

INVESTIGATION OF ELECTROSTATICS IN A HYDROGEN BOND NETWORK

A Senior Honors Thesis

by

FATEMA SULTANA UDDIN

Submitted to the Office of Honors Programs
& Academic Scholarships
Texas A&M University
in partial fulfillment of the requirements of the

UNIVERSITY UNDERGRADUATE
RESEARCH FELLOWS

April 2001

Group: Biochemistry

INVESTIGATION OF ELECTROSTATICS IN A HYDROGEN BOND NETWORK

A Senior Honors Thesis

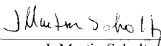
by

FATEMA SULTANA UDDIN

Submitted to the Office of Honors Programs
& Academic Scholarships
Texas A&M University
in partial fulfillment of the requirements of the

UNIVERSITY UNDERGRADUATE
RESEARCH FELLOWS

Approved as to style and content by:



J. Martin Scholtz
(Fellows Advisor)



Edward A. Funkhouser
(Executive Director)

April 2001

Group: Biochemistry

ABSTRACT

Investigating the Role of Electrostatics in a Hydrogen Bond Network. (April 2001)

Fatema Sultana Uddin
Department of Biology
Texas A&M University

Fellows Advisor: Dr. J. Martin Scholtz
Department of Medical Biochemistry and Genetics

Proteins are important for many reasons—chief among them are that many enzymes are proteins. Without enzymes, most biochemical reactions would not take place. Comparisons between proteins or amino acid sequences that differ only slightly from one another can provide estimates of energy gained or lost as a result of those differences. In this project, the use of site-directed mutagenesis will allow us to introduce minor changes into a protein. The differences in the conformation stability of the wild type and mutant proteins permit us to study the influence of a specific amino acid's side chain to its environment.

In a protein, there are several interactions that can be studied. Our particular area of focus on this project is electrostatics. We work with one protein in particular, HPr, which is isolated from two different bacterial sources. The two proteins differ with respect to amino acid sequence, yet are able to retain a singular function. Similarly, the

region of interest, although defined by different amino acids, still retains comparable folding patterns between the proteins. In either protein, a negatively charged amino acid is fixed in position by its interaction with other neighboring amino acids. We believe that the amino acid is important in stabilizing the area via electrostatic interactions between both negatively and positively charged amino acids. We will test this hypothesis by making several mutant proteins and measuring their conformational stability.

The mutant D30T bsHPr showed a decrease in energy of ~ 1.0 kcal/mol. An attempt to determine the pKa of the wild type residue was unsuccessful because studies could not be conducted at low pH values. D72K bsHPr was also made. This was a difficult protein to purify; studies that were done on it were inconclusive. In general, it can be concluded that the amino acid residues within the wild type form stabilizing interactions. The absence of charged residues at key locations within a defined region lead to decreases in energy for the protein.

I hope that I make my parents as happy as they make me feel.

ACKNOWLEDGMENTS

I would like to thank Dr. J. Martin Scholtz, my advisor and incredibly patient mentor. Thank you for your guidance. I would also like to thank the graduate students that have helped me along the way, especially Ron Peterson, Jason Schmittschmitt, and Roy Alston, but not to exclude the others of the Scholtz/Pace laboratories.

TABLE OF CONTENTS

	Page
ABSTRACT.....	iii
DEDICATION.....	v
ACKNOWLEDGMENTS.....	vi
TABLE OF CONTENTS.....	vii
LIST OF FIGURES.....	viii
CHAPTER	
I INTRODUCTION.....	1
II METHODS.....	16
III RESULTS.....	30
IV DISCUSSION.....	35
V CONCLUSION.....	39
REFERENCES.....	41
VITA.....	42

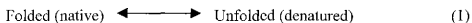
LIST OF FIGURES

FIGURES		Page
1	Conformational stability.....	2
2	A ribbon diagram of <i>bsHPr</i>	8
3	A ribbon diagram of <i>ecHPr</i>	9
4	The hydrogen bond network involving Asp69 in <i>bsHPr</i>	10
5	The hydrogen bond network involving Asp69 in <i>ecHPr</i>	11
6	Residues that surround Asp69 in <i>bsHPr</i>	13
7	Residues that surround Asp69 in <i>ecHPr</i>	14
8	The amino acids.....	15
9	The structure of polyacrylamide.....	21
10	The structure of sodium dodecyl sulfate (SDS).....	22
11	The structure of diethylaminoethyl (DEAE) and carboxymethyl (CM).....	23
12	The structures of urea and guanidinium chloride.....	26
13	The linear extrapolation method.....	27
14	Urea denaturation curve of D30T <i>bsHPr</i>	32
15	Urea denaturation curve of D30T <i>bsHPr</i> as compared to wild type.....	33

CHAPTER I
INTRODUCTION

Conformational Stability

In order to understand why and how a protein adopts a specific three-dimensional shape, we need to understand the forces that determine its conformational stability. Conformational energy can be defined as the free energy difference between the folded and unfolded states of a protein. It is the free energy for the reaction (1):



The goal of our studies is to try to increase the conformational stability of a protein. The function of an enzyme or any protein is intimately tied to its structure; understanding how a protein folds, then, necessarily provides us comprehension of its three-dimensional shape. It follows, then, that a more complete understanding of structure would allow us to at least take advantage of, if not improve, a protein's function. This could have important implications in medicine, the performance of biochemical tests, and industry. If conformational stability is the energy difference between the folded and unfolded states of a protein, then the only way to increase its value is to increase the energy difference between the folded and unfolded states. There are three ways that we can do this: increase the energy of the unfolded state, decrease the energy of the folded state, or increase the energy of the transition state. This thesis follows the style and format of *Biochemistry*.

energy of the folded state, or change the energy of both states. Thus the question becomes how to affect those changes—the answer lies in perturbing the forces involved in stabilizing the protein.

There are many forces that stabilize the folded state of a protein (Figure 1). These forces include electrostatics, hydrogen bonds, hydrophobics, and van der Waals interactions.

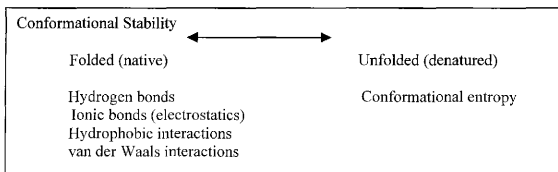


FIGURE 1: Conformational stability. Conformational stability is defined as the free energy difference between the folded and unfolded states of a protein. This figure provides an illustration of that definition. It also identifies some of the major forces that stabilize either the native or denatured states of the protein.

