2010-06-29

The Structure and Stability of Alpha-Helical, Orthogonal-Bundle Proteins on Surfaces

Shuai Wei
Brigham Young University - Provo

The Structure and Stability of Alpha-helical, Orthogonal-bundle Proteins on Surfaces

Shuai Wei

A thesis submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

Thomas A. Knotts IV, Chair
Dean R. Wheeler
W.Vincent Wilding

Department of Chemical Engineering

Brigham Young University

August 2010

ABSTRACT

The Structure and Stability of Alpha-helical, Orthogonal-bundle Proteins on Surfaces

Shuai Wei

Department of Chemical Engineering

Master of Science

The interaction of proteins with surfaces is a major problem involved in protein microarrays. Understanding protein/surface interactions is key to improving the performance of protein microarrays, but current understanding of the behavior of proteins on surfaces is lacking. Prevailing theories on the subject, which suggest that proteins should be stabilized when tethered to surfaces, do not explain the experimentally observed fact that proteins are often denatured on surfaces. In an attempt to develop some predictive capabilities with respect to protein/surface interactions, it was asked in previous works if the stabilization/destabilization of proteins on surfaces could be correlated to secondary structure and found that no link existed. However, further investigation has revealed that proteins with similar tertiary structure show predictable stabilization patterns. In this research, it is reported how five, alpha-helical, orthogonal-bundle proteins behave on the surface compared to the bulk. By measuring stabilization using melting temperatures and the Gibbs energies of folding, it is shown that the stability of proteins tethered to surfaces can be correlated to the shape of the loop region where the tether is placed and the free rotation ability of the part of proteins near surfaces. It is also shown that any destabilization that occurs because of the surface is an enthalpic effect and that surfaces always stabilize proteins entropically. Furthermore, the entropical stabilization effect comes from unfolded states of the tethered protein, while the enthalpical destabilization effect is from the folded states of protein. A further analysis of surface induced change of folding mechanism is also studied with a multi-state protein 7LZM in this research. The result showed that by tethering a protein on a surface, the melting temperature of part of the protein changed, which leads to a miss of state.

Keywords: simulation, thermodynamics, tertiary structure, interaction, protein microarray

ACKNOWLEDGMENTS

In the first place, I want to address deepest appreciation to Dr. Thomas A. Knotts IV for his amazing ideas, advice and supervision to my project and thesis. Dr. Knotts introduced me to the field of bio-simulation. His enthusiasm and creativity regarding to the research inspired me to acquire more and more fundamental knowledge in the protein world. Under his supervision, the research worked so organized and productive. Working in this lab was a wonderful experience in my life. Without Dr. Knotts guidance and persistent help, it would have been impossible for me to finish this thesis.

I would like to thank my committee members for their precious time and advice for my project. I want to thank Terry Schmitt for teaching me a lot of simulation techniques to make the work so interesting.

I also would like to thank my wife Wei Zhou and my parents for their support and encouragement all the time. With their support, all difficulties turned out to be so easy to overcome. As a graduate student major in molecular biology, Wei helped me so much for understanding those complicated biology processes. I appreciate all her effort and ideas for helping me out.

Contents

List of Figures

List of Tables

1 PROTEIN FOLDING AND STABILITY ON PROTEIN MICROARRAYS

1.1 Introduction

A protein microarray is a high-throughput diagnostic device that can perform thousands of biological assays in parallel. This technology is developing as a powerful tool for proteomics and clinical applications in recent years since the emergence of the two papers by Macbeath and Schreiber[1] and Zhu *et al.*[2] in the beginning of this century. It is created by depositing proteins onto a solid substrate with a different type of protein located at each addressable point on the "chip" to facilitate identification. With proper fluorescent labeling, the technology can identify proteins in serum, determine concentrations, screen drug candidates, detect protein/ligand interactions, or ascertain function. [3, 4, 5, 6]

While protein arrays have great potential in both research and clinical settings, the technology is currently limited by poor performance.[7] It is difficult to obtain reproducible, quantitative results. As such, regulatory agencies are reluctant to approve, and end-users are reluctant to use, the technology in its current state.

The key to any protein array is to deposit protein on the surface in a way that preserves function. There are basically two different techniques through which proteins are deposited on surfaces. The molecule can either adsorb non-covalently to the substrate or can be tethered to the surface by covalent linkage. The main challenge using either method is that proteins can change conformation when bound or adsorbed to a surface[8, 9, 10]. Since protein structure leads directly to protein function, transformations that do occur prevent the proteins on the surface from producing the desired outcome. Covalent tethering is emerging as the favored method to create protein arrays, as conformation changes are more pronounced for adsorbed proteins, but significant challenges remain. The difficulty is that no method exists to predict à *priori* how a particular tethered protein will behave.

Thus rational design of protein arrays is not possible and ad hoc choices must be made about variables such as the location of the tether site on the protein or the type of surface to use.

Although many researches have been involved in understanding what affects protein behaviors on a chip in experimental and simulational method, there is still little known about the underlying biophysics. The lack of capabilities in predicting protein stabilities and obtaining protein structures are hurdles obstructing researchers from understanding protein behavior in presence of a surface. Simulational researches could be better choices because protein structures are available to observe and compare during simulations. Also, interesting properties of proteins on surfaces such as thermodynamic values, conformations, and stabilities are measurable, which provide us tools to evaluate theories in predicting protein behaviors on surfaces. Details about protein microarrays, previous experimental understanding and simulation and theoretical understanding are shown in the following section "Background". Analyses about simulation efficiency and details of properties calculations are discussed in chapter 2.

A coarse-grain model by Karanicolas and Brooks [11, 12, 13, 14, 15] is used in this research, which has been proven to be able to reproduce protein folding mechanisms and has the required simulation efficiency. Since previous results showed that proteins could be stabilized or destabilized on surfaces and protein stabilities can not be correlated to their secondary structure motifs, a hypothesis is stated as protein stabilities on surfaces could be correlated to their tertiary structure motifs. Five proteins, from the alpha-helical, orthogonal bundle motif, are simulated in this work to test the hypothesis.

The results in this work show, for the first time, that protein stability on surfaces can be correlated to tertiary structural motif for alpha-helical, orthogonal-bundle proteins. The important factors to consider when selecting a tether site are the shape of the loop region and the volume available for the protein to rotate on the surface. For loop regions that have large rotation volumes, sites can always be found which stabilize the protein. A thermodynamic analysis shows that proteins are always stabilized entropically when tethered to a surface and that any destabilization is an enthalpic effect. Taken as a whole, the results offer hope for rational prediction of protein surface interactions and a rigorous thermodynamic

understanding of the origins of stabilization/destabilization of surface. Simulation details and data analyses are shown in chapter 3.

Further research on a multi-states protein 7LZM revealed that the melting temperature of some of its parts could be changed when it is tethered to a surface. The change of melting temperature of some parts of the protein, affects the number of folding metastates, which leads to the change of folding mechanism. Details of simulation in the multi-states protein are talked about in chapter 4.

1.2 Background

1.2.1 Protein Microarray

According to a nomenclature proposed by Kodadek [16], there are two basic kinds of protein microarrays. They could either be employed for studies of native protein activities (Figure 1.1) or serve as an analytical tool of monitoring protein levels in a given biological sample (Figure 1.2), both in massively parallel fashions. In the first application, native ligands are arrayed in defined spots, and fluorescently labeled proteins in solution are tested through the array. Spots that 'light up' would be due to binding of the labeled protein. In the other procedure, one specific kind of ligand is arrayed on the chip, then levels of corresponding proteins in solution could be detected according to the binding level.

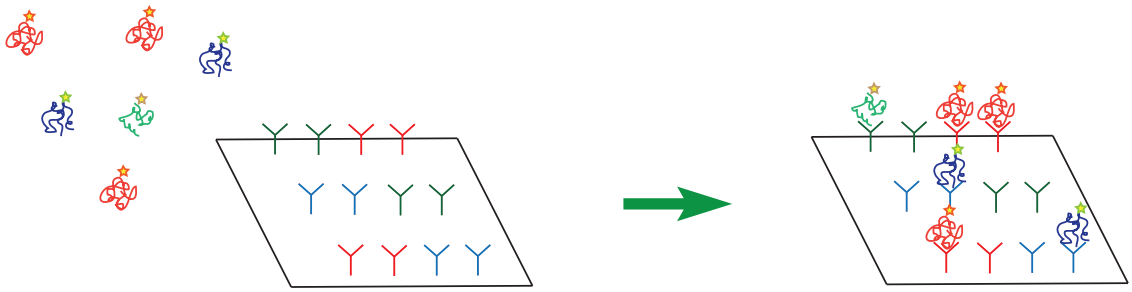


Figure 1.1: Protein microarray for studies of native proteins activities

Several platforms have been developed for measuring different parameters from a minute amount of sample.[7] Protein arrays have the potential to be applied in basic research, drug target discovery and validation, drug development, and clinical diagnostics. The first

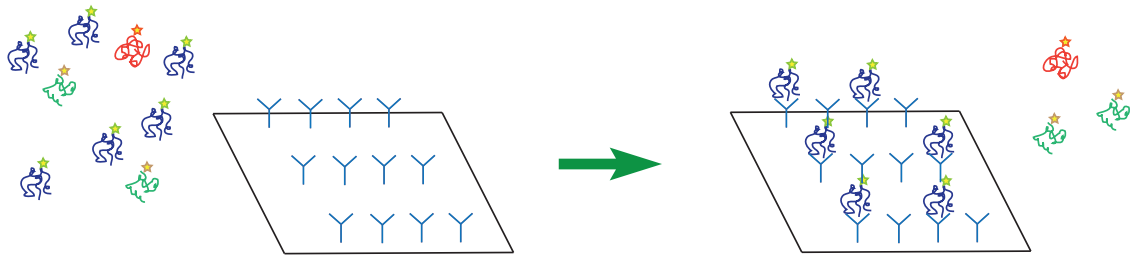


Figure 1.2: Protein microarray for monitoring protein levels

high-density antibody microarrays were studied by Haab *et al.*[17]; they were used to test whether a linear relationship could be detected between an antibody and antigen pair in an array format. Soen *et al.*[18] have fabricated one analytical microarray using peptide-MHC complexes to detect and characterize antigen-specific T-cell populations. Hsu *et al.*[19] have built up a lectin chip with 21 lectins for use in profiling the surface lipopolysaccharides in cacterial cells. The lectins were able to capture the bacterial cells onto the chip while labeled *E.coli* cells were incubated.

Despite being studied for several years, researchers are still far from being confident about protein arrays' performance. For example, results from antibody arrays are not always conclusive due to different arraying technologies.[20] Some antibodies have been shown to be active in standard assaying techniques such as ELISA, while the activity can not be measured on surfaces.[21] Also, signal intensities can vary as much as 43% on the same chip[22]. Moreover, since antibody arrays are the most advanced of the technologies[6], and antibody structure is fairly similar across the entire class of molecules, arraying other proteins, such as cytokines[20, 23, 24, 25], is even more challenging. In short, despite the promise of protein arrays as a clinical tool to improve quality of life, current technology cannot produce arrays that perform to levels commensurate for use in clinical settings.[20, 6, 26, 27, 28]

Thus, before protein microarrays become the mainstream tool for biological applications, some bottleneck problems need to be solved.[29] A major challenge of creating a protein microarray is that proteins always denature when deposited on the chip due to the effect from the surface [6, 8, 9, 10]. Since protein structures lead to protein functions, such transformation will prevent proteins from performing their desired roles. Since protein/surface

interactions are very important in processes like protein microarrays, many studies, either experimental or simulational, have been done to figure out what affect protein stabilities on the surface.

Nevertheless, there is still little known about the underlying biophysics, and prevailing theories have been shown to break down under careful scientic testing. Moreover, current ability to control and manipulate protein adsorption and function on surfaces is limited because current models cannot yet predict the behaviors of proteins at interfaces. Because maintaining protein stability on surfaces is essential to array function, several researchers have been involved in studying protein/surface interactions. These are described in the next section.

1.2.2 Experimental Understanding

Researchers have studied the behavior of polypeptides at interfaces for decades.[9] One major hurdle of implementing experimental studies for protein/ surface interactions on protein microarrays is the lack of a method to predict protein stabilities on surfaces. Generally two regimes of control are desired, the prevention of non-specific protein adsorption, and the precise placement of the proteins on the surface in a manner that preserves functionality. It is often necessary to combine both of them at the same time.

There are many kinds of methods for binding proteins on surfaces.[6, 30] The simplist one is the surface adsorption, which has been used in the standard enzyme-linked immunosorbent assay (ELISA) and Westen blot for many years. It is generally mediated by electrostatic charges[17] or hydrophobic interactions [31]. Despite its simplicity, the main drawbacks of this method are the possibility of denaturating proteins and non-specific protein adsorption on the surface. [6] Covalent binding of proteins to substrate surfaces is a more efficient and robust approach.[32] The surfaces usually carry reactive groups, such as epoxides, aldehydes, succinimidyl esters or isothiocyanates, which react with nucleophilic groups (e.g., amino, thiol or hydrocyl groups) of amino acid residues. Covalent immobilization via random attachment also tends to denature arrayed proteins. Researchers have also developed a method

of affinity interaction by specific tags, which provides a means of immobilizing proteins in a defined orientation on a tag-capture surface, often retaining full protein activity.[2]

Perhaps the most popular technology to control protein/surface interactions are self-assembled-monolayers (SAMs) of proteins, which are effective at preventing non-specific adsorption as well as directing desired protein placement on the surface.[33, 34] However, their inherent instability prevents their effective use outside of the research setting, particularly their use in medical diagnostics which require an extended shelf life.[35] Polymer coated surfaces, usually PEG-based, are used extensively to prevent fouling of surfaces and to control protein placement on the surface through appropriate functionalization of the polymer.

Despite the success of SAMs, polymers, and other coatings on surfaces, predicting protein behavior on surfaces, remains difficult. A striking example is the fact that some antibodies, which have a high degree of similarity from one molecule to the next, are active in solution but not on surfaces while others are not affected by the substrate.[21]

The other hurdle obstructing us from obtaining more experimental knowledge of protein surfaces interactions is that it is hard to determine protein structures. Typical techniques for obtaining structures of proteins such as NMR and X-ray crystallography, are not adaptable to surface-bound proteins. Some techniques, such as surface plasmon resonance (SPR), dual-polarisation interferometry (DPI), ellipsometry, circular dichroism spectroscopy (CD), and FTIR can be used to provide a gross estimate of protein structure but cannot provide mechanistic understanding or atomic-level structural resolutions.[36] Because of this, relatively little is known about how to predict the behavior of a protein on a surface *a priori*, or how to control the function of absorbed molecules.

1.2.3 Theoretical and Simulation Understanding

Since experimental methods cannot provide enough information, several groups have done simulations and theoretical work to understand protein/ surface interactions, both with atomistic and coarse-grain methods.

Some groups have implemented atomistic simulations of proteins on surfaces. For example, Latour and coworkers have investigated both model peptides [37] and biologically

relevant proteins, such as fibrinogen[38], using SAMs with many different functionalizations using an all-atom representation. In each study, they report both agreement and conflict between simulation and experiment. Jiang [39] also showed conflict with experimental results in energies of adsorption and monolayer structure. Kubiak *et al.* [40] atomistically simulated Egg-white Lysozyme in three different systems, and found that lysozyme has a preferred orientation for absorption to surfaces. However, they also reported that the time scale in this simulation is still too short for the protein to unfold on the surface. Due to computational limitations, all-atom models are restricted to probing global orientation or local structural changes, which fails to meet our need for protein properties in large time and space scales.

To calculate protein stability, the protein must fold and unfold many times during the simulation to get histigrams of energy and density of states. To achieve the necessary sampling time scale, researchers have removed degrees of freedom by using coarse-grain approaches, fixing configurations in space, and removing solvent molecules.[41, 42, 43, 44, 45, 46, 47] For example, Sun *et al.* [41] employed an implicit solvation in simulations to decide the orientation of proteins when absorbed to surfaces. Zhou *et al.* developed a united-residue model to study the adsorption and orientation of two antibodies on surfaces with Monte Carlo simulations. Carlsson *et al.* [45] reported a study about lysozyme adsorption to charged surfaces by Monte Carlo simulation. The lysozyme in his simulation was modeled as a large hard sphere and each of 32 charged amino acids is represented by a small charged site on the surface of the sphere.

