Simulating Protein Adsorption Phenomena for Experimental Comparison

Michael Deighan

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2014

Reading Committee:

Jim Pfaendtner, Chair

Lilo Pozzo

Lutz Maibaum

Program Authorized to Offer Degree:
Chemical Engineering

University of Washington

Abstract

Simulating Protein Adsorption
for Experimental Comparison

Michael Deighan

Chair of the Supervisory Committee:
Assistant Professor Jim Pfaendtner
Department of Chemical Engineering

Many biological processes and technological applications involve proteins coming into contact with a solid surface. Generally, we know that proteins experience some degree of conformational change at the solid/liquid interface, and can measure these changes in the lab. However, while many experimental techniques exist for characterizing surface-bound proteins, none have been able to resolve high-precision structures. Computer simulation offers a unique route to determining how proteins adsorb. Herein, we apply a popular statistical sampling technique – Parallel Tempering Metadynamics – to all-atom molecular dynamics simulations of explicitly solvated proteins interacting with solid surfaces. We show that by biasing specific degrees of freedom – or collective variables – a protein can be influenced to exhaustively explore conformational space both on and off a surface. The results from these simulations can be post-processed to reveal details such as: surface-bound conformations, orientations, and finer structural details like interatomic distances and Ramachandran angles – which, in turn, can be compared to, and validated by, experimental measurements. Ultimately, this work should convey the descriptive power that can arise from a mutually beneficial partnership between surface science and computer simulation in the context of biomolecular adsorption.

Table of Contents

List of Figures

List of Tables

Acknowledgements

Five years in retrospect seems like so much and so little all at once. I would like to say thank you to the many people who have helped me on my journey. First, I would like to thank François Baneyx, as well as some of his current and former research group – David Chiu, Brandon Coyle, James Matthaei, Brent Nannenga, Juan Miguel Puertas, and Weibin Zhou – for taking me under their collective wing during my first summer in Seattle and giving me my first taste of graduate school.

I am also grateful to other friends and grad students for making everyday life much more interesting; namely, Cori Bucherl, Andrew Collord, John-Rene Ella, Alaina Floyd, Tyler House, Razieh Kalifehzadeh, Drew Keefe, Kasia Nowinski, Ikechukwu Nwaneshiudu, Jeff Richards & Jackie Unger, Beau Richardson & Taryn Borges, Qing Shao, Trent Silbaugh & Krissy Mikhova, and Chris Walcott.

I think I have had the pleasure of interacting with nearly every single member of the Pfaendtner Research Group to date. The fondness I have for the PRG is that of a brother for his siblings. Their sincere intellectual and emotional support helped me through many difficult experiences. Thank you grad students: Patrick Burney, Vance Jaeger, Kelly Fleming, Blake Hough, Kayla Vanous, Stephanie Hoffmann, and Moses Cho – postdocs: Cristina Russo, Shaohui Zheng, and Sean Fischer – and a special shout-out to my two favorite undergrads: Zack Jarin (good luck in Chicago!) and Melissa Gile (viel Spaß in Deutschland!), you kept me sane!

I am profoundly grateful to Jim Pfaendtner for his patience, support, and insight throughout the course of my PhD. Five years is a substantial amount of time for both the advisor and student, especially when they represent $^2/_3$ of the inaugural crew (Patrick rounds it out) – we have come a long way. I doubt I would have ever found the resolve to finish if not for his energy and encouragement. Thank you, Jim.

Finally, I cannot properly put into words how thankful I am for my family: brothers Joe and Dave, and parents Richard and Hilary. My parents deserve their own special recognition. They were in my heart every step of the way, sticking with me through every triumph and tribulation. Their unconditional love has kept me, and continues to keep me, firmly grounded in this universe. With my whole self, thank you Mom and Dad. I love you.

Dedication

To all my family - even the fuzzy ones.

Introduction

The idea of a boundary is an invention of human intuition that allows us to rationally evaluate physical systems. For example, we use the concept of a boundary to set limits for differential equations, define the confines of an apparatus, and segregate different phases and states. Therefore, the natural world on many (if not all) scales can be imagined as a rich, heterogeneous medium. Phases partitioned by boundaries interact with both themselves and one another at interfaces, where two or more boundaries meet. Consequently, a tremendous amount of complexity emerges from a myriad of interfacing schemes that occur among or within physical systems (e.g., subatomic or atomic interactions, chemical bonds, living organisms, or planetary sytems). Furthermore, the drive to understand and predict how different physical systems interface has provided inspiration for some of humanitys greatest discoveries and achievements.

When discussing biological interfaces, proteins dominate the conversation. Proteins (for terseness, but also biomolecules such as DNA and RNA), in and of themselves, are exemplars of interfacing – the functionality of many depends on intrachain interactions and their relationship with the surrounding environment. In addition to their multifarious biological role, proteins are both utilized and studied by a number of fields that focus on: filtration, [1] separations, [2] materials development, [3] coatings, [4] and chromatography [5] to name a few examples. In many of these examples, one can imagine that a protein is in contact not only with a bulk solvent environment, but also with a secondary phase – in many cases, a solid surface. To consider the molecular-scale behavior of a protein in solution is a sizeable endeavor, but to attempt to predict how a protein simultaneously interacts with two distinct phases – essentially, the intersection of three interfaces (protein/liquid, protein/solid, and liquid/solid) –constitutes a separate area of study altogether. Conventionally, the structure of a protein can be ascertained either in solution from nuclear magnetic resonance (NMR) spectroscopy, or in the crystal state with X-ray crystallography. No such technique exists for resolving whole surface-bound proteins, and therein lies the central problem posed at the onset of this work. Specifically, educing the molecular-scale conformation(s) of a whole protein at a liquid/solid interface is a task that could not be done five years ago by experimental analysis alone, and still cannot be done today. As of the time of this writing, the Protein Data Bank (PDB) contains over 100,000 entries –none of which,

to our knowledge, are of a protein in an adsorbed state.

Several computational research groups became aware of this deficiency during the first decade of the 21st century. [6–10] With the advent of powerful microprocessors and sophisticated modeling techniques, molecular simulation was poised to make increasingly impactful contributions in many fields, including surface science. Early simulations of proteins at surfaces constituted a small protein in proximity to a surface, and then allowing the system to settle into thermodynamic equilibrium using a molecular mechanics (MM), Monte Carlo (MC), or molecular dynamics (MD) algorithm. [11–22] Later studies employed computational techniques such as: the RosettaSurface algorithm, [23] which produces energy-minimized conformations of surface-bound proteins; Umbrella Sampling, [24] where a peptide is restrained at different distances from a surface, providing an idea of the potential energy landscape for adsorption; biased replica exchange, [25] which adds an umbrella sampling-derived potential field to a replica exchange (a.k.a. parallel tempering, discussed below) to encourage better sampling; and steered molecular dynamics, [26] which "pulls" a protein from a surface (analogous to atomic force microscopy) in order to identify interesting energetic features of the protein/surface complex. However, while these early simulations shed light on simple aspects of protein adsorption (e.g., conformational change, surface affinity of different residues, or binding energy), they lacked the ability to produce converged statistical samples of adsorption phenomena, which precluded quantitative comparison with experiment. [27]

Sampling challenges arose in early simulations mostly due to time- and length-scale impediments inherent to molecular simulation (discussed below). Compounding those two fundamental issues was also the question of system size. Protein adsorption simulations require a sufficient amount of lateral space for a protein to undergo conformational change and diffuse across a surface without influencing itself through the periodic boundary. Similarly, enough overhead space must be provided in the simulation environment for a protein to be able to diffuse beyond the influence of the surface. There, it may freely orient and serve as a solution phase comparison to the surface-bound alternative. As a result, all-atom, explicitly solvated protein adsorption simulations require large simulation environments, which can be extremely expensive to converge.

The work presented in this document begins by addressing these sampling challenges. Specifically, Chapter 1 expands on a large body of work authored by the developers of Metadynamics. Therein, we demonstrate that the cost of a simulation can be greatly reduced by subtly biasing a system's potential energy, setting the stage for large-scale protein adsorption simulations. Chapter 2 takes the proof of concept established in Chapter 1 and applies it to several simulations containing short LK peptide sequences in the presence of two self-assembled monolayer surfaces. With this work, we were able to demonstrate the applicability of the sampling method developed in Chapter 1 to study protein adsorption, provide insight on peptide adsorption at the molecular level, compare existing force-fields and methodologies, and – most importantly – link our

results to experimental data, adding to a growing body of literature that justifies the proposed partnership between experiment and simulation. [27–31] In Chapters 3 and 4 we comprehensively evaluate the salivary protein statherin in the presence of a hydroxyapatite surface. Statherin was an ideal candidate for study, owing to the extensive experimental attention it has garnered. Chapter 3 focuses on a truncated statherin segment – namely, the first 15 N-terminal residues that constitute the protein's surface-binding domain. In Chapter 4, the adsorption behavior of the entire 43-residue statherin protein is simulated. In the case of both SN15 and statherin, we are able to compare simulation results directly to over 20 individual experimental measurements. Our findings provide insight for both theorists and experimentalists on protein adsorption phenomena. The overarching theme of this work declares this sophisticated MD strategy as a workhorse that can untangle fine, molecular-scale details within a complex system. Moreover, its simple implementation and exclusive use of free software should encourage curious scientists in a range of disciplines to try their hand at molecular modeling. In the final chapter, the overall significance of this work is discussed, including directions for future work.

Molecular Simulation Primer

Advances in computational speed over the past few decades have enabled the development of sophisticated algorithms geared toward elucidating the behavior of microscopic systems. Many different approaches have arisen from fundamental and empirical principles, and continue to evolve toward becoming more inclusive, descriptive, and efficient at simulating physical reality. Physics-based computational models in use today can be categorized according to which time and length scales they approximate most effectively. Quantum mechanical (QM) and *ab initio* algorithms account for both electronic and nuclear degrees of freedom (DOFs) in their calculations, and must be integrated over very short time steps (10^{-15} seconds) in order to remain consistent with the First Law of Thermodynamics – as is the case, more or less, for the all molecular modeling algorithms. Atomistic methods such as MM, MC, and MD forgo calculating electronic DOFs to gain access to longer time scales and larger length scales (up to microseconds and nanometers). These methods represent atoms as points in space that interact with one another according to simple potential energy functions. Greater time and length scales can be simulated with coarse graining (CG) methods that agglomerate functional groups, monomers, or whole molecules into singular entities. These models can be used in conjunction with MM, MC, or MD algorithms to explore the physical properties of macromolecular structures (e.g., collagen bundles or the lipid bilayer). [32, 33] Additionally, the accessible timescale for atomistic methods can be extended even further (approaching milliseconds) by using non-standard computing architectures – namely, graphics processing units (GPUs) or the ANTON super computer. [34,35] Ultimately,

continuum models (e.g., the Navier-Stokes equation) replace discrete, particle-based models for simulating systems that contain molar quantities of molecules.