Electrostatic interactions between two charged atoms are potentially the strongest noncovalent forces. The attraction between ions of the same or opposite charge can extend over great distances. Charge-charge interactions can therefore influence the structures of individual biomolecules as well as their interaction with other charged molecules. Hydrogen bonds are electrostatic interactions between a weakly acidic donor group and an acceptor atom that bears a lone pair of electrons (2). Alpha helices and beta-pleated sheets provide examples of known hydrogen bonding within a protein; the hydrogen bonds are formed between backbone amide and carbonyl groups. The amide

group donates the hydrogen while the carbonyl oxygen (which has a two pairs of lone electrons) functions as the acceptor. Internal hydrogen bonds within a folded protein are arranged such that a maximum number of stabilizing interactions are formed. Since an unfolded protein forms all of its hydrogen bonds with solvent molecules (mostly water—a molecule capable of being either the donor or acceptor within a hydrogen bond), the energy contribution of hydrogen bonding can be determined by calculating the energy difference between the folded and unfolded states (2). Hydrophobic considerations are more easily defined in terms of the hydrophobic effect: the name that designates those influences that cause nonpolar groups and substances to minimize their contact with polar molecules, e.g., water. It is believed that water molecules rearrange themselves around a nonpolar molecule; this rearrangement disrupts the highly ordered and extensively hydrogen bonded network of liquid water. Thus, the hydrophobic effect is driven by favorable changes in entropy. The primary force that stabilizes the unfolded state is conformational entropy.

The most common changes made to a protein involve single changes in the amino acid sequence; these changes are introduced to a specific site via site-directed mutagenesis. Changes can also be made in a protein via covalent modification of an amino acid side chain. In either case, the change made within a protein is specific. We make alterations to test specific ideas about what role certain amino acids play in the stabilization of the wild type protein. Thus, when the difference in energy between the wild type and mutant proteins is determined, we can draw specific conclusions about the role of a given amino acid with respect to the forces that it is believed to contribute to.

A brief review of thermodynamics

Thermodynamics describe the relationships between various forms of energy. The feasibility of a physical process is determined by the thermodynamics of its system; however, it is important to note that it does not determine the rate of the process. There are three laws of thermodynamics of which only the first two will be discussed here: first, energy can neither be created nor destroyed, only conserved; and second, spontaneous processes occur in directions that increase the overall disorder of the universe. The first law merely states that energy is conserved. The change in energy (U) is the difference between the heat (q) absorbed by a system from its surroundings and the work (w) done by the system on its surroundings:

$$\Delta U = U_{\text{final}} - U_{\text{initial}} = q - w \quad (2)$$

Work is defined as a force, while heat reflects the amount of random molecular motion. If heat is released by a system ($-q$) then the process is exothermic; in the case that heat is absorbed ($+q$), the process is endothermic. Thus, work done by the system is a positive quantity. Enthalpy is the measured heat that is generated or absorbed by a system. An enthalpically favored process is one that is exothermic ($\Delta H_{\text{system}} < 0$).

The second law of thermodynamics allows us to determine whether a process will occur spontaneously or not. Spontaneous processes are characterized by the conversion of order to chaos. According to the second law of thermodynamics, disorder

is the favored state of the universe. This is based on probabilities; it is far more likely to find a random arrangement of components than it is to find them in an ordered array. For example, when a drop of dye is added to a beaker of water, the dye drop does not remain in aggregate; rather, its molecules diffuse throughout the water in a random arrangement. It is more probable to find the random array than the single drop. The molecular rearrangement of the dye drop is a spontaneous process. The laws of random chance cause any system of reasonable size to spontaneously adopt its most probable arrangement, the one in which the entropy is a maximum, simply because this state is so overwhelmingly probable (2). Any spontaneous process causes the entropy of the universe to increase ($\Delta S_{\text{universe}} > 0$). This is not to say that a system cannot become ordered or increase its amount of order. In fact, the ordering of a system can only occur at with an increase in the disorder of its surroundings.

Although a process may be entropically favored, it may not occur spontaneously because of its unfavorable enthalpic change. Thus, some relationship between enthalpy and entropy is required to definitively characterize a process as spontaneously occurring or not. That relationship was formulated by J. Willard Gibbs in 1878 and is known as Gibbs free energy; it is an indicator of spontaneity for constant temperatures and pressure processes. It can be defined with the following equation:

$$\Delta G = \Delta H - T\Delta S \quad (3)$$

where G is free energy, H is enthalpy, T is temperature, and S is entropy. Provided that $\Delta G \leq 0$, then a process is deemed as spontaneously occurring. Spontaneous processes are said to be exergonic (they can be utilized to do work). Processes at equilibrium have ΔG values of zero. It is also important to note that ΔG varies with temperature. For biological systems, ΔG represents the maximum amount of recoverable work, meaning that ΔG is an indication of the greatest difference between two states of a system, i.e., the folded and unfolded states of protein.

The model system, HPr

The protein that we study is one that is isolated from two different bacterial sources, *Escherichia coli* and *Bacillus subtilis* (Figures 2, 3). Its role in either of these bacterial systems is in sugar transport. In *B. subtilis*, the histidine-containing protein (HPr) can be phosphorylated at two different sites: either at an active site histidine residue or at a regulatory serine. It has been shown via thermodynamic analyses of single and double mutants that the two phosphorylation sites are not energetically coupled in the HPr protein. On the other hand, *E. coli* HPr has only one phosphorylation site. The often reversible phosphorylation of enzymes as a means of regulation is very common in biological systems. Examples of such regulation include that of glycogen phosphorylase and isocitrate dehydrogenase (3). It is thought that phosphorylation causes a protein to undergo a conformational change and thus alter its active site to either release or bind a substrate, thus effectively turning on or shutting off a biochemical pathway.

In either *E. coli* or *B. subtilis*, HPr retains the same function despite having different amino acid sequences. This is in part because the proteins have comparable folding patterns—each has three alpha helices and four beta sheets. These proteins may utilize similar structural characteristics differently. This may be due to electrostatic and charge-charge interactions. Electrostatic interactions or any other type of favorable or unfavorable interactions exist as a consequence of the amino acids present within a given environment. Studying the contribution an amino acid makes necessarily involves changing it; site-directed mutagenesis allows us to do that. At position 69 in each protein is a conserved aspartic acid residue (Figures 4, 5). It is believed to interact with the backbone atoms of residues 30 and 31 in either protein in a hydrogen bond network as well as with the residue present at position 45. In both cases, it is remarkable that virtually the same interaction is formed using residues present at the same positions although those residues are not conserved between the proteins. It has been demonstrated that Asp69 is integral to the formation of this favorable interaction because when the residue is altered, the protein loses a significant amount of conformational stability. Thus, we are interested in learning if the formation of this hydrogen bond network is contingent on the presence of specific charged residues within the neighborhood of Asp69. This hypothesis was tested by altering the charges present in the vicinity of Asp69 and measuring the consequent difference in conformational stability.