The thermodynamic perspective explaining the influence of the surface on the stability of proteins, theorized by Dill *et al.* [42, 48], predicts that proteins are always stabilized when tethered to short-ranged, repulsive surfaces. The reason is summarized in Figure 1.3. As depicted, the number of unfolded conformations available to tethered peptides is less than in the bulk, because configurations are confined by the surface. This decreases the entropy of unfolded protein on surfaces which destabilizes the unfolded state, favoring the folding process. Assuming that the enthalpy of folding is approximately the same on and off the surface, a decrease in the entropic cost of folding decreases the Gibbs energy of folding for the tethered protein relative to the bulk protein resulting in stabilization of the proteins. In

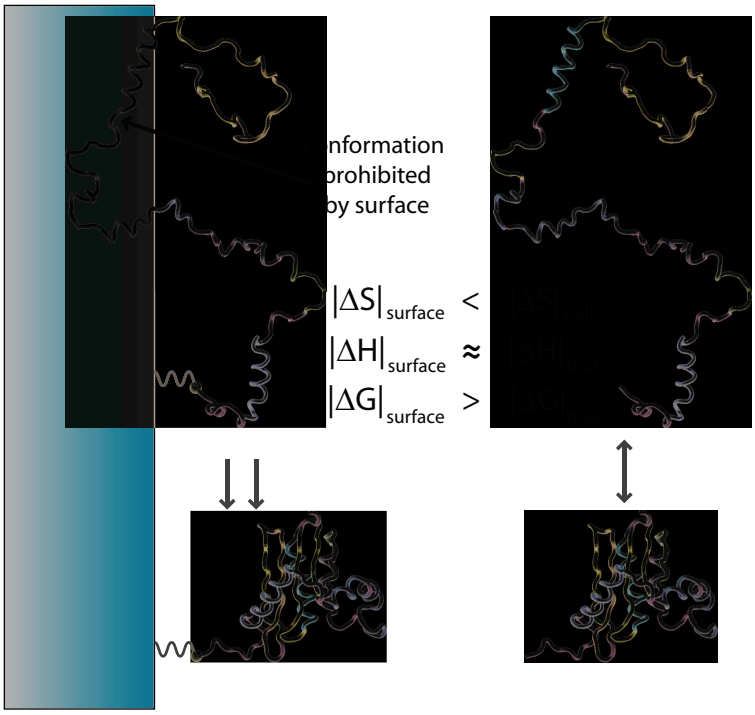


Figure 1.3: Dill's Theory behind the stabilizing influence of the surface on tethered proteins

short, the theory shows that the entropic cost of folding is greater in the bulk case than the on the surface because unfolded bulk peptides have more entropy to lose than the surface proteins. A decrease in the entropic cost results in a more negative value of ΔG_f.

Although Dill's theory is one of the most widely cited theories, some simulations provide different results. In one example, work by Friedel *et al.*[49, 50] showed simulations of a four-strand, beta-barrel protein both in the bulk and on surfaces with different tethering sites in outer loop regions. The results show that the protein could be stabilized or destabilized on the surface depending on the tethering site. Results also showed that if the tethering is done to a site on the interior of the molecule, the protein is always destabilized. One more recent work by Zhuang *et al.*[51] also shows similar variation of protein stability when tethering the src-SH3 protein on surfaces with different sites. Similar results were seen in prior work by Knotts *et al.*[52] done on the all-alpha, three-helix-bundle protein from *Staphylococcus aureus*. In this study both the mechanical and thermal stabilities of the peptide were reduced when the protein was tethered to the surface.

Recent work by Knotts *et al.*[53] showed a study of four proteins: protein A, 434 repressor, SH3, and Protein G, which have different secondary or tertiary structure motifs. All four proteins were simulated both in the bulk and on an inert surfaces with N- and C- termini, and the thermal stability of the proteins was probed using configurational-temperature-density-of-states simulations. The work showed that proteins could be stabilized or destabilized on surfaces. Also, it was shown that only all-alpha proteins displayed the surface-induced destabilization, while the proteins with beta-content displayed only stabilization. This is consistant with results from Zhuang *et al.*[51]. Another important result from the work by Knotts *et al.*[53] is that the stability cannot be correlated to secondary structures, because protein A and 434 repressor, which are both all-α peptides, displayed different behaviors. Furthermore, as the results for protein A suggested, the same protein can behave differently on surfaces depending on different tethering sites.

Previous studies show that there are successful methods to study protein stability in inhomogeneous systems, and additional work on more proteins is required for understanding and predicting protein stabilities on surfaces.

1.3 Summary

Although researchers realize the importance of understanding protein stability on surfaces, there is still a lack of convincing theory for predicting protein behavior on surfaces. Even though methods for binding proteins to surfaces in predetermined manners have been developed, and there are some successful instances of binding proteins with conserved stabilities, experimental understanding is limited by the difficulty in producing protein structures on surfaces. Again, there are no theories guiding how to tether a protein to a surface and keep its function.

Simulation methods provide a better means of exploring this topic. Atomistic simulations have shown to be ineffective in rendering protein folding information due to their large computational requirement. Some coarse-grain models have been proven to provide results that are consistant with experiments. With these methods, a lot of research work has been done. Results show that proteins could be stabilized or destabilized on surfaces,

which conflicts with Dill's theory, and that the folding mechanism could be changed if the tethering site is in an structured part of the transition area[51]. Moreover, protein stabilities can not be correlated to their secondary structure motifs. Since protein structure leads to protein function, a deeper study and analysis of protein stabilities on surfaces with respect to their tertiary structure is needed.

2 EFFICIENTLY SIMULATING PROTEIN FOLDING: REPLICA EXCHANGE

In this chapter, general simulation tools and thermodynamic calculation methods are introduced. After that, the replica exchange simulation method is also discussed in the context of rugged energy landscapes. To meet the simulation speed requirement, message passing interface (MPI) is used and results are validated.

2.1 General Approach and Thermodynamic Quantities Calculation

2.1.1 General Approach

In this work, Brooks' Go-like model[11, 12, 13], which is a suitable method for predicting protein folding properties, is used. This model extends earlier ones by introducing different energy scales to describe non-bonded interactions between side chains, hydrogen bonding in regular secondary structure, and sequence-dependent virtual dihedral potentials to keep proteins in appropriate conformations. The resulting energy surface can mimic that of the real protein more closely than earlier models, which employed fewer energy scales or targeted specific encoding of the backbone structure with virtual dihedral potentials. [11, 12, 13, 14, 15] This model has been shown to give good agreement between simulation and experimental folding studies.[12]

To compare the effect of the surface on protein stability, certain proteins discussed below were simulated in the bulk and tethered to the surface in several locations: at both the N- and C- termini and in each of the loop/turn regions connecting secondary structural elements. In each case, the melting temperature (T_m), Gibbs energy of folding (ΔG_f), enthalpy of folding (ΔH_f), and entropy of folding ($T\Delta S_f$) were calculated. In addition, some order parameters were calculated for analysis of the correlation of protein structures

and stabilities on surfaces. Also, the fraction of each secondary structure that was folded was measured for analysis of the folding mechanism.

To make the comparison easier, T_m^* was defined, which is the melting point of the protein in bulk. By definition, $T_m/T_m^* = 1$ for the protein in the bulk. If $T_m/T_m^* < 1$, the protein is destabilized by the surface at the tether location indicated. If $T_m/T_m^* > 1$, the protein is stabilized. Also, $\Delta\Delta G = \Delta G - \Delta G^*$ was defined, in which, ΔG^* is the Gibbs energy of the protein at T_m^* in bulk. Then, if $\Delta\Delta G$ is negative, the protein is stabilized by tethering. If the value of $\Delta\Delta G$ is positive, the surface has a destabilizing influence. Since all ΔG^* are 0 by definition, $\Delta\Delta G$ as talked above, should be equal to ΔG.

2.1.2 Thermodynamic Quantities Calculation

The metrics used to quantify stability were calculated from simulation using standard methods from statistical mechanics. The melting point is the temperature of the peak in the heat capacity curve. The heat capacity, C, is related to the fluctuations of the potential energy of the system according to

$$C(T) = \frac{\langle U^2 \rangle_T - \langle U \rangle_T^2}{RT^2} \tag{2.1}$$

where R is the gas constant, T is the temperature, and the <>'s denote the average of the corresponding quantities. The average of any arbitrary quantity, X, can be found from

$$X(T) = \langle X \rangle_T = \frac{\sum_U X(U)\Omega(U)e^{-\beta U}}{\sum_U \Omega(U)e^{-\beta U}}. \tag{2.2}$$

The key quantity needed to evaluate Equation 2.2 is the density of states, $\Omega(U)$, which is calculated using the Weighted Histogram Analysis Method (WHAM) [54] on the data obtained from replica exchange simulations.

Each of the proteins investigated fold through a two-state mechanism. For two-state folders, the Gibbs energy of folding is calculated from

$$\Delta G_f = G_{\text{folded}} - G_{\text{unfolded}} = -k_B T \ln\left(\frac{P_f}{1 - P_f}\right), \tag{2.3}$$

where P_f is the probability of the folded state at temperature T. The values of P_f are determined by classifying the configurations sampled throughout the simulation into "folded" and "unfolded" ensembles based upon the instantaneous fractional nativeness, q. The fractional nativeness is the ratio of the number of native contacts formed at a particular instance to the total number of native contacts possible. A protein is considered folded if $q > q(T_m)$ where T_m is the melting temperature of the protein. This treatment yields $\Delta G_f = 0$ for the protein at its melting temperature–a relationship which must be true by definition as mentioned above.

The enthalpy change ΔH_f associated with the folding is calculated as the difference of the potential energy between the folded and unfolded states. (Strictly, $H = U + PV$, but the changes in the PV term are assumed to be negligible as has been done previously[55].) The change in entropy is then obtained from $T\Delta S = \Delta H - \Delta G$.

2.2 Replica Exchange

One of the key challenges in the computer simulation of proteins at the atomic level is the sampling of conformational space. Many commonly used sampling protocols, such as Monte Carlo (MC) and molecular dynamics (MD), usually get trapped in local minima in rugged energy landscapes, because they cannot cross high free-energy barriers between conformational states. Replica exchange (RE), also known as parallel tempering, provides an efficient sampling method to solve that problem using a series of replicas of a system of interest. Each replica is typically in the canonical ensemble, and usually each replica is at a different temperature. [56, 57] To accomplish barrier crossings, replicas at different temperatures exchange complete configurations, which is also called "swaps". Periodically, coordinates are exchanged by using a Metropolis criterion that ensures that at any given temperature a canonical distribution is realized. Swaps were attempted every 2,000 steps and accepted with probability: Figure 2.1 shows how replica exchange works on a single processor.

To get the thermodynamic information, a lot of samples are needed. Even with WHAM, it is still needed to sample about 24 replicas in different temperatures to cover the

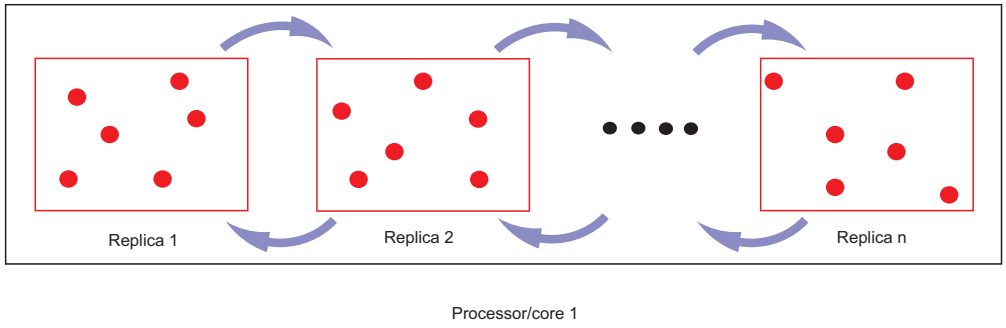


Figure 2.1: Replica exchange simulation

temperature range of interest. Since it is needed to sample both folded and unfolded states many times, a large time scale is also needed. Furthermore, for statistical consideration, many simulations need to be run to get an accurate expectation value. To fulfill those requirements, a large amount of computational time is required. 6 peptides are in use, and 8 simualtions of different sites are needed for each peptides on average. Also, it is expected to get 6 copies for each simulation to get an average value. It takes 11 days on average to run one single simulation. Even if all simulations ran without any error, 8.7 years are needed.

2.3 MPI

Fortunately, besides more sampling efficiency brought by the method itself, the RE simulation is also easy to employ in highly efficiency parallel computation with large clusters of CPU. That means, MD simulations run for each replica on an individual processor in parallel. That means, the need of computational time can be tranfered into the need of computational resources. In this case, it is not required to force one computer to run all 24 replicas for a long time, because it is easy to assign each replica to different processors, as shown in Figure 2.2. The only problem is how to implement replica swapping as described in RE simulation. In certain steps, swaps are proposed between replicas, which are on different processors in this case. To realize that, the MPI is refered, which is introduced as follows.

The Message Passing Interface (MPI) [58, 59], is a specification for an application programming interface (API) that allows many computers to communicate with one another.

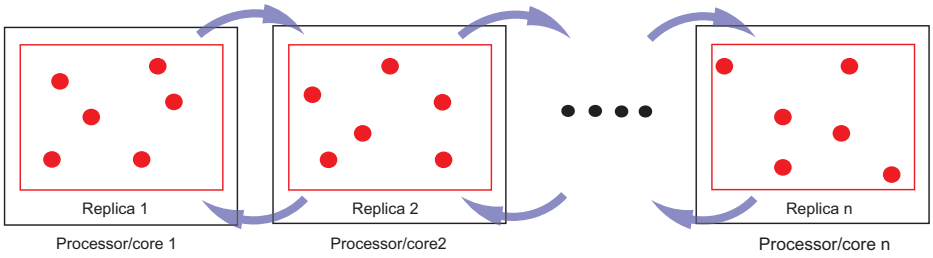


Figure 2.2: Replica exchange with MPI

This feature is just what needed in parallel simulation of RE. With proper design in MPI, replicas can be simulated on each processor with MD, and exchange their coordinates in certain iterations as expected. How does that help this research? If calculated in ideal condition, only 1/24 of the 8.7 years, which is 132 days, is required. That is more than acceptable if the script can be writen out in a couple of months, and leaving enough time to analyze data.

It took one month to write the script for replica exchange. After that, it took about half a month to make it work with the group scripts and get it to run on the supercomputer. Simulations of RE on a single processor were already available in our group scripts, including MD simulations of many replicas, and the swap algorithm. Therefore, what is needed to do is to assign each replica into an individual computer, and write the swap algorithm using MPI to combine processors. First, replicas are assigned to different processors by using submission files in multi-processor format. After that, a function of swapping is developed in MPI, and the script is written in the C++ programming language. In each 2,000 iterations, every processor calls the function of swapping, to send or receive a swapping request. If the request is accepted by the metropolis criterion, two processors exchange their coordinates. If not, they just keep their own replicas and continue to run MD simulation until the next swapping call.

The detail of MPI programming for replica exchange is shown as Appendix A.

2.3.1 Performance of the MPI Scripts

To validate the MPI scripts for the replica exchange, 3WRP was taken as an example by comparing the predicted melting temperatures with and without MPI. The melting temperatures of protein 3WRP are measured as $303.57 \pm 0.78K$ with MPI and $304.62 \pm 1.01K$ without MPI. The data are very close to each other, with only 0.34% difference. Furthermore, the data are consistant with very little flucturation. For all other proteins, we get the similar accuracy and consistancy with MPI.

Since results are very accurate, the next thing that is needed is to know how much time could be saved with MPI in simualtions. Originally, running the RE simulation on a single processor takes 10.98 days. With the assistance of MPI and faster new processors in the BYU Supercomputer Lab, it just took 7.55 hours for the same job. Now to calculate the total time needed, using 8 hours as the average for each simulation. It reduced to 96 days with MPI from 8.7 years without it.

To summerize, using MPI speeds up simulations very much, and gives consistantly accurate results. Furthermore, it saved tons of time for debugging simulations. Also, much more data than needed could be obtained just to make results more convincing.

3 THE STRUCTURE AND STABILITY OF ALPHA-HELICAL, ORTHOGONAL-BUNDLE PROTEINS ON SURFACES

In this chapter, the details of the hypothesis in this research is stated. To test this hypothesis, five chosen Alpha-helical, Orthogonal-bundle proteins are simulated. Also, thermodynamic quantities are calculated and further discussion are shown based on the melting temperature results. After that, an analysis of thermodynamic quantities in more detail is implemented.

3.1 Hypothesis

Previous studies[49, 50, 52] demonstrated the fact that tethering must be done in the outer loop regions of proteins, and the secondary structure is not a good predictor of proteins stability on surfaces, it is predicted that if proteins are tethered in outer loop regions only, and if all proteins are chosen from the same tertiary structure motif, they may display similar behaviors on the surface. Specifically, efforts are concentrated on alpha-helical, orthogonal-bundle proteins.

Alpha-helical, orthogonal-bundle are terms from CATH classification method, which groups proteins based on their structure motifs. Protein research progress instigates categorization of structurally related proteins. As a result, structure-based classifications, such CATH, can be effective at identifying unanticipated relationships in known structures and in optimal cases function can also be assigned. Analysis of the structural families generated by CATH reveals the prominent features of protein structure space.[60, 61]

CATH is a hierarchical classification of protein domain structures, which clusters proteins at four major levels: Class (C), Architecture (A), Topology (T) and Homologous superfamily (H)[60, 61]. The Class (C-level) is determined according to the secondary structure (alpha helix or beta sheet) composition and packing within the structure. Typically,

there are three majors in the Class level: Mainly-Alpha, Mainly-Beta, and Alpha/Beta. Under the Class level (C-level), is the Architecture (A-level) level, which describes the overall shape of the domain structure as determined by the orientations of the secondary structures. Under the Mainly-Alpha class, there are different architectures like the up-down bundle, orthogonal bundle, alpha/alpha barrel, and horseshoe. They represent different A-levels just because of their various orientational organization (or shape) of the alpha helices. We acknowledge that this type of protein comprises only a small fraction of all available proteins. This type is chosen as a starting point because of their simplicity and size.

