2013-03-19

Protein-Surface Interactions with Coarse-Grain Simulation Methods

Shuai Wei
Brigham Young University - Provo

Protein-Surface Interactions with Coarse-Grain Simulation Methods

Shuai Wei

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

Doctor of Philosophy

Thomas A. Knotts, Chair
Dean R. Wheeler
W. Vincent Wilding
John D. Hedengren
Randy S. Lewis

Department of Chemical Engineering

Brigham Young University

March 2013

ABSTRACT

Protein-Surface Interactions with Coarse-Grain Simulation Methods

Shuai Wei
Department of Chemical Engineering
Doctor of Philosophy

The interaction of proteins with surfaces is a major process 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. This document outlines several studies done to develop a model which is capable of predicting the stabilization and destabilization of proteins tethered to surfaces. As the start point of the research, part of this research showed that the stability of five mainly-alpha, orthogonal-bundle 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. To test the expandability of the protein stability prediction pattern derived for mainly-alpha, orthogonal-bundle proteins, same analysis is performed for proteins from other structure motifs. Besides the study in these small two-state proteins, a further analysis of surface-induced change of folding mechanism is also studied with a multi-state lysozyme protein 7LZM. The result showed that by tethering a protein on a surface, the melting temperature of a part of the protein changed, which leads to an avoidance of the meta-stable state. Besides the change of folding mechanism, by tethering the lysozyme protein to a certain site, the protein could both keep a stable structure and a good orientation, allowing active sites to be available to other proteins in bulk solution. All the work described above are done with a purely repulsive surface model which was widely used to roughly simulate solid surfaces in protein microarrays. For a next-level understanding of protein-surface interactions, a novel coarse-grain surface model was developed, parameterized, and validated according to experimental results from different groups. A case study of interaction between lysozyme protein 7LZM and three types of surfaces with the novel model has been performed. The results showed that protein stabilities and structures are dependent on the types of surfaces and their different hydrophobicities. This result is consistent with previously published experimental work[1].

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 was very 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 variety of simulation techniques to make the work interesting.

I also would like to thank my wife Wei Zhou and my parents for their constant support and encouragement. With their support, all difficulties turned out to be easy to overcome. As a graduate student major in molecular biology and statistics, Wei helped me understand the complicated biology processes. I appreciate all her efforts and ideas.

Table of Contents

List of Tables

List of Figures

Chapter 1

Protein Folding and Stability on Protein Microarrays

1.1 Introduction

Protein-surface interactions are the key phenomenon involved in many technologies such as diagnostics, bioanalytes, and biomaterials [2, 3, 4, 5, 6]. Protein microarrays, in particular, are a prominent example because of their potential to revolutionize medicine, biocatalysis, and biosensors. To create protein microarrays, specific proteins are deposited at addressable locations on a solid substrate. The deposited proteins perform functions such as binding or reacting with complementary molecules found in solutions placed on the chip. For example, antibody-antigen microarrays are created by depositing different antibody proteins on the surface. A sample containing unknown antigens can then be incubated on the surface to screen for identification using fluorescent labels. Such technologies have been used in the research setting to identify proteins in serum, determine levels of protein expression, screen drug candidates, ascertain functions, and detect protein-ligand interactions [3, 4, 5, 6].

However, despite the promise of protein microarrays, the use of the technology is limited as it is difficult to obtain reproducible, qualitative results [7]. Placing a protein on a surface in a manner that maintains protein native structure and stability is the key to protein microarray performance. This is complicated because surfaces can induce conformational changes in proteins [7, 8, 9, 10, 11, 12, 13]. Since protein structure leads directly to protein function, transformations that do occur prevent the proteins on the surface from producing the desired outcome. Designing around this fact is difficult because no method exists to predict à *priori* how a particular tethered protein will behave.

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. Covalent tethering has emerged as the favored method to

create protein arrays, as conformation changes are more pronounced for adsorbed proteins, but significant challenges still remain. Thus rational design of protein arrays is difficult with current understanding 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 researchers, using both simulation and experiment, have been involved in understanding what affects protein behaviors on a chip, there is still little known about the underlying biophysics. The lack of capabilities in predicting protein stabilities and structures on surfaces hinder researchers from optimizing protein arrays for improved performance. *Therefore, the overall goal of this work is to develop a fundamental understanding of the mechanism by which surfaces affect protein stabilities, so that rational design of protein arrays can occur.*

This document is organized as follows. The remainder of Chapter 1 discusses details about protein microarrays and previous experimental, simulation, and theoretical studies of protein-surface interactions. Chapter 2 focuses on the simulation methods used in the work. Simulation was used because experimental techniques do not exist which can probe protein structure on surfaces with atomic level resolution. Chapters 3-6 contain the bulk of the results for the work. Chapter 3 describes how stability can be correlated to the type of loop region where the tether is placed in the protein. The results of this work show, for the first time, that protein stability on surfaces can be correlated to tertiary structure for a specific family of proteins called alpha-helical, orthogonal-bundles. A paper *Predicting Stability of Alpha-helical, Orthogonal-bundle Proteins on Surfaces* related to Chapter 3 has been published on the Journal of Chemical Physics in 2010. Chapter 4 explains how well the protein stability prediction pattern derived for mainly-alpha, orthogonal-bundle proteins are suitable to other tertiary motifs. Chapter 5 extends the concepts outlined in Chapters 3 and 4 to explain how large proteins that fold through a multi-state mechanism are affected by the surface and how correct tether placement can both stabilize the protein on the surface and keep the active site available to the bulk solution. A paper *Effects of Tethering a Multistate Folding Protein to a Surface* related to Chapter 5 has been published on the Journal of Chemical Physics in 2011. Chapter 6 describes how the study of protein-surface interactions is moved to the next level by developing a new model that more realistically captures experimentally-observed

adsorption energies than models used previously. This document ends with Chapter 7 where the results are summarized and ideas about future work are offered. Taken as a whole, the work reported in this dissertation presents a rigorous thermodynamic understanding of the origins of protein stabilization/destabilization of surfaces and offers hope that rational prediction of protein behavior on surfaces is possible.

1.2 Background

1.2.1 Protein Structure

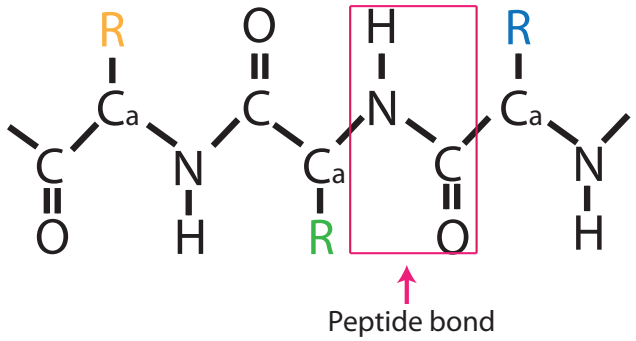


Figure 1.1: Protein primary structure.

Proteins are large biological molecules consisting of one or more chains of amino acids as shown in Figure 1.1. Each amino acid has a carboxyl and an amino group, connected to a carbon atom (C_α). A different function group side chain (R) is also connected to the C_α. There are 20 different side chains in standard amino acids found in nature. Individual amino acids are connected to each other by forming peptide bonds between the carboxyl and amino groups of adjacent amino acid residues. Because it is a linear chain, one end of the protein molecule ends in a carboxyl group (C-terminus), and the other ends with an amine group (N-terminus).

The interaction of the amines, carbonyls, and side chains with each other and the solvent causes proteins to self aggregate or "fold" into complex three-dimensional structure.

Figure 1.2: Protein secondary structures: α-helix and β-sheets.

Different terms are given to these structures. The primary structure is simply the chain of amino acids as described above. The first level of folding produces what are called secondary structures. Figure 1.2 depicts the two main types of secondary structures: α-helix and β-sheets. Unstructured portions of the molecule, pictured in green in Figure 1.2, are termed coils and loops and connect different secondary structures together. A given protein molecules will be composed of several secondary structure elements linked by loops. These will also aggregate together or fold into more complex shapes called tertiary structure. This final structure is important and it imparts to the protein its specific function. If the protein structure breaks (unfolds) then function is lost.

1.2.2 Protein Microarrays and Their Limitations.

A protein microarray is a high-throughput diagnostic device that can perform thousands of biological assays in parallel. This technology has been identified as a powerful tool to facilitate next generation proteomics and patient-tailored medicine [14, 15]. An array 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 a label-based detection method [3, 16, 4, 5, 6] – such as fluorescent dyes, radioisotopes, epitope tags – or label-free techniques [17, 18, 19, 20] – such as surface plasmon resonance, carbon nanotubes and nanowires, and microcantilevers – researchers can identify proteins or determine protein concentrations in samples.

Several protein array platforms have been developed. The first high-density antibody microarrays were studied by Haab *et al.*, and were used to test whether a linear relationship could be detected between an antibody and antigen pair in an array format [21]. Soen *et al.* fabricated an analytical microarray using peptide-MHC complexes to detect and characterize antigen-specific T-cell populations [22]. Hsu *et al.* have built up a lectin chip with 21 lectins for use in profiling the surface lipopolysaccharides in bacterial cells [23]. The lectins were able to capture the bacterial cells onto the chip when labeled *E.coli* cells were incubated on the chip.

There are two basic types of protein microarrays: [24, 25] functional protein microarrays and protein-detecting microarrays. Functional protein microarrays are employed to qualitatively distinguish proteins (Figure 1.3), and the protein-detecting microarrays serve as quantitative analytical tool of monitoring protein levels in a given biological sample (Figure 1.4). Both perform massively parallel assays. In the first application, native ligands are arrayed in defined spots, and fluorescently labeled proteins in solution are screened as they are incubated on the chip. Those spots where complementary proteins bind will fluoresce thus allowing identification. In the second format, one specific kind of ligand is arrayed on the chip so that levels of corresponding proteins in solution may be detected and calculated according to the binding concentration.

Although 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.

For example, results from antibody arrays are not always conclusive due to different arraying technologies [26]. Some antibodies have been shown to be active in standard assaying techniques such as ELISA while the activity can not be measured on surfaces [27]. Also,

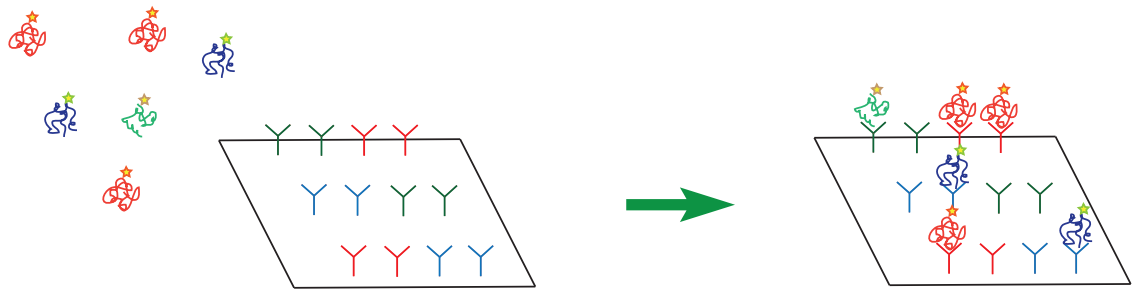


Figure 1.3: Protein microarray to identify proteins in a sample.

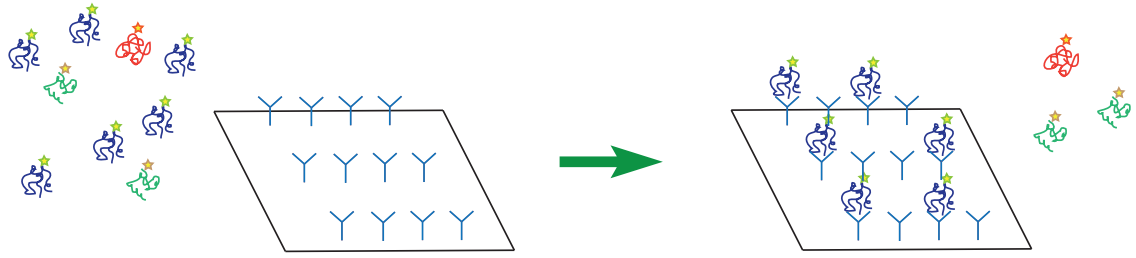


Figure 1.4: Protein microarray for monitoring protein levels in a sample.

signal intensities can vary as much as 43% on the same chip [28]. Moreover, since antibody arrays are the most advanced of the technologies [29], arraying other proteins, such as cytokines [26, 30, 31, 32], is even more challenging. In short, despite the promise of protein arrays as a diagnostic tool, current technology cannot produce arrays that perform at levels commensurate for use in clinical settings [26, 6, 8, 29, 33].

There are two main factors leading to the poor performance of microarrays. One is the cost of expressing and purifying proteins; the other is the difficulty in maintaining protein stabilities and orientation on chip surfaces. To address the first problem, He *et al.* developed and simplified an *in situ* synthesis method for protein arrays that improves proteins expression and convenience of handling [34, 35, 36]. By using this method, protein microarrays can be produced at the precise time needed, which improves protein stabilities by eliminating the storage time of the chip. However, the second factor, described in detail in the next section, has not received much attention because of the lack of experimental

techniques which can probe protein structure on the surface with atomic and molecular-level resolution.

1.2.3 Experimental Understanding of Protein-surface Interactions

Two general conditions for protein-surface interactions are needed for functional microarrays. The first is to create a surface that allows molecules in the bulk to interact with complementary molecules on the surface but prevent non-specific adsorption. The second is to place the diagnostic proteins on the surface at the precise location desired while preserving the structure and function of the molecule. Researchers have studied the behavior of polypeptides at interfaces for decades with these goals in mind. However, as will be described shortly, this research has not produced better microarrays because no experimental techniques exist which can measure stability while also providing the molecular-level resolution needed to correlate structure and stability.

There are many methods for placing proteins on surfaces [6, 36]. The simplest is the adsorption of the protein directly to the surface. This technique has been used in the standard enzyme-linked immunosorbent assay (ELISA) and Westen blot for many years. It is generally mediated by electrostatic charges [21] or hydrophobic interactions [37]. Despite its simplicity, the main drawback of this method is the high probability of denaturation of the diagnostic proteins and non-specific protein adsorption [6]. Covalent binding of proteins to substrate surfaces is a more efficient and robust approach [38, 39, 34, 25].The surfaces are designed with reactive groups, such as epoxides, aldehydes, succinimidyl esters, or isothiocyanates, which react with nucleophilic groups (e.g., amino, thiol or hydroxyl groups) of amino acid residues. However, when the immobilization is done via random attachment, proteins tend to denature. 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 [15]. It produces better biological activity of immobilized protein than protein microarrays that employs non-site-specific immobilization techniques by providing controlled protein orientation on solid surfaces. However, it is still not clear how to make the selection of the tether site to keep the protein stable.

Perhaps the most popular technology to control protein-surface interactions are self-assembled-monolayers (SAMs) of alkanethiols on gold surfaces and alkylchlorosilanes or alkylalkoxysilanes on glass surfaces. Both are effective at preventing non-specific adsorption as well as directing desired protein placement on the surface [40, 41]. However, their inherent instability prevents their effective use outside of the research setting particularly as medical diagnostics which require an extended shelf life [42]. Polymer coated surfaces, usually polyethylene glycol (PEG)-based, are used extensively to prevent fouling of surfaces and to control protein placement on the surface through appropriate functionalization of the polymer. It has been shown by several research works [43, 44, 45] that optimized polymer coating improves the performance of protein microarrays. 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 are active in solution but not on surfaces while others are not affected by the substrate despite both two groups of antibodies having similar structure [27].

One difficulty of understanding protein-surface interactions is that experimental methods are limited. 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 fourier transform infrared spectroscopy (FTIR) can be used to provide a gross estimate of protein structure but cannot provide mechanistic understanding or atomic-level structural resolutions [46]. Tsapikouni and Missirlis [47] summarized the experimental techniques to study protein surface interactions from micro-to-nano scale. In their work, the wide-spread use of atomic force microscopy (AFM) is introduced as a turning point in studying protein structure and measuring the interaction force, but even this advanced technique does not provide the needed resolution.

Most experimental studies have focused on developing a thermodynamic understanding of protein-surface interactions is essential for understanding the protein recognition phenomenon in biology and describing possible protein adsorption mechanisms [48]. There are two main methods for thermodynamic analysis of protein binding: chromatography with van't Hoff analysis [49, 50, 51, 52] and microcalorimetry. Researchers have used reverse

phase chromatography (RPC) and hydrophobic interaction chromatography (HIC) to analyze the hydrophobic characteristics of biomolecules [53, 54]. With the van't Hoff method, the enthalpy of protein adsorption on the surface is calculated from the free energy change at different temperatures. The temperature change has to be in a narrow range to assume that the thermodynamic parameters and protein folding mechanisms are invariable. The other method, microcalorimetry, emerged in the last two decades as an alternate for directly determining adsorption enthalpy [48, 1]. There are some observable differences in enthalpy results between these two methods and the reason is well explained by Chen *et al.* [48] Microcalorimetry measures all energy changes happen during protein-surface interactions– including dilution heat, binding heat, desolvation energy, and protein melting– while the van't Hoff method does not include the contribution of protein conformation change.

Latour's group used atomic force microscopy (AFM) and surface plasmon resonance (SPR) to measure protein adsorption free energy [55]. Since only surfaces that can be formed as thin layers on metallic biosensor substrates on nanometer scales can be used for SPR, they evaluated the transferability of the energy measured on SPR to the AFM technique that can be used with a variety of surfaces. Their results showed a linear correlation between data from AFM and that from SPR for a similar set of protein-surface systems.

In summary, a variety of experimental techniques exist which can measure thermodynamic properties of protein-surface interactions but none of these provide details on the structure of the protein or the surface at the resolution needed to correlate structure and stability. What is needed is a way to measure both the thermodynamic stability of the system and the structural details of the molecule at the same time. Such would allow predictive models to be developed that could be used to rationally design protein microarrays according to the specific protein being tethered to the surface. Molecular simulation can provide the needed capability, and was used to accomplish the goals of this work. The following section summarizes previous simulation research into protein surface interactions.

1.2.4 Theoretical and Simulation Understanding

Since experimental methods cannot provide atomic-level information, several groups have done simulations and theoretical work to understand protein-surface interactions both

with atomistic and coarse-grain methods. In an all-atom simulation, each atom in the system is explicitely represented by a particle and particles interact with each other through empirically-parameterized mathematical descriptions called force fields. Coarse-grained simulation is a method that simplifies molecular structure by grouping atoms into single sites and using an implicit solvent, and is thus less accurate than atomic force fields. In general, all-atom simulations can be highly accurate because aqueous solvation effects are explicitly addressed which enables solvent molecules to participate in the molecular system. [56] However, in all-atom methods, simulation size and time scales are limited due to the high computational resource demands. This can be overcome by using a coarse-grain model. Even though there is a cost in accuracy, well-parameterized coarse-grained model, such as that of Karanicolas and Brooks [57, 58, 59, 60, 61], can capture the most important effects with regard to proteins-surfaces interactions and have been shown to reproduce experimental results.

1.2.4.1 Progress Using Atomistic Simulations

Some groups have implemented atomistic simulations of proteins on surfaces. For example, Latour and coworkers have investigated both model peptides [62] and biologically relevant proteins, such as fibrinogen [63], using SAMs of many different functionalizations using an all-atom representation. In each study, they report both agreement and conflict between simulation and experiment. Jiang *et al.* used atomistic models [64] to study the structure of phosphorylcholine self-assembled monolayers (PC-SAMs) for protein-surface interactions. Jiang [65] also showed conflict with experimental results in energies of adsorption and monolayer structure. Kubiak *et al.* [66] atomistically simulated Egg-white Lysozyme in three different systems, and found that lysozyme has a preferred orientation for absorption to surfaces. Wei *et al.* [67] performed a series of atomistic MD simulations with explicit water to study lysozyme-polyethylene interactions with different initial lysozyme orientations. However, they realized that the time scale in this simulation was too short to adequately sample the configurations for proper understanding of the protein-surface interaction.

In recent years, Kokh *et al.* [68] developed an atomistic force field for modeling protein-metal surface interactions. They validated their work by comparing the adsorption free

energies and potential of mean force (PMFs) of capped amino acids to previously published MD simulations by Hoefling *et al.* [69] Latour and coworkers performed very careful atomistic simulations considering pressure for a system with constrained atoms and methods to assess electrostatic effects [70, 71]. In another review paper from the same group, it was emphasized that a novel coarse-grain method is needed to bridge time and length scales to properly study protein-surface interaction [72].

The Latour group, whose research was just mentioned, is among the premier groups studying protein-surface interactions. They use a combined experimental and all-atom simulation approach and in a recent review pointed out three main issues that must be considered in a molecular simulation of protein-surface interactions: [56] force field parameterization, solvation effects, and sampling ergodicity which suggests that all accessible microstates are equiprobable over a long period of simulation time. Latour sought to capture these effects by developing an experimentally-validated, all-atom force field for protein-surface interactions based upon novel experimental techniques measuring the free energy of adsorption for 11 naturally occurring amino acids [73].

A recent paper [74] illustrates the current limitations of all-atom models to reproduce experimentally-observed behavior. Simulations of a variety of proteins on both hydrophobic and hydrophilic surfaces were done using an all-atom model with three different parameterizations (termed force fields): CHARMM-22, AMBER94, and OPLS-AA. The study concluded that CHARMM-22 generally produced more reliable conformational behavior of proteins on surfaces than other force fields. However, it is mentioned that an improved force field is needed for fully understanding protein adsorption behavior to material surfaces. The work is a significant advancement in studying adsorption, but due to computational limitations, the group later realized that a coarse-grain model is needed. The all atom model could probe global orientation of a protein with respect to the surface and local structural changes, but was too computationally slow to probe the larger structural changes needed to ascertain stability.

1.2.4.2 Progress Using Coarse-grain Simulations

As mentioned previously, predicting protein stabilities on surfaces is key to designing better arrays. To calculate protein stability, the protein must fold and unfold many times during the simulation. It is hard to achieve such sampling using atomistic models, so researchers have developed different coarse-grain approaches to solve the problem. The general idea of coarse-grain model is to reduce the degrees of freedom modeled in the system. This can be done a variety of ways, such as fixing configurations in space and removing solvent molecules [75, 76, 77, 78, 79, 80, 81]. For example, Sun *et al.* [75] employed an implicit solvent 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.* [79] reported a study about lysozyme adsorption to charged surfaces by Monte Carlo simulation. The lysozyme in their simulation was modeled as a large hard sphere with charges placed on the surface of the sphere at locations corresponding to charged amino acids in the real protein.

One important way that coarse-grain simulations have improved understanding of protein-surface interactions is by providing a thermodynamic perspective. Several years ago, Dill *et al.* predicted that proteins are always stabilized when tethered to non-attractive, purely-repulsive surfaces (those found in protein arrays) [76, 82]. The reason is summarized in Figure 3.9. 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 the surface 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 (ΔG_f) for the tethered protein relative to the bulk protein resulting in stabilization on the surface. In 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 (more stable) value of ΔG_f.

Coarse-grain simulations have helped show that Dill's theory does not take into account several important phenomena and does not describe reality. In one example, work by

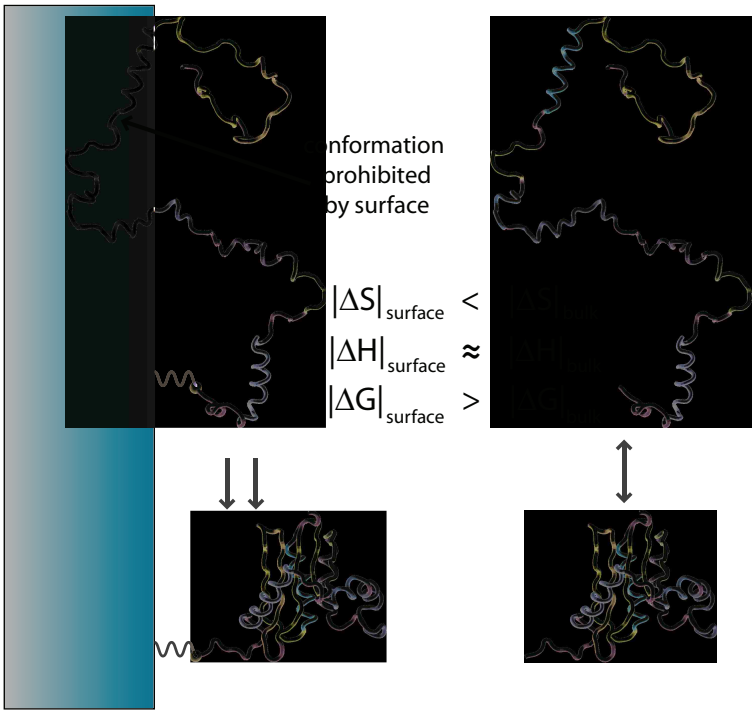


Figure 1.5: Theory behind the stabilizing influence of the surface on tethered proteins.

Friedel *et al.* [83, 84] used 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, which was contrary to the "always stabilized" idea proposed in the theory. Results also showed that if the tethering is done to a site on the interior of the molecule the protein is always destabilized. More recent work by Zhuang *et al.* also shows variation of protein stability when tethering the src-SH3 protein on surfaces with different sites [85]. Similar results were seen in prior work by Knotts *et al.* done on the all-alpha, three-helix-bundle protein from *Staphylococcus aureus* [86]. In this study both the mechanical and thermal stabilities of the peptide were reduced when the protein was tethered to the surface. In recent years, Freed and Cramer [87] did a study of protein-surface interaction maps. A good estimates of electrostatic interaction strengths were generated by their work, which were consistent with the primary mechanism of interaction in ion-exchange chromatography.

Other work by Knotts *et al.* used four proteins: protein A, 434 repressor, SH3, and Protein G, which have different secondary and tertiary structure motifs [88]. All four proteins were simulated both in the bulk and tethered to an inert (non-attractive) surface at the $N-$ and $C-$ termini of the protein molecule, and the thermal stability of the proteins was probed using configurational-temperature-density-of-states simulations. The work showed that proteins could be both stabilized and destabilized on surfaces. It was also shown that only all-alpha proteins displayed the surface-induced destabilization, while the proteins with beta-content displayed only stabilization. This is consistent with results from Zhuang *et al.* [85] Another important result from the work by Knotts *et al.* is that the stability cannot be correlated to secondary structure as protein A and 434 repressor, which are both all-α helical peptides, displayed different behaviors [88]. Furthermore, as the results for protein A suggested, the same protein can behave differently on surfaces depending on different tethering sites.

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 a predetermined manner have been developed, and there are some successful instances of binding proteins with conserved stabilities, most tethering does not result in functional proteins on the surface. This is in large part because there are no theories guiding how to tether a protein to a surface and maintain its function.

Simulation is an adept means of exploring this topic if used properly. Atomistic simulations have shown to be ineffective in rendering protein folding information due to their large computational requirement, but some coarse-grain models have proven to provide results that are consistent with experiments. Results using such methods have shown that proteins could be stabilized or destabilized on surfaces, and that the folding mechanism could be changed if the tethering site is in a part of the molecule involved in transition between the folded and unfolded structure [85]. Moreover, protein stabilities cannot be correlated to 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.

In short, despite the success in understanding the behavior of a select few proteins on a surface, no method currently exists to predict how to tether a protein to a surface to maintain stability. This study seeks to address this in two ways. The first thrust is to develop a prediction method that can identify which amino acid in the protein to use when tethering to the surface based upon the structure of the protein. The second thrust is to develop a new model for protein-surface interactions that takes into account the fact that many surfaces are attractive to proteins. This second effort is needed to move the study of protein-surface interactions to the next level.

Chapter 2

Efficiently Simulating Protein Folding

In this chapter, general simulation tools and thermodynamic calculation methods are introduced. After that, the replica exchange simulation method is discussed in the context of rugged energy landscapes and how these relate to protein folding. This chapter also describes how a new parallelized code was developed to achieve the simulation times needed to study protein stability. Finally, the umbrella sampling method, used for calculating protein adsorption free energies is described as it was used to develop the new model for protein-surface interactions that was mentioned in the previous chapter.

2.1 General Approach and Thermodynamic Quantities Calculation

2.1.1 Monte Carlo and Molecular Dynamics Simulation Methods

Monte Carlo (MC) is a simulation technique based on repeated stochastic sampling. Statistical mechanics teaches that the probability of a particular system configuration is dependent on system potential energy and temperature. The statistical ensemble representing a probability distribution of microscopic states of the system that all have the same number of particles (N), volume (V), and temperature (T) is called the canonical (NVT) ensemble. Monte Carlo simulations proceed by randomly generating configurations of the system, evaluating the energy of the new configuration, and accepting the new state as the next in the ensemble according to it probability. The specific algorithm used is that of Metropolis [89]. This algorithm moves the system from high to low energy. If the sampling is adequate, thermodynamic properties of the simulation system can be determined by calculating ensemble averages.