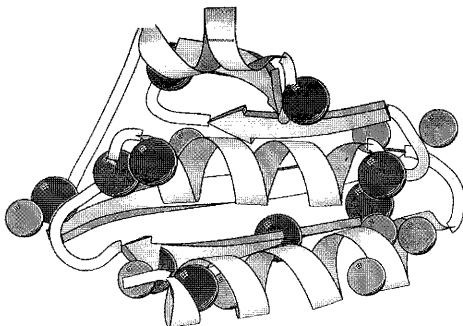


FIGURE 2: A ribbon diagram of *bsHPr*. This protein contains 88 amino acid residues. It contains three alpha helices and four beta sheets. It is a convenient model for protein study because it obeys a two-state folding equilibrium ($F \rightleftharpoons U$). The larger spheres indicate positively charged residues, while negatively charged residues are indicated by smaller ones.

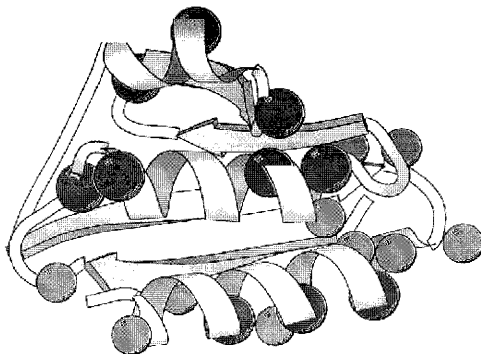


FIGURE 3: A ribbon diagram of *echPr*. This protein contains 85 amino acid residues. It contains three alpha helices and four beta sheets. It is a convenient model for protein study because it obeys a two-state folding equilibrium ($F \rightleftharpoons U$). The larger spheres indicate positively charged residues, while negatively charged residues are indicated by smaller ones.

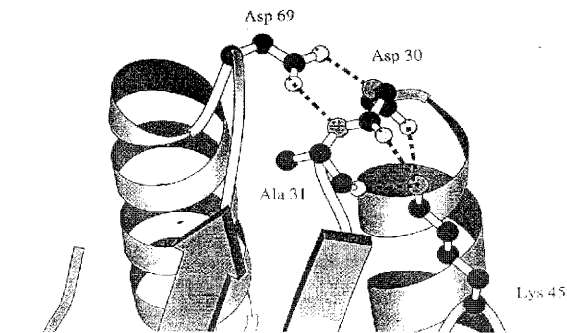


FIGURE 4: The hydrogen bond network involving Asp69 in *bsHPr*. Alpha helices A and C are shown on the right and left, respectively. The side chain of Asp69 accepts hydrogen bonds from backbone NH of residues Asp30 and Ala31, while the side chain of Lys45 appears to donate hydrogen bonds to the backbone carbonyl (C=O) of residues 30 and 31.

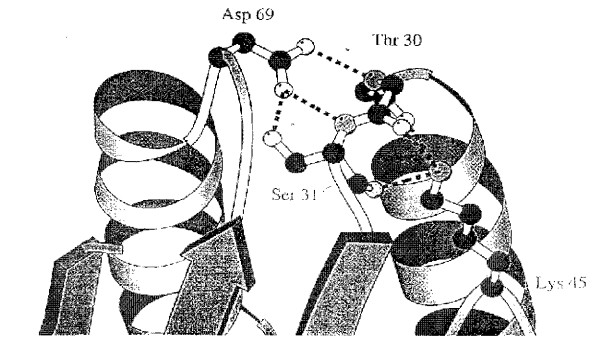


FIGURE 5: The hydrogen bond network involving Asp69 in *ecHPr*. Alpha helices A and C are shown on the right and left, respectively. The side chain of Asp69 accepts hydrogen bonds from backbone NH of Thr30 and Ser31, while the side chain of Lys45 appears to donate hydrogen bonds to the backbone carbonyl (C=O) of residues 30 and 31. A hydrogen bond is formed between side chains Ser31 and Asp69.

Mutant design

Within this section and throughout this paper, mutants are labeled according to the mutation they contain. For example, D30T *bs*HPr is a mutant made in *B. subtilis* HPr; in the wild type at position 30 there is an aspartic acid (D) that has been changed to a threonine (T). A mutant made in *E. coli* would have the letters “*ec*” precede HPr.

Mutations are made within a protein in a deliberate manner. We believe that formation of the hydrogen bond network is dependent on the charge-charge interactions between Asp69 and the surrounding charged amino acid side chains (Figures 6, 7). As such, the mutants that were designed had changes made to them at positions 30, 68, and 72. Additionally, since we wanted to test the hypothesis that electrostatics play a role in hydrogen bond formation, we introduced uncharged and oppositely charged residues in lieu of the wild type ones. Site-directed mutagenesis allowed us to introduce a new amino acid in place of the original; thus we had to decide which amino acid would be appropriate in our study. It is convenient that we study two proteins because we are able to easily decide what the new amino acid should be. When deciding to make a change in *B. subtilis* HPr at position 30, we simply look to see which amino acid is present in the *E. coli* protein. If the amino acids differ, then we have found the change to be made. Thus, we made the mutant D30T *bs*HPr, changing the aspartic acid residue at position thirty to a threonine; at position 30 in wild type *E. coli* HPr, there is a threonine present. This is how we designed all three of our mutants: D30T *bs*IIPr, D72K *bs*HPr, and E68A *ec*HPr (Figure 8). In hindsight, it may have been easier to make mutants that were similar in that they either all eliminated a charge or introduced the same opposite charge.

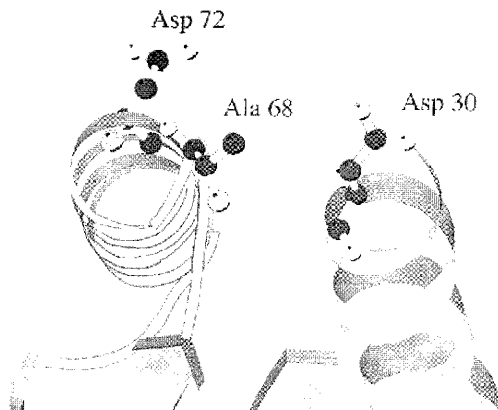


FIGURE 6: Residues that surround Asp69 in *bsHPr*. This ribbon diagram shows using ball and stick form the residues present at positions 30, 68, and 72. The charges that surround Asp69 are thought to contribute to its ability to form the favorable interactions diagrammed in Figure 4.