In this research, folding properties on surfaces of a group of proteins within the same tertiary family are going to be summerized, furthermore, *a priori* how proteins behave on a surface based on their structure classification or structure motifs are tried to be predicted. The work proposed here is based upon these findings. Formally, the hypothesis of the research work is that all peptides classified as orthogonal bundle, alpha-helical motifs, will behave similarily on a surface. All-alpha proteins are chosen because this particular structural motif is present in large proportions in antibody-binding proteinsan important class of proteins involved in protein/surface interactions. We also have extensive experience in simulation of all-alpha proteins. The orthogonal-bundle architecture was chosen because it is believed such proteins will behave more consistently across this architecture than other options. This is because orthogonal-bundle proteins are more sphere-like than proteins in other architectures. In summary, all-alpha, orthogonal-bundle proteins are important from a biotechnology standpoint and are expected to give consistent results, providing a useful starting point for discovery.

3.2 Method

3.2.1 Proteins

Five different proteins were used to test the hypothesis. These were identified with the CATH classification method[60, 61]. Each of the five proteins have the same class, mainly alpha, and the same architecture, orthogonal bundle. The five proteins, shown in Figure 3.1,

are the N-terminal domain of phage 434 repressor (PDB ID: 1R69), cytochrome C-552 from *Nitrosomonas europaea* (1A56), retinoblastoma tumor suppressor (1AD6), cytochrome C6 (1A2S), and myoglobin (5MBN). The size of each protein ranges from 64 to 163 residues.

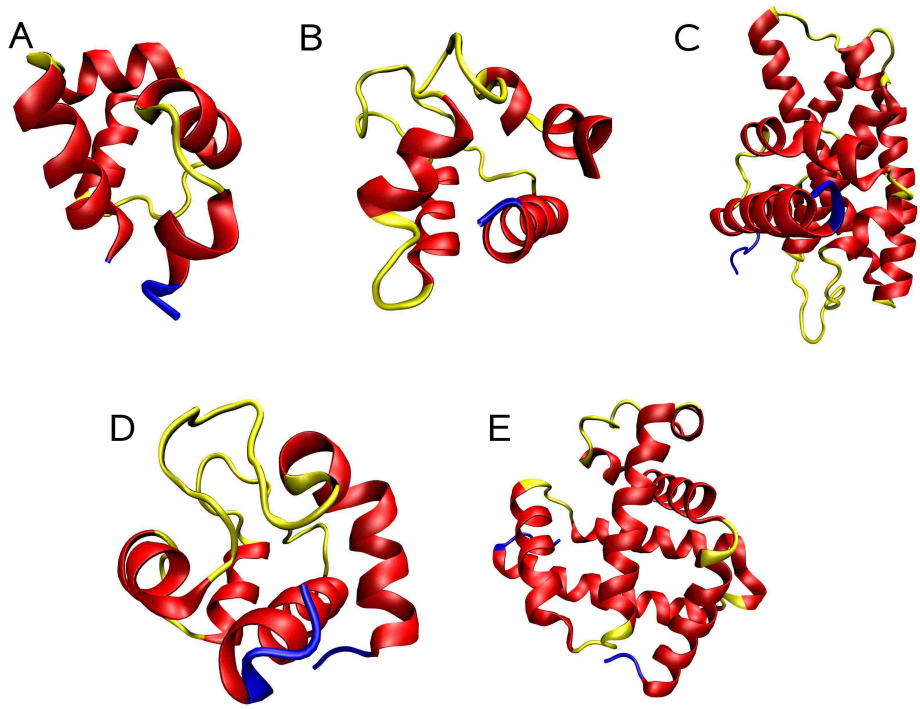


Figure 3.1: Schematic representation of the five alpha-helical, orthogonal-bundle proteins: A. 1R69, B. 1A56, C. 1AD6, D. 1A2S, E. 5MBN

Orthogonal-bundle proteins were chosen as they provide a convenient starting place to investigate the behavior of families of proteins on the surface. This family of proteins are composed of α-helices connected by loop regions. The helices lie at approximately 90° with respect to each other. By comparison, *all-alpha, up-down bundles* (a family of proteins with the same CATH class but different architecture) are composed of α-helices which lie in a roughly parallel orientation resulting in an elongated structure rather than a globular structure. The globular nature of orthogonal-bundle proteins is such that the loop regions are found on the exterior of the molecule, a condition that has been shown to be necessary to maintain the native structure of the protein when tethered to the surface [50, 51].

For computational efficiency, a coarse-grain model is used to represent the proteins. The specific implementation is the Gō-like model of Karanicolas and Brooks [11, 12, 13, 14,

15]. In this formalism, each residue is represented by one site placed at the C_α position. The model extends earlier Gō-like models by introducing different energy scales to describe hydrogen bonding between side chains, and sequence-dependent dihedral potentials. Previous models employed fewer energy scales and set dihedral parameters based upon the PDB structure and not the sequence. As such, the resulting energy surface mimics that of real proteins more closely than earlier models. Moreover, the model has been shown to give good agreement with experimental folding studies [12, 13, 14, 15, 62]. Input files were generated using the MMTSB website `http://www.mmtsb.org`.

Table 3.1 contains a residue-level, structural analysis of the five proteins used in the study. The residues comprising each helix and loop are listed. The categorization of each residue as either "loop" or "helix" was performed with VMD [63] which uses the STRIDE algorithm [64]. The number of helices among the proteins ranges from 4 to 9. The lengths listed are the distance between the first and last residue for each structural element. For example, the length of Helix 1 of 1R69 is 16.3005 Å which is the distance between residues 2-13.

3.2.2 Experimental Design

To compare the effect of the surface on protein stability, the five proteins mentioned above were simulated in the bulk (no surface) and tethered to the surface at several locations in each of the loop regions identified in Table 3.1. In each case, the melting temperature (T_m), Gibbs energy of folding (ΔG_f), enthalpy of folding (ΔH_f), and entropy of folding ($T\Delta S_f$) were determined. In addition, order parameters such as the fraction of native contacts formed and the radius of gyration were calculated to analyze the correlation between structures and stabilities.

Comparing the stability of tethered proteins to bulk proteins is done using T_m and ΔG_f for each case. For the melting temperatures, results are presented with the temperatures scaled by the melting temperature of the protein in bulk (T_m/T_m^*). If this scaled temperature is less than 1, the protein is destabilized by the surface. If the scaled temperature is greater than 1, the protein is stabilized by the surface. Comparing Gibbs energies of folding in

Table 3.1: Residue-level secondary structure analysis of five α-helical, orthogonal-bundle proteins

1R69											
	Coil 1	Helix 1	Loop 1	Helix 2	Loop 2	Helix 3	Loop 3	Helix 4	Loop 4	Helix 5	Coil 2
Sites	1	2-13	14-16	17-24	25-27	28-36	37-43	44-52	53-55	56-61	62-63
Length	0	16.3005	6.8399	10.4675	7.1107	12.5014	15.0222	12.3014	7.148	8.5078	3.8104
1A56											
	Helix 1	Loop 1	Helix 2	Loop 2	Helix 3	Loop 3	Helix 4	Loop 4	Helix 5	Coil 1	
Sites	1-8	9	10-13	14-24	25-32	33-37	38-48	49-66	67-80	81	
Length	10.5426	0	5.9274	4.6505	10.8597	7.0102	14.0612	16.0364	19.5559	0	
1AD6											
	Coil 1	Helix 1	Loop 1	Helix 2	Loop 2	Helix 3	Loop 3	Helix 4	Loop 4	Helix 5	Loop 5
Sites	1-6	7-14	15-20	21-29	30-33	34-57	58-60	61-91	92-96	97-101	102
Length	14.6039	10.9725	9.8860	12.2118	5.2275	33.3364	6.7516	44.6546	10.948	5.8727	0
	Helix 6	Loop 6	Helix 7	Loop 7	Helix 8	Loop 8	Helix 9	Coil 2			
Sites	103-120	121-137	138-144	145-147	148-161	162-166	167-183	184-185			
Length	25.5257	12.3804	10.3599	6.7828	20.2056	9.4146	24.1438	3.7462			
1A2S											
	Coil 1	Helix 1	Loop 1	Helix 2	Loop 2	Helix 3	Loop 3	Helix 4	Coil 2		
Sites	1-3	4-19	20-33	34-40	41-46	47-55	56-69	70-86	87-89		
Length	7.2506	16.4795	12.0588	9.5048	9.0018	12.4227	12.7552	24.2138	5.4674		
5MBN											
	Coil 1	Helix 1	Loop 1	Helix 2	Loop 2	3-10-Helix 1	Loop 3	Helix 3	Loop 4	Helix 4	Loop 5
Sites	1-3	4-18	19-20	21-35	36	37-42	43-51	52-57	58	59-77	78-81
Length	6.3033	20.6028	3.7954	21.0518	0	9.6629	15.3579	8.5878	0	27.1246	5.1646
	Helix 5	Loop 6	Helix 6	3-10-Helix 2	Loop 7	Helix 7	Coil 2				
Sites	82-96	97-100	101-119	120-122	123-124	125-149	150-153				
Length	20.9421	9.3471	27.1248	5.074	3.8155	36.1229	9.1842				

different environments is commonly done by defining the quantity $\Delta\Delta G$. For the present purposes, $\Delta\Delta G = \Delta G_f^{surface} - \Delta G_f^{bulk}$ which is the difference between the Gibbs energy of folding on the surface and in the bulk. As Gibbs energy of folding is a temperature-dependent property, the data presented later are tabulated at the melting temperature of the protein in the bulk. At this temperature, $\Delta G_f^{bulk} = 0$ by definition and $\Delta\Delta G = \Delta G_f^{surface}$. The double-$\Delta$ notation is therefore dropped and Gibbs energies are reported as simply ΔG_f. For tethered proteins, if $\Delta G_f < 0$, the protein is stabilized, and if $\Delta G_f > 0$, the protein is destabilized. Also, the lower (more negative) the value of ΔG is, the more stable the protein.

3.2.3 Simulation Protocols

To prevent the simulation from becoming trapped in local energy minima, simulations were performed using the replica exchange (RE) algorithm [57, 65]. Twenty-four replicas were used for each protein, and the canonical ensemble was generated using the Nosé-Hoover-Chain [66, 67, 68] integration method with 3 thermostats of mass 10^{-26} kg Å^2. The time step was 1 fs, and each simulation contained 10 million steps of equilibrium followed by 30 million steps of production. Swaps were attempted every 2000 steps and accepted with probability:

$$P_{acc}(swap) = \min\{1, exp(-\Delta\beta\Delta U)\} \tag{3.1}$$

where $\beta = \frac{1}{k_B T}$, k_B is Boltzmann's constant, U is the potential energy of the system. Temperature increment between adjacent boxes ranged from 2.5 to 10 degrees. The smaller increments were used close to the melting temperature and the larger increment farther away.

3.2.4 Order Parameters

In order to correlate the stability of the protein to different patterns in the structure of the molecule, several order parameters were defined. Order parameter selection is a trial-and-error process, and several parameters were calculated to describe protein stability as a function of measurable variables. The lengths reported in Table 3.1 were one type of parameter tried. Others included the number of residues in the loop segment, the lengths

of the helices adjacent to the tether point, the angle formed by adjacent helices, the free rotation volume and the free rotation angle.

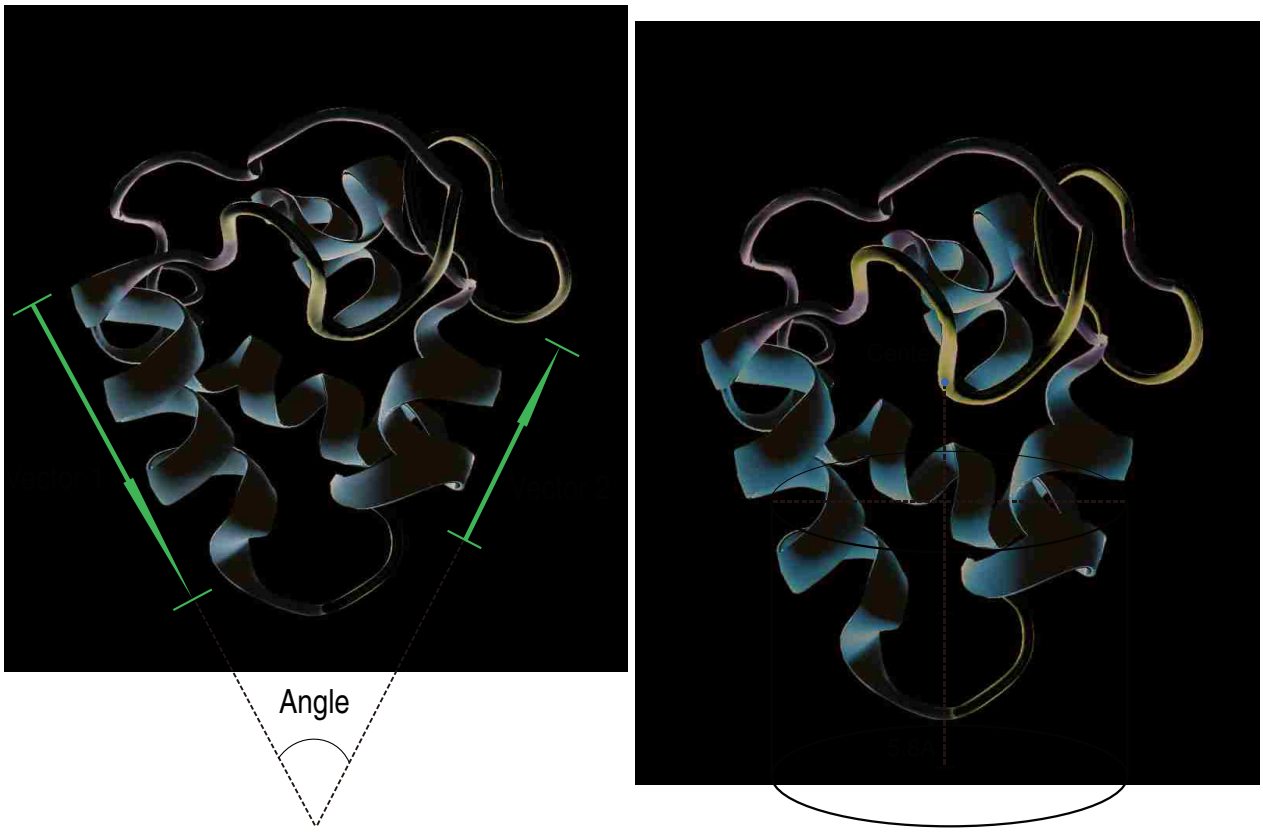


Figure 3.2: Order parameters defined as (a) angle and length (b) free rotation volume

The last four parameters are described in Figure 3.2 and Figure 3.3.

The length of a helix is the distance between the first and last sites comprising the helix as found in Table 3.1. To define the angle made by adjacent helices, a vector is defined for each helix. Each vector extends away from the tether point and is formed between the two points within the helix that are farthest away from each other but lie on the same side of the structure. Choosing sites on the same side of the helix creates a vector that is parallel to the vector running directly through the middle of the helix. With the two vectors defined, the angle between the helices is found from definition of the dot product.

For the first three investigated order parameters: 1) the angle formed by consecutive helices, 2) the distance between the consecutive helices, and 3) the presence/absence of β-turns in the loop regions, no correlation was found between stabilization/destabilization

and these parameters. For example, it was hypothesized that if the protein was tethered in a loop region where the adjoining helices made an acute angle that the protein would be stabilized. However, this does not explain results where two sites in the same loop show different behavior such as site 41 of 1A2S, which was destabilized, and site 46, which was stabilized (See Figure 3.5). It was also thought that stability is related to the number of residues in the loop region as a longer loop was expected to allow the protein more flexibility to accommodate surface interactions without disrupting the positions of the helices forming the bundle. For example, all the tethering sites in Loop 3 (18 residues in length) of 1A56 showed stability. In fact, all the sites in loop regions with more than 10 residues resulted in stability that was equal to or greater than that found in the bulk. However, for loops less than 10 residues in length, varied behavior was seen. For example, site 15 (destabilized) and site 25 (stabilized) of 1R69, are found in separate loops of 3 residues in length but have different stability.

Actually, these order parameters were tried when fewer tether sites were tested. After failures of distinguishing loop regions by using these order parameters, simulations of more tether sites were implemented to clarify the trend. Finally, two kinds of differences between tether sites were realized: the difference between tether sites in the same loop region and the difference between loop regions. As shown below, tether site positions in three kind of loop region shapes were used to distinguish tether sites in the same loop, and the rotational volume or angle were tried to tell the differences between loop regions.