Each random sampling step of MC simulation needs to follow the Metropolis criterion, which is based on the ratio of probabilities of the new configuration and the old one. It is

ideal that the whole energy landscape could be fully accessed with sufficient sampling steps and appropriate sampling acceptance ratio. If that is the case, thermodynamic properties of the simulation system would be derived by calculating ensemble averages.

The molecular dynamics (MD) method is a simulation technique based on force driven deterministic moves. Essentially, the forces between the atoms in the system are calculated and used in Newton's equations of motion to propagate the system through space and time. As with Monte Carlo, the end result is that the system will move from a high-energy state to a low-energy state. Macroscopic thermodynamic properties of the system can be calculated from the MD simulation as time averages.

One important thing that has to happen in both MC and MD is that the system needs to sample all the relevant conformations of the system. In protein simulations, this means that the folded and unfolded states of the proteins, and all the states in between, need to occur multiple times in the simulations. This is difficult to do in protein simulations because the energy landscape of the system is rugged. For a complex system such as protein folding on solid surfaces, the potential energy surface contains numerous local minima separated by relatively high energy barriers. As shown in the Figure 2.1, with commonly used sampling protocols, such as MC and MD, the system could be trapped in some of the local minima, and not sample all the states between the folded and unfolded molecule. Should this occur, the thermodynamic quantities that are calculated from the simulations would not be correct. To prevent the system from becoming trapped in low-energy minima, two advanced sampling methods are use in this research: replica exchange and umbrella sampling.

2.2 Replica Exchange

Replica exchange (RE), also known as parallel tempering, is a method developed by Sugita and Okamoto [90] to overcome energy barriers during simulations. It provides an efficient sampling method to solve the problem using a series of replicas of the system of interest. Each replica is simulated in the canonical ensemble at a different temperature [91, 92]. To accomplish barrier crossings, replicas at different temperatures exchange complete configurations (a process called "swapping"). Swaps are accepted using a Metropolis criterion that ensures that at any given temperature a canonical distribution is realized. Swaps

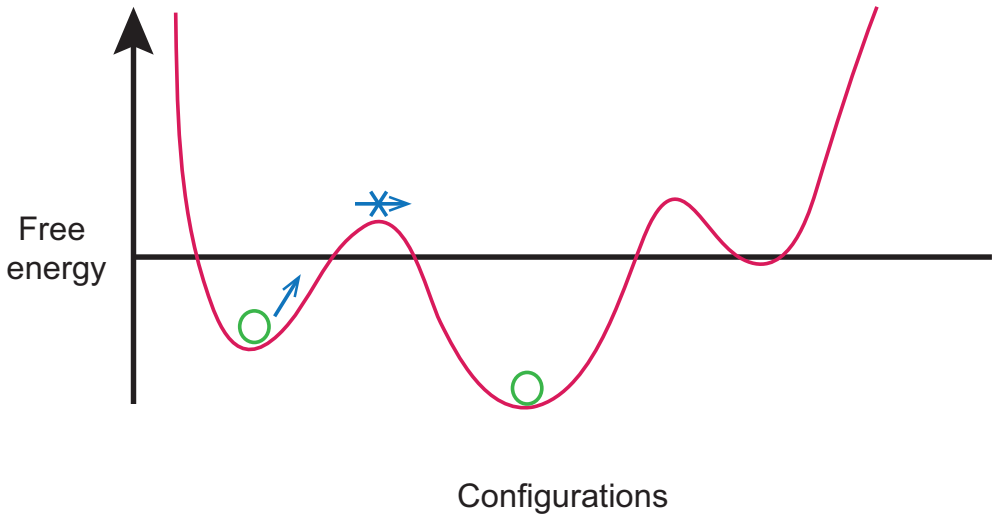


Figure 2.1: Typical energy landscape of a protein.

between each two adjacent replicates X and X' at temperatures T_m and T_n with potential energy of U_i and U_j are accepted with probability:

$$P_{acc}(swap) = \min\{1, exp(-\Delta\beta\Delta U)\} \tag{2.1}$$

where $\beta = \frac{1}{k_B T}$, k_B is Boltzmann's constant, and U is the potential energy of the system.

The equation 2.1 is derived from the basic Metropolis rule that

$$P_{X->X'} = \frac{P_{X'}}{P_X} \tag{2.2}$$

where

$$P_{X'} = e^{-\beta_m U_i} e^{-\beta_n U_j} \tag{2.3}$$

and

$$P_X = e^{-\beta_m U_j} e^{-\beta_n U_i} \tag{2.4}$$

Figure 2.2 shows how replica exchange works on a single processor. Replica exchange works because the systems at higher temperature will not become trapped in local energy minima so that swapping down to lower temperatures will allow the system to traverse energy

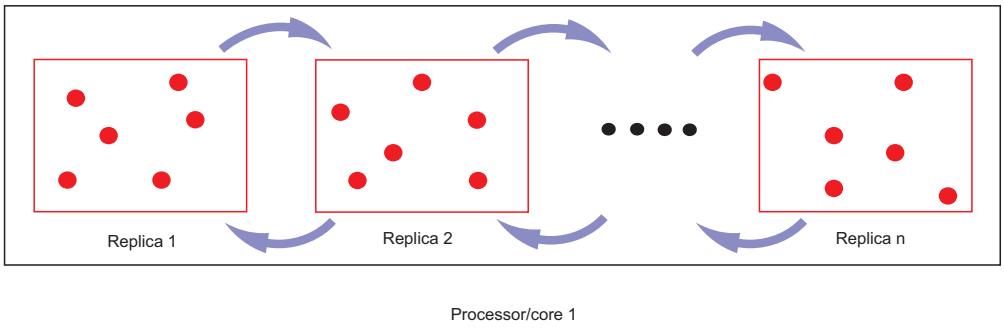


Figure 2.2: Replica exchange simulation.

barriers. The replica exchange method provides a powerful tool to ensure proper sampling of protein configurations in the simulation and therefore leads to an easy way to study protein-surface interactions. However this comes at a cost as significantly longer computation time is required comparing to the traditional MC or MD methods. In general, 24 replicas at different temperatures are needed to generate a complete folding curve for the proteins used in this study. The number of replicas 24 is chosen with regard to the typical temperature range and the supercomputer architecture. Furthermore, for statistical consideration, many simulations need to be run to prove significance. In short, a very large amount of simulation time is needed to accomplish the aims of the research.

It was realized early on that the computational time required to simulate protein folding in simulation needed to be reduced if the aims of the research were to be accomplished. Parallel computing provided the answer. With parallel programming, different parts of a simulation work simultaneously on multiple processors instead of working in a sequential way on the same computing unit.

2.2.1 Parallel Programing with MPI

Fortunately, the RE simulation method is easy to parallelize. This is done by running each simulation replicate on a different computer. Using parallel programming means that the need for computational time can be changed into a need for 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 2.3. The only challenge is found in implementing the swaps described earlier. Recall that at certain steps in the replica exchange method, swaps are proposed between replicas. In the case of parallel coding, the replicas are found on different processors. Communication between the computers is thus required for the swapping.

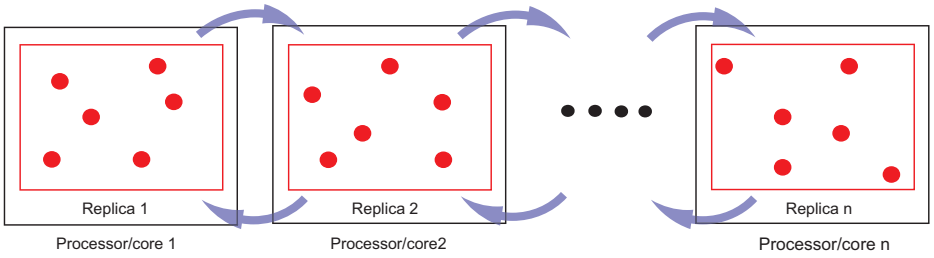


Figure 2.3: Replica exchange with MPI.

Message Passing Interface (MPI), [93, 94] is a specification for an application programming interface (API) that allows many computers to communicate with one another. This feature is just what is needed in parallel simulation of RE. With proper design in MPI, replicas can be simulated on each processor with MD and exchange their coordinates at certain iterations as expected. The time savings are large in that a 24-box simulation will take 1/24th of the time it would on a single processor.

The benefits of parallization of the replica exchange simulation method were so large, that the first part of this research concerned implementing such an approach. It took about two and a half months to write the code to perform replica exchange simulations in parallel and test the program on the BYU supercomputer. The scripts needed to analyze the data were also written at this time. Code for performing RE simulations on a single processor were already available in our group, including the basic MD simulation algorithm and the swap algorithm; therefore, work consisted of using MPI to assign each replica to an individual

computer and communicate between the computers when a swap move was proposed. The general algorithm is as follows:

First, replicas are assigned to different processors by using submission files in multi-processor format. Each 2,000 iterations, every processor then calls a function that controls the swap. This includes sending the energy and temperature of one box to the other and evaluating if the swap is accepted according to Equation 2.1. If the swap is accepted, the coordinates between the two boxes are then swapped. If the move is not accepted, each simulation just keeps its own replica and continues to run MD simulation until the next swapping call. The details of the MPI programming for replica exchange are found in Appendix A.

2.2.2 Performance of the MPI Scripts

To validate the MPI implementation of the replica exchange method, the protein 3WRP was simulated using the single-processor code and the parallel code. The melting temperature of the protein was calculated from the results of each case. A value of 303.57 ± 0.78 K was obtained with MPI and 304.62 ± 1.01 K without MPI. There is no statistically significant difference between the two values. Tests with other proteins showed the same consistency.

After validating the accuracy of the parallel replica exchange code, the next step was to determine the time savings offered by the approach. The protein 3WRP was used again in these tests. Without MPI code, the simulation 40 ns of 3WRP on a single processor took 10.98 days. The MPI code accomplished the same feat in just 7.55 hours. As before, simulations with other proteins showed similar time savings using MPI code.

To summarize, using MPI speeds up simulations by a very large degree while still providing accurate results. This code played a large part in the ability of this research to accomplish the goal of understanding protein-surface interactions. Simulations of longer time and larger protein size could be performed.

2.3 Umbrella Sampling

Umbrella sampling is another method to overcome high energy barriers found in a simulation, and is particularly useful for a protein adsorbing onto a surface. In umbrella sampling, an artificial driving force is added to the simulation to faciliate escape from local minima and more fully sample the whole phase space. By using umbrella sampling, a biased-probability distribution $P_{ij,biased}$ is sampled during simulation. This artificially applied force field is then removed to obtain unbiased thermodynamic properties. The main property obtained from an umbrella sampling simulation is the potential of mean force or free energy for each state of the system. This information can be used to calculate changes in free energy as the system moves from one state to another.

As will be shown in Chapter 5, a harmonic potential is the umbrella force that is added to the simulation to study protein adsorption. This harmonic potential holds the protein at different distances away from the surface to sample all of the states involved as a protein adsorbs to the surface. Distances sampled ranged from 1 to 100 Å. At 100 Å, the interaction between the protein and the surface has decayed to zero. After the simulation, the data are analyzed to remove the bias and calculate the free energy of the system as a function of the distance that the protein is from the surface. Using these data, the free energy of adsorption (ΔG_{ads}) can be obtained.

Chapter 3

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

As mentioned in Chapter 1, the goal of this research is to develop a fundamental understanding of the mechanism by which surfaces affect protein stabilities, so that rational design of protein arrays can occur. This chapter describes efforts to accomplish this goal by correlating the structure of a protein to the stability of the protein when tethered to surfaces. The focus is on a class of proteins called alpha-helical, orthogonal-bundles.

3.1 Background and Hypothesis

Previous studies [83, 84, 86] demonstrated that protein stability is related to tether site in two ways. First, tethering must be done to a residue found on the surface (rather than the interior) of the protein. Second, tethering must not interfere with the folding pathway. The hypothesis stated below is formulated to comply with these two requirements. In order to cast the results in a usable light, a second goal of the research is to correlate the stability patterns to easily-identifiable structural motifs. Prior work showed that stability is not correlated to secondary structure, so this work examines the suitability of tertiary structure for prediction. Formally, the hypothesis tested is that proteins with the same tertiary structure, when tethered only in loop regions, will be stabilized on surfaces. Loop regions are chosen as they are usually found on the outside of the protein and are less likely to interfere with intermediate states found in the folding pathway. As many tertiary structural motifs exist, this study focuses on only one type alpha-helical, orthogonal bundle. As will be described later, this is a classification of proteins with a large number of members. Specifically the hypothesis of this work is alpha-helical peptides with orthogonal-bundle

tertiary structure will be stabilized when tethered to the surface in the loop regions joining adjacent helices.

3.2 Methods

3.2.1 Proteins

Five different proteins were used to test the hypothesis. These were identified with the CATH classification method [95, 96]. CATH (**C**lass, **A**rchitecture, **T**opology, **H**omologous superfamily) is a method of categorizing proteins. The first two levels do so according to structure. *Class* describes the overall secondary structure and *Architecture* the tertiary structure (groupings of structural elements). 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 these proteins ranges from 64 to 163 residues.

Mainly-alpha, orthogonal-bundle proteins provide a convenient starting place to investigate the behavior of families of proteins on surfaces. It may appear that studying only orthogonal-bundle proteins limits the scope of subsequent conclusions; however, such specificity is needed to find useful patterns in the results as attempts to create correlations using broader descriptors have proven difficult [88]. Moreover, this architecture contains over 71% of all the proteins in the mainly-alpha class in the CATH universe.

Mainly-alpha, orthogonal bundle proteins are composed of alpha-helices connected by loop regions. The helices lie at approximately 90° with respect to each other. By comparison, *mainly-alpha, up-down bundles* (a family of proteins with the same CATH class but different architecture) are composed of alpha-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 [84, 85].

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 [58, 59, 60, 61].

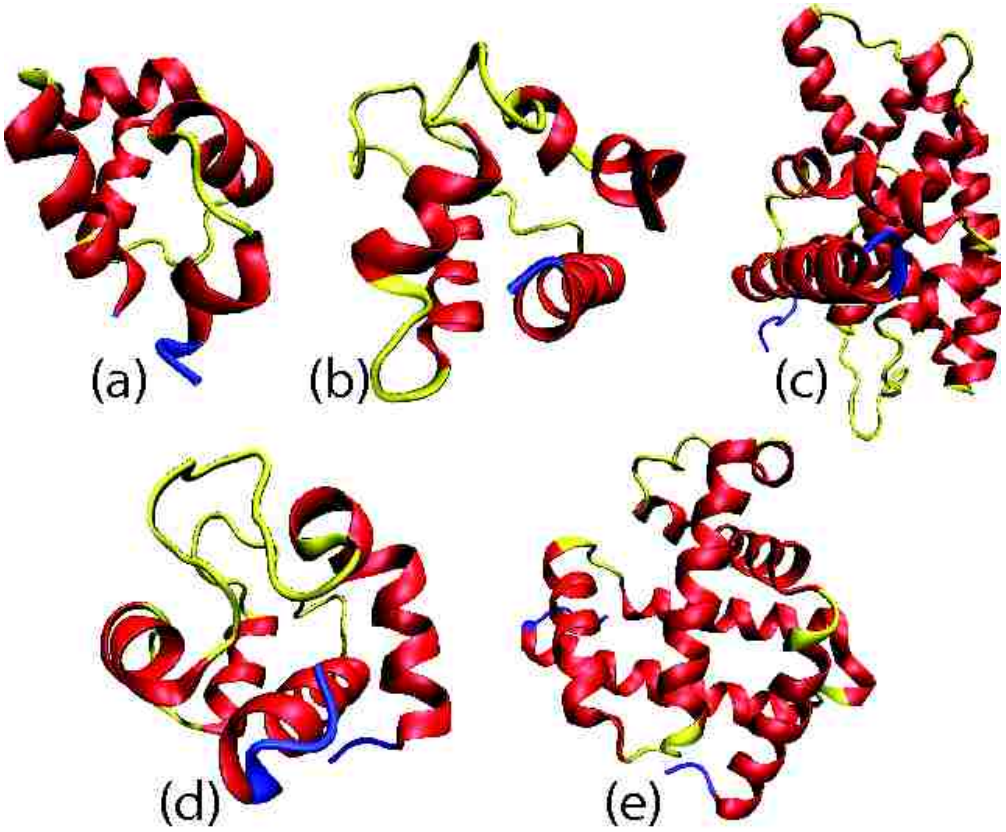


Figure 3.1: Schematic representation of the five alpha-helical, orthogonal-bundle proteins: (a) 1R69, (b) 1A56, (c) 1AD6, (d) 1A2S, (e) 5MBN.

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 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 [58, 59, 60, 61, 97]. 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 [98] which uses the STRIDE algorithm [99]. 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 16.30 Å which is the distance between residues 2 and 13.

3.2.2 Surface Model

The peptides were attached to a short-range, repulsive surface, located at the $z = 0$ plane, that did not interact with the peptide until a residue came into close proximity. The influence of the surface on the system is described by a Lennard-Jones type potential:

$$V_{\text{surface}} = \sum_{i}^{N} \left\{ \epsilon_{sur} \left[\left(\frac{\sigma_{sur}}{z_{is}} \right)^{9} - 7.5 \left(\frac{\sigma_{sur}}{z_{is}} \right)^{3} + c \right] \right\} \tag{3.1}$$

where z_{is} is the distance between site i and the surface, $\epsilon_{sur} = 0.0363$ kcal/mol, and the value of σ_{sur} is residue specific. Previous work has shown that the exact value of ϵ_{sur} has little effect on the behavior of the protein [86]. The parameter c is chosen such that the potential falls smoothly to zero at $z_{is} = \left(\frac{2}{5}\right)^{\frac{1}{6}} \sigma_{sur}$. This interaction remains zero for distances greater than this cutoff. The peptide is bound to the surface by a harmonic restraint with an interaction potential of the form:

$$V_{\text{restraint}} = \frac{1}{2} k_r r^2 \tag{3.2}$$

where k_r is the parameter describing the strength of the restraint and r is the distance of the restrained site from its original position of (0, 0, 5.8) Å. For each type of surface, $k_r = 100$ kcal/mol.

3.2.3 Experimental Design

The effect of the surface on protein stability was quantified by simulating the five proteins 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

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				

and the radius of gyration were calculated to analyze the correlation between structures and stabilities.

Comparing the stability of tethered protein to bulk protein is done using T_m and ΔG_f. 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 different environments is commonly done by defining the quantity $\Delta\Delta G$. For the present purposes, $\Delta\Delta G = \Delta G_f^{\text{surface}} - \Delta G_f^{\text{bulk}}$ which is the difference between the Gibbs energy of folding on the surface and in the bulk. As the 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, $\Delta G_f^{\text{bulk}} = 0$ by definition and $\Delta\Delta G = \Delta G_f^{\text{surface}}$. 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.

3.2.4 Simulation Protocols

To prevent the simulation from becoming trapped in local energy minima, simulations were performed using the replica exchange (RE) algorithm [100, 101]. Twenty-four replicas were used for each protein, and the canonical ensemble was generated using the Nosé-Hoover-Chain [102, 103, 104] 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 equilibration followed by 30 million steps of production. Swaps were attempted every 2000 steps, and 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. By tracking the potential energies, it is known that 10 million steps of equilibrium is long enough for all simulation systems described in this chapter and Chapter 4, 5, and 6.

3.2.5 Thermodynamic Quantities Calculation

The metrics used to quantify stability were calculated from simulation data 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

$$C(T) = \frac{\langle U^2 \rangle_T - \langle U \rangle_T^2}{RT^2} \tag{3.3}$$

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

$$X(T) = \langle X \rangle_T = \frac{\sum_U X(U)\Omega(U)e^{-\beta U}}{\sum_U \Omega(U)e^{-\beta U}}. \tag{3.4}$$

where $\beta = \frac{1}{k_b T}$ and k_B is Boltzmann's constant. The key quantity needed to evaluate Equation 3.4 is the density of states, $\Omega(U)$, which is calculated using the Weighted Histogram Analysis Method (WHAM) [105] on the data obtained from replica exchange simulations.

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

$$\Delta G_f = G_{\text{folded}} - G_{\text{unfolded}} = -k_B T \ln\left(\frac{P_f}{1 - P_f}\right), \tag{3.5}$$

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 a protein at its melting temperature–a relationship which must be true by definition.

The enthalpy change, ΔH_f, associated with 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 [106, 86, 88]. The change in entropy is then obtained from $T\Delta S_f = \Delta H_f - \Delta G_f$.

To evaluate the reproducibility of the results, a total of $N = 6$ independent RE simulations were performed for each protein in each environment. Results for an arbitrary property, P, are presented as the average, $\langle P \rangle$, of the N replicate values. Uncertainties are calculated from these N quantities as $\sigma_{\langle P \rangle}/\sqrt{N-1}$, where $\sigma_{\langle P \rangle}$ is the standard deviation of the N averaged property values. The error bars reported later are $\pm\sigma_{\langle P \rangle}/\sqrt{N-1}$.

3.2.6 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. 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

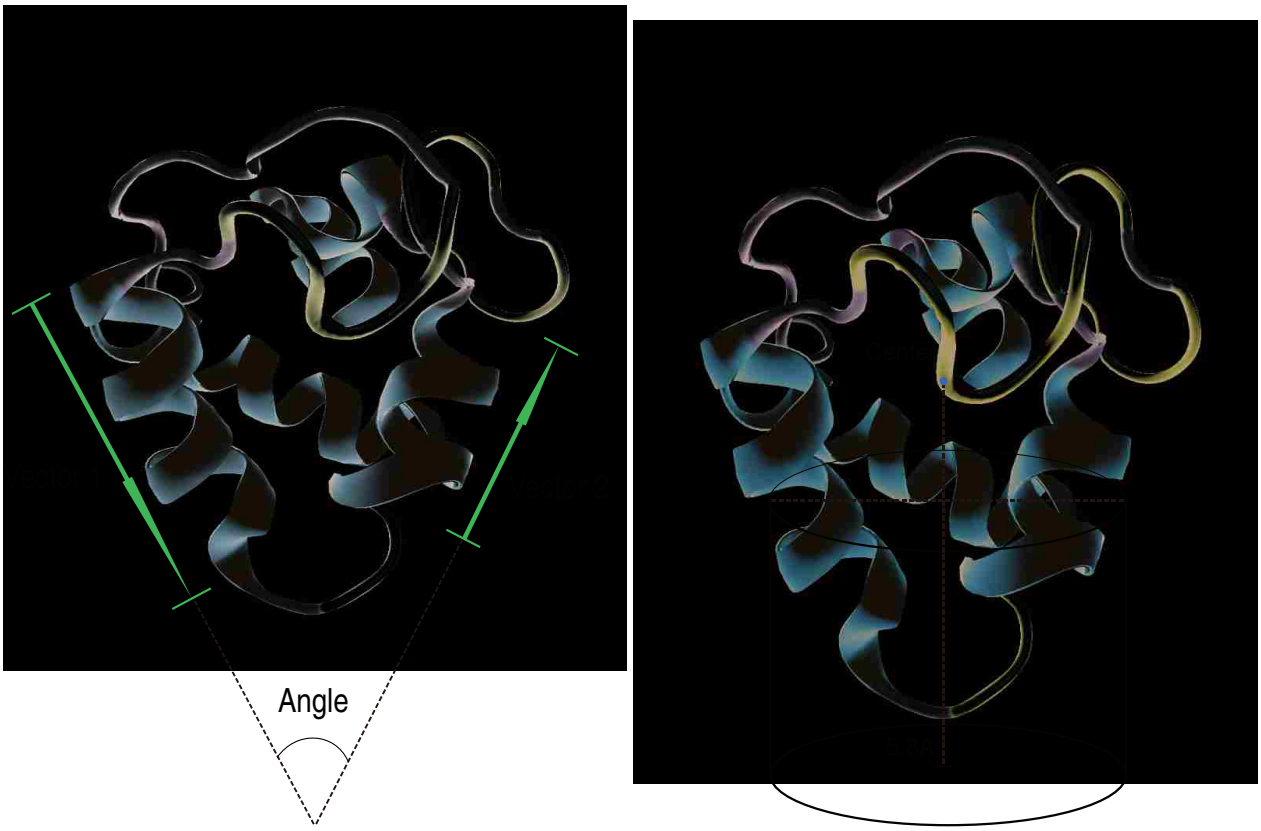


Figure 3.2: Order parameters defined as (a) angle and length (b) free rotation volume.

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. [107] The protein volume, V_p, is the sum of the residue volumes. Then the VFAFR per atom is then given by

$$VFAFR = \frac{V_c - V_p}{V_c}. \tag{3.6}$$

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 complement 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.

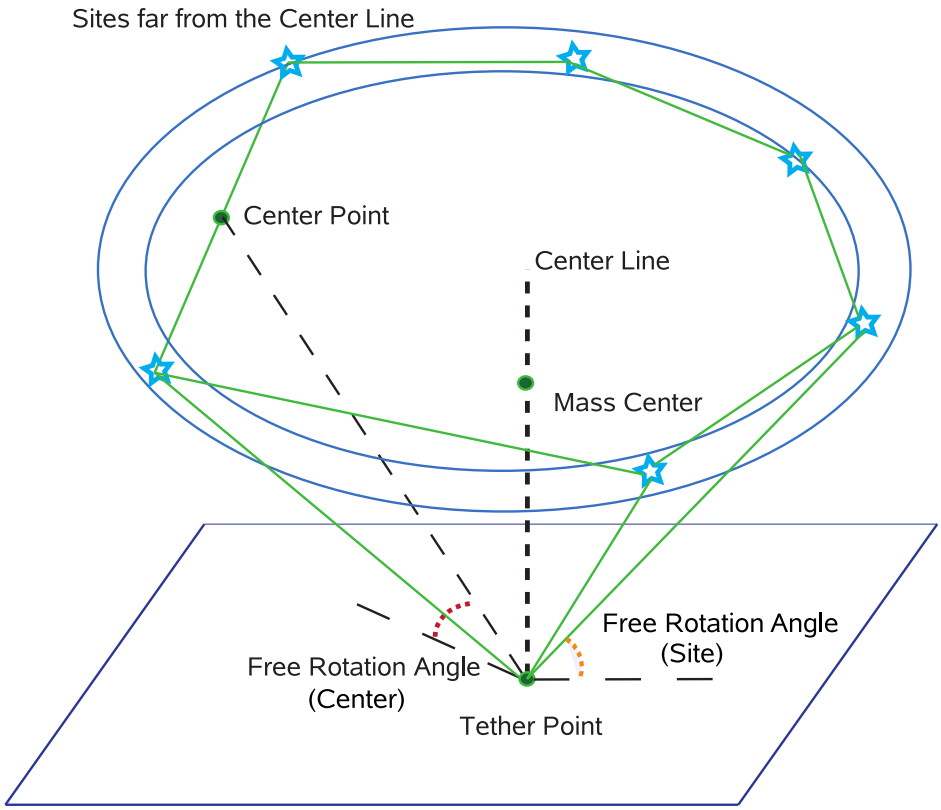


Figure 3.3: Definition of the free rotation angle for proteins.

3.3 Results and Discussion

3.3.1 Melting Temperatures

As described in the previous sections, stabilities were quantified using several thermodynamic quantities and order parameters. Figure 3.4 shows C, Q, and R_g as a function of temperature for 1R69. (The other four proteins studied show similar behavior in C, Q, and

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

Figure 3.5 shows a summary of the melting temperatures, identified from the respective heat capacity curves, of the five 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 each of the loop regions joining adjacent helical 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".

The melting temperatures show a large amount of variability according to which site is tethered. Tethering the protein resulted in melting temperatures which were approximately equal to or greater than the bulk values in 31 of the 42 cases (74%). Included in these 31 are the 4 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 from 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 glance, the data in Figure 3.5 indicate that the hypothesis is incorrect. However, a careful examination of the data reveals an interesting pattern.