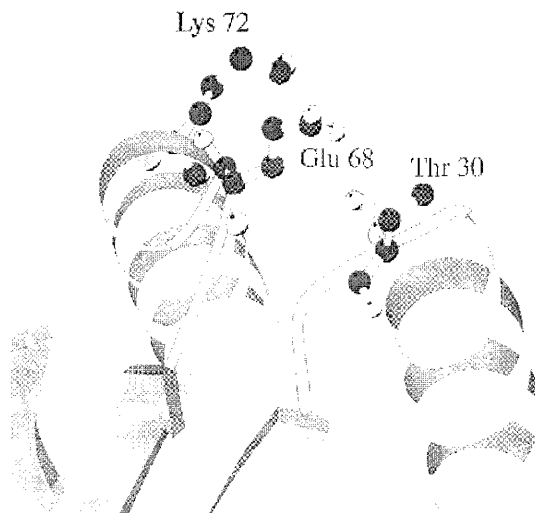


FIGURE 7: Residues that surround Asp69 in eChPr. This ribbon diagram shows in ball and stick representation the residues present at positions 30, 68, and 72. The charges that surround Asp69 are thought to contribute to its ability to form the favorable interactions diagrammed in Figure 5.

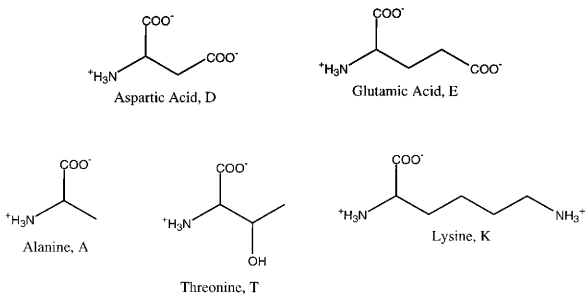


FIGURE 8: The amino acids. This figure shows the amino acids aspartic acid, glutamic acid, alanine, threonine, and lysine. Only threonine and alanine are neutral in charge; both aspartic and glutamic acids are negatively charged while lysine is positively charged. The formation of electrostatic interactions is contingent upon the presence of charged residues; thus, those side chains that carry a charge greatly influence the shape and function of a protein.

CHAPTER 2

METHODS

The procedures followed for this project can be generally described in three sections. First, we make a mutant. This involves introducing a change in the amino acid sequence by site-directed mutagenesis and amplifying the results with the aid of polymerase chain reaction. Once the DNA has been amplified, a bacterial culture is transformed with the DNA. There is large-scale growth of the cells, and finally the cells are harvested for the protein of interest. In the second general step, the protein is purified, first on an ion exchange column (separation based on anion exchange) and then by gel filtration (separation based on shape and size). In the final phase, the stability of the protein is measured by denaturing it. The denaturant can be either urea or temperature. Linear extrapolation provides estimates of ΔG

Making a mutant

The process of making a mutant begins with designing the proper primer. Primers are short nucleotide sequences that contain the desired mutation. Amino acids are coded for by any combination of three nucleotides (and a string of amino acids is known as a polypeptide). Thus, when designing a primer, changes are made by selecting the amino acid to be altered and then changing the nucleotides that code for it. The mutation site is generally made in the middle of the primer sequence. Ideally, the nucleotides that flank the mutation site should be rich in Guanine-Cytosine pairs; G-C pairs have three

hydrogen bonds between them and are therefore stronger than their Adenine-Thymine counterparts that only have two hydrogen bonds between them. The number of nucleotides that abut the mutation site is determined by calculating an approximate annealing temperature of 60°C. The annealing temperature, T, can be determined as follows:

$$T = 4^{\circ}\text{C} \times (\# \text{ G-C pairs}) + 2^{\circ}\text{C} \times (\# \text{ A-T pairs}). \quad (4)$$

Once the primer is designed, they can be ordered on-line at: www.idtdna.com.

Polymerase chain reaction (PCR) has become an invaluable tool for researchers. It allows the amplification of specific DNA segments of up to 6 kb (6000 base pairs). In this technique, a denatured DNA sample is incubated with DNA polymerase and two oligonucleotide primers that direct the DNA polymerase to synthesize new complimentary strands (also included in the reaction is a mix of dNTPs—a source of free nucleosides). Each cycle approximately doubles the amount of DNA present; thus after a short number of cycles the amount of DNA has increased exponentially. In each cycle, the two strands of the DNA template are separated by heat denaturation, the primers are annealed to their complimentary segments on the DNA, and the DNA polymerase directs the synthesis of the complimentary strands. Heat stable DNA polymerase is used so that the need to add fresh enzyme at the end of each cycle is avoided (2). The ingredients for the PCR reaction are obtained from Stratagene. The following cycle was used in the making of our mutants:

98°C for 5 minutes

*98°C for 1 minute (the denaturation step)

*Annealing temperature (~60°C) for 1 minute (allows primer to bind denatured
DNA template)

*68 - 72°C for 1 minute (extension)

72°C for 7 minutes

4°C stops the reaction (4°C is also the storage temperature of the DNA)

*steps constitute the cyclic portion of the reaction

Once the PCR reactions are complete, a restriction enzyme is added to the eppendorf tube so that any DNA not containing the mutation is digested.

So far we have designed and hopefully amplified a mutant piece of DNA that codes for a mutant protein of interest. Now we need to be able to produce the protein that we want to study. Thank goodness for bacteria which are willing to overlook our abuse and do just that. Transformation is the uptake of naked DNA from its surroundings by a bacterial cell. Cells treated with calcium chloride have perforated cell walls and membranes that allow them to take up environment; such cells are deemed supercompetent (as opposed to incompetent). Aliquots of amplified DNA, supercompetent cells, and LB broth (media required for bacterial growth) are incubated together at 37°C for one hour. After one hour, the cells are plated on LB plates that have been spread with ampicillin. The presence of ampicillin on the plate allows us to select

for only those cells that contain new DNA. The plates are then incubated overnight. Hopefully something will grow. If in the morning colonies pockmark your plate, then members of a single colony are further grown at 37°C in 5mL of LB; again ampicillin is added as a selection factor. The QIAprep Spin Plasmid kit provides the components of the next step: isolation of our plasmid. Once the miniprep is completed, an agarose gel is run on the sample collected to verify that indeed a piece of DNA has been isolated. If the agarose gel verifies that DNA was isolated, then a sample of the DNA is sent to the GeneTech Laboratories located on main campus of Texas A&M for sequencing (the nucleotide order of the plasmid or DNA segment is determined). When the sequencing results are compared to a wild type sequence, the only difference should appear at the site of the mutation; thus, sequencing results need to be checked before any further steps are taken to ensure that the mutation was indeed successful and that what will subsequently be grown contains the desired mutation.

