range from 2.4 ºC to 7 ºC while the ΔT_m for RNase Sa3 variants range from 1.4 ºC to 3 ºC. The ΔΔGs range from 0.5 to 1.5 kcal/mol for RNase Sa2 and from 0.5 to 1.1 kcal/mol for RNase Sa3 (Table 10). For both proteins, the double proline variants gave the largest increase in stability (1.1-1.5 kcal/mol). The effects of proline mutations on stability has generally been observed to be additive,[140; 142; 144] although Prajapati et al.[201] reported a non-additive behavior for thiredoxin variants. In our double mutations, the prolines introduced are adjacent to each other, yet our results still support this additivity.

```
                                                    :::: 
                                                 .... .... 
Sa1  ---DVSGTVCLSALPPEATDTLNLIASDGPFPYSSQDGVVFQNRESVLPTQ     47
Sa2  -ADPALADVCRTKLPSQAQDTLALIAKNGPYPYNNRDGVVFENRESRLPKK     49
Sa3  ASVKAVGRVCYSALPSQAHDTLDLIDEGGPFPYSSQDGVVFQNREGLLPAH     50

     ::::         ....                              ****      .... 
Sa1  SYGYYHEYTVITPGARTRGTRRIITGEATQEDYYTGDHYATFSLIDQTC     96
Sa2  GNGYYHEXTVVTPGSNDRGTRRVVTG-GYGEQYWSPDHYATFQEIDPRC     97
Sa3  STGYYHEYTVITPGSPTRGARRIITGQQWQEDYYTADHYASFRRVDFAC     99
```

Fig. 20. Sequence alignment of RNase Sa family. Shown are type I (.), II (:) and IV (*) turns. Residues in bold were studied here.

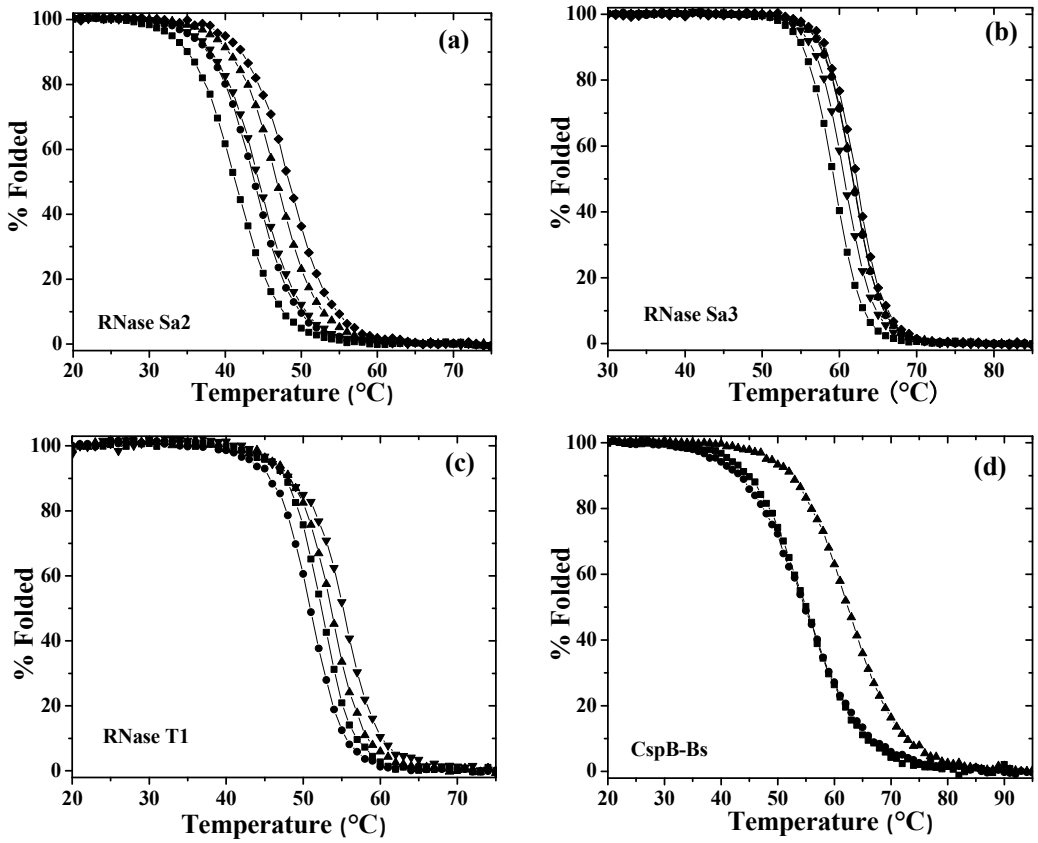

Fig. 21. Thermal and urea denaturation curves for the turn variants. (a) RNaseSa2: WT (■), N33P (●), G50P (▲), N51P (▼), G50P/N51P (♦); (b) RNaseSa3: WT (■), S34P (●), S51P (▲), T52P (▼), S51P/T52P (♦); (c) RNaseT1: WT (■), S8G (●), S37G (▲), S63P (▼); (d) CspB-Bs: WT (■), Q23G (●), N55P (▲); (e) RNaseHI: WT (■), A93P (●), G123P (▲), A125G (▼); (f) TSα: WT (■), T24P (●), A158G (▲), A189G (▼); (g) MBP: WT (■), G13P (●), N205G (▲), A206P (▼); (h) ubiquitin: WT (■), S19P (●).

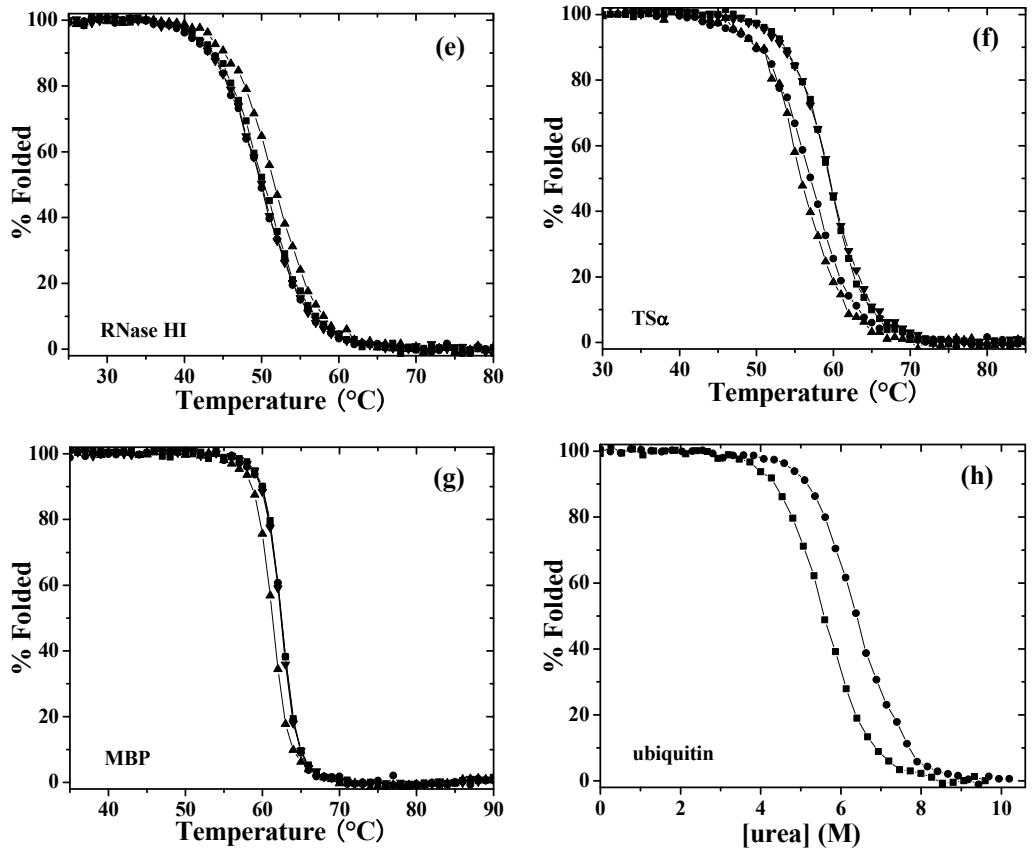

Fig. 21. Continued.

RNase T1

Three variants were designed in RNase T1. The three variants showed different stabilizing effects: S37G and S63P are more stable than the wild-type RNase T1, while S8G is less stable (Fig. 21c). Non-linear least squares fit of the data gives a T_m of 52.3 ºC for the wild-type RNase T1. For S37G and S63P, the T_ms are 1.5 ºC and 2.7 ºC higher than the wild-type RNase T1, which correspond to a conformational stability increase of 0.4 and 0.8 kcal/mol respectively. In contrast, the stability of S8G is 0.5 kcal/mol lower than the wild-type RNase T1 (Table 10).

CspB-Bs

Cold shock protein has been extensively studied as a model for optimizing charge-charge interactions in proteins. Two variants, Q23G and N55P, were designed here. Q23G has the same thermal stability as that of the wild-type CspB-Bs, while N55P is significantly more stable than wild-type CspB-Bs (Fig. 21d). Analysis of thermal denaturation curve gives a T_m of 55.0 ºC for the wild-type CspB-Bs, very close to previously reported value. [179] The N55 position was previously shown to favor negative charged amino acids; mutation to E or D gave a ΔT_m of 4 ºC.[179] In this study, the T_m for N55P variant is 62.4 ºC, 7.4 ºC higher than the wild-type CspB-Bs, this translates to an increase in conformational stability of 1 kcal/mol (Table 10).

RNase HI

RNase HI aggregates at neutral pH. To assure >90% reversibility, the denaturation of wild-type RNase HI and variants were performed at pH 5.5 in 1M GuHCl (Fig. 21e). For the three designed variants, only G123P has higher thermal

stability than wild-type RNase HI. The A93P and A125G variants have a melting temperature similar to wild-type RNase HI. The T_m of G123P is 51.7 ºC, 1.4 ºC higher than that of wild-type RNase HI. This ΔT_m corresponds to 0.3 kcal/mol increase in conformational stability (Table 10).

TSα

The thermal stability of TSα and its three variants were determined by far-UV CD (Fig. 21f). Two of the variants are less stable than the wild-type TSα, while the T_m of A189G variant is similar to that of wild-type TSα. The decrease in T_m for T24P and A158G variants are 2.5 ºC and 3.4 ºC, which correspond to decrease in conformational stability of 0.7 and 0.9 kcal/mol at 25 ºC respectively (Table 10).

MBP

MBP has two domains, however, the thermal and chemical denaturation of MBP both indicate a two-state unfolding mechanism.[161] None of the variants are more stable than wild-type MBP (Fig. 21g). Two variants, G13P and A206P have the same T_m as wild-type MBP, while the variant N205G has a lower T_m. The T_m decrease for N205G is 1.2 ºC; this corresponds to a decrease in conformational stability of 0.7 kcal/mol (Table 10).

Ubiquitin

Thermal denaturation of ubiquitin at pH >5 is not reversible. Therefore, we measured the conformational stability of ubiquitin and its S19P variant using urea denaturation . The yeast ubiquitin, which is studied here, differs from human ubiquitin by only 4 residues, yet human ubiquitin is more stable than yeast ubiquitin [202]. In human

Table 11. Parameters characterizing the urea denaturation of wild-type and S19P yeast ubiquitin

Variant	Turn - type	Position	$D_{1/2}$ [a]	$\Delta G^o(H_2O)$[b]	$\Delta\Delta G^o$[c]
wild-type	-	-	5.44±0.05	5.4±0.1	-
S19P	I	i+1	6.33±0.10	6.3±0.1	0.9

Experiments were done at 25ºC, 30 mM NaOAc (pH 5.0)

[a] $D_{1/2}$ (in M) is urea concentration at the midpoint of the unfolding transition curve.

[b] ΔG^o (H_2O) = m*$D_{1/2}$ (in kcal/mol), wild-type m value 1.00 kcal/mol/M was used.

[c] $\Delta\Delta G^o = \Delta G^o$ (variant) - ΔG^o (wild-type). A positive value indicates an increase in stability.

ubiquitin, position 19 is a proline. The stability of the designed S19P variant is higher than the wild-type ubiquitin (Fig. 21h). The measured conformational stability of wild-type ubiquitin is 5.4 kcal/mol, in good agreement with previous reported value.[86] The designed S19P variant unfolds at a higher urea concentration, resulting in a 0.9 kcal/mol increase in conformational stability (Table 11).

Discussion

Thermodynamic effects of mutations

The "proline rule" proposed by Suzuki et al. states that the increases in the frequency of proline occurrence at the second sites of type I or type II β-turns can enhance protein thermostability.[142] Their study and studies by others[144; 203] were based on the analysis of mesophilic enzymes with their thermophilic counterparts, which suggested that the increased number of proline residues found in thermophilic proteins could contribute to the higher thermal stability. While this approach is especially useful in large proteins when there is significant difference in the proline profiles, we have shown one example for a small protein HPr.[204] The thermophilic Bst HPr has a proline at position 56 while mesophilic Bs HPr has an alanine at this position. This proline contributes about 0.7 kcal/mol to the higher stability of Bst HPr, and is in the first position of a type II β-turn. Most of our proline substitutions reported here are in the second positions of type I and type II β-turns, and also increased protein stability (Table 10, Table 11).

The stabilizing effect of the proline substitution is generally attributed to the fact that proline constrains the backbone configuration thus decrease the conformational entropy in the denatured state. Unlike other residues, proline has a constrained ϕ main chain dihedral angle of -63±15º. For type I and type II β-turns, the ϕ angle of -60º for the second position makes these positions suitable for proline introduction. As for the first position of type II β-turn, there is no structural preference. Hence, a major consideration for proline introduction is to have compatible ϕ and ψ angles, especially the ϕ angle.[145; 205] All of the wide-type residue ϕ angles of our proline variants fall into the range of ~ -60º to ~ -90º, regardless of the positions (Fig. 22a).

The decrease in conformational entropy in the denatured state for proline substitution suggests that we should observe a negative ΔΔS for the proline mutants. However, the ΔΔS of our results vary considerably, ranging from -54 cal/mol/K to 10 cal/mol/K, (Table 10). Interestingly, for RNase Sa2 and RNase Sa3, the ΔΔS of the same position mutants have the similar values but with different signs. Another observation is that the value of most ΔΔS is larger than expected, according to Nemethy et al.[206] The same deviation was also observed for a series of proline variants by Prajapati et al.[145] This suggests that other factors may also contribute to the observed ΔΔS.

For the glycine variants, this strategy worked very well in RNase Sa[140]. All three such substitutions in RNase Sa as well as S37G of RNase T1in this study gave increase in Stability. Structural analysis indicates that wild-type residues assume ϕ, ψ angles that's in the L-α region on the Ramachandran plot (0º<ϕ<90º, 0º<ψ<90º). [207] The reason

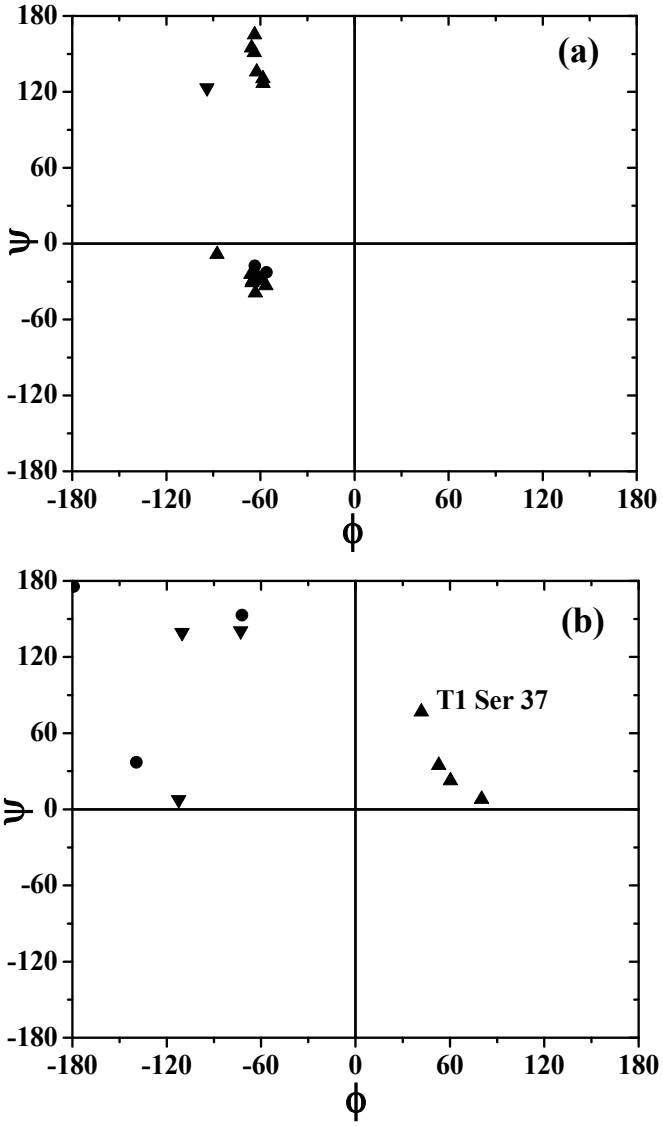

Fig. 22. Ramachandran plot for all the wild-type turn residues. Shown here are plots for residues that were mutated to (a) prolines or (b) glycines indicating those that are stabilizing (▲), destabilizing (▼) or have no effects (●).For comparison, the ϕ ψ data for the wild-type turn residues of RNaseSa from Trevino et al.[140] were also calculated and plotted here.

for this increase is probably because when a non-glycine residue resides in this region, there is a steric hindrance between the β carbon and the peptide backbone. Therefore, in such cases, a substitution to glycine will remove the van der Waals repulsion as well as relieve the hindrance and may thus slightly increase the conformational entropy in the native state. The latter has the same effect as to reduce the entropy of unfolded state, (i.e, a reduced ΔS for unfolding,) therefore increasing protein stability. There are several studies involving glycine substitution that show increases in stability of 1 – 2 kcal/mol.[208-210] By definition, the second position of type II or type I' β-turns has the ϕ,ψ angles in the L-α region of Ramanchandra plot, thus a non-glycine residue at such turn positions will make a negative contribution to the protein stability. However, since glycine has the highest backbone entropy of all residues, substitutions from glycine to a more rigid residue are entropically favored. Indeed, Matthews et al.[141] found that glycine to alanine substitution in T4 lysozyme leads to a small increase in stability (0.4 kcal/mol). Therefore, substitutions to glycine would be mostly destabilizing, which is observed in most of the glycine variants in this study. Although statistically favored, the last position of type I and type IV turns does not have constrained ϕ,ψ angles, suggesting that the effect of substitution to glycine in such position will not be as predictable. (Fig. 22b). Hence, the substitution S37G resulted in an increase in stability, while most of other glycine variants did not. In addition, the change of conformational entropy upon substitution to glycine suggests a negative ΔΔS for those residues in L-α region and a positive ΔΔS for the others. However, our results here show almost all negative ΔΔS, indicating other factors are also involved.