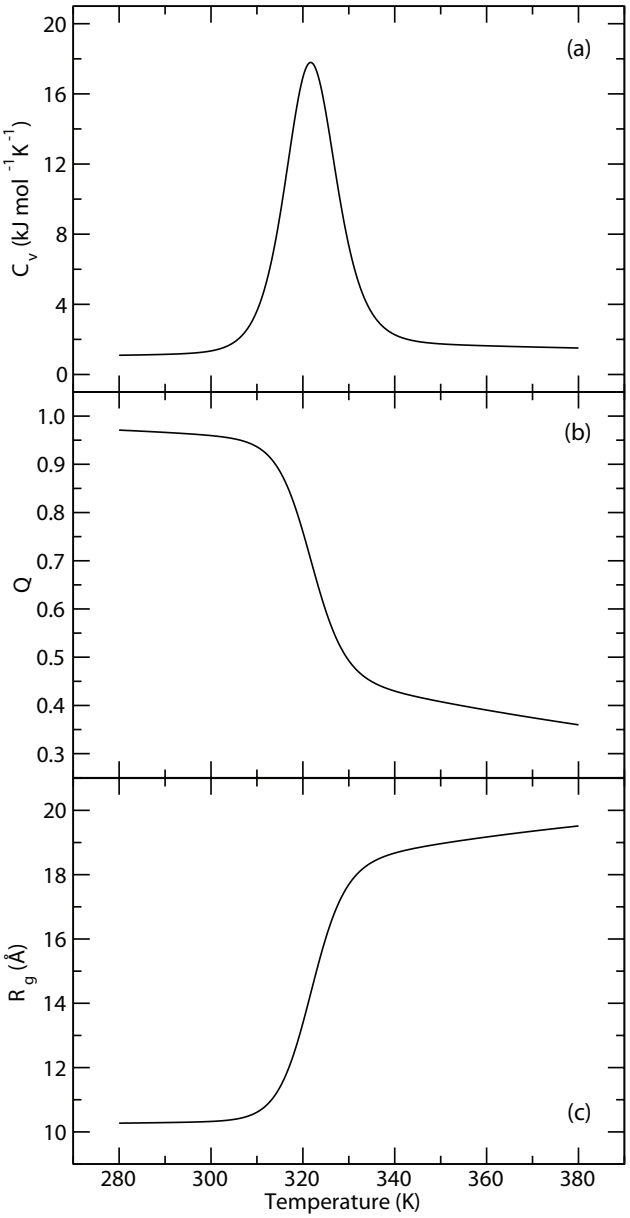


Figure 3.4: Heat capacity (Panel a), fractional nativeness (Panel b), and radius of gyration (Panel c) as a function of temperature for 1R69.

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.

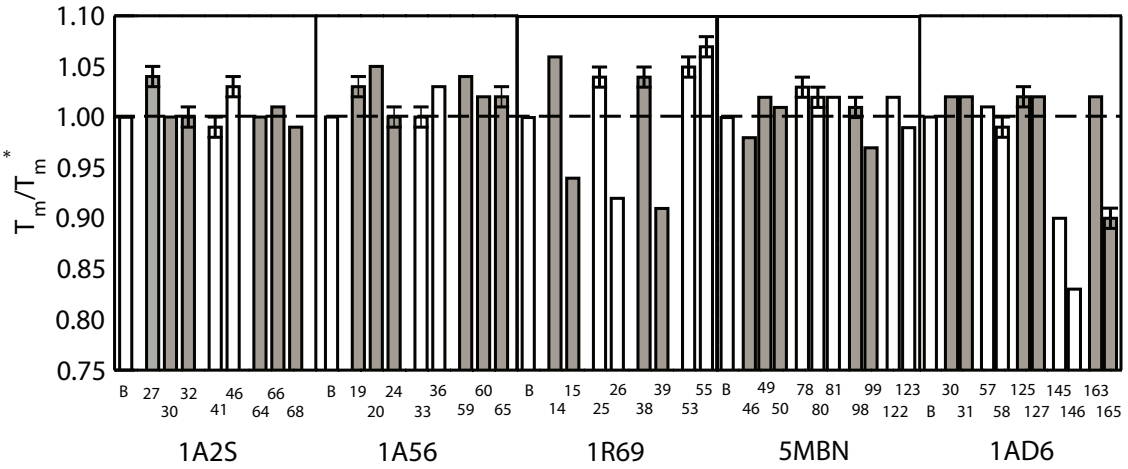


Figure 3.5: Scaled melting temperatures of 1A2S, 1A56, 1R69, 5MBN, and 1AD6 in the bulk and tethered to the surface in various locations.

3.3.3 Categorization of Tethering Sites

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 [86, 83, 84, 88] as well as the degree to which the tethered site disrupts the transition state along the folding pathway. [85] 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 example, Figure 3.6 shows angles formed by consecutive helices as a function of the melting temperatures. Notice that there is no pattern between angle and the melting temperature. As another 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, varied behavior was seen for loops less than 10 residues in length. For example, site 15 (destabilized) and site 25 (stabilized) of 1R69, are found in separate loops of 3 residues in length but have different stabilities (See Figure 3.5).

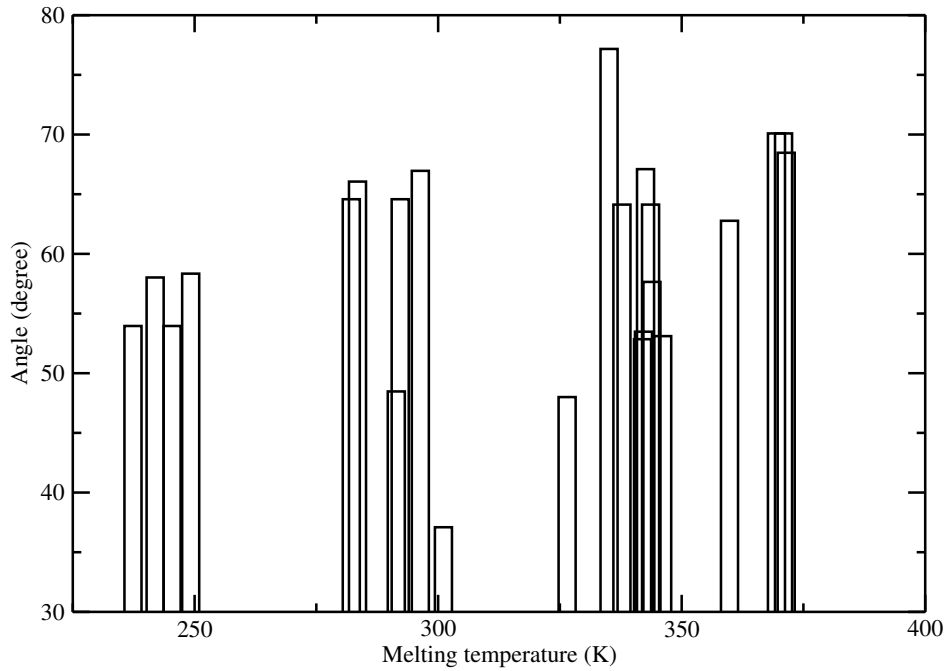


Figure 3.6: Angles formed by consecutive helices as a function of the melting temperatures.

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 3.7 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 Loop 1 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.

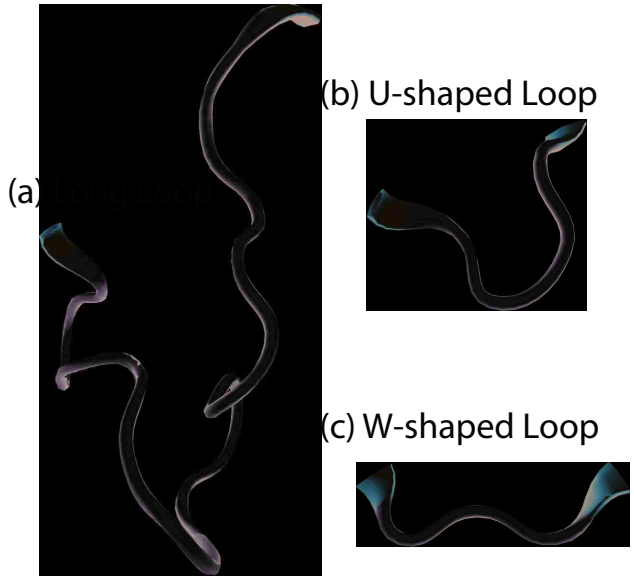


Figure 3.7: Types of loop regions found in alpha-helical, orthogonal-bundle proteins: (a) Long loop, (b) U-shaped loop, (c) W-shaped loop.

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 at 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 3.8. 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 and vibrate with respect to the surface. By tethering at site 163, one side of the molecule forms a flat foundation, but the other portion has a V-shape. By tethering at site 145, 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. 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.

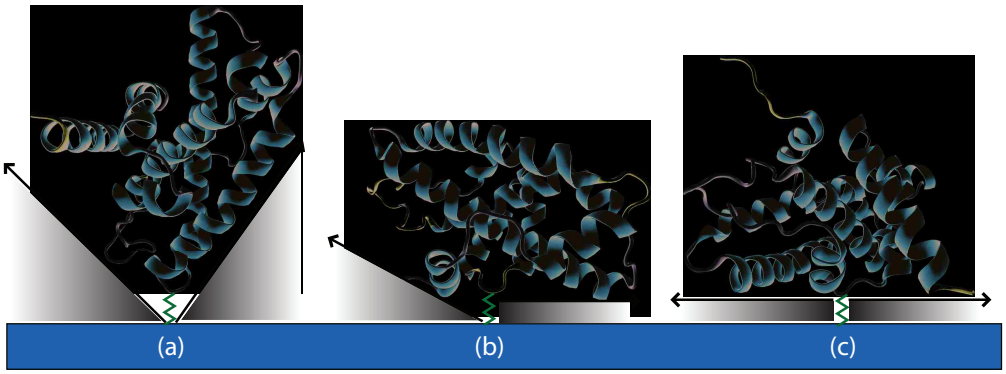


Figure 3.8: Free volume available for rotation for 1AD6 according to tether site: (a) Site 57, (b) Site 163, (c) Site 145.

The origins of the destabilization lies in the restriction of the movement of the protein. In the bulk, proteins are free to rotate through all possible angles. When tethered to the surface, the ability of the protein to rotate is inhibited by the surface itself with the degree of restriction being dependent upon the structure of the protein at the tether point. At a given temperature, the bond, angle, dihedral, and rotational vibrations of the protein seek to populate a characteristic distribution of frequencies. For tether sites which restrict the rotational freedom of the protein, such as site 145, the energy which usually partitions into the rotational modes is transferred to the bond, angle, and dihedral vibrations. The increase of energy into these vibrations causes the protein to unfold at a lower temperature than would be expected. Once unfolded, rotational modes can be populated 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 3.8). The values

were calculated assuming a quasi-harmonic approximation [108] using the Wordom analysis package [108]. 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

3.3.3.1 Free Rotation Volume

Since the lack of free rotation volume of tethering at site 145 is believed to be the major reason why this protein is destabilized on surfaces, the first metric for quantifying this difference is the measurement of the fraction of free rotation volume. However, to tether at site 145 showed about 78% of the free rotation volume as defined above, which is not consistent with the observation. The reason for the failure is because the definition of the free rotation volume is ineffective to describe the difference in free rotation or vibration ability between different tethering sites.

When the fraction of free rotation volumes was calculated, both the radius and the height of the cylinder are variables. The radius is the largest distance between the atom and

the axis, while the height is 0.7 times the length between tethering site and the mass center, plus 5.8 Å. That is an easy way to define the volume but it does not provide comparable results for different tethering cases. For example, with the site 145 tethered, the radius of the cylinder is very large due to a couple of long strands close to the surface. This leads to a very large cylinder volume compared to other cases. Most volume between these long strands in the cylinder is taken as free rotation volume but it is not accessible to the protein by rotation or vibration because the couple of long strands formed a frame that blocked the way. The free rotation volume that is interested is the volume between the frame formed by strands and the surfaces, which could be accessed by rotation or vibration of the protein.

3.3.3.2 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 14.95 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 very low 19.72. As described above, the protein is stable by tethering site 163 but unstable by tethering site 145. Even though these two sites are distinguishable somehow, the sensitivity of this metric is not very convincing due the small difference between the two angles.

Instead of using sites to define the free rotation angle, a better description of rotation or vibration ability of the protein could be the angle formed by the center point between two sites and the surface as shown in red in the Figure 3.3. This definition describes how much angle the protein can rotate or vibrate in the area between two long strands. If two long strands are close to each other, there would be not much more angle the protein could rotate or vibrate between them than below each strand. However, if the two strands formed a very large angle between them, the protein would have much more ability to access the area between them, which therefore generate larger free rotation angle. The result shown in the column of "Angle from center" in Table 3.3 proves this idea. 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 better sensitivity of this method can be seen because a larger difference of these two

cases are calculated. To make this metric as a criterion, the cutoff is needed to be measured from a larger amount of data of different sites.

For the present purposes, the easiest way to determine if the tether site has adequate rotation volume 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 shown above 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 3.8 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, but the added complexity, while increasing sensitivity, is beyond the scope of the present purposes. Moreover, as site 145 was the only instance of destabilization from inhibited rotation, the cutoff value for any metric marking the limits of stabilization/destabilization is imprecise at best. As such, quantifying the free volume available for rotation is left to future work where it can be addressed with the required detail.

Though the above analysis concerning loop structure 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 needed 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.

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	80.60	36.39	53.96
	Site 33			
	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 Site 64	76.54	28.55	48.47
	Site 66 Site 68	79.28	50.77	66.06

3. **W-shaped Loops**

   (a) Tethering in "concave-up" regions will result in stabilization if the protein can vibrate freely on the surface.

   (b) Tethering in "concave-down" regions will result in destabilization.

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 $|\Delta G_f|$. The thermodynamic perspective explaining the influence of the surface on the stability of proteins, theorized by Dill *et al.* [109, 76], predicts that proteins are always stabilized when tethered to short-ranged, repulsive surfaces. The reason is summarized in Figure 3.9. 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 surface proteins 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 on the surface. In short, the theory suggests that the entropic cost of folding is greater in the bulk than on the surface because unfolded bulk peptides have more entropy to lose than unfolded surface proteins. A decrease in the entropic cost results in a more negative value of ΔG_f.

The results in Figure 3.5 indicate that proteins are not always stabilized when tethered to surfaces as the theory predicts. 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 [86, 84, 88], 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, which has received little treatment in the literature, 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, that is, 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,

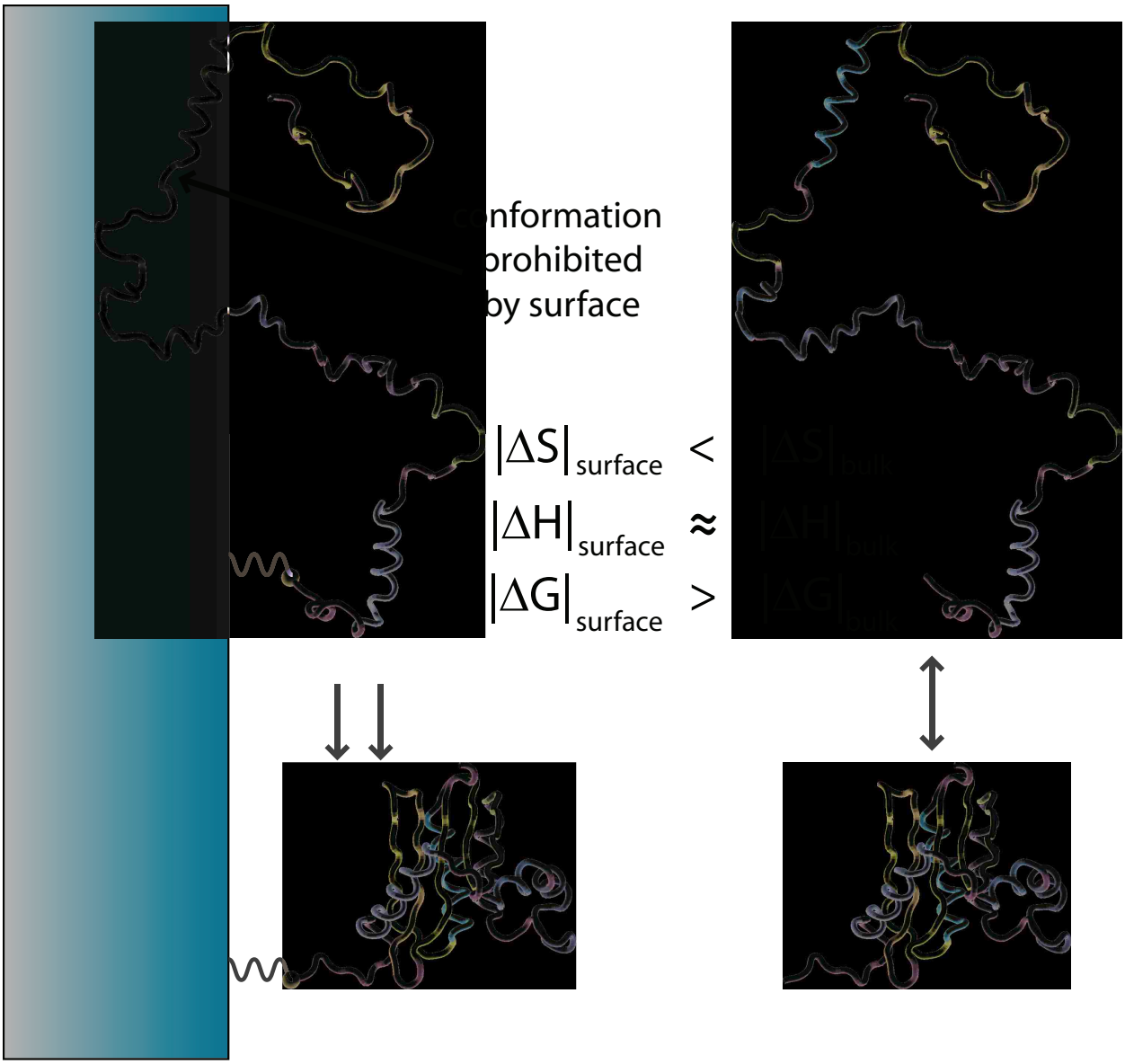


Figure 3.9: Theory behind the stabilizing influence of the surface on tethered proteins.

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.

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. 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 3.10 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 between the *change that occurs upon folding* (Δ) and 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 a similar relationship holds 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, values of ΔG_f are also shown in the Figure 3.10.

Figure 3.10 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 rather than the *unfolded* state. The data show that the unfolded states are always affected *in the same direction* by the surface, but the folded states show varied behavior. In each case, the enthalpy of the unfolded state on the surface is less (more favorable) than the enthalpy of the unfolded state in the bulk, but the enthalpy of the folded state on the surface is sometimes greater than and sometimes less than the value found in the bulk. The reason for the unfolded-state behavior is straightforward. Whether in the bulk or on the surface, 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.

Table 3.4: Thermodynamic quantities of proteins.

Protein	Location	ΔG_f (kJ/mol)	ΔH_f (kJ/mol)	$T\Delta S_f$ (kJ/mol)	Shape
	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
1R69	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
	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
1A56	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
	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
1AD6	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
	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

In contrast to the unfolded state, the surface affects the folded state in ways that do not always allow for favorable contact to be made. For sites that result in overall stabilization (sites 14, 25 38 53, and 55), $\delta H_{\text{folded}} < 0$ suggesting that the surface improves the ability of the protein to make favorable contacts. The degree of stabilization in these cases is similar in magnitude to that seen for the unfolded state (i.e. $\delta H_{\text{folded}} \approx \delta H_{\text{unfolded}}$), 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 (sites 15, 26, and 39), $\delta H_{\text{folded}} > 0$ meaning that the surface inhibits the

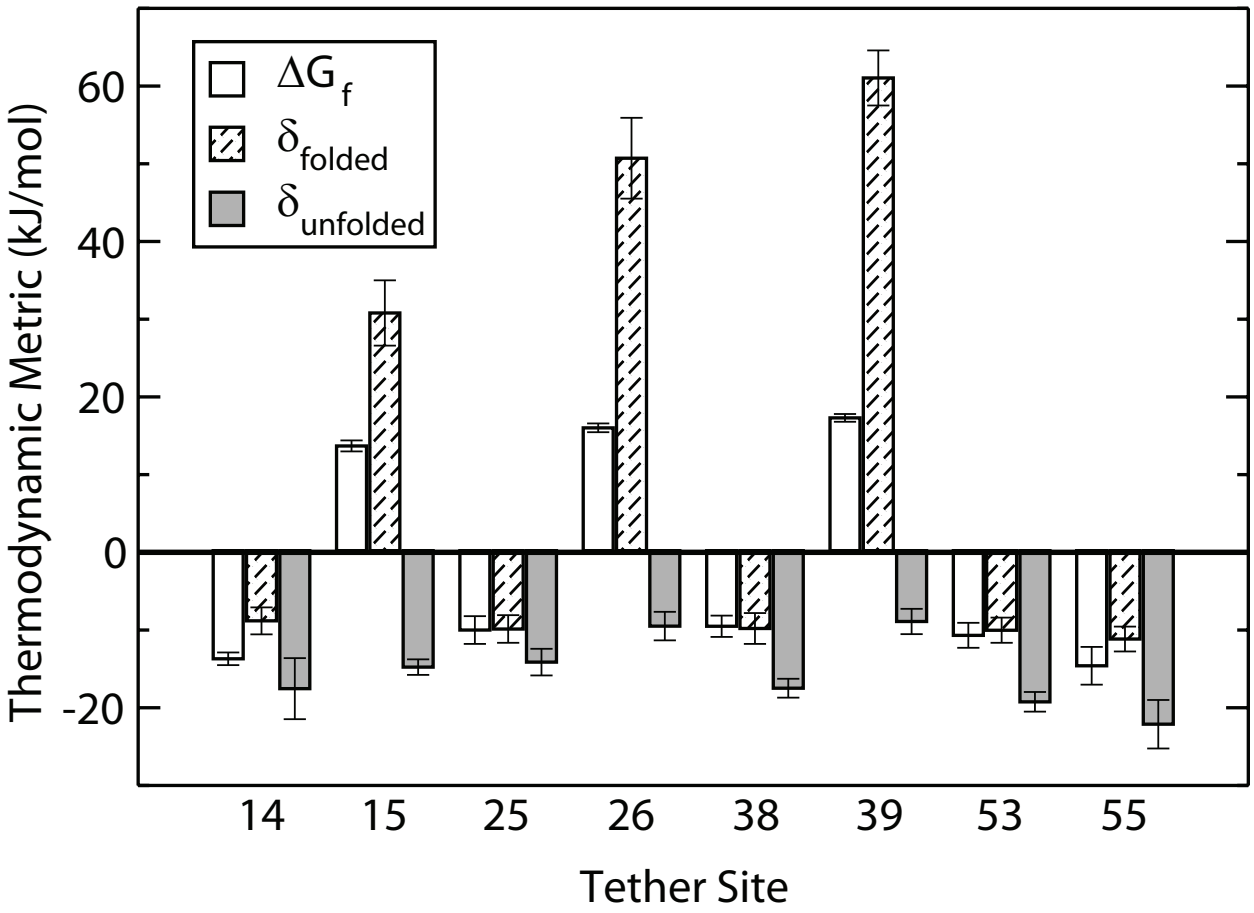


Figure 3.10: Influence of the surface on the enthalpy of the folded and unfolded state of 1R69 at $T = T^*$.

formation of favorable contacts. The extent of destabilization of the folded state is so great that $\Delta H_f^{\text{surface}} > \Delta H_f^{\text{bulk}}$.

It is instructive to discuss the enthalpic effect in a similar manner as was done for entropy (see Figure 3.9 and the accompanying discussion). Just as folding "costs" Gibbs energy entropically, it "profits" enthalpically. The surface either allows the protein to energetically profit from folding by the same amount that it would in the bulk or it reduces the profit. When the latter occurs, the Gibbs energy of folding on the surface is greater than it would be in the bulk and the result is destabilization.

The above analysis significantly improves current understanding of protein-surface interactions 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, and the effect is concentrated on the unfolded state. Proteins will be stabilized or destabilized depending upon the extent to which the surface affects the ability of the protein to form correct enthalpic contacts, and the effect is concentrated on the folded state. The enthalpic effect is related to the loop analysis described previously. Sites in long loops, U-shaped loops, and the concave-up regions of W-shaped loops, which have adequate free rotation volume, allow the folded state of the protein to exist on the surface as is does in the bulk. As such, $\Delta H_f^{\text{surface}} \approx \Delta H_f^{\text{bulk}}$ 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. Consequently, $\Delta H_f^{\text{surface}} > \Delta H_f^{\text{bulk}}$ and the result is enthalpic destabilization.

3.3.5 Applicability of the Results

This study used a coarse-grain representation to model the proteins of interest. As with all simulation results, the findings presented above are only valid if the model can reproduce the relevant physics involved. The model used in this study (as stated in the Methods section) has been shown to reproduce realistic folding mechanisms compared to experimental results and has been extensively used to study protein folding [58, 59, 60, 61, 97, 97]. It is therefore reasonable to assume that the *trends* seen in this study are applicable to real situations.

Because coarse-graining reduces the number of degrees of freedom of the system, and solvent interactions are included only implicitly, the numerical values reported for the thermodynamic properties are not expected to correspond exactly to those that would be seen if experimental results could be obtained. Moreover, as neither the surface nor the protein model explicitly account for chemical composition, the results presented above are only valid for weakly-interacting surfaces and not for surfaces of mixed chemical composition or those that are highly attractive or repulsive. Despite these limitations, the model is expected to approximate the biophysics involved sufficiently enough to give realistic trends. In a field where atomic simulations are not possible due to computational limitations, and current experimental techniques cannot probe the relevant phenomena, the reported results are welcomed findings.

3.4 Summary

The results of the chapter 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 surfaces 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 on surfaces.

Chapter 4

The Stability of Proteins in Other Tertiary Motifs on Surfaces

The previous chapter outlined research that showed how to predict the stabilities of alpha-helical, orthogonal-bundle proteins on surfaces using only the structure of the protein. According to the prediction pattern, this class of proteins could be stabilized on a surface if the tethered protein has large free rotation volume and the tethering is done at a "concave-up" residue in the loop region. The question is if the prediction pattern derived for mainly-alpha, orthogonal-bundle proteins is transferable to proteins with other tertiary structure. This questions is addressed in this chapter using several proteins from three other tertiary motifs.

4.1 Method

Identifying if there are structural patterns that can be correlated to stability when proteins other than alpha-helical, orthogonal bundles are tethered to the surface was done in a similar way as was described in Chapter 3. Various proteins are selected and tethered to the surface at different locations. For these proteins, the tethering sites were selected based upon the results described above. Namely, the tether sites were selected based upon the VFAFR and the type of loop region to see if the patterns found previously are transferable to other classes of proteins.

4.1.1 Proteins with Different Tertiary Motifs

As was done in Chapter 3, the proteins were chosen according to the CATH classification method [110, 111]. The three tertiary motifs considered in this research, as shown in Table 4.1, were updown-bundle (Figure 4.1), beta-barrel (Figure 4.2), and beta-sandwich (Figure 4.3). These were chosen because they are the most abundant tertiary motifs in the

mainly-alpha and mainly-beta secondary motifs. Together with mainly-alpha, orthogonal-bundle motif studied in chapter 3, 96.57% of mainly-alpha and 76.13% of mainly-beta protein domains are represented.

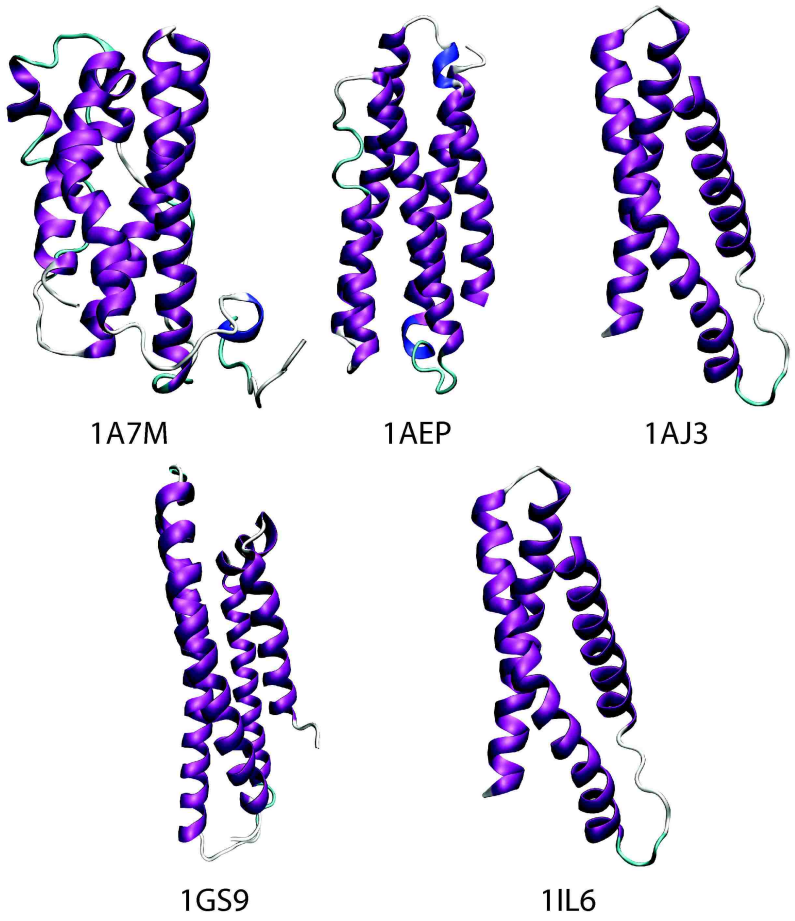


Figure 4.1: Alpha-helical, updown-bundle proteins.