Once it has been determined that the amplified DNA contains the mutation we want, we can grow a large amount of bacteria that contain it. The media in which the cells are grown is centrifuged; once the supernatant has been discarded, the pellet, which contains the cells, is re-suspended in buffer. This cell solution is then subjected to a pressure of approximately 650 psi, which shears the cell walls, thus releasing the cellular content without damaging it. The cell homogenate is centrifuged; hopefully the proteins remain in the supernatant, which is dialyzed overnight in buffer. Dialysis is a process that separates molecules according to size through the use of semi-permeable membranes containing pores of less than macromolecular dimensions. The pores block

the passages of larger molecules, but allow small molecules of solvent, salt, and small metabolites to diffuse across the membrane.

Purify the protein

Purification of the protein is accomplished with the aid of column chromatography and gel electrophoresis. In any chromatographic procedure, the solution loaded onto the column is known as the mobile phase. The column contains the stationary phase, which consists of a porous solid material, which in some cases may be associated with a bound liquid. As the mobile phase interacts with the stationary, the progress of certain materials is slowed; hence, a separation of mobile phase contents based on different migratory rates that can be collected individually as labeled fractions. Different chromatographs exist as a function of their differing mobile and stationary phases: gas-liquid chromatography employs gaseous mobile phase and a liquid stationary phase while with liquid-liquid chromatography both phases are liquid with one bound to an inert solid. Chromatography can be further characterized according to the nature of the stationary phase; the two types of chromatographs that we use are ion exchange and then gel filtration.

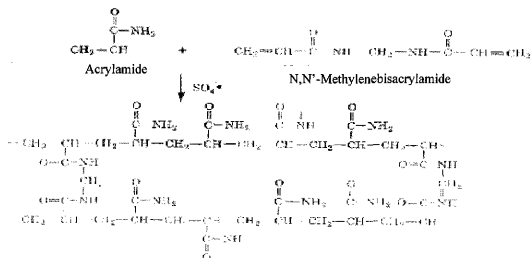


FIGURE 9: The structure of polyacrylamide. Gel electrophoresis is used to assess the quantity and purity of a protein sample. The gel used in polyacrylamide gel electrophoresis is a polymer made by the free radical reaction of acrylamide and N,N-methylenebisacrylamide. Similar to gel filtration chromatography, electrophoresis is able to achieve separation of components based on size as well as charge.

Electrophoresis is the migration of ions within an electric field and is used for the analytical separation of molecules. The separation of molecules is based on gel filtration properties as well as the charge nature of the solutes. Contrary to gel filtration chromatography, smaller molecules migrate more quickly through a gel subjected to electrophoresis. In polyacrylamide gel electrophoresis (PAGE), the gels are made by radical polymerization of acrylamide and N,N-methylenebisacrylamide (Figure 9). In order to ensure that all loaded samples contain the same charge and are migrating through the gel based on size, protein samples are first mixed with a detergent, sodium dodecyl sulfate (SDS) (Figure 10). Detergents such as SDS are strong protein denaturants; they are amphipathic molecules that once bound to a protein, force it to assume a rod-like shape. The large negative charge of the detergent masks the intrinsic

charges of the protein such that those proteins loaded on an SDS-containing gel are separated according to their molecular mass.

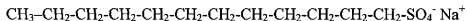


FIGURE 10: The structure of sodium dodecyl sulfate (SDS). Like other detergents, SDS is a strong protein denaturant. Detergents are able to denature proteins because of their amphipathic nature. The long hydrophobic tail of SDS interacts with hydrophobic regions within a protein, forcing the protein to expose its nonpolar interior. The large charge group of SDS interacts with the solvent and masks the native charges of the protein.

Typically, fractions collected off a chromatograph as well as a standard molecular weight marker are run on an SDS-PAGE gel before fractions are pooled; this ensures that only those fractions containing the appropriate protein are collected. Purity of the protein can also be assessed from these gels; usually the presence of a single band in a given well is a good indication of the purity of the sample. Furthermore, the intensity of the band can also be used to assess the relative concentration of sample within a fraction.

In an ion exchange column, the bound group on the column packing can either be positively or negatively charged. Anion exchangers contain bound positive groups, and thus, they attract anions to those sites. Cation exchangers, on the other hand, are just the opposite; bound negative groups attract cationic solutes. The first column that we run our solution over is an anion exchange column whose bound group is diethylaminoethyl (DEAE) (Figure 11). Our negatively charged protein hopefully binds to the column and then is eluted off with a gradient solution of increasing salt concentration. It is expected

that the desired protein is not the only negatively charged solute in the protein soup, but the stationary phase is selected as such to ensure the greatest binding strength between it and the desired solute. Furthermore, during elution, those solutes that are only marginally attracted to the matrix are eluted off the column first; they flow through the column faster. This occurs because those particles are not as highly attracted to the matrix beads are able to move faster through the column while the progress of other solutes is retarded as it flows through more beads to which it they are strongly attracted to. Fractions are collected in test tubes; as the fraction is collected, a spectrophotometer measures the absorbance of each sample. Based on key absorption peaks at expected times and analysis of SDS-PAGE, the fractions that correspond to those peaks and presumably contain the desired protein are pooled and freeze-dried. They are later resuspended and then run over a second chromatograph.

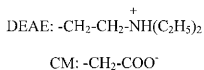


FIGURE 11: The structures of diethylaminoethyl (DEAE) and carboxymethyl (CM). Ion exchange columns achieve separation of mobile phase components based on charge. The bound group on ion exchange column packing can either be positively or negatively charged. Anion exchangers contain bound positive groups, such as DEAE, and thus, they attract anions to those sites. Cation exchangers such as CM are just the opposite; they bind positively charged groups.

Gel filtration chromatography is also known as size exclusion or molecular sieve chromatography (2). Separation of particles on this column is based on size and shape of the solute. The stationary phase consists of beads of a hydrated, sponge-like material containing pores that span a relatively narrow size range of molecular dimensions (2).

Thus, as a solution passes through the column, solutes that exceed the size of the pores are excluded from them; their progress through the column is quicker compared to those solutes that find themselves weaving through the interior of the beads. The path length of the larger solutes is shorter than those of smaller ones, and thus they come off the column sooner. Gel filtration is necessary after a solution has passed through an ion exchange column because it allows the protein to be “desalted.” Since the solution was eluted off the ion exchange column with a salt gradient, gel filtration allows us to remove any excess salt in order to obtain a purer protein sample. The exclusion limit of the gel is equivalent to the molecular mass of the smallest molecule unable to penetrate the pores of that gel. Again, fractions are collected and absorbencies are measured using a spectrophotometer. The appropriate fractions are pooled and freeze-dried. The sample’s purity is assessed using gel electrophoresis; hopefully a pure protein has been obtained and an assessment of its stability can begin.