Structural interpretation of the stability effects of mutations

Based on the backbone conformational entropy change due to the substitution, the expected ΔΔG values for substitution to proline range from 0.7 kcal/mol (for threonine to proline substitution) to 2.1 kcal/mol (for glycine to proline substitution) (Table 12). The comparison of the predicted values to the experimental results suggests a general correlation and reasonable agreement, yet the experimental values are smaller in all cases, as has been previously shown by Matthews et al.[141] They attributed this as the simplification of theory, and indeed the work by Yun et al.[211] suggested that the conformational entropy change ΔΔS may be overestimated. Alternately, this discrepancy may reflect the difference in solvation energy for proline and the residues substituted. For those proline substitutions that did not have stabilizing effects, the targeted residues are either mostly buried (i.e. T24P in TSα) or involve a glycine to proline substitution (i.e. G13P in MBP). Substitution in such positions will lead to the loss of hydrophobic contact (for large residues) or a steric conflict (for small residues). The only exception is the RNase HI A93P variant. This residue is mostly exposed, yet the substitution to proline has almost no effect on stability. The side chain of the two neighboring residues, threonine 92 and aspartic acid 94, each forms a hydrogen bond with the backbone nitrogen of lysine 96. Therefore, the constraint introduced by proline at this position may disrupt this hydrogen bonding network, counteracting the stabilizing entropic effect.

The stability effect of a substitution to glycine has larger variation, although the predicted ΔΔG based on conformational entropy change are all negative (Table 12). For buried positions, the substitution should have a further destabilizing effect because of the

Table 12. Structural parameters and expected entropic changes for turn variants

Protein	Mutation	B-factor[a]	SideChain Burial%[b]	$-T\Delta\Delta S_{BB}$[c]	$\Delta\Delta G^{o}_{m}$[d]
Favored proline substitution					
CspB-Bs	N55P	34	6.5	1.31	1.0
ubiquitin	S19P	7	5.7	1.21	0.9
RNaseSa2	N33P	34	-7.2	1.25	0.5
	G50P	27	100(24.6)	2.00	1.2
	N51P	31	6.3	1.21	0.7
RNaseSa3	S34P	15	16	1.33	0.9
	S51P	30	6.6	1.33	0.9
	T52P	30	22.5	0.70	0.5
RNaseT1	S63P	10	38.6	1.30	0.8
RNaseHI	A93P	20	18.8	1.29	-0.1
	G123P	29	100(20.6)	2.07	0.3
TSα	T24P	25	61.8	0.70	-0.7
MBP	G13P	19	100(71.9)	2.13	0.0
	A206P	24	68.6	1.34	-0.1
Favored glycine substitution					
CspB-Bs	Q23G	56	55.4	-0.89	0.0
RNaseT1	S8G	16	-3.1	-0.78	-0.5
	S37G	12	24.1	-0.78	0.4
RNaseHI	A125G	30	45.7	-0.78	-0.1
TSα	A158G	36	100	-0.80	-0.9
	A189G	36	17.2	-0.80	0.0
MBP	N205G	27	31	-0.80	-0.7

[a] Average B-factors for backbone atoms.
[b] Accessibility values for the wild-type residue. Values in the parentheses are the average for the two neighboring residues. The program *pfis* was used to calculate the ASAs. [75]
[c] $\Delta\Delta G$ contribution due to the difference of backbone conformational entropy calculated according to Nemethy et al (in kcal/mol).[206] Temperatures used are T_ms for wild-type proteins and 25 ºC for S19P ubiquitin.
[d] Measured stability change as in Table 10 (in kcal/mol).

hydrophobic effect, however this is not observed. The reason may be attributed to the higher flexibility for these positions substituted (CspB-Bs Q23G, RNaseHI A125G, TSa A189G) as reflected by their higher *B*- factor values (Table 12). A more flexible region can easily rearrange to compensate the loss of hydrophobic contact introduced by a mutation. In addition, the conformational entropy may have very small difference for different types of amino acid residues in such positions. Therefore, different amino acid residues in such positions may have very small difference in their contribution to protein stability. It is indeed observed in this study that some of our glycine substitution variants (CspB-Bs Q23G, RNaseHI A125G, TSa A189G) show no stability difference from wild-type proteins. These results are in agreement with previous reports by others where researchers have observed large stability changes from mutations only in rigid regions.[167; 209; 212]

Besides the backbone conformational entropy, the side chain conformational entropy has also been suggested to have a role in the stability change observed in proline substitutions. Agah et al. argued that their H26P mutation in Parvalbumin gains more stability increase than a A62P substitution because histidine has a higher side chain entropy than alanine, resulting in a higher entropy loss. [203] However, such analysis would only be appropriate when the side chain of mutated residue is buried; for a solvent exposed residue, the side chain entropy should have no difference either in the native state or in the unfolded state. The further gain of stability in their H26P variant may be due to improved charge-charge interactions. In addition, in a buried position, the hydrophobic contribution should also be included to account for the stability change as

discussed above. Our ability to predict the ΔΔG for such substitutions made in this study, therefore, will depend on our ability to correctly assess the side chain entropy and hydrophobic interaction contributions from polar or non-polar groups. However, such assessments are still in debate.

Conclusion

By using a statistical approach, we have redesigned the solvent exposed turns in eight proteins ranging in size from 66 to 370 residues. In most of the proline substitutions, an increase in stability was observed, yet, the increase in stability is only observed in one variant that was mutated to glycine. Our main finding is that at certain positions on turns, substitutions to proline have a stabilizing effect presumably because of their constrained ϕ, ψ angles. In other positions the substitution to proline or glycine will also have a stabilizing effect only if the wild-type residues are in the specific ϕ, ψ space and are not buried. These substitutions either constrain the backbone to decrease the entropy in denatured state, or relieve steric hindrance in the native state. Both approaches reduce the conformational entropy change upon unfolding and stabilize the protein. Our strategy worked well for substitutions on residues that are mostly solvent exposed, especially for substitutions to proline. In addition, in our initial identification process, we intentionally eliminated the charged amino acid residues as our targets. However, such precaution is not completely necessary. In fact, it has been reported in almost all proline substitution studies that some mutations from charged amino acid residues to proline increased the protein stability.[144; 145; 213] The work by Makhatadze and associates[127] has now greatly improved our ability to predict the effect charged amino

acid residues on protein stability and can potentially help us locate those positions that are insensitive to the charges for our purpose. In conjunction with their charge-charge energy calculation, our strategy can be improved and easily applied in engineering protein stability.

CHAPTER V

INCREASING PROTEIN STABILITY: IMPORTANCE OF THE DENATURED STATE DEPENDS ON THE STABILIZING MECHANISM

Introduction

Proteins carry out the most important functions in all biological systems. Organisms can be classified as mesophiles, thermophiles or hyperthermophiles based on the difference in optimal growth temperatures.[214] The stabilities of proteins from these organisms have different thermodynamic properties.[215] Thermophilic proteins are of great interest in that they provide the insight into the general mechanisms of protein folding and stabilization. Comparative studies show that thermophilic proteins and their mesophilic homologues often adopt similar three-dimensional structures. [216] A number of factors have been identified to account for the increased stability of thermophilic proteins, which include tighter packing of the hydrophobic core,[129; 217] more and shorter hydrogen bonding,[216] and more favorable charge-charge interactions. [218; 219] Thus it has been difficult to identify specific amino acid residues that contribute to the difference in stability between thermophilic proteins and their mesophilic homologues, because there are often many differences in the primary structure.

RNase Sa is a small enzyme with only 96 amino acid residues and one disulfide bond. The stability and thermodynamics of folding for RNase Sa and its variants have been extensively studied.[71; 75; 165; 220-223] In our previous studies of the stability of RNase Sa, a number of mutations at different positions have been found to increase the stability.

These can be categorized into three groups: (1) those that increase favorable charge-charge interactions (D25K, E74K); [130] (2) those that favorably change the entropic contribution to stability at a β-turn site (S31P, S42G, S48P, T76P and Q77G) [140] and (3) a mutation that removes a buried charged amino acid residue (D79F). Except for the last case, all the residues that were mutated were more than 80% solvent exposed. We have summarized these findings in Table 1. It has been shown that the combination of such stabilizing mutations is often additive; therefore, we combined these mutations to design two RNase Sa variants in which either (1) only the 7 surface mutations are introduced or (2) all 8 mutations are introduced. We then measured the stability of both variants to determine which thermodynamic parameters are the most important in contributing to the increased stability. We hope that by creating a "thermophilic" variant of RNase Sa, we can learn more about the increased stability of thermophilic proteins.

Table 13. Mutations introduced to make RNase Sa 8S and their expected contribution to stability

Mutation Type	Mutation	ΔT_m[a] (ºC)	$\Delta\Delta G^o$[b] (kcal/mol)	$\Delta\Delta G^o$[c] (kcal/mol)
Surface Charge[d]	D25K	3.0[d]	0.9[d]	-
	E74K	3.9[d]	1.1[d]	-
subtotal[g]	**2S**	6.9	2.0	0.8
β-Turn	S31P	2.5[e]	0.7[e]	-
	S42G	2.6[e]	0.7[e]	-
	S48P	4.5[e]	1.3[e]	-
	T76P	3.5[e]	1.0[e]	-
	Q77G	2.9[e]	0.8[e]	-
subtotal[g]	**7S**	22.9	6.5	5.2
Buried Charge	D79F	14[f]	3.0[f]	2.6
total[g]	**8S**	36.9	9.5	8.2

[a] Change in T_m reported previously. [90; 130; 140]
[b] Change in conformational stability reported previously. [90; 130; 140]
[c] Experimental change in conformational stability determined by thermal denaturation.
[d] From Grimsley et al. 1999 [130].
[e] From Trevino et al. 2007 [140].
[f] From Trevino et al. 2005 [90].
[g] Subtotal and total entries assume complete additivity of the effect of mutation on stability

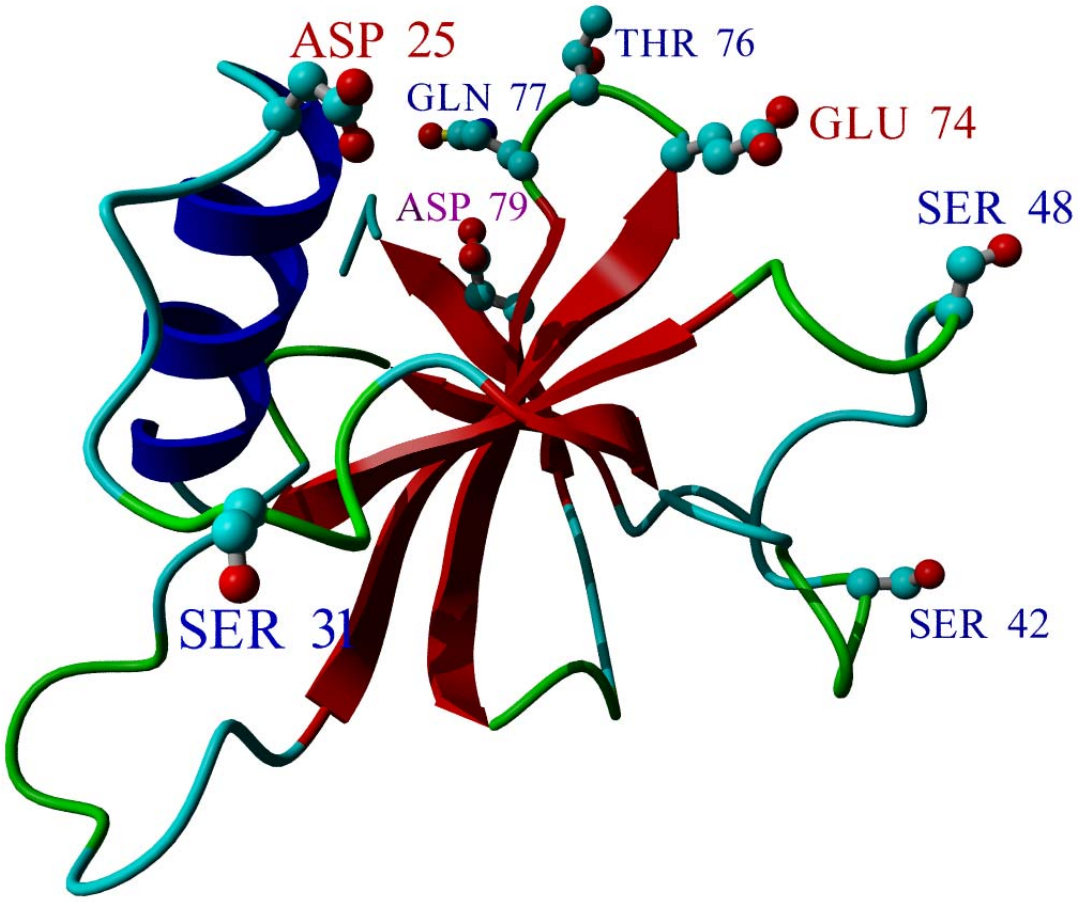

Fig. 23. Ribbon diagram of the RNase Sa structure indicating the mutated residues. The figure was generated using the YASARA program [106] and the 1.2 Å crystal structure of RNase Sa (PDB code 1RGG).[196] The residues are labeled in different colors for different type of mutations to make: red, surface charge; blue, turn; magenta, buried charge.

Results

Design of mutations

Previous studies have identified eight mutations that increased the stability of RNase Sa.[90; 130; 140] In a variant containing all of these mutations, and assuming that the contributions to stability are completely additive, we would expect an increase in T_m to ~ 86.3 ºC, which is 36.9 ºC higher than the wild-type protein (Table 13), and an increase in conformational stability of 9.5kcal/mol. The eight mutations above were introduced in steps. Three variants were designated as 2S (D25K, E74K, where 2S indicates 2 stabilizing mutations), 7S (2S + S31P, S42G, S48P, T76P and Q77G) and 8S (7S + D79F). Eight of the mutations are on the surface, while only D79F is buried (Fig. 23). In order to study the stabilizing mechanisms, six single variants were designed into 7S or 8S, as described below.

Structure characterization of variants

Circular dichroism (CD) spectroscopy was used to monitor any perturbation in the structure induced by the mutations. The CD spectra of all the variants were very similar to that of wild-type RNase Sa (data not shown). These observations suggest that the overall structures are not affected by the mutations introduced.

Thermal denaturation

The thermal unfolding of RNase Sa and its variants were followed using CD. The denaturation curves were analyzed to determine the temperature at the midpoint of the denaturation curves, T_m and the enthalpy change at T_m, ΔH_m. The denaturation curves in Fig. 24 show that all three variants unfold at higher temperatures than wild-type

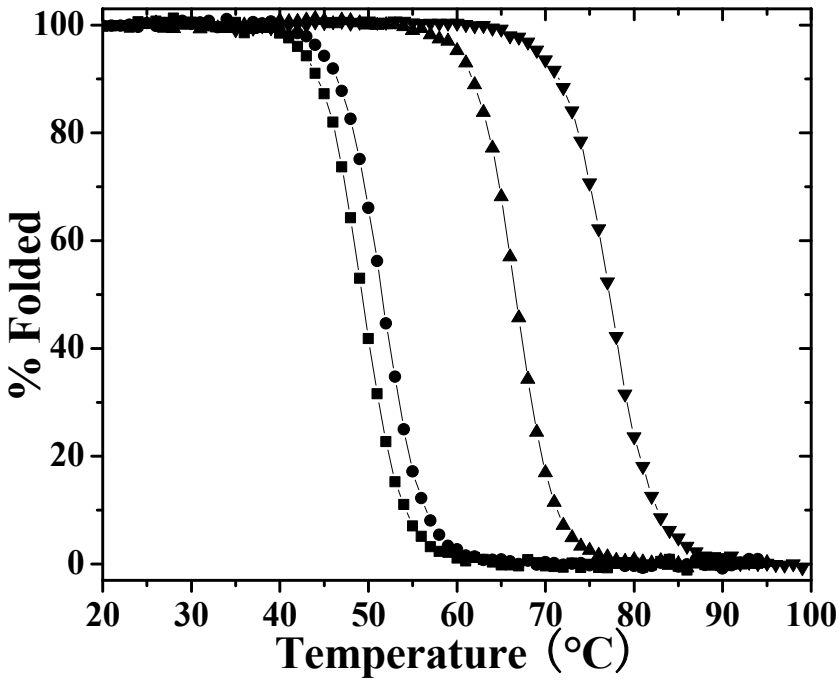

Fig. 24. Thermal unfolding curves for RNase Sa and stabilizing variants. Shown in figure are WT(■), 2S (●), 7S (▲) and 8S (▼).