4.1.2 Simulation Model

For computational efficiency, the Gō-like model of Karanicolas and Brooks [58, 59, 60, 61] is used as in previous research (see Chapter 3 for details). Input files are generated using the MMTSB website `http://www.mmtsb.org`.

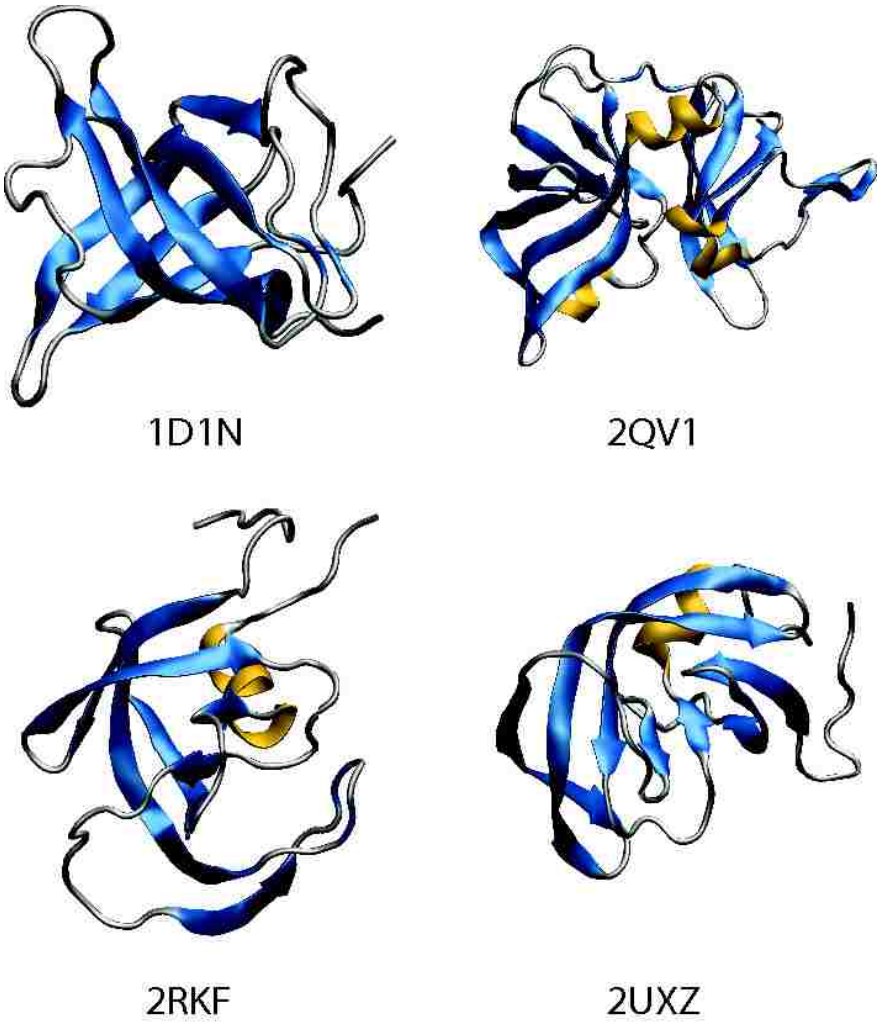


Figure 4.2: Beta-barrel proteins.

4.1.3 Surface Model

The peptides are attached to a short-range, repulsive surface that does not interact with the peptide until a residue came into close proximity. The influence of the surface on the system is described by a Lennard-Jones type potential:

$$V_{\text{surface}} = \sum_{i}^{N} \left\{ \epsilon_{sur} \left[\left(\frac{\sigma_{sur}}{z_{is}} \right)^{9} - 7.5 \left(\frac{\sigma_{sur}}{z_{is}} \right)^{3} + c \right] \right\} \tag{4.1}$$

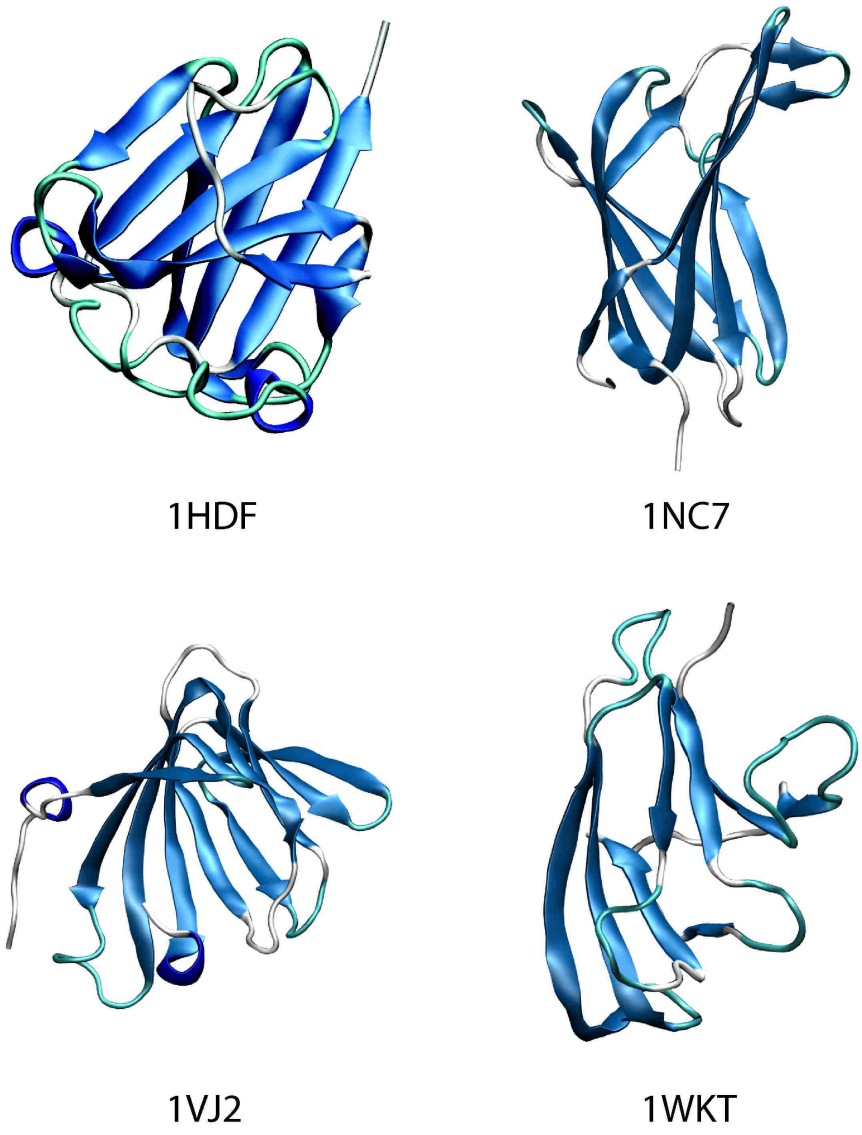


Figure 4.3: Beta-sandwich proteins.

where z_{is} is the distance between site i and the surface, $\epsilon_{sur} = 0.0363$ kcal/mol, and the value of σ_{sur} is residue specific. Previous work has shown that the exact value of ϵ_{sur} has little effect on the behavior of the protein [86]. The parameter c is chosen such that the potential falls smoothly to zero at $z_{is} = \left(\frac{2}{5}\right)^{\frac{1}{6}} \sigma_{sur}$. This interaction remains zero for distances greater than this cutoff. The peptide will be bound to the surface by a harmonic restraint with an

Table 4.1: Protein structure motifs for further study.

Secondary	Tertiary	PDB ID
Mainly α	Up-down Bundle	1A7M
		1AEP
		1AJ3
		1GS9
		1IL6
Mainly β	β-Barrel	1D1N
		2RKF
		2UXZ
		2QV1
	β-Sandwich	1HDF
		1NC7
		1VJ2
		1WKT

interaction potential of the form:

$$V_{\text{restraint}} = \frac{1}{2} k_r r^2 \tag{4.2}$$

where k_r is the parameter describing the strength of the restraint and r is the distance of the restrained site from its original position of (0, 0, 5.8) Å.

4.1.4 Experimental Design and Simulation Protocols

The comparison of stabilities of proteins tethered to surfaces and proteins in the bulk can be done using the melting temperature T_m. 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.

To prevent the simulation from becoming trapped in local energy minima, simulations were performed using the replica exchange (RE) algorithm [100, 101]. Twenty-four replicas were used for each protein, and the canonical ensemble was generated using the Nosé-Hoover-Chain [102, 103, 104] integration method with 3 thermostats of mass 10^{-26} kgÅ^2. The time step was 1 fs and each simulation contained 10 million steps of equilibration followed by 30 million steps of production. Swaps were attempted every 2000 steps, and temperature

increments 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.

4.2 Results and Discussion

4.2.1 Up-down Proteins

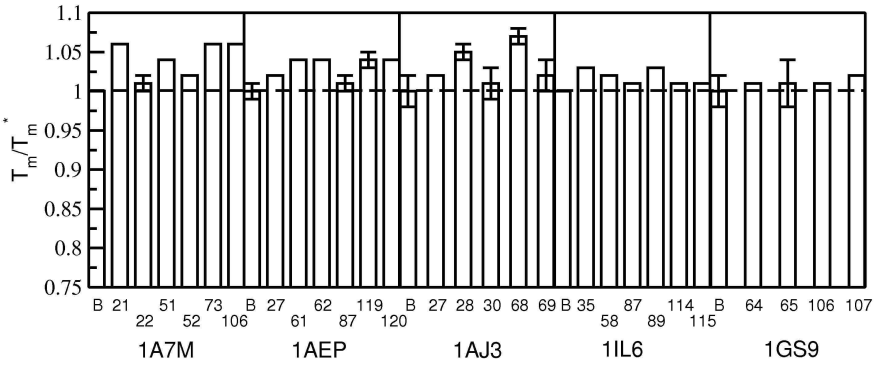


Figure 4.4: Scaled melting temperature of updown-bundle proteins in the bulk and on surfaces.

Figure 4.4 shows the scaled melting temperatures for the five mainly-alpha, updown-bundle proteins 1A7M, 1AEP, 1AJ3, 1IL6, and 1GS9. Several tethering sites are chosen for each protein that are expected to be both stabilized and destabilized based upon their VFAFR and loop structure. As shown in Figure 4.4 all sites resulted in stabilization. Thus it appears that up-down bundles behave slightly differently than orthogonal bundles. The latter exhibited both stabilization and destabilization (see Chapter 3) while the former only stabilization.

Site 52 in protein 1A7M is an example of a residue that is predicted to be unstable (it is a concave down site in a W-shape loop region) but is actually stabilized. The structure is shown in Figure 4.4. These results suggest that the shape of the loop region is of secondary importance for these proteins and that the VFAFR is the dominant parameters governing stabilization. Due to their elongated shape, most tethering sites in updown-bundle proteins have very large VFAFR. The parallel shape of alpha helices makes it is easier for loop regions

to access to surfaces, so the surface has less effect on the interhelical forces stabilizing the protein. In summary, for mainly-alpha, updown-bundle proteins, the VFAFR is the main factor that the correlates to stability on the surface.

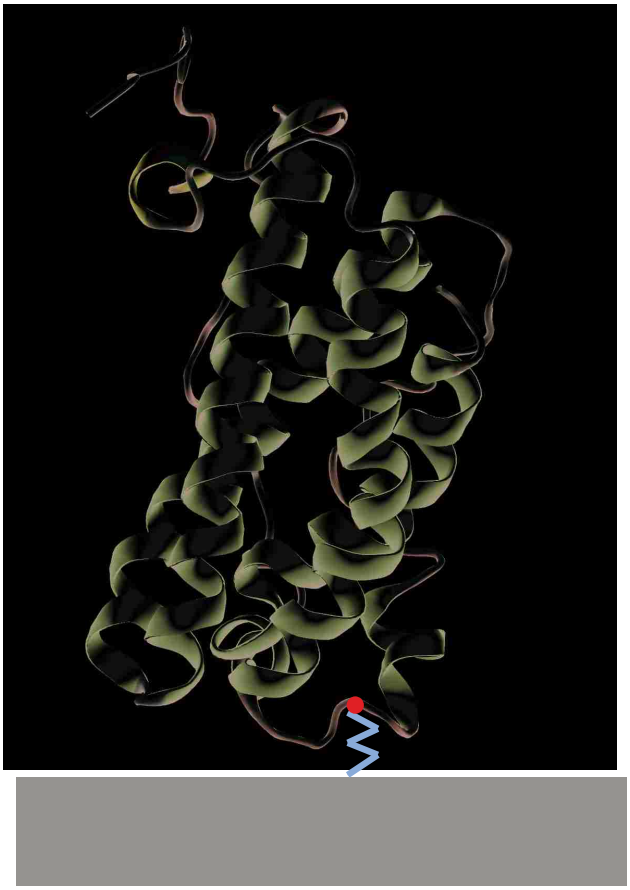


Figure 4.5: Protein 1A7M is tethered on surface with site 52.

In contrast, when dealing with mainly-alpha, orthogonal-bundle proteins, it is important to determine the shape of the loop region at the tethering point. This is because this class of proteins in general has a large VFAFR only for concave-up residues. For example, in the specific case of site 145 in protein 1AD6, the free rotation volume is not even large enough when tethering to concave-up sites. Compare this to mainly-alpha, updown-bundle proteins. Here, the VFAFR is always very large so that both concave-up and concave-down

sites results in stabilization on the surface. In summary, for mainly-alpha proteins, both the VFAFR and the shape of tethering loop region are important factors for predicting protein stabilities on surfaces, and the rotation volume of the tethering loop region should be considered first.

4.2.2 Beta-sandwich Proteins

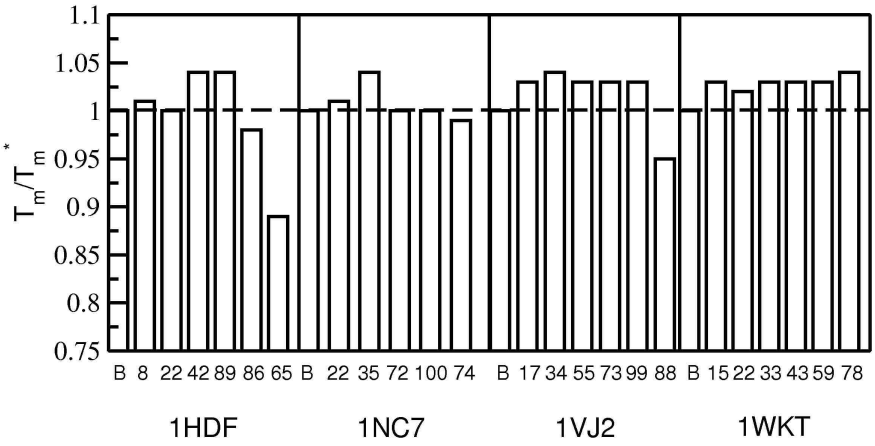


Figure 4.6: Scaled melting temperature of beta-sandwich proteins in the bulk and on surfaces.

Figure 4.6 shows the scaled melting temperatures for four beta-sandwich proteins (1HDF, 1NC7, 1VJ2, and 1WKT) in the bulk and tethered to the surface at different locations. Residues for tethering were selected that were expected to be both stabilized and destabilized based upon the VFAFR and loop structure. Those residues expected to be destabilized because they were tethered in the concave-down region of the loop are sites 86 and 65 in 1HDF, site 74 in 1NC7, and site 88 in 1VJ2. The scaled melting temperatures for each of these sites is less than one. This means that the protein tethered at these residues was destabilized as predicted from the rules generated from the mainly-alpha, orthogonal-bundle results. Also in agreement with the rules is that all the others sites, which were predicted to stabilize proteins, did result in stabilization when simulated. Beta-sandwich proteins have a globular shape similar to that of mainly-alpha, orthogonal-bundle proteins, and thus have

sufficiently-large VFAFR to be stabilized when tethered to a surface in a concave-up loop. Also like the alpha-helical, orthogonal-bundle proteins, the VFAFR is not large enough when the proteins are tethered in concave-down loop regions.

4.2.3 Beta-barrel Proteins

Four beta-barrel proteins were simulated both in the bulk and on surfaces to examine another class of proteins. As before the proteins were tethered to the surface in multiple locations that were expected to be both stabilized and destabilized based upon the loop structure and the VFAFR. Unstable sites were predicted in each of the proteins. These were site 84 in 1D1N, site 68 in 2QV1, site 38 in 2UXZ and sites 36 and 8 in 2RKF.

In accordance with prediction, tethering to most of the sites in 1D1N resulted in stabilization. The one exception was site 84. Though this site is found in a long loop region, tethering results in very little free rotation volume as shown in Figure 4.8. The simulations show that tethering at this site does destabilize the protein as predicted. A low VFAFR is also observed for site 68 for the protein 2RKF as shown in Figure 4.9 and simulation also shows that this site results in destabilization. In fact, the prediction method holds true for all but one of the tethering sites investigated. This was site 51 in 2UXZ. As shown in Figure 4.10, this site is in a U-shape loop region and there is large free rotation volume for the protein with this site tethered. It is therefore expected to be stabilized. But the simulations show that tethering at this site actually destabilizes the protein.

One reason for this unexpected behavior could be that the two beta strands connected by the loop region are long and extend relatively far from the rest of the protein (see Figure 4.10). This means there are fewer attractive forces stabilizing these strands, so the surface can easily affect the structure of this region of the protein which leads to the destabilization of the protein. However, future work is needed to firmly establish the cause.

4.3 Summary

The purpose of this chapter was to show how the rules for stability on the surface, generated using alpha-helical, orthogonal-bundle proteins, transferred to other tertiary structure motifs. Three other protein families were investigated. The results show that the prediction

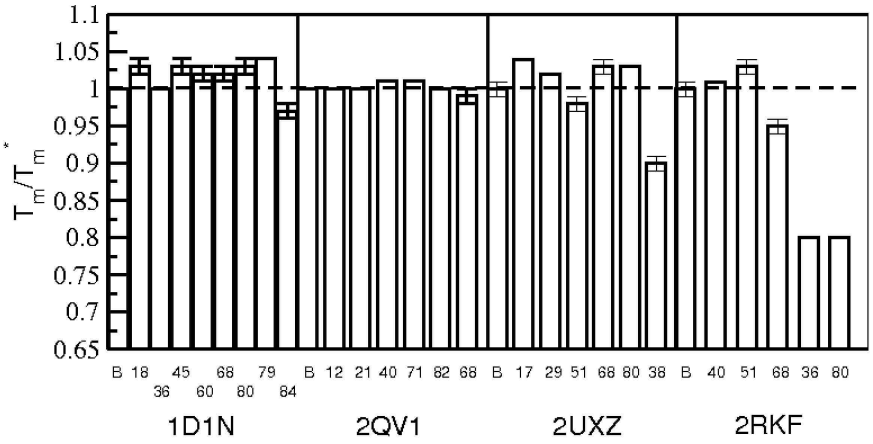


Figure 4.7: Scaled melting temperature of beta-barrel proteins in the bulk and on surfaces.

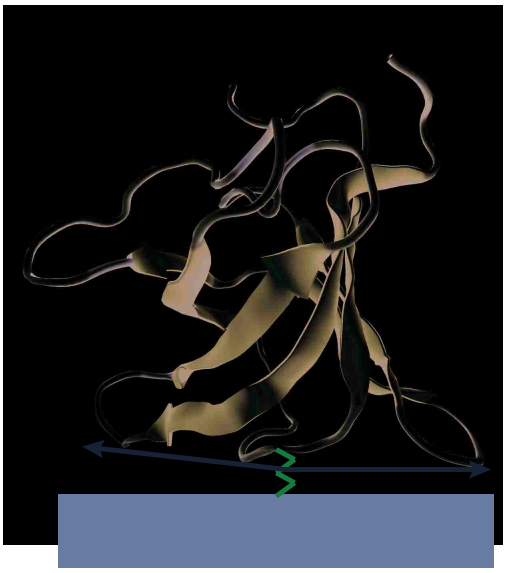


Figure 4.8: Protein 1D1N is tethered on surface with site 84.

patterns could generally be applied to many proteins regardless of tertiary structure motifs. It was also observed that beta-barrel proteins seem more unstable than other classes. The VFAFR for beta-barrel proteins, in general, is fairly low. This is not the case for other classes of proteins. In fact, up-down-bundles, by nature of the loops being found at the ends of their elongated structures, rarely have low VFAFR. The results also indicate that mainly-alpha

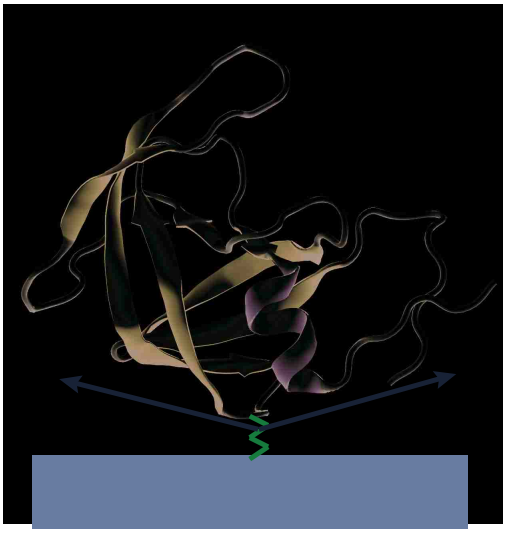


Figure 4.9: Protein 2RKF is tethered on surface with site 68.

Figure 4.10: Protein 2UXZ is tethered on surface with site 51.

proteins are generally more stable than mainly-beta proteins. In summary, over 119 tethering locations were tested on 18 proteins. Out of these 119 cases, only one failed to follow the rules for stabilization. Thus, taken as a whole, the results of this and the previous chapters

shows that most proteins will be stabilized on the surface if two conditions are met: 1) the tethering is done in a concave-up residue of the loop region and 2) the tethering results in a sufficiently large VFAFR.

Chapter 5

Surface Induced Changes to Folding Mechanism

The results presented in the previous two chapters, as well as previous studies mentioned in Chapter 1, have offered many valuable insights into the folding of proteins on surfaces. However, a careful reading of the literature reveals that all such studies involve proteins which fold through a two-state mechanism. A two-state mechanism means that the protein is only found in one of two states: folded and unfolded. Most small proteins fold through such a two-state mechanism, but as proteins increase in size, many fold through multistate mechanisms which include one or more partially-folded but stable intermediates.

The purpose of this chapter is to investigate how surfaces affect the folding mechanisms of multistate folders. Specific focus is placed upon finding the tethering configurations which stabilize the multistate folder on the surface and keep the active site available to the bulk phase–two necessary requirements for proper array function. The remainder of this chapter is structured as follows. First, a description of the protein used in the study, completed with an analysis of secondary and tertiary structure, is presented. This is followed by an outline of the methods used to compare folding on the surface and in the bulk. The results, which illustrate how the surface can drastically alter the folding mechanism of a protein, are then presented followed by a discussion about the impact of the findings. In all, the results of this chapter show that the effect of surfaces on the folding of multistate folders involves rich phenomena not seen in previous studies involving two-state folders.

5.1 Methods

5.1.1 Protein

Bacteriophage T4 lysozyme (PDB ID: 7LZM) was used as a model system to determine how tether site location can affect the folding mechanism of surface-tethered proteins.

This protein is 164 residues in length and folds through a multistate mechanism as will be described later. 7LZM is classified as a mainly-alpha, orthogonal-bundle protein by the CATH classification scheme [112, 95, 96]. It is composed of ten alpha-helices, one 3_{10} helix, and a four-stranded beta-sheet. Figure 6.8 shows a cartoon schematic of 7LZM with the secondary structural elements as identified using the STRIDE [99] program implemented in VMD [98]. Effort has been taken to code the secondary structure type using color. Helices are represented by cool colors (*e.g.* blues and violets) and strands with warm colors (*e.g.* reds and oranges). The protein has two sections. Helices 1 and 4-11 form a orthogonal bundle. Helix 2 along with the beta-sheet composed of strands 1-4 form an alpha/beta roll. Helix 3 connects the bundle with the roll. The residues comprising each secondary structure and loop region are found in Table 5.1.

Table 5.1: Residue-level secondary structure analysis of 7LZM with site numbers corresponding to the associated residue number in the primary amino acid sequence of the protein.

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

Lysozyme has been very carefully studied with experiments as a model system for protein folding [113, 114, 115, 116, 117, 118]. These works have shown that there is an intermediate detected in the folding pathway and therefore a multi-stage folding mechanism could be expected. Folding includes an early hydrophobic collapse of the alpha-domain of the protein [116, 119] and a stage of slow folding of the beta-domain after that. In addition, 20% of the molecules fold in a fast way without an observable intermediate [117]. Many efforts have been put into understanding the importance of the intermediate state and the refolding

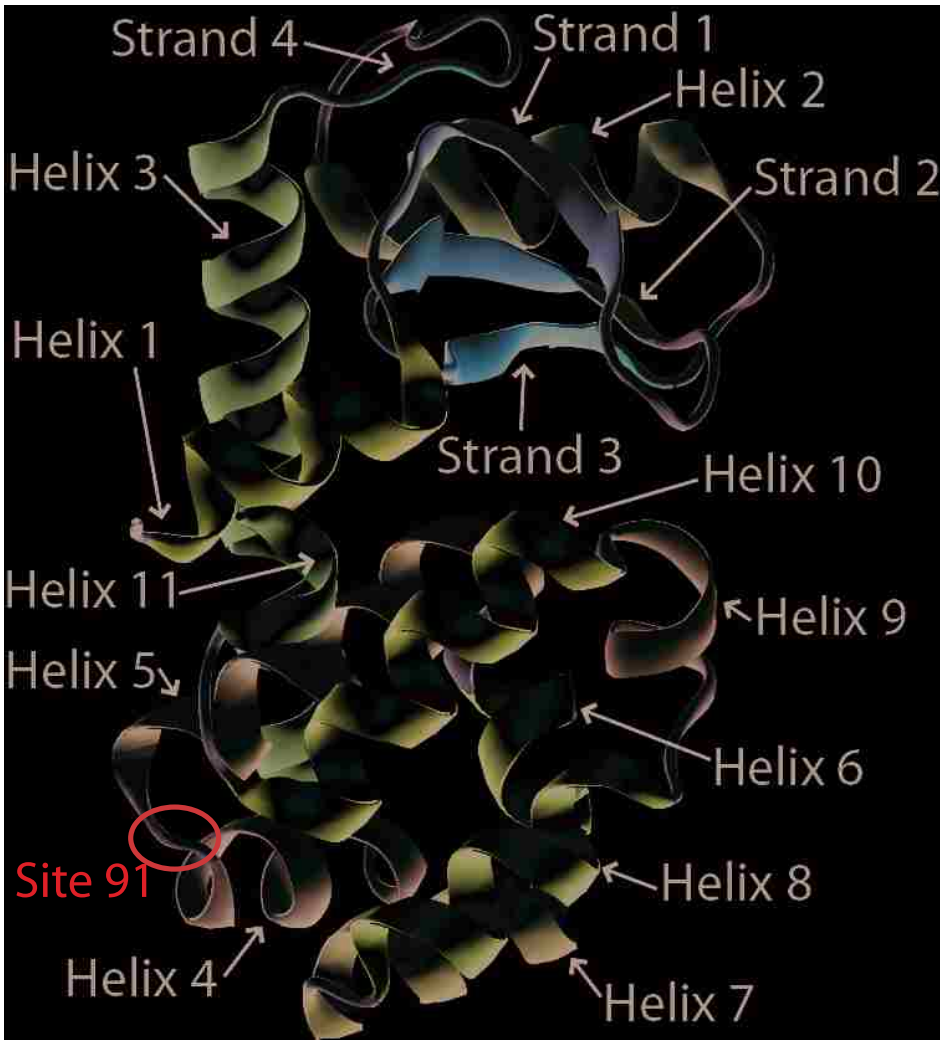


Figure 5.1: Schematic representation of 7LZM.

kinetic process of lysozyme [120]. However, for the sake of further practical implementation, it is still important to characterize the intermediate in this research. Furthermore, because it is still not easy to obtain a high revolution configuration of the intermediate by experimental means, a simulation method that can capture protein folding mechanism is used in this research. As was done in the previous two chapters, the protein is modeled using the Gō-like model of Karanicolas and Brooks [58, 59, 60, 61]. Input files were generated using tools available on the MMTSB website[1] and simulations were performed using code developed "in-house."