As shown below, the ability of the protein to vibrate and rotate on the surface is important in stability. The volume fraction available for rotation (VFAFR), the metric used to follow this phenomenon, is seen in Panel b of Figure 3.2. The VFAFR is protein and tether site specific and is calculated by first defining a cylinder which contains the portion of the protein which can interact with the surface. The axis of the cylinder line connects the tethering point and the mass center of the protein. The length of the cylinder, l, is 70% of the length of the center line (about one third of the diameter of the protein) plus 5.8 Å (the length of the tethering bond). The radius of the cylinder if found by first identifying all the atoms that lie between two planes placed perpendicular to the cylinder axis at the ends of

the cylinder. The distance between each of these atoms and the cylinder axis is calculated and the radius (r) of the cylinder is taken to be the largest of these values.

To calculate the VFAFR, the volume of the residues found within the cylinder must be subtracted from the volume of the cylinder. The volume of the cylinder is $V_c = \pi r^2 l$. The volume of each of the residues found within the cylinder is calculated using Voroni tessellations on the full atomic coordinates using the software PROVAT [69]. The protein volume, V_p, is the sum of the residue volumes. Then the VFAFR per atom is then given by

$$VFAFR = \frac{V_c - V_p}{V_c}. \tag{3.2}$$

As observed, small rotation angles are always formed by those atoms far from the center line which cross the mass center of the protein and the tethering site, and they are the limitation for the rotation of proteins. Therefore, the average free rotation angles formed by those atoms that are far from the center line could be an acceptable metric that distinguish site 145 from others. The free rotation angle is defined (in orange in Figure 3.3) as the compliment angle of the one that is formed by a ray through one site on the protein and a ray through the mass center, both of which are across at the tethering site. More serious consideration of the calculation of free rotation angles were discussed in the following result section.

3.3 Results and Discussion

3.3.1 Melting Temperatures

As mentioned in the previous section, stabilities of tethered and bulk proteins are compared to prove the hypothesis. The stabilities are related to the heat capacity and native contacts as previously described. Figure 3.4 shows C_v, Q, and R_g as a function of temperature for 1R69. (The other four proteins studied show similar behavior in C_v, Q, and R_g, but the results are not shown for conciseness.) A single, sharp peak is present in the heat capacity (panel (a)). The location of this peak is the melting temperature of the protein. The fractional nativeness and radius of gyration, panels (b) and (c), display a sigmoidal shape indicating an

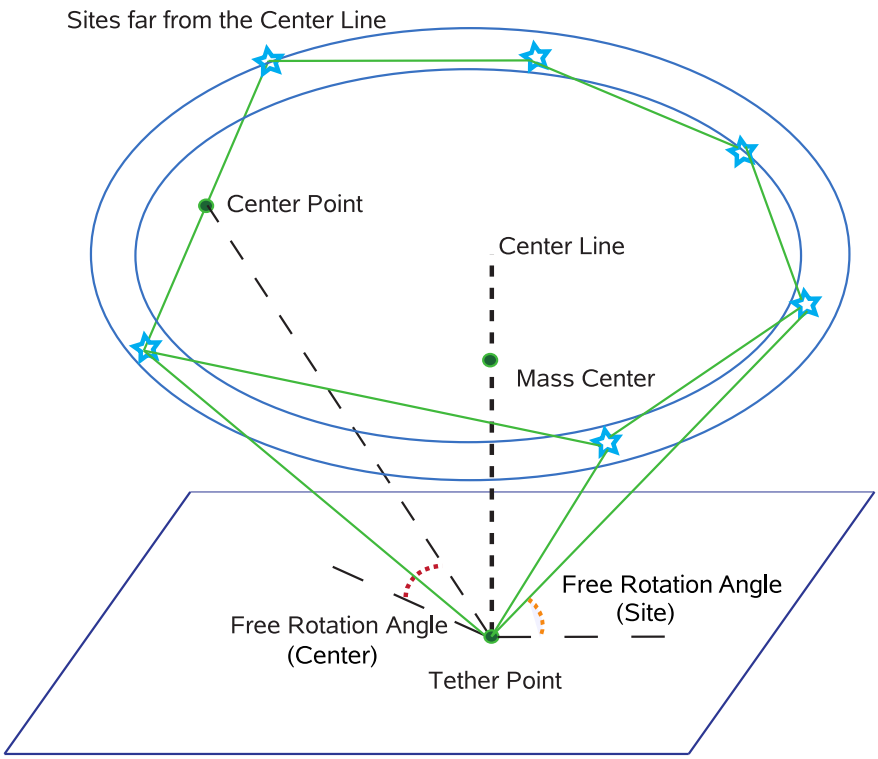


Figure 3.3: Defination of the free rotation angle for proteins

abrupt transition from the folded to the unfolded state. The inflection point in each of these curves occurs at the melting temperature obtained from the heat capacity curve in panel (a). The fact that the heat capacity curve displays only one peak, and the melting temperature identified by C_v coincides with the transition temperature of the order parameters Q and R_g, indicates 1R69 follows a two-state folding model. As such, our assumption of two state folding to calculate ΔG_f is reasonable.

The calculated melting temperature for 1R69 is $49.13°C$. Ku *et al.*[70] estimated the melting temperature for this protein to be between $30°C$ and $90°C$ using different experimental techniques. Therefore, the melting temperature calculated for 1R69 with this model is reasonable.

Figure 3.5 shows a summary of the melting temperatures of all the proteins in the bulk and tethered to the surface at multiple locations. The value reported on the ordinate is T_m/T_m^* where T_m is the melting temperature on the surface when tethered at the site indicated on the abscissa and T_m^* is the melting temperature of the protein in bulk. The proteins were tethered to the surface in the each of the loop regions joining adjacent helical

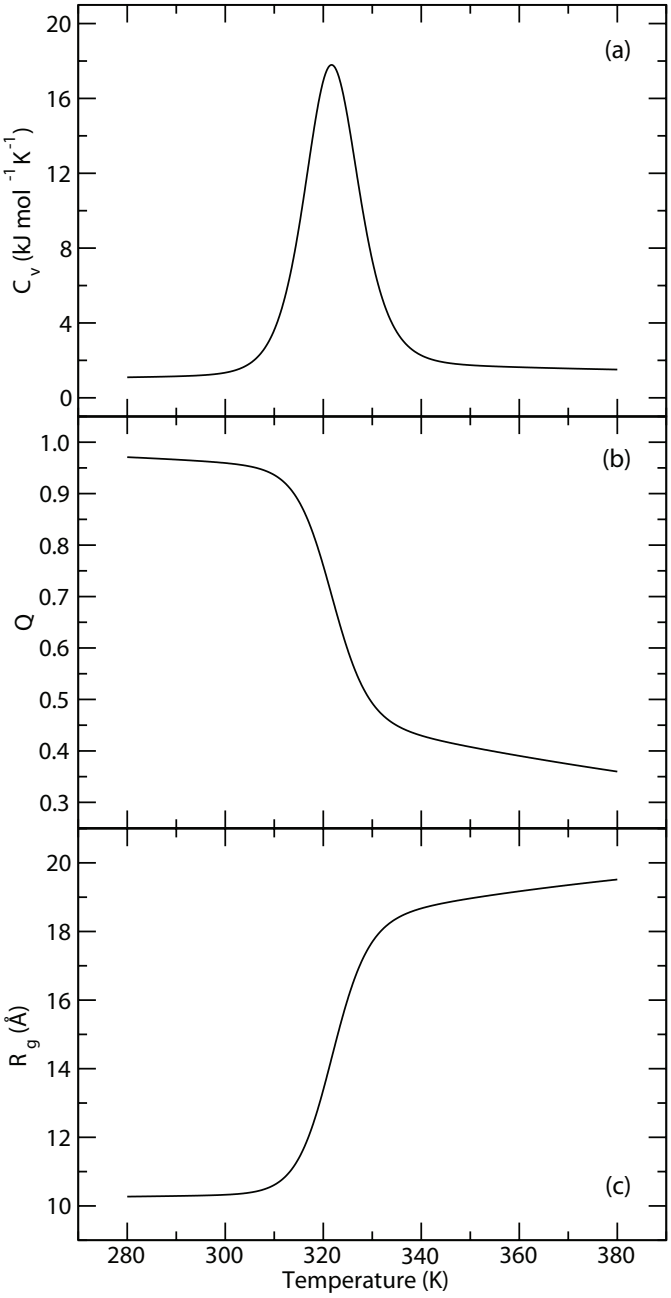


Figure 3.4: Heat capacity (panel a), fractional nativeness (panel b), and radius of gyration (panel c) as a function of temperature for 1R69

segments. The results are grouped by alternating colors. Adjacent bars of the same color indicate that each of the listed tether sites are found in the same loop region. For example, tether sites 27, 30, and 32 are found in Loop 1, sites 41 and 46 in Loop 2, and sites 64, 66, and 68 in Loop 3 of 1A2S. The bulk value is designated by the letter "B".

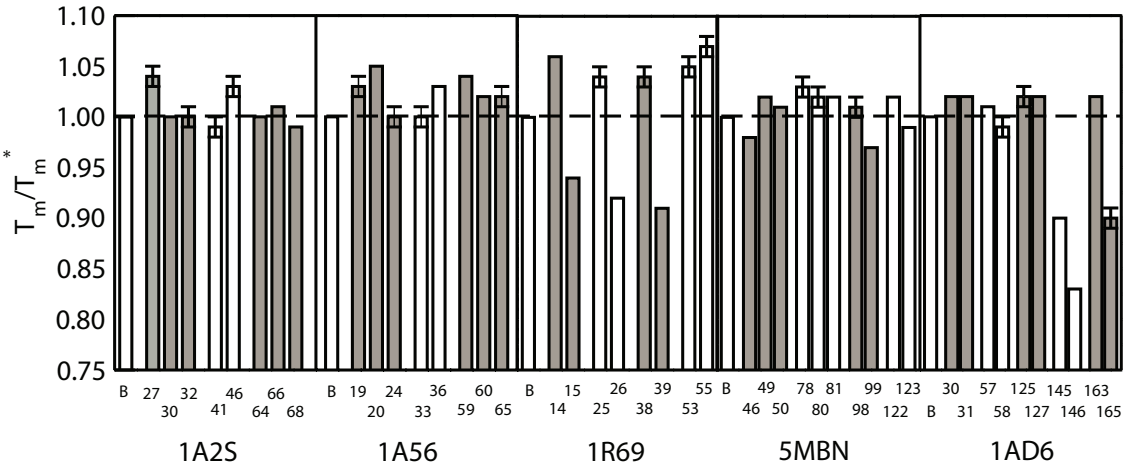


Figure 3.5: Scaled melting temperatures of 1R69, 1A56, 1AD6, 1A2S, and 5MBN in the bulk and tethered to the surface in various locations

The melting temperatures show a large amount of variability according to which site is tethered. Tethering the protein resulted in stabilities which were approximately equal to or greater than the bulk for 31 of the 42 (74%) cases. Included in these 31 are situations, such as site 24 of 1A56, where the error bars are such that the scaled melting temperature cannot be shown to be statistically different than the value of 1.

3.3.2 Analysis of Hypothesis

The hypothesis of this work was that all-alpha, orthogonal bundle proteins, when tethered to the surface only in the loop regions adjoining adjacent helices, will be stabilized compared to the bulk value. At first glace, the data in Figure 3.5 indicate that the hypothesis is incorrect. However, a careful examination of the data reveal an interesting pattern. In 18 of the 19 loop regions investigated, tethering sites can be found which stabilize the protein on the surface. The only exception is the loop formed by sites 145 to 147 of 1AD6. In this region, no site could be found which stabilized the protein. In general, however, it appears that loop-region sites can be found that result in stabilization of tethered all-alpha, orthogonal bundle proteins. The next section examines in more detail why certain tethering sites result in stabilization while others do not.

3.3.3 Categorization of Tethering Sites and Rotational Order Parameters

Previous theoretical work has shown that stabilization of proteins on surfaces is related to how the tethering site affects both the entropy and enthalpy of the protein [52, 49, 50, 53] as well the degree to which the tethered site disrupts the transition state along the folding pathway [51]. The difficulty with applying this knowledge in a predictive manner is that either the folding pathway must be known or experiments or simulations have to be performed to ascertain the mechanism. It would be ideal if design heuristics could be developed which when applied to the crystal structure of the protein of interest would result in a list of tethering sites that would maintain the stability of the protein on the surface. In this section, several geometric order parameters, and their ability to predict the stability of orthogonal-bundle proteins on surfaces, are described.

The first order parameters investigated were: 1) the angle formed by consecutive helices, 2) the distance between the consecutive helices, 3) the number of residues comprising the loop region, and 4) the presence/absence of β-turns in the loop regions. However, no correlation was found between stabilization/destabilization and these parameters for all instances. For example, it was hypothesized that if the protein was tethered in a loop region where the adjoining helices made an acute angle that the protein would be stabilized. However, this does not explain results where two sites in the same loop show different behavior such as site 41 of 1A2S, which was destabilized, and site 46, which was stabilized (See Figure 3.5). It was also thought that stability is related to the number of residues in the loop region as a longer loop would be expected to allow the protein more flexibility to accommodate surface interactions without disrupting the positions of the helices forming the bundle. For example, all the tethering sites in Loop 3 (18 residues in length) of 1A56 showed stability. In fact, all the sites in loop regions with more than 10 residues resulted in stability that was equal to or greater than that found in the bulk. However, for loops less than 10 residues in length, varied behavior was seen. For example, site 15 (destabilized) and site 25 (stabilized) of 1R69, are found in separate loops of 3 residues in length but have different stability.

Further examination revealed that for short loops, loops less than 10 residues in length, the local structure of the loop must be taken into account. Figure 3.6 shows the different classifications of loop regions. Panel (A) is the long loop just described. In this type of loop, stabilization occurs as long as the tether site is not next to one of the helices. Panel (B) shows a U-shaped loop. If a loop is composed of less than 10 residues, but the tether is placed in a U-shaped loop, then the protein is stabilized. This is the case for Loops 1 and of 1AD6 and Loop 2 of 5MBN. Panel (C) shows a W-shaped loop. For this type the placement of the tether is important. If the tether is placed in the "concave up" portion of the loop, the protein will be stabilized on the surface. If the tether is place in the "concave down" portion of the loop, the protein will be destabilized on the surface.

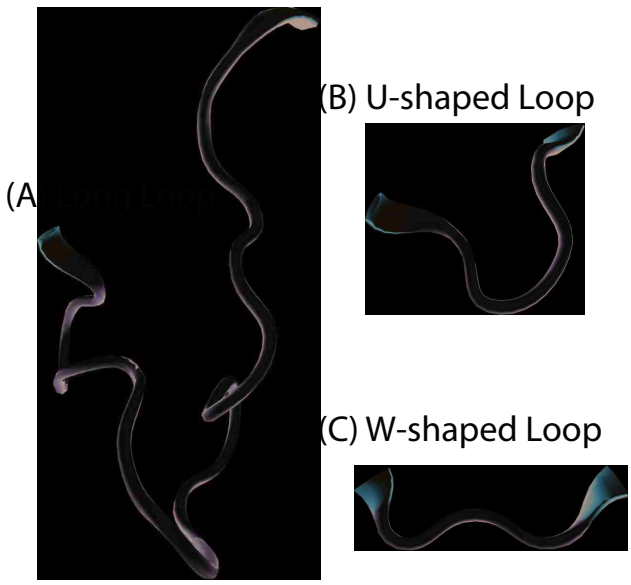


Figure 3.6: Samples of shapes for loop regions: (A) Long loop, (B) U-shaped loop, (C) W-shaped loop

The "loop structure" discussion just described accounts for all of the stability patterns except one. Loop 4 of 1AD6, consisting of sites 145-147, forms a W-shape, but stabilization does not occur when tethered to any of sites involved. Analysis of this anomaly reveals another factor affecting the stability of proteins on surfaces, and the idea is depicted in Figure 3.7. The protein is 1AD6. Panel (a) shows the protein tethered at site 57, panel (b) at site 163, and panel (c) at site 145. For each configuration, the shaded region shows

the volume available for the protein to rotate and vibrate on the surface. Notice that the protein tethered at site 57 forms a V-shape which allows the protein a large amount of volume to rotate an vibrate with respect to the surface. The opposite is true for tethering at site 145. In this configuration, the three helices nearest the surface form a flat base which severely restricts the ability of the protein to rotate and vibrate on the surface. Tethering at site 163 gives the protein a hybrid shape between the two extremes just described. In this configuration, one side of the molecule forms a flat foundation, but the other portion has a V-shape. For each of these sites, the protein is tethered in either a U-loop or the concave-up region of a W-loop and would expected to be stabilized; however, site 145 is destabilized.

