HOW NATURE FINE TUNES PROTEIN STABILITY

A Senior Honors Thesis

by

MEGAN MARIE WICKSTROM

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 2007

Major: Biology

ABSTRACT

How Nature Fine Tunes Protein Stability (April 2007)

Megan Wickstrom
Department of Biology
Texas A&M University

Fellows Co-Advisors:
Dr. C. Nick Pace
Department of Medical Biochemistry and Genetics
Dr. Lisa Pérez
Department of Chemistry

The purpose of this project was to gain a better understanding of the means used to keep large proteins from becoming too stable. As globular proteins become larger, they bury a larger fraction of their side chains and peptide groups. Based on past research performed on forces stabilizing proteins, they should become much more stable as they become larger. However, this is not observed. This suggests that evolution has strategies to keep large proteins from becoming too stable. Theoretical and computational methods were used to compare and analyze the structure and stability of 8 globular proteins, ranging in size from 36 to 370 residues. For each of the proteins,

computational methods were used to estimate the groups buried and the number of hydrogen bonds formed during folding. The estimated conformational stabilities were then able to be determined. The estimated stabilities were larger than the observed stabilities, and the difference increased with increasing size. The most surprising finding was that the burial of charged groups also increased with increasing size from less than 25% in the small proteins to over 50% in the larger proteins. This suggests that burying charged groups in the interior of the protein is the primary strategy used to fine tune protein stability. To confirm these findings, it is recommended that the computational methods used in this study be extended to a larger sample of proteins and that experimental methods be used to gain a better understanding of the cost of burying charged groups in protein folding.

DEDICATION

This thesis is dedicated to my parents who have always been there for me. I have always looked up to and admired them and I appreciate their loving support. Also, this thesis is dedicated to my fiancé who has been a constant source of motivation and inspiration.

ACKNOWLEDGMENTS

I wish to express my sincere thanks and appreciation to my supervisor and mentor, Dr. C. Nick Pace, for his attention, guidance, and support throughout the course of this work. A special thanks to my co-advisor, Dr. Lisa Pérez, for her help and assistance in using the computational methods during this research. I also appreciate the Laboratory for Molecular Simulation for allowing me to use their computers and numerous programs. Many thanks to Dr. J. Martin Scholtz and Saul Trevino for their assistance during my research.

I wish to thank the Office of Honors Programs and Academic Scholarships and the University Honors Research Fellows Program for giving me the opportunity to perform this research and write this thesis. More specifically, thank you to Mrs. Cindy Raisor, University Honors Research Fellows Program Coordinator, and Dr. Kurt Ritter, Associate Director of Honors Programs, for their constant guidance and patience.

TABLE OF CONTENTS

Page

ABSTRACT………………………………………………………….. ii

DEDICATION………………………………………..……….…….. iv

ACKNOWLEDGMENTS…...…………………………………..…. v

TABLE OF CONTENTS…………………………………………..…. vi

LIST OF FIGURES………………………………………………….…….. viii

LIST OF TABLES…………………………………………………… ix

CHAPTER

I. INTRODUCTION………………………………………… 1

II. LITERATURE REVIEW………………………………….. 4

Introduction…………………………………………….…... 4
Enthalpic Contribution to Protein Stability………………. 4
Buried Charge and Protein Folding………………………. 5
Destabilization of Larger Proteins………………………... 7
Large Protein Destabilizing Effects……………………..... 8
Conclusion………………………………………………... 9

III. METHODS………………………………………………… 10

PFIS Methods……………………………………….…..... 13
CHARMm_19 Methods……………………………….. 14

IV. RESULTS…………………………………………………….. 16

Side Chain Burial……………………………………….….. 16
Hydrogen Bonding………..……………………………… 16
Polar and Charged Burials………………………………... 21
van der Waals Contributions…………...………………… 22

V. DISCUSSION……………………………………………… 24

Nature's Need for Large Protein Instability..…………….… 24
Overall Side Chain Burial Effects...……………………… 25
Hydrogen Bonding Effects...…………...………………… 25
van der Waals Contribution Effects with Partial Charges... 27
van der Waals Contribution Effects with Partial and Full Charges…………………………………………………… 29
Summary……………………………………………………. 30

REFERENCES…………………………………...………… 31

APPENDIX……………………………………………………... 34

CURRICULUM VITA……………………………………………….. 35

LIST OF FIGURES

FIGURE		Page
1	Ribbon diagram of B1 Immunoglobulin-Binding Protein G (1PGB)………………………………………………………...	11
2	Ribbon diagram of Ubiquitin (1UBQ)…………...……………...	11
3	Ribbon diagram of Bacteriophage T4 Lysozyme (2LZM)...……...	12
4	Ribbon diagram of Maltodextrin-Binding Protein (1OMP)..........	12
5	Percent burial of charged, polar, and nonpolar side chains vs. number of residues………………………………………………	26
6	Number of hydrogen bonds vs. number of residues……………..	27
7	van der Waals contribution to stability with only partial charges included vs. number of residues……………...........................	28
8	van der Waals contribution to stability with both partial and full charges included vs. number of residues………………………...	30

LIST OF TABLES

TABLE		Page
1a	Burial of nonpolar side chains, peptide groups, polar groups and ionizable groups in the folding of 1YRF, 1PGB, 1UBQ, and 1RGG…………………………………………………………..	18
1b	Burial of nonpolar side chains, peptide groups, polar groups, and ionizable groups in the folding of 1A2P, 2LZM, 1L8W, and 1OMP…………………………………………………………..	19
2	Estimates of the major forces contributing to protein stability…...	20
3	Polar and Charged Group Burial……………………..………..	21
4	van der Waals contributions for the native and extended structures with only partial charges included ……….…………....	23
5	van der Waals contributions for the native and extended structures with both partial and full charges included……………	23

I. INTRODUCTION[1]

The chain of amino acids making up a protein must fold, in the presence of water, to a unique, roughly spherical, three-dimensional structure, called the native state, in order to carry out their biological functions. These proteins are called globular proteins and are by far the most abundant class of proteins (Pace et al. 2005). The folded proteins pack most of their side chains inside to form a hydrophobic core. The other hydrophilic or polar side chains are on the surface of the protein so that they can interact with the water. Based on research performed over the past 50 years, large proteins should be more stable than small proteins, but this is not observed. In fact, small and large proteins have about the same stability, therefore it is important to determine how and why nature favors the destabilization of larger proteins.

It is now known that proteins could have become more stable, but they do not. Evolution must have strategies to keep large proteins from becoming too stable (Pace et al. 2005). How does nature destabilize big proteins? Answering this question will

---

This thesis follows the style and format of *Protein Science*.

provide a better understanding of how nature fine tunes protein stability. More specifically, what forces are used to destabilize larger proteins so that they do not become too stable for their biological function?