The work presented in this document makes use of MD for simulating proteins, surfaces, and explicit solvent molecules. MD is a straightforward algorithm that scientists can use with facility to examine many different physical systems at the molecular scale. Although it consists of, in its basic form, simple iterations of Newton's equations of motion, MD aptly addresses the intrinsic nonlinearity of the natural world. In other words, MD calculations take into account the coupling and interdependence of many degrees of freedom in a physical system, whose behavior may be nearly impossible to describe analytically. However, serious limitations become apparent when considering the time and length scales that are accessible to MD, especially in the context of complex biological processes (e.g., protein folding or conformational change at a surface). Additionally, computational cost scales proportionally with N·log(N) (N = number of atoms), which impedes the simulation of systems larger than a few million atoms. [36–39] These limitations underscore one of the central challenges to simulation research: efficiently obtaining a representative statistical sample of phase space from a complex molecular model. The following section briefly describes the specific method used for the work presented in this thesis.

Sampling Technique Primer

As noted above, the time and length scales accessible to MD restrict fast and efficient sampling by conventional means (i.e., brute force calculation). The work presented in the following chapters involves moderately sized systems (< 100,000 atoms), suitable for MD. However, protein conformational change and adsorption are complicated processes that can take place over seconds or much much longer – a corresponding simulation would certainly extend beyond the lifetime of even the most masochistic graduate students. How, then, is it possible to circumvent an obstacle as fundamental as time? Many researchers have asked themselves a similar question. In response, several highly effective "enhanced sampling" (ES), or "generalized ensemble" (GE), methods have been developed over the past few decades. These methods all take advantage of the statistical mechanical framework underpinning MD. The GEMD methods most pertinent to my research are briefly described below.

Parallel Tempering Metadynamics

This work makes use of the Parallel Tempering Metadynamics in the Well-Tempered Ensemble (PTMetaD-WTE) enhanced sampling method. PTMetaD is an amalgam of two highly effective GE methods, described separately here – the WTE is introduced in Chapter 2. Recall from the previous section that, in MD,

an atoms electronic degrees of freedom are averaged out. Instead, bonds and non-bonded interactions are represented by different pairwise potentials, constituting the basis of the MD force field, detailed below and as a cartoon in Figure 0.1.

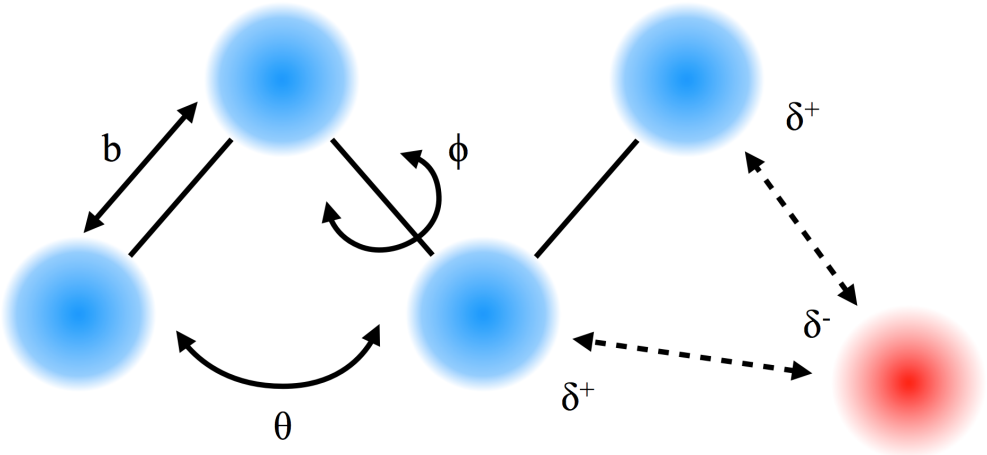


Figure 0.1. Cartoon representing the conventional molecular dynamics force field. Bonds (b), bond angles (θ), and dihedral angles (ϕ) are all represented by Equations 0.2, 0.3, and 0.4, respectively. Non-bonded atoms (red and blue), and atoms more than 4 bonds away within the same molecule, "see" each other as soft spheres. They interact via Van der Waals and electrostatic forces, whose governing equations are Equations 0.5 and 0.6, respectively. Most force fields assign a single, static point charge to each atom (δ), which, in part, approximates electron density.

$$U = U_{bond} + U_{angle} + U_{dihedral} + U_{VdW} + U_{Coulomb} \tag{0.1}$$

$$U_{bond} = \frac{1}{2} k_{ij}^{b} \left(r_{ij} - r_{ij}^{o} \right)^2 \tag{0.2}$$

$$U_{angles} = \frac{1}{2} k_{ij}^{\theta} \left(\theta_{ijk} - \theta_{ijk}^{o} \right)^2 \tag{0.3}$$

$$U_{dihedral} = k_{\phi} \left(1 + cos \left(n\phi - \phi^{o} \right) \right) \tag{0.4}$$

$$U_{VdW} = 4\epsilon_{ij} \left(\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right) \tag{0.5}$$

$$U_{Coulomb} = \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}^2} \tag{0.6}$$

Metadynamics (MetaD) augments Equation 0.1, the total potential energy, in a specific way. [40] In practice, the MetaD bias energy acts on slow DOFs inherent to a system in order to accelerate the sampling of conformational space. Slow DOFs are represented by user-defined collective variables (CVs), which describe one or more intuitively slow processes (e.g., breaking/forming of hydrogen bonds, collapse/dispersion of

hydrophobic residues, or adsorption/desorption with respect to a surface). MetaD bias is periodically added to a simulation in the form of Gaussian hills, forming a history-dependent energy function that forces a system into regions of unexplored conformational space, and eventually approximates the underlying free energy landscape with respect to the chosen CV(s). In Well-Tempered Metadynamics (WTM), each new bias addition decays with respect to the history-dependent potential, [41] resulting in a smoothly converging sum as simulation time approaches infinity:

$$U_{meta}\left(s(r),t\right) = \sum_{n=1}^{t'=n\tau<t} \left(\omega \cdot \exp\left[\frac{-U_{meta}\left(s(r),t'\right)}{k_B \Delta T}\right] \cdot \prod_{i=1}^{N_{CV}} \exp\left[\frac{-\left(s_i(r) - s_i(r(t'))\right)}{2\sigma_i^2}\right]\right) \tag{0.7}$$

The sum U_{meta} represents the total bias potential added to Equation 0.1, where each term within the sum takes the shape of a N_{CV}-dimensional Gaussian hill whose height is scaled by the decaying prefactor $\omega \cdot \exp\left[\frac{-U_{meta}\left(s(r),t'\right)}{k_B \Delta T}\right]$. Please note that the WTM bias does not perfectly compensate the underlying free energy landscape to yield a flat histogram. Instead, it produces a scaled impression that can be scaled to reproduce an impression of the free energy landscape. Some analogies exist that explain this concept. For example: filling an empty swimming pool with grains of sand, or water flooding a valley to a high enough mark for a boat to freely pass to the next valley. For WTM, imagine a garden in a snowstorm. Let the area of the garden represent a plane formed by the intersection of two CVs, and its topography (topiary) represent the underlying free energy landscape. Each snow flake, in this example, is a single packet of Metadynamics bias energy. We are unable to see the true underlying landscape, but we can estimate what it looks like based on how snow settles onto it. After enough snow has fallen, the features of the garden become smoothed out, but an impression of its underlying topography is still evident. The formalism of WTM allows us to examine the smoothed surface and determine where the footpaths are, what kinds of plants are growing, how many gnomes survived, and so on. We used the Well-Tempered version of MetaD in work reported in this thesis.

Parallel Tempering (PT) improves sampling efficiency in a more general way than MetaD. [42, 43] A PT simulation comprises several identical replica systems, each equilibrated at incrementally greater temperatures. Periodically, throughout a simulation, the coordinates of replicas adjacent in "temperature space" are swapped according to the Metropolis criterion:

$$P = \min\left\{1, \exp\left[\left(\beta_j - \beta_i\right)\left(U(q_j) - U(q_i)\right) + \beta_i\left(V_i(s(q_i)) - V_i(s(q_j))\right) + \beta_j\left(V_j(s(q_j)) - V_j(s(s_i))\right)\right]\right\} \tag{0.8}$$

This method allows a protein at a colder temperature to gradually diffuse to hotter temperatures where they can overcome all relevant energy barriers, and subsequently return to colder temperatures assuming a different configurational state. In the limiting case of an infinitely long simulation using a perfect force

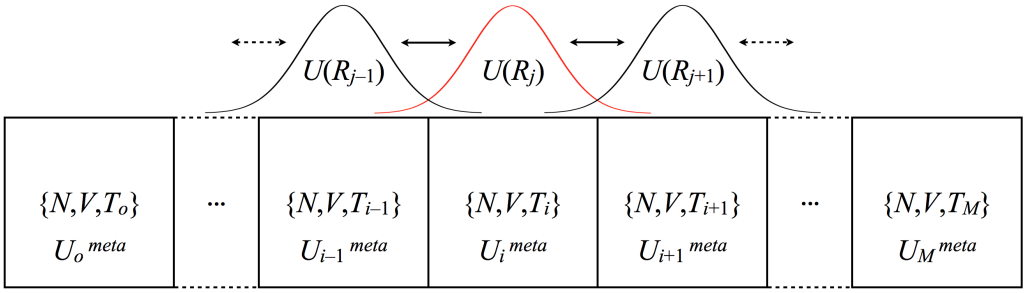


Figure 0.2. Graphical representation of the PTMetaD scheme. A series of M replicas of identical volume (V) and composition (N) hold discrete places in a temperature distribution ranging from T_O to T_M. Periodically, the coordinates (R_j) of two adjacent replicas are allowed to switch temperatures if their respective potential energies (U_j) satisfy the Metropolis criterion in Equation 0.8. The overall exchange rate between two replicas is determined by how substantially their respective potential energy distributions overlap. Meanwhile, MetaD bias (U_i^{meta}) accumulates at each position in temperature space, but is not swapped.

field, an accurate multidimensional histogram of phase space will eventually emerge at all temperatures. Combining PT and MetaD improves the efficacy of both methods. [44] MetaD drives a system to explore specific regions of CV space, whereas PT enables the system to cross any "hidden" energy barriers that would otherwise be invisible to the MetaD bias. A graphical description of PTMetaD is provided in Figure 0.2. The two methods work seamlessly in concert to produce a comprehensive description of a system's underlying free energy landscape, which can be processed to yield the equilibrium Boltzmann distribution for any observable. [45] In the following chapter, we describe how generally biasing a systems potential energy (establishing the Well-Tempered Ensemble) can greatly reduce computational cost of large-scale PTMetaD simulations.

Chapter 1

Efficient Simulation of Explicitly Solvated Proteins in the Well-Tempered Ensemble[1]

The task of quickly and accurately exploring large regions of phase space is a pressing challenge for biomolecular simulations, especially when considering large systems (e.g., proteins or other organic polymers). Schemes that address this challenge, frequently referred to as enhanced sampling methods, typically fall into two broad categories. In one class of methods, specic slow degrees of freedom (i.e., collective variables, or CVs) are biased in order to accelerate the dynamics of the system, forcing an extensive exploration of conformational space. A prominent example of this class is the metadynamics method, which has become increasingly popular over the past decade. [40, 46] The Metadynamics methodology is centered around the periodic addition of a Gaussian-shaped packets of bias potential to a system's Hamiltonian. Over the course of a simulation, the history-dependent bias sum influences a system to sample new regions of relevant phase space, producing a free energy profile along a user-defined set of CVs. However, Metadynamics suffers from a general problem: hidden degrees of freedom, which may not be accurately described by the chosen CVs, can frustrate exploration of phase space and sometimes iterfere with the extent of convergence and accuracy of results.