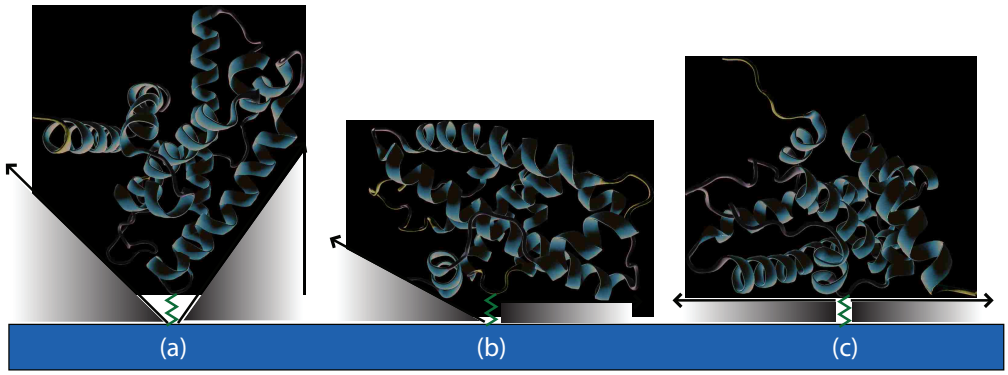


Figure 3.7: Free volume available for rotation for 1AD6 according to tether site. Panel (a): site 57, Panel (b) site 163, and Panel (c) site 145

The origins of the destabilization lies in the restriction of the movement of the protein. At a given temperature, the bond, angle, dihedral and rotational vibrations of the protein seek to populate a characteristic distribution of frequencies. For the protein as a whole (relative to the surface), the amplitudes of these motions are commensurate to the size of the protein meaning that they are large compared the motions of the atoms relative to each other. For tether site 145, the whole-protein vibrations are severely inhibited. In order to populate the desired rotational/vibrational states, the protein unfolded at a lower temperature than would be expected and once unfolded can rotate and vibrate with ease.

Evidence to the fact that the vibrations increase for tethering configurations which restrict rotational movement is found in Table 3.2. Listed are the vibrational entropies, at

240 K and 440 K, for tethering 1AD6 at sites 57, 163, and 145 (See Figure 3.7). The values were calculated assuming a quasi-harmonic approximation [71] using the Wordom analysis package [71]. The lower temperature, 240 K, is below the melting temperature of the protein. The higher temperature, 440 K, is above the melting temperature. At 440 K, the entropies are very similar indicating that the unfolded state of the protein, regardless of tether site, partitions energy into vibrations in roughly equal amounts. At the lower temperature, the vibrational entropy increases as the volume available for rotation decreases. Specifically, the flat base produced by tethering at site 145, which restricts the rotational ability of the protein, has the most vibrational entropy. Site 57, which produces a V-shape with the largest amount of rotational ability, has the least amount of vibrational entropy. The mixed, flat-V shape created by tethering at site 163, which has an intermediate ability to rotate relative to the surface, has an intermediate amount of vibrational entropy.

Table 3.2: Vibrational entropy of 1AD6 for various tether sites

Site	S_{vib} (kJ mol^{-1} K^{-1})	
	240 K	440 K
57	6.89 ± 0.04	18.039 ± 0.005
163	7.78 ± 0.21	18.080 ± 0.001
145	9.63 ± 0.02	18.030 ± 0.001

For the present purposes, the easiest way to determine if the tether site has adequate rotation volume without performing a simulation is to view the protein in a molecular viewer such as VMD. Sites which could be problematic can quickly be discerned using such an approach. Several attempts were made to quantify the ability of the protein to vibrate on the surface, but determining a simple, quantifiable metric which delineates between the types of shapes pictured in Figure 3.7 is difficult. Any averaging of angles or distances, which is usually required for simple metrics, reduces ability to distinguish the difference between site 163 [Panel (b)] and site 145 [Panel (c)]. More sophisticated metrics were investigated as follows.

Free Rotation Volume

Since the lack of free rotation volume of site 145 is believed to be the major reason why protein is destabilized on surfaces, the first metric for quantifying this difference is the measurement of the fraction of free rotation volume. However, the site 145 showed about 78% of fraction of free rotation volume, which is not consistent to the observation. The reason for the failure of this method is stated as follows. When the fraction of free rotation volumes was calculated, the radius and the height of the cylinder are various in each case. The radius is the largest distance between the atom and the axis, while the height is 0.7 times of length between tethering site and the mass center, plus 5.8 $\mathring{A}$. However, with site 145 tethered, some long strands of the protein spread away from the axis enlarge the radius extraordinarily. A lot of volumes of the same or higher height as atoms are also counted in the free rotation volume, but actually, they should not be. The free rotation volume in our consideration should be the part shown in the figure, near the surface.

Free Rotation Angle

The free rotation angle could be a better metric for describing the difference between the site 145 on 1AD6 to all others than the free rotation volume metric. The result in Table 3.3 as the column of "Angle from site" showed a lowest value of free rotation angle of site 145 on protein 1AD6. However, the value of site 163, a site with which the protein tethered and showed partially flat bottom and partially V-shape bottom, is also high. Even though these two sites are distinguishable somehow, the sensitivity of this metric is not quite well. Also, if the definition of the free rotation angle is considered more seriously, the limitation of rotation of a protein is closely but not accurately from those sites that are far from the center line, but the lines through each two of those sites. This angle, which is shown in red in the Figure 3.3, is larger than that formed by site.

Table 3.3: Order parameters for protein rotation

Protein	Location	VFAFR (%)	Angle from site (°)	Angle from center (°)
1R69	Site 14	81.05	36.74	53.48
	Site 15			
	Site 25	76.43	43.22	66.96
	Site 26			
	Site 38	77.17	27.64	52.90
	Site 39			
	Site 53	64.13	40.84	64.50
	Site 55			
1A56	Site 19	81.82	41.65	58.34
	Site 20			
	Site 24			
	Site 33	80.60	36.39	53.96
	Site 36			
	Site 59	67.88	41.25	58.02
	Site 60			
	Site 65			
5MBN	Site 46	79.61	52.17	67.10
	Site 50			
	Site 78	75.56	39.89	57.65
	Site 80			
	Site 81			
	Site 98	77.34	34.51	52.85
	Site 99			
	Site 122	74.52	34.80	53.11
	Site 123			

Table 3.3: Continued

Protein	Location	VFAFR (%)	Angle from site (°)	Angle from center (°)
1AD6	Site 30 Site 31	85.48	50.61	70.09
	Site 57 Site 58	72.08	41.82	62.77
	Site 125 Site 127	97.71	50.60	68.47
	Site 145 Site 146	78.51	14.95	37.09
	Site 163 Site 165	81.56	19.72	48.00
1A2S	Site 27 Site 30 Site 32	84.35	47.32	64.58
	Site 41 Site 46	76.54	28.55	48.47
	Site 64 Site 66 Site 68	79.28	50.77	66.06

The closer of two sites, the similar of angles formed by their center point to angles formed by the two sites could be. The further of two sites, the smaller of the angle formed by their center point than angles formed by those two sites could be. That is, if sites spread evenly in all directions as in case of site 145 tethered the average of free rotation angles is larger but in a small range than the average of angles formed by sites. However, in the case of site 163 tethered, the bottom of the protein is more flat in some direction than the other, those sites that are far from the center line crowded in two ends. Thus, sites in different

ends could be very far from each other, and then the free rotation angle by their center point could be much larger than both of those angles form by the two sites. The result shown in the column of "Angle from center" in Table 3.3 proved the analysis shown above. The site 145 has the least free rotation angle (37.09°), and the next least one the site 163, which has a value of 48.00°. The good sensitivity of this method can be seen. To make this metric as a criterion, the cutoff is needed to be measured from a larger amount of data of different sites.

Though the above analysis concerning tethering loops and vibration/rotation on the surface involves only five proteins, the consistency and logic is such that the following heuristics for designing protein-surface interactions *of alpha-helical, orthogonal-bundle proteins* are presented. It is recognized that these are preliminary and are based upon a limited data set, but formalization provides a starting point for future investigations. Moreover, for a field where little is known, these heuristics provide a first step towards rational design of protein/surface interactions.

1. Long Loops: Tethering in loop regions of greater than 10 residues in length will result in stabilization of the protein on the surface.
2. "U-shaped" loops Tethering in U-shaped loop regions of less than 10 residues in length will result in stabilization if the protein can vibrate freely on the surface.
3. "W-shaped" loops
   (a) Tethering in "concave-up" regions will result in stabilization if the protein can vibrate freely on the surface.
   (b) Destabilized for tethering in "concave-down" regions.

3.3.4 Thermodynamic Analysis

The influence of the surface on the stability of proteins can be explained in terms of common thermodynamic properties. For stable proteins, $\Delta G_f = \Delta H_f - T\Delta S_f < 0$. In other words, the more stable the protein, the more negative the value of ΔG_f or the greater the value of $|\Delta G_f|$. As theory from Dill *et al.* [42, 48] stated before, proteins are always stabilized when

tethered to short-ranged, repulsive surfaces, because the entropic cost of folding is greater in the bulk case than the on the surface.

The results in Figure 3.5 indicate that proteins are not always stabilized when tethered to surfaces as would be expected from the theory just explained. Prior work has shown that the *entropic* portion of the argument is valid, namely that the entropic cost of folding for tethered proteins is less than in the bulk [52, 50, 53], so any destabilization must be an *enthalpic* effect. One of the assumptions upon which the theory is based is that ΔH_f is the same on and off the surface. The validity of this assumption is now addressed.

Table 3.4 shows a summary of ΔG_f, ΔH_f, and $T\Delta S_f$ for each protein in the bulk and tethered to the surface at the same sites depicted in Figure 3.5. For reference, the type of loop for each site is also listed. The temperature in each case is the melting temperature of the protein in the bulk. As such, $\Delta G_f = 0$ for each protein in the bulk. Comparing the ΔG_f values with the corresponding T_m/T_m^* values of Figure 3.5 shows that the data are consistent. Tethering sites which result in an increase in the melting temperature of the protein on the surface compared to the bulk have negative values for ΔG_f. Similarly, sites which result in melting temperatures that are less than the bulk value have positive values for ΔG_f. As T_m and ΔG_f are calculated in two, distinct and independent ways, the agreement between the two values attests to the reliability of the results.

Table 3.4: Thermodynamic quantities of proteins

Protein	Location	ΔG_f (kJ/mol)	ΔH_f (kJ/mol)	$T\Delta S_f$ (kJ/mol)	Shape
1R69	Bulk	0.0	-236.8 ± 1.7	-236.8 ± 1.7	-
	Site 14	-13.7 ± 0.8	-228.0 ± 4.3	-214.3 ± 4.0	W
	Site 15	13.7 ± 0.7	-191.1 ± 4.2	-204.8 ± 3.8	W
	Site 25	-10.0 ± 1.8	-232.5 ± 1.1	-222.5 ± 1.7	W
	Site 26	16.0 ± 0.6	-176.6 ± 5.1	-192.6 ± 4.8	W
	Site 38	-9.5 ± 1.4	-227.9 ± 2.1	-218.4 ± 2.5	W
	Site 39	17.3 ± 0.5	-166.8 ± 4.0	-184.1 ± 4.4	W
	Site 53	-10.7 ± 1.6	-227.5 ± 1.3	-216.9 ± 0.7	W
	Site 55	-14.6 ± 2.4	-225.8 ± 3.0	-211.2 ± 4.8	W
1A56	Bulk	0.0	-105.1 ± 1.7	-105.0 ± 1.9	-
	Site 19	-2.6 ± 1.1	-96.9 ± 5.6	-94.2 ± 4.7	Long
	Site 20	-4.3 ± 0.4	-93.6 ± 2.3	-89.4 ± 2.1	Long
	Site 24	0.5 ± 0.7	-88.2 ± 3.2	-88.7 ± 3.8	Long
	Site 33	1.2 ± 0.4	-77.2 ± 2.4	-78.4 ± 2.4	W
	Site 36	-2.9 ± 0.4	-93.9 ± 4.2	-91.0 ± 3.8	W
	Site 59	-4.1 ± 0.5	-97.8 ± 3.7	-93.7 ± 3.6	Long
	Site 60	-0.8 ± 0.7	-86.7 ± 4.5	-85.9 ± 4.1	Long
	Site 65	-2.4 ± 0.6	-99.2 ± 2.6	-96.8 ± 2.1	Long
1AD6	Bulk	0.0	-502.8 ± 6.9	-502.2 ± 6.6	-
	Site 30	-9.3 ± 1.5	-494.3 ± 13.8	-485.0 ± 14.3	U
	Site 31	-11.9 ± 2.2	-489.9 ± 9.8	-478.0 ± 9.8	U
	Site 57	-6.5 ± 1.0	-466.7 ± 13.4	-460.2 ± 13.7	U
	Site 58	-0.7 ± 2.4	-432.1 ± 37.7	-431.4 ± 36.0	U
	Site 125	-8.2 ± 2.5	-487.9 ± 18.9	-479.7 ± 16.5	Long
	Site 127	-13.0 ± 1.4	-494.1 ± 8.1	-481.1 ± 9.4	Long

Table 3.4: Continued

1AD6	Site 145	30.2 ± 0.8	-282.2 ± 10.9	-312.4 ± 10.3	W
	Site 146	40.0 ± 0.7	-262.0 ± 11.51	-302.0 ± 12.1	W
	Site 163	-10.1 ± 1.9	-497.0 ± 24.7	-487.0 ± 25.6	W
	Site 165	28.4 ± 1.8	-264.6 ± 13.0	-293.0 ± 12.9	W
1A2S	Bulk	0.0	-128.2 ± 1.7	-128.2 ± 2.0	-
	Site 27	-4.4 ± 0.7	-125.5 ± 3.2	-121.1 ± 3.2	Long
	Site 30	-0.5 ± 0.7	-112.0 ± 4.6	-111.6 ± 5.0	Long
	Site 32	1.0 ± 0.6	-105.7 ± 1.6	-106.7 ± 1.2	Long
	Site 41	1.0 ± 0.5	-107.4 ± 2.0	-108.4 ± 2.3	W
	Site 46	-4.5 ± 1.0	-131.7 ± 3.4	-127.3 ± 4.0	W
	Site 64	-0.4 ± 0.5	-125.4 ± 1.2	-125.0 ± 1.4	Long
	Site 66	-0.9 ± 0.3	-112.2 ± 1.8	-111.2 ± 1.6	Long
	Site 68	1.3 ± 0.4	-120.4 ± 4.3	-121.7 ± 3.9	Long
5MBN	Bulk	0.0	-577.1 ± 11.5	-577.0 ± 11.6	-
	Site 46	11.5 ± 1.1	-468.2 ± 7.4	-479.7 ± 8.0	W
	Site 50	-6.7 ± 2.9	-553.3 ± 16.6	-546.6 ± 16.0	W
	Site 78	-14.9 ± 2.8	-517.8 ± 25.2	-502.8 ± 27.8	U
	Site 80	-8.4 ± 3.4	-516.6 ± 11.7	-508.2 ± 14.9	U
	Site 81	-10.3 ± 2.4	-533.9 ± 14.0	-523.6 ± 16.1	U
	Site 98	-5.3 ± 5.0	-529.4 ± 22.2	-524.1 ± 27.1	W
	Site 99	17.5 ± 0.2	-456.5 ± 10.4	-474.0 ± 10.3	W
	Site 122	-12.3 ± 2.0	-545.7 ± 16.5	-533.4 ± 18.2	W
	Site 123	2.9 ± 1.9	-504.0 ± 13.3	-506.8 ± 12.9	W

As reported in the Table, all the surface-tethered proteins studied in this work have $T\Delta S_f$ values that are statistically equal to or larger than (less negative) the bulk values. This agrees with the theory as a reduction in the loss of entropy causes a decrease in ΔG_f, indicating stabilization. For tethering sites that stabilize the protein on the surface, the

value for ΔH_f is approximately equal in the bulk and on the surface. For sites that result in destabilization, the ΔH_f value is greater (less negative) on the surface than in the bulk. Thus, in the limit that ΔH_f is equal on and off the surface (the situation described by the theory), stabilization occurs. Away from this limit, destabilization occurs.

Further analysis provides additional insights. In general, the change in enthalpy upon folding is related to both the enthalpy of the folded state and the unfolded state ($\Delta H_f = H_{\text{folded}} - H_{\text{unfolded}}$). Figure 3.8 shows the influence of the surface on the folded-state and unfolded-state enthalpies for 1R69 at $T = T^*$. Depicted is the difference between the enthalpy on the surface and in the bulk for both folded and unfolded protein. Specifically, $\delta H_{\text{folded}} \equiv H_{\text{folded}}^{\text{surface}} - H_{\text{folded}}^{\text{bulk}}$ and $\delta H_{\text{unfolded}} \equiv H_{\text{unfolded}}^{\text{surface}} - H_{\text{unfolded}}^{\text{bulk}}$. The symbol δ is used in place of Δ to prevent confusion with the *change that occurs upon folding* (Δ) with the *difference between the value on and off the surface* (δ). If $H_{\text{folded}}^{\text{surface}} \approx H_{\text{folded}}^{\text{bulk}}$ then $\delta H_{\text{folded}} \approx 0$ and similarly for the unfolded values. If the surface stabilizes the state (either folded or unfolded), the corresponding δ-value will be negative. If the surface destabilizes the state, the corresponding δ-value will be positive. For convenience, the corresponding values of ΔG_f are also shown on the figure.