Protein folding is the process by which a protein forms first secondary structure and then tertiary structure, leading to a functional protein. The conformational stability of a globular protein is defined as the free energy, ΔG, difference between the native folded state and the denatured unfolded state under physiological conditions. The reaction is:

Folded (Native) ⇔ Unfolded (Denatured)

In this project I have focused on a group of 8 globular proteins. These proteins start with the smallest having 36 residues and then increase in size to the largest at 370 residues. Sometimes, proteins have more than one globular unit in a single polypeptide chain and these units are then called domains. In this study, I have analyzed only single domain proteins. The forces and interactions which I analyzed extensively are the side chain interactions of the charged, polar, and nonpolar side chains. I also estimated the contribution to the conformational stability of all eight proteins of van der Waals

interaction. With this data, I was able to determine the contribution of other important factors such as the hydrophobic effect and hydrogen bonding, and with these estimate the stability of the proteins.

There are two main questions that I will be considering. One, why is it necessary to destabilize larger proteins. Two, what is the primary factor used to destabilize larger proteins?

II. LITERATURE REVIEW

Introduction

The central goal of this project is to determine how nature fine tunes protein stability. I will be duplicating methods which were used by other researchers to perform specific calculations which deal with protein folding. Past research suggested the importance of charge burial and why nature does not want more stable large proteins. One project has even focused on the effects of destabilizing larger proteins. However, this research had a small sample and only analyzed two proteins. These different studies have been conducted at various locations around the world; California, Massachusetts, Texas and all the way to Finland.

Enthalpic Contribution to Protein Stability

Themis Lazaridis and fellow co-workers wrote a paper which examined the basic energy effects that cause proteins to fold into their native states. The 1995 paper, *Enthalpic Contribution To Protein Stability: Insights From Atom-Based Calculations And Statistical Mechanics* is one of the first projects to actually perform detailed calculations associated with protein folding. These calculations were used to understand the protein folding interactions such as enthalpy, entropy and free energy. This project

makes no references to the reason why proteins need to be destabilized, especially the larger ones. However, since this paper was one of the first to perform the necessary computational calculations, I duplicated many of their methods.

I used the same computer program, CHARMm_19 which was written at Harvard University. One of the most important things that I determined from there paper was that I needed to calculate the energy of not only the native structure but also the unfolded state to determine the overall energy difference of protein folding. If the unfolded state calculations were not performed then the total energy calculations would be ambiguous. Using Lazaridis' methods, the most important calculated energies that I determined for the native proteins and unfolded chains were the total energy, bond energy, Coulombic energy and van der Waals energy. However, for this particular project I was most interested in the van der Waals contributions.

Buried Charge and Protein Folding

In November 2000, multiple scientists led by Tommi Kajander published a paper which dealt with the stabilizing and destabilizing effects of buried charges during protein folding. Their paper, *Buried Charged Surface in Proteins* analyzes the work they performed to determine that buried charge is a more important factor in the energetics of

protein folding, than previously acknowledged. They extended previous research by showing that buried charge is much more common then generally thought. They worked with a large sample of proteins, with a wide range of molecular weights. They showed that the amount of buried charge increases with protein size but, this becomes apparent only in larger proteins. Also, they discussed the particular roles of buried charge in the protein such as; structural roles in forming hydrogen bonds, electrostatic interactions, forming folding intermediates, role in activity, role in allostery and contributing to the destabilizing effects of proteins. This project suggests that proteins do have specific electrostatic arrangements because the protein must be organized to accommodate the ionic groups to increase or decrease the stability of the protein.

This paper only briefly considers exactly which buried charges contribute to stabilizing and destabilizing proteins. For example, the paper determines that, "Buried charged residues in a native structure can be stabilizing with respect to the denatured state of the protein in which the charge is not buried" (Kajander 2000). On the other hand, they discuss that in destabilizing environments such as in or near the active site there are numerous ions and ionic groups which contribute to the catalytic function. This project does acknowledge the lack of information of protein folding and the actual

effects of the charge burials. They state, that the burial of the charged groups may provide a way to fine tune the balance between stability and instability. Different proteins require different amounts and types of stabilization and/or destabilization. There needs to be more extensive research with a larger sample of proteins, that includes more larger proteins.

Destabilization of Larger Proteins

Peter Fields was one of the first researchers to determine that nature does in fact, disfavor the stability of larger proteins. In his 2001 paper, *Protein function at thermal extremes: balancing stability and flexibility* Fields points out that the maintenance of an appropriate balance between molecular stability and structural flexibility is the key to protein function. Fields touches on some past research and states that, "Researchers have found no new amino acids, covalent modifications or structural motifs that explain the ability of [certain] molecules to function in such harsh environments." (Fields 2001). With this, Fields believes that slight redistributions of the same intramolecular interactions, at different temperatures, must cause protein stabilization. There are multiple reasons that nature needs stable proteins, for instance, to ensure the appropriate

geometry for ligand binding or to avoid denaturation. On the other hand, flexibility is essential to allow catalysis at a metabolically appropriate rate.

Field's paper determined the need for protein stability and/or flexibility when dealing with small and large proteins at different temperatures. However, my research extends Field's project by determining what natural effects cause the destabilization of larger proteins.

Large Protein Destabilizing Effects

In 2005, C. Nick Pace and co-workers wrote a paper titled, *Fine Tuning Protein Stability*. Their research concentrated on the effects that stabilize and destabilize proteins, which is basically the same project that I am researching. However, their study considered only two proteins, one being relatively small with 36 residues and the other being one of the largest, single-domain globular proteins with 340 amino acids. Very similar to their study, my protein sample contains the same small and large globular proteins but I also added 6 of intermediate size. I wanted to investigate the stability of different proteins as a function of their sizes. This should enable me to determine how nature destabilizes proteins as they increase in size.

Conclusion

Using different types of computational methods, I will try and determine the natural effects which fine tune protein stability. In the past 5 years, research has shown us that nature does in fact want stable proteins, but at the same time, some proteins need to be flexible. The goal of my research was to evaluate the actual forces that cause larger proteins to become destabilized and therefore capable of performing their biological functions.