[1] `http://www.mmtsb.org/`

5.1.2 Surface Model

The protein was attached to a short-range, repulsive surface, located at the $z = 0$ plane, which did not interact with the peptide until a residue came into close proximity. The influence of the surface on the system is described by a Lennard-Jones type potential:

$$U_{\text{surface}} = \sum_{i}^{N} \left\{ \epsilon_{sur} \left[\left(\frac{\sigma_{sur}}{z_{is}} \right)^{9} - 7.5 \left(\frac{\sigma_{sur}}{z_{is}} \right)^{3} + c \right] \right\} \tag{5.1}$$

where z_{is} is the distance between site i and the surface, $\epsilon_{sur} = 0.0363$ kcal/mol, and the value of σ_{sur} is residue specific. Previous work has shown that the exact value of ϵ_{sur} has little effect on the behavior of the protein [86]. The parameter c is chosen such that the potential falls smoothly to zero at $z_{is} = \left(\frac{2}{5}\right)^{\frac{1}{6}} \sigma_{sur}$. This interaction remains zero for distances greater than this cutoff. The peptide is bound to the surface by a harmonic restraint with an interaction potential of the form:

$$U_{\text{restraint}} = \frac{1}{2} k_r r^2 \tag{5.2}$$

where $k_r = 100$ kcal/mol is the parameter describing the strength of the restraint and r is the distance of the restrained site from its original position of (0, 0, 5.8) Å.

5.1.3 Simulation Protocols

Simulations were performed using the replica exchange (RE) algorithm to prevent the simulation from becoming trapped in local energy minima [100, 101]. Twenty-four replicas were used for each system, and the canonical ensemble was generated using the Nosé-Hoover Chain [102, 103, 104] 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. Equilibrium, as defined by the convergence of the potential and kinetic energies of the system at the temperature of interest, occurred in roughly 3 million steps. Temperature steps between boxes ranged from 2.5 to 10 degrees across the temperature range with the smaller increments found near transition temperatures. The data were analyzed using the Weighted Histogram Analysis Method (WHAM) [105].

5.1.4 Metrics to Quantify Folding Behavior

One metric useful in identifying and quantifying folding transitions is the heat capacity. Two-state folders will display a single, sharp peak in a plot of heat capacity vs. temperature. The temperature at which the peak occurs is the melting temperature of such proteins. For multistate folders, more than one peak is present. For example, two peaks indicate a three-state folder. For such proteins, the temperatures at which the peaks occur are often termed "transition" temperatures. Another useful quantity for understanding protein folding is the *fractional nativeness* or *fraction folded*, q. Fractional nativeness measures how much of the protein is found in the native state at any particular time during the simulation. It is defined as the number of native contacts considered to be correctly formed at time t divided by the total number of contacts possible. Following the scheme of Karanicolas and Brooks [58, 59, 60, 61], a native contact is considered present if the distance between the two sites comprising the contact is less than 1.2 times the distance between the sites in the PDB structure.

Fractional nativeness is a useful metric to study protein folding for multiple reasons. First, during transitions, q decreases very quickly and a plot of q vs. T will show a sigmoidal shape. A transition temperature can be determined by identifying the location of the inflection point in such a curve. Agreement between the transition temperature found in this manner, and the corresponding value obtained from the heat capacity curve helps validate the appropriateness of the protein model. These two methods for determining transition temperatures involve different types of data (thermal vs. structural) and less robust coarse-grain models often show a lack of agreement.

The other reasons q is a useful metric are: 1) it is analogous to data that can be obtained from experimental methods such as circular dichroism and 2) it can be used to help determine the structure of the stable states involved in the folding pathway. The latter is possible by classifying each type of native contact found in the molecule and calculating the q for individual secondary and tertiary structures. For example, among the native contacts found in 7LZM are those between one site found in Helix 4 and one site in Helix 5. This would be classified as a tertiary contact, denoted as q_{H4-H5}, since it involves sites from two different secondary structures. Alternatively, a native contact could act to stabilize a single

secondary structure element, such as one formed from two sites found in Helix 3 (q_{H3-H3}). Each q_{x-x} for the molecule will undergo a sigmoidal transition at the same temperature for a two-state folder. However, for a multistate folder, different structural elements will unfold at different temperatures, and the individual q_{x-x} curves can be used to determine which portions of the molecule undergo changes at each transition.

5.1.5 Experimental Design

To determine how the surface affects the folding mechanisms of multistate folders, 7LZM was simulated in the bulk (the control) and tethered to the surface at different residues. The tether sites were chosen based upon work described in Chapters 3 and 4 where it was shown that tethering in loop regions of most proteins can lead to predictable stability patterns. As such, 7LZM was tethered to the surface in each of the loop regions joining secondary structural elements as denoted in Table 5.1. In each case, the heat capacity and fractional nativeness curves were calculated to facilitate comparisons between each surface case and the control. Two main features were looked for during comparisons. The first were changes in the location of the peaks in the C vs. T curve or the location of the sigmoidal transitions in the q vs. T curves. Such changes indicate an increase or decrease in transition temperature and a corresponding increase or decrease in stability. The second was a change in the number of transitions which would indicate a surface-induced change in the folding mechanism of the protein. When changes were identified, q_{x-x} curves were used to explain the differences.

5.2 Results

5.2.1 Folding in the Bulk

Panel A of Figure 5.2 shows the heat capacity as a function of temperature for 7LZM in the bulk. Three peaks are present in the curve indicating that 7LZM folds through a four-state process. Since a coarse-grain model with implicit solutions is used in this research, these transition temperatures can not be taken as a realistic temperature scale. However, the temperatures still capture different transition states as they shifted in the same scale. The first transition occurs at approximately 265 K, the second at 301 K, and the third at

335 K. The most prominent transition, the highest and most narrow peak, is at 335 K. The peak at the 301 K transition is small and broad. The peak at 265 K is sharper than that at 301 K but is about a third of the size of the larger peak at 335 K.

The structure of the protein in each of the stable states along the folding pathway in the bulk was determined by plotting q_{x-x} versus temperature for each type of native contact found in 7LZM. Figure 5.3 shows the results of this effort. For clarity, the figure is constructed such that contacts are grouped according to transition temperature. Panel (A) contains the contacts for the 265 K transition, Panel (B) for the 301 K transition, and Panels (C) and (D) for the 335 K transition. For convenience, the discussion that follows is taken from the perspective of melting (starting at low temperature and moving to higher temperature), but this *direction* is arbitrary as the results, obtained from equilibrium simulations, displayed no dependence on the starting structure of the protein (*i.e.* folded or unfolded).

During the transition at 265 K, part of the alpha/beta roll formed by Helix 2 and Strands 1-4 comes apart. The tertiary contacts holding the roll together–those between Strands 1-3 and Helices 2-3, the inter-strand contacts holding the sheet together, and the contacts between Helices 2 and 3–come apart, but the secondary structure of Helices 2 and 3 remains largely intact. During the transition at 301 K, Helix 1 breaks apart from the bundle. On the surface, the fact that only Helix 1 is involved in the second transition, may seem insignificant, but such is not the case. When Helix 1 is bound to the bundle formed by Helices 4-11, the top portion of the molecule forms a loop which reduces the conformations accessible to the melted alpha/beta roll. However, when Helix 1 breaks away from the bundle, the residues comprising the entire top portion of the molecule, Helices 1-3 and Strands 1-5, can sample random coil configurations with ease. During the highest-temperature transition at 335 K the entire lower portion of the molecule, the bundle formed by Helices 4-11, breaks apart.

Figure 5.4 shows the conformations of the protein in each of the four states found between the transitions. These are labeled, beginning at the low temperature, states 1-b, 2-b, 3-b, and 4-b where the "b" signifies "bulk." State 1-b is located at low temperatures and represents the folded state of the protein. State 2-b shows the protein after the alpha/beta roll has melted. Notice that because Helix 1 keeps contact with the helical bundle, the

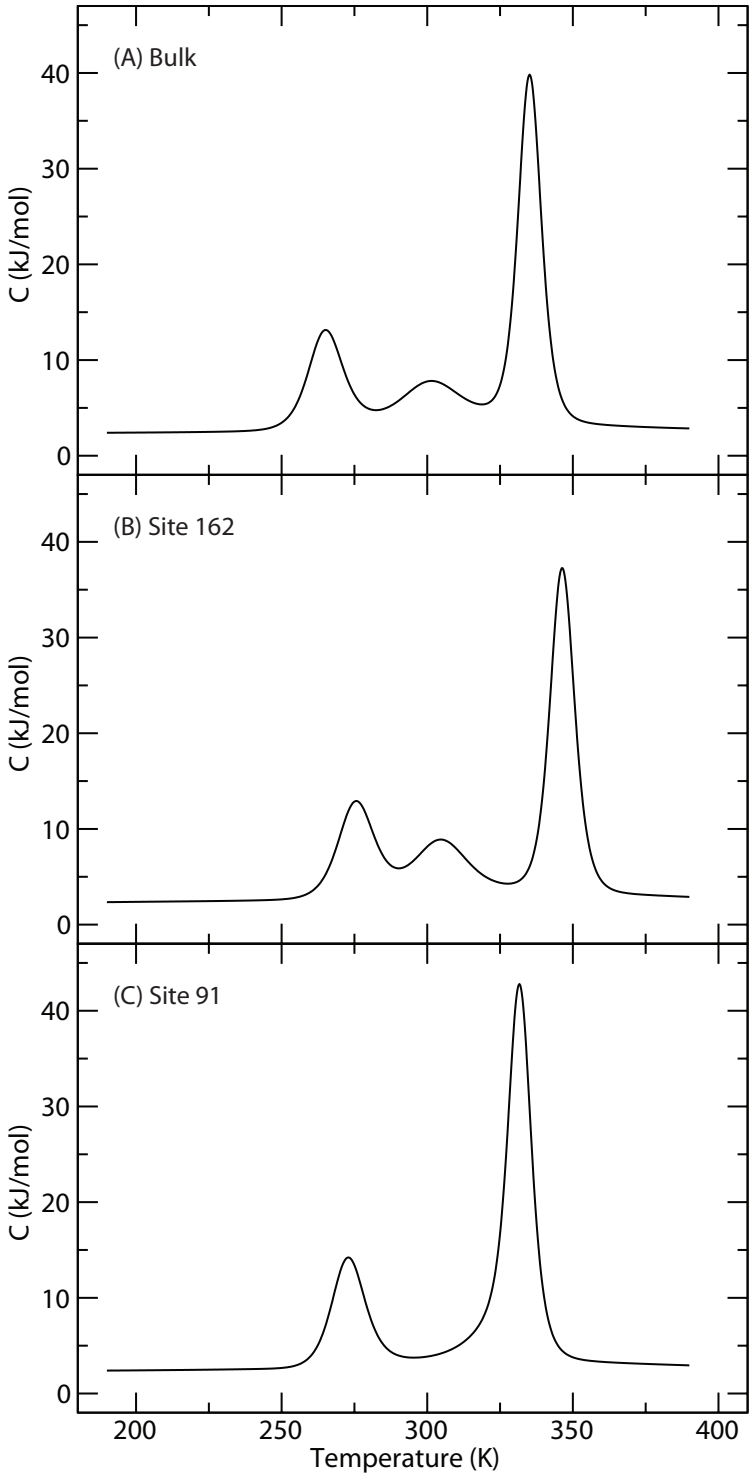


Figure 5.2: Heat capacity as a function of temperature for 7LZM in the bulk (Panel A) and tethered to the surface at sites 162 (Panel B) and 91 (Panel C).

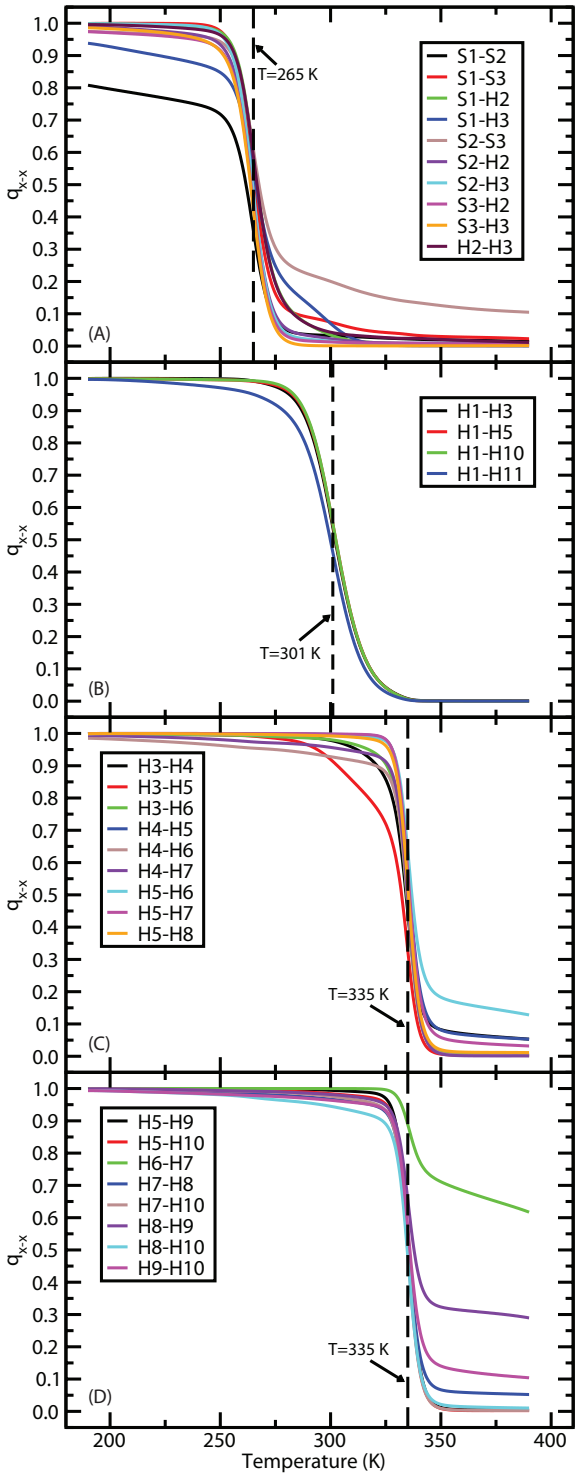


Figure 5.3: Folding/Unfolding transitions of individual tertiary structures as a function of temperature for 7LZM in the bulk. Transitions occurring at ≈ 265 K are found in Panel (A), those at ≈ 301 K in Panel (B), and those at ≈ 335 K in Panels (C) and (D).

residues forming the roll have limited mobility and remain close to their positions found in the native state. State 3-b shows the conformation of the molecule once Helix 1 breaks away from the bundle during the transition at 301 K. After this transition, the entire upper portion of the molecule assumes coil-like configurations. State 4-b is the unfolded state of the protein after the highest-temperature transition at 335 K. The tertiary, helical bundle formed by Helixes 4-11 comes apart and the individual helices themselves lose most of their secondary structure.

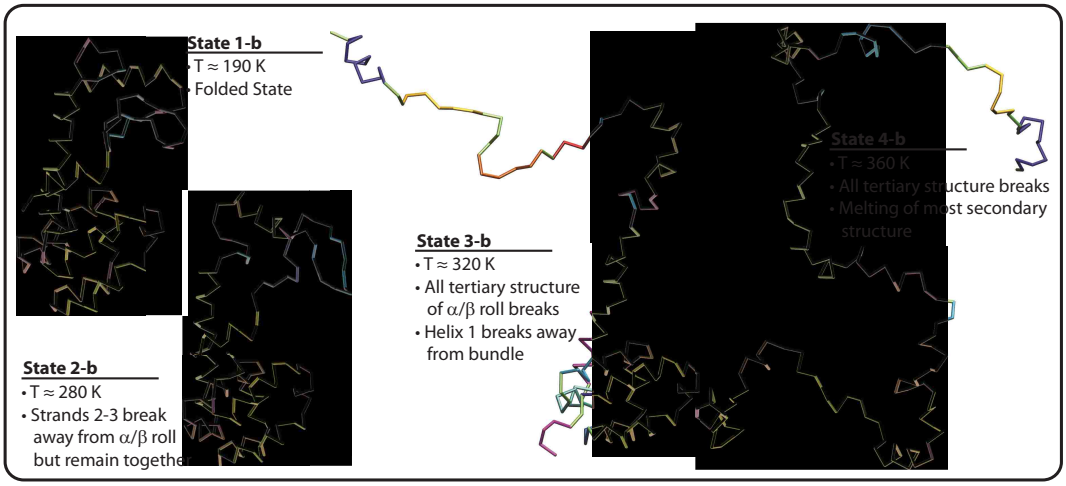


Figure 5.4: Conformations of 7LZM in the bulk at 190, 280, 320, and 360 K.

The size and shape of the peaks in the heat capacity are related to the amount of secondary and tertiary structure affected at the particular transition in question. During the 335 K transition, a large amount of secondary and tertiary structure melts which results in a high, sharp peak. The 301 K transition involves only the breaking away of Helix 1 from the rest of the bundle. As such, the heat capacity peak is small and broad. The 265 K transition, where the alpha/beta roll melts, involves more structure than the 301 K transition but less than the 335 K transition. As such, the shape and size of the peak lies in between that of the two extremes.

5.2.2 Folding on the Surface

As discussed above, the protein was tethered to the surface in each of the loop regions joining two different secondary structures (see Table 5.1). Each simulation was analyzed, but only two cases are discussed below for brevity. All but one of the tether sites resulted in heat capacity curves similar to that shown for site 162 (Panel B of Figure 5.2), three peaks with the most prominent occurring at the highest temperature. In some cases, as for site 162, the locations of the peaks were slightly higher compared the bulk indicating that the surface stabilized the protein. In other cases, the locations of the peaks were shifted to slightly lower temperatures than in the bulk indicating that the surface destabilized the protein. However, whether stabilized or destabilized, the structure of the protein in each state is the same as in the bulk (see Figure 5.4) for all surface cases with three peaks. Notice that the states shown in Figure 4 and 6 are typical structures in those temperatures, which show the common characters of the ensemble of states.

An interesting phenomenon is observed for tethering at site 91; the folding mechanism changes. As shown in Panel C of Figure 5.2, one peak is missing in the heat capacity curve for tethering at this site compared to the the bulk situation. The missing peak indicates that one of the stable intermediates along the folding pathway of 7LZM has been eliminated.

Explicit evidence to this fact is seen in Figure 5.5 which contains the plots of q_{x-x} as a function of temperature for tethering at site 91. The lowest-temperature transition, occurring at 273 K, is similar to that found in the bulk. Part of the alpha/beta roll formed by Helix 2 and Strands 1-4 comes apart, Helices 2 and 3 maintain their secondary structure, and Helix 1 maintains it secondary structure as well as its tertiary contacts with the bundle formed by Helices 4-11. In the next transition, all of the native contacts break. This is analogous to the highest-temperature transition found in the bulk where the resulting molecule has no tertiary structure and little secondary structure. The transition that is eliminated in site 91 compared to the bulk is the breaking away of Helix 1 from the rest of the bundle. In the bulk, this occurred at a much lower temperature than the melting of the rest of the bundle, but on the surface, Helix 1 and the bundle all melt at approximately the same temperature.

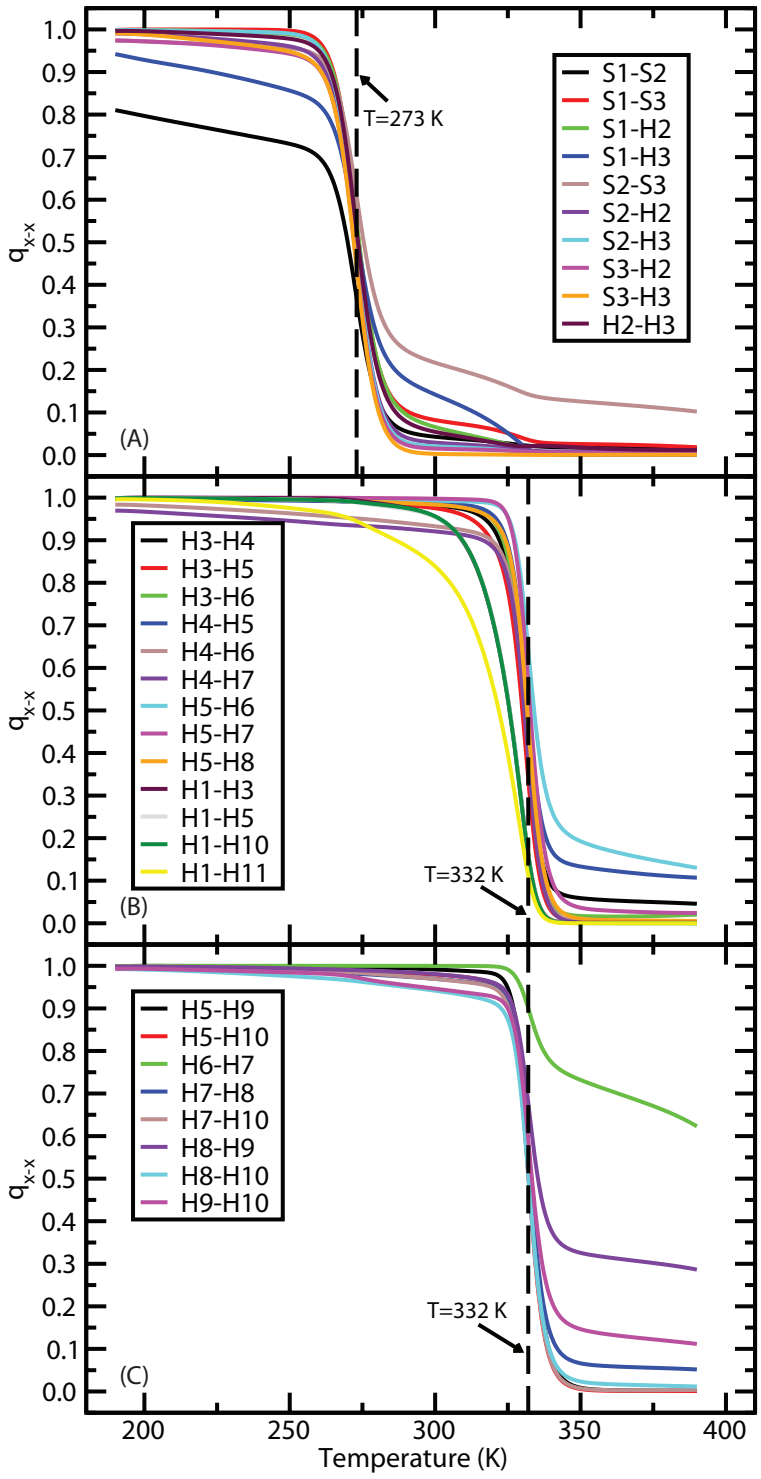


Figure 5.5: Folding/Unfolding transitions of individual tertiary structures as a function of temperature for 7LZM when tethered to the surface with residue 91. Transitions occurring at ≈ 273 K are found in Panel (A) and those at ≈ 332 K in Panels (B) and (C).

Figure 5.6 shows the conformations of 7LZM, when tethered at site 91, in each of the stable states found between the transitions. These are labeled, beginning at the low temperature, 1-s91, 2-s91, and 3-s91 where the "s91" signifies "site 91." State 1-s91 is the folded state of the protein found at low temperatures and corresponds to state 1-b in the bulk. State 2-s91 is the protein after the alpha/beta roll has melted and corresponds to state 2-b in the bulk where Helix 1 keeps contact with the helical bundle. State 3-s91 is the unfolded state of the protein found at high temperatures and corresponds to state 4-b in the bulk. When tethered to the surface at site 91, no state is found that is analogous to state 3-b.

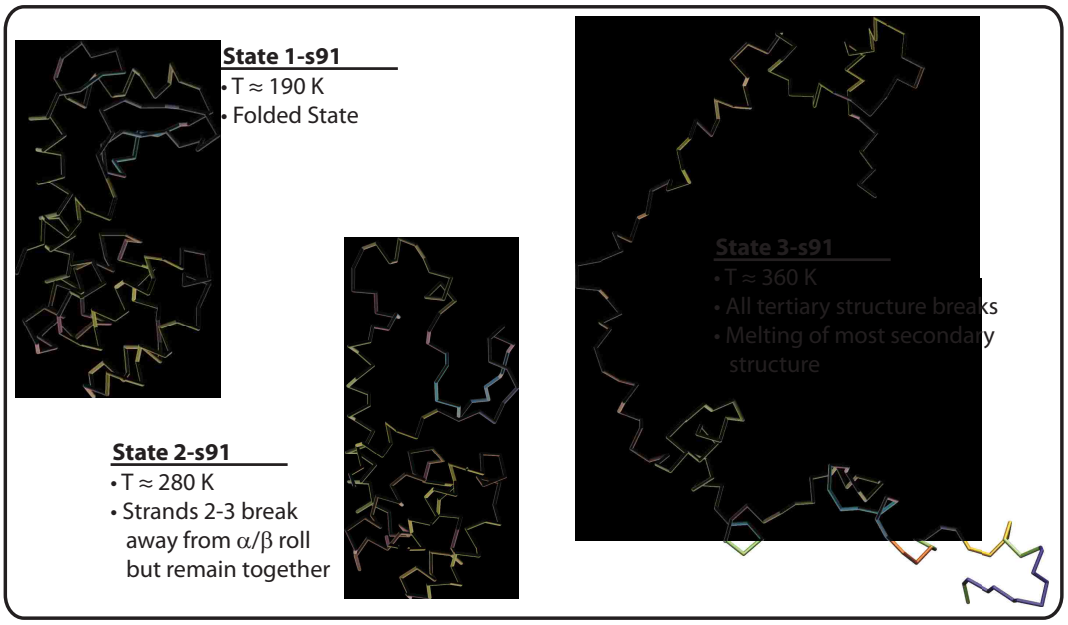


Figure 5.6: Conformations of 7LZM tethered to the surface at site 91 at temperatures of 190, 280, and 360 K.

5.3 Discussion

5.3.1 Protein Array Design

As explained previously, one of the major reasons for studying protein-surface interactions is optimization of protein arrays. Over the last several years, multiple techniques have been developed to successfully tether proteins to the surface at specific residues. For example, DNA and unnatural amino acids have been shown to be effective at site-specific tethering. [29, 33] However, the question as to the particular residue to tether in each case has not yet been answered.

The results of this study offer hope that simulation can help answer this question. In the bulk, 7LZM folds through a pathway with four states (see Panel A of Figure 5.2 and Figures 5.3 and 5.4). When tethered to the surface in the loop regions of the molecule, these same states are seen in the majority of the situations. However, for site 91, the folding pathway changes significantly affecting the two lower-temperature transitions. The middle transition is eliminated in its entirety and the lowest-temperature transition is muted. This suggests that tether site selection for multistate folders can be used as an optimization parameter when designing technologies involving protein-surface interactions.

To understand how this can occur, consider the relationship between the structure and the function of the molecule. 7LZM is a lysozyme–an enzyme which cleaves peptide bonds. The active site of this enzyme is found in the "cleft" formed between the two portions of the molecule. Panel A of Figure 5.7 shows a space-filled representation of 7LZM with the location of the active site in red. The orientation of the protein is the same as in Figure 6.8 where the top portion of the molecule is the alpha/beta roll formed by Helix 2 and Strands 1-4 and the bottom portion of the molecule is the helical bundle formed by Helices 1, 4-11. The location of residue 91 is shown in green and is located on the side opposite the active site cleft. For reference, Panel B of Figure 5.7 shows the molecule rotated approximately 200 degrees about the vertical axis compared to the orientation in Panel A. Panel C show how the molecule would be orientated when bound to the surface at site 91. Notice that the active site cleft is still accessible to the bulk phase.

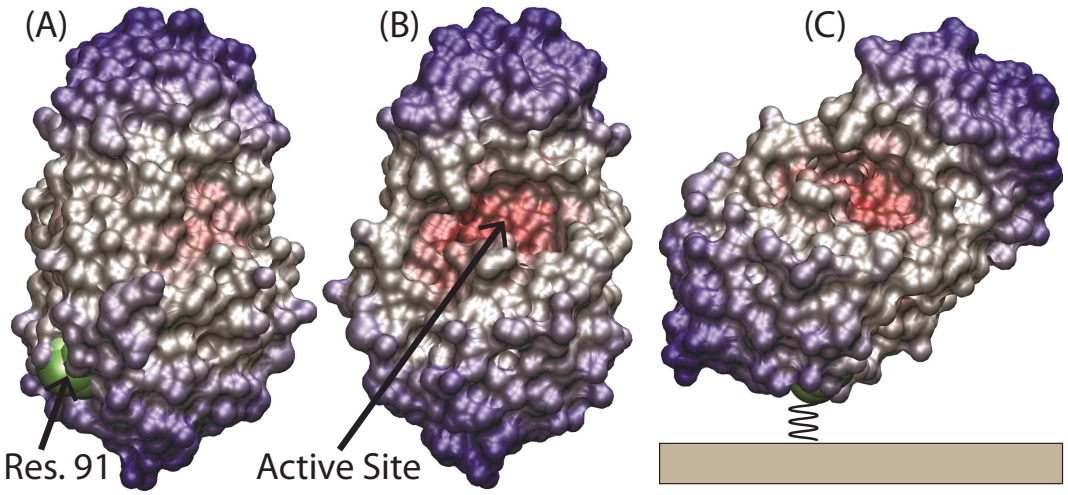


Figure 5.7: Relationship between reside 91 and the active site of 7LZM. Panel (A): Space-fill representation in the same orientation as in Figure 5.1; Panel (B): Depiction of the active site obtained by rotating the molecule approximately 200 degrees about the vertical axis compared to the orientation found in Panel (A); Panel (C): Position of the molecule when tethered at site 91.