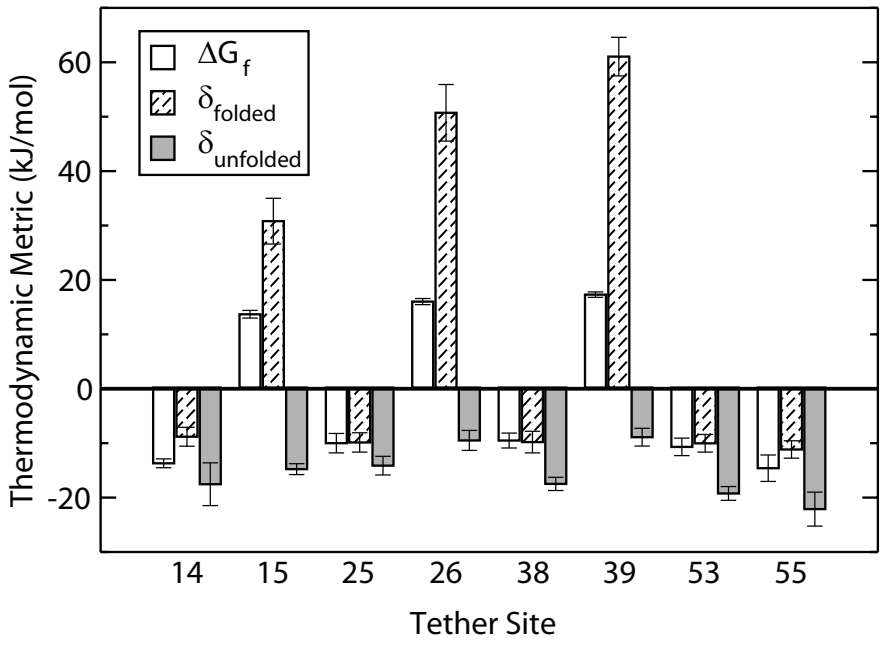


Figure 3.8: Influence of surface on the enthalpy of the folded and unfolded state of 1R69 at $T = T^*$

Figure 3.8 is evidence that tethering configurations which result in destabilization of the protein are caused by the effects of the surface on the *folded* state of the protein. In each case, the data show that the enthalpy of the unfolded state on the surface is less (more favorable) than the enthalpy of the unfolded state in the bulk. The reason is straightforward. In the bulk, when proteins are unfolded, the entropy drives the system to sample configurations with little structure. The enthalpy drives the system to fold to increase hydrogen bonding and reduce hydrophobic/hydrophilic contacts in favor of hydrophobic/hydrophobic and hydrophilic/hydrophilic contacts. Enthalpically-favorable contacts can only form as the distance between complimentary sites decreases. When the protein is on the surface, the average distance between sites of the unfolded protein is reduced which causes a reduction of the enthalpy.

In contrast to the unfolded state, the surface affects the folding state in ways that do not always stabilize the protein. For sites that result in stabilization, $\delta H_{\text{folded}} < 0$ suggesting that the surface improves the ability of the protein to make favorable contacts. The degree of stabilization is similar to that seen for the unfolded state, and the result is that $\Delta H_f^{\text{surface}} \approx \Delta H_f^{\text{bulk}}$ as previously described (see Table 3.4). For destabilized configurations, $\delta H_{\text{folded}} > 0$ meaning that the surface inhibits the formation of favorable contacts. The extent of destabilization of the folded state is so great that $\Delta H_f^{\text{surface}} > \Delta H_f^{\text{bulk}}$.

The above analysis improves current understanding of protein/surface interactions significantly and provides the most complete picture to date of the the thermodynamics involved. To summarize the findings, entropy works to stabilize tethered proteins on the surface as expected from theory. Unexpectedly, proteins will be stabilized or destabilized depending upon the interaction of the *folded state* of the protein with the surface. Said another way, the surface affects the *unfolded* state entropically but affects the *folded* state enthalpically. Sites in long loops or in U-shaped and the concave-up regions of W-shaped loop which have adequate free rotation volume, allow the folded state of the protein to exists on the surface as is does in the bulk and the result is entropic stabilization. Sites in the concave-down region of W-shaped loops and those which restrict rotation and vibration, inhibit the ability of the protein to exist in its native state and the result is enthalpic destabilization.

3.4 Summary

The results in this work show, for the first time, that protein stability on surfaces can be correlated to tertiary structural elements for alpha-helical, orthogonal bundle proteins. The important factors to consider when selecting a tether site are the shape of the loop region and the volume available for the protein to rotate on the surface. For loop regions that have large rotation volumes, sites can always be found which stabilize the protein. A thermodynamic analysis shows that proteins are always stabilized entropically when tethered to surface and that any destabilization is an enthalpic effect. Taken as a whole, the results offer hope for rational design of protein surface interactions and a rigorous thermodynamic understanding of the origins of stabilization/destabilization of surface.

Future efforts are needed to fully understand the implications of the results found in this work. The next major step is to investigate other classes of tertiary structure to determine if correlations between tether site and stability can be found. Moreover, additional work is needed to further characterize the amount of rotation/vibration volume needed by a protein to remain stable on the surface. Though preliminary, the results presented provide the needed starting point for these future investigations.

4 SURFACE INDUCED CHANGES TO FOLDING MECHANISM

As introduced before, when a tethering is done to a site involved in an intermediate state, the folding mechanism could be changed. To understand how surfaces change folding mechanism of alpha-helical, orthogonal-bundle proteins, a multi-metastates protein 7LZM is introduced. Thermodynamic analyses are also accomplished for this protein, and the structures are compared in different temperatures on and off the surface.

4.1 Introduction

Proteins with a small number of residues are always two-state folders, which means every part of the protein melts at the same temperature. However, some proteins may have multiple state configurations along the temperature domain, which means, different parts of a protein melt at different temperatures. Between each melting temperature, there is a metastable intermediate state.

In this part of work, the surface effect on the stability and folding mechanism of multistate proteins is talked.

4.2 Methods

We focus on a peptide, Enterobacteria phage t4 (7LZM), which has two intermediate states in bulk. The structue is shown in Figure 4.1. To compare the change of melting temperature of each part, this protein is simulated in the bulk (no surface) and tethered to a surface at each loop region. By categorizing native contacts of each secondary structure in 7LZM, the change of intermediate states or even the folding mechanism can be analyzed.

Considering computational efficiency, the Go-like model of Karanicolas and Brooks [11, 12, 13, 14, 15] is used. Input files were generated from the MMTSB:`www.mmtsb.org`.

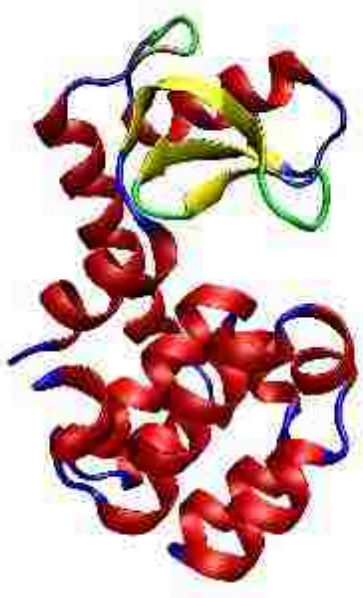
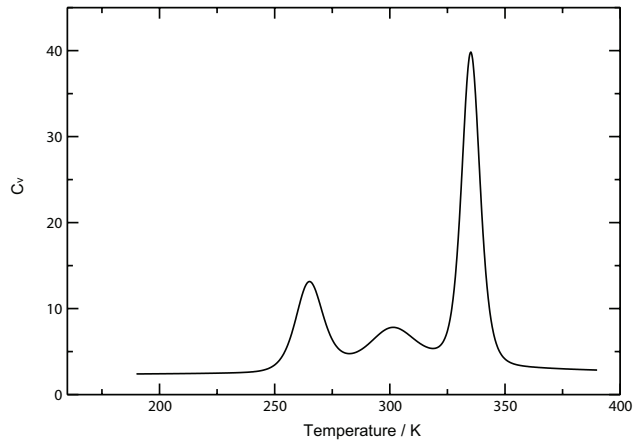


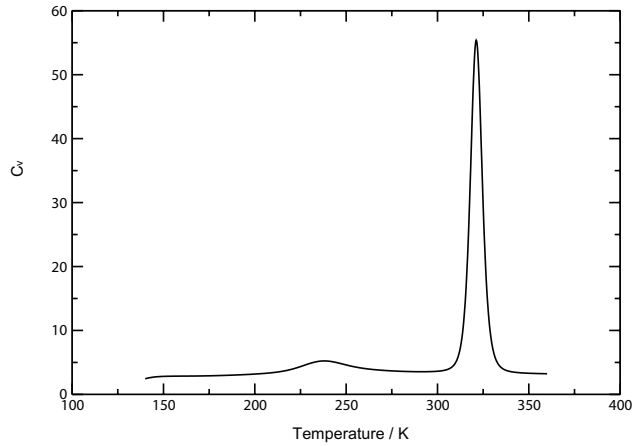
Figure 4.1: Native conformation of protein 7LZM

This protein model extends earlier Go models by introducing different energy scales to describe hydrogen bonding between side chains and sequence-dependent virtual dihedral potentials to keep proteins in appropriate conformations. The resulting energy surface can mimic that of the real protein more closely than earlier models, which employed fewer energy scales or targeted specific encoding of the backbone structure with virtual dihedral potentials.[11, 12, 13, 14, 15] This model has been shown to give good agreement between simulation and experimental folding studies.

To prevent the simulation from becoming trapped in local energy minima,[72] simulations were performed using the replica exchange (RE) algorithm [56, 57]. For each protein, 24 replicas were used. The canonical ensemble was generated using the Nose Hoover Chain [66, 67, 68] integration method, with 3 thermostats of mass $10^{-26} kg \cdot \mathring{A}^2$. The time step was $1 fs$, and each simulation contained 10 million steps of equilibrium followed by 30 million steps of production. Temperature steps between boxes range from 2.5 to 10 degrees accross the temperature range.



(a)



(b)

Figure 4.2: The heat capacity curve in temperature domain of protein 7LZM (a) in the bulk and (b) on the surface

4.3 Results

As shown in Figure 4.2(a), there are three peaks along the heat capacity curve for a temperature range of the protein in bulk. As shown in figure 4.2(b), the heat capacity curve of 7LZM tethered to a surface with the residue site number of 91 has one peak less than in bulk. That is, the metastate conformations are different when it is tethered on a surface at that site. Also, the position of the high peak changes when the protein is tethered to the surface at that site. Based on that investigation, it was hypothesized that the folding mechanism of 7LZM is changed on a surface.

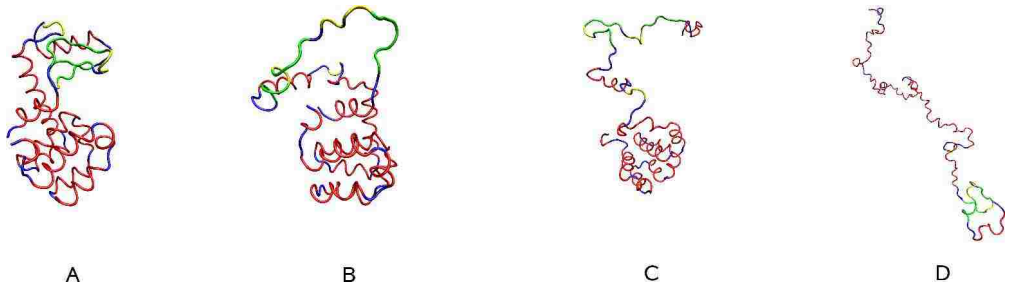


Figure 4.3: Conformations of protein 7LZM in the bulk in different temperatures: (A) 190 K; (B) 280 K; (C) 320 K; (D) 380 K

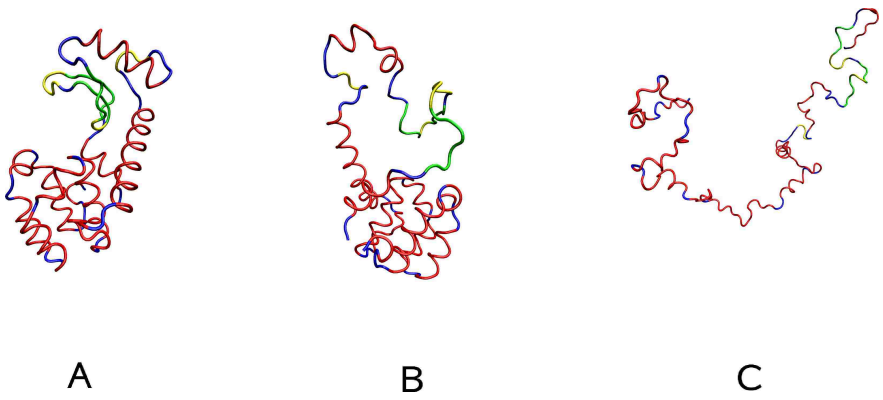


Figure 4.4: Conformations of protein 7LZM on the surface in different temperatures: (A) 190 K; (B) 280 K; (C) 380 K

To validate this hypothesis, it is recorded and compared that comformations of each state for proteins in bulk and on surfaces in Figure 4.3 and 4.4. Comparing these two figures, it was found that proteins are folded at the lowest temperature ($190K$) both in the bulk and on the surface and unfolded, or melted, at the highest temperatures($380K$) in both conditions. Also, the metastate of the protein on the surface at $280K$ (Figure 4.4(B)) is similar in shape to the first metastate (Figure 4.3(B)) in the bulk at the same temperature. The most important characteristic of these two metastates is that the upper part of the protein still keeps a "ring" shape. That means, at $280K$, all tertiary structures in the upper part have melted, but the contacts formed by the ending sections of the ring and the lower part of protein are still stable.

The protein on the surface lacks of the second kind of metastate (at $320K$) as in bulk. At this temperature, the upper ring of the protein in the bulk opens, as shown in Figure 4.3(C). However, such a transformation of the protein does not show up on the surface. Furthermore, since conformation at ($380K$) are the same as shown above, it is obvious that some native contacts melted at the same higher temperature as the remaining parts. The change of the melting temperature lead to the disappearance of one metastate when the protein is tethered to the surface. In other words, tethering 7LZM to the surface at this site, changes the mechanism of folding by raising the melting point of some parts.

Table 4.1: Categorization and lengths of parts for 7LZM

7LZM								
	Coil 1	Helix 1	Coil 2	Beta 1	Turn 1	Coil 3	Beta 2	Turn 2
Sites	1-2	3-11	12-13	14-19	20-23	24	25-28	29-30
	Beta 3	Coil 4	Helix 2	Coil 5	Turn 3	Coil 6	Helix 3	Coil 7
Sites	31-34	35-38	39-50	51-53	54-56	57-59	60-80	81-82
	Helix 4	Coil 8	Helix 5	Coil 9	Helix 6	Coil 10	Helix 7	Coil 11
Sites	83-90	91-92	93-106	107	108-113	114	115-123	124-125
	Helix 8	Coil 12	Helix 9	Coil 13	Helix 10	Coil 14	Helix 11	Coil 15
Sites	126-134	135-136	137-141	142	143-155	156-158	159-161	162

To get explicit evidence to describe this change, native contacts for each part of the protein along the temperature range are monitered. First, parts of the protein as in Table 4.1 were categorized, and then the native contacts of each pair of segments in a temperature range were recorded. The categorization is based on secondary structure motifs like α helix, β sheets, and coils/ turnings between them.

The solid line Figure 4.5 is the curve of native contacts between Turn 3 (site 54-56) and Coil 6 (site 57-59) along a temperature range, and the dashed line is its scaled derivative. As shown in this figure, the number of native contacts decrease as the temperature increased, which means the two coordinated parts melt. The temperature of the peak point in the derivitive curve is the melting point of this small section in the protein.