III. METHODS

The atomic coordinates for the atoms in the proteins were downloaded from the Research Collaboratory for Structural Bioinformatics Protein Data Bank, RCSB PDB (Berman et al. 2000). The proteins in order of increasing size and their PDB filenames are as follows:

1. Villin Head Piece: (1YRF) 36 Residues
2. B1 Immunoglobulin-Binding Domain Protein G: (1PGB) 56 Residues (Figure 1)
3. Ubiquitin: (1UBQ) 76 Residues (Figure 2)
4. Guanyloribonuclease: (1RGG) 96 Residues
5. Barnase Wildtype: (1A2P) 108 Residues
6. Bacteriophage T4 Lysozyme: (2LZM) 164 Residues (Figure 3)
7. Lyme Disease Variable Surface Antigen: (1L8W) 295 Residues
8. Maltodextrin-Binding protein: (1OMP) 370 Residues (Figure 4)

These protein crystal structures were determined through X-ray crystallography and/or NMR spectroscopy by, Chiu et al. (2005), Gallagher et al (1994) and Ferreon and Bolen

(2004), Vigay-Kumar et al. (1987), Pace et al. (1998), Mauguen et al. (1982), Weaver and Matthews (1987), Eicken et al. (2002), and Sharff et al. (1992), and respectively.

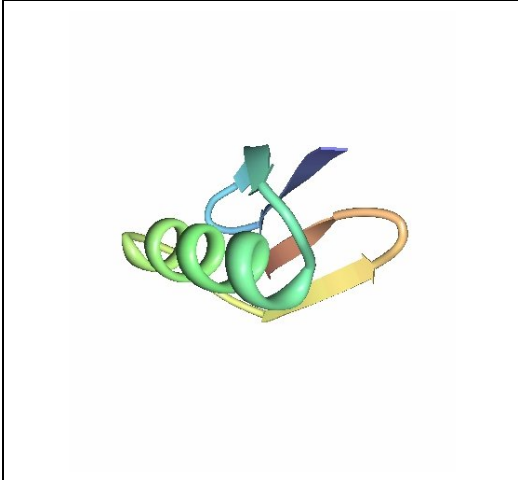

Figure 1. Ribbon diagram of B1 Immunoglobulin-Binding Protein G (1PGB)

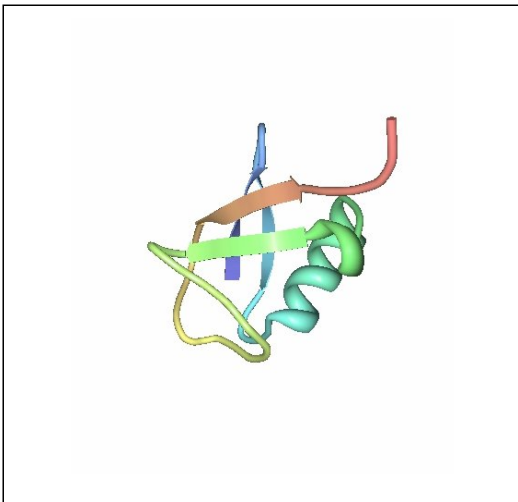

Figure 2. Ribbon diagram of Ubiquitin (1UBQ)

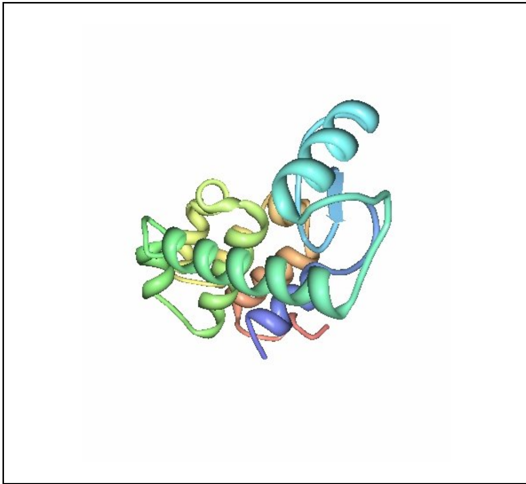

Figure 3. Ribbon diagram of Bacteriophage T4 Lysozyme (2LZM)

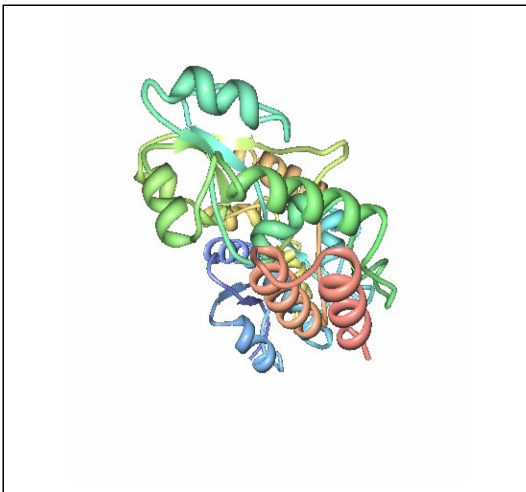

Figure 4. Ribbon diagram of Maltodextrin-Binding Protein (1OMP)

PFIS Methods

The first computational methodology used to study protein stability was an in-house computer program. The program, called PFIS, was written by Eric J. Hebert in 1998 at Texas A&M University and runs on UNIX, which is a common operating system. The PFIS program was used to obtain the conformational stability, hydrogen bonding, and accessible surface area of the proteins of interest.

Through the accessible surface area calculation summary, I was able to determine the number of ionizable groups and the percentage of these charged side chains which are buried within the hydrophobic core. By analyzing the conformational stability results I was able to find the number of nonpolar side chains along with their percent burial. This is the same for the peptide groups and polar side chains along with their burials within the protein core. By multiplying the number of nonpolar side chains by their specified ΔG values (Pace 1995) I was able to do energy calculations for the hydrophobic effect. The contribution of the hydrophobic effect to the stability is the sum of the contributions of the individual side chains. For a more complete description of this see Pace et al. (1998).

The hydrogen bonding results provides the total number of hydrogen bonds within each protein and was used to estimate their contribution to the estimated stability. The hydrogen bonding summary also analyzed the main-chain to main-chain, main-chain to side-chain, and the side-chain to side-chain hydrogen bonds. This helped to determine the number of polar and charged side chains which were forming hydrogen bonds.

CHARMm_19 Methods

The second computational methodology used was the CHARMm_19 program as implemented in the Insight II 2005 Suite of Software (Accelrys 2001-2007). I wanted to analyze the van der Waals contribution to the energy within the native and extended structures with partial and full charges. To calculate these energies, I followed the procedure of Lazaridis et al. (1995) on the proteins of interest. I repeated their steps by using the X-ray crystal structures and/or NMR spectroscopy for the native state and extended chain models.

As in Lazaridis' research, I added polar hydrogen positions to the native protein by using the HBUILD program of CHARMm. To relieve any possible bad interactions, I minimized the structures using 300 steepest descent steps. When constructing the

extended protein I followed the same guidelines (Lazaridis et al. 1995). I extended the chains with Φ = -140 and ψ=135, and minimized the structures with 300 steepest descent steps. Any resulting deformations by the prolines were repaired by local rearrangements during the minimizations.