Table 14. Parameters characterizing thermal unfolding of RNase Sa and stabilized variants

Variant	ΔH_m [a] (kcal/mol)	ΔS_m [b] (cal/mol/K)	T_m (ºC)[c]	ΔT_m [d] (ºC)	ΔG^o (25ºC)[e] (kcal/mol)	$\Delta\Delta G^o$ (25ºC)[f] (kcal/mol)	$\Delta\Delta G^o$ (65ºC)[g] (kcal/mol)	ΔC_p [h] (kcal /mol/K)
WT	93	287	49.3	-	5.5	-	-	1.52[i]
2S	92	284	51.5	2.2	5.8	0.3	0.8	1.52 [i]
7S	105	309	66.6	17.3	8.4	2.9	5.6	1.68
8S	97	277	77.2	27.9	10.1	4.6	8.2	1.07

Experiments were done in 30mM MOPS (pH 7).

[a] ΔH_m, enthalpy of unfolding at T_m.The error is ±5 kcal/mol.

[b] $\Delta S_m = \Delta H_m/T_m$. The error is ±15 cal/mol/K.

[c] $T_m = T$, where $\Delta G^o = 0$. The error is ±0.3C.

[d] $\Delta T_m = T_m$ (variant) – T_m (WT).

[e] ΔG^o, free energy of unfolding calculated using Eq. 5.3 at 25ºC. The error is ±0.2 kcal/mol.

[f] $\Delta\Delta G^o = \Delta G^o$ (variant) - ΔG^o (WT), Positive values indicate increase in stability. The error is ±0.3 kcal/mol

[g] $\Delta\Delta G^o = \Delta G^o$ (variant) - ΔG^o (WT), calculated using Eq. 5.3 at 65 ºC, the average for the T_ms of WT, 7S and 8S. The error is ±0.3 kcal/mol.

[h] ΔC_p, heat capacity change upon unfolding, determined by Kirchoff analysis of thermal denaturation data at pH 7; the errors are ±0.04 and ±0.02 kcal/mol/Kfor 7S and 8S respectively.

[i] Data from Pace et al. (1998) [165]. The ΔC_p of WT was used for 2S variant.

Table 15. Parameters characterizing the thermal unfolding of RNase Sa 7S, 8S and their variants

Variant	ΔH_m (kcal/mol)	ΔS_m (cal/mol/K)	T_m (ºC)	ΔT_m (ºC)	ΔG^o (25ºC) (kcal/mol)	$\Delta\Delta G^o$ (kcal/mol)	ΔC_p [a] (kcal /mol/K)
7S	105	309	66.6	-	8.4	-	1.68
I92D7S	73	237	34.0	-32.6	2.0	-6.4	1.41
D79A7S	102	293	75.3	8.7	10.2	1.8	1.15
8S	97	277	77.2	-	10.1	-	1.07
I92D8S	84	261	50.2	-26.9	5.2	-4.9	1.41
I70D8S	78	247	41.5	-35.6	3.6	-6.5	1.08
I92A8S	102	303	63.9	-13.3	8.7	-1.4	1.33
Y80A8S	87	263	58.0	-19.2	6.5	-3.6	1.29

Experiments were done in 30 mM MOPS (pH 7). See Table 14 for a definition of parameters.

[a] ΔC_p is the heat capacity change upon unfolding, determined by plotting ΔH_m as function of T_m, determined at different pH. See Fig. 25. The mean error in ΔC_p is ±0.03 kcal/mol/K.

RNase Sa, indicating an increase in stability. The thermodynamic parameters for wild-type RNase Sa and the three variants are given in Table 14. The values of T_m and ΔH_m for wild-type RNase Sa given here are in good agreement with previous reported values.[75] The increase in stability for the variants is shown by ΔT_m, which ranges from 2.2 ºC (2S) to 27.9 ºC (8S). This corresponds to an increase in stability from 0.8 to 8.2 kcal/mol at 65 ºC.

The two variants are much more stable than wild-type RNase Sa. We then further studied their ability to tolerate highly destabilizing mutations by making six additional mutations on the background of 7S and 8S. Their thermal denaturations were also followed by CD at pH 7. The results for 7S, 8S and these additional mutants are shown in Table 15. The mutations that changed a buried non-polar group to a charged group (I92D7S, I92D8S and I70D8S), or to a smaller non-polar group (I92A8S and Y80A8S), all resulted in significant decreases in stability. The stability of 7S, 8S and their variants were also studied as a function of pH. This allows for ΔH_m to be determined as a function of T_m, giving the heat capacity change for folding, ΔC_p, according to the Kirchoff equation :

$$d(\Delta H_m)/d(T_m) = \Delta C_p \qquad (5.1)$$

The plots of ΔH_m as a function of T_m for 7S and its variants are shown in Fig. 25a, and for 8S and its variants are shown in Fig. 25b. The resulting ΔC_p values for all proteins are given in Table 15. The ΔC_p values for 7S and 8S are 1.68 kcal/mol/K and 1.07 kcal/mol/K, respectively. The two variants of 7S both have lower ΔC_p values, while for 8S, all but one variant have higher ΔC_p values.

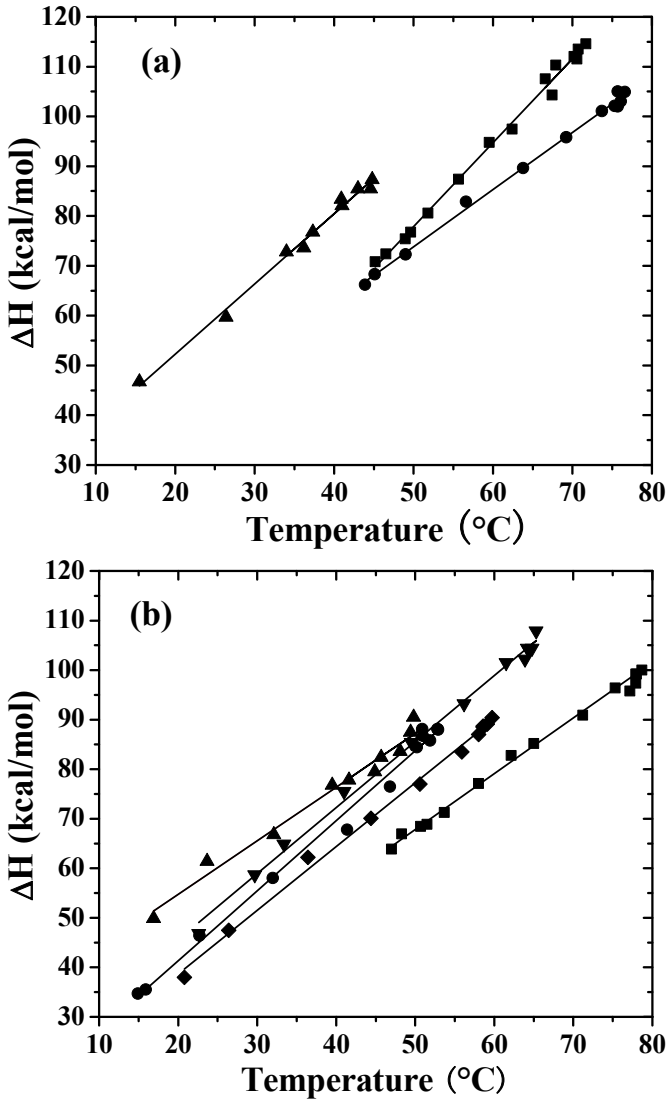

Fig. 25. Kirchoff analysis of RNase Sa variants. (a) 7S (■) and variants D79A7S (●), I92D7S (▲). (b) 8S (■) and variants I92D8S (●), I70D8S (▲), I92A8S (▼), Y80A8S (♦).

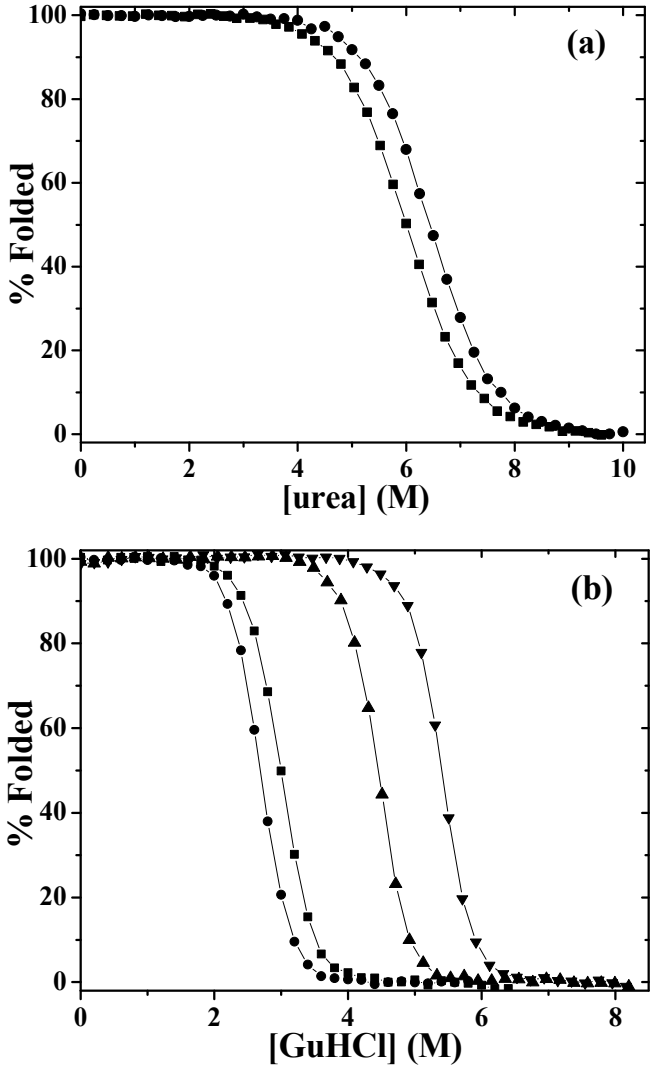

Fig. 26. Chemical denaturant unfolding curves for RNase Sa and stabilized variants. (a) urea and (b) GuHCl unfolding curves for WT (■), 2S (●), 7S (▲) and 8S (▼).

Chemical denaturation of RNase Sa and its variants at pH 7, 25ºC

The effect of mutations on the conformational stability of RNase Sa was also determined by urea and GuHCl denaturation. In urea, the 2S variant unfolded at a higher urea concentration than wild-type RNase Sa (Fig. 26a). Urea denaturations of 7S and 8S variants were attempted; however, the unfolding transitions for both proteins were incomplete, suggesting that they are too stable to be unfolded by urea. When GuHCl was used as a denaturant, both 7S and 8S unfolded at much higher GuHCl concentrations than the wild-type protein, while the 2S variant unfolded at a lower GuHCl concentration (Fig. 26b). Both urea and GuHCl curves for wild-type RNase Sa and the variants were analyzed assuming a two-state folding mechanism. From the data in the transition region, the unfolding free energy, ΔG, was calculated and found to vary linearly with chemical concentration. Therefore, the data were analyzed using the linear extrapolation method:

$$\Delta G = \Delta G\ (H_2O) - m\ [\text{denaturant}] \qquad (5.2)$$

where *m* is the measure of the dependence of ΔG on chemical concentration, and ΔG (H_2O) is the estimate of the conformational stability of the protein that assumes that the linear dependence of ΔG on chemical concentration observed in the transition region continues to 0 M denaturant. [166; 224] The parameters characterizing the chemical denaturation of wild-type RNase Sa and the 2S, 7S and 8S variants are given in Table 16. The 2S variant is 0.2 kcal/mol less stable than wild-type RNase Sa as determined by GuHCl, yet 0.8 kcal/mol more stable than wild-type RNase Sa, as determined by urea.

Table 16. Parameters characterizing chemical unfolding of RNase Sa and stabilized variants

Variant	denaturant	*m*[a] (kcal /mol/M)	$D_{1/2}$ [b] (M)	ΔG^o (H_2O) [c] (kcal/mol)	$\Delta\Delta G^o$ (kcal/mol)
WT	GuHCl	2.43	2.99±0.02	7.3±0.1	-
2S	GuHCl	2.61	2.70±0.04	7.1±0.1	-0.2
7S	GuHCl	2.52	4.44±0.01	11.2±0.1	3.9
8S	GuHCl	2.51	5.40±0.01	13.6±0.2	6.3
WT	Urea	0.96	6.06±0.05	5.8 ± 0.1	-
2S	Urea	1.02	6.51±0.11	6.6±0.2	0.8

[a] Mean error for *m* values from GuHCl denatratuion is ±0.04 kcal/mol/M and the mean error for *m* values from urea denaturation is ±0.02 kcal/mol/M.

[b] $D_{1/2}$ is the urea or GuHCl concentration at the midpoint of the unfolding transition curve.

[c] ΔG (H_2O) = m * $D_{1/2}$ The individual *m* value for each variant was used for the calculation. The mean error is 0.2 kcal/mol.

Table 17. Parameters characterizing GuHCl unfolding of RNase Sa 7S, 8S and their variants

Variant	*m*[a] (kcal/ mol/M)	$D_{1/2}$[b] (M)	ΔG^o (H_2O)[c] (kcal/mol)	$\Delta\Delta G^{od}$ (kcal/mol)
7S	2.52	4.44±0.01	11.2±0.1	-
I92D7S	2.67	0.83±0.02	2.2±0.1	-9.0
D79A7S	2.40	5.05±0.01	12.1±0.3	0.9
8S	2.51	5.40±0.01	13.6±0.2	-
I92D8S	2.73	2.01±0.05	5.5±0.2	-8.1
I70D8S	2.87	1.30±0.01	3.7±0.1	-9.9
I92A8S	2.78	3.22±0.01	9.2±0.2	-4.4
Y80A8S	2.69	2.74±0.01	7.4±0.1	-6.2

[a] Mean error for *m* values is ±0.04 kcal/mol/M.
[b] $D_{1/2}$ is GuHCl concentration at the midpoint of the unfolding transition curve.
[c] ΔG^o (H_2O) = m * $D_{1/2}$, individual m value for each variant was used for the calculation.
[d] $\Delta\Delta G^o = \Delta G^o$(variant) - ΔG^o(parent).

Both the 7S and 8S variants showed significant increases in stability as determined by GuHCl induced denaturation.

The conformational stability for the single mutants of 7S and 8S were also determined in GuHCl, since urea could not unfold them completely. These results are given in Table 17. All of the variants are less stable than their respective parent proteins, except D79A7S, which is 0.9 kcal/mol more stable than 7S.

NaCl dependence of the thermal stability of RNase Sa and the 2S variant

The conformational stability of RNase Sa and the 2S variant as determined by urea or GuHCl did not agree with each other. Since GuHCl is also a salt, this prompted us to determine the stability of these two proteins in the presence of NaCl. The results are presented in Fig. 27. Both proteins are stabilized by NaCl. However, the increase in stability for 2S in the presence of NaCl is smaller than for wild-type RNase Sa.