An alternate approach is to manipulate some or all degrees of freedom in a more general way. For our purposes, we focus on accomplishing this by increasing a system's temperature. The archetypical example of this class is parallel tempering (PT). [42, 43] In PT, a series of identical replicas are simulated together along a distribution of temperatures. Periodically, adjacent replicas in temperature space are allowed to exchange their coordinates in accordance with the Metropolis criterion. [47] A measure of effciency for PT is the characteristic time required for the coldest replica to diffuse to the hottest temperature and return (i.e., the round-trip time, RTT). [48]

[1] Reproduced in part with permission from M. Deighan, M. Bonomi, J. Pfaendtner. Efficient simulation of explicitly solvated proteins in the well-tempered ensemble. *J. Chem. Theory Comput.*, 8:2189-2192, 2012. Copyright 2012 American Chemical Society.

A more effcient approach, which incorporates the attributes of both classes mentioned above, is to combine metadynamics with PT (PTMetaD). [44] This method allows a system to overcome hidden energy barriers and comprehensively explore CV space. However, although this approach is very powerful and significantly reduces the required simulation time *per replica*, it is not a panacea. Systems containing more than a few dozen amino acids still remain prohibitively large for the PTMetaD algorithm due to the large number of replicas that would be required to simulate those systems.

This scaling challenge is rooted in the fact that in order to achieve effcient exchange between adjacent replicas, suffcient overlap between their potential energy (PE) distributions is needed. In the well-tempered ensemble (WTE) the PE overlap is increased by amplifying energy fluctuations with a tunable factor γ, resulting in a decreased RTT and quicker convergence. [49] More importantly, since WTE preserves the average energy of the original ensemble, canonical averages of all other obsrvables can be accurately obtained by reweighing techniques. [45, 50]

To investigate the performance of PTMetaD and WTE combined (PTMetaD-WTE), we have applied it to an explicitly solvated 20-residue protein and performed over 20 μs of aggregate sampling. We compare PTMetaD-WTE to a full PTMetaD simulation and show that a reduction of the overall computational cost can be achieved by properly tuning γ.

All simulations were carried out using the Gromacs 4.5.3 molecular dynamics engine, [51] the AMBER99SB force field, [52] and the PLUMED 1.2.2 plug-in. [53] Calculations were performed on a 4.9 nm^3 simulation box containing the 20-residue tryptophan-cage protein (PDB entry 1L2Y [54]), 3717 TIP3P water molecules, and a chlorine ion for charge neutrality. To study the properties of PTMetaD-WTE, we performed the following simulations: (I) 100-replica PTMetaD, (II; $\gamma = 12$) 100-replica PTMetaD-WTE, (III; $\gamma = 12$) 50-replica PTMetaD-WTE, (IV; $\gamma = 12$) 10-replica PTMetaD-WTE simulation, and (V; $\gamma = 24$) 10-replica PTMetaD-WTE. Simulations I and II were completed with 50 ns/replica, III had 100 ns/replica, and both IV and V had 250 ns/replica. Therefore, we report a total of 20 μs of aggregate sampling (5 μs each for IIII and 2.5 μs each for IV and V). Details pertaining to preparing and running PTMetaD-WTE simulations are provided in Appendix I. For the well-tempered metadynamics, [41] we used CVs that have often been used for studying protein folding: [46] (1) the formation of the secondary structure (i.e., hydrogen bonds) and (2) the hydrophobic core of the protein (see Appendix I for more information regarding the PTMetaD-WTE parameters).

The rate of convergence of our calculations as a function of the total computational time is computed using the FES from the PTMetaD run (simulation I) as a reference, shown in Figure 1.1. Unlike small model systems (e.g., alanine dipeptide) from which an exhaustively sampled FES can be obtained with umbrella sampling or metadynamics, [41] the complexity of the trp-cage system precludes us from obtaining such a

perfectly sampled reference. However, we believe that the FES from simulation I is a good approximation of the real FES, given that we not only observed hundreds of recrossings between the folded and unfolded states, but also saw the free energy difference between both states clearly converge, which we show in Figure A.1 in Appendix I. Furthermore, Figure A.1 shows the quantitative value of folding free energy we obtained is in reasonable agreement with that determined from experiments, a nding that was previously reported for this same protein and force field. [55] These three criteria all point toward convergence of the reference simulation; however, in principle, it would be equally appropriate to assess convergence based on the behavior of other experimental observables whose ensemble averages could be obtained from reweighting the PTMetaD-WTE simulations in a straightforward way. [29, 53]

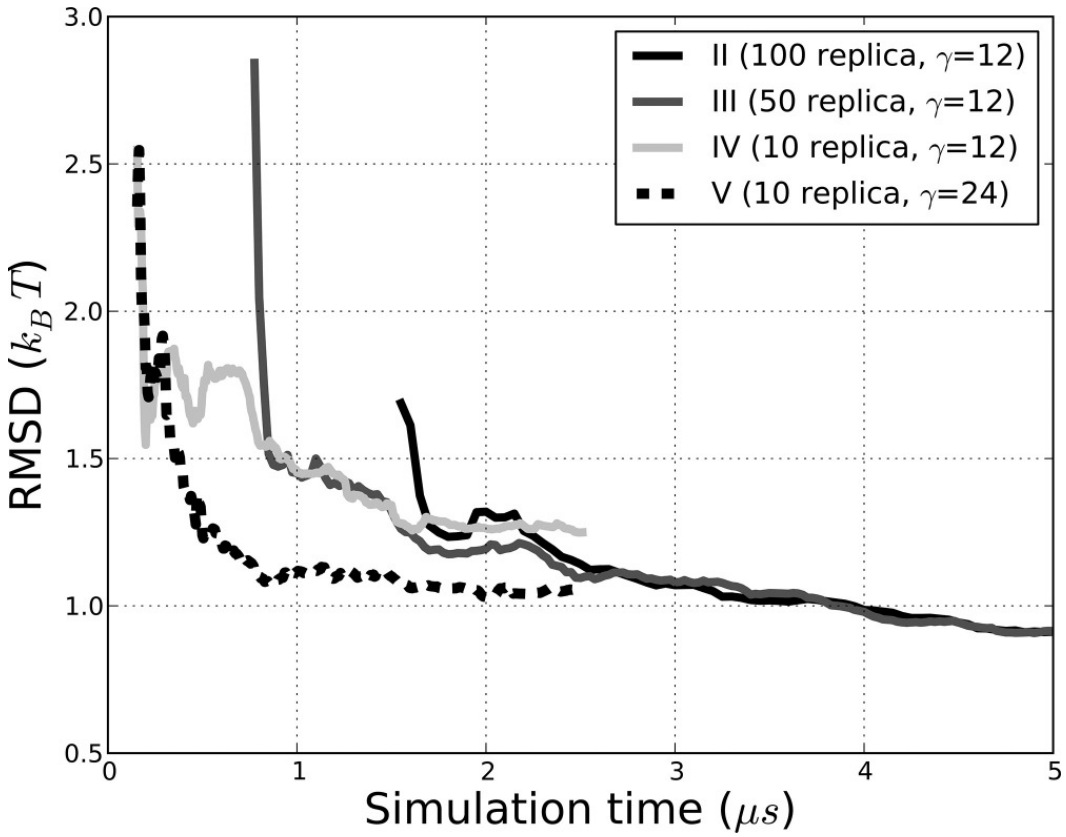


Figure 1.1. Convergence of the simulations using the RMSD of the FES between PTMetaD-WTE and a reference (details in Appendix I). The reference is the FES from simulation I, and serves as an estimate of the true FES for TRP cage folding. The four simulations shown are labeled in the gure inset. Total simulation time is dened as the simulation time per replica multiplied by the number of replicas.

Table 1.1 shows that simulations with similar RTTs (runs II-IV) have a similar convergence trend with respect to the total simulation time. In contrast, simulation V shows improved convergence and a reduction of the overall computational cost. Simulation V has a smaller RTT value, presumably made possible by increasing the adjustable WTE parameter γ to 24. This especially convincing when comparing simulations

Table 1.1. Performance Metrics for Enhanced Sampling Simulations

RUN	Replicas	γ	RTT[a]	RMSD[b]	t^c_{FE}
I	100	1	6.3 (0.2)	0.34	10.1 (0.6)
II	100	12	5.1 (0.1)	0.38	11.2 (0.7)
III	50	12	5.4 (0.2)	0.37	9.1 (0.8)
IV	10	12	5.7 (0.5)	0.38	11.5 (2.4)
V	10	24	4.0 (0.3)	0.46	8.4 (1.4)

[a] Average round trip time in ns, with the standard error in parentheses
[b] Comparison to the lowest energy NMR structure in Å
[c] Average time per folding/unfolding event in ns, with the standard error in parentheses

IV and V, as the only difference between them is the value of γ. The two factors that most strongly control the convergence of the PTMetaD-WTE simulations are the RTT and the total simulation time. As we show in Figure 1.1, similar total simulation times can be achieved by balancing the number of replicas with the simulation time per replica. Over 2.5 μs of total simulation time, we see nearly identical convergence behavior for simulations II-IV. PTMetaD simulations that comprise a small number of replicas are not generally achievable. The incorporation of the WTE framework offers users the ability to tune the number of replicas and RTT to the availability of a particular resource.

It is important to note that the benefits gained from the walk through temperature space correlate with the presence of energy barriers greater than $k_B T$ at the temperature of interest. Even with Metadynamics applied to the two coarse folding CVs, we expect there to be some "hidden variables" preventing diffusive exploration of phase space once the bias potential is converged. To demonstrate that the hidden barriers are energetic and not entropic, we performed a serial simulation of Well-Tempered Metadynamics at 300 K for a simulation time of 500 ns. This simulation contained only 10 folding events (47 ns/event based on total simulation time) compared with 10 ns/event observed for PTMetaD-WTE. Additionally, a comparison of the FESs from the reference PTMetaD simulation (I), the fastest converging PTMetaD-WTE simulation after 500 ns of total sampling (V), and the single replica simulation after 500 ns in Figure A.2 shows that although the folded state can be identified, the unfolded region is dramatically overfilled due to the comparative lack of sampling. Therefore, even for studies that do not make use of the higher temperature data, the PTMetaD-WTE algorithm can offer a signicant reduction in computational cost.

In spite of our complicated system model and the additional time-dependent metadynamics bias applied to it, we have effectively reproduced all of the scaling behavior that is expected from many years of experience with standard PT simulations. [56,57] Future work will explore the convergence properties of PTMetaD-WTE calculations with a model system that permits a more systematic exploration of its relevant parameters. In

light of recent work showing a dramatic loss of accuracy in biased simulations that depart too far from their original ensemble, [50] our future work should also provide quantitative guidance for estimating a reasonable upper bound for γ that is appropriate for explicitly solvated systems. A related consideration is the extra computational cost required to converge the WTE bias potential at higher values of γ.