Measure the stability of the protein—determining ΔG

Once a pure protein has been isolated and purified, a measure of the stability it has gained or lost as a result of the mutation made to it needs to be made. Based on wavelength scans of the protein, we can determine that it is indeed an HPr protein and that it is pure. Additional information from a wavelength scan can also be determined, including the almost obvious but essential fact that the protein is folded. Conformation stability has been defined as the free energy difference between the folded and unfolded states of the protein. In order to obtain a value for ΔG , we need to denature the protein

(obtain measurements in its unfolded state). A protein can be unfolded using a variety of denaturants; all denaturants share the common feature that they disrupt any or a combination of the forces that stabilize the folded state of protein. Among the more commonly used denaturants or denaturing conditions are:

1. temperature—when a protein is heated, conformationally sensitive properties such as optical rotation, viscosity, and UV absorption are altered; a protein's T_m , or melting temperature, is the midpoint of the process during which the protein is unfolded
2. pH variations—it is well known that proteins function at many different pH levels, with each protein having an optimum pH; placing an enzyme in an environment that alters the ionic states its amino acid constituents almost invariably will disrupt the protein's native conformation because of the affect the environment has on electrostatic and hydrogen bond formation
3. detergents—these agents have already been described as strong protein denaturants; these molecules hydrophobically associate with the nonpolar residues of a protein, thereby interfering with the hydrophobic interactions formed by the protein in its natural state
4. high concentrations of water-soluble organic substances—e.g., alcohol—these substances also interfere with stabilizing hydrophobic interactions
5. salts and chaotropes—the influence of salts is variable; some stabilize the unfolded state by interacting with exposed ionic groups, while others such as

guanidinium chloride (one of the most commonly used denaturants) act as chaotropic agents. The other chaotrope that is not a salt is urea (Figure 12). Chaotropic agents increase the solubility of nonpolar substances in water; their effectiveness as denaturing agents stems from their ability to disrupt hydrophobic interactions (2).

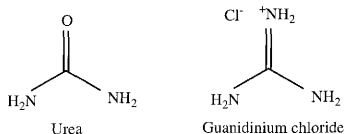


FIGURE 12: The structures of urea and guanidinium chloride. Among the more commonly used denaturants are guanidinium chloride and urea. Salts such as guanidinium chloride may stabilize the unfolded state of a protein by interacting with exposed ionic groups. Urea acts as a chaotropic agent, increasing the solubility of nonpolar substances in water.

The denaturant of choice is urea. A urea denaturation is more or less a titration experiment. A 10M stock of urea is made and mixed with a small amount of the purified protein. Shots from this stock are added to a cuvette containing an aliquot of purified protein suspended in buffer. The cuvette is located within a cell holder of a spectroscopic instrument. Before a shot of concentrated urea is added to the cuvette, an equivalent amount of buffer/protein is removed. Once the urea is added, the solution is allowed to mix and a measurement is taken. Circular dichroism is used to follow the unfolding of the protein. The amount of folded protein is being measured as the

concentration of denaturant increases. The end of the experiment is reached once the CD signal in the post-transition has become steady (a flat line); hence, most, if not all, of the protein remaining in the cuvette is denatured (Figure 13). The experiments for this project were conducted at pH 7 and 25°C. The protein's unfolding is followed at 222nm because that is the wavelength at which the greatest difference in absorbencies between the folded and unfolded protein occurs.

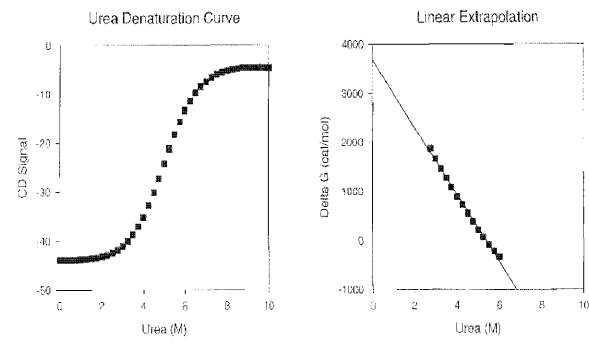


FIGURE 13: The linear extrapolation method. A sample urea denaturation curve plots the amount of folded protein present measured by some spectroscopic signal (CD) as a function of increasing denaturant. Parameters that define the pre-transition, transition, and post-transition regions are best fit to a line in the linear extrapolation method. Linear extrapolation allows us to estimate values of delta G for the protein.

The assumption that HPr follow a two state folding mechanism is important; it allows us to calculate the amount of denatured protein using:

$$F_d = (y_n - y_{\text{obs}})/(y_n - y_d) \quad (5)$$

where y_{obs} is the observed variable parameter (CD signal) and y_n and y_d are the values of y characteristic of the native and denatured states of the protein. Using the determined value F_d , ΔG can be calculated:

$$\Delta G = -RT \ln [F_d/(1 - F_d)] \quad (6)$$

$$= -RT \ln [(y_n - y_{\text{obs}})/(y_n - y_d)] \quad (7)$$

where R is the gas constant and T is the absolute temperature. In most cases, ΔG is directly proportional to the concentration of the denaturant; the data from the transition region is fit to the equation

$$\Delta G = \Delta G(\text{H}_2\text{O}) - m[\text{D}] \quad (8)$$

using the method of least-squares analysis. $\Delta G(\text{H}_2\text{O})$ is the value of ΔG in the absence of denaturant and m is a measure of the dependence of ΔG on the concentration of denaturant, $[\text{D}]$. Note that the equation is of the general form of a line, $y = mx + b$ (4, 5). Once a value for ΔG has been obtained, it can be compared to the wild type or other calculated values of ΔG to determine whether the protein being studied is more or less stable.

The pre- and post-transition regions of the urea denaturation are not ignored. Similarly to the transition region, parameters of these regions are best fit to a line; together with the transition region, they are all fit to a single line of the equation,

$$y = \{(y_n + m_n[D]) + (y_d + m_d[D]) \times \exp[m \times ([D]-[D]_{1/2})/RT]\} / \quad (9)$$

$$(1 + \exp[m \times ([D]-[D]_{1/2})/RT]) \quad (10)$$

where m_n and m_d are the slope of the pre- and post-transition (native and denatured) regions, respectively, and $[D]_{1/2}$ measures the midpoint of the transition region (5).

CHAPTER 3

RESULTS

D30T bsHPR

This mutant was successfully isolated and purified. A sample of the mutant was sent to Gene Technologies Laboratory located on the main campus of Texas A&M University. They sequenced the gene product we provided to them and upon reading the sequence, we were able to verify that we were successful in making a mutation at position 30 in the wild type protein. At position 30, the nucleotides GAC code for an aspartic acid; sequencing results showed that the mutation ACC, which codes for a threonine, had been made. Once the protein was isolated from a large-scale growth of *B. subtilis*, it was purified on both DEAE and gel filtration columns. The protein was denatured with a stock of 10M urea; denaturation condition were pH 7.0 and 25°C (Figure 14). The protein stock had a concentration of 3mg/ml, and a 10mM potassium phosphate buffer was used. Not much can be determined from a single curve plot because there is no comparison to the wild type to help estimate whether the mutant gained or lost stability as a result of the mutation that was made. However, using the linear extrapolation method, ΔG for D30T *bsHPr* was determined to be 4.1 kcal/mol. When plotted against the wild type protein, we can see that our mutant is less stable; it takes more urea to denature the wild type protein than it does to denature the mutant (Figure 15). Calculations indicate a decrease in energy of 0.9 kcal/mol for the mutant from the wild type protein. We attempted to determine the pKa of the wild type residue,

but were unable to; the protein precipitated out of solution when studies were conducted at low pHs.