Discussion

Stability of 7S and 8S variants

One of the goals for protein engineering is to understand how thermophilic organisms achieve high protein stability and to use this knowledge to make mesophilic proteins more stable by rational design. However, it is usually not easy to find single changes in amino acid sequence that will increase the conformational stability of a protein by as much as 1 kcal/mol. A computer algorithm was used to design a

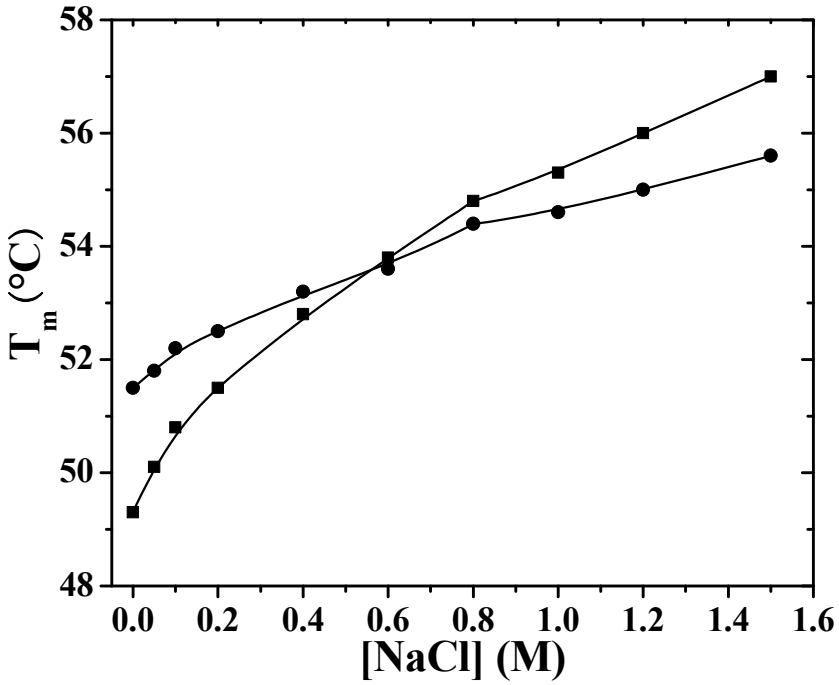

Fig. 27. Salt dependence of the melting temperature for RNase Sa variants. The variants are WT(■) and 2S (●). The lines are to guide eyes and have no meaning.

hyperstable variant of the protein Gβ1[217], and by making seven mutations they successfully increased the stability by 4.3 kcal/mol. Alternately, the combination of a series of stabilizing mutations identified by different research groups resulted in a staphylococcal nuclease variant that was 5.6 kcal/mol more stable than the wild-type protein.[117] In another approach, the stability of *Bacillus subtilis* cold shock protein was increased 5 kcal/mol by evolutionary optimization selection using phage display. [129] All of these approaches indicate that optimization of core packing is a common feature in making a protein more stable. Nevertheless, mutational effects are usually difficult to predict, especially when mutations involve buried amino acid residues.[15] To minimize this problem, 7 of the 8 stabilizing mutations introduced in this study are on the protein surface. These mutations are unlikely to introduce significant structural changes, thus their effects on stability should be more predictable than those in buried positions, and their effects are more likely to be additive. The first two mutations D25K, E74K are expected to increase the stability by 2.0 kcal/mol than that of wild-type RNase Sa, but the measured stability increase is only 0.8 kcal/mol (Table 13). This reflects the complex nature of charge-charge interactions: the effects are not additive because they are not simple pair-wise interactions. Changes in one charge-charge interaction site can affect the electrostatics in the whole protein.[179] Alternately, charge-charge interaction may also influence the denatured state.[225] The β-turn mutations, on the other hand, give better additivity of the single mutations, suggesting that any cooperative effects among these mutations are negligible. The measured increase in stability for the mutation of D79F in 8S is also smaller than expected (3.8 vs 2.6 kcal/mol), this difference is presumably

because the stabilizing effect of D79F results from both the removal of an unfavorable charge-charge interactions and the reduction of the large solvation penalty for burial of a charged amino acid.[90] The combination of the above stabilizing effects results in two highly stable RNase Sa variants: 7S and 8S. The 7S and 8S variants have ΔT_ms 17.1 and 26.9 ºC higher than that of wild-type RNase Sa. These translate to increases in stabilities of 5.6 and 8.2 kcal/mol respectively when calculated at 65 ºC, which is the average T_m of the variants. These increases in stability correspond well with the expected values from the contributions of the single mutations. The results strongly suggest that the stabilization by optimizing turns and optimizing charge-charge interactions are not influenced by each other, and thus are additive.

The higher stability of 7S and 8S confirms that RNase Sa can be made more stable. All three RNase Sa homologs (RNase Sa, Sa2 and Sa3) have a relatively low stability of 3 – 5.8 kcal/mol.[165] Our study showed that Sa2 and Sa3 can also be stabilized by more than 2 kcal/mol by same turn optimization. (see chapter IV). In fact, a low stability of 5 – 15 kcal/mol is generally observed for all proteins.[226] Such low stability may be an essential requirement so that proteins can be easily degraded when necessary[13]. Higher stability would thus hinder the degradation and in some cases the result would be deleterious. Indeed, in order to express the stable variants RNase Sa 7S and 8S in *E. coli.*, we had to remove the active site residue to inactivate the enzyme so that they are not toxic to the expression host. Nevertheless, nature may find other ways to keep a protein stable without harmful effects. For example, barnase, another ribonuclease from *Bacillus amyloliquefaciens*, has a ΔG of 9 kcal/mol, and is more

stable than most other microbial ribonucleases.[165] In this case, the bacteria evolved an inhibitor protein, barstar, that specifically binds barnase at sub nano- molar concentrations, keeping this protein from being active and thus too toxic to the host.[227] In line with this, all attempts to express barnase in *E. coli* without the co-expression of barstar failed presumably because of its high toxicity. [228]

Stabilities as determined by urea or GuHCl denaturation

The conformational stabilities of the 7S and 8S variants as determined by thermal or solvent denaturation are all significantly higher than that of wild-type, but their values vary. Protein stability determined using urea or GuHCl usually agree well with each other, supporting the general validity of the linear extrapolation method.[15] However, in rare cases, deviation from linearity have been observed with GuHCl.[229; 230] Makhatadze (1999)[231] has attributed this curvature to the electrolytic nature of GuHCl. The overall effect of GuHCl on protein stability is a combination of the denaturing effects originated from the guanidinium ion (Gu^+) and the salt effect contributed by the chloride ion (Cl^-). Therefore, the effect of GuHCl on protein stability may be similar to that of the mixture of urea and NaCl, as suggested by Smith and Scholtz (1996).[232]

The presence of salt can have varied effects on protein stability. These include: (1) Debye-Hückel screening of electrostatic interactions, (2) salting-in and salting-out effects and (3) the preferential binding of ions to either the native or the denatured state.[233; 234] We have previously shown that the conformational stability of five microbial ribonucleases including RNase Sa can be increased by NaCl.[165] This increase in stability was attributed to the preferential binding of cations and/or anions to the native state of

the protein. Therefore, the higher conformational stability of the RNase Sa variants determined by GuHCl denaturation are most likely overestimated when compared to urea and thermal denaturation results, since chloride ion stabilizes these proteins. [225] In addition, the stability dependence of wild-type RNase Sa on NaCl is bigger than that of the 2S variant. In the presence of higher concentration of NaCl, this leads to higher melting temperatures of wild-type RNase Sa in comparison with 2S variant (Fig. 27). Consequently, the ΔG value of wild-type RNase Sa determined by GuHCl is higher than that of the 2S variant, while the opposite was observed when urea was used (Table 16).

pH dependence of stabilities of 7S and 8S variants

The far-UV CD spectra of the RNase Sa 7S and 8S variant are very similar from pH 1.65 to pH 10, indicating that the protein structures remain unchanged over this pH range. The T_m and ΔH_m values were measured from pH 1.65 to pH 10 for both RNase Sa 7S and 8S. These values were used with the ΔC_p values in Table 3 to calculate the ΔG values at 25 ºC as a function of pH using the modified Gibbs-Helmholtz equation:

$$\Delta G\,(T) = \Delta H_m\,(1-T/T_m) + \Delta C_p\,[(T-T_m) - T\ln(T/T_m)] \qquad (5.3)$$

Results of the pH dependence of T_m and ΔG for both variants are shown in Fig. 28. Both variants reach their maximum stability near pH 5, as observed in all microbial ribonucleases. At their pH of maximum stability, the ΔG (25 ºC) of 7S and 8S are 9.9 kcal/mol and 10.7 kcal/mol respectively, both larger than the 7.0 kcal/mol measured for wild-type RNase Sa. Over the entire pH range, the stability of 8S is higher than that of 7S, and both variants have significantly higher stability than previously reported wild-type values.[165]

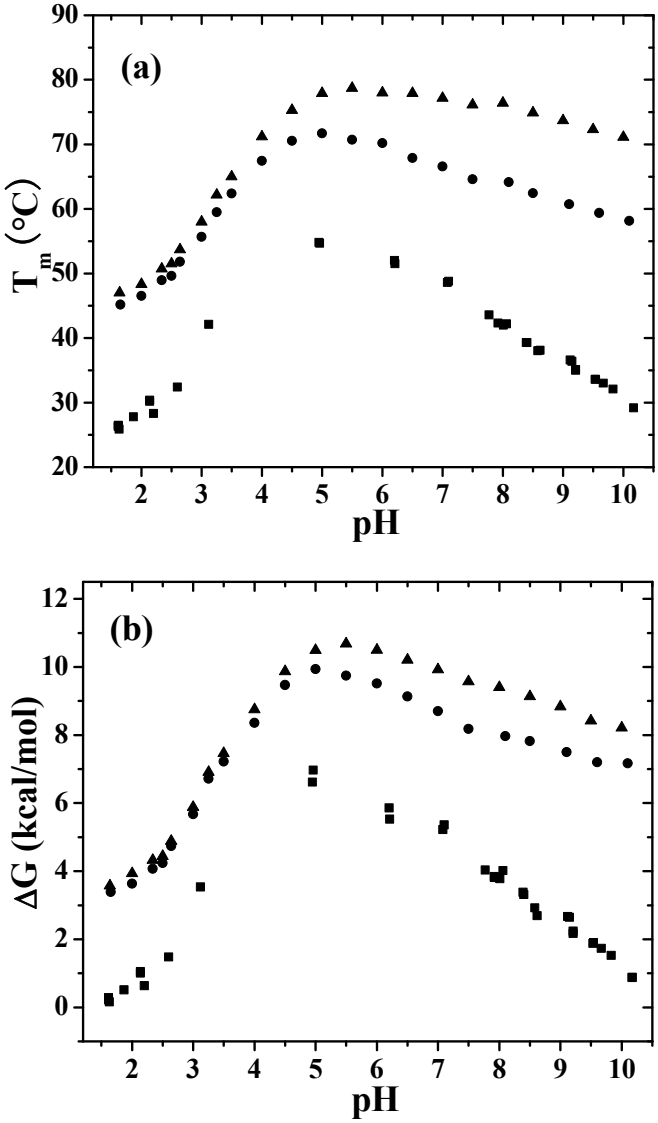

Fig. 28. pH dependence of protein stability for WT, 7S and 8S RNase Sa. Depencence of T_m (a) and (b) ΔG (25ºC) for WT (■), 7S (●) and 8S (▲) variants. ΔG (25ºC) for 7S and 8S were calculated from the T_m values using Eq. 5.3. WT data were from Pace et al.[165]

The pH perturbations of protein stability are the consequence of the ionizable groups in a protein having different pK values for the denatured states or native state. The stabilities of 7S and 8S decrease markedly when the pH decreases, similar to that observed for the wild-type RNase Sa.[165] This is because most of the carboxyl groups of the RNase Sa have lower pK values in the native state than in the denatured state.[235] As a result, the denatured state binds protons more tightly than the native state, thus the equilibrium will be shifted towards the denatured state when the proton concentration increases. The dependence of ΔG (25 ºC) on pH for 7S and 8S are very similar to that of wild-type RNase Sa at low pH. An analysis of such dependence gives a similar number of protons taken up upon unfolding (≈3). At higher pH values, the decrease in stabilities results from some of the charged groups having higher pK values in the native state than in the denatured state. For wild-type RNase Sa, there are five such groups: Asp 25, Asp 79, Glu 54, His 53 and the N terminus.[235] Three of these groups have pKs that are more than 1.5 units higher than the pKs measured in the denatured state, accounting for the stability decrease at the higher pH range in wild-type RNase Sa.[165] For 7S and 8S, the stability dependence on pH between pH 5 and pH 10 is relatively smaller (Fig. 28). This can be partly attributed to the elimination of Asp 25 (pK = 4.87 in wild-type RNase Sa[235]). More importantly, the two charge reversals may reduce the native state pK values of the other four that are elevated in the wild-type protein, especially the His 53. In both RNase T1[236] and barnase[237], the pH dependence between pH 5 and pH 9 has been shown to depend mainly on the different pK values of histidine side chains in native and denatured state. We have previously designed a charge reversal variant of RNase Sa, in

which five Asp or Glu of wild-type RNase Sa have been replaced with Lys (5K). The native state pK values of all 5K charged group side chains were determined, and they are all lower than those measured in wild-type RNase Sa. [222] The two charge reversal mutations, D25K and E74K in 7S and 8S may have a similar effect on the side chain pK values, thus make the stabilities of the two variants less sensitive to pH at higher pH range.

ΔC_p of unfolding for 7S and 8S variants

The traditional way to determine ΔC_p is to use differential scanning calorimetry. The ΔH_{cal} and T_m values are determined in the presence of a perturbant such as urea, GuHCl, salt, or a variation in pH , then fit linearly using Eq. 5.3 to obtain ΔC_p.[238] Usually the pH is varied to change T_m, and the measurements are generally restricted to the acidic region, using carboxyl buffers to minimize the contribution of proton ionization to the observed ΔH_{cal}. In most cases, ΔH_{cal} is found to vary linearly with T_m . However, many proteins bind anions at low pH and this probably contributes to the observed ΔC_p value and makes such measurements unpredictable.[239; 240] For example, in the case of RNase A, the ΔC_p value reported in the literature vary from 1.74 to 3.25 kcal/mol/K depending on the buffer used, [241] these values are larger than the generally accepted value of 1.15 kcal/mol/K[242]. In some cases, ΔC_p can be determined as described above from thermal denaturation curves.

We have previously determined the ΔC_p value for wild-type RNase Sa and other microbial ribonucleases by measuring the ΔH_m by a van't Hoff analysis of thermal denaturation curves and fitting it with the measured T_m using Eq. 5.1.[165] This method

provided a good estimate of ΔC_p for wild-type RNase Sa (1.52 kcal/mol/K). The calculated ΔG (25ºC) at pH 7 using this value for ΔC_p in Eq. 5.3 is in good agreement with that determined by urea denaturation (Table 14,16). In this report we used the same approach to estimate the ΔC_p values for the 7S and 8S variants as well as the variants based on them. These results are given in Table 15. This analysis gave a ΔC_p value of 1.68 kcal/mol/K for 7S and 1.07 kcal/mol/K for 8S, a difference of 36%. Such a change in ΔC_p values caused by a single mutation is rare. One example is the substitution of a buried polar residue (Gln41) to a nonpolar residue (Leu or Val) in ubiquitin, that increased ΔC_p by 25%. [178] However, the magnitude of the change of our observation is even larger and, most interestingly, the sign of change is the opposite to that observed in ubiquitin.

Pace and Laurents proposed another approach for estimating ΔC_p values using both thermal and chemical denaturation data.[174] In this method, the ΔG values determined at lower temperatures by chemical denaturation are constrained by the ΔG values determined in the transition region of thermal denaturation and globally fit with Eq. 5.3 to obtain the ΔC_p. This is now very common used method.[243-247] Because the 7S and 8S variants are too stable to characterize at pH 7 using urea, a series of urea denaturation curves for 7S and 8S were determined at pH 3 at temperatures ranging from 10 ºC to 31 ºC. A global fit of these urea data constrained by the T_m and ΔH_m values determined by thermal denaturation at pH 3 yields ΔC_p of 1.47 and 1.01 kcal/mol/K for

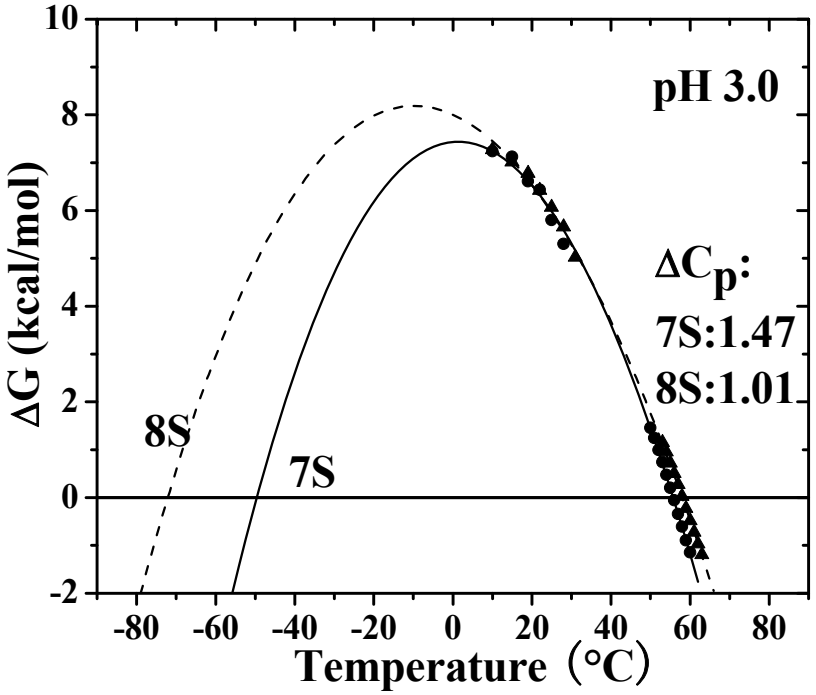

Fig. 29. Stability curves for 7S and 8S variants at pH 3. The data at low temperatures are from results of a series of urea denaturation experiments and the data at higher temperatures are from thermal denaturation experiments without urea. Curves are fit with Eq. 5.3 for 7S (●) and 8S (▲) to get the respective ΔC_P values.