Figure 1.2 shows the FES as a function of the metadynamics CVs obtained from four of the simulations (IIV). A two state system with a tightly dened folded is correctly described by each simulation. Additionally, we verified that the additional WTE bias did not adversely affect the FES obtained for the metadynamics CVs. Formally, the FESs in Figure 1.2 were obtained by computing the probability distribution of the H-Bond and Hydrophobic Core CVs after removing the effect of the bias on the potential energy. Comparing the FESs with and without the removal of the WTE bias showed the results to be equal within about 0.1 kcal/mol. The fact that WTE preserves the canonical averages guarantees a strong overlap of the biased and unbiased ensembles and facilitates the reconstruction of the latter. Conversely, the WTE energy distribution is not signicantly altered by the Metadynamics bias along other CVs. As shown in Figure A.3, the distribution of the potential energy does not change appeciably from the beginning to the end of thePTMetaD-WTE simulation.

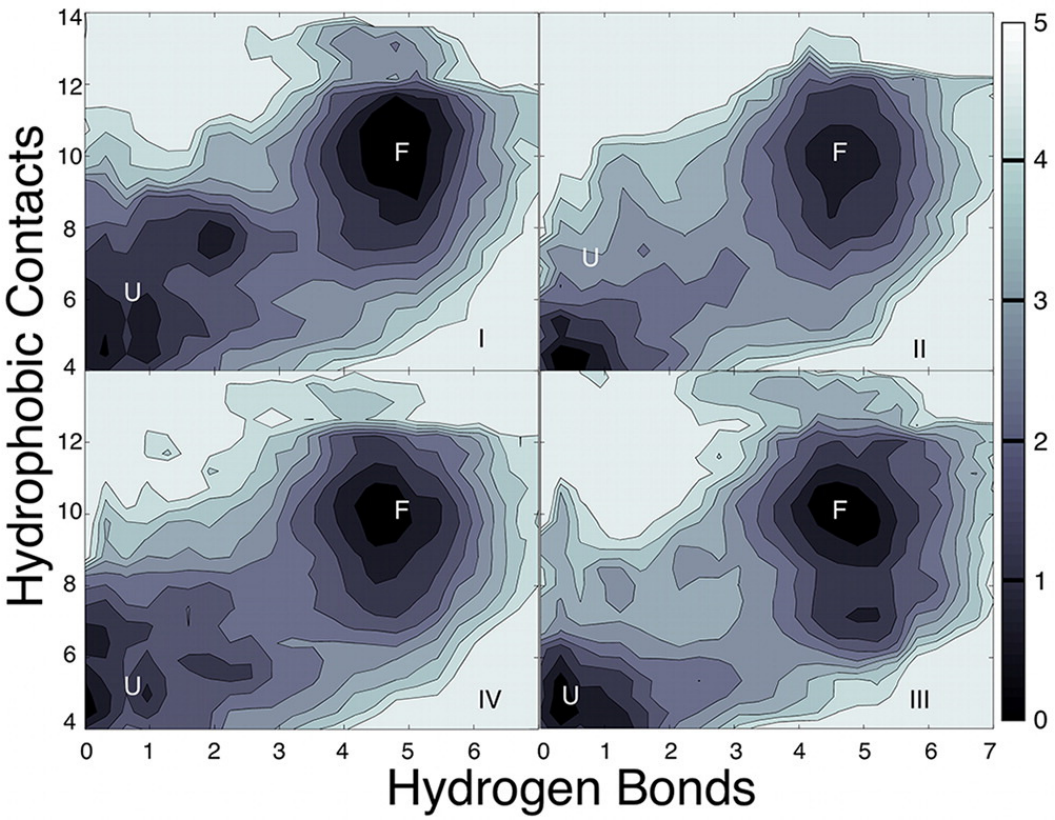


Figure 1.2. Free energy surfaces for runs I-IV. The folded region is marked in each by the white F. Contour lines are plotted every 0.5 kcal/mol with the energy scale shown at the right side of the figure.

In addition to ensuring the similarity of the FESs, we assessed how PTMetaD-WTE affects the folding/unfolding rate of the protein relative to PTMetaD. As shown in Table 1.1, the average time per folding event (described in Appendix I) is essentially the same for every simulation, with the PTMetaD simulation (I) falling somewhere in the range of observed values from PTMetaD-WTE. This demonstrates that the additional bias on the PE does not introduce an appreciable slowing of the conformational dynamics of the system.

In summary, we reported initial observations of the convergence and properties of all-atom PTMetaD-WTE simulations. By exploiting the properties of the recently introduced WTE, we were able to establish amplified potential energy uctuations in a system containing an explicitly solvated tryptophan-cage protein while actively biasing two additional CVs. The convergence properties of PTMetaD-WTE depend both on the total simulation time and the RTT. Without any rigorous tuning of the adjustable parameter γ, we observed that a trp-cage simulation can be reduced to 10 total replicas, which achieves a similar RTT and total computational cost to 100-replica PTMetaD and PTMetaD-WTE simulations. This is signicant in that the 10-replica simulation would otherwise experience no exchange events, and therefore have an innite RTT, without the imposed WTE framework. A reduction of the overall computational cost can also be achieved by further tuning the value of the WTE parameter γ; thereby enhancing the exchange probability between replicas and reducing the RTT. The agreement found in our simulations shows that PTMetaD-WTE in all-atom simulations of biomolecules is a robust improvement to current sampling schemes. Finally, this technique presents a bridge toward the enhanced sampling of systems that are far larger than what has previously been considered, greatly extending the applicability of the metadynamics method.

Chapter 2

Exhaustively Sampling Peptide Adsorption with Metadynamics[1]

Introduction

The adsorption of a protein at an interface is a ubiquitous phenomenon with wide ranging applications in biomedicine, biomaterials, chromatography, and adhesion. Considerable progress has been made in developing new technologies that mimic or interface with biological systems with the overall goal of inhibiting or controlling protein adsorption. Realizing the goal of rationally designing the biomolecule/surface interface will require a detailed understanding of the molecular scale behavior that takes place there. Moreover, unraveling such mechanistic details necessitates quantitative structural and thermodynamic information that describes adsorbed proteins. Compared to solution phase (NMR) or crystal state (X-rays) techniques, the experimental toolkit for resolving atomistically detailed structures of biomolecules at surfaces is lacking. This further motivates the use of molecular modeling to complement experiments and provide additional information not readily measurable in the lab. However, as others have already pointed out [8], extending modern molecular modeling techniques to protein adsorption absolutely requires a model system that has been well-characterized experimentally.

DeGrado and Lear originally designed two sequences of leucine and lysine residues (LK peptides) for the purpose of evaluating induced conformational changes upon adsorption. [58] One sequence, LKα14 (LKKL-LKLLKKLLKL), is expected to adopt an α-helical conformation at certain interfaces. In its folded state, polar lysine and nonpolar leucine residues are partitioned to opposite hemicylinders of the helix. Another sequence, LKβ15 (LKLKLKLKLKLKLKL), is a model structure that exhibits β-sheet character. In its

[1] Reproduced in part with permission from M. Deighan and J. Pfaendtner. Exhaustively sampling peptide adsorption with metadynamics. *Langmuir*, 29:7999-8009, 2012. Copyright 2012 American Chemical Society.

ordered state, every other residue is oriented in the same direction, giving LKβ15 amphipathic qualities similar to those of LKα14. The structures of both peptides are shown in Figure 2.1. Extensive work has been done to study the orientation and conformation of LK peptides adsorbed to polar and nonpolar surfaces. Sum frequency generation (SFG) [59–62] spectroscopy and near edge X-ray adsorption fine structure (NEXAFS) [63, 64] spectroscopy experiments have provided convincing evidence that leucine residues preferentially orient themselves toward a hydrophobic surface upon adsorption; whereas lysine residues behave similarly when exposed to a hydrophilic surface. Results from some of these studies and solid-state nuclear magnetic resonance (ssNMR) [65] also suggest that the peptides retain their presumed secondary structures upon adsorption to methyl- or carboxyl-terminated self-assembled monolayers (SAMs). X-ray photoelectron spectroscopy (XPS) and time of flight secondary ion mass spectrometry (ToF-SIMS) experiments reinforce these findings. [66] In the same work, Apte and coworkers observed that leucine and lysine residues predominantly orient toward hydrophobic and hydrophilic surfaces, respectively. Additionally, they showed that the peptides have a stronger affinity for the hydrophilic surface. LK peptide adsorption kinetics have been studied with surface plasmon resonance (SPR); however, the experiment showed that LK peptides adsorb irreversibly to methyl- and carboxyl-functionalized SAM surfaces, preventing the calculation of an adsorption equilibrium constant. [67] What can be inferred from SPR is that LK peptides have a very high affinity for the two SAM surfaces and will adsorb in higher concentrations onto the carboxyl-terminated SAM. This behavior suggests that the free energy of adsorption of a LK peptide onto a carboxyl-terminated SAM is greater than that of the methyl-terminated SAM.

A major drawback to the experimental methods mentioned above concerns the overall lack of atomic resolution in their data. For example, experimental measurements can be used to calculate changes in binding free energies, but provide little to no insight regarding how specific protein/surface interactions influence the adsorption free energy profile. Computer simulations can alleviate this issue by confirming or enhancing experimental results with atomic detail. To date, many simulations have been reported that examine proteins interacting with surfaces of varying structure and functionality. Protein adsorption studies have addressed such topics as: amyloid aggregation on surfaces [68, 69], biomineralization [26, 70], nonfouling surfaces [71, 72], dimer formation upon adsorption [73], protein immobilization [74], adsorption to a patterned surface [75], and other areas of interest related to conformational change in the presence of a surface. [25, 76–80] Regarding LK peptides, Collier and coworkers have recently examined the behavior of LKα14 and two LKβ7 peptides restrained to respective SAM surfaces in a study that addressed the role of empirical force fields in adsorption simulations. [81]

Molecular simulation – molecular dynamics (MD) in particular – could powerfully augment experimental studies if it were more broadly applied. However, in light of the strong binding energy felt at the pro-

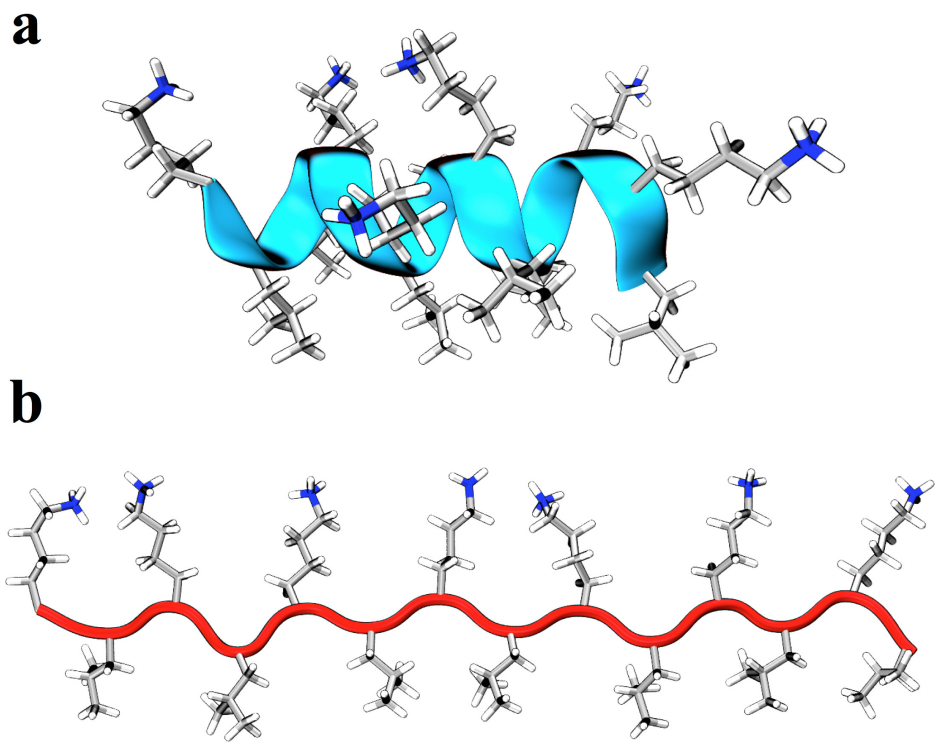


Figure 2.1. The model peptides (a) LKα14 and (b) LKβ15

tein/surface interface as well as the conformational changes expected to occur in a protein, classical MD is severely limited in accessible timescales. Even microsecond long simulations are expected to be lacking in one or more key areas, precluding quantitative evaluation of the free energy change on adsorption and the thermodynamics of surface-induced conformational change. In principle, a Monte Carlo (MC) based approach could be used to sample peptide adsorption, but the high configurational entropy of the peptide in addition to the need for an explicit water model would dramatically reduce its efficiency.