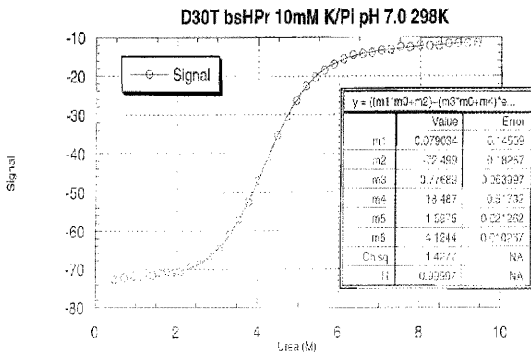


FIGURE 14: Urea denaturation curve of D30T *bsHPr*. This figure shows the denaturation of D30T *bsHPr* carried out at 25°C and pH 7.0 in 10mM potassium phosphate buffer. The stock concentration of urea was 10M, and the protein concentration used was 3mg/ml. The chart on the graph indicates the slopes and intercepts of the pre-transition, transition, and post-transition regions. These parameters will be best fit to a line via the linear extrapolation method, thus allowing us to gain an estimate of ΔG for this mutant. The denaturation curve will also then compared to wild type *bsHPr* data.

Fraction Folded vs Urea wt *bsHPr* and D30T *bsHPr*

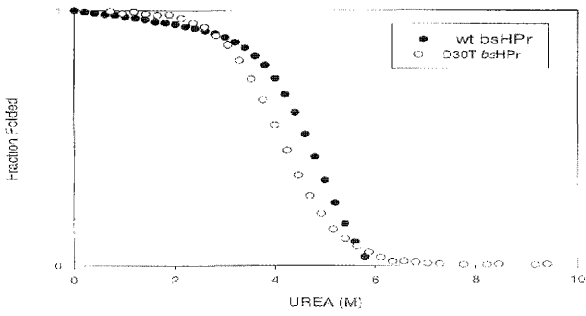


FIGURE 15: Urea denaturation curve of D30T *bsHPr* as compared to wild type. This figure shows the denaturation of D30T *bsHPr* carried out at 25°C and pH 7.0 in 10mM potassium phosphate buffer (dark circles). The concentration of denaturant was 10M. When compared to wild type (open circles), denatured under the same conditions, the calculated ΔG value for the mutant is ~ 1.0 kcal/mol less, thus indicating a decrease in conformational energy. It can be inferred from this that the wild type protein is more stable than the mutant. It is easier to think of stability in terms of how much urea is required to denature the protein; it takes more urea to denature the more stable protein, hence, the wild type is more stable.

D72K bsHPr

There were two attempts to make this protein. In the first attempt, sequencing of the gene showed that at position 72, AAG replaced the wild type nucleotides, GAT. This was confirmation that the mutation made at position 72 in the wild type protein was from an aspartic acid to a lysine. Initial studies of an isolated protein showed that it was probably purified in its unfolded form. Although sequencing results indicated otherwise, we became suspicious that this was not an HPr protein. In order to verify the identity of the protein, a wavelength scan was done. Also, the protein was subjected to SDS-PAGE. The gel verified the purity of the protein, but the wavelength scan was inconclusive. The characteristic absorbance at 222nm of HPr proteins was absent. The thought occurred that the protein had formed an inclusion body. Future work may include purifying the protein from such an aggregate. The protein would be extracted with a strong protein denaturant such as guanidinium chloride and then re-purifying it. Before that, though, a second attempt to make the protein was undertaken. The DNA from the first preparation was used again to make the mutant, since its sequence confirmed that it contained the mutation. In retrospect, it may have been better to start over. Regardless, this effort proved fruitless yet again, as troubles were encountered with the DEAE column. Between loading the column and running a gel, the protein remained outside of ideal storage conditions for about two days. The protein probably became denatured; in any case, it was not detected on the subsequently run gel.

CHAPTER 4

DISCUSSION

D30T bsHPR

Energetically, the mutant is much less stable than the wild type protein. Wild type *B. subtilis* HPr has a ΔG value of 5.0 kcal/mol. The calculated ΔG value of the mutant was 4.1 kcal/mol. Thus, there was a decrease in energy of 0.9 kcal/mol. From estimates using model compound data, suggest that this mutant should have gained -1.0 kcal/mol of energy as a combined effect of hydrophobic and entropy considerations. The change made to the protein introduced an uncharged threonine in place of a negatively charged aspartic acid. The loss of charge makes threonine at least marginally less polar than aspartic acid, hence the increase in hydrophobicity. Entropic considerations arise out of the fact that an aspartic acid forms more bonds between its atoms and thus may have more opportunity for bond rearrangement and rotary motion than the side chain of threonine. Minimizing those motions should take away from the entropy that stabilizes the unfolded state of a protein, thus making the mutant more stable. Alas, this is not what was observed on the urea denaturation. Given that there is no place for a covalent or hydrogen bond to form, this leaves us with the loss or gain of electrostatic interactions as the only explanation for the decrease in energy. In order to verify this hypothesis, we would have to determine the pKa of the residue.

It is well understood that electrostatic interactions are important in protein stability. Hence, it can be further understood that the ionic states of amino acids

influence the three-dimensional structure and biochemical functions of a protein. Some amino acids are acids, while others are basic in nature. Their protonation or deprotonation is dependent on their immediate environment and the pH of the solution. Electrostatic effects within a protein involve changes in their ionization tendencies, or pKa values. The pKa of an ionizable group can be defined as the pH at the midpoint of that group's titration. In other words, the pKa is the point during a titration at which there are equal amounts of the protonated and deprotonated species. The values of amino acid side chain pKas can differ between proteins for two reasons: first, the alpha amino and carboxyl groups lose their charges once linked by peptide bonds (they no longer exert strong inductive effects) and second, the position within the protein of the side chain may differ between proteins with respect to their microenvironment (6). There are several other factors that influence pKa values, among them, solvent accessibility and polarity. Determining the pKa value of a side chain within a protein can tell us about its interactions and its environment. When comparing calculated pKa values of an amino acid side chain within the wild type and mutant proteins, a decrease in pKa value suggests that in the wild type the amino acid formed a stabilizing interaction, whereas an increase in pKa value hints at a more stabilizing interaction within the mutant. With regards to environment, if many positively charged residues surround the amino acid side chain, it is more likely to ionize (lose a proton); the resulting negative charge can be stabilized. In general, the formation of favorable electrostatic interactions upon ionization increases the tendency of any group to ionize. No change in pKa value can be suggestive of two things: first, the residue that is

involved in the electrostatic event in the folded state forms the same or similar interaction in the unfolded state, or second, there was no electrostatic interaction.