7S and 8S respectively (Fig. 29). The good agreement in magnitude and shape of ΔC_p values between both sets of data indicates that both methods are good for determining the ΔC_p values for RNase Sa variants studied here. For another studied protein ribosomal protein L30E, a similar agreement was also observed, but in that case the difference in ΔC_p values from both methods was larger than what we observed here.[246]

Evidence for residual structure in 8S variant

The data above shows that there is a dramatic difference between the ΔC_p values for RNase Sa variants 7S and 8S. It is well known that ΔC_p is correlated with the change in accessible surface area upon unfolding (ΔASA). Several equations have been developed to predict ΔC_p from ΔASA.[175-177] For example, Myers et al.[177] showed that ΔC_p = -119 + 0.20*(ΔASA) with an r value of 0.98. However, equations of such type cannot account for the big difference of 8S ΔCp from that of 7S and wild-type. The native state ASA is unlikely to change with a single mutation, therefore, the big decrease in ΔC_p suggests that the denatured state of 8S may have a smaller ASA than 7S. This implies that the denatured state of 8S may contain residual structure, making it more compact. Although the denatured state of proteins was first proposed to behave as a random coil[2], the existence of residual structures in denatured state is now clear.[3; 4; 6; 10; 11] An analysis of *m*-values and ΔC_p values of another microbial ribonuclease RNase T1 and its D76N variant suggested that a compact denatured state may exist in this protein.[76] In some thermophilic proteins, residual structure has been suggested in the denatured state, accounting for the reduced ΔC_p and higher stability. [247-249]

Residual structure in the denatured state has been suggested by both experiments and theory.[4; 6; 149; 243; 247; 250; 251] However, direct characterization of structure in denatured state is difficult. Global parameters such as *m*-values or radius of hydration are generally not useful in detecting the nature of residual structures. In one case, an indirect mutational strategy was used to probe for residual structure in the denatured state of *T.thermophilus* RNase H. [247] The premise here was that the introduction of a mutation that disrupts residual structure should increase ΔC_p because such disruption will increase the ASA of the denatured state. The strategy was to replace a buried hydrophohic residue with a charged reside. When not involved in a residual structure, such a substitution has been previously shown to decrease ΔC_p.[178; 252] The substitution of a larger buried hydrophobic residue to a smaller hydrophobic residue should have little effects on ΔC_p if the substitution is not in the residual structure. However, if these mutations disrupt the proposed residuals structure in denatured state, the ΔC_p in the variant will be larger.

Although the main purpose of this strategy is to probe for the existence of residual structure in denatured state, we can also learn the origin of of the denatured state. Modeling analysis revealed that the side chain of Ile 92 is very close to the Phe 79 side chain (<4 Å) in modeled 8S structure (Fig. 30). Therefore, an I92D mutation in both 7S and 8S should lead to different response in ΔC_p values if the denatured states of both variants are different. As a control, the proposed mutation on the Ile 70 residue that's not in direct contact with residue 79 should not increase the ΔC_p value. In addition, mutations on other neighboring residues may also help reveal the roles of these residues

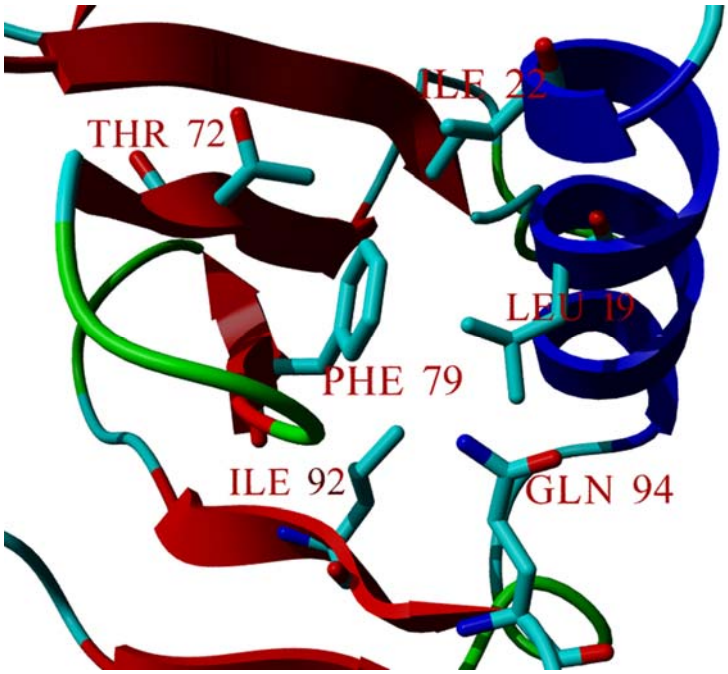

Fig. 30. Environment of the Phe 79 in the modeled 8S variant. The mutations were introduced in program Swiss PDB viewer[154] and the resulting structure was energy minimized in Insight II. Residues that are within 4 Å of Phe 79 side chain are shown. The final figure was generated using program YASARA[106].

in the denatured state. Therefore, the following mutations were designed: I92D, I70D, I92A, Y80A on 8S and I92D, D79A on 7S which were all made within 5 Å of residue 79. As expected, all mutations lead to significant change in stabilities (Table 16, 17). D79A7S is the only variant that is more stable than its parent protein and its T_m is very similar to that of 8S. The ΔC_p values were obtained by Kirchoff analysis on thermal denaturation data of these variants and they are shown in Table 15.

The I92D mutation in 8S resulted in an increase in ΔC_p while the same mutation in 7S led to a decrease in ΔC_p. Furthermore, the I92D7S and I92D8S variants have nearly the same ΔC_p values. In contrast, the control mutation has no effect to the ΔC_p value of 8S. These data suggest that I92D mutation on 8S may disrupt some residual structure present in the denatured state. These data suggest that there is less detectable residual structure in the 7S variant. Consequently, that both mutations have similar denatured states in I92D7S and I92D8S is supported by their similar ΔC_p values.

There is evidence suggesting that residual structure often includes aromatic amino acid residues.[5; 8] For 8S, one wonders if any residual structure studied is promoted by the introduced Phe mutation , or by the removal of the unfavorable charged amino acid, Asp 79. The D79A mutation was introduced in 7S to answer this question. The resulting D79A7S variant has a ΔC_p very similar to that of 8S, suggesting that residual structure may also exists in the denatured state of this variant. This indicates that the residual structure is more likely caused by the removal of the unfavorable charge-charge interactions with neighboring residues, and not the introduction of an aromatic group. Therefore, this residual structure could consist of amino acid residues close to residue

79, including a neighboring aromatic cluster and the Ile 92 residue. The preliminary study of both Y80A and I92A mutations in 8S showed a smaller increase in ΔC_p as well. These results suggest that these two residues may also be involved in any residual structure of the denatured state of 8S. However, the detailed nature of this residual structure still requires further investigation.

Taken together, these studies argue that the 8S variant forms residual structure in its denatured state, lowering its ΔC_p. Furthermore, this residual structure may be induced by the removal of unfavorable charge-charge interaction.

Interpretation of the stabilizing mechanisms for 7S and 8S variants

The modified Gibbs-Helmholtz equation can be used to define a protein's stability curve. From these stability curves, two important parameters can be derived: ΔG_S (the maximum stability) and T_S, the temperature of maximum stability (ΔS = zero); the latter is decribed as: [253]

$$T_S = T_m * \exp\left[\Delta H/(T_m * \Delta C_p)\right] \quad (5.4)$$

Fig. 31a shows three ways that proteins can increase their stability:[254] (I) Raising the stability curve to increase the overall stability of unfolding at all temperatures. This can be achieved by increasing the ΔH_S without the change in the ΔS. (II) Lowering the ΔC_p for unfolding, resulting in a broad stability curve and moving the T_m to higher values. (III) Shifting the stability curve to higher temperature through a reduction of ΔS. All three methods, either independently or in combination, have been observed in nature for the pairs of mesophilic and thermophilic proteins. Razvi and Scholtz [255] have

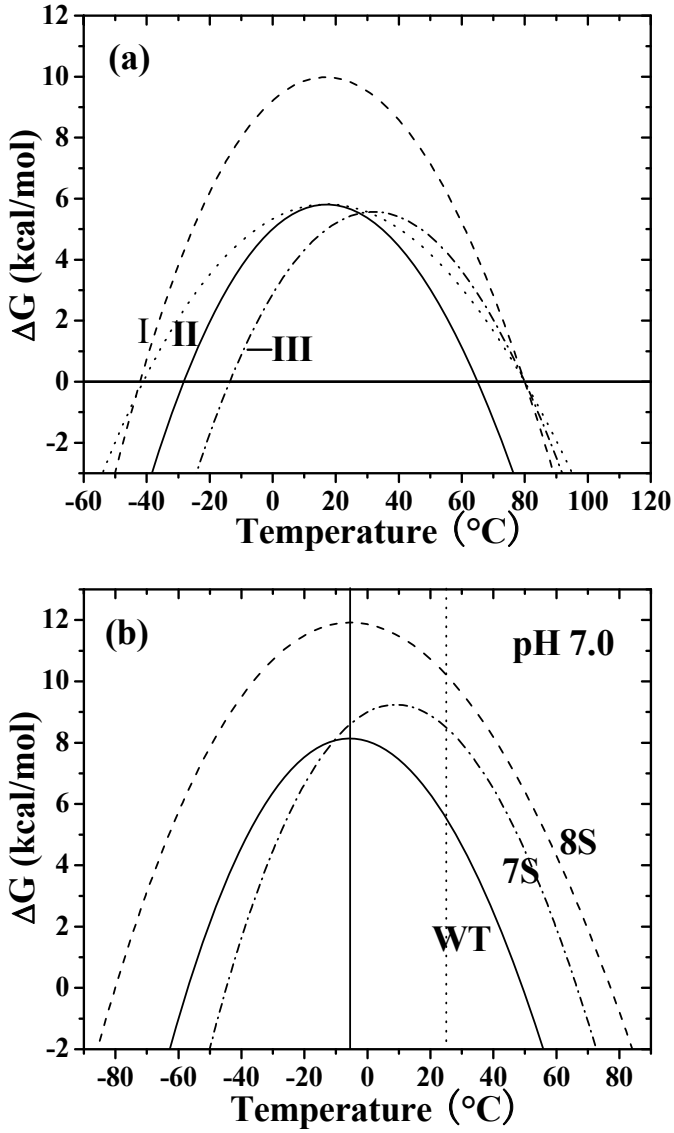

Fig. 31. Stability curves showing different methods for increasing protein stability. (a) Three thermostabilization mechanisms for a protein: reference (solid line), method I (dashed line), method II (dotted line), method III (dash-dotted line). See text for detailed discussion. (b) Stability curves for WT (solid line), 7S (dash-dotted line) and 8S (dashed line) RNase Sa at pH 7 constructed using Eq. 5.3 with parameters from Table 14. The two vertical lines represent -6.2 ºC (solid line, the T_S of WT) and 25 ºC (dotted line, room temperature).

recently conducted a literature survey on 26 pairs of mesophilic and thermophilic proteins. Their results suggest that Method I is the most common way that proteins use to increase their stability. More than 70 percent of the thermophilic proteins in their study used this method. Method II, the second most used way, is to lower the ΔC_p, making the stability curve flatter. The least frequently observed way is method III, where the thermophilic proteins exhibit higher T_S.

Method I is the most common method because it is relatively easy for a protein to improve its ΔG by optimization of various forces like hydrogen bonding, salt bridges or increasing hydrophobic interactions. These can be accomplished with only a single amino acid change. To use Method II there are also many options to change protein sequence thus optimizing charge-charge interactions [246] or promoting the formation of residual structure in the denatured state.[247; 256] The stability curves for wild-type RNase Sa and the 7S, 8S variants at pH 7 were determined using the parameters in Table 14 and shown in Fig. 31b. Clearly, the 8S variant in this study uses a combination of both Method I and Method II. As shown in Table 18 and Figure 31b, the 8S variant has a lower ΔC_p and a higher ΔG_S in comparison with the wild-type protein. However, the T_S of 8S is very similar to that of wild-type RNase Sa. Our mutational study strongly suggests that this smaller ΔC_p is caused by residual structure in the denatured state of 8S. One may also argue that the formation of residual structure which leads to a more compact denatured state may preserve some native interaction.[248] Consequently, this will lower the free energy of the denatured state and lead to the destabilization of the protein

Table 18. Parameters characterizing the stability of RNase Sa, 7S and 8S

Variant	T_m [a](ºC)	$\Delta G(25ºC)$ [a] (kcal/mol)	T_S [b](ºC)	ΔG_S [c] (kcal/mol)	ΔC_p [d] (kcal /mol/K)	ΔC_p [e] (kcal /mol/K)
WT	49.3	5.5	-6.2	8.2	1.52[f]	-
7S	66.6	8.4	9.5	9.1	1.68	1.47
8S	77.1	10.1	-4.4	11.7	1.07	1.01

All values were determined at pH 7.0 except the ΔC_p values for 7S and 8S in last column, which were determined at pH 3.

[a] Same as in Table 14.

[b] T_S, Temperature of maximum stability, calculated using Eq. 5.4 with ΔC_p values determined by Kirchoff analysis at pH 7 (column 6).

[c] ΔG_S, stability at T_S at pH 7, calculated using Eq. 5.3.

[d] ΔC_p, heat capacity change upon unfolding, determined by Kirchoff analysis of thermal denaturation data at pH 7.

[e] Determined by global fit of urea and thermal denaturation data at pH 3, errors are ±0.02 and ±0.01 kcal/mol/K for 7S and 8S respectively.

[f] From Pace et al. (1998) [165].

at all temperatures. However, this analysis seems to neglect the higher entropy and the larger contribution of repulsive charge-charge interactions in a more compact denatured state.[149; 225] Both factors are capable of increasing the free energy of denatured state.

Method III, on the other hand, is very rare.[219; 255] Thermal stabilization using this method might require some unique changes in sequence, because a higher T_S is achieved by reducing the ΔS of unfolding in this scenario. The reduction in ΔS can be achieved by either reduced entropy in the denatured state or enhanced entropy in native state. Therefore it will require specific changes in the sequence, to effect the entropy of these two states in different ways. For example, the Pros and Glys introduced in this study should reduce the ΔS by either constraining the loop in the denatured state or releasing the space hindrance of the turn in the native state. Moreover, the Pros constrain the denatured state and may lead to more rigid denatured state. This type of unique changes in sequence may be the reason why Method III is not frequently observed. Clearly, the 7S variant in this study used this rare method to gain the higher thermal stability relative to wild-type RNase Sa. The T_S of 7S is 15 ºC higher than that of wild-type RNase Sa, while the ΔG_S value is similar to that of wild-type RNase Sa (Table 18). To our knowledge, our result is the first experimental example for modulating T_S to increase protein thermal stability by rational entropic optimization.

Conclusion

We designed two RNase Sa variants, 7S and 8S, by changing 7 or 8 amino acid residues. These two variants appear to have three-dimensional structures similar to the

wild-type protein. Both variants have higher stability than that of wild-type RNase Sa. However, the two variants use different mechanisms to obtain their thermal stabilities. The 7S variant achieves stability by decreasing the entropy in such a way that the increase in the global stability is only moderate, and thus 7S is suggested to have a more rigid denatured state because of the proline mutations introduced. In contrast, the 8S variant achieves its higher stability by increasing the global stability as well as lowering the ΔC_p, most likely by increasing residual structure in the denatured state. This is interesting because the two variants differ in only one buried amino acid residue. This pair of proteins will be useful in determining the role of denatured state in defining protein stability.

CHAPTER VI

SUMMARY

The conformational stabilities of 11 proteins and their 46 variants have been determined to further our understanding of the non-covalent forces that contribute to protein stability and use current knowledge of these forces to design more stable proteins. Two of these proteins, the small 36 residue VHP, and the large 337-residue VlsE, were used as model systems for studying the contribution of the hydrophobic effect to protein stability. Specifically, mutations of large hydrophobic residues to smaller ones were made in these two proteins. They were Leu to Ala, Val to Ala, and Met to Ala in VHP, and Ile to Val in VlsE. The stabilities of the VHP variants, as determined by thermal and chemical denaturation, were in good agreement. The stabilities of VlsE variants were determined by urea denaturation. All variants for both proteins had lower stabilities than the wild type proteins. A previous study of four proteins has shown that the average $\Delta\Delta G$ values are -1.3, -2.5, -3.5 and -3.8 kcal/mol for Ile to Val, Val to Ala, Leu to Ala and Phe to Ala mutations, respectively. However, the average $\Delta\Delta G$ values for each type of mutations (Leu/Val/Phe to Ala) of the VHP variants are less than those values. In addition, the average loss in stability for removing a single $-CH_2-$ for the VHP variants is 0.6 kcal/mol, which is also smaller than the reported value of 1.3 kcal/mol. On the other hand, the average loss in stability for removing a single $-CH_2-$ for the VlsE variants is 1.6 kcal/mol, which is slightly larger than the reported value. Several structural energy correlations were attempted. Such

analysis indicates that the disagreement between the measured $\Delta\Delta G$ values here and those generally observed can be reconciled by the difference in hydrophobicity and modeled cavity, as previously reported for T4 lysozyme variants.[39] The results here suggest that the stabilizing hydrophobic effect for both VHP and VlsE can be described using the same mechanism. These studies indicate that the VHP and VlsE are not different from the 'typical' proteins usually studied, which are typically 100-150 amino acid residues in length.

One strategy for increasing protein stability is to target residues in β-turns, where non-proline and non-glycine residues at certain positions are substituted to the statistically favored proline and glycine.[140] We further investigated this strategy using eight proteins ranging from 66 to 370 amino acid residues in length. We made 21 variants (14 proline and 7 glycine) and measured their stabilities. Of the 14 proline variants, 10 were more stable than the wild-type proteins, with increases in stability ranging from 0.3 to 1.2 kcal/mol. Of the glycine variants, only one was more stable than the wild-type protein. For all the residues that were substituted to proline, the ϕ angles of wild-type residues are between -90º and -60º, however, only the substitutions of those residues that are highly solvent exposed gave a significant stability increase. In contrast, the mutation to glycine appears to only have a stabilizing effect when the wild-type residue falls in the L-α region on the Ramachandra plot ($0º<\phi<90º$, $0º<\psi<90º$). The molecular basis for the stabilizing effects of substitutions of proline or glycine is to reduce the conformational entropy of unfolding by reducing the entropy of the denatured

state, or by removing unfavorable van der Waals repulsion. Our results suggest that this strategy could provide a simple and efficient way for increasing protein stability.