So-called enhanced sampling methods have been successfully used to overcome some of the limitations of MD, particularly those related to simulating protein adsorption. In fact, many of these methods have already been applied to study peptide adsorption. Specifically, these are techniques that bias one or more relevant, coarse degrees of freedom (i.e., collective variables, or CVs) such as steered molecular dynamics (SMD) [26, 82, 83], umbrella sampling (US) [24, 78], or variations of parallel tempering molecular dynamics (PT), which accelerate the sampling of all degrees of freedom in a general way. [25, 77, 81] A detailed review of enhanced sampling approaches in general can be found elsewhere. [27, 84]

The challenges inherent to sampling peptide adsorption are unique. Overcoming the combination of driving forces and energetic barriers inherent to peptide adsorption requires a method that can simultaneously:

1) drive the peptide on and off the surface, 2) quickly explore relevant conformational space, and 3) overcome "hidden" energy barriers that would otherwise inhibit sampling. Few (if any) enhanced sampling methods have been able to accomplish all three of these points when applied to peptide adsorption. Work done by Vellore and coworkers [28] is a good example of a conscientious application of enhanced sampling methods, where they combined a free energy method with PT in order to obtain a more detailed description of their system. Above all, we must emphasize that modeling protein adsorption is not a simple problem, nor is it likely that there exists an elegant solution to it.

There is a substantial amount of important work that is yet to be done in this area of research. Our approach to the challenge of accurately modeling peptide adsorption uses a variant of the Metadynamics method. [40] Metadynamics works by applying a time-dependent bias to one or more collective variables that describe, in reduced dimension, descriptive characteristics of a system (e.g., α-helicity or the distance from a surface). The algorithm biases a MD simulation through the periodic addition of a small repulsive potential (which is Gaussian-shaped and centered on the CV at the time of addition) to the Hamiltonian. This results in a net force that over time influences the system to explore many different states along the CV dimension. The cumulative effect is a dramatic increase in the number of states that are explored within a single MD simulation. Metadynamics, its common variants, and a wide range of applications have been thoroughly discussed in several review articles. [46,85,86] In this work, we used a combination of the parallel tempering and well-tempered metadynamics methods (PTMetaD) [44] to which we recently applied the well-tempered ensemble (WTE) methodology. [49] Our PTMetaD-WTE approach [87] is an effective way to overcome well-known limitations of system size that plague PT-based approaches while increasing the overall sampling efficiency. Herein, we demonstrate that the application of PTMetaD-WTE produces converged, all-atom protein adsorption simulations yielding valuable structural and thermodynamic information. To offer some perspective regarding how a different enhanced sampling method biases protein adsorption, binding energy profiles from PTMetaD-WTE are compared to those of several umbrella sampling simulations. The effectiveness of each method and the importance of sampling technique in simulating protein adsorption are discussed as well. Additionally, we examine the impact the choice of molecular force field has on a simulation – a subject that has also been discussed by Collier and coworkers [81] for older force fields. Three popular force fields (AMBER99SB [52], CHARMM22 with CMAP correction [88], and OPLS-AA/L [89]) were applied to identical systems and subsequently compared to one another. We also evaluate how closely each simulation reproduces experimental observations.

The comparison to experimental observations is included below in the "Side Chain Orientation" subsection, and more detailed results from each force field are provided in Appendix II.

Methods

2.2.1 System Setup

Table 2.1. Physical System Details

System	Force Field	Peptide	SAM	Waters	Box Dimensions (nm^3)
I	AMBER99SB	LKα14	COOH	9402	$7.8 \times 6.5 \times 7.6$
II	AMBER99SB	LKα14	CH_3	9508	$7.8 \times 6.5 \times 7.6$
III	AMBER99SB	LKβ15	COOH	9372	$7.8 \times 6.5 \times 7.6$
IV	AMBER99SB	LKβ15	CH_3	9476	$7.8 \times 6.5 \times 7.6$
V	CHARMM22[a]	LKα14	CH_3	9507	$7.8 \times 6.5 \times 7.6$
VI	OPLS-AA/L	LKα14	CH_3	9507	$7.8 \times 6.5 \times 7.6$

[a] With CMAP correction

A detailed list of system specifications is shown in Table 2.1. In general, a simulation environment consisted of a single LK peptide, a SAM surface, TIP3P waters, and enough sodium or chlorine ions to ensure a net charge of zero. LKα14 and LKβ15 peptides were generated with the psfgen plugin for VMD. [90] The peptides' N-termini were capped with an acetyl group, mimicking their experimental counterparts. [59–67] In order to be consistent with experiment, the protonation state of the peptides was defined according to a physiological pH of 7.4. Lysine residues were simulated in the $NH_3{}^+$ side chain protonation state, and C-termini were simulated in the unprotonated state (carboxylate). The Latour Research Group provides downloadable SAM structures on their website, (`http://www.clemson.edu/ces/latourlabs/Jmol/Surfaces.html`) which we used in our simulations. The two SAM surfaces, composed of dodecane thiol (methyl-terminated) and mercaptododecanoic acid (carboxyl-terminated) chains, were chosen due to their popularity in the literature as model hydrophobic and hydrophilic surfaces, respectively. The SAM surfaces' component alkanethiol chains were organized in $\left(\sqrt{3} \times \sqrt{3}\right)$ R3° geometry with 5 Å spacing. Each chain was tilted 30° from the Z-axis to mimic how alkanethiols pack on a gold surface. [91] The carboxyl-terminated SAM comprised equal proportions of randomly distributed protonated and deprotonated chains, corresponding to a bulk pH of 7.4. [92]

The Gromacs 4.5.5 molecular dynamics engine [51] and the PLUMED 1.3 plugin [53] were used to perform all simulations. System setup began by energetically minimizing a lone, unsolvated SAM surface with a steepest decent algorithm for 5000 steps. Force field parameters for the methyl, carboxylate, and carboxylic acid terminated SAMs were borrowed from AMBER99SB's leucine, glutamate, and glutamic acid residues, respectively. To prevent alkane thiol chains from diffusing into solution or melting at high temperatures, their thiol groups and first 10 CH_2 groups were frozen in place and remained frozen for the entire study. Following

the energy minimization, the SAM was thermally equilibrated at 300K with a global stochastic thermostat for 1 ns. [93] A steepest descent energy minimization was performed on a separate system containing an LK peptide, explicit TIP3P water, and ions. The LK peptide system was then equilibrated at 300K and 1 bar for 1 ns with the same stochastic thermostat and Berendsen barostat. [94] During this step, the X and Y simulation box dimensions of the peptide environment were kept equal to those of the SAM environment. The equilibrated LK peptide system was then grafted above the SAM, energy minimized, and thermally equilibrated once more at 300K for 1 ns with a time step of 2 fs. A sufficient volume of liquid water was provided along the Z dimension (normal to the surface) for an LK peptide to be able to diffuse beyond the 10 Å non-bonded force field cutoff distance of the SAM. A harmonic restraint, placed on the peptide's center-of-mass relative to the Z-axis, prevented any spurious interactions between the peptide and the bottom of the SAM. Generally, a peptide could diffuse to a point about 40 Å away from a surface, or 15 Å from the top of the simulation box. Production simulations with Metadynamics used periodicity in all three dimensions to enable particle mesh Ewald summation for electrostatic calculations.

Ensembles of 25 identical replicas were created for systems I through VI, each set spanning a temperature range of 300K to 500K. Replicas were spaced according to the algorithm developed by Prakash and coworkers [95] and equilibrated before production runs. We also ensured that each of the 25 replicas in each system would initiate from unique configurations. Simulations ran until convergence, which amounted to 50 ns/replica for LKα14 simulations and 75 ns/replica for LKβ15 simulations. The computational cost each LKα14 and LKβ15 simulation was about 100,000 CPU hours and 150,000 CPU hours, respectively. It is important to note that when optimizing computational cost for a PT simulation, it is important to: 1) choose a maximum temperature that is close to the peptide's melting point to avoid sampling uninteresting regions of phase space, and 2) have replicas swap frequently in order to reduce the round trip time of any given replica. [96]

2.2.2 Enhanced Sampling and Free Energy Calculations

Our PTMetaD-WTE simulation protocol is described in our prior work and briefly summarized here. First, a short well-tempered Metadynamics (WTM) [41] simulation is performed using only the system's potential energy (PE) as a collective variable. This increases the overlap in the PE distribution between adjacent replicas, thereby reducing the systems' round-trip-time and concomitantly increasing sampling efficiency. After this, the metadynamics bias is implemented as a static biasing potential on the PE, allowing subsequent simulations to bias other metadynamics CVs. The only change made from our previous work was to use a γ_{WTE} of 20 for preparing the WTE bias potential. This value was chosen to achieve an approximate

exchange success probability of 40% between replicas. It is important to note that without the WTE bias, the exchange probability would have been essentially 0%. Moreover, about 100 replicas would have been needed for a conventional PT simulation to achieve the same exchange probability as the PTMetaD-WTE simulations.

Production runs were performed with a static WTE potential energy bias in effect; its sole purpose being to amplify PE fluctuations. Two collective variables were actively biased during these simulations with WTM, which uses an exponentially decaying bias to improve the speed of convergence of a metadynamics simulation. The bias takes the form of a summation of many Gaussian hills centered on the CV coordinate at the time of addition:

$$V(S,t) = \sum_{n=1}^{t'=n\tau<t} \dot{\omega} \cdot \tau \cdot e^{-\frac{V(S,t')}{k_B \Delta T}} \cdot e^{-\sum_{i=1}^{d} \frac{(S_i(R)-S_i(R(t')))^2}{2\sigma_i^2}} \tag{2.1}$$

The first two factors in Equation 2.11, $\dot{\omega}$ and τ, represent the initial hill height deposition rate ($4\ \mathrm{kJ{\cdot}mol^{-1}{\cdot}ps^{-1}$) and time step between bias additions (0.5 ps), respectively. The third factor in the equation is a bias-dependent exponential that scales the hill height to prevent overfilling the free energy landscape. A virtual temperature difference, ΔT, dictates how quickly the hill height decays anywhere in CV space; it is scaled with the so-called bias factor, $\gamma = (T + \Delta T)/T$, which we set to be 10. The temperature, T, in this case is the temperature of the coolest replica (300K). The fourth factor in Equation 2.1 is a Gaussian hill of width σ that is centered on the ith CV coordinate at time t. More information related to Metadynamics input parameters is available elsewhere. [53] It is important to note that the WTM bias can be reweighted to reproduce the equilibrium Boltzmann distribution. [45] From the reweighted bias, the equilibrium distribution of any other observable in a system can be calculated. We have already demonstrated conclusively [87] that the well-tempered ensemble setup we employ does not change in any way the final free energies that are obtained.