Recall that tethering at site 91 made the protein behave more like a two-state folder. In effect, the surface constrained the protein in such a way that the partially-unfolded intermediate found in the bulk, where the alpha/beta roll and Helices 1 and 3 were melted, was no longer present. The consequence is that the protein remains in a more folded configuration a greater amount of time. In the bulk, the entire top portion of the molecule melts at the transition occurring at 301 K. However, when tethered to the surface at site 91, the top portion of the molecule does not unfold to such a degree until approximately 332 K when the entire molecule unfolds. For 7LZM to function, both portions of the molecule must be folded so that the active-site cleft is formed. By eliminating the transition occurring at 301 K, the surface greatly enhances the ability of the protein to sample conformations closer to the native state which can be expected to lead to enhanced enzymatic activity. It is true that some denaturation of the protein happens at the lower temperature transition, which would decrease the activity compared to the perfectly-folded protein, but compared with the bulk case the situation on the surface will lead to more native-like behavior. In short, tethering at site 91 is predicted by the model to stabilize the active state of the protein *and*

keep the catalytic site available to the bulk–two conditions which are essential for proper array function.

5.3.2 Applicability of the Results

This study investigated the behavior of only one multistate protein, so more studies are needed to fully investigate the universality of the reported phenomena. Due to the diversity of structure in proteins, it is likely that other proteins will display different and unique, protein-specific behavior. However, the methods outlined above to study 7LZM are very versatile, so investigation of many proteins of interest is possible.

The results presented above are also limited by the coarse-grained representations of the protein and the surface. For example, the chemical composition of the surface was not modeled explicitly, so the results are only valid for weakly-interacting surfaces. Though this is the type most frequently encountered in protein-surface applications, different behavior is likely to be observed for highly attractive or repulsive surfaces. Despite this, the approach presented above has been extensively used before to study protein-surface interactions [86, 83, 84, 88, 85, 121] and protein folding [58, 59, 60, 61, 97], so it is reasonable to assume that the *trends* observed in the results represent realistic behavior. The numerical values reported are not expected to correspond exactly to those that would be observed experimentally, due to the reduction in the degrees of freedom resulting from the coarse-graining of the system and the use of an implicit solvent model, but the *comparisons* between the bulk and surface behavior are valid as the same model was used in both instances.

To ensure statistical significance, multiple independent simulations for each system were performed where each was initialized with a different seed from the random number generator. The results from these simulations were determined to be reproducible and consistent. This suggests that the features of each plot, as discussed above, adequately represent the behavior of the system within the accuracy of the model.

5.4 Summary

The results in this chapter showed, for the first time, how surfaces can change the folding mechanism of a protein that folds through multiple states. In the bulk, 7LZM folds

through a four-state mechanism. When tethered to the surface, this same mechanism is seen for most of the tether locations; however, when tethered at residue 91 the mechanism changes. One of the intermediates is eliminated creating a three-state mechanism. It was found that this change stabilized the active site of the protein in a way that would increase function at higher temperatures. Moreover, it was shown that tethering at site 91 presented the active site away from the surface so that substrates would have access to the tethered enzyme from the bulk phase. The results offer hope that rational design of protein arrays is possible once a molecular-level understanding of the relevant phenomena is obtained.

Chapter 6

New coarse-grained Model for Protein Surface-interaction

Despite the fundamental understanding gained from the coarse-grained simulations described in the previous chapters as well as those of prior studies [83, 84, 86], the technique to date has suffered from a major limitation: the surface has been modeled in only a rudimentary way even by coarse graining standards. Specifically, the vast majority of all coarse-grained simulations have used either a hard surface or a short-range repulsive surface. This approach captures what is arguably the most important feature of protein-surface interactions, which is the reduced conformational entropy caused by a decrease in the phase space available to the protein, but it has three implications. First, the model does not take into account the solvent exclusion (or desolvation) which occurs between the peptide and the surface as adsorption occurs. Second, the surface affects all residues equally regardless of the chemical specificity of the residues comprising the proteins or the composition of the surface. Third, attractive surfaces have not been adequately studied and in the few cases where such have been investigated, they continue to suffer from the first two weakness. In essence, these weakness means that the enthalpic contribution to the free energy from the interaction of the protein with the surface has not been adequately described to date.

In regard to the first weakness mentioned above, experiments have shown that a barrier exists to adsorption of proteins to surfaces. When a protein is adsorbing to the surface, solvent molecules between the protein and the surface need to be excluded to achieve the adsorption. The energy required for this solvent exclusion (or desolvation) causes a small energy barrier as the protein come close to the surface but before it completely adsorbs..

In regard to the second and third weakness described above, recent experiments have shown that peptides are not repelled by relevant surfaces regardless of the composition of the surface or the identity of the amino acids involved [73]. Rather, all peptide-surface

interactions are attractive with the degree of attraction varying according to the chemical nature of the interaction. As mentioned above, research into attractive surfaces is limited, but examples do exist. Knotts *et al.* [88] showed that attractive surfaces cause entropic frustration to occur in the folding mechanism of proteins tethered to attractive surfaces and Shea *et al.* [122] recently used an attractive surface model to study peptides aggregate morphology. But the conclusions drawn from these studies are limited because they do not take into account chemical specificity and are not parameterized against experimental data.

The examples listed above demonstrate that a new model is needed to make further advances in the simulation and understanding of protein-surface interactions. This chapter describes efforts to create a new coarse-grained model for protein-surface interactions which is parameterized against experimental data, takes into account the chemical composition of the protein and the surface, includes the desolvation penalty which is present as proteins adsorb onto surfaces, and replicates experimental adsorption free energies for proteins not used in the parameterization. The remainder of the chapter is constructed as follows. First, the key features for the new model will be discussed and the exact mathematical description will be outlined. This will be followed by a description of the procedure followed to parameterize the model including a depiction of the experimental data used to this end. Next, the results of the parameterization will be presented, and this will be followed by an explanation of how the model was validated against experimental data not used in the parameterization. The chapter will end with a discussion of the implications of the work.

6.1 Model

The model proposed here was developed to have several key features. These are:

1) The model should be coarse-grained to make investigation into protein stability computationally possible.

2) The model should be concise and easy to understand.

3) The model should capture characteristics such as solvent exclusion, chemically-specific adsorption, and hydrophobic effects.

4) The model should be parameterized against experimental data for adsorption free energies.

5) The model should reproduce experimental adsorption free energies for large, biologically-relevant proteins.

6) The model should provide realistic protein folding mechanisms on surfaces.

As producing realistic folding mechanisms is a key feature needed in the new model, it was decided to build the model based on the principles explained by Karanicolas and Brooks [58, 59, 60, 61] whose existing model has met with considerable success in bulk simulations. Specifically, all of the intra-protein interactions are the same as those used previously to model protein folding. New terms were added to the model to account for the protein-surface interactions. These terms were specifically designed to reproduce two key features found in both experiments and atomistic simulations. The first is the presence of a minimum in the free energy of adsorption curve between a protein and a surface at short distances. The second is the presence of an energy barrier, caused by solvent exclusion effects,between this minimum and long distances where the protein does not feel the presence of the surface.

The proposed surface potential is found in Equation 6.1. The first three terms in the force field equation capture the repulsion energy of proteins close to the surface (represented by the 9th-power term), the adsorption well (represented by the 7th-power term), and the energy barrier (represented the cubic term with parameter θ_3). According to Wei and Latour's previous analysis [55], the hydrophobicity of a surface affects protein/surface interaction mainly by changing the ability of the incoming molecule to induce exclusion of the solvent molecules between the surface and the protein.Therefore, a second cubic term is included to specify the hydrophobic effect for different surfaces with parameters θ_s and χ_s. For a similar reason, the hydrophobic effect of different amino acids also has a cubic dependence on the distance with parameters θ_p and χ_p.

$$V_{surface} = \sum_{i}^{N} \left\{ \pi\rho\sigma^3\epsilon_{sur} \left[\theta_1 \left(\frac{\sigma_{sur}}{z_{is}} \right)^9 - \theta_2 \left(\frac{\sigma_{sur}}{z_{is}} \right)^7 + \theta_3 \left(\frac{\sigma_{sur}}{z_{is}} \right)^3 - (\theta_s\chi_s + \theta_p\chi_p) \left(\frac{\sigma_{sur}}{z_{is}} \right)^3 \right] \right\} \tag{6.1}$$

Also in Equation 6.1, z_{is} is the distance between site i and the surface, σ_{sur} is a residue specific van der Waals parameter, and ϵ_{sur} is parameter controlling the overall strength of the interaction. θ_1 and θ_3 are the two parameters controlling the repulsive part of the potential, θ_2 is the parameter governing the depth of the attractive well, χ_s values are hydropathy indices assigned to distinguish different surfaces, and χ_p values are hydropathy indices of amino acids. χ_s and χ_p are experimental values which characterize the hydrophobic/hydrophilic values of chemical moieties and are not fitting parameters. The χ_p values for each amino acid are found in Table 6.1.

Table 6.1: Hydropathy index of guest amino acids χ_p.

G	A	P	V	L	I	M	F	Y	W
−0.4	1.8	1.6	4.2	3.8	4.5	1.9	2.8	−1.3	−0.9
S	T	C	N	Q	K	H	R	D	E
−0.8	−0.7	2.5	−3.5	−3.5	−3.9	−3.2	−4.5	−3.5	−3.5

Values for each θ are obtained by fitting simulation results to match experimental data in the manner described below.

6.2 Method

6.2.1 Systems

Different simulation systems are used in this research to parameterize the surface model and validate the transferability of the parameters to other systems. Below is the description of each and an appropriate designation.

1) Eleven host-guest style peptides and three types of surfaces with different hydrophobicities are used to parameterize the new surface model. These correspond to peptides used in the experimental work of Latour *et al.* who have measured the change in free energy upon adsorption for several such peptides. A host-guest peptide refers to a peptide with the sequence of TGTG-X-GTGT (where G and T are glycine and threonine amino acid residues respectively and X represents a different "guest" amino acid residue) [73]. The eleven guest peptides used in this research are (X = D, F, G, K, L, N, R, S, T, V, and W). For the surface, rather than parameterize specific chemical moieties that could represent different surfaces, it was decided to create three different classes: hydrophobic, hydrophilic, and moderately hydrophilic. Experiments have shown that most surfaces, regardless of the exact chemical composition, have interactions that can be grouped into one of these classifications [73].

2) Three proteins: hen egg-white lysozyme protein (PDB ID: 2LYZ), myoglobin (PDB ID: 1MBN), and cytochrome c (PDB ID: 1HRC), are used to validate the model parameters. These proteins are chosen because the melting temperature, folding mechanism, and adsorption free energy to different surfaces of these proteins are available from several experimental and simulation studies [1, 120, 48, 123, 124, 125, 126]. Note that no reparameterization of the new model is done with these proteins. All parameterization uses only the host-guest peptides. 2LYZ, 1MBN, and 1HRC are only used to test the transferability of the parameters to real systems.

6.2.2 Protein Model

As mentioned above, the Gō-like model of Karanocolis and Brooks [58, 59, 60, 61] is used to represent intramolecular interactions. This includes both bonded and non-bonded, intra-protein interactions. In this formalism, each residue is represented by one site placed at the C_α position of the residue. This model is used because it keeps the number of interaction sites per amino acid to a minimum but uses different energy scales to describe hydrogen bonding between side chains and sequence-dependent dihedral potentials. Less-realistic models employ fewer energy scales and set dihedral parameters based upon the PDB structure and not sequence. As such, the resulting energy surface mimics that of real

proteins, and the model has been shown to give good agreement with experimental folding studies. [58, 59, 60, 61, 97]

For the eleven host-guest peptides, initial structures are generated from atomistic simulation with the CHARMM simulation program [127, 128, 129] with implicit solvent. To ensure adequate sampling, the system was heated, cooled, and equilibrated to find the most-favorable conformation of the peptides at 298 K. The Gō-like model input files are obtained from the MMTSB website `http://www.mmtsb.org` using coordinates from the equilibrated, all-atom structures, but native contacts are removed to inhibit the formation of unrealistic secondary structures and keep the host-guest peptides in a coiled state as expected form experiments. [73]

6.2.3 Simulation Temperature

One of the limitations of coarse-grained simulation is that the removal of degrees of freedom alters the energy of the system so that the temperature scale becomes relative. This means, for example, that the melting temperature of the protein can be scaled to any value desired by arbitrarily increasing the strength of the intramolecular interactions. This ability to scale the melting temperature requires careful selection of the simulation temperature used in parameterization. The goal was to make this value correspond to an experimental temperature of 298 K. This value is widely used in protein-related experiments because most biologically-relevant proteins are stable at this temperature. Previous simulation work has shown the melting temperature of lysozyme, modeled according to Karanocolis and Brooks, is about 265 K [130]. The experimentally-measured value is about 348 K for a difference of approximately 83 K. Also, previous simulation research has shown that several other proteins also melt at 298 K [121, 130].

To be consistent with the experiments, simulations must be performed at a temperature low enough for the proteins of interest to maintain their native structures. A temperature of 215 K is used to accomplish this purpose for the guest-host peptides. This value was obtained based on lysozyme data. The value of 215 K was obtained by subtracting 83 from 298. Simulations have shown that the exact value of the temperature used to capture protein adsorption is not important as long as it is far enough below the melting

temperature of the protein. Moreover, calculations of the over 30 proteins described in this dissertation and modeled using the method of Karanocolis and Brooks have all shown to be stable at 215 K [57, 58, 59, 60].

6.2.4 Parameterization

Experimentally, the nature of the surface can be controlled by changing the chemical moieties presented on the surface. To capture this in the model, the parameterization of the $\theta's$ in Equation 6.1 was done against data from Latour *et al.* [73] who measured adsorption potentials for 108 different protein-surface combinations. This was done by matching experimental adsorption free energies with those calculated from simulation. The free energy of adsorption ΔG_{ads} is defined as the change in free energy that occurs by moving a protein from a position far away from the surface to a favorable position close to the surface. Experimental values for ΔG_{ads} were determined by Latour *et al.* [55] The simulation values were obtained using umbrella sampling [131] to induce protein adsorption to the surface. Using a potential of the form

$$V_{umb} = k_u \left(\xi - \xi_h\right)^2 \tag{6.2}$$

where k_u is a force constant, ξ_h is the desired distance between the center of mass of a protein and the surface for a particular umbrella, and ξ is the instantaneous distance from the surface. The umbrella's ranged from $\xi_h = 1$ to 100 Å in increments of 0.25 Å. 100 Å is far enough away from the surface that the protein-surface interaction has decayed to zero.

The canonical ensemble is used, and the temperature is maintained at 215 K by Nosé-Hoover-Chain integration method with 3 thermostates of mass 10^{-26} kgÅ^2 [102, 103, 104]. Each simulation consists of 10 million steps of equilibrium and 30 million steps of production with a step size of 1 fs. The general parameterization scheme consisted of an umbrella simulation of one of the host-proteins, analysis of the simulation data to obtain ΔG_{ads}, comparison of this value to experimental data, ad hoc selection of new parameters, and repeating until the agreement between simulation and experiment was satisfactory.

6.2.5 Characterization and Validation

Once the model was parameterized using the iterative method just described, validation was done by simulating larger, more-relevant proteins for which experimental adsorption data on different types of surfaces is available. For the validation simulations– which were done for lysozyme (PDB ID: 2LYZ), myoglobin (PDB ID: 1MBN), and cytochrom c (PDB ID: 1HRC)– the model parameters were not changed to test the transferability of the model. Structures of these three proteins are shown in Figure 6.1. For each protein, umbrella sampling was performed to calculate the adsorption free energy of the protein on the surface and the protein-surface distance ranged from 1 to 450 Å with an interval size of 0.5 Å. At 450 Å the protein is far enough from the surface that the interaction decays to zero. Simulations were done in the canonical ensemble, and the temperature was maintained at 215 K by the Nosé-Hoover-Chain integration method [102, 103, 104]. Each simulation consisted of 10 million steps of equilibrium and 30 million steps of production with a step size of 1 fs. The surface hydrophobic index was set equal to 4.5 to mimic the hydrophobic surfaces used in experiments [1, 48].

6.3 Result and Discussion

6.3.1 Parameterization

To demonstrate the functional form of the qualitative features seen in experimental adsorption curves, Figure 6.2 compares the potential of mean force generated using a purely repulsive surface model (as done in previous research) with the new surface model prior to parameterization ($\theta_{1,2,3} = 1$ and $\theta_{s,p} = 0$). The new model produces features in the experiment not seen using the previous repulsive model. These include the attractive well and the small energy barrier found before dropping into the energy well.

The first round of parameterization used threonine as the guest residue in the host-guest peptide. The purpose of this round was to produce a rough estimate of the model without regard to specific residues. Threonine was chosen because all-atom simulations had been performed previously for this host-guest peptide and the hydropathy index of threonine

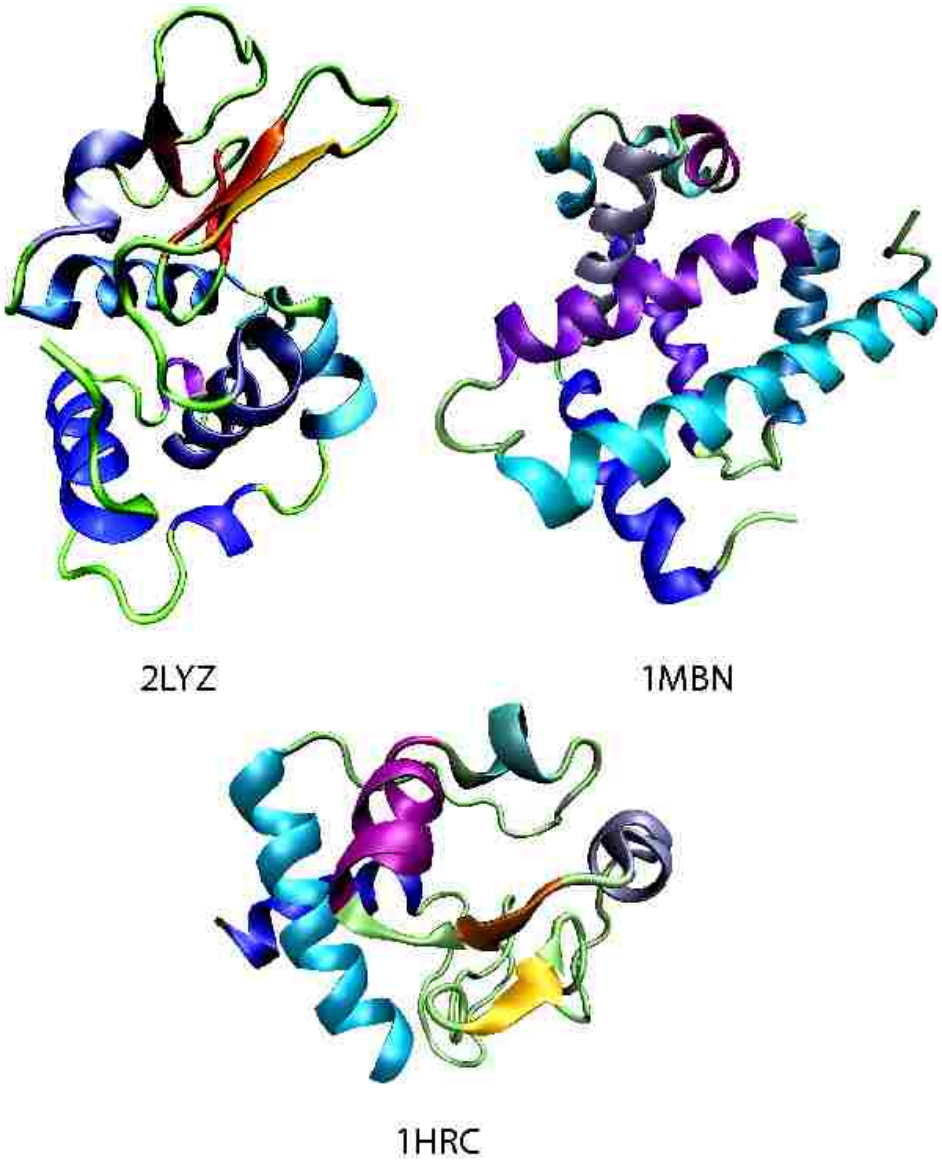
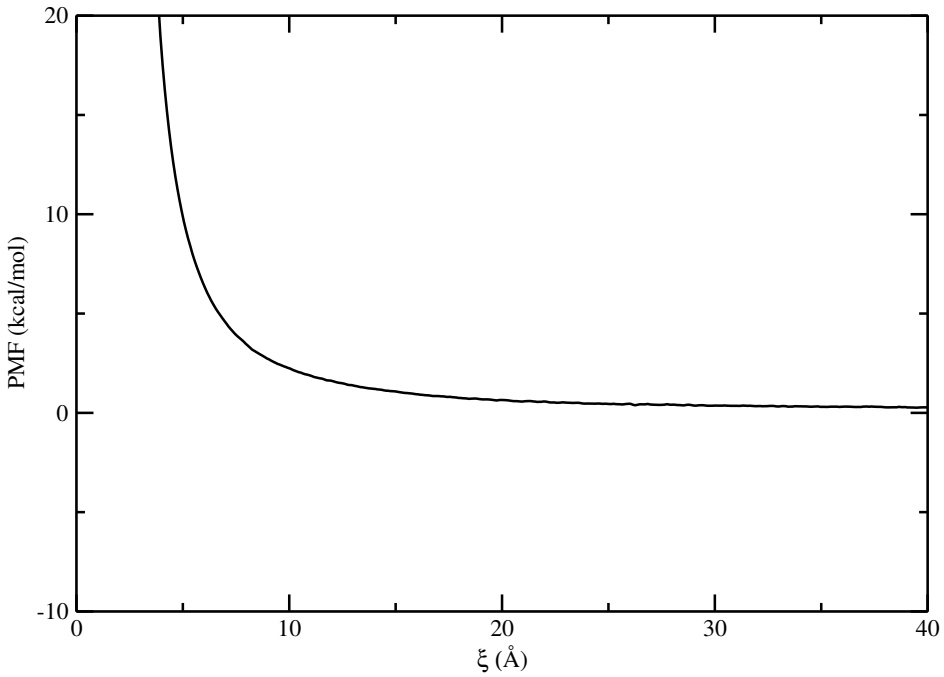


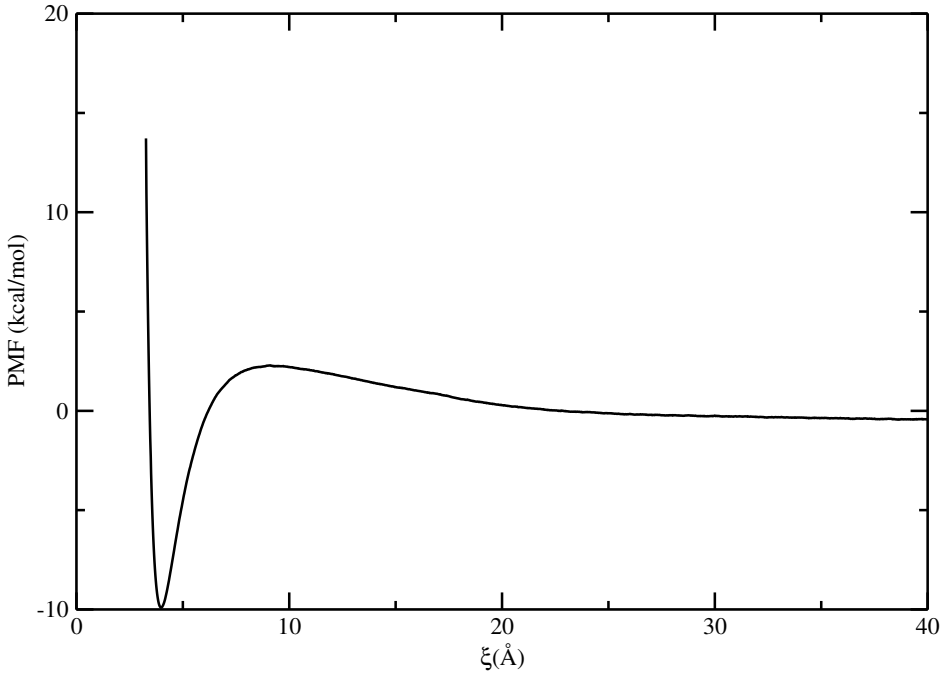
Figure 6.1: Structure of the three large proteins used for model validation.

(−0.7) is close to the middle of the range for all amino acids (−4.5 to 4.5). Specifically TGTG-T-GTGT was used to obtain the three parameters ($\theta_{1,2,3}$) in Equation 6.1.

The resulting potential of mean force curve of protein adsorption on surface is found in Figure 6.3 (with $\theta_s = 0$ and $\theta_p = 0$). The red dotted line data were generated by Latour



(A) Potential of mean force curve of the previous, purely repulsive surface model.



(B) Potential of mean force curve of the new surface model.

Figure 6.2: Potential of mean force curve for a purely-repulsive model (Panel A) and the new model (Panel B).

et al. [132] using all atom simulation. The new, coarse-grained model reproduces the data from Latour *et al.* very well. Both show a double energy well and small desolvation barriers at similar values of ξ. The depth of the energy well is slightly smaller than the atomistic data, and the atomic overlap repulsion occurs at a slightly different location, but these are minor issues. As shown in Table 6.2, the ΔG_{ads} from the coarse-grained simulation with X = T is -2.70 kcal/mol and the experimental value is -2.76 ± 0.28 kcal/mol. Moreover, the barrier heights are similar to those from Latour *et al.*. Since it is a preliminary estimation of model parameters, detail statistical analysis with error bars is not performed.

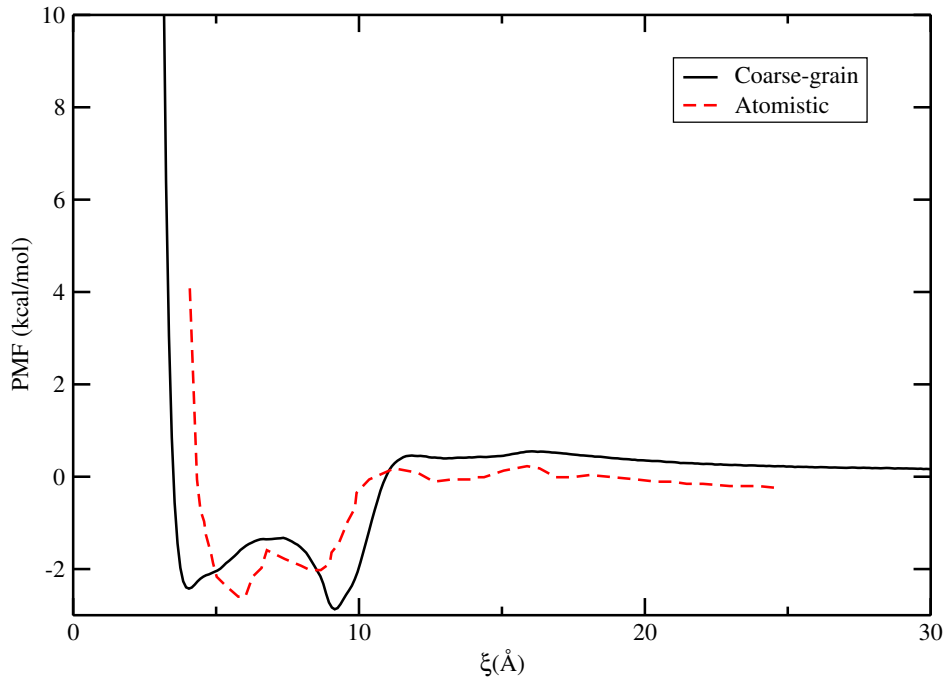


Figure 6.3: Potential of mean force curves from atomistic simulation and from the new surface model .

As just described, parameterizing θ_1, θ_2, and θ_3, which gives just the overall shape of the potential and does not take into account exact chemistries, gave remarkably agreement between the new model and experimental ΔG_{ads} values. Using the set of parameters obtained with X = T, ΔG_{ads} values were obtained for proteins with guest residues X = D, K, and V

on the SAM-CH_3 surface. As shown in Table 6.2 the simulations produce adsorption free energies in very good agreement with the experimental values for three of the four peptides (X = D, K, and T) on the SAM-CH_3 surface, but the ΔG_{ads} of the peptide with X = V was slightly off. To correct this problem, the next round of parameterization included the chemistry of the surface.

Table 6.2: ΔG_{ads} for X = D, K, T, and V on the SAM-CH_3 surface parameterizing only θ_1, θ_2, and θ_3 but leaving θ_s and θ_p equal to zero.