Finally, we combined eight previously indentified stabilizing mutations for RNase Sa and designed two 'thermophilic' RNase Sa variants ,7S and 8S. We determined the thermodynamic parameters of these two variants and other variants based on them. At pH 7, the melting temperatures, T_m, are 17.3 (7S) and 27.9 ºC (8S) higher than wild-type RNase Sa. The increase in stability for both mutants is observed over a wide range of pH. When the stabilities of 7S and 8S are determined by GuHCl, they are 3.9 (7S) and 6.3 (8S) kcal/mol more stable than wild-type. The change in heat capacity, ΔC_p, of both variants was determined by Kirchoff analysis of thermal denaturation data and a global fit of the urea and thermal denaturation data. The results from the two methods are in good agreement for both variants. The 7S variant has a ΔC_p of 1.68 kcal/mol/K, similar to that of wild-type RNase Sa, but ΔC_p for the 8S variant is significantly lower (1.07 kcal/mol/K). An analysis of the stability curve reveals that the stabilization of the 7S variant is achieved by increasing the temperature of maximum stability (T_S), shifting the stability curve to higher temperature with only a small change in the maximum stability, ΔG_S. In contrast, the 8S variant achieves its thermostabilization through a larger ΔG_S as well as a reduced ΔC_p ,which shifts the stability curve higher and making it flatter. Furthermore, a mutagenesis study suggests that there is residual structure in the denatured state of 8S which contributes to the lower ΔC_p. Interestingly, the 7S and 8S variants differ by only one buried amino acid. This

single amino acid change is capable of changing the stabilizing mechanism, perhaps by modulating the denatured state.

REFERENCES

1. Anfinsen, C. (1973). Principles that govern the folding of protein chains. *Science.* **181**, 223-30.
2. Tanford, C. (1968). Protein denaturation. *Adv. Protein Chem.* **23**, 121-282.
3. Shirley, B. A., Stanssens, P., Steyaert, J. & Pace, C. N. (1989). Conformational stability and activity of ribonuclease T1 and mutants. Gln25----Lys, Glu58----Ala, and the double mutant. *J. Biol. Chem.* **264**, 11621-11625.
4. Stigter, D., Alonso, D. O. & Dill, K. A. (1991). Protein stability: electrostatics and compact denatured states. *Proc. Natl Acad. Sci. USA* **88**, 4176-4180.
5. Neri, D., Billeter, M., Wider, G. & Wuthrich, K. (1992). NMR retermination of residual structure in a urea-denatured protein, the 434-Repressor. *Science* **257**, 1559-1563.
6. Shortle, D. (1996). The denatured state (the other half of the folding equation) and its role in protein stability. *FASEB J.* **10**, 27-34.
7. Nolting, B., Golbik, R., Soler-Gonzalez, A. S. & Fersht, A. R. (1997). Circular dichroism of denatured barstar suggests residual structure. *Biochemistry* **36**, 9899-9905.
8. Klein-Seetharaman, J., Oikawa, M., Grimshaw, S. B., Wirmer, J., Duchardt, E., Ueda, T., Imoto, T., Smith, L. J., Dobson, C. M. & Schwalbe, H. (2002). Long-range interactions within a nonnative protein. *Science* **295**, 1719-1722.

9. Mayor, U., Grossmann, J. G., Foster, N. W., Freund, S. M. & Fersht, A. R. (2003). The denatured state of Engrailed Homeodomain under denaturing and native conditions. *J. Mol. Biol.* **333**, 977-91.

10. Fitzkee, N. C. & Rose, G. D. (2004). Reassessing random-coil statistics in unfolded proteins. *Proc. Natl Acad. Sci. USA* **101**, 12497-12502.

11. McCarney, E. R., Kohn, J. E. & Plaxco, K. W. (2005). Is there or isn't there? The case for (and against) residual structure in chemically denatured proteins. *Crit. Rev. Biochem. Mol. Biol.* **40**, 181-189.

12. Pace, C. N. (1975). The stability of globular proteins. *CRC Crit Rev Biochem.* **3**, 1-43.

13. 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.

14. Shoichet, B. K., Baase, W. A., Kuroki, R. & Matthews, B. W. (1995). A relationship between protein stability and protein function. *Proc. Natl Acad. Sci. USA* **92**, 452-456.

15. Pace, C. N. (1990). Measuring and increasing protein stability. *Trends Biotechnol.* **8**, 93-98.

16. Wu, H. (1931). Studies on denaturation of proteins XIII. A theory of denaturation. *Chin J Physiol.* **5**, 321-344.

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

18. Burley, S. K. & Petsko, G. A. (1988). Weakly polar interactions in proteins. *Adv. Protein Chem.* **39**, 125-189.

19. Levitt, M. (1974). Energy refinement of hen egg-white lysozyme. *J. Mol. Biol.* **82**, 393-420.

20. Creighton, T. E. (1993). *Proteins: structures and molecular properties*. 2nd ed., W. H Freeman and Company, New York.

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

22. Chothia, C. (1975). Structural invariants in protein folding. *Nature* **254**, 304-308.

23. Sandberg, W. S. & Terwilliger, T. C. (1991). Repacking protein interiors. *Trends Biotechnol.* **9**, 59-63.

24. Willis, M. A., Bishop, B., Regan, L. & Brunger, A. T. (2000). Dramatic structural and thermodynamic consequences of repacking a protein's hydrophobic core. *Structure* **8**, 1319-1328.

25. Richards, F. M. (1997). Protein stability: still an unsolved problem. *Cell Mol Life Sci.* **53**, 790-802.

26. Dill, K. A. (1990). Dominant forces in protein folding. *Biochemistry* **29**, 7133-7155.

27. Tanford, C. (1962). Contribution of hydrophobic interactions to the stability of the globular conformation of proteins. *J. Am. Chem. Soc.* **84**, 4240-4247.

28. Chothia, C. (1976). The nature of the accessible and buried surfaces in proteins. *J. Mol. Biol.* **105**, 1-14.

29. Rose, G. D., Geselowitz, A. R., Lesser, G. J., Lee, R. H. & Zehfus, M. H. (1985). Hydrophobicity of amino acid residues in globular proteins. *Science* **229**, 834-8.

30. Nozaki, Y. & Tanford, C. (1971). The solubility of amino acids and two glycine peptides in aqueous ethanol and dioxane solutions. establishment of a hydrophobicity scale. *J. Biol. Chem.* **246**, 2211-2217.

31. Radzicka, A. & Wolfenden, R. (1988). Comparing the polarites of the amino acids; side-chain distribution coefficients between the vapor phase, cyclohexane, 1-octanol, and neutral squeous solution. *Biochemistry* **27**, 1644-1670.

32. Fauchere, J. L. & Pliska, V. E. (1983). Hydrophobic parameters of amino-acid side-chains from the partitioning of N-acetyl-amino-acid amides. *Eur. J. Med. Chem.* **18**, 369-375.

33. Kyte, J. & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105-132.

34. Yutani, K., Ogasahara, K., Tsujita, T. & Sugino, Y. (1987). Dependence of conformational stability on hydrophobicity of the amino acid residue in a series of variant proteins aubstituted at a unique position of tryptophan synthase alpha subunit. *Proc. Natl Acad. Sci. USA* **84**, 4441-4444.

35. Kellis, J. T., Nyberg, K., Sail, D. & Fersht, A. R. (1988). Contribution of hydrophobic interactions to protein stability. *Nature* **333**, 784-786.

36. Kellis, J. T., Nyberg, K. & Fersht, A. R. (1989). Energetics of complementary side chain packing in a protein hydrophobic core. *Biochemistry* **28**, 4914-4922.

37. Shortle, D., Stites, W. E. & Meeker, A. K. (1990). Contributions of the large hydrophobic amino acids to the stability of staphylococcal nuclease. *Biochemistry* **29**, 8033-8041.

38. Lim, W. A. & Sauer, R. T. (1991). The role of internal packing interactions in determining the structure and stability of a protein. *J. Mol. Biol.* **219**, 359-376.

39. 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.

40. 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.

41. Takano, K., Ogasahara, K., Kaneda, H., Yamagata, Y., Fujii, S., Kanaya, E., Kikuchi, M., Oobatake, M. & Yutani, K. (1995). Contribution of hydrophobic residues to the stability of human lysozyme: calorimetric studies and X-ray structural analysis of the five isoleucine to valine mutants. *J. Mol. Biol.* **254**, 62-76.

42. Takano, K., Yamagata, Y., Fujii, S. & Yutani, K. (1997). Contribution of the hydrophobic effect to the stability of human lysozyme: calorimetric studies and X-ray structural analyses of the nine valine to alanine mutants. *Biochemistry* **36**, 688-698.

43. Lazar, G. A. & Handel, T. M. (1998). Hydrophobic core packing and protein design. *Curr. Opin. Chem. Biol.* **2**, 675-679.

44. Pace, C. N. (1992). Contribution of the hydrophobic effect to globular protein stability. *J. Mol. Biol.* **226**, 29-35.

45. Zhou, H. Z., Yaoqi (2002). Stability scale and atomic solvation parameters extracted from 1023 mutation experiments. *Proteins: Struct. Funct. Genet.* **49**, 483-492.

46. 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.

47. Serrano, L., Kellis, J. T., Cann, P., Matouschek, A. & Fersht, A. R. (1992). The folding of an enzyme : II. Substructure of barnase and the contribution of different interactions to protein stability. *J. Mol. Biol.* **224**, 783-804.

48. Hubbard, S. J. & Argos, P. (1995). Evidence on close packing and cavities in proteins. *Curr. Opin. Biotechnol.* **6**, 375-381.

49. Vlassi, M., Cesareni, G. & Kokkinidis, M. (1999). A correlation between the loss of hydrophobic core packing interactions and protein stability. *J. Mol. Biol.* **285**, 817-827.

50. Ratnaparkhi, G. S. & Varadarajan, R. (2000). Thermodynamic and structural studies of cavity formation in proteins suggest that loss of packing interactions rather than the hydrophobic effect dominates the observed energetics. *Biochemistry* **39**, 12365-12374.

51. Chen, J. & Stites, W. E. (2001). Packing is a key selection factor in the evolution of protein hydrophobic cores. *Biochemistry* **40**, 15280-15289.

52. Loladze, V. V., Ermolenko, D. N. & Makhatadze, G. I. (2002). Thermodynamic consequences of burial of polar and non-polar amino acid residues in the protein interior. *J. Mol. Biol.* **320**, 343-357.

53. Bueno, M., Campos, L. A., Estrada, J. & Sancho, J. (2006). Energetics of aliphatic deletions in protein cores. *Protein Sci.* **15**, 1858-1872.

54. Xu, J., Baase, W. A., Baldwin, E. & Matthews, B. W. (1998). The response of T4 lysozyme to large-to-small substitutions within the core and its relation to the hydrophobic effect. *Protein Sci.* **7**, 158-177.

55. Buckle, A. M., Cramer, P. & Fersht, A. R. (1996). Structural and energetic responses to cavity-creating mutations in hydrophobic cores: observation of a buried water molecule and the hydrophilic nature of such hydrophobic cavities. *Biochemistry* **35**, 4298-4305.

56. Pauling, L. & Corey, R. B. (1951). The pleated sheet, a new layer configuration of polypeptide chains. *Proc. Natl Acad. Sci. USA* **37**, 251-256.

57. Pauling, L., Corey, R. B. & Branson, H. R. (1951). The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain. *Proc. Natl Acad. Sci. USA* **37**, 205-211.

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

59. Kondratova, M. S. & Efimov, A. V. (2002). A systematic analysis of buried polar side chains and their interactions in α-helical proteins. *Mol Biol (Mosk).* **V36**, 117-123.

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

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

62. Honig, B. & Yang, A.-S. (1995). Free energy balance in protein folding. *Adv. Protein Chem.* **46**, 27-58.

63. 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.

64. Sippl, M. J., Ortner, M., Jaritz, M., Lackner, P. & Flockner, H. (1996). Helmholtz free energies of atom pair interactions in proteins. *Fold Des.* **1**, 289-298.

65. Klotz, I. M. & Franzen, J. S. (1962). Hydrogen bonds between model peptide groups in solution. *J. Am. Chem. Soc.* **84**, 3461-3466.

66. Kresheck, G. C. & Klotz, I. M. (1966). The thermodynamics of transfer of amides from an apolar to an aqueous solution. *Biochemistry* **8**, 8-12.

67. Gill, S. J. & Noll, L. (1972). Calorimetric study of association of diketopiperazine in water. *J. Phys. Chem.* **76**, 3065-3068.

68. Habermann, S. M. & Murphy, K. P. (1996). Energetics of hydrogen bonding in proteins: a model compound study. *Protein Sci.* **5**, 1229-1239.

69. Shirley, B. A., Stanssens, P., Hahn, U. & Pace, C. N. (1992). Contribution of hydrogen bonding to the conformational stability of ribonuclease T1. *Biochemistry* **31**, 725-732.

70. Fersht, A. R. (1987). The hydrogen bond in molecular recogition. *Trends Biochem. Sci* **12**, 301-304.

71. 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.

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

73. 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.

74. Munoz, V. & Serrano, L. (1995). Elucidating the folding problem of helical peptides using empirical parameters. II. helix macrodipole effects and rational modification of the helical content of natural peptides. *J. Mol. Biol.* **245**, 275-296.

75. 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.

76. Giletto, A. & Pace, C. N. (1999). Buried, charged, non-ion-paired aspartic acid 76 contributes favorably to the conformational stability of ribonuclease T-1. *Biochemistry* **38**, 13379-13384.

77. Deechongkit, S., Nguyen, H., Powers, E. T., Dawson, P. E., Gruebele, M. & Kelly, J. W. (2004). Context-dependent contributions of backbone hydrogen bonding to [beta]-sheet folding energetics. *Nature* **430**, 101-105.

78. Wang, M., Wales, T. E. & Fitzgerald, M. C. (2006). Conserved thermodynamic contributions of backbone hydrogen bonds in a protein fold. *Proc. Natl Acad. Sci. USA* **103**, 2600-2604.

79. Garcia-Moreno, B. E., Dwyer, J. J., Gittis, A. G., Lattman, E. E., Spencer, D. S. & Stites, W. E. (1997). Experimental measurement of the effective dielectric in the hydrophobic core of a protein. *Biophysical Chemistry* **64**, 211-224.

80. Tanford, C. (1961). *Physical chemistry of macromolecules*, John Wiley & Sons, Inc., New York.

81. Wada, A. & Nakamura, H. (1981). Nature of the charge distribution in proteins. *Nature* **293**, 757-758.

82. 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.

83. Fersht, A. R. (1972). Conformational equilibria in α- and δ-chymotrypsin: the energetics and importance of the salt-bridge. *J. Mol. Biol.* **64**, 497-509.

84. Anderson, D. E., Becktel, W. J. & Dahlquist, F. W. (1990). pH-induced denaturation of proteins: a single salt bridge contributes 3-5 kcal/mol to the free energy of folding of T4 lysozyme. *Biochemistry* **29**, 2403-2408.

85. Dao-Pin, S., Sauer, U., Nicholson, H. & Matthews, B. W. (1991). Contribution of engineered surface salt bridges to the stability of T4 lysozyme determined by directed mutagenesis. *Biochemistry* **30**, 7142-7153.

86. 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.

87. Barlow, D. J. & Thornton, J. M. (1983). Ion-pairs in proteins. *J. Mol. Biol.* **168**, 867-885.

88. Honig, B. H., Hubbell, W. L. & Flewelling, R. F. (1986). Electrostatic interactions in membranes and proteins. *Annu. Rev. Biophys. Biomol. Struct.* **15**, 163-193.

89. Waldburger, C. D., Schildbach, J. F. & Sauer, R. T. (1995). Are buried salt bridges important for protein stability and conformational specificity? *Nat. Struct. Biol.* **2**, 122-128.

90. 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.

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

92. Privalov, P. L. (1979). Stability of proteins: small globular proteins. *Adv. Protein Chem.* **33**, 167-241.

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

94. Lee, K. H., Xie, D., Freire, E. & Amzel, L. M. (1994). Estimation of changes in side chain conformational entropy in binding and folding: general methods and application to helix formation. *Proteins: Struct. Funct. Genet.* **20**, 68-84.

95. Trevor, P. C. (2000). Side-chain conformational entropy in protein unfolded states. *Proteins: Struct. Funct. Genet.* **40**, 443-450.

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

97. 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.

98. McKnight, J. C., Doering, D. S., Matsudaira, P. T. & Kim, P. S. (1996). A Thermostable 35-Residue Subdomain within Villin Headpiece. *J. Mol. Biol.* **260**, 126-134.

99. Chiu, T. K., Kubelka, J., Herbst-Irmer, R., Eaton, W. A., Hofrichter, J. & Davies, D. R. (2005). High-resolution x-ray crystal structures of the villin headpiece subdomain, an ultrafast folding protein. *Proc. Natl Acad. Sci. USA* **102**, 7517-7522.