It was unclear how the cutoffs were implemented by Lazaridis et al. Consequently, I used a cut off of 999 which is the maximum allowed in CHARMm. The partial charges for both the native and extended structures were added so that each side chain was neutral. Van der Waals energies in kcal/mol were then calculated. I then used the full charges for both structures and again obtained the van der Waals energies. Using the van der Waals contributions in the native and extended structures I was able to determine their contribution to the conformational stability.

I did have a problem with the Lyme Disease Variable Surface Antigen (1L8W) protein. I was not able to obtain the van der Waals contributions for this protein even after numerous attempts at troubleshooting CHARMm_19. Therefore, my data for the second methodology is incomplete because it does not contain the information for the 1L8W protein.

IV. RESULTS

Side Chain Burial

The data using the first methodology, PFIS, is given in Tables 1-3. Obviously, as the protein increases in size there is an increase in the number residues. However, are there more nonpolar, polar, or charged groups? Nonpolar side chains are defined as alanine, cysteine, isoleucine, leucine, methionine, phenylalanine, proline, tryptophan, tyrosine, and valine. I determined the number of these residues along with the number buried in each protein (Table 1a/1b). Polar side chains are defined as: serine, threonine, tyrosine, asparagine, and glutamine. I again determined the number of these residues along with the number buried in each protein (Table 1a/1b). Ionizable side chains are defined as: aspartic acid, glutamic acid, arginine, histidine, and lysine. I found the number of these residues along with the number buried in each protein (Table 1a/1b). I also analyzed the number of peptide groups along with their burial within all 8 proteins

Hydrogen Bonding

The estimates of the major forces contributing to protein stability are provided in Table 2. The hydrophobic effect energies were calculated with the use of the ΔG values

from Pace et al. (1995). The total hydrogen bonds in each protein were given by the PFIS hydrogen bonding summary. It is assumed that each hydrogen bond contributes 1 kcal/mole to stability. For example, Maltodextrin-Binding protein (1OMP) has a total of 337 hydrogen bonds and thus hydrogen bonding contributes 337 kcal/mole to the total stability of the protein. This same basic idea was used to find the contribution of conformational entropy to the conformational stability (Table 2). It is assumed that for every residue, conformational entropy contributes 1.7 kcal/mole towards total protein instability. The net estimated stability was calculated with the following formula:

Estimated Stability = Hydrophobic Effect + Hydrogen Bonding - Conformational Entropy

The measured stabilities given in Table 2 were provided by Pace (2005). I also modified the predicted stabilities by taking into account the cost of burying charged groups and this is referred to as the Revised Predicted Stabilities. The specifics for finding the revised predicted stabilities are given in the Table 2 caption. Note that the revised predicted stabilities are much closer to the measured stabilities than the predicted stabilities. This is the most important finding in this thesis. Larger proteins are destabilized mainly by burying more charged groups than small proteins.

Table 1a. Burial of nonpolar side chains, peptide groups, polar groups, and ionizable groups in the folding of 1YRF, 1PGB, 1UBQ, and 1RGG

	1YRF (36)		**1PGB (56)**		**1UBQ (76)**		**1RGG (96)**	
Nonpolar Side Chain	Number Present	Number Buried	Number Present	Number Buried	Number Present	Number Buried	Number Present	Number Buried
Alanine	3	1.2	6	4.1	2	0.9	6	3.2
Cysteine	0	0	0	0	0	0	2	1.3
Isoleucine	0	0	1	0.6	7	6.6	5	4.3
Leucine	5	3.2	3	2.7	9	7.4	6	5.1
Methionine	2	0.2	1	0.4	1	0.8	0	0
Phenylalanine	4	2.8	2	2	2	1.5	2	2.7
Proline	1	0.5	0	0	3	1.8	6	2.7
Tryptophan	1	0.3	1	0.8	0	0	0	0
Tyrosine	0	0	3	2.4	1	0.8	8	6.3
Valine	1	0.8	4	2.6	4	3.7	6	4.6
Total	17	9 (53%)	21	15.6 (74%)	29	23.5 (81%)	41	30.2 (74%)
Peptide Groups	36	24.7 (68%)	56	40.5 (72%)	76	54.32 (71%)	96	68.67 (71%)
Side Chain Polar Groups[a]	11	4.2 (38%)	22	11.4 (52%)	27	15 (56%)	42	26.99 (64%)
Ionizable Groups[b]	18	4 (22%)	29	11 (37%)	42	14 (33%)	41	22 (53%)

[a] The accessibility of the following polar groups were determined: the –OH group for the serine, threonine, and tyrosine side chains, and the O and NH_2 groups for the asparagine and glutamine side chains.

[b] The accessibility of the following charged groups were determined: both O atoms for the carboxyl groups, both NH_2 groups in the arginine side chain, both NH groups in histidine side chains, and the NH_2 groups in the lysine side chain and at the amino terminus.

Table 1b. Burial of nonpolar side chains, peptide groups, polar groups, and ionizable groups in the folding of 1A2P, 2LZM, 1L8W, and 1OMP

	1A2P (108)		**2LZM (164)**		**1L8W (295)**		**1OMP (370)**	
Nonpolar Side Chain	Number Present	Number Buried	Number Present	Number Buried	Number Present	Number Buried	Number Present	Number Buried
Alanine	7	5.3	15	11.5	58	45.6	44	36.3
Cysteine	0	0	2	1.8	0	0	0	0
Isoleucine	8	6.7	10	9.5	15	14.2	23	20.6
Leucine	7	6.1	16	13.9	16	14.4	30	28.7
Methionine	0	0	5	4.4	0	0	6	5.0
Phenylalanine	4	3	5	4.3	10	7.6	15	14.4
Proline	3	1.8	3	1.1	5	3.8	21	15.7
Tryptophan	3	2.6	3	2.6	0	0	8	7
Tyrosine	7	5.8	6	4.9	2	1.1	15	12.5
Valine	4	2.8	9	7.3	24	22	20	18.8
Total	43	34.1 (79%)	74	61.3 (83%)	130	109 (84%)	182	159 (88%)
Peptide Groups	108	79.33 (73%)	164	134.55 (82%)	295	252.7 (85%)	370	311.8 (84%)
Side Chain Polar Groups[a]	43	27.79 (65%)	57	32.82 (58%)	57	33.1 (58%)	108	74.3 (69%)
Ionizable Groups[b]	51	29 (56%)	80	43 (54%)	131	65.5 (50%)	156	80 (51%)

[a] The accessibility of the following polar groups were determined: the –OH group for the serine, threonine, and tyrosine side chains, and the O and NH_2 groups for the asparagine and glutamine side chains.