	ΔG_{ads} kcal/mol			
	D	K	T	V
Experimental	-3.54 ± 0.60	-3.34 ± 0.39	-2.76 ± 0.28	-4.40 ± 0.31
Simulation	-3.21	-3.54	-2.70	-3.00
	Hydrophobic	Hydrophobic	Moderately-hydrophilic	Hydrophilic

As mentioned previously, surface chemistry is included in the model with the term containing θ_s. The exact chemistry is controlled by χ_s, the hydropathy index of the surface. Rather than parameterize every possible chemical group that could be used to create a surface, it was decided to parameterize "types" of surface. Experimental data has shown that there are basically three types of surfaces involved in protein-surface interactions: hydrophobic, hydrophilic, and moderately hydrophilic. The exact chemical groups used to create such surfaces is not important. Rather, the actual type of surface created is the dominant factor. Figure 6.4 shows three potential of mean force curves used to parameterize θ_s and θ_p. These correspond to simulation of a host-guest peptide with X = Asp on the three different types of surfaces. The χ_s values used to describe each type of surface are found in Table 6.4. Notice that the hydrophobic surface provides the deepest adsorption well and the flattest energy barrier, while the hydrophilic surface provides the shallowest adsorption well and highest energy barrier. These trends, though counter-intuitive, are consistent with results from Latour's experimental research. It might be expected that Asp, being a hydrophilic residue, would be most attractive to a hydrophilic surface. However, the experiments show that the opposite is true. The new model captures this unexpected behavior both qualitatively and

quantitatively as the ΔG_{ads} produced in the simulation is very close to the experimental values. (The exact values will be described shortly in the discussion about Table 6.5).

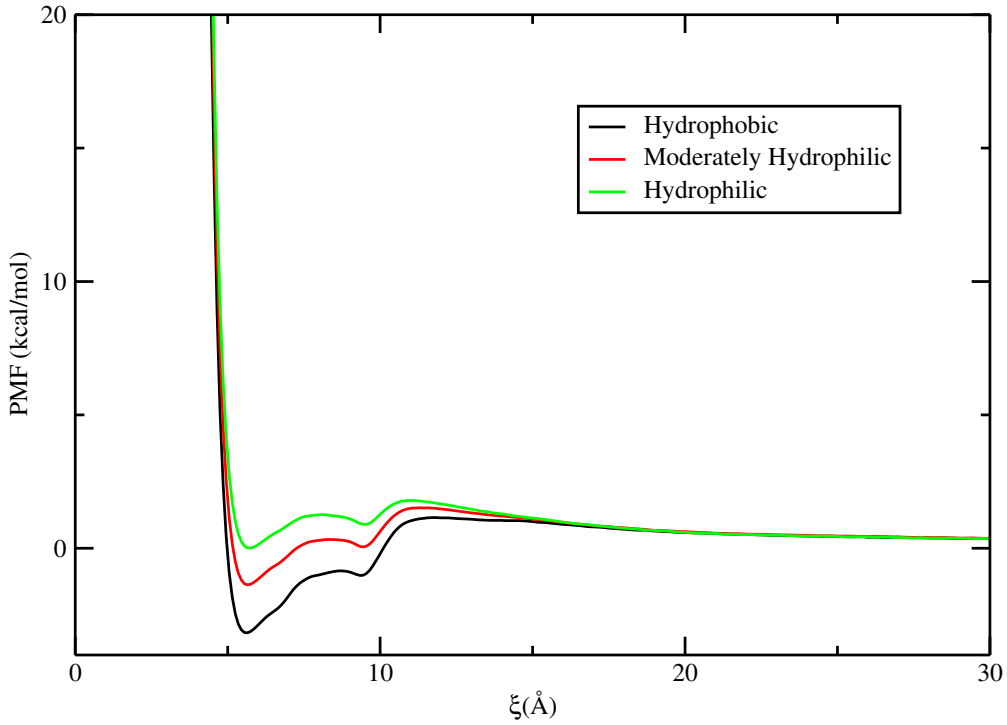


Figure 6.4: Potential of mean force curves with different surfaces.

The set of parameters used to produce the results found in Figure 6.4 are listed in Table 6.3. As mentioned previously, the χ values for the surface are found in Table 6.4 and those for the residues are found in Table 6.5. To test the transferability of the model, several guest-host systems were simulated. These are summarized in Table 6.5. The free energies of adsorption from both the new model and experiment are found in the table. The rows, showing different amino acids, are groups according to the residue type with polar residues at the top followed by non-polar residues and charged residues. The columns, represent different types of surfaces with the hydrophobic surface on the left followed by moderately hydrophilic and hydrophilic surfaces.

Table 6.3: Parameters for the surface model.

θ_1	θ_2	θ_3	θ_s	θ_p
0.2340	0.4936	0.1333	0.0067	0.0333

Table 6.4: χ_s for different surfaces.

Hydrophobic	Moderately-Hydrophilic	Hydrophilic
4.5	1.5	-1.0

Notice that the results are in good agreement with experiment for proteins with polar guest residues on all three types of surfaces. In addition, good agreement with experiment is found for all peptides on hydrophobic surfaces. Small deviations can be noticed for peptides with nonpolar guest residues on the hydrophilic surface. However, overall the agreement is remarkable considering that the fitting is only done with the five θ parameters and all the chemistry is contained in the experimental χ values.

Table 6.5: Comparison of adsorption free energy ΔG_{ads} (kcal/mol) between the simulation with the surface model and experiments for Host-Guest proteins. [73]

	Hydrophobic		Moderately-Hydrophilic		Hydrophilic	
-X-	Experiment	Simulation	Experiment	Simulation	Experiment	Simulation
			Nonpolar Guest Residues			
-L-	-3.87 ± 0.69	-4.17	-1.04 ± 0.30	-2.81	-0.40 ± 0.28	-1.60
-F-	-4.16 ± 0.16	-4.29	-2.44 ± 0.40	-2.89	-0.30 ± 0.13	-1.62
-V-	-4.40 ± 0.31	-3.82	-0.16 ± 0.10	-2.66	-0.26 ± 0.06	-1.37
-W-	-3.89 ± 0.34	-3.40	-1.94 ± 0.45	-1.80	-1.72 ± 0.33	-0.62
			Polar Guest Residues			
-T-	-2.76 ± 0.28	-3.21	-0.16 ± 0.09	-1.74	-0.28 ± 0.15	-0.23
-G-	-3.40 ± 0.39	-3.67	-1.86 ± 0.20	-2.15	-0.30 ± 0.20	-0.64
-S-	-2.75 ± 0.23	-2.91	-1.49 ± 0.47	-1.24	-0.34 ± 0.11	-0.25
-N-	-4.33 ± 0.62	-3.95	-1.64 ± 0.23	-2.30	-0.59 ± 0.11	-0.61
			Charged Guest residues			
-R-	-4.15 ± 0.55	-3.99	-1.60 ± 0.80	-2.64	-0.20 ± 0.10	-1.48
-K-	-3.34 ± 0.39	-2.74	-0.12 ± 0.07	-0.87	-0.19 ± 0.07	$+0.62$
-D-	-3.54 ± 0.60	-3.04	-1.93 ± 0.52	-1.35	-0.44 ± 0.14	-0.05

6.3.2 Validation

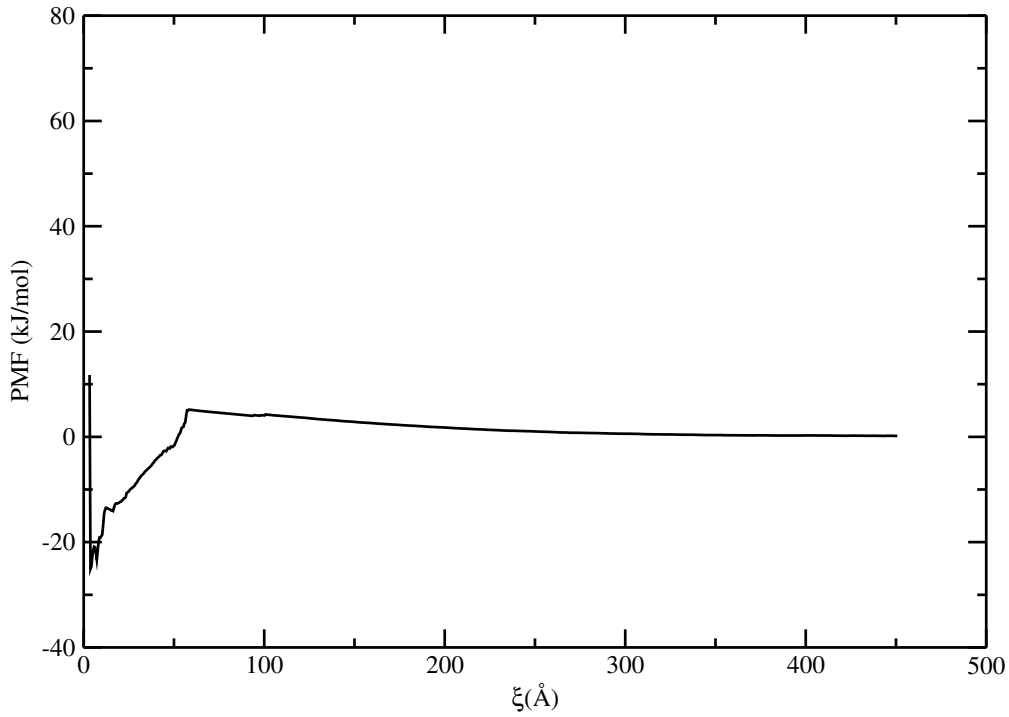


Figure 6.5: Potential of mean force curve for adsorption of lysozyme onto a hydrophobic surface.

The parameters used to generate the data in Table 6.3 were optimized against experimental data for the host-guest proteins. To test the transferability of the parameters, especially for real, scientifically-relevant proteins, simulations were done to calculate the ΔG_{ads} for three large proteins: lysozyme, myoglobin, and cytochrome C. No new reparameterization was done in these simulations. The parameters found in Table 6.1 were used unaltered along with a hydrophobic surface ($\chi_s = 4.5$). A hydrophobic surface was used because experimental data is available for the proteins of interest on such a surface. The potential of mean force curves for adsorption of each peptide onto the surface are found in Figures 6.5, 6.6, and 6.7 for lysozyme, myoglobin, and cytochrome C respectively.

The adsorption free energy calculated from the lysozyme curve is −25.1 kJ/mol. The adsorption free energy of lysozyme onto hydrophobic surfaces has been measure using multiple techniques by various groups. One group used a PEG-600 surface and found that the adsorption energy ranged from −23.1 to −27.8 kJ/mol [133] depending upon ionic density. Other groups have reported values of −37.3 kJ/mol [1] and −19.5 kJ/mol [48]. Thus, the experimental data place the value between −19.5 kJ/mol and −37.3 kJ/mol. The simulation with the new model yielded a value of −25.1 kJ/mol which falls in the range set by experimental results.

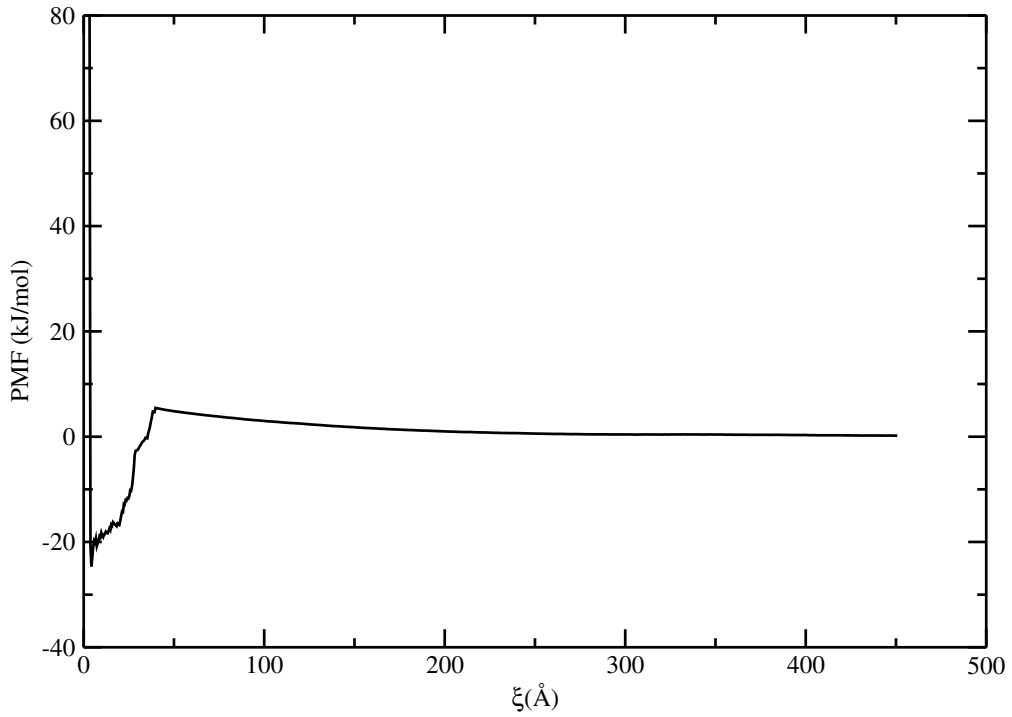


Figure 6.6: Potential of mean force curve for adsorption of myoglobin onto a hydrophobic surface.

Experimentally measured adsorption free energy of myoglobin from different groups are also varied. One group measured a value of −19.0 kJ/mol on hydrophobic surfaces. [48] Others place the value at −27.6 kJ/mol (transformed from adsorption constant K of Lang-

muir isotherm) on a biomimetic hydroxyapatite surface [123] and −32.0 kJ/mol (transformed from adsorption constant K of Langmuir isotherm) on hydrophobic surfaces. [124] Figure 6.6 shows that the new model predicts a value of −24.6 kJ/mol which is within the range of the experimentally-measured values.

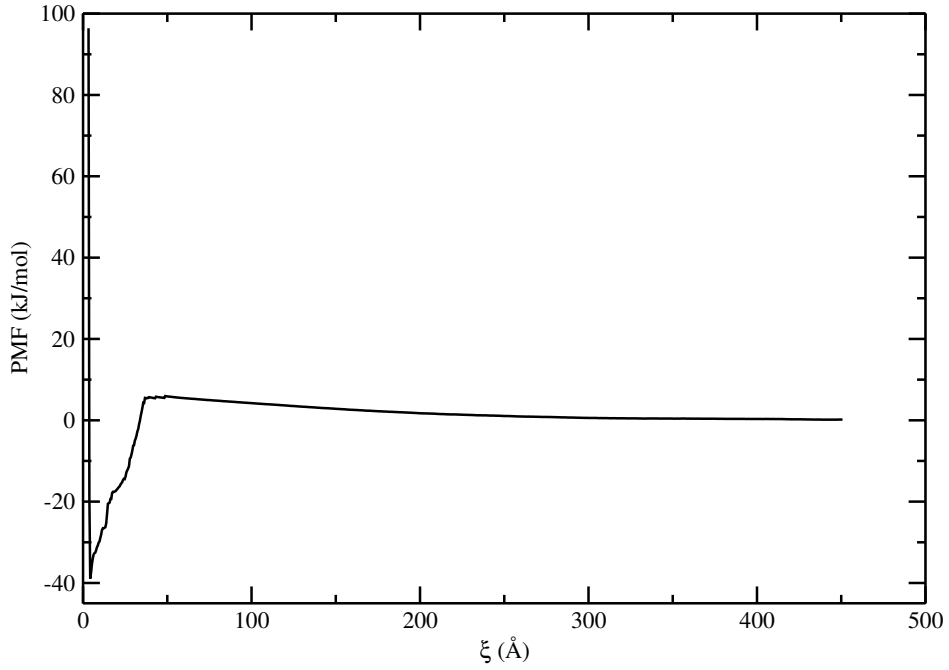


Figure 6.7: Potential of mean force curve for adsorption of cytochrom c onto a hydrophobic surface.

Finally, simulation of cytochrome C on a hydrophobic surface gives a ΔG_{ads} value of −39.1 kJ/mol (See Figure 6.7). Experiments place the value between −36.8 and −49.4 kJ/mol. [125, 126] Thus, the simulation value again lies within the experimentally-determined range.

The results are summarized in Table 6.6 which shows ΔG_{ads} obtained from simulation and experiment for all three proteins. These data demonstrate that the new model is in remarkable agreement with experiment. The model not only predicts the qualitative trend

(cytochrome C is most strongly attractive to the surface in both simulation and experiment) but it also produces quantitative agreement in each case.

Table 6.6: Comparison of adsorption free energy, ΔG_{ads} (kJ/mol), between simulation and experiments for large proteins.

Protein (PDB ID)	Simulation	Experiments
Lysozyme (2LYZ)	-25.1	-19.5 to -37.3
Myoglobin (1MBN)	-24.6	-19.0 to -32.0
Cytochrom c (1HRC)	-39.1	-36.8 to -49.4

6.3.3 Case Study of the Protein Lysozyme

Lysozyme has been so extensively studied experimentally that additional validation of the model is possible. Figure 6.8 depicts the structure of lysozyme (PDB ID: 7LZM) colored by secondary structures. The protein contains two sections: the upper part that is formed by several β-strands and α-helices and the lower part that is formed by a group of α-helices.

Experiments have shown that the stability of lysozyme changes depending upon the type of surface to which it is tethered. Hanson *et al.* [1] tethered lysozyme to three different types of surfaces (P2HEMA, PEMA, and POMA). These surfaces correspond to hydrophobic, moderately hydrophilic, and hydrophilic surfaces respectively. It was found that the protein was totally unfolded on the hydrophobic surface, partly folded on the moderately-hydrophillic surface, and folded on the hydrophilic surface.

To see if the model captures this behavior, simulations of 7LZM were done by tethering site 91 to three different surfaces: hydrophobic, moderately hydrophilic, and hydrophilic surface. Site 91 was chosen based upon the results from Chapter 5 where it was shown that tethering protein 7LZM on a purely repulsive surface at site 91 stabilized the protein compared to the protein in the bulk solution. Figure 6.9 shows the equilibrium structure of 7LZM on a hydrophobic surface. The protein is totally melted on this type of surface (with $\chi_s = 4.5$) at a temperature 215 K. The protein is stable at this temperature in the bulk. As

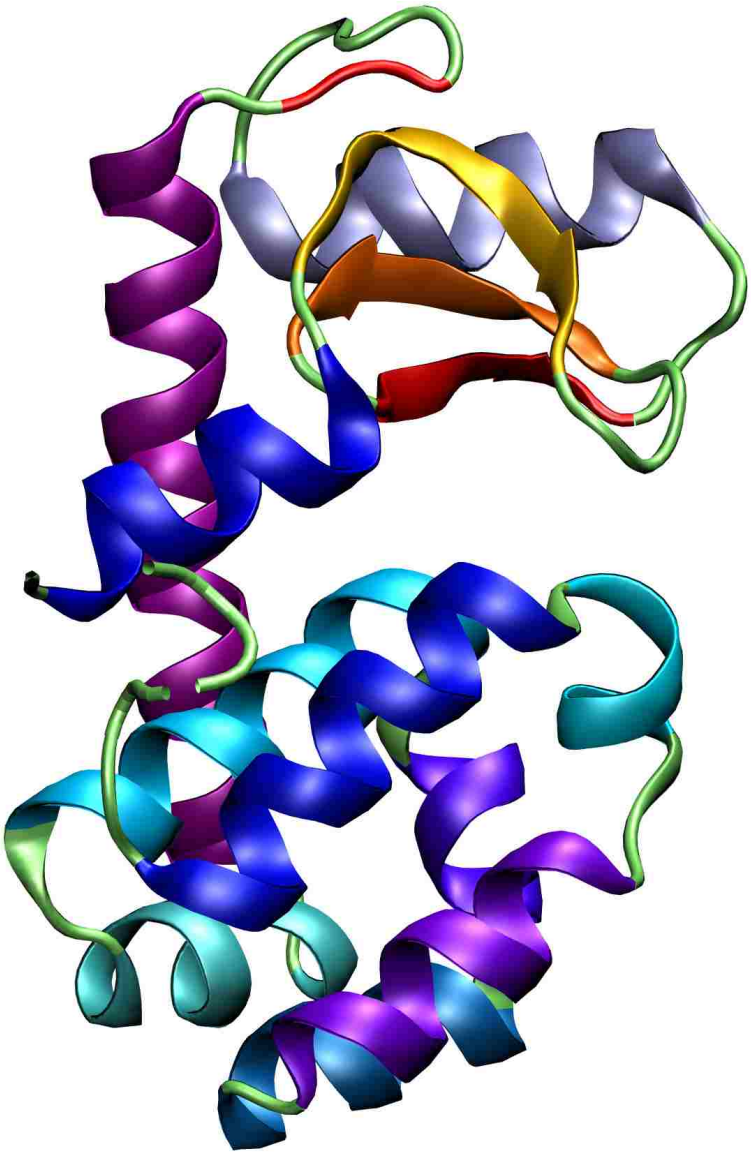


Figure 6.8: Structure of 7LZM colored by secondary structures.

shown in Figure 6.10, at the same temperature 215 K, the protein is only partially stable on a moderately hydrophilic surface (with $\chi_s = 1.5$) as the upper part of the protein is melted while the lower part is still folded. Finally, as depicted in Figure 6.11, the protein maintains a folded structure when it is tethered to a hydrophilic surface (with $\chi_s = -1.0$) with the same tethering site and simulation temperature.

These results are in agreement with the experimental results described previously. No reparameterization was done to achieve these results. The only difference between the simulations is the χ_s value, and yet both simulation and experiment show that the protein unfolds on a hydrophobic surface, remains folded on a hydrophilic surface, and partially unfolds on a moderately hydrophilic surface.

The reason for the different behavior is found in how attractive the protein is to the surface. The hydrophobic surface provides a deep adsorption well for the protein when it is close to the surface and a low energy barrier as it approaches the surface. This causes all of the sites of the protein to want to be close to the surface and the result is a random coil structure confined to a single plane. When the protein is tethered to a hydrophilic surface, a large energy barrier keeps the protein farther away from the surface. Moreover, the adsorption well in this case is shallow. The overall result is the protein remains stable on the surface. The moderately hydrophilic surface, where the protein partially unfolds, is a case in between. The energy barrier is lower than the hydrophilic surface and higher of than that of the hydrophobic surface. Moreover, the adsorption well is deeper than with the hydrophilic surface and shallower than with the hydrophobic surface.

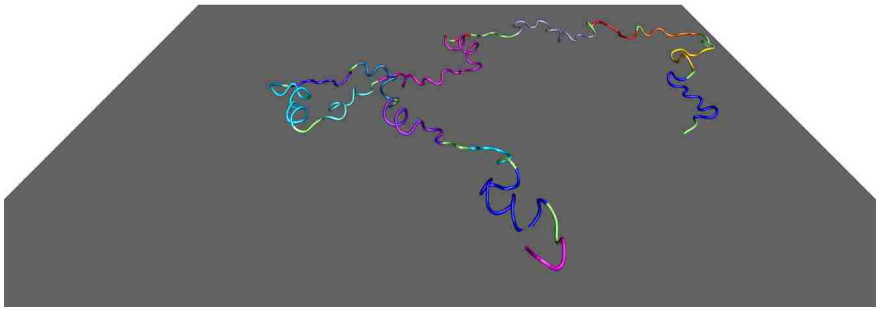


Figure 6.9: Structure of 7LZM tethered on a hydrophobic surface with site 91.

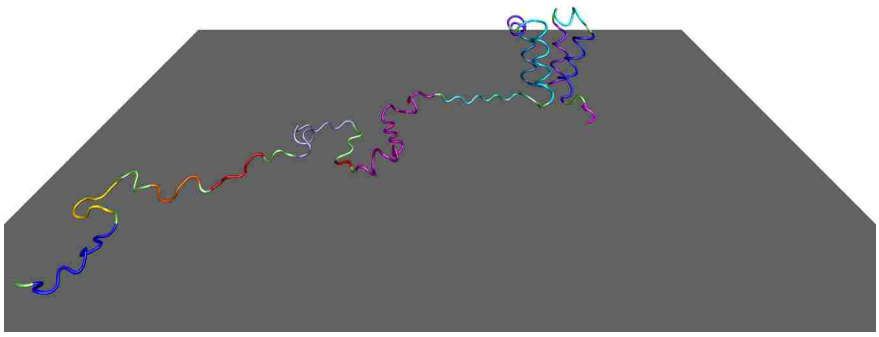


Figure 6.10: Structure of 7LZM tethered on a moderately hydrophilic surface with site 91.

Of note is the portion of the molecule that unfolds on the moderately hydrophilic surface. It is the upper portion of the molecule that contains the α-β roll. Results in Chapter 5 showed that this portion of the molecule is weaker than the bottom portion composed of the orthogonal bundle of α helices. It appears that the α-β roll is the first to unfold regardless of the forces causing the denaturation. In Chapter 5, the unfolding was accomplished using temperature as the denaturant. In this chapter, the forces between the surface and the protein cause the unfolding. The agreement demonstrates the consistency of the simulation approach and the ability of the simulation to accurately capture the effects of different environments.

6.4 Conclusion

This chapter described a new coarse-grained surface model for protein-surface interactions. The model successfully reproduces several important features of protein-surface interactions including the adsorption well, the barrier due to solvent exclusion between the protein and the surface, and residue-level hydrophobic effects. These were lacking in models

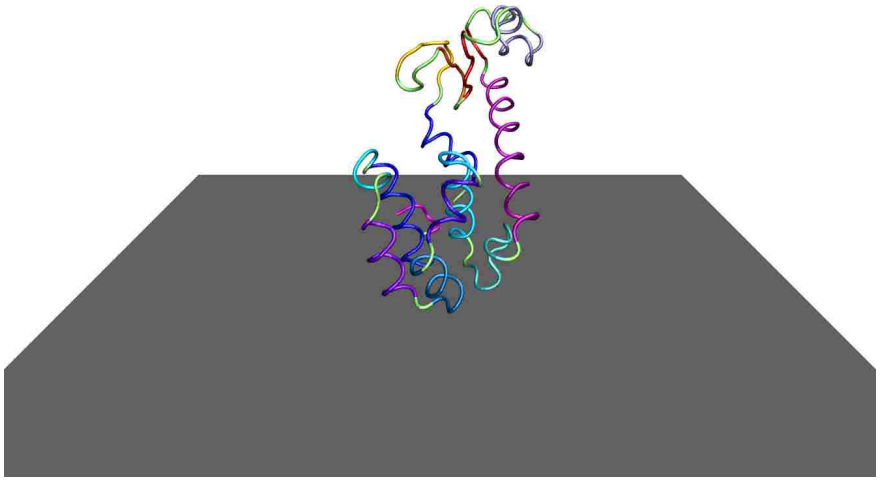


Figure 6.11: Structure of 7LZM tethered on a hydrophilic surface with site 91.

used previously. The performance of the surface model was validated by comparing adsorption free energy of three different proteins with the available experiments data from different groups. The simulation and experiment agree both qualitatively and quantitatively. Additional qualitative validation was done using lysozyme where it was shown that the new model accurately predicts the folding/unfolding behavior of the protein on three different types of surfaces.

This new model will open up new avenues in the simulation of protein/surface interactions. Because the model accurately captures the adsorption well, simulation no longer has to be restricted to systems where the protein is physically tethered to the surface. Now the effects of protein adsorption can be studied. Such effects are important for understanding non-specific protein-surface interactions that are not only found in protein microarrays but in a variety of biomedical applications.

Chapter 7

Conclusion

7.1 Summary

Protein surface interactions are important in a variety of applications but previous research has not provided the needed understanding to predict how a particular system will behave. This work has shown, for the first time how protein stability can be correlated to structure. Specifically, Chapter 3 explained how the shape of the tethering loop region and the ability of the protein to rotate and vibrate on the surface affects stability. Three types of loop regions were discovered: a long loop, a U-shaped loop, and a W-shaped loop. Proteins were stabilized when tethered in each type except the concave-down region of the W-shaped loop. Moreover, a thermodynamic analysis showed that the entropic part of the free energy always helps to stabilize the protein on surfaces and that destabilization in mainly an enthalpic effect.

Further research, described in Chapter 4, showed that the stability of most proteins, not just those possessing alpha-helical, orthogonal bundle structure, can be correlated to loop structure and surface rotation ability. It was also found that the complicated shapes of beta-barrel proteins made them susceptible to destabilization. The results also suggest that mainly-alpha proteins are generally more stable than mainly-beta proteins.

The results in Chapter 5 showed, for the first time, how surfaces can change the folding mechanism of a protein that folds through multiple states. The folding mechanism of 7LZM changes when the protein is tethered on the surface with site 91. One of the intermediates is eliminated creating a three-state mechanism instead of a four-state mechanism as in the bulk condition. It was found that this change stabilized the active site of the protein in a way that would increase function at higher temperatures. Moreover, it was shown that tethering at site 91 presented the active site away from the surface so that substrates would have access

to the tethered enzyme from the bulk phase. The results offer hope that rational design of protein arrays is possible once a molecular-level understanding of the relevant phenomena is obtained.