100. Frank, B. S., Vardar, D., Buckley, D. A. & McKnight, C. J. (2002). The role of aromatic residues in the hydrophobic core of the villin headpiece subdomain. *Protein Sci* **11**, 680-687.

101. Kubelka, J., Eaton, W. A. & Hofrichter, J. (2003). Experimental tests of villin subdomain folding simulations. *J. Mol. Biol.* **329**, 625-630.

102. Wickstrom, L., Okur, A., Song, K., Hornak, V., Raleigh, D. P. & Simmerling, C. L. (2006). The unfolded state of the villin headpiece helical subdomain: computational studies of the role of locally stabilized structure. *J. Mol. Biol.* **360**, 1094-1107.

103. Jones, K. & Wittung-Stafshede, P. (2003). The largest protein observed to fold by two-state kinetic mechanism does not obey contact-order correlation. *J. Am. Chem. Soc.* **125**, 9606-9607.

104. Eicken, C., Sharma, V., Klabunde, T., Lawrenz, M. B., Hardham, J. M., Norris, S. J. & Sacchettini, J. C. (2002). Crystal structure of lyme disease variable surface antigen VlsE of *Borrelia burgdorferi*. *J. Biol. Chem.* **277**, 21691-21696.

105. Jones, K., Guidry, J. & Wittung-Stafshede, P. (2001). Characterization of surface antigen from lyme disease spirochete *Borrelia burgdorferi*. *Biochem Biophys Res Commun.* **289**, 389-394.

106. Krieger, E., Koraimann, G. & Vriend, G. (2002). Increasing the precision of comparative models with YASARA NOVA a-self-parameterizing force field. *Proteins: Struct. Funct. Bioinf.* **47**, 393-402.

107. Pace, C. N. & McGrath, T. (1980). Substrate stabilization of lysozyme to thermal and guanidine hydrochloride denaturation. *J. Biol. Chem.* **255**, 3862-3865.

108. Pace, C. N. & Grimsley, G. R. (1988). Ribonuclease T1 is stabilized by cation and anion binding. *Biochemistry* **27**, 3242-3246.

109. Mello, C. C. & Barrick, D. (2003). Measuring the stability of partly folded proteins using TMAO. *Protein Sci.* **12**, 1522-1529.

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

111. Alber, T. & Wozniak, J. A. (1985). A genetic screen for mutations that increase the thermal stability of phage T4 lysozyme. *Proc. Natl Acad. Sci. USA* **82**, 747-750.

112. Kather, I., Jakob, R., Dobbek, H. & Schmid, F. X. (2008). Changing the determinants of protein stability from covalent to non-covalent interactions by *in vitro* evolution: a structural and energetic analysis. *J. Mol. Biol.* **381**, 1040-1054.

113. DiTursi, M. K., Kwon, S.-J., Reeder, P. J. & Dordick, J. S. (2006). Bioinformatics-driven, rational engineering of protein thermostability. *Protein Eng. Des. Sel.* **19**, 517-524.

114. Amin, N., Liu, A. D., Ramer, S., Aehle, W., Meijer, D., Metin, M., Wong, S., Gualfetti, P. & Schellenberger, V. (2004). Construction of stabilized proteins by combinatorial consensus mutagenesis. *Protein Eng. Des. Sel.* **17**, 787-793.

115. Akasako, A., Haruki, M., Oobatake, M. & Kanaya, S. (1997). Conformational stabilities of *Escherichia coli* RNase HI variants with a series of amino acid

substitutions at a cavity within the hydrophobic core. *J. Biol. Chem.* **272**, 18686-18693.

116. Bueno, M., Cremades, N., Neira, J. L. & Sancho, J. (2006). Filling small, empty protein cavities: structural and energetic consequences. *J. Mol. Biol.* **358**, 701-712.
117. Chen, J., Lu, Z., Sakon, J. & Stites, W. E. (2000). Increasing the thermostability of staphylococcal nuclease: implications for the origin of protein thermostability. *J. Mol. Biol.* **303**, 125-130.
118. Hernandez-Rocamora, V. M., Maestro, B., Molla-Morales, A. & Sanz, J. M. (2008). Rational stabilization of the C-LytA affinity tag by protein engineering. *Protein Eng. Des. Sel.* **21**, 709-720.
119. Peterson, R. W., Nicholson, E. M., Thapar, R., Klevit, R. E. & Scholtz, J. M. (1999). Increased helix and protein stability through the introduction of a new tertiary hydrogen bond. *J. Mol. Biol.* **286**, 1609-1619.
120. Jager, M., Zhang, Y., Bieschke, J., Nguyen, H., Dendle, M., Bowman, M. E., Noel, J. P., Gruebele, M. & Kelly, J. W. (2006). Structure-function-folding relationship in a WW domain. *Proc. Natl Acad. Sci. USA* **103**, 10648-10653.
121. Villegas, V., Viguera, A. R., Aviles, F. X. & Serrano, L. (1996). Stabilization of proteins by rational design of [alpha]-helix stability using helix/coil transition theory. *Fold Des.* **1**, 29-34.
122. Thapar, R., Nicholson, E. M., Rajagopal, P., Waygood, E. B., Scholtz, J. M. & Klevit, R. E. (1996). Influence of N-Cap Mutations on the Structure and Stability of Escherichia coli HPr. *Biochemistry* **35**, 11268-11277.

123. Bi, Y., Cho, J. H., Kim, E. Y., Shan, B., Schindelin, H. & Raleigh, D. P. (2007). Rational design, structural and thermodynamic characterization of a hyperstable variant of the villin headpiece helical subdomain. *Biochemistry* **46**, 7497-7505.

124. Simpson, E. R., Meldrum, J. K., Bofill, R., Crespo, M. D., Holmes, E. & Searle, M. S. (2005). Engineering enhanced protein atability through beta-turn optimization: insights for the design of stable peptide beta-hairpin systems. *Angew Chem Int Ed Engl.* **44**, 4939-4944.

125. Loladze, V. V., Ibarra-Molero, B., Sanchez-Ruiz, J. M. & Makhatadze, G. I. (1999). Engineering a thermostable protein via optimization of charge-charge interactions on the protein surface. *Biochemistry* **38**, 16419-16423.

126. Spector, S., Wang, M., Carp, S. A., Robblee, J., Hendsch, Z. S., Fairman, R., Tidor, B. & Raleigh, D. P. (2000). Rational modification of protein stability by the mutation of charged surface residues. *Biochemistry* **39**, 872-879.

127. Schweiker, K. L. & Makhatadze, G. I. (2008). Protein stabilization by the rational design of surface charge-charge interactions. In *Protein structure, stability, and interactions* (Shriver, J. W., ed.), Vol. 490. Humana Press, New York.

128. Schwehm, J. M., Fitch, C. A., Dang, B. N., Garcia-Moreno, B. & Stites, W. E. (2003). Changes in stability upon charge reversal and neutralization substitution in staphylococcal nuclease are dominated by favorable electrostatic effects. *Biochemistry* **42**, 1118-1128.

129. Wunderlich, M., Martin, A. & Schmid, F. X. (2005). Stabilization of the cold shock protein CspB from *Bacillus subtilis* by evolutionary optimization of coulombic interactions. *J. Mol. Biol.* **347**, 1063-1076.

130. 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.

131. 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.

132. Langsetmo, K., Fuchs, J. A. & Woodward, C. (1991). The conserved, buried aspartic acid in oxidized *Escherichia coli* thioredoxin has a pKa of 7.5. Its titration produces a related shift in global stability. *Biochemistry* **30**, 7603 - 7609.

133. Sola, M., Lopez-Hernandez, E., Cronet, P., Lacroix, E., Serrano, L., Coll, M. & Parraga, A. (2000). Towards understanding a molecular switch mechanism: thermodynamic and crystallographic studies of the signal transduction protein CheY. *J. Mol. Biol.* **303**, 213-225.

134. Inoue, M., Yamada, H., Hashimoto, Y., Yasukochi, T., Hamaguchi, K., Miki, T., Horiuchi, T. & Imoto, T. (1992). Stabilization of a protein by removal of unfavorable abnormal pKa: substitution of undissociable residue for glutamic acid-35 in chicken lysozyme. *Biochemistry* **31**, 8816-8821.

135. Perry, L. J. & Wetzel, R. (1984). Disulfide bond engineered into T4 lysozyme: stabilization of the protein towards thermal inactivation. *Science* **226**, 555-557.

136. Wells, J. A. & Powers, D. B. (1986). *In vivo* formation and stability of engineered disulfide bonds in subtilisin. *J. Biol. Chem.* **261**, 6564-6570.

137. Robinson, C. R. & Sauer, R. T. (2000). Striking stabilization of arc repressor by an engineered disulfide bond. *Biochemistry* **39**, 12494-12502.

138. Guzzi, R., Andolfi, L., Cannistraro, S., Verbeet, M. P., Canters, G. W. & Sportelli, L. (2004). Thermal stability of wild type and disulfide bridge containing mutant of poplar plastocyanin. *Biophys. Chem.* **112**, 35-43.

139. Chakraborty, K., Thakurela, S., Prajapati, R. S., Indu, S., Ali, P. S. S., Ramakrishnan, C. & Varadarajan, R. (2005). Protein stabilization by introduction of cross-strand disulfides. *Biochemistry* **44**, 14638-14646.

140. Trevino, S. R., Schaefer, S., Scholtz, J. M. & Pace, C. N. (2007). Increasing protein conformational stability by optimizing [beta]-turn sequence. *J. Mol. Biol.* **373**, 211-218.

141. Matthews, B. W., Nicholson, H. & Becktel, W. J. (1987). Enhanced protein thermostability from site-directed mutations that decrease the entropy of unfolding. *Proc. Natl Acad. Sci. USA* **84**, 6663-6667.

142. Watanabe, K., Masuda, T., Ohashi, H., Mihara, H. & Suzuki, Y. (1994). Multiple proline substitutions cumulatively thermostabilize *Bacillus Cereus* ATCC7064 Oligo-1,6-Glucosidase. *Eur. J. Biochem.* **226**, 277-283.

143. Watanabe, K., Kitamura, K. & Suzuki, Y. (1996). Analysis of the critical sites for protein thermostabilization by proline substitution in oligo-1,6-glucosidase from

Bacillus coagulans ATCC 7050 and the evolutionary consideration of proline residues. *Appl Environ Microbiol.* **62**, 2066-2073.

144. Bogin, O., Peretz, M., Hacham, Y., Korkhin, Y., Frolow, F., Kalb, A. J. & Burstein, Y. (1998). Enhanced thermal stability of Clostridium beijerinckii alcohol dehydrogenase after strategic substitution of amino acid residues with prolines from the homologous thermophilic Thermoanaerobacter brockii alcohol dehydrogenase. *Protein Sci.* **7**, 1156-1163.
145. Prajapati, R. S., Das, M., Sreeramulu, S., Sirajuddin, M., Srinivasan, S., Krishnamurthy, V., Ranjani, R., Ramakrishnan, C. & Varadarajan, R. (2007). Thermodynamic effects of proline introduction on protein stability. *Proteins: Struct. Funct. Bioinf.* **66**, 480-491.
146. Sandberg, W. S. & Terwilliger, T. C. (1991). Energetics of repacking a protein interior. *Proc. Natl Acad. Sci. USA* **88**, 1706-1710.
147. Hendsch, Z. S. & Tidor, B. (1994). Do salt bridges stabilize proteins? A continuum electrostatic analysis. *Protein Sci.* **3**, 211-226.
148. Trefethen, J. M., Pace, C. N., Scholtz, J. M. & Brems, D. N. (2005). Charge-charge interactions in the denatured state influence the folding kinetics of ribonuclease Sa. *Protein Sci* **14**, 1934-1938.
149. Anil, B., Li, Y., Cho, J. H. & Raleigh, D. P. (2006). The unfolded state of NTL9 is compact in the absence of denaturant. *Biochemisty* **45**, 10110-10116.
150. Hebert, E. J., Grimsley, G. R., Hartley, R. W., Horn, G., Schell, D., Garcia, S., Both, V., Sevcik, J. & Pace, C. N. (1997). Purification of ribonucleases Sa, Sa2,

and Sa3 after expression in *Escherichia coli*. *Protein Expression and Purif.* **11**, 162-168.

151. Studier, F. W. & Moffatt, B. A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* **189**, 113-130.

152. Dumon-Seignovert, L., Cariot, G. & Vuillard, L. (2004). The toxicity of recombinant proteins in *Escherichia coli*: a comparison of overexpression in BL21(DE3), C41(DE3), and C43(DE3). *Protein Expression and Purif.* **37**, 203-206.

153. Studier, F. W. (2005). Protein production by auto-induction in high-density shaking cultures. *Protein Expression and Purif.* **41**, 207-234.

154. Guex, N. & Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-Pdb Viewer: An environment for comparative protein modeling. *Electrophoresis* **18**, 2714-2723.

155. Thompson, J. D., Gibson,T.J., Plewniak,F., Jeanmougin,F. and Higgins,D.G. (1997). The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **24**, 4876-4882.

156. Miroux, B. & Walker, J. E. (1996). Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J. Mol. Biol.* **260**, 289-298.

157. Chiu, J., March, P. E., Lee, R. & Tillett, D. (2004). Site-directed, Ligase-Independent Mutagenesis (SLIM): a single-tube methodology approaching 100% efficiency in 4 h. *Nucleic Acids Res.* **32**, e174.

158. Schindler, T., Graumann, P. L., Perl, D., Ma, S., Schmid, F. X. & Marahiel, M. A. (1999). The family of cold shock proteins of *Bacillus subtilis:* stability and dynamics *in vitro* and *in vivo*. *J. Biol. Chem.* **274**, 3407-3413.

159. Makhatadze, G. I., Lopez, M. M., Richardson-Iii, J. M. & Thomas, S. T. (1998). Anion binding to the ubiquitin molecule. *Protein Sci.* **7**, 689-697.

160. Bilsel, O., Zitzewitz, J. A., Bowers, K. E. & Matthews, C. R. (1999). Folding mechanism of the α-subunit of tryptophan synthase, an α/β barrel protein: global analysis highlights the interconversion of multiple native, intermediate, and unfolded forms through parallel channels. *Biochemistry* **38**, 1018-1029.

161. Sheshadri, S., Lingaraju, G. M. & Varadarajan, R. (1999). Denaturant mediated unfolding of both native and molten globule states of maltose binding protein are accompanied by large deltaCp's. *Protein Sci.* **8**, 1689-1695.

162. Goedken, E. R., Raschke, T. M. & Marqusee, S. (1997). Importance of the c-terminal helix to the stability and enzymatic activity of *Escherichia coli* ribonuclease HI. *Biochemistry* **36**, 7256-7263.

163. Grimsley, G. R. & Pace, C. N. (2003). Spectrophotometric determination of protein concentration. In *Current protocols in protein science* (Hoang, T., ed.). John Wiley and Sons, Inc., Hoboken, NJ.

164. Pace, C. N. & Scholtz, J. M. (1996). Measuring the conformational stability of a protein. In *Protein structure: a practical approach.* (Creighton, T. E., ed.), pp. 299-321. IRL Press, Oxford, MA.

165. 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.

166. Greene, R. F. & Pace, C. N. (1974). Urea and guanidine hydrochloride denaturation of ribonuclease, lysozyme, agr-chymotrypsin, and {beta}-lactoglobulin. *J. Biol. Chem.* **249**, 5388-5393.

167. Matthews, B. W. (1993). Structural and genetic analysis of the protein stability. *Annual Review of Biochemistry* **62**, 139-160.

168. Tang, Y., Grey, M. J., McKnight, J., Palmer Iii, A. G. & Raleigh, D. P. (2006). Multistate folding of the villin headpiece domain. *J. Mol. Biol.* **355**, 1066-1077.

169. Kubelka, J., Chiu, T. K., Davies, D. R., Eaton, W. A. & Hofrichter, J. (2006). Sub-microsecond protein folding. *J. Mol. Biol.* **359**, 546-553.

170. Vardar, D., Buckley, D. A., Frank, B. S. & McKnight, C. J. (1999). NMR structure of an F-actin-binding "headpiece" motif from villin. *J. Mol. Biol.* **294**, 1299-1310.

171. Homouz, D., Perham, M., Samiotakis, A., Cheung, M. S. & Wittung-Stafshede, P. (2008). Crowded, cell-like environment induces shape changes in aspherical protein. *Proc. Natl Acad. Sci. USA* **105**, 11754-11759.

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

173. Bi, Y., Tang, Y., Raleigh, D. P. & Cho, J.-H. (2006). Efficient high level expression of peptides and proteins as fusion proteins with the N-terminal domain of L9: Application to the villin headpiece helical subdomain. *Protein Expression and Purif.* **47**, 234–240.

174. Pace, C. N. & Laurents, D. V. (1989). A new method for determining the heat capacity change for protein folding. *Biochemistry* **28**, 2520-2525.

175. Murphy, K. P. & Freire, E. (1992). Thermodynamics of structural stability and cooperative folding behavior in proteins. *Adv. Protein Chem.* **43**, 313-361.

176. Spolar, R. S., Livingstone, J. R. & Record, M. T. (1992). Use of liquid hydrocarbon and amide transfer data to estimate contributions to thermodynamic functions of protein folding from the removal of nonpolar and polar surface from water. *Biochemistry* **31**, 3947-3955.