[b] The accessibility of the following charged groups were determined: both O atoms for the carboxyl groups, both NH_2 groups in the arginine side chain, both NH groups in histidine side chains, and the NH_2 groups in the lysine side chain and at the amino terminus.

Table 2. Estimates of the major forces contributing to protein stability*

Factor	1YRF	1PGB	1UBQ	1RGG	1A2P	2LZM	1L8W	1OMP
Number of Residues	36	56	76	96	108	164	295	370
Hydrophobic Effect[a]	54	84	146	147	178	307	463	708
Hydrogen Bonding[b]	27	54	65	87	106	167	258	337
Conformational Entropy[c]	60	95	129	163	184	279	461	629
Predicted Stability[d]	21	43	82	71	100	195	260	416
Revised Predicted Stability[e]	5	-1	26	-17	-16	23	-4	96
Measured Stability[f]	3.3	2.8	6.7	5.8	8.7	5.5	4.5	4.8

* Numbers are given in kcal/mole.

[a] The free energy of transfer, ΔG_{tr}, from water to cyclohexane was used to estimate the contribution of the burial of the nonpolar side chains to the conformational stability of the proteins. The ΔG_{tr} values are summarized in Pace (1995). The number of buried nonpolar side chains is given in Table 1. The contribution of the hydrophobic effect to the stability is the sum of the contributions of the individual side chains. (See Pace et al. (1998) for a more complete description of this and the following contributions)

[b] We assume that each hydrogen bond contributes 1 kcal/mole to stability.

[c] The conformational entropy was calculated assuming 1.7 kcal/mole per residue.

[d] Net estimated stability equals the sum of the contribution of the hydrophobic effect and hydrogen bonding, and the subtraction of conformational entropy.

[e] This was calculated by taking the net estimated stability and then subtracting the number of buried charges times a rough estimate of the cost of burying a charge (4 kcal)

[f] The measured stabilities were provided by (Pace 2005).

Polar and Charged Burials

Table 3 shows the results for polar and charged group burial. The methods used to obtain this data are provided in the Table 3 footnotes. One of the most important findings is that the burial of charged groups increases from 22% for the smallest protein to over 50% in the larger proteins. The burial of polar side chains does not increase as drastically.

Table 3. Polar and charged group burial

Group	1YRF	1PGB	1UBQ	1RGG	1A2P	2LZM	1L8W	1OMP
# Polar[a]	81	132	174	226	254	380	647	412
# Buried	52 (64%)	92 (69%)	124 (71%)	164 (73%)	186 (73%)	302 (79%)	540 (83%)	348 (84%)
# Ionizable[b]	18	29	42	41	51	80	131	156
# Buried	4 (22%)	11 (37%)	14 (33%)	22 (53%)	29 (55%)	43 (54%)	66 (50%)	80 (51%)

[a] The polar groups were determined as follows: the NH and O groups of the peptide bonds of all residues except proline were counted. Just the O group of the carbonyl of the proline residues was counted. The O and NH_2 groups for the asparagine and glutamine side chains and the OH groups of serine, threonine, and tyrosine side chains were counted.

[b] The accessibility of the following charged groups were determined: both O atoms for aspartic acid, glutamic acid and C-terminal carboxyls, both NH_2 groups in the arginine side chains, both NH groups in the histidine side chains, and the NH_2 groups in the lysine side chains and at the amino terminus.

van der Waals Contributions

The data for the second methodology, CHARMm_19 is given in Tables 4-5. I calculated the van der Waals contributions because they make an important contribution to protein stability. Table 4 provides the van der Waals contributions in kcal/mole for the native and extended structures with only the partial charges included. Adding partial charges to the proteins measures, in part, the contribution of hydrogen bonding. Table 4 also shows the conformational stability or free energy (ΔG) difference between native and extended structures. Table 5 gives the van der Waals contributions for the native and extended structures with both partial and full charges included. Full charges show the charge-charge interaction or electrostatic contribution to protein stability. Therefore, adding partial and full charges to the proteins estimates the total contribution to the stability of the protein. As in Table 4, I again show the conformational stability (free energy difference) between the native and extended structures. This is important because it analyzes how much more stable the native structure of a protein is than the unfolded or extended structure.

Table 4. van der Waals contributions for the native and extended structures with only partial charges included*

	1YRF	IPGB	1UBQ	1RGG	1A2P	2LZM	1OMP
Native Structure	248	418	580	709	820	1382	3241
Extended Structure	85	117	149	201	255	358	804
Difference between Native & Extended Structures	163	301	431	508	565	1024	2437

*Numbers are given in kcal/mole

Table 5. van der Waals contributions for the native and extended structures with both partial and full charges included*

	1YRF	IPGB	1UBQ	1RGG	1A2P	2LZM	1OMP
Native Structure	485	807	1141	1394	1573	2702	6331
Extended Structure	167	231	289	394	497	706	1584
Difference between Native and Extended Structures	318	576	852	1000	1076	1996	4747

*Numbers are given in kcal/mole

V. DISCUSSION

Nature's Need for Large Protein Instability

Why does nature favor destabilization of big globular proteins? In 2001, Fields was one of the first to consider why nature would want less stable larger proteins. "The key to protein function…is the maintenance of an appropriate balance between molecular stability on the one hand and structural flexibility on the other" (Fields 2001). There are numerous reasons why nature would want stable proteins. Fields states that the reasons might be to ensure appropriate geometry for ligand binding or to avoid denaturation. However, flexibility is essential to allow catalysis at a metabolically appropriate rate and also if the protein ultimately needs to be degraded (Fields 2001). The biological functions of the proteins require some flexibility and therefore it would be more difficult if the proteins were too stable. If necessary, the protein may need to be destabilized so that it can be degraded. This suggests that nature intentionally uses methods to assist in the destabilization of large proteins.

Overall Side Chain Burial Effects

I have plotted some of the important data from Table 1a and 1b. Figure 5 shows the percent burial of charged, polar, and nonpolar side chains versus the number of residues. For the most part, the graph shows that as the number of residues increase so does the percent burial of each of the side chains. Something important to note is how the polar and charged side chain burials have a significant increase around 100 residues and then slightly start to decrease. Also around this same protein size, the nonpolar side chains do the opposite. This protein size is similar to Guanyloribonuclease (1RGG) and Barnase Wildtype (1A2P) in my protein sample. I suggest that this research be extended to include more proteins in order to analyze these findings more extensively.