All of the derivative lines were plotted in the same figure for both in bulk and on the surface in Figure 4.6 and 4.7. There are three peaks of the protein in bulk and two peaks on the surface, which is consistant with the melting temperature results. That means there

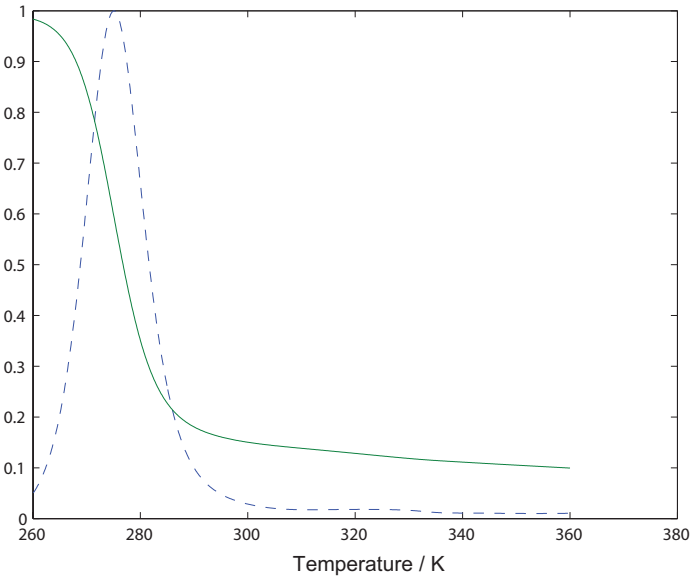


Figure 4.5: Native contacts and the derivative curve for Turn 3 and Coil 6

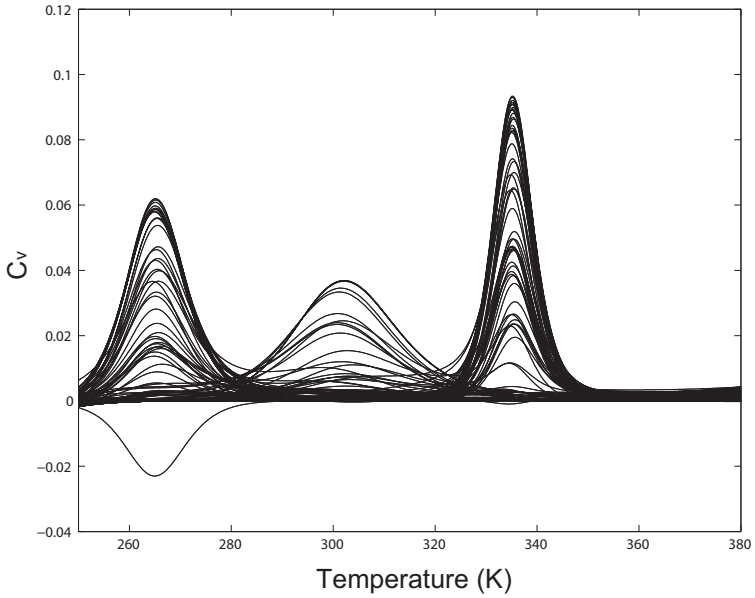


Figure 4.6: The derivative curve of all native contacts of 7LZM in bulk.

must be three groups of native contacts, while each group of native contacts breaks at a given temperature. Then those native contacts were grouped based on their peak positions. Also groups of native contacts were got according to their heat capacity peak position, as shown in Table 4.2. In this table, there are three groups of the protein in the bulk, and two groups on the surface. That is consistant with the melting temperature results and shapes of metastates shown above. Also, there is little difference in group I in both conditions.

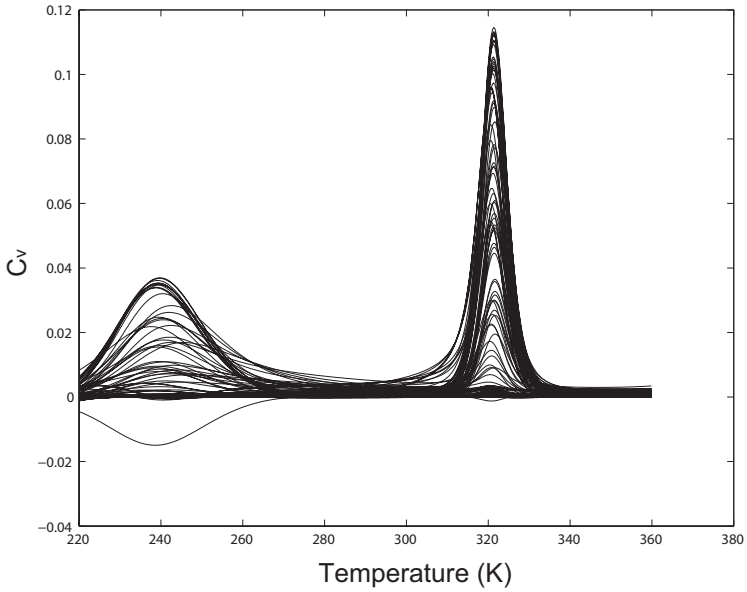


Figure 4.7: The derivative curve of all native contacts of 7LZM on surface with site 91 tethered

The group II of the protein on the surface is the combination of group II and group III of the protein in the bulk. It is a proof to the prediction that the folding mechanism of the protein on the surface changes due to the change in melting temperature of a small part of the protein. That part, as shown in Table 4.2 Group II of protein in the bulk, is the connection between the ring and the lower part of the protein, as shown in Figure 4.3(B) and Figure 4.4(B). The melting of this part of the protein provided the conformation change as shown from Figure 4.3(B) to 4.3(C).

4.4 Summary

7LZM, a multi-states protein, changes the melting temperature of some of its parts when it is tethered to a surface. Also, through the comparison of the conformation shape in the bulk and on the surface, it was found out that one metastate disappeared on surfaces with site 91 tethered. That is because the melting temperature of the second group of native contacts increased and the heat capacity peaks merge with one another. The change of melting temperature of some parts of a protein, affects the number of metastates, which leads to the change of folding mechanism.

Table 4.2: Groups in melting temperatures for 7LZM

7LZM in the Bulk							
GROUP I	H1-C2	H1-T2	C2-B2	C2-T2	C2-H3	B1-B1	B1-C3
	B1-B2	B1-T2	B1-B3	B1-H2	B1-T3	B1-C6	B1-H3
	T1-T1	T1-B2	T1-H9	T1-C13	C3-B3	C3-C4	B2-T2
	B2-B3	B2-C4	B2-H2	B2-C6	B2-H3	T2-H3	B3-H2
	B3-H3	C4-H2	H2-H2	H2-C5	H2-T3	H2-C6	H2-H3
	C5-C6	C5-H3	T3-C6	C6-H3			
GROUP II	C1-H1	C1-C14	C1-H11	C1-C15	H1-H1	H1-H3	H1-H5
	H1-H10	H1-C14	H1-H11	H5-H11	H10-H11	H11-C15	
GROUP III	H3-H3	H3-C7	H3-H4	H3-H5	H3-H6	C7-H4	C7-H6
	H4-H4	H4-C8	H4-H5	H4-H6	H4-H7	C8-H5	C8-H8
	C8-H10	H5-H5	H5-C9	H5-H6	H5-C10	H5-H7	H5-H8
	H5-H9	H5-C13	H5-H10	H5-C14	C9-H6	H6-H6	H6-C10
	H6-H7	C10-H7	C10-H8	C10-C12	C10-H9	H7-H7	H7-C11
	H7-H8	H7-H10	C11-H8	H8-H8	H8-C12	H8-H9	H8-H10
	C12-H9	H9-H9	H9-C13	H9-H10	C13-H10	H10-H10	H10-C14
7LZM on the Surface							
GROUP I	H1-T2	C2-B2	C2-T2	B1-B1	B1-C3	B1-B2	B1-T2
	B1-B3	B1-H2	B1-T3	B1-C6	B1-H3	T1-T1	T1-C3
	T1-B2	T1-H9	T1-C13	C3-B3	C3-C4	B2-T2	B2-B3
	B2-C4	B2-H2	B2-C6	B2-H3	T2-H3	B3-H2	B3-H3
	C4-H2	H2-H2	H2-C5	H2-T3	H2-C6	H2-H3	C5-C6
	C5-H3	T3-C6	C6-H3				
GROUP II	C1-H1	C1-C14	C1-H11	C1-C15	H1-H1	H1-C2	H1-H3
	H1-H5	H1-H10	H1-C14	H1-H11	C2-H3	B3-C4	H3-H3
	H3-C7	H3-H4	H3-H5	H3-H6	C7-H4	C7-H6	H4-H4
	H4-C8	H4-H5	H4-H6	H4-H7	C8-H5	C8-H8	C8-H10
	H5-H5	H5-C9	H5-H6	H5-C10	H5-H7	H5-H8	H5-H9
	H5-C13	H5-H10	H5-C14	H5-H11	C9-H6	H6-H6	H6-C10
	H6-H7	C10-H7	C10-H8	C10-C12	C10-H9	H7-H7	H7-C11
	H7-H8	H7-H10	C11-H8	H8-H8	H8-C12	H8-H9	H8-H10
	C12-H9	H9-H9	H9-C13	H9-H10	C13-H10	H10-H10	H10-C14
	H10-H11	C14-H11	C14-C15	H11-C15			

5 CONCLUSION

The stability of protein changes when they are tethered on surfaces. For α-helical, two-folded state orthogonal bundle proteins, predictable stabilization patterns can be found. If a protein is tethered with a loop that have large free rotation volume, at least one stabilizing site can be found in the loop region. The position of the stabilizing site is correlated to the shape of the loop.

5.1 Summary

The stability of protein changes when they are tethered on surfaces. For α-helical, two-folded state orthogonal bundle proteins, predictable stabilization patterns can be found. If a protein is tethered with a loop that have large free rotation volume, at least one stabilizing site can be found in the loop region. The position of the stabilizing site is correlated to the shape of the loop.

In a long loop, it is not easy to find a site where the protein tethered on a surface that distabilizes the protein, until sites which are very near to the neighbour helix are tested. For a u-shape loop, most sites in the loop region behave in a stabilizing way. For a w-shape loop, the outer sites are always stabilizing sites for a protein, while the inner ones always perform destabilization. In another word, if the tethering site bends out of the protein molecule bulk, the protein is always stabilized, while if the tethering site bend into the protein molecule bulk, the protein is destabilized. Tethering with sites at the outer position provides more space for protein rotation, thus more stabilization than with sites bend into the protein.

Since the entropy part always helps to stabilize the protein on surfaces, Dill's theory is right when enthalpy does not change. However, the enthalpy part changes in several cases, and always in a destabilization direction, which is not the same as assumed in Dill's theory.

For a multi-states protein, like 7LZM, the melting temperature of some parts changes when it is tethered on the surface. Also, through the comparison of the conformation shape as in the bulk and on the surface, it is noticed that one metastate disappeared on surfaces with site 91 tethered. That is because the melting temperature of second group of native contacts increased to the temperature of the remaining part. The change in melting temperature of some parts of the protein, leads to the change of metastates, which results in the change of folding mechanism.

5.2 Future Work

To investigate more truth about protein/surface interaction, it is needed to simulate proteins from other tertiary structure motifs. According to the CATH classification method, several secondary and tertiary structure motifs shown in table 5.1 are waiting for further study. Take up-down bundle as an example, based on the conclusion from this work, proteins in up-down bundle tertiary structure should be stabilized on surfaces if a convex site is chosen to tether with, due to the large free rotation angle of each loop region. That is because, in these proteins, strands are in the up and down directions which forms acute angles.

Table 5.1: Protein structure motifs for further study

Secondary	Tertiary
Mainly α	Up-down Bundle
	Alpha Horseshoe
	Alpha solenoid
	Alpha/alpha barrel
Mainly β	Ribbon
	Single Sheet
	Beta Barrel
	Roll
Mixed $\alpha - \beta$	2-Layer Sandwich
	Apha-Beta Barrel

References

[1] G MacBeath and SL Schreiber. Printing proteins as microarrays for high-throughput function determination. *Science*, 289:1760–1763, 2000.

[2] H Zhu, M Bilgin, R Bangham, D Hall, A Casamayor, P Bertone, N Lan, R Jansen, S Bidlingmaier, T Houfek, T Mitchell, P Miller, RA Dean, M Gerstein, and M Snyder. Global analysis of protein activities using proteome chips. *Science*, 293:2101–2105, 2001.

[3] U Bilitewski. Protein-sensing assay formats and devices. *Analytica Chimica Acta*, 568:232 – 247, 2006.

[4] H Zhu and M Snyder. Protein chip technology. *Current Opinion in Chemical Biology*, 7:55 – 63, 2003.

[5] M Uttamchandani and SQ Yao. Peptide microarrays: Next generation biochips for detection, diagnostics and high-throughput screening. *Current Pharmaceutical Design*, 14:2428–2438, 2008.

[6] M Cretich, F Damina, G Pirria, and M Chiari. Protein and peptide arrays: Recent trends and new directions. *Biomolecular Engineering*, 23:77–88, 2006.

[7] T Joos and J Bachmann. Protein microarrays: potentials and limitations. *Frontiers in bioscience*, 14:4376–4385, 2009.

[8] NK Adam. *The Physics and Chemistry of Surfaces*. Oxford University Press, London, 1941.

[9] DF Cheesman and JT Davies. Physicochemical and Biological Aspects of Proteins at interfaces. *Advance Protein Chemistry*, pages 439–501, 1954.

[10] A Rothen. Films of protein in biological processes. *Advance Protein Chemistry*, pages 123–137, 1947.

[11] J Karanicolas and III CL Brooks. The origins of asymmetry in the folding transition states of protein L and protein G. *Protein Science*, 11:2351–2361, 2002.

[12] J Karanicolas and III CL Brooks. Improved Go-like models demonstrate the robustness of protein folding mechanisms towards non-native interactions. *Journal of Molecular Biology*, 334:309–325, 2003.

[13] J Karanicolasa and III CL Brooks. The structural basis for biphasic kinetics in the folding of the WW domain from a formin-binding protein: Lessons for protein design? *Proceedings of the National Acamedy of Sciences of the United States of America*, 100:3954–3959, 2003.

[14] J Karanicolasa and III CL Brooks. Integrating folding kinetics and protein function: Biphasic kinetics and dual binding specificity in a WW domain. *Proceedings of the National Acamedy of Sciences of the United States of America*, 101:3432–3437, 2004.

[15] Jr. RD Hills and III CL Brooks. Insights from Coarse-Grained Go Models for Protein Folding and Dynamics. *International Journal of Molecular Sciences*, 10:889–905, 2009.

[16] T Kodadek. Protein microarrays: prospects and problems. *Chemistry & Biology*, 8:105–115, FEB 2001.

[17] BB Haab, MJ Dunham, and PO Brown. Protein microarrays for highly parallel detection and quantitation of specific proteins and antibodies in complex solutions. *Genome Biology*, 2, 2001.

[18] Y Soen, DS Chen, DL Kraft, MM Davis, and PO Brown. Detection and characterization of cellular immune responses using peptide-MHC microarrays. *Plos Biology*, 1:429–438, 2003.

[19] KL Hsu, KT Pilobello, and LK Mahal. Analyzing the dynamic bacterial glycome with a lectin microarray approach. *Nature Chemical Biology*, 2(3):153–157, MAR 2006.

[20] I Balboni, SM Chan, M Kattah, JD Tenenbaum, AJ Butte, and PJ Utz. Multiplexed protein array platforms for analysis of autoimmune diseases. *Annual Review of Immunology*, 24:391–418, 2006.

[21] P Angenendt, J Glklera, D Murphya, H Lehracha, and DJ CahillCorresponding. Toward optimized antibody microarrays: a comparison of current microarray support materials. *Analytical Biochemistry*, 309:253–260, 2002.

[22] P Angenendt, J Glokler, J Sobek, H Lehrach, and DJ Cahill. Next generation of protein microarray support materials: Evaluation for protein and antibody microarray applications. *Journal of Chromatography A*, 1009:97–104, 2003.

[23] M Feldmann. Development of anti-TNF therapy for rheumatoid arthritis. *Nature Reviews Immunology*, 2:364–371, 2002.

[24] M Feldmann and L Steinman. Design of effective immunotherapy for human autoimmunity. *Nature*, 435:612–619, 2005.

[25] J Vilcek and M Feldmann. Historical review: Cytokines as therapeutics and targets of therapeutics. *Trends in Pharmacological Sciences*, 25(4):201–209, APR 2004.

[26] JJ Gray. The interaction of proteins with solid surfaces. *Current Opinion in Structural Biology*, 14:110–115, 2004.

[27] P Peluso, DS Wilson, D Do, H Tran, M Venkatasubbaiah, D Quincy, B Heidecker, K Poindexter, N Tolani, M Phelan, K Witte, LS Jung, P Wagner, and S Nock. Optimizing antibody immobilization strategies for the construction of protein microarrays. *Analytical Biochemistry*, 312:113–124, 2003.

[28] R Wacker, H Schroder, and CM Niemeyer. Performance of antibody microarrays fabricated by either DNA-directed immobilization, direct spotting, or streptavidin-biotin attachment: a comparative study. *Analytical Biochemistry*, 330:281–287, 2004.

[29] N Nath, R Hurst, B Hook, P Meisenheimer, KQ Zhao, N Nassif, RF Bulleit, and DR Storts. Improving protein array performance: Focus on washing and storage conditions. *Journal of Proteome Research*, 7:4475–4482, 2008.

[30] O Stoevesandt, MJ Taussig, and M He. Protein microarrays: high-throughput tools for proteomics. *Expert Review of Proteomics*, pages 145–157, 2009.

[31] K Bssow, D Cahill, W Nietfeld, D Bancroft, E Scherzinger, H Lehrach, and G Walter. A method for global protein expression and antibody screening on high-density filters of an arrayed cDNA library. *Nucleic Acids Research*, 26, 1998.

[32] C Mateo, G Fernndez-Lorente, O Abian, R Fernndez-Lafuente, and JM Guisn. Multifunctional epoxy supports: A new tool to improve the covalent immobilization of proteins. The promotion of physical adsorptions of proteins on the supports before their covalent linkage. *Biomacromolecules*, 1:739–745, 2000.

[33] E Delamarche, A Bernard, H Schmid, A Bietsch, B Michel, and H Biebuyck. Microfluidic networks for chemical patterning of substrate: Design and application to bioassays. *Journal of the American Chemical Society*, 120:500–508, 1998.