Finally, Chapter 6 described a new coarse-grain model for protein-surface interactions. It successfully reproduced several important features that have been observed experimentally: the adsorption well, the solvation/desolvation effect, and residue-level hydrophobic effects. None of these were found in models used previously. The performance of the new model was validated by comparing adsorption free energy of three different proteins with available experiments data from different groups. The results were in remarkable agreement with experiment both qualitatively and quantitatively. Further validation was done using the protein lysozyme (PDB ID: 7LZM) on three different types of surfaces. Both experiment and simulation without reparameterization showed the same behavior with lysozyme unfolding on a hydophobic surface, remaining folded on a hydrophilic surface, and partially unfolding on a hydrophobic surface. This model, because it accurately replicates the attractive well found between proteins and surfaces, has the capability to examine protein adsorption in a manner not previously possible.

7.2 Future Work

Fundamental study of protein-surface interactions is in its infancy and many avenues for future work remain. One of these that relates directly to the work described in Chapters 3 and 4 concerns studying more protein structures. Those presented above encompass a majority of the proteins, but more work is needed for a comprehensive understanding. These are found in the mixed alpha/beta class in the CATH classification method. Table 7.1 shows the tertiary motifs that are found in this class that need to be studied.

Table 7.1: Protein structure motifs for further study.

Secondary	Tertiary
Mixed $\alpha - \beta$	2-Layer Sandwich
	3-Layer Sandwich
	Apha-Beta Barrel

Since the novel coarse-grain model that has been developed is well validated, various applications of protein-surfaces interactions could be performed. Comparisons of results generated using the new model with those from previous models needs to be performed to validate the previous results. Studies comparing adsorption to tethering also need to be done.

Future efforts with parameterization could also focus on charged surfaces. This would help improve agreement between the simulation and experiment on surfaces created with $-COOH$ and $-NH_2$ functional groups. Improvement of the protein model to include charge effects is needed for this to occur. Charge effects could be accounted by an additional term in the potential equation. Parameterizations of parameters based on the Latour's benchmark experimental work[73] are needed to be performed for different types of charge models. Finally, further parameterization efforts could improve the numerical agreement between simulation and experiment for the melting temperature of the proteins.

Finally, since the motivation for the entire research is to develop predictive simulation tools that can ultimately improve the performance of protein microarrays, the research will eventually need to address protein-protein interactions. These are foundational to microarray function. Such studies could be based on the protein-surface interaction model explained in Chapter 6, but further development is needed for protein-protein interactions. Charge effects, desolvation effects, and the van der Waals forces between a protein and its ligand need to be taken into account.

Bibliography

[1] M Hanson, KK Unger, R Denoyel, and J Rouquerol. Interactions of lysozyme with hydrophilic and hydrophobic polymethacrylate stationary phases in reversed-phase chromatography(RPC). *J. Biochem. Biophys. Methods*, 29(3-4):283–294, 1994. ii, 9, 86, 89, 97, 99

[2] P Jonkheijm, D Weinrich, H Schroeder, CM Niemeyer, and H Waldmann. Chemical strategies for generating protein biochips. *Angew. Chem. Int. Ed.*, 47:9618–9647, 2008. 1

[3] U Bilitewski. Protein-sensing assay formats and devices. *Anal. Chim. Acta.*, 568:232 – 247, 2006. 1, 5

[4] H Zhu and M Snyder. Protein chip technology. *Curr. Opin. Chem. Bio.*, 7:55 – 63, 2003. 1, 5

[5] M Uttamchandani and SQ Yao. Peptide microarrays: Next generation biochips for detection, diagnostics and high-throughput screening. *Curr. Pharm. Design*, 14:2428–2438, 2008. 1, 5

[6] M Cretich, F Damina, G Pirria, and M Chiari. Protein and peptide arrays: Recent trends and new directions. *Biomol. Eng.*, 23:77–88, 2006. 1, 5, 6, 7

[7] T Joos and J Bachmann. Protein microarrays: potentials and limitations. *Front. Biosci.*, 14:4376–4385, 2009. 1, 5

[8] JJ Gray. The interaction of proteins with solid surfaces. *Curr. Opin. Struc. Biol.*, 14(1):110–115, 2004. 1, 6

[9] K Nakanishi, T Sakiyama, and K Imamura. On the adsorption of proteins on solid surfaces, a common but very complicated phenomenon. *J. Biosci. Bioeng.*, 91:233–244, 2001. 1

[10] MFM Engel, AJWG Visser, and CPM van Mierlo. Conformation and orientation of a protein folding intermediate trapped by adsorption. *P. Natl. Acad. Sci. USA*, 101(31):11316–11321, 2004. 1

[11] P Billsten, M Wahlgren, T Arnebrant, J Mcguire, and H Elwing. Structural-Changes of T4 Lysozyme upon Adsorption to Silica Nanoparticles Measured by Circular-Dichroism. *J. Colloid Interf. Sci.*, 175:77–82, 1995. 1

[12] C Czeslik and R Winter. Effect of temperature on the conformation of lysozyme adsorbed to silica particles. *Phys. Chem. Chem. Phys.*, 3(2):235–239, 2001. 1

[13] H Larsericsdotter, S Oscarsson, and J Buijs. Thermodynamic analysis of lysozyme adsorbed to silica. *J. Colloid Interf. Sci.*, 276:261–268, 2004. 1

[14] G MacBeath and SL Schreiber. Printing proteins as microarrays for high-throughput function determination. *Science*, 289:1760–1763, 2000. 4

[15] 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. 4, 7

[16] V Espina, EC Woodhouse, J Wulfkuhle, HD Asmussen, EF Petricoin, and LA Liotta. Protein microarray detection strategies: focus on direct detection technologies. *J. Immunol. Methods*, 290(1-2):121–133, 2004. 5

[17] PY Li, L Bo, J Gerstenmaier, and BT Cunningham. A new method for label-free imaging of biomolecular interactions. *Sens. Actuators, B*, 99(1):6–13, 2004. 5

[18] X Yu, D Xu, and Q Cheng. Label-free detection methods for protein microarrays. *Proteomics*, 6(20):5493–5503, 2006. 5

[19] S Ray, G Mehta, and S Srivastava. Label-free detection techniques for protein microarrays: Prospects, merits and challenges. *Proteomics*, 10(4, SI):731–748, 2010. 5

[20] H Chandra, PJ Reddy, and S Srivastava. Protein microarrays and novel detection platforms. *Expert Rev. Proteomics*, 8(1):61–79, FEB 2011. 5

[21] BB Haab, MJ Dunham, and PO Brown. Protein microarrays for highly parallel detection and quantitation of specific proteins and antibodies in complex solutions. *Genome Biol.*, 2, 2001. 5, 7

[22] Y Soen, DS Chen, DL Kraft, MM Davis, and PO Brown. Detection and characterization of cellular immune responses using peptide-MHC microarrays. *Plos Biol.*, 1(3):429–438, 2003. 5

[23] KL Hsu and LK Mahal. A lectin microarray approach for the rapid analysis of bacterial glycans. *Nat. Protoc.*, 1:543–549, 2006. 5

[24] T Kodadek. Protein microarrays: prospects and problems. *Chem. Biol.*, 8:105–115, 2001. 5

[25] L Berrade, AE Garcia, and JA Camarero. Protein Microarrays: Novel Developments and Applications. *Pharm. Res.*, 28(7):1480–1499, 2011. 5, 7

[26] I Balboni, S M. Chan, M Kattah, JD Tenenbaum, AJ Butte, and PJ Utz. Multiplexed protein array platforms for analysis of autoimmune diseases. *Annu. Rev. Immunol.*, 24:391–418, 2006. 5, 6

[27] P Angenendt, J Glokler, D Murphy, H Lehrach, and DJ Cahill. Toward optimized antibody microarrays: a comparison of current microarray support materials. *Anal. Biochem.*, 309(2):253–260, 2002. 5, 8

[28] 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. *J. Chromatography A*, 1009(1-2):97–104, 2003. 6

[29] Peluso, P and Wilson, DS and Do, D and Tran, H and Venkatasubbaiah, M and Quincy, D and Heidecker, B and Poindexter, K and Tolani, N and Phelan, M and Witte, K and Jung, LS and Wagner, P and Nock, S. Optimizing antibody immobilization strategies for the construction of protein microarrays. *Anal. Biochem.*, 312:113–124, 2003. 6, 78

[30] M Feldmann. Development of anti-TNF therapy for rheumatoid arthritis. *Nat. Rev. Immunol.*, 2:364–371, 2002. 6

[31] M Feldmann and L Steinman. Design of effective immunotherapy for human autoimmunity. *Nature*, 435:612–619, 2005. 6

[32] J Vilcek and M Feldmann. Historical review: Cytokines as therapeutics and targets of therapeutics. *Trends Pharmacol. Sci.*, 25:201–209, 2004. 6

[33] 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. *Anal. Biochem.*, 330:281–287, JUL 15 2004. 6, 78

[34] M He, O Stoevesandt, and MJ Taussig. In situ synthesis of protein arrays. *Curr. Opin. Biotechnol.*, 19(1):4–9, 2008. 6, 7

[35] M He. Cell-free protein synthesis: applications in proteomics and biotechnology. *New Biotechnology*, 25(2-3):126–132, 2008. 6

[36] O Stoevesandt, MJ Taussig, and M He. Protein microarrays: high-throughput tools for proteomics. *Expert Rev. Proteomic.*, pages 145–157, 2009. 6, 7

[37] K Bssow, C 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 Res.*, 26, 1998. 7

[38] 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. 7

[39] Camarero JA Albala JS Coleman MA, Beernink PT. Applications of functional protein microarrays: identifying protein-protein interactions in an array format. *Methods Mol Biol.*, 385:121–30, 2007. 7

[40] 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. *J. Am. Chem. Soc.*, 120:500–508, 1998. 8

[41] KL Prime and GM Whitesides. Adsorption of Proteins onto Surfaces Containing End-Attached Oligo(Ethylene Oxide)- A Model System Using Self-Assembled Monolayers. *J. Am. Chem. Soc.*, 115:10714–10721, 1993. 8

[42] JC Love, LA Estroff, JK Kriebel, RG Nuzzo, and GM Whitesides. Self-assembled monolayers of thiolates on metals as a form of nanotechnology. *Chem. Rev.*, 105:1103–1169, 2005. 8

[43] M Cretich, G Pirri, F Damin, I Solinas, and M Chiari. A new polymeric coating for protein microarrays. *Anal. Biochem.*, 332(1):67–74, 2004. 8

[44] W Kusnezow, A Jacob, A Walijew, F Diehl, and JD Hoheisel. Antibody microarrays: An evaluation of production parameters. *Proteomics*, 3(3):254–264, 2003. 8

[45] A Hucknall, S Rangarajan, and A Chilkoti. In Pursuit of Zero: Polymer Brushes that Resist the Adsorption of Proteins. *Adv. Mater.*, 21(23):2441–2446, 2009. 8

[46] 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. *Eur. Biophys. J.*, 37:639–45, 2008. 8

[47] TS Tsapikouni and YF Missirlis. Protein-material interactions: From micro-to-nano scale. *Mater. Sci. Eng., B*, 152(1-3):2–7, 2008. 8

[48] WY Chen, HM Huang, CC Lin, FY Lin, and YC Chan. Effect of temperature on hydrophobic interaction between proteins and hydrophobic adsorbents: Studies by isothermal titration calorimetry and the van't Hoff equation. *Langmuir*, 19, 2003. 8, 9, 86, 89, 97

[49] RI Boysen, Y Wang, HH Keah, and MTW Hearn. Observations on the origin of the non-linear van'tHoff behaviour of polypeptides in hydrophobic environments. *Biophys. Chem.*, 77(2-3):79–97, 1999. 8

[50] D Haidacher, A Vailaya, and C Horvath. Temperature effects in hydrophobic interaction chromatography. *Proc. Natl. Acad. Sci.*, 93(6):2290–2295, 1996. 8

[51] MTW Hearn and GL Zhao. Investigations into the thermodynamics of polypeptide interaction with nonpolar ligands. *Anal. Chem.*, 71(21):4874–4885, 1999. 8

[52] A Vailaya and C Horvath. Retention thermodynamics in hydrophobic interaction chromatography. *Ind. Eng. Chem. Res.*, 35(9):2964–2981, 1996. 8

[53] AW Purcell, MI Aguilar, and MTW Hearn. High-performance liquid-chromatography of amino-acids, peptides and proteins. 115. Thermodynamic behavior of peptides in reversed-phase chromatography. *J. Chromatogr.*, 593(1-2):103–117, 1992. 9

[54] P Karsnas and T Lindblom. Characterization of hydrophobic interaction and hydrophobic interaction chromatography media by multivariate-analysis. *J. Chromatogr.*, 599(1-2):131–136, 1992. 9

[55] Y Wei and RA Latour. Correlation between Desorption Force Measured by Atomic Force Microscopy and Adsorption Free Energy Measured by Surface Plasmon Resonance Spectroscopy for Peptide-Surface Interactions. *Langmuir*, 26:18852–18861, 2010. 9, 84, 88

[56] RA Latour. Molecular simulation of protein-surface interactions: Benefits, problems, solutions, and future directions. *Biointerfaces*, 3:FC2–FC12, 2008. 10, 11

[57] J Karanicolas and CL Brooks, III. The origins of asymmetry in the folding transition states of protein L and protein G. *Protein Sci.*, 11:2351–2361, 2002. 10, 88

[58] J Karanicolas and CL Brooks III. Improved Gō-like models demonstrate the robustness of protein folding mechanisms towards non-native interactions. *J. Mol. Biol.*, 334:309–325, 2003. 10, 24, 25, 51, 54, 67, 69, 80, 84, 86, 87, 88

[59] J Karanicolas and CL Brooks III. The structural basis for biphasic kinetics in the folding of the WW domain from a formin-binding protein: Lessons for protein design? *Proc. Nat. Acad. Sci.*, 100:3954–3959, 2003. 10, 24, 25, 51, 54, 67, 69, 80, 84, 86, 87, 88

[60] J Karanicolas and CL Brooks III. Integrating folding kinetics and protein function: Biphasic kinetics and dual binding specificity in a WW domain. *Proc. Nat. Acad. Sci.*, 101:3432–3437, 2004. 10, 24, 25, 51, 54, 67, 69, 80, 84, 86, 87, 88

[61] Ronald D Hills, Jr and CL Brooks, III. Insights from coarse-grained Gō Models for Protein Folding and Dynamics. *Int. J. Mol. Sci.*, 10:889–905, 2009. 10, 24, 25, 51, 54, 67, 69, 80, 84, 86, 87

[62] VP Raut, MA Agashe, SJ Stuart, and RA Latour. Molecular dynamics simulations of peptide-surface interactions. *Langmuir*, 21(4):1629–1639, 2005. 10

[63] M Agashe, V Raut, SJ Stuart, and RA Latour. Molecular simulation to characterize the adsorption behavior of a fibrinogen gamma-chain fragment. *Langmuir*, 21:1103–1117, 2005. 10

[64] JC Hower, Y He, MT Bernards, and SY Jiang. Understanding the nonfouling mechanism of surfaces through molecular simulations of sugar-based self-assembled monolayers. *J. Chem. Phys.*, 125, 2006. 10

[65] 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. 10

[66] P Mulheran and K Kubiak. Protein adsorption mechanisms on solid surfaces: lysozyme-on-mica. *Mol. Simulat.*, 35:561 – 566, 2009. 10

[67] T Wei, MA Carignano, and I Szleifer. Molecular Dynamics Simulation of Lysozyme Adsorption/Desorption on Hydrophobic Surfaces. *J. Phys. Chem. B*, 116(34):10189–10194, 2012. 10

[68] DB Kokh, S Corni, PJ Winn, M Hoefling, KE Gottschalk, and RC Wade. ProMetCS: An Atomistic Force Field for Modeling Protein-Metal Surface Interactions in a Continuum Aqueous Solvent. *J. Chem. Theory Comput.*, 6(5):1753–1768, 2010. 10

[69] M Hoefling, F Iori, S Corni, and KE Gottschalk. Interaction of Amino Acids with the Au(111) Surface: Adsorption Free Energies from Molecular Dynamics Simulations. *Langmuir*, 26(11):8347–8351, 2010. 11

[70] G Collier, NA Vellore, RA Latour, and SJ Stuart. Development of molecular simulation methods to accurately represent protein-surface interactions: Method assessment for the calculation of electrostatic effects. *Biointerphases*, 4:57–64, 2009. 11

[71] JA Yancey, NA Vellore, G Collier, SJ Stuart, and RA Latour. Development of molecular simulation methods to accurately represent protein-surface interactions: The effect of pressure and its determination for a system with constrained atoms. *Biointerphases*, 5:85–95, 2010. 11

[72] O Cohavi, S Corni, F De Rienzo, R Di Felice, KE Gottschalk, M Hoefling, D Kokh, E Molinari, G Schreiber, A Vaskevich, and RC Wade. Protein-surface interactions: challenging experiments and computations. *J Mol Recognit*, 23(3):259–262, 2010. 11

[73] Y Wei and RA Latour. Benchmark Experimental Data Set and Assessment of Adsorption Free Energy for Peptide-Surface Interactions. *Langmuir*, 25:5637–5646, 2009. 11, 82, 86, 87, 88, 95, 106

[74] G Collier, NA Vellore, JA Yancey, SJ Stuart, and RA Latour. Comparison Between Empirical Protein Force Fields for the Simulation of the Adsorption Behavior of Structured LK Peptides on Functionalized Surfaces. *Biointerphases*, 7(1-4), 2012. 11

[75] Y Sun, WJ Welsh, and RA Latour. Prediction of the orientations of adsorbed protein using an empirical energy function with implicit solvation. *Langmuir*, 21:5616–5626, 2005. 12

[76] HX Zhou and KA Dill. Stabilization of proteins in confined spaces. *Biochemistry*, 40:11289–11293, 2001. 12, 45

[77] J Zhou, S Chen, and SY Jiang. Orientation of adsorbed antibodies on charged surfaces by computer simulation based on a united-residue model. *Langmuir*, 19:3472–3478, 2003. 12

[78] 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. *J. Phys. Chem. B*, 105:3610–3613, 2001. 12

[79] F Carlsson, E Hyltner, T Arnebrant, M Malmsten, and P Linse. Lysozyme adsorption to charged surfaces. A Monte Carlo study. *J. Phys. Chem. B.* 12

[80] V Castells and PR van Tassel. Conformational transition free energy profiles of an adsorbed, lattice model protein by multicanonical Monte Carlo simulation. *J. Chem. Phys.*, 122, 2005. 12

[81] V Castells, SX Yang, and PR Van Tassel. Surface-induced conformational changes in lattice model proteins by Monte Carlo simulation. *Phys. Rev. E*, 65, 2002. 12

[82] F Fang and I Szleifer. Controlled release of proteins from polymer-modified surfaces. *Proc. Nat. Acad. Sci.*, 103:5769–5774, 2006. 12

[83] M Friedel, A Baumketner, and JE Shea. Effects of surface tethering on protein folding mechanisms. *Proc. Nat. Acad. Sci.*, 103:8396–8401, 2006. 13, 23, 36, 80, 82

[84] M Friedel, A Baumketner, and JE Shea. Stability of a protein tethered to a surface. *J. Chem. Phys.*, 126:095101, 2007. 13, 23, 24, 36, 45, 80, 82

[85] Z Zhuang, AI Jewett, P Soto, and JE Shea. The effect of surface tethering on the folding of the src-SH3 protein domain. *Phys. Biol.*, 6:015004, 2009. 13, 14, 24, 36, 80

[86] TA Knotts IV, N Rathore, and JJ de Pablo. Structure and stability of a model three-helix-bundle protein on tailored surfaces. *Proteins*, 61:385–397, 2005. 13, 23, 26, 30, 36, 45, 56, 68, 80, 82

[87] AS Freed and SM Cramer. Protein-Surface Interaction Maps for Ion-Exchange Chromatography. *Langmuir*, 27(7):3561–3568, 2011. 13

[88] TA Knotts IV, N Rathore, and JJ de Pablo. An entropic perspective of protein stability on surfaces. *Biophys. J.*, 94:4473–4483, 2008. 14, 24, 30, 36, 45, 80, 83

[89] N Metropolis and S Ulam. The Monte Carlo Method. 1949. 16

[90] Y Sugita and Y Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.*, 314(1-2):141–151, 1999. 17

[91] ME Fisher and M Randeria. Location of Renormalization-Group Fixed-Points. *Phys. Rev. Lett.*, 56:2333, 1986. 17

[92] K Hukushima and K Nemoto. Exchange Monte Carlo method and application to spin glass simulations. *J. Phys. Soc. Jpn.*, 65:1604–1608, 1996. 17

[93] W Gropp, E Lusk, N Doss, and A Skjellum. A high-performance, portable implementation of the mpi message passing interface standard. *Parallel Comput.*, 22:789 – 828, 1996. 20

[94] W Gropp and E Lusk. A high-performance mpi implementation on a shared-memory vector supercomputer. *Parallel Comput.*, 22:1513 – 1526, 1997. 20

[95] 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 Res.*, 37:D310–D314, 2009. 24, 66

[96] CA Orengo, AD Michie, DT Jones, MB Swindells, and JM Thornton. CATH: A Hierarchic Classification of Protein Domain Structures. *Structure*, 5:1093–1108, 1997. 24, 66

[97] TJ Schmitt, JE Clark, and TA Knotts IV. Thermal and mechanical multistate folding of ribonuclease H. *J. Chem. Phys.*, 131:235101, 2009. 25, 51, 80, 87

[98] W Humphrey, A Dalke, and K Schulten. VMD-Visual Molecular Dynamics. *J. Mol. Graphics*, 14:33–38, 1996. 25, 66

[99] D Frishman and P Argos. Knowledge-based protein secondary structure assignment. *Proteins*, 23:566–579, 1995. 25, 66

[100] Y Sugita and Y Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.*, 314:141–151, 1999. 28, 57, 68

[101] Y Sugita, A Kitao, and Y Okamoto. Multidimensional replica-exchange method for free-energy calculations. *J. Chem. Phys.*, 113:6042–6051, 2000. 28, 57, 68

[102] S Nosé. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.*, 81:511–519, 1984. 28, 57, 68, 88, 89

[103] WG Hoover. Canonical dynamics: equilibrium phase-space distributions. *Phys. Rev. A*, 31:1695, 1985. 28, 57, 68, 88, 89

[104] S Nosé. Constant-temperature molecular dynamics. *J. Phys.-Condens, Mat.*, 2:SA115–SA119, 1990. 28, 57, 68, 88, 89

[105] 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. *J. Comput. Chem.*, 13:1011–1021, 1992. 29, 68

[106] SC Hue. Modeling protein density of states: Additive hydrophobic effects are insufficient for calorimetric two-state cooperativity. *Proteins*, 40:543–571, 2000. 30

[107] SP Gore, DF Burke, and TL Blundell. PROVAT: a tool for Voronoi tessellation analysis of protein structures and complexes. *Bioinformatics*, 21:3316–3317, 2005. 32

[108] 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. 40

[109] KA Dill and DOV Alonso. Conformational entropy and protein stability. *Colloquim Mosbach 1988: Prot. Struct. Prot. Eng.*, 39:51–58, 1988. 45

[110] 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 Res.*, 2008. 53

[111] CA Orengo, AD Michie, DT Jones, MB Swindells, and JM Thornton. CATH: A Hierarchic Classification of Protein Domain Structures. *Structure*, 5:1093–1108, 1997. 53

[112] HM Berman, J Westbrook, Z Feng, G Gilliland, TN Bhat, H Weissig, IN Shindyalov, and PE Bourne. The Protein Data Bank. *Nucleic Acids Res.*, 22:235–242, 2000. 66

[113] C Tanford, KC Aune, and A Ikai. Kinetics of unfolding and refolding of proteins : Iii. results for lysozyme. *J. Mol. Biol.*, 73(2):185 – 197, 1973. 66

[114] K Kuwajima, Y Hiraoka, M Ikeguchi, and S Sugai. Comparison of the transient folding intermediates in lysozyme and alpha-lactalbumin. *Biochemistry*, 24(4):874–881, 1985. 66

[115] AF Chaffotte, Y Guillou, and ME Goldberg. Kinetic resolution of peptide-bond and side-chain far-uv circular-dichroism during the folding of hen egg-white lysozyme. *Biochemistry*, 31(40):9694–9702, 1992. 66

[116] SE Radford, CM Dobson, and PA Evans. The folding of hen lysozyme involves partially structured intermediates and multiple pathways. *Nature*, 358(6384):302–307, 1992. 66

[117] T Kiefhaber. Kinetic traps in lysozyme folding. *Proc. Nat. Acad. Sci.*, 92(20):9029–9033, 1995. 66

[118] MJ Parker, J Spence, and AR Clarke. An integrated kinetic-analysis of intermediates and transition-states in protein-folding reactions. *J. Mol. Biol.*, 253(5):771–786, 1995. 66

[119] A Miranker, CV Robinson, SE Randord, RT Aplin, and CM Dobson. Detection of transient protein-folding populations by mass-spectrometry. *Science*, 262(5135):896–900, 1993. 66

[120] G Wildegger and T Kiefhaber. Three-state model for lysozyme folding: Triangular folding mechanism with an energetically trapped intermediate. *J. Mol. Biol.*, 270(2):294–304, 1997. 67, 86

[121] S Wei and TA Knotts IV. Predicting stability of alpha-helical, orthogonal-bundle proteins on surfaces. *J. Chem. Phys.*, 133:115102, 2010. 80, 87

[122] A Morriss-Andrews, G Bellesia, and JE Shea. Effects of surface interactions on peptide aggregate morphology. *J. Chem. Phys.*, 135, 2011. 83

[123] M Iafisco, B Palazzo, G Falini, M Di Foggia, S Bonora, S Nicolis, L Casella, and N Roveri. Adsorption and conformational change of myoglobin on biomimetic hydroxyapatite nanocrystals functionalized with alendronate. *Langmuir*, 24(9):4924–4930, 2008. 86, 98

[124] JE Lee and SS Saavedra. Molecular orientation in heme protein films adsorbed to hydrophilic and hydrophobic glass surfaces. *Langmuir*, 12(16):4025–4032, 1996. 86, 98

[125] JH Santos, N Matsuda, ZM Qi, T Yoshida, A Takatsu, and K Kato. Experimental evidence of the reversibility of the first stage of protein adsorption at a hydrophobic quartz surface near the isoelectric point. *Surf. Interface Anal.*, 35(5):432–436, 2003. 86, 98

[126] JS Salafsky and KB Eisenthal. Protein adsorption at interfaces detected by second harmonic generation. *J. Phys. Chem. B*, 104(32):7752–7755, 2000. 86, 98

[127] BR Brooks, RE Bruccoleri, DJ Olafson, DJ States, S Swaminathan, and M Karplus. Charmm: A program for macromolecular energy, minimization, and dynamics calculations. *J Comput. Chem.*, 4:187–217, 1983. 87

[128] AD MacKerel Jr., CL Brooks III, L Nilsson, B Roux, Y Won, and M Karplus. *CHARMM: The Energy Function and Its Parameterization with an Overview of the Program*, volume 1 of *The Encyclopedia of Computational Chemistry*, pages 271–277. John Wiley & Sons: Chichester, 1998. 87

[129] BR Brooks, CL Brooks, III, AD Mackerell, Jr., L Nilsson, RJ Petrella, B Roux, Y Won, G Archontis, C Bartels, S Boresch, A Caflisch, L Caves, Q Cui, AR Dinner, M Feig, S Fischer, J Gao, M Hodoscek, W Im, K Kuczera, T Lazaridis, J Ma, V Ovchinnikov, E Paci, RW Pastor, CB Post, JZ Pu, M Schaefer, B Tidor, RM Venable, HL Woodcock, X Wu, W Yang, DM York, and M Karplus. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.*, 30:1545–1614, 2009. 87

[130] S Wei and TA Knotts IV. Effects of tethering a multistate folding protein to a surface. *J. Chem. Phys.*, 134, 2011. 87

[131] TC Beutler and WF Vangunsteren. The Computation of a Potential of Mean Force - Choice of the Biasing Potential in the Umbrella Sampling Technique. *J. Chem. Phys.*, 100:1492–1497, 1994. 88

[132] NA Vellore, JA Yancey, G Collier, RA Latour, and SJ Stuart. Assessment of the transferability of a protein force field for the simulation of peptide-surface interactions. *Langmuir*, 26:7396–7404, 2010. 92

[133] XP Geng, MR Zheng, BH Wang, ZM Lei, and XD Geng. Fractions of thermodynamic functions for native lysozyme adsorption onto moderately hydrophobic surface. *J. Therm. Anal. Calorim.*, 93(2):503–508, 2008. 97

Appendix A

Detail of MPI 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
}
```