In order to determine a pKa value, $\Delta\Delta G$ values (wild type minus mutant ΔG values) are plotted against pH. $\Delta\Delta G$ values can be calculated using the following:

$$\Delta(\Delta G) = \Delta(\Delta G)([H^+] = \infty) + RT \ln \left(\frac{[1+K_{i,N}/[H^+]]}{[1+K_{i,D}/[H^+]]} \right)$$

where $K_{i,N}$ and $K_{i,D}$ are the initial equilibrium constants of the native and denatured states. Several denaturations need to be done to determine enough $\Delta\Delta G$ values to establish pre-, post- and transition regions of a denaturation curve. In order to expedite the process, thermal denaturations are done instead; they use temperature instead of urea as the denaturant. Thermal denaturations are of additional use to us because they allow the confirmation of data obtained and calculated through urea denaturations. In determining pKa values, samples of protein are denatured at a number of intervals within a pH range. If, as was the case with our protein, low enough pH settings do not facilitate making the necessary measurements, a pre-transition baseline cannot be established and, hence, a pKa value cannot be reliably calculated.

In order to determine the pKa of the aspartic acid side chain, we may make a mutant that replaces the side chain with a histidine. The side chain of histidine ionizes at about pH 7.0, thus eliminating the difficulty in conducting experiments at low pHs.

D72K bsHPr

Based on the fact that there has not yet been a mutant that could not be isolated, we believe now that the protein is in all likelihood housed within an inclusion body. An inclusion body is a package of synthesized protein and other cellular components. It is not known why inclusion bodies form, but it is thought to be because large amounts of protein are made and the relatively insoluble unfolded polypeptide accumulates before it can be properly folded (7). There are advantages to the formation of an inclusion body such as that they aid in purification of a protein, and they protect the protein from being degraded within the cell. Future work on this protein will include running gels on the pellets of the centrifuged homogenates that resulted from this protein preparation. If the protein is present, we will isolate and purify it, and then test its stability.

If the protein has not gone into an inclusion body, there are still other options to explore. We could work with the already isolated protein: conducting mass spectroscopy on the protein would allow us to verify its identity through molecular weight comparisons. We could place the protein in a more stable background before running denaturations on it; a stable background could be a buffer with a high salt concentration. Finally, the last resort would be to simply (although experience would suggest otherwise) remake the protein.

CHAPTER 5

CONCLUSION

Given that D30T *bsHPr* lost ~1 kcal/mol of energy, it can be concluded that the aspartic acid residue present in the wild type protein forms a stabilizing interaction. By replacing a charged residue with an uncharged one (threonine), we may have demonstrated that the charge on the aspartic acid residue at position 30 contributes to the formation of the hydrogen bond network. It is difficult to draw definite conclusions from works that are still in progress. However, we can still speculate as to why certain residues interact in the manner they do and hope that our hypotheses are later verified.

In *bsHPr*, Asp 69 is surrounded by two additional aspartic acid residues (at positions 30 and 72); having three negatively charged residues in such close proximity may at first seem an error in nature's judgement, but on closer inspection can be rationalized. The repulsive forces at work between like-charged groups may force the residues into other favorable interactions within the protein, among which include the hydrogen bond network that has Asp69 as its hinge.

On the other hand, in *ecHPr*, the conserved Asp 69 is surrounded by both positive and negative charges at positions 72 and 68, respectively. The negative charge may act to repulse Asp 69 such that both of the negatively charged residues form favorable interactions with the resident positively charged residue and with other charged residues within the protein (the hydrogen bond network).

In conclusion, although one may get the impression that the forces that impart to a protein its conformational stability work independent of one another, it is not an unreasonable speculation that they instead work with and as a function of one another. The dynamics of protein folding are far from being solved, but are becoming better understood.

REFERENCES

1. Pace, C. N. (1990) *Trends in Biotechnol.* 8, 93-98.
2. Voet, D., Voet, J., and Pratt, C. (1999) in *Fundamentals of Biochemistry* John Wiley & Sons, Inc., New York.
3. Huffine, M. E., and Scholtz, J. M. (1996) *J. Biol. Chem.* 271, 28898-28902.
4. Pace, C. N., and Shaw, K., L. (2000) *Proteins: Structure, Function, and Genetics Suppl 4*, 1-7.
5. Pace, C. N., and Scholtz, J. M. (1997) in *Protein Structure: A Practical Approach* (Creighton, T. E., Ed.) pp 299-321, Oxford University Press, New York, NY.
6. Voet, D., and Voet, J., (1995) in *Biochemistry* pp 61-62, John Wiley & Sons, Inc., New York.
7. Creighton, T. E. (1993) in *Proteins: Structure and Molecular Properties* pp 62-63, W. H. Freeman and Company, New York.

VITA

Fatema Sultana Uddin
5221 White Oak
El Paso, Texas 79932

Education:

Texas A&M University College Station, Texas 1998 – present
*Bachelor of Science degree in **Biology** with a minor in **Psychology***
Expected date of graduation: **December 2001**
Cumulative GPA: 4.00

Honors:

Dean's Honor Roll (fall 98 – present)
Presidential Endowed Scholarship holder (fall 98 – present)
University Scholar (fall 99 – present)
Barry M. Goldwater Scholarship nominee (fall 00)
Gathwright Award Winner (spring 01)
Howard Hughes Undergraduate Medical Intern (fall 98 – spring 99)

Professional Experience:

Del Sol Medical Center El Paso, Texas summer 2000
Co-author

Helped author "Marked Luekopenia, Severe Neutropnia, and Reactive Plasmacytosis Resmbling Multiple Myeloma Secondary to Asacol"; under review for publication

Columbia Regional Oncology Center El Paso, Texas summer 1999
Research Assistant

Researched the possible link between gastric carcinoma and locality with the director of the Oncology Center; hopeful of publication

Columbia Medical Center El Paso, Texas 1998
Research Assistant

Research the possibility of finding an antibody specific to *Giardia lamblia*

Extracurricular Activities:

ASPIRE mentor
Golden Key National Honor Society
Phi Eta Sigma
Phi Kappa Phi
University Scholar Mentor Group participant
Research with Dr. J. Martin Scholtz
College Bowl competitor
Muslim Student Association