[34] KL Prime and GM Whitesides. Adsorption of Proteins onto Surfaces Containing End-Attached Oligo(Ethylene Oxide)- A Model System Using Self-Assembled Monolayers. *Journal of the American Chemical Society*, 115:10714–10721, 1993.

[35] JC Love, LA Estroff, JK Kriebel, RG Nuzzo, and GM Whitesides. Self-assembled monolayers of thiolates on metals as a form of nanotechnology. *Chemical Reviews*, 105:1103–1169, 2005.

[36] AP Le Brun, SA Holt, DS Shah, CF Majkrzak, and JH Lakey. Monitoring the assembly of antibody-binding membrane protein arrays using polarised neutron reflection. *European Biophysics Journal*, 37:639–45, 2008.

[37] VP Raut, MA Agashe, SJ Stuart, and RA Latour. Molecular dynamics simulations of peptide-surface interactions. *Langmuir*, 22:2402–2402 , 2006.

[38] M Agashe, VR, SJ Stuart, and RA Latour. Molecular simulation to characterize the adsorption behavior of a fibrinogen gamma-chain fragment. *Langmuir*, 21:1103–1117, 2005.

[39] J Zhou, S Chen, and S Jiang. Orientation of adsorbed antibodies on charged surfaces by computer simulation based on a united-residue model. *Langmuir*, 19:3472–3478, 2003.

[40] P Mulheran and K Kubiak. Protein adsorption mechanisms on solid surfaces: lysozyme-on-mica. *Molecular Simulation*, 35:561 – 566, 2009.

[41] Y Sun, WJ William, and LA Robert. Prediction of the orientations of adsorbed protein using an empirical energy function with implicit solvation. *Langmuir*, 21:5616–5626, 2005.

[42] HX Zhou and KA Dill. Stabilization of proteins in confined spaces. *Biochemistry*, 40:11289–11293, 2001.

[43] J Zhou, S Chen, and S Jiang. Orientation of adsorbed antibodies on charged surfaces by computer simulation based on a united-residue model. *Langmuir*, 19:3472–3478, 2003.

[44] S Ravichandran, JD Madura, and J Talbot. A Brownian dynamics study of the initial stages of hen egg-white lysozyme adsorption at a solid interface. *The Journal of Physical Chemistry B*, 105:3610–3613, 2001.

[45] F Carlsson, E Hyltner, T Arnebrant, M Malmsten, and P Linse. Lysozyme adsorption to charged surfaces. A Monte Carlo study. *Journal of Physical Chemistry B*, 108:9871–9881, 2004.

[46] V Castells and PR Van Tassel. Conformational transition free energy profiles of an adsorbed, lattice model protein by multicanonical Monte Carlo simulation. *Journal of Chemical Physics*, 122, 2005.

[47] S Yang V Castells and PR Van Tassel. Surface-induced conformational changes in lattice model proteins by Monte Carlo simulation. *Physical Review E*, 65, 2002.

[48] F Fang and I Szleifer. Controlled release of proteins from polymer-modified surfaces. *Proceedings of the National Acamedy of Sciences of the United States of America*, 103:5769–5774, 2006.

[49] M Friedel, A Baumketner, and JE Shea. Effects of surface tethering on protein folding mechanisms. *Proceedings of the National Acamedy of Sciences of the United States of America*, 103:8396–8401, 2006.

[50] M Friedel, A Baumketner, and JE Shea. Stability of a protein tethered to a surface. *Journal of Chemical Physics*, 126, 2007.

[51] Z Zhuang, AI Jewett, P Soto, and JE Shea. The effect of surface tethering on the folding of the src-SH3 protein domain. *Physical Biology*, 6:015004, 2009.

[52] IV TA Knotts, N Rathore, and JJ de Pablo. Structure and stability of a model three-helix-bundle protein on tailored surfaces. *Proteins-Structure Function and Bioinformatics*, 61:385–397, 2005.

[53] TA Knotts IV, N Rathore, and JJ de Pablo. An entropic perspective of protein stability on surfaces. *Biophysical Journal*, 94:4473–4483, 2008.

[54] S Kumar, D Bouzida, RH Swendsen, PA Kollman, and JM Rosenberg. The weighted histogram analysis method for free-energy calculations on biomolecules .1. The method. *Journal of Computational Chemistry*, 13:1011–1021, 1992.

[55] HS Chan. Modeling protein density of states: Additive hydrophobic effects are insufficient for calorimetric two-state cooperativity. *Proteins-Structure Function and Genetics*, 40:543–571, 2000.

[56] ME Fisher and M Randeria. Location of Renormalization-Group Fixed-Points. *Physical Review Letters*, 56:2333, 1986.

[57] K Hukushima and K Nemoto. Exchange Monte Carlo method and application to spin glass simulations. *Journal of the Physical Society of Japan*, 65:1604–1608, 1996.

[58] W Gropp, E Lusk, N Doss, and A Skjellum. A high-performance, portable implementation of the mpi message passing interface standard. *Parallel Computing*, 22:789 – 828, 1996.

[59] W Gropp and E Lusk. A high-performance mpi implementation on a shared-memory vector supercomputer. *Parallel Computing*, 22:1513 – 1526, 1997.

[60] AL Cuff, I Sillitoe, T Lewis, OC Redfern, R Garratt, J Thornton, and CA Orengo. The CATH classification revisited–architectures reviewed and new ways to characterize structural divergence in superfamilies. *Nucleic Acids Research*, 2008.

[61] CA Orengo, AD Michie, DT Jones, MB Swindells, and JM Thornton. CATH: A Hierarchic Classification of Protein Domain Structures. *Structure*, 5:1093–1108, 1997.

[62] TJ Schmitt, JE Clark, and IV TA Knotts. Thermal and mechanical multistate folding of ribonuclease H. *J. Chem. Phys.*, 131:235101, 2009.

[63] W Humphrey, A Dalke, and K Schulten. VMD-Visual Molecular Dynamics. *J. Mol. Graphics*, 14:33–38, 1996.

[64] D Frishman and P Argos. Knowledge-based protein secondary structure assignment. *Proteins*, 23:566–579, 1995.

[65] Y Sugita and Y Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.*, 314:141–151, 1999.

[66] S Nos. A unified formulation of the constant temperature molecular dynamics methods. *Journal of Chemical Physics*, 81:511–519, 1984.

[67] WG Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A*, 31:1695+, 1985.

[68] S Nos. Constant-temperature molecular dynamics. *Journal of Physics: Condensed Matter*, pages SA115–SA119, 1990.

[69] SP Gore, DF Burke, and TL Blundell. PROVAT: a tool for Voronoi tessellation analysis of protein structures and complexes. *Bioinformatics*, 21:3316–3317, 2005.

[70] T Ku, P Lu, C Chan, T Wang, S Lai, P Lyu, and N Hsiao. Predicting melting temperature directly from protein sequences.

[71] M Seeber, M Cecchini, F Rao, G Settanni, and A Caflisch. Wordom: a program for efficient analysis of molecular dynamics simulations. *Bioinformatics*, 23:2625–2627, 2007.

[72] Y Sugita, A Kitao, and Y Okamoto. Multidimensional replica-exchange method for free-energy calculations, 2000.

A DETAIL OF PARALLEL CODING

In the following script, swap requests are proposed from certain processors (odd or even ones) to their upper neighbor processors. For example, if there are 8 processors (p1 to p8), swap requests are initially proposed from odd processor, p1, p3, p5, and p7, to their upper neighbors p2, p4, p6, and p8. And then, in a certain number of steps, swap requests are proposed from even processors, p2, p4 and p6, to p3, p5, and p7. Since processor p8 does not have an upper neighbor processor, it does not propose such a swap request.

```
 if(mpi.my_rank!=mpi.p-1 && flag ==1)//flag=1 mean the box is sending a request to rank+1
{...
 MPI_Send(&swap_request,sizeof(swap_request),MPI_CHAR,mpi.my_rank+1,1,MPI_COMM_WORLD);
...}
```

At the same time, some processors must receive those requests information from their lower neighbors, and then determine if they will accept the deal or not, based on the Metropolis criterion. The following scripts show how that works.

```
 else if(mpi.my_rank!=0 && flag==0)//flag 0 means the box is receiving a request from rank-1
  {...
MPI_Recv(&swap_request,sizeof(swap_request),MPI_CHAR,mpi.my_rank-1,1,MPI_COMM_WORLD,&status);

double accep_crit = exp(delta_beta*delta_energy);

    if(accep_crit > ran2()){//Swap is accepted
      swap_accept.accept = 1;
     ...}
...}
```

In the script above, 'MPI-Send' and 'MPI-Recv' are MPI functions for sending and receiving information from each other, based on their ranks, which are their processor numbers). They are effective in the range of 24 processors, which are defined in the 'MPI-COMM-WORLD'.

After the scprits above finish their job, the acceptance decision is made and sent back to the requesting processors. If requests are accepted, processors work on changing their replicas, while if not, they will keep running their own replicas. The commands used are just the same ones as shown above for swapping requests. After each swap, requesting processors are changed between odd and even ranks by changing flags. Therefore, each processor can communicate to both upper and lower neighbor processors. The whole function script is shown as follows.

```
#ifdef MPI
#include "defines.h"
void nblist         (int);
double ran2         (void);
#ifdef STATUS
void curr_status    (int,int);
#endif
int swap_box_mpi(int flag)
{
  int k = 0;
  MPI_Status status;
  #ifdef STATUS
  curr_status(k,5);
  #endif
  if(mpi.my_rank!=mpi.p-1 && flag ==1)//flag=1 mean the box is sending a request to rank+1
  {
    struct msg_swap_request swap_request;
    swap_request.potens         =       en[k].potens;
    swap_request.kT             =       sim.kT[k];
    struct msg_swap_accept swap_accept;
    MPI_Send(&swap_request,sizeof(swap_request),MPI_CHAR,mpi.my_rank+1,1,MPI_COMM_WORLD);
    MPI_Recv(&swap_accept, sizeof(swap_accept), MPI_CHAR,mpi.my_rank+1,2,MPI_COMM_WORLD,&status);
    if(swap_accept.accept==0){
      return 0;
    }
    else if (swap_accept.accept==1){
        for(int i =0; i< box[k].boxns; i++){
        atom_temp[k][i]         = atom[k][i];         /* Back up coordinates of box k                  */
        atnopbc_temp[k][i]      = atnopbc[k][i];  /* Back up coordinates of box k                      */
        ff_temp[k][i]           = ff[k][i];
        vv_temp[k][i]           = vv[k][i];
      }
      en_temp[k]                        = en[k];
      #ifdef PRESSURE
      pvir_temp[k]                      = pvir[k];
      #endif
      MPI_Recv(atom[k],   box[k].boxns * sizeof(struct atoms),MPI_CHAR,mpi.my_rank+1,3,MPI_COMM_WORLD,&status);
      MPI_Recv(atnopbc[k],box[k].boxns * sizeof(struct atoms),MPI_CHAR,mpi.my_rank+1,4,MPI_COMM_WORLD,&status);
      MPI_Recv(ff[k],     box[k].boxns * sizeof(struct veloc),MPI_CHAR,mpi.my_rank+1,5,MPI_COMM_WORLD,&status);
      MPI_Recv(vv[k],     box[k].boxns * sizeof(struct veloc),MPI_CHAR,mpi.my_rank+2,6,MPI_COMM_WORLD,&status);
      MPI_Recv(&en[k],    sizeof(struct energy),MPI_CHAR,mpi.my_rank+1,11,MPI_COMM_WORLD,&status);
      #ifdef PRESSURE
      MPI_Recv(&pvir[k],    sizeof(struct virial),MPI_CHAR,mpi.my_rank+1,13,MPI_COMM_WORLD,&status);
      #endif
     for(int i=0; i< box[k].boxns; i++){
        vv[k][i].x  /= swap_accept.scale;
        vv[k][i].y  /= swap_accept.scale;
        vv[k][i].z  /= swap_accept.scale;
        uu[k][i]     = vv[k][i];
        }
      MPI_Send(atom_temp[k],   box[k].boxns * sizeof(struct atoms),MPI_CHAR,mpi.my_rank+1, 7,MPI_COMM_WORLD);
      MPI_Send(atnopbc_temp[k],box[k].boxns * sizeof(struct atoms),MPI_CHAR,mpi.my_rank+1, 8,MPI_COMM_WORLD);
      MPI_Send(ff_temp[k],     box[k].boxns * sizeof(struct veloc),MPI_CHAR,mpi.my_rank+1, 9,MPI_COMM_WORLD);
      MPI_Send(vv_temp[k],     box[k].boxns * sizeof(struct veloc),MPI_CHAR,mpi.my_rank+1,10,MPI_COMM_WORLD);
      MPI_Send(&en_temp[k],     sizeof(struct energy),MPI_CHAR,mpi.my_rank+1,12,MPI_COMM_WORLD);
      #ifdef PRESSURE
      MPI_Send(&pvir[k],    sizeof(struct virial),MPI_CHAR,mpi.my_rank+1,14,MPI_COMM_WORLD,&status);
      #endif
```

```
      #ifdef NLIST
      nblist(k);
      #endif
      return 1;
    }// if swap accepted then for the sending box
  }// end of mpi.rank !=p-1
  else if(mpi.my_rank!=0 && flag==0)//flag 0 means the box is receiving a request from rank-1
  {
    struct msg_swap_request swap_request;
    struct msg_swap_accept swap_accept;
    MPI_Recv(&swap_request,sizeof(swap_request),MPI_CHAR,mpi.my_rank-1,1,MPI_COMM_WORLD,&status);
    double e2 = swap_request.potens;
    double delta_energy = e2 - en[k].potens;
    double T2 = swap_request.kT;
    double delta_beta = 1.0/T2-1.0/ sim.kT[k];
    double accep_crit = exp(delta_beta*delta_energy);
    if(accep_crit > ran2()){//Swap is accepted
      swap_accept.accept = 1;
      swap_accept.scale  = sqrt(sim.kT[k]/T2);
      MPI_Send(&swap_accept,sizeof(swap_accept),   MPI_CHAR,mpi.my_rank-1,2,MPI_COMM_WORLD);
      MPI_Send(atom[k],   box[k].boxns * sizeof(struct atoms),MPI_CHAR,mpi.my_rank-1,3,MPI_COMM_WORLD);
      MPI_Send(atnopbc[k],box[k].boxns * sizeof(struct atoms),MPI_CHAR,mpi.my_rank-1,4,MPI_COMM_WORLD);
      MPI_Send(ff[k],     box[k].boxns * sizeof(struct veloc),MPI_CHAR,mpi.my_rank-1,5,MPI_COMM_WORLD);
      MPI_Send(vv[k],     box[k].boxns * sizeof(struct veloc),MPI_CHAR,mpi.my_rank-1,6,MPI_COMM_WORLD);
      MPI_Send(&en[k],     sizeof(struct energy),MPI_CHAR,mpi.my_rank-1,11,MPI_COMM_WORLD);
      #ifdef PRESSURE
      MPI_Send(&pvir[k],    sizeof(struct virial),MPI_CHAR,mpi.my_rank-1,13,MPI_COMM_WORLD,&status);
      #endif
      MPI_Recv(atom[k],   box[k].boxns * sizeof(struct atoms),MPI_CHAR,mpi.my_rank-1, 7,MPI_COMM_WORLD,&status);
      MPI_Recv(atnopbc[k],box[k].boxns * sizeof(struct atoms),MPI_CHAR,mpi.my_rank-1, 8,MPI_COMM_WORLD,&status);
      MPI_Recv(ff[k],     box[k].boxns * sizeof(struct veloc),MPI_CHAR,mpi.my_rank-1, 9,MPI_COMM_WORLD,&status);
      MPI_Recv(vv[k],     box[k].boxns * sizeof(struct veloc),MPI_CHAR,mpi.my_rank-1,10,MPI_COMM_WORLD,&status);
      MPI_Recv(&en[k],     sizeof(struct energy),MPI_CHAR,mpi.my_rank-1,12,MPI_COMM_WORLD,&status);
      #ifdef PRESSURE
      MPI_Recv(&pvir[k],    sizeof(struct virial),MPI_CHAR,mpi.my_rank-1,14,MPI_COMM_WORLD,&status);
      #endif
      for(int i=0; i< box[k].boxns; i++){
        vv[k][i].x  *= swap_accept.scale;
        vv[k][i].y  *= swap_accept.scale;
        vv[k][i].z  *= swap_accept.scale;
        uu[k][i]     = vv[k][i];
      }
      #ifdef NLIST
      nblist(k);
      #endif
      return 1;
    }//end of accept for receiving box
    else { //swap is rejected
      swap_accept.accept=0;
      swap_accept.potens=e2;// e2 belongs to myrank-1
      MPI_Send(&swap_accept,sizeof(swap_accept),   MPI_CHAR,mpi.my_rank-1,2,MPI_COMM_WORLD);
      return 0;
    }// end of reject for receiving box
  }// end of rank!=0 loop
  return 2;
}
#endif
}
```