Taking note of SPR experiments [67], we expected the free energy change of adsorption to be quite high. To ensure sufficient sampling of the binding process, we biased the distance along the Z-axis from the peptide's center of mass to the surface plane (Gaussian hill width of 0.1).

Two additional CVs had to be devised for LKα14 and LKβ15, respectively, so that each peptide would be biased to explore relevant conformational space while adsorbing and desorbing. The LKα14 peptide in simulations I, II, V, and VI was biased with respect to its α-helical hydrogen bond contacts. Its collective variable took the form of a switching function that compared a set of backbone hydrogen-oxygen distances (r_i) against a reference distance ($r_o = 2.5$ Å):

$$S_{Hbond}(R) = \sum_{i=1}^{N_{HB}} \begin{cases} 1 & r_i \leq 0 \\ \frac{1-(r_i/r_o)^n}{1-(r_i/r_o)^m} & r_i \geq 0 \end{cases} \tag{2.2}$$

The tuning parameters n and m as well as the hill width were set as 8, 12, and 0.4, respectively. The LKβ15 in simulations III and IV was biased with respect to its ϕ and ψ backbone dihedral angles. Its collective variable measured the dihedral offset describing the general orientation of leucine and lysine residues with respect to one another, and took the form:

$$S_{DH}(\Phi, \Psi) = \frac{1}{2} \sum_{i=1}^{N_{phi}} \left(1 + \cos\left(\phi_i - \phi^{ref}\right)\right) + \frac{1}{2} \sum_{i=1}^{N_{psi}} \left(1 + \cos\left(\psi_i - \psi^{ref}\right)\right) \tag{2.3}$$

The maximum value of Equation 2.3 corresponds to a state where all leucine side chains are pointed in the opposite direction of all lysine side chains. Reference angles ϕ^{ref} and ψ^{ref}, and hill width were defined as -2.36 and 2.36 radians, and 0.1, respectively.

Collective variables for effectively biasing protein adsorption should represent descriptive parameters that distinguish between states of interest (i.e., slow degrees of freedom). As reported elsewhere [44], when applying CV-based methods to study conformational change in peptides or proteins, it is likely that the chosen CVs are not perfect descriptors. In these cases it is possible that nominally hidden free energy barriers exist that hinder the complete exploration of relevant phase space. We previously demonstrated that for a model peptide in water [87], the PT portion of the PTMetaD-WTE algorithm increases overall sampling efficiency. PT played an equally necessary role in this work, given the likely presence of conformational impediments at the peptide/surface interface.

Several umbrella sampling simulations were also carried out and compared to the results obtained from the PTMetaD-WTE simulations. One-dimensional sampling was performed on the Z-distance CV of a system containing a LKα14 peptide and methyl-terminated SAM surface (identical to the system from simulations II, V, and VI). A harmonic restraining potential, $V = \frac{1}{2}k\left(z_{COM} - z_o\right)^2$, acted on the center of mass of the peptide, biasing the peptide to desorb from the surface over the course of several hundred 75 ps US simulation windows. The energy constant k was set as 100 kJ·mol^{-1}·Å^{-2} and the restraint distance was incremented by 0.08 Å from a surface separation distance of 6 to 35 Å. Five simulations were run: two beginning from the adsorbed folded state, and three beginning from the adsorbed unfolded state. The first 5 ps of each umbrella window was considered equilibration time and omitted from our analysis. The weighted histogram analysis method (WHAM) [97] was used to calculate potential of mean force (PMF) curves for

each simulation. We note here that increasing the window time to 300 ps did not appreciably change the US results, validating 75 ps as an acceptable window time. We concluded that the same degree of conformational change is accessible on either timescale.

2.2.3 Clustering Algorithm

There is an inherent degeneracy associated with some collective variables due to their nonspecific, combinatorial nature. We observed this degeneracy in the conformational CVs that were biased in each simulation, motivating us to perform an iterative clustering analysis on peptide structures in the adsorbed state. The clustering algorithm is carried out first by drawing a random sample of 5000 structures from a parent set. A reference structure is selected from the set, to which all 5000 structures are subsequently aligned. The structures are subjected to a 2D C_α alignment [98] to the reference structure in order to preserve the orientation of each peptide with respect to the surface. The Gromacs tool g_cluster then sorts the set of 5000 according to the gromos clustering algorithm [99] using a C_α RMSD cutoff of 1.5 Å. The gromos algorithm selects cluster centers by systematically identifying the structures in a set with the most neighbors within the RMSD cutoff. Once the largest center has been identified, it and its neighbors are defined as a cluster and excluded from future comparison or assignment. The selection process is performed until every structure is assigned to exactly one cluster. Keeping with the constraint that a structure may only belong to one cluster: while the logic used by gromos guarantees that each cluster center is beyond the RMSD cutoff of all other cluster centers, it is possible that some constituent structures may be incorrectly assigned to their clusters. In other words, although a structure may fall within the cutoff of one cluster center, it may more closely match another cluster center that is defined later by the gromos algorithm. To address this, we retain the cluster centers defined by gromos and reevaluate cluster membership on a case-by-case basis according to the same C_α RMSD cutoff of 1.5 Å. During the reevaluation, assignments are made for the structures from the 5000-member sample in addition to the remainder of the parent set. Structures that do not fall within the cutoff of any cluster center are extracted and subjected to a new iteration of the entire algorithm. Iterations are performed until all structures are assigned to a cluster. Following the final iteration, the Metadynamics bias is reweighted [45] to yield the equilibrium Boltzmann distribution of the clusters, from which low free energy conformations can be identified.

Results and Discussion

2.3.1 Convergence of Metadynamics Calculations

For methods such as Metadynamics that apply a time-dependent bias, it is essential to assess the convergence of the calculations. Unlike the case of an idealized model system (e.g., alanine dipeptide) – whose reference free energy is known and can serve as a quantitative check – it is extremely difficult to obtain an exact reference free energy for the adsorption of a peptide that may fold and unfold. Therefore, we assessed the convergence of our simulations with an approach that we previously applied to biased folding with metadynamics [87, 100] and observed the following: 1) full exploration of physically interesting regions of CV space, 2) the free energy difference between adsorbed and solvated states converges, and 3) throughout each simulation, the peptide trajectories along the CV dimension show diffusion through states of interest. Figure 2.2 shows the free energy difference between adsorbed and desorbed states for our simulations as a function of time. A peptide was considered to be completely adsorbed if its center of mass was less than 11 Å away from the surface; it was considered solvated if its center of mass was farther than 26 Å from the surface. We also observed dozens of folding/unfolding and adsorption/desorption events in each of our simulations. Given this substantial body of corroborating evidence, we are confident that PTMetaD-WTE enables fast convergence when applied to these systems. The calculated free energy differences also follow the expectation laid out in previous experiments [67] that both LK peptides have a higher affinity for the hydrophilic, carboxyl-terminated SAM.

2.3.2 Reweighted FES and Clustering Analysis of I-IV

Figure 2.3 shows the reweighted free energy surfaces of the original biased CVs from simulations I through IV. The adsorbed state is clearly visible as a deep free energy minimum at a distance between 6 and 11 Å from the SAM surface; we use this observation to define the cutoff distance for the adsorbed state. LKα14 occupies multiple stable states at the surface in simulations I and II. When comparing the two simulations, it becomes apparent that the hydrophobic surface of II promotes a tight adsorption interaction for all bound structures, depicted by the steep free energy gradient between the adsorbed and solvated states that spans all hydrogen bond values In contrast, the hydrophilic surface exhibits a gentler gradient and affects the binding energy profile almost a full nanometer away from the surface, even for the folded peptide. The reason for this behavior can be understood by considering how the different surface functional groups interact with the leucine and lysine residues of LKα14. Tight association between LKα14 and the hydrophobic methyl termini of the SAM in simulation II is directly related to the entropic maximization that occurs when the

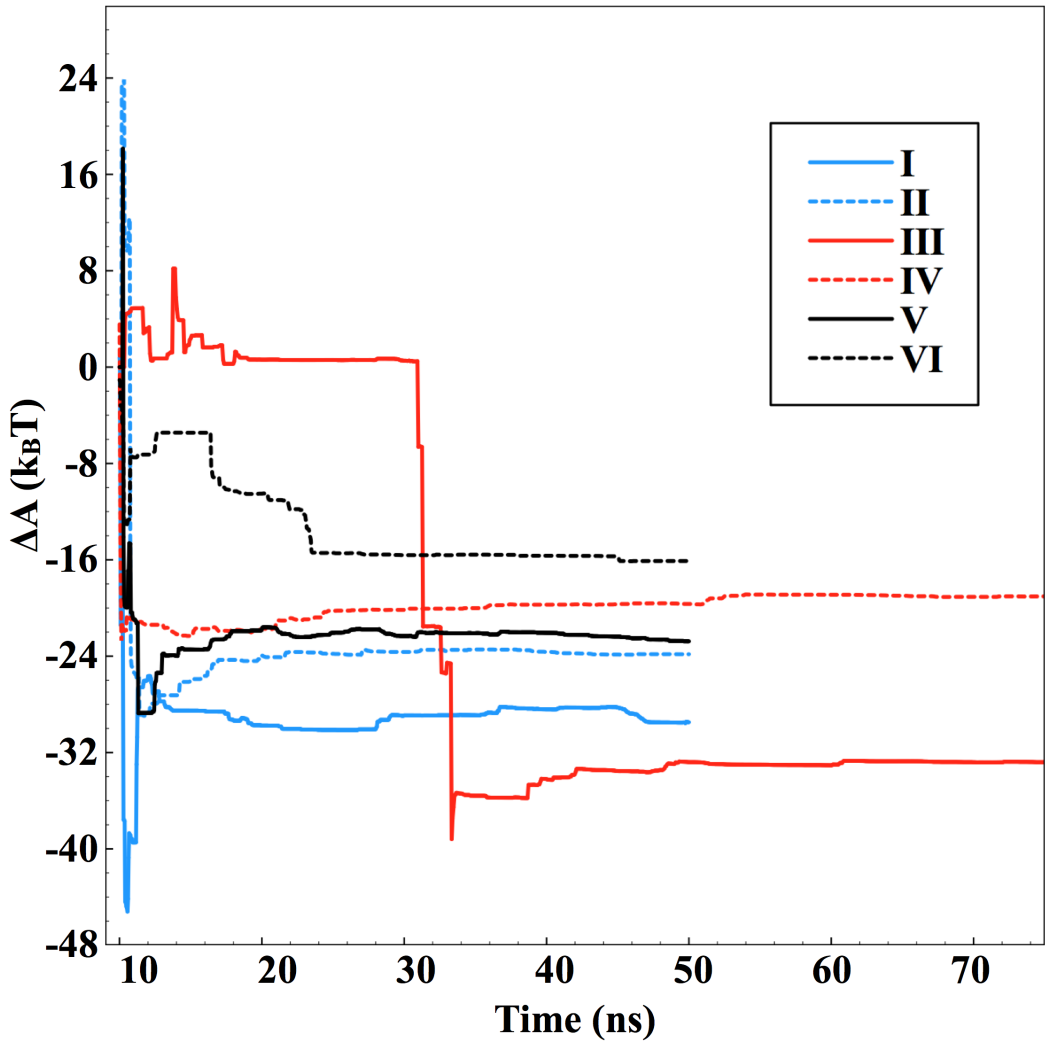


Figure 2.2. Time series of the free energy difference between the adsorbed and solvated states at 300K, showing the convergence of (**I**) LKα14 on carboxyl SAM, (**II**) LKα14 on methyl SAM, (**III**) LKβ15 on carboxyl SAM, (**IV**) LKβ15 on methyl SAM, (**V**) LKα14 on methyl SAM with CHARMM22/CMAP, and (**VI**) LKα14 on methyl SAM with OPLS-AA/L.