Hydrogen Bonding Effects

Figure 6 displays a graph of the number of hydrogen bonds versus the number of residues. The graph in Figure 6 shows that as the number of residues increase so does the number of hydrogen bonds. It has been estimated that on average there is one hydrogen bond per residue (Pace et al. 2005). However, I have determined that for my sample of proteins there are approximately 1.4 hydrogen bonds per residue. This is a

very interesting finding because it could ultimately indicate that hydrogen bonding contributes more to protein stability than has previously thought.

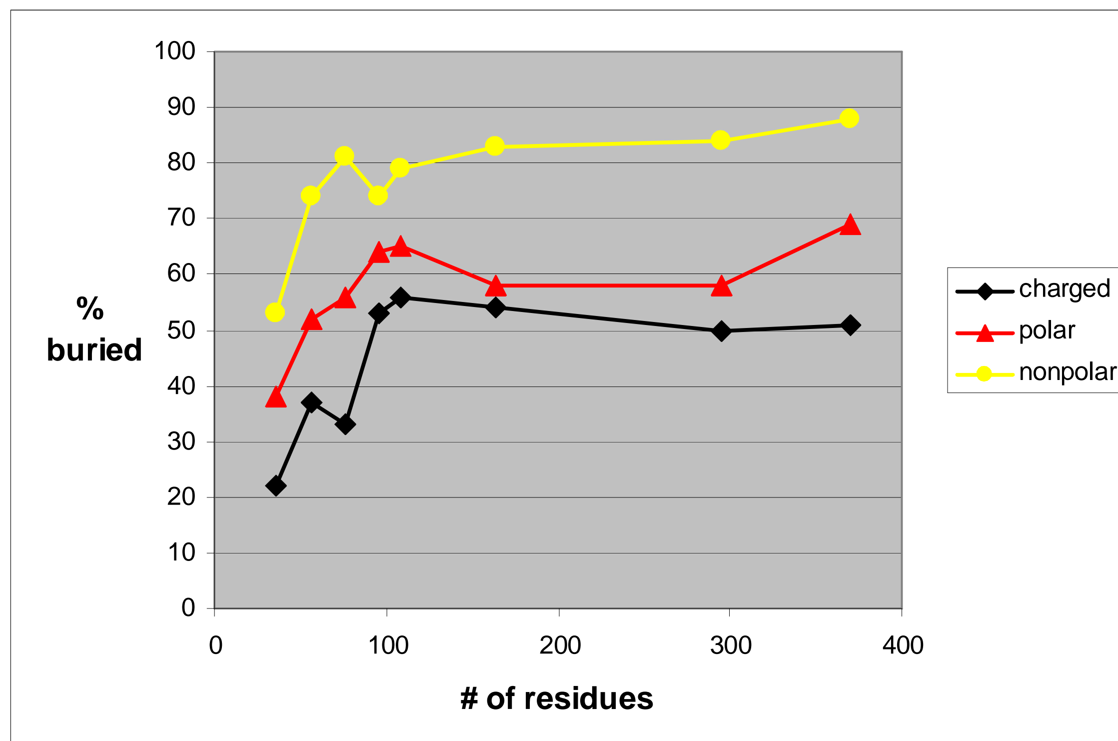

Figure 5. Percent burial of charged, polar, and nonpolar side chains vs. number of residues

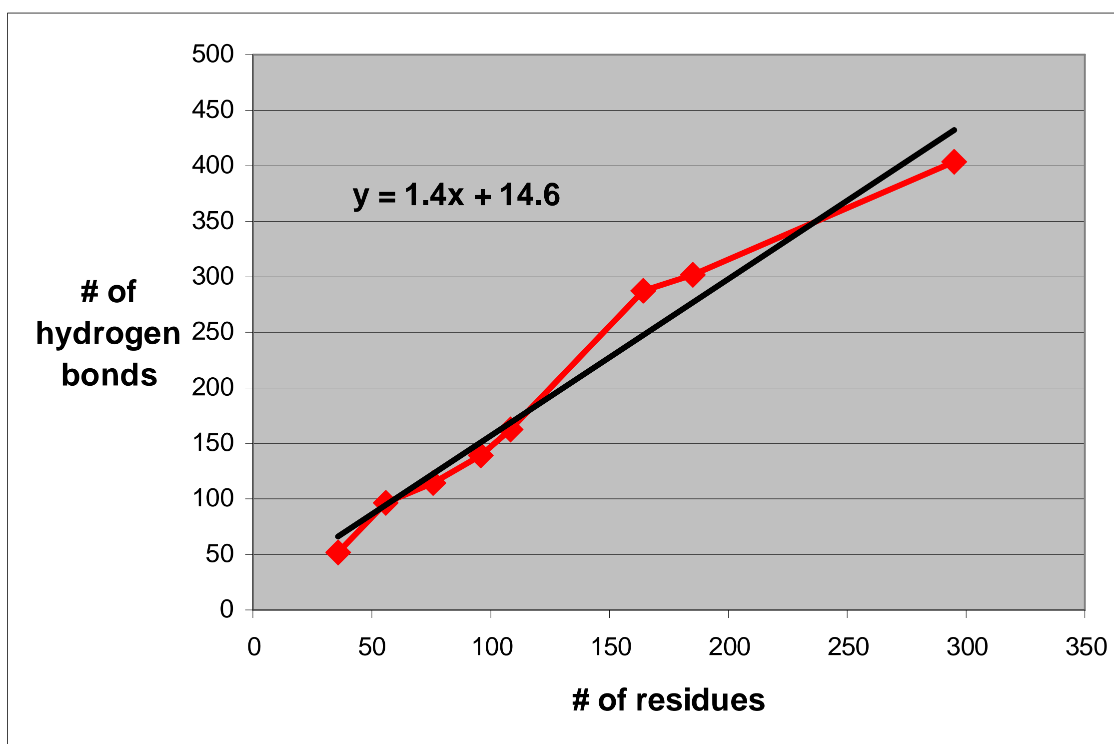

Figure 6. Number of hydrogen bonds vs. number of residues

van der Waals Contribution Effects With Partial Charges

Figure 7 displays the van der Waals contributions versus the number of residues. More specifically, it displays the conformational stability or free energy (ΔG) difference of the 7 proteins used in the second methodology. Figure 7 shows that as the number of residues increase so does the van der Waals contributions. First, this graph shows that the native state of each protein is always much more stable than the extended unfolded state. This is due to protein folding and the formation and packing of a stable

hydrophobic core. Second, the graph shows that the larger proteins are predicted to be much more stable than the smaller proteins. But experimental studies have shown that larger proteins are not significantly more stable than smaller proteins.