177. Myers, J. K., Pace, C. N. & Scholtz, J. M. (1995). Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci* **4**, 2138-2148.

178. Loladze, V. V., Ermolenko, D. N. & Makhatadze, G. I. (2001). Heat capacity changes upon burial of polar and nonpolar groups in proteins. *Protein Sci.* **10**, 1343-1352.

179. 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.

180. Wickstrom, L., Bi, Y., Hornak, V., Raleigh, D. P. & Simmerling, C. (2007). Reconciling the solution and X-ray structures of the villin headpiece helical subdomain: molecular dynamics simulations and double mutant cycles reveal a stabilizing cation−π interaction. *Biochemistry* **46**, 3624-3634.

181. Kumar, M. D. S., Bava, K. A., Gromiha, M. M., Prabakaran, P., Kitajima, K., Uedaira, H. & Sarai, A. (2006). ProTherm and ProNIT: thermodynamic databases for proteins and protein-nucleic acid interactions. *Nucleic Acids Res.* **34**, D204-206.

182. Zhou, H. & Zhou, Y. (2004). Quantifying the effect of burial of amino acid residues on protein stability. *Proteins: Struct. Funct. Genet.* **54**, 315-322.

183. Dong, H., Mukaiyama, A., Tadokoro, T., Koga, Y., Takano, K. & Kanaya, S. (2008). Hydrophobic effect on the stability and folding of a hyperthermophilic protein. *J. Mol. Biol.* **378**, 264-272.

184. Takano, K., Yamagata, Y. & Yutani, K. (2003). Buried water molecules contribute to the conformational stability of a protein. *Protein Eng. Des. Sel.* **16**, 5-9.

185. Machicado, C., Bueno, M. & Sancho, J. (2002). Predicting the structure of protein cavities created by mutation. *Protein Eng. Des. Sel.* **15**, 669-675.

186. Pace, C. N., Shirley, B. A., McNutt, M. & Gajiwala, K. (1996). Forces contributing to the conformational stability of proteins. *FASEB J.* **76**, 75-83.

187. Schweiker, K. L., Zarrine-Afsar, A., Davidson, A. R. & Makhatadze, G. I. (2007). Computational design of the Fyn SH3 domain with increased stability through optimization of surface charge charge interactions. *Protein Sci.* **16**, 2694-2702.

188. Predki, P. F., Agrawal, V., Brunger, A. T. & Regan, L. (1996). Amino-acid substitutions in a surface turn modulate protein stability. *Nat. Struct. Biol.* **3**, 54-58.

189. Lee, J., Dubey, V. K., Longo, L. M. & Blaber, M. (2008). A logical or redundancy within the Asx-Pro-Asx-Gly type I [beta]-turn motif. *J. Mol. Biol.* **377**, 1251-1264.

190. Ohage, E. C., Graml, W., Walter, M. M., Steinbacher, S. & Steipe, B. (1997). {beta}-Turn propensities as paradigms for the analysis of structural motifs to engineer protein stability. *Protein Sci.* **6**, 233-241.

191. Hutchinson, E. G. & Thornton, J. M. (1996). PROMOTIF--A program to identify and analyze structural motifs in proteins. *Protein Sci.* **5**, 212-220.

192. Guruprasad, K. & Rajkumar, S. (2000). Beta-and gamma-turns in proteins revisited: a new set of amino acid turn-type dependent positional preferences and potentials. *J. Biosci.* **25**, 143-156.

193. Schindelin, H., Marahiel, M. A. & Heinemann, U. (1993). Universal nucleic acid-binding domain revealed by crystal structure of the B. subtilis major cold-shock protein. *Nature* **364**, 164-168.

194. Ermolenko, D. N., Dangi, B., Gvritishvili, A., Gronenborn, A. M. & Makhatadze, G. I. (2007). Elimination of the C-cap in ubiquitin--structure, dynamics and thermodynamic consequences. *Biophys. Chem.* **126**, 25-35.

195. Sevcik, J., Dauter, Z. & Wilson, K. S. (2004). Crystal structure reveals two alternative conformations in the active site of ribonuclease Sa2. *Acta Crystallogr., Sect D: Biol. Crystallogr.* **60**, 1198-1204.

196. Sevcik, J., Urbanikova, L., Leland, P. A. & Raines, R. T. (2002). X-ray structure of two crystalline forms of a streptomycete ribonuclease with cytotoxic activity. *J. Biol. Chem.* **277**, 47325-47330.

197. Martinez-Oyanedel, J., Choe, H.-W., Heinemann, U. & Saenger, W. (1991). Ribonuclease T1 with free recognition and catalytic site: Crystal structure analysis at 1.5 Å resolution. *J. Mol. Biol.* **222**, 335-352.

198. Katayanagi, K., Miyagawa, M., Matsushima, M., Ishikawa, M., Kanaya, S., Nakamura, H., Ikehara, M., Matsuzaki, T. & Morikawa, K. (1992). Structural details of ribonuclease H from Escherichia coli as refined to an atomic resolution. *J. Mol. Biol.* **223**, 1029-1052.

199. Nishio, K., Morimoto, Y., Ishizuka, M., Ogasahara, K., Tsukihara, T. & Yutani, K. (2005). Conformational changes in the α-subunit coupled to binding of the β2-Subunit of tryptophan synthase from *Escherichia coli*: crystal structure of the tryptophan synthase α-subunit alone. *Biochemistry* **44**, 1184-1192.

200. Quiocho, F., Spurlino, J. & Rodseth, L. (1997). Extensive features of tight oligosaccharide binding revealed in high-resolution structures of the maltodextrin transport/chemosensory receptor. *Structure* **5**, 997-1015.

201. Prajapati, R. S., Sirajuddin, M., Durani, V., Sreeramulu, S. & Varadarajan, R. (2006). Contribution of cation−π interactions to protein stability. *Biochemistry* **45**, 15000-15010.

202. Wintrode, P. L., Makhatadze, G. I. & Privalov, P. L. (1994). Thermodynamics of ubiquitin unfolding. *Proteins: Struct. Funct. Genet.* **18**, 246-253.

203. Agah, S., Larson, J. D. & Henzl, M. T. (2003). Impact of proline residues on parvalbumin stability. *Biochemistry* **42**, 10886-10895.

204. Razvi, A. (2005). A comparative study of HPr proteins from extremophilic organisms. Dissertation, Texas A&M University; College Station.

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

206. Nemethy, G., Leach, S. J. & Scheraga, H. A. (1966). The influence of amino acid side chains on the free energy of helix-coil transitions. *J. Phys. Chem.* **70**, 998-1004.

207. Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* **7**, 95-99.

208. Stites, W. E., Meeker, A. K. & Shortle, D. (1994). Evidence for strained interactions between side-chains and the polypeptide backbone. *J. Mol. Biol.* **235**, 27-32.

209. Masumoto, K., Ueda, T., Motoshima, H. & Imoto, T. (2000). Relationship between local structure and stability in hen egg white lysozyme mutant with alanine substituted for glycine. *Protein Eng. Des. Sel.* **13**, 691-695.

210. Kimura, S., Kanaya, S. & Nakamura, H. (1992). Thermostabilization of *Escherichia coli* ribonuclease HI by replacing left-handed helical Lys95 with Gly or Asn. *J. Biol. Chem.* **267**, 22014-22017.

211. Yun, R. H., Anderson, A. & Hermans, J. (1991). Proline in alpha-helix: stability and conformation studied by dynamics simulation. *Proteins: Struct. Funct. Bioinf.* **10**, 219-228.

212. Takano, K., Yamagata, Y. & Yutani, K. (2001). Role of amino acid residues in left-handed helical conformation for the conformational stability of a protein. *Proteins: Struct. Funct. Bioinf.* **45**, 274-280.

213. Watanabe, K., Chishiro, K., Kitamura, K. & Suzuki, Y. (1991). Proline residues responsible for thermostability occur with high frequency in the loop regions of an extremely thermostable oligo-1,6- glucosidase from Bacillus thermoglucosidasius KP1006. *J. Biol. Chem.* **266**, 24287-24294.

214. Vieille, C. & Zeikus, G. J. (2001). Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiol Mol Biol Rev.* **65**, 1-43.

215. Zeikus, J. G., Vieille, C. & Savchenko, A. (1998). Thermozymes: biotechnology and structure–function relationships. *Extremophiles* **2**, 179-183.

216. Szilágyi, A. & Závodszky, P. (2000). Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey. *Structure* **8**, 493-504.

217. Malakauskas, S. M. & Mayo, S. L. (1998). Design, structure and stability of a hyperthermophilic protein variant. *Nat. Struct. Biol.* **5**, 470-475.

218. Pace, C. N. (2000). Single surface stabilizer. *Nat. Struct. Biol.* **7**, 345-6.

219. Makhatadze, G. I., Loladze, V. V., Gribenko, A. V. & Lopez, M. M. (2004). Mechanism of thermostabilization in a designed cold shock protein with optimized surface electrostatic interactions. *J. Mol. Biol.* **336**, 929-942.

220. Shaw, K. L., Grimsley, G. R., Yakovlev, G. I., Makarov, A. A. & Pace, C. N. (2001). The effect of net charge on the solubility, activity, and stability of ribonuclease Sa. *Protein Sci* **10**, 1206-1215.

221. Huyghues-Despointes, B. M., Thurlkill, R. L., Daily, M. D., Schell, D., Briggs, J. M., Antosiewicz, J. M., Pace, C. N. & Scholtz, J. M. (2003). pK values of histidine residues in ribonuclease Sa: effect of salt and net charge. *J. Mol. Biol.* **325**, 1093-105.

222. Laurents, D. V., Huyghues-Despointes, B. M., Bruix, M., Thurlkill, R. L., Schell, D., Newsom, S., Grimsley, G. R., Shaw, K. L., Trevino, S., Rico, M., Briggs, J. M., Antosiewicz, J. M., Scholtz, J. M. & Pace, C. N. (2003). Charge-charge interactions are key determinants of the pK values of ionizable groups in ribonuclease Sa (pI=3.5) and a basic variant (pI=10.2). *J. Mol. Biol.* **325**, 1077-92.

223. Yakovlev, G. I., Mitkevich, V. A., Shaw, K. L., Trevino, S., Newsom, S., Pace, C. N. & Makarov, A. A. (2003). Contribution of active site residues to the activity and thermal stability of ribonuclease Sa. *Protein Sci.* **12**, 2367-2373.

224. Pace, C. N. (1986). Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol.* **131**, 266-280.

225. 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.

226. Pace, C. N. (1990). Conformational stability of globular proteins. *Trends Biochem. Sci* **15**, 14-17.

227. Hartley, R. & Rogerson, D. J. (1972). Production and purification of the extracellular ribonuclease of Bacillus amyloliquefaciens (barnase) and its intracellular inhibitor (barstar). II. Barstar. *Prep Biochem.* **2**, 243-50.

228. Hartley, R. & Rogerson, D. J. (1972). Production and purification of the extracellular ribonuclease of Bacillus amyloliquefaciens (barnase) and its intracellular inhibitor (barstar). I. Barnase. *Prep Biochem.* **2**, 229-42.

229. Pace, C. N. & Vanderberg, K. E. (1979). Determining globular protein stability: guanidine hydrochloride denaturation of myoglobin. *Biochemistry* **18**, 288-292.

230. Ferreon, A. C. M. & Bolen, D. W. (2004). Thermodynamics of denaturant-induced unfolding of a protein that exhibits variable two-state denaturation. *Biochemistry* **43**, 13357-13369.

231. Makhatadze, G. I. (1999). Thermodynamics of protein interactions with urea and guanidinium hydrochloride. *J. Phys. Chem. B,* **103**, 4781-4785.

232. Smith, J. S. & Scholtz, J. M. (1996). Guanidine hydrochloride unfolding of peptide helices: separation of denaturant and salt effects. *Biochemistry* **35**, 7292-7297.

233. von Hippel, P. H. & Wong, K.-Y. (1965). On the conformational stability of globular proteins. *J. Biol. Chem.* **240**, 3909-3923.

234. Baldwin, R. L. (1996). How Hofmeister ion interactions affect protein stability. *Biophys J* **71**, 2056-2063.

235. Laurents, D. V., Huyghues-Despointes, B. M. P., Bruix, M., Thurlkill, R. L., Schell, D., Newsom, S., Grimsley, G. R., Shaw, K. L., Trevino, S. & Rico, M. (2003). Charge-charge interactions are key determinants of the pK values of ionizable groups in ribonuclease Sa (pI=3.5) and a basic variant (pI=10.2). *J. Mol. Biol.* **325**, 1077-1092.

236. McNutt, M., Mullins, L. S., Raushel, F. M. & Pace, C. N. (1990). Contribution of histidine residues to the conformational stability of ribonuclease T1 and mutant Glu58->Ala. *Biochemistry* **29**, 7572-7576.

237. Pace, C. N., Laurents, D. V. & Erickson, R. E. (1992). Urea denaturation of barnase: pH dependence and characterization of the unfolded state. *Biochemistry* **31**, 2728-2734.

238. Privalov, P. L. & Khechinashvili, N. N. (1974). A thermodynamic approach to the problem of stabilization of globular protein structure: a calorimetric study. *J. Mol. Biol.* **86**, 665-684.

239. McCrary, B. S., Bedell, J., Edmondson, S. P. & Shriver, J. W. (1998). Linkage of protonation and anion binding to the folding of Sac7d. *J. Mol. Biol.* **276**, 203-224.

240. Guinto, E. R. & Di Cera, E. (1996). Large heat capacity change in a protein−monovalent cation interaction. *Biochemistry* **35**, 8800-8804.

241. Liu, Y. & Sturtevant, J. M. (1996). The observed change in heat capacity accompanying the thermal unfolding of proteins depends on the composition of the

solution and on the method employed to change the temperature of unfolding. *Biochemistry* **35**, 3059-3062.

242. Pace, C. N., Grimsley, G. R., Thomas, S. T. & Makhatadze, G. I. (1999). Heat capacity change for ribonuclease A folding. *Protein Sci* **8**, 1500-1504.

243. Wildes, D., Anderson, L. M., Sabogal, A. & Marqusee, S. (2006). Native state energetics of the Src SH2 domain: evidence for a partially structured state in the denatured ensemble. *Protein Sci.* **15**, 1769-1779.

244. Street, T. O., Courtemanche, N. & Barrick, D. (2008). Protein folding and stability using denaturants. In *Methods in cell biology*, Vol. 84, pp. 295-325. Academic Press, San Diego, CA.

245. McCrary, B. S., Edmondson, S. P. & Shriver, J. W. (1996). Hyperthermophile protein folding thermodynamics: differential scanning calorimetry and chemical denaturation of Sac7d. *J. Mol. Biol.* **264**, 784-805.

246. Lee, C.-F., Allen, M. D., Bycroft, M. & Wong, K.-B. (2005). Electrostatic interactions contribute to reduced heat capacity change of unfolding in a thermophilic ribosomal protein L30e. *J. Mol. Biol.* **348**, 419-431.

247. Robic, S., Guzman-Casado, M., Sanchez-Ruiz, J. M. & Marqusee, S. (2003). Role of residual structure in the unfolded state of a thermophilic protein. *Proc. Natl Acad. Sci. USA* **100**, 11345-11349.

248. Zhou, H., X. (2002). Toward the physical basis of thermophilic proteins: linking of enriched polar interactions and reduced heat capacity of unfolding. *Biophys J* **83**, 3126-3133.

249. Guzman-Casado, M., Parody-Morreale, A., Robic, S., Marqusee, S. & Sanchez-Ruiz, J. M. (2003). Energetic evidence for formation of a pH-dependent hydrophobic cluster in the denatured state of *Thermus thermophilus* ribonuclease H. *J. Mol. Biol.* **329**, 731-743.

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

251. Dill, K. A. & Shortle, D. (1991). Denatured states of proteins. *Annu. Rev. Biochem.* **60**, 795-825.

252. Ladbury, J. E., Wynn, R., Thomson, J. A. & Sturtevant, J. M. (1995). Substitution of charged residues into the hydrophobic core of *Escherichia coli* thioredoxin results in a change in heat capacity of the native protein. *Biochemistry* **34**, 2148-2152.

253. Becktel, W. J. & Schellman, J. A. (1987). Protein stability curves. *Biopolymers* **26**, 1859-1877.

254. Nojima, H., Ikai, A., Oshima, T. & Noda, H. (1977). Reversible thermal unfolding of thermostable phosphoglycerate kinase. Thermostability associated with mean zero enthalpy change. *J. Mol. Biol.* **116**, 429-442.

255. Razvi, A. & Scholtz, J. M. (2006). Lessons in stability from thermophilic proteins. *Protein Sci* **15**, 1569-78.

256. Asada, Y., Sawano, M., Ogasahara, K., Nakamura, J., Ota, M., Kuroishi, C., Sugahara, M., Yutani, K. & Kunishima, N. (2005). Stabilization mechanism of the

tryptophan synthase {alpha}-subunit from *Thermus thermophilus* HB8: X-Ray crystallographic analysis and calorimetry. *J. Biochem.* **138**, 343-353.

VITA

Name: Hailong Fu

Address: c/o Dr. J. Martin Scholtz

440 Reynolds Medical Building

1114 TAMU, College Station, TX 77843-1114

E-mail address: hailong@tamu.edu

Education: B.S., Molecular and Cellular Biology, University of Science and Technology of China, 1999

Ph.D., Biochemistry, Texas A&M University, 2009