peptide's leucine residues interact with the surface, freeing loosely bound surface waters. Leucine side chains and methyl groups are nearly charge neutral, explaining their strong affinity for each other at the interface. The absence of Coulombic interactions between the leucines and the surface is the most likely reason why the surface free energy trough of simulation II is shallower than that of simulation I. Given these observations, it becomes apparent that LKα14 can adopt a more tightly folded conformation with greater ease on the methyl surface compared to the carboxyl surface. In this case, intramolecular forces (predominantly hydrogen bonds) within LKα14 play a more significant role in influencing the ensemble of accessible equilibrium states that the peptide explores at the hydrophobic surface, acting against only surface van der Waals forces instead of van der Waals and Coulombic forces. An explanation of the FES extracted from simulation I can be carried

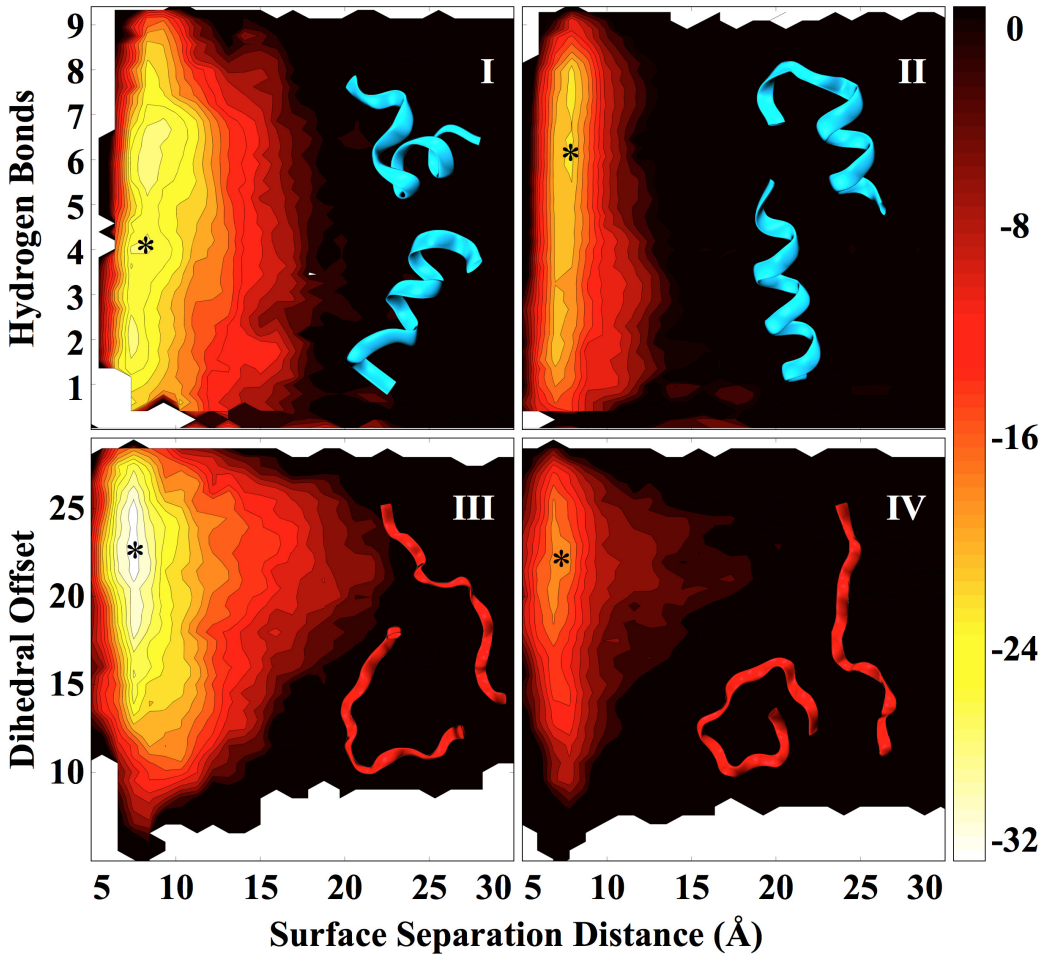


Figure 2.3. Free energy surfaces of the biased Metadynamics collective variables for simulations (**I**) LKα14 on carboxyl SAM, (**II**) LKα14 on methyl SAM, (**III**) LKβ15 on carboxyl SAM, and (**IV**) LKβ15 on methyl SAM. Asterisks mark the regions of free energy space from which corresponding representative structures were drawn. The colorbar is in units of k_BT with isolines spaced every $2k_BT$.

out along the same vein. Negatively charged carboxyl head groups attract positively charged lysine residues to the surface, and the Coulombic forcesbetween the two groups contribute to the deeper surface free energy trough. Solvent effects in this case are the most likely reason for the softer gradient between the adsorbed and solvated states, as electrostatic forces hold a tight boundary layer of water at the SAM/liquid interface. This phenomenon is directly observable from a comparison of trajectory snapshots of the solid/liquid interfaces from simulations I and II. In these frames, we were able to see waters bound 1 Å closer to the carboxyl surface than to the methyl surface.

In comparison to simulations I and II, the free energy minima shown in simulations III and IV are broad swaths of CV space that are also located between 6 and 11 Å from the surface. At a glance, the LKβ15 simulations appear to share some similarities with the LKα14 ones. The FES of simulation IV

(LKβ15 and methyl SAM) is a shallow well with a steep gradient between the adsorbed and solvated states, although the gradient is not as steep as that of simulation II. Simulation III (LKβ15 and carboxyl SAM) has a deeper free energy well than IV at the surface with a gentler gradient leading away into solution. The defining characteristics of the two simulations can also be attributed to the hydrophobic effect. An absence of Coulombic interactions between the leucine residues and methyl head groups of simulation IV should account for its comparatively shallower FES. Water has a higher affinity for the carboxyl surface head groups in simulation III and most likely forms a barrier that impedes rapid adsorption of LKβ15 onto the surface, similar to simulation I.

The FESs of systems III and IV in Figure 2.3 suggest that that LKβ15 conforms to a single preferred state upon adsorption. However, a random selection of structures lying within k_BT of the free energy minima of each FES revealed that many different structures populate the adsorbed state. Examples of these structures are shown in Figure 2.3, where the regions of free energy space from which they were extracted are marked with asterisks. In a broad sense, the conformational CVs are effective at influencing the peptides to explore a wide range of extended and collapsed states, but are clearly unable to resolve finer structural details. To address this, we carried out a reweighted clustering analysis on systems I through IV in order to elucidate low free energy peptide conformations in the adsorbed state. Only structures whose centers of mass fell within 6 and 11 Å from the surface were considered and subjected to theclustering algorithm described above. Many clusters were identified within each dataset (I: 270, II: 365, III: 813, IV: 1051) and, as expected, more than one type of structure populates the free energy minima in Figure 2.3. Moreover, we observed clusters that span large areas of CV space; some of which contained structures that occupy different free energy minima. This can be explained by the combinatorial nature of the hydrogen bond and dihedral offset CVs for LKα14 and LKβ15, respectively. In other words, small differences in certain structural features, which may not greatly affect a structure's RMSD from its cluster center, can result in a much larger shift along a CV coordinate.

The structural degeneracy along both conformational CV coordinates made it impractical to identify cluster-occupied regions on the FESs in Figure 2.3. To provide a better visual representation of the general structure of the adsorbed peptides, and an idea of the distribution of α and β character at a surface, we carried out an analysis of the peptides' backbone dihedral angles. First, ϕ and ψ backbone dihedral angles were calculated from trajectory data for peptides whose surface separation distances were between 6 and 11 Å (i.e., the adsorbed state). The angles were subsequently reweighted to produce the Ramachandran free energy surfaces shown in Figure 2.4, all of which converged after roughly the same amount of sampling time as their counterparts in Figure 2.3. Clusters were also subjected to the reweighting algorithm in order to find low free energy structures in the adsorbed state. The five lowest free energy cluster centers are shown

for systems I and II in Figure 2.5, and systems III and IV in Figure 2.6. Additionally, (ϕ, ψ) dihedral angle pairs were calculated for each cluster center and are represented as dots that overlay the Ramachandran plots beside the structures in Figure 2.5 and Figure 2.6.