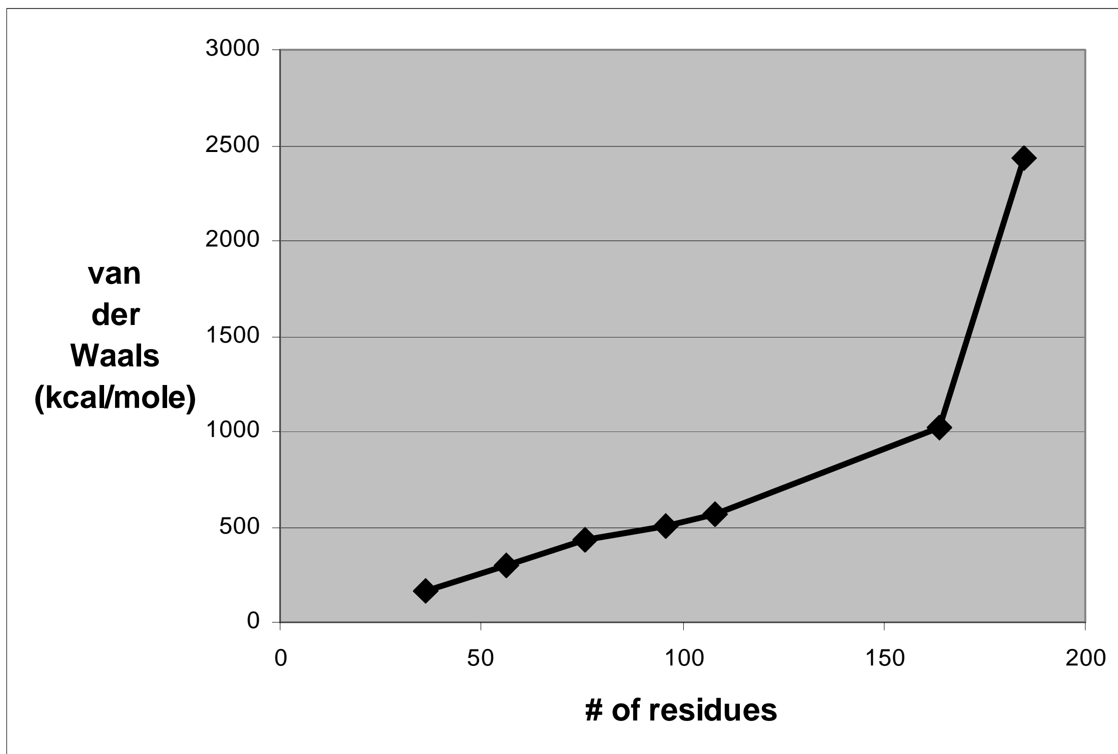

Figure 7. van der Waals contribution to stability with only partial charges included vs. number of residues

van der Waals Contribution Effects with Partial and Full Charges

Figure 8 is similar to Figure 7 in that it displays a graph with the van der Waals contributions versus the number of residues. However, this graph shows the contributions to protein stability with the addition of full and partial charges. Again, this shows the conformational stability or the energy difference between the native and extended structures. The graph in Figure 8 also shows that with an increase in residue number there is a continuing increase in the van der Waals contributions. The van der Waals interactions in Figure 8 are approximately double those in Figure 7. This shows that partial charges and full charges make similar contributions to protein stability. Partial charges show the contribution of hydrogen bonds and full charges show the contribution of electrostatic interactions. Therefore, hydrogen bonding and electrostatic interactions both contribute significantly to protein stability in about the same amount.

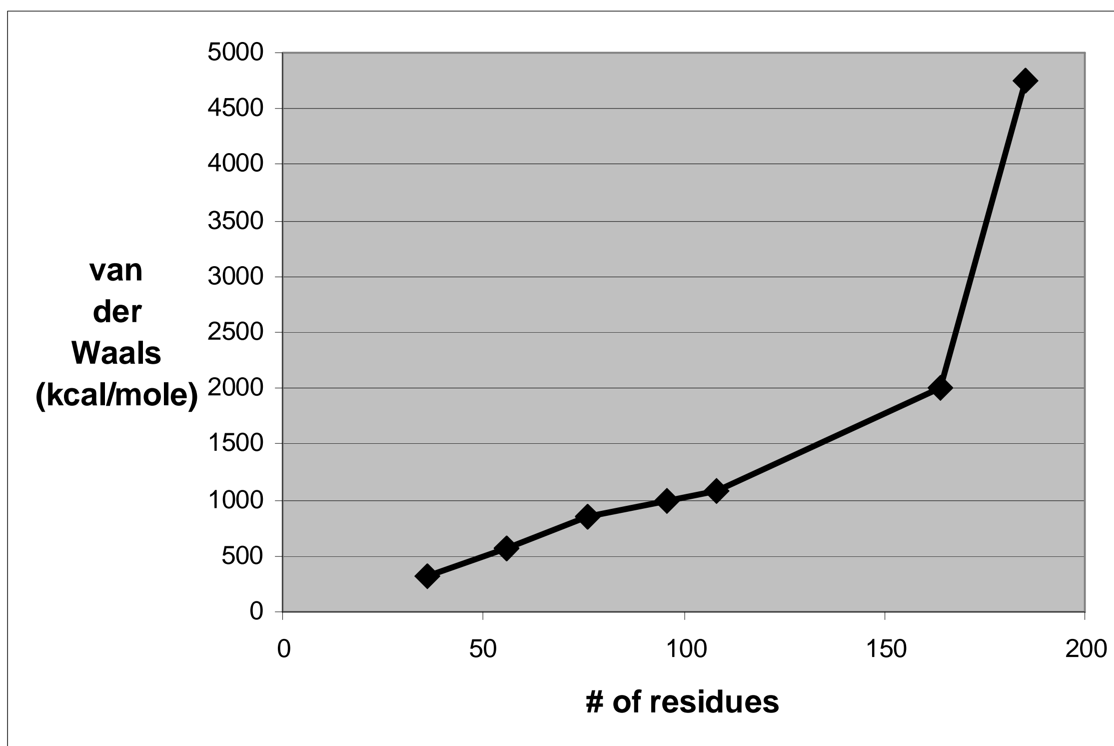

Figure 8. van der Waals contribution to stability with both partial and full charges included vs. number of residues

Summary

By comparing the stability and structure of my 8 proteins, I feel that I have gained a better understanding of the means used to keep larger proteins from becoming too stable. This project was aimed at determining the chief tactic used to destabilize proteins. My hypothesis was that burying charged groups in the interior of globular proteins has the greatest effect on big protein destabilization. The importance of charge burial was first cited by Kajander et al. in their 2001 paper. "Charge burial may provide

a way to fine tune the balance between stability and instability both globally and locally."

I believe that my results show that burying charged groups in the interior of the protein is the primary strategy used to fine tune protein stability. Based on my results, I recommend that more extensive and involved research in this area of protein destabilizing forces be conducted. I recommend an extension of this project to include a larger protein sample with more larger proteins included in the sample. This will provide more conclusive evidence that charged side chain burial is the primary strategy used to destabilize proteins.

REFERENCES

Accelrys. 2001-2007. Insight II: Suite of Software.

Berman M., J.W., Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, P.E. Bourne. 2000. The Protein Data Bank. *Nucleic Acids Research* **28:** 235-242.

Chiu T.K., J.K., Regine Herbst-Irmer, William A. Eaton, James Hofrichter, and David R. Davies. 2005. High-resolution x-ray crystal structures of the villin headpiece subdomain, an ultrafast folding protein. *Proceedings of the National Academy of Sciences* **102:** 7517-7522.

Eicken C., V.S., Thomas Klabunde, Matthew B. Lawrenz, John M. Hardham, Steven J. Norris, and James C. Sacchettini. 2002. Crystal Structure of Lyme Disease Variable Surface Antigen V1sE of *Borrelia burgdorferi*. *The Journal of Biological Chemistry* **277:** 21691-21696.

Ferreon A.C.M., B.D.W. 2004. Thermodynamics of Denaturant-Induced Unfolding of a Protein That Exhibits Variable Two-State Denaturation. *Biochemistry* **43:** 13357-13369.

Fields P. 2001. Protein function at thermal extremes: balancing stability and flexibility. *Comparative Biochemistry and Physicology* **129:** 417-431.

Gallagher T., P.A., Philip Bryan, and Gary L. Gilliland. 1994. Two Crystal Structures of the B1 Immunoglobulin-Binding Domain of Streptococcal Protein G and Comparison with NMR. *Biochemistry* **1994:** 4721-4729.

Hebert, E.J. 1998. PFIS Program. *Biochemistry* **37:** 271-286.

Kajander T., P.C.K., Sari Helin Passila, Daniel C. Cohen, Lari Lehito, William Adolfsen, Jim Warwicker, Ursula Schell, and Adrian Goldman. 2000. Buried Charged Surface Proteins. *Structure* **8:** 1203-1214.

Lazaridis T., G.A., M.K. 1995. Enthalpic Contribution To Protein Stability: Insights From Atom-Based Calculations And Statistical Mechanics. *Advances in Protein Chemistry* **47:** 231-306.

Mauguen Y., R.W.H., E. J. Dodson, G.G. Dodson, G. Bricogne, Cyrus Chothia and Jack A. 1982. Molecular structure of a new family of ribonucleases. *Nature* **297**.

Pace C.N., A.R., and J. Martin Scholtz. 2005. Fine Tuning Protein Stability. Texas A&M Health Science Center, College Station.

Pace C.N., E.J.H., Kevin L. Shaw, David Schell, Valentin Both, Daniela Krajcikova, Jozef Sevcik, Keith S. Wilson, Zbigniew Dauter, Robert W. Hartley and Gerald R. Grimsley. 1998. Conformational Stability and Thermodynamics of Folding of Ribonucleases Sa, Sa2 and Sa3. *Journal of Molecular Biology* **279:** 271-286.

Pace, C.N. 1995. Evaluating Contribution of Hydrogen Bonding and Hydrophobic Bonding to Protein Folding. *Methods In Enzymology* **259:** 539-554.

Sharff A.J , L.E.R., John C. Spurlino, and Florante A. Quiocho. 1992. Crystallographic Evidence of a Large Ligand-Induced Hinge-Twist Motion between the Two Domains of the Maltodextrin Binding Protein Involved in Active Transport and Chemotaxis. *Biochemistry* **31:** 10657-10663.

Vijay-Kumar S., C.E.B., W.J.C. 1987. Structure of Ubiquitin Refined at 1.8 Angstroms Resolution. *Journal of Molecular Biology* **194:** 531-544.

Weaver L.H., M.B.W. 1987. Structure of Bacteriophage T4 Lysozyme Refined at 1.7 Angstroms Resolution. *Journal of Molecular Biology* **193:** 189-199.

APPENDIX

Copyright permission did not need to be obtained for the use of Figures 1-4. The following was provided by an RCSB representative:

The contents of Protein Database (PDB) are in the public domain. Online and printed resources are welcome to include PDB data and images from the RCSB PDB website. They may be sold, as long as the images and data are not for sale as commercial items themselves. However, it is expected that the authors of an entry as well as the RCSB PDB be properly cited whenever their work is referred to.

http://www.rcsb.org/pdb/home/home.do

CURRICULUM VITA

MEGAN MARIE WICKSTROM
aggie.megs@tamu.edu

EDUCATION

Texas A&M University	Bachelor's of Science Degree
College Station, Texas	Major: Biology
Expected Graduation: August 2007	GPR: 3.854/4.0
University of Texas Arlington	Bachelor's of Science Degree
Arlington, Texas	Major: Biology
Transferred Date: May 2005	GPR 3.61/4.0

LEADERSHIP EXPERIENCE

Order of Omega (Greek Honor Society), Texas A&M University

- President (2006-present), Treasurer (2005-2006)
- Recruited, interviewed, accepted and initiated the new fall members. Monitored all events including socials, philanthropy events and Greek Gala.

Delta Delta Delta, University of Texas Arlington

- Vice President (2005), Secretary (2004)
- Served as Head of Standards (disciplinary sector) and preserved all rituals.

HONORS, SCHOLARSHIPS AND AWARDS

- Academic Achievement Scholarship (2006-2007)
- College of Science Academic Excellence Award (2003-2004, 2004-2005)
- Freshman Honors List for Academic Year 2003-2004
- Honors College Presidential Scholarship (2004)
- Honors Incentive Award (2006-2007)

- Hazelwood Scholarship for Transfer Students (2005-2006)
- National Dean's List (2003-2004, 2004-2005)
- Phi Kappa Phi Honor Society (2007)

ORGANIZATIONS AND HONOR SOCIETIES

- Alpha Epsilon Delta: Pre-Medical Honor Society (2005-2007)
- Honors Invitational Program Leader (April 2006-April 2007)
- Greek Ambassadors (April 2006-April 2007)
- Golden Key Honor Society (2006-2007)
- The National Society of Collegiate Scholars (2004-2007)
- Phi Kappa Phi Honor Society (2007)
- Sigma Xi: The Scientific Research Society (2007)

VOLUNTEER

- John Peter Smith Hospital: Inpatient Care Unit (Dec 03-Aug 05)
- Medical Ministry International Mission Trip: Lima, Peru (Spring Break 05)
- College Station Medical Center: Intensive Care Unit (Aug 05-Dec 05)
- Southern Care Hospice (June 06-Jan 07)
- Meals on Wheels (Feb 07-present)