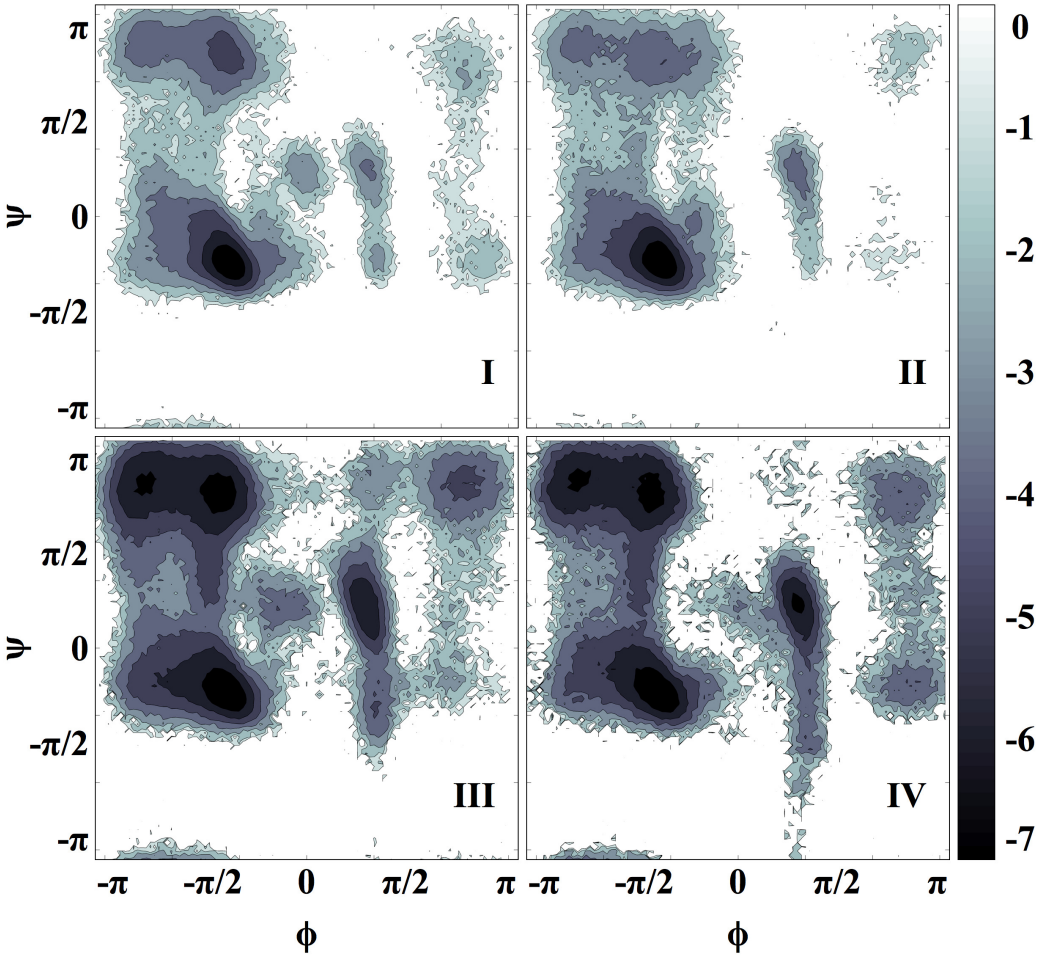
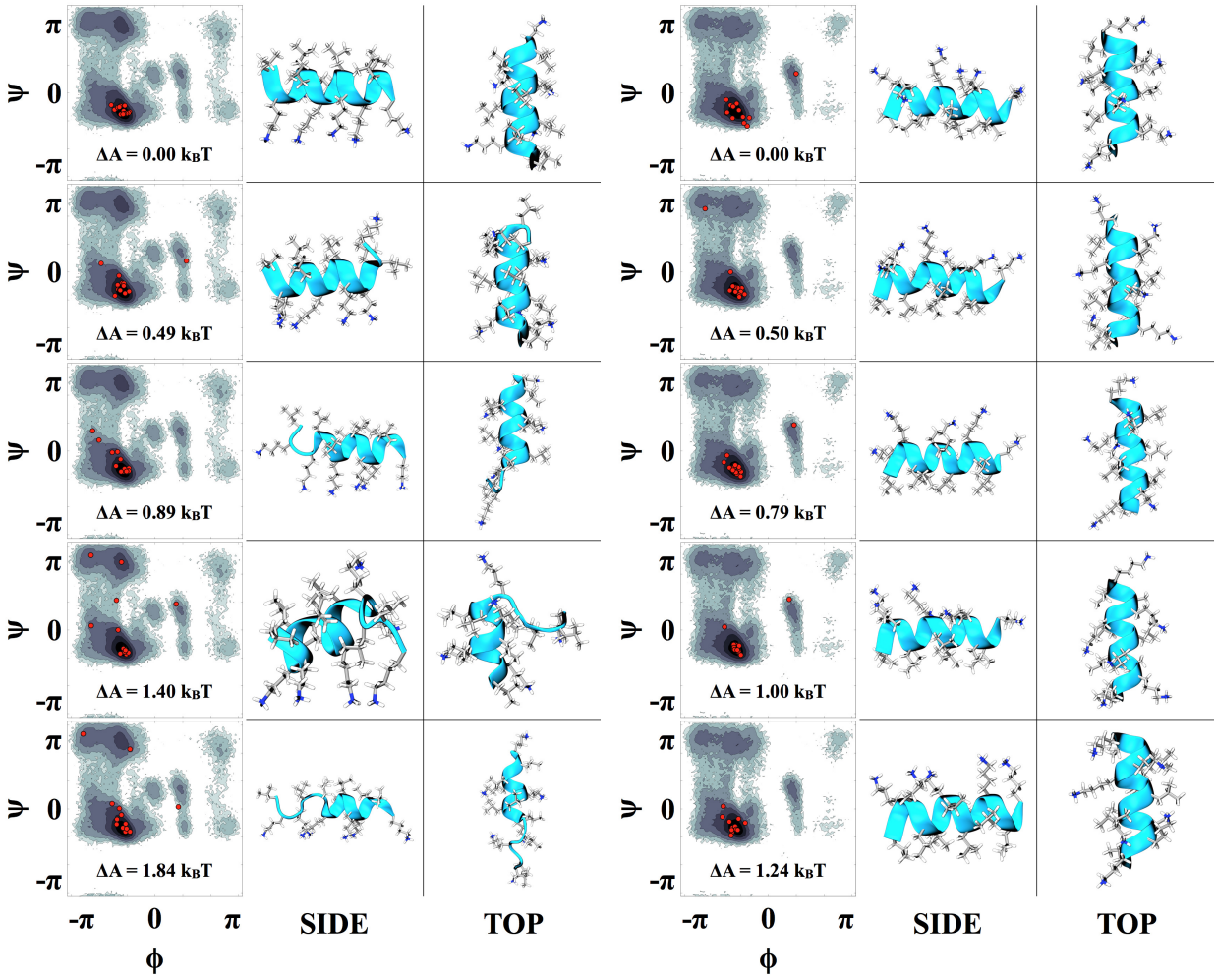


Figure 2.4. Reweighted Ramachandran free energy surfaces for adsorbed structures from simulations (**I**) LKα14 on carboxyl SAM, (**II**) LKα14 on methyl SAM, (**III**) LKβ15 on carboxyl SAM, and (**IV**) LKβ15 on methyl SAM. The colorbar is in units of k_BT with isolines spaced every k_BT.

The peptides in systems I and II are preferentially α-helical in the adsorbed state; however, the methyl-terminated surface of system II allows LKα14 to assume a wider variety of tight α-helical conformations. The contrast between the two systems can be explained by the presence and absence of Coulombic interactions. Comparatively weaker binding forces allow the peptide to be more flexible and mobile at the hydrophobic interface, resulting in the exploration of a greater number of conformational states (delineated by the cluster RMSD cutoff). Similar behavior can be inferred for systems III and IV. Although most of the clusters observed in systems III and IV are coils or β-like, sometimes with helical character, the structures observed in system III generally adopt conformations that are more extended than those in system IV. This

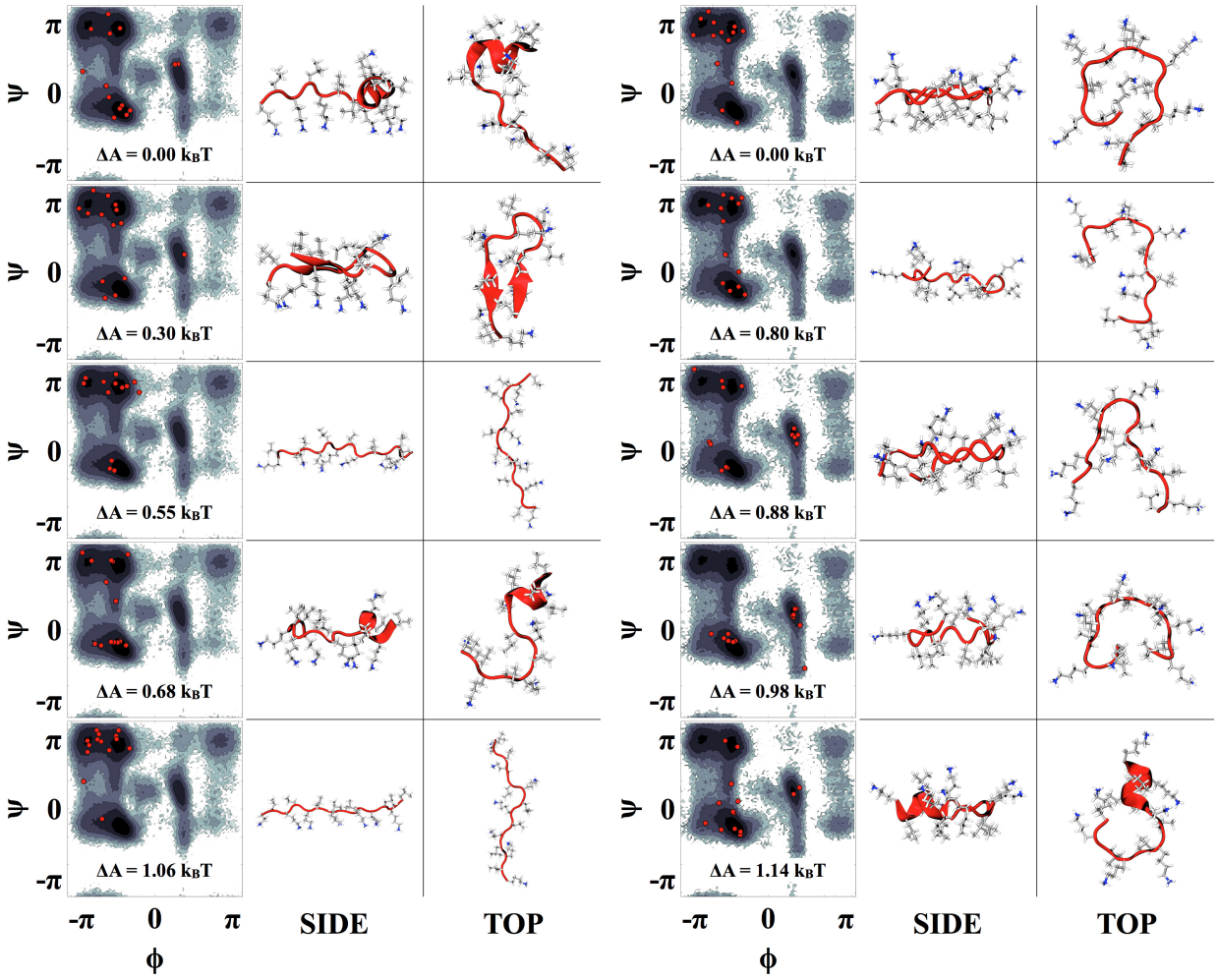


(a) LKα14 on a COOH-terminated SAM

(b) LKα14 on a CH_3-terminated SAM

Figure 2.5. The most probable adsorbed conformations of LKα14 found in (**a**) simulation I , and (**b**) simulation II, and their associated free energy differences (with respect to the lowest free energy state) calculated from a reweighted clustering analysis. Left column: cluster center (ϕ, ψ) dihedral angles (shown as red dots) overlaid on the Ramachandran FES. Center column: side view of the cluster center structure. Right column: top view of the cluster center structure.

behavior can be attributed again to the greater mobility of leucines on the methyl surface. Surface van der Waals forces of the methyl-terminated SAM are weak enough to allow the LKβ15 peptide to easily transition between extended and compact states, whereas the comparatively stronger electrostatic forces of the carboxyl-terminated SAM stabilize peptide conformations that are less entropically favorable. Additionally, as we discuss below, the structures' orientations in the adsorbed state generally agree with what has been observed experimentally.



(a) LKβ15 on a COOH-terminated SAM

(b) LKβ15 on a CH_3-terminated SAM

Figure 2.6. The most probable adsorbed conformations of LKβ15 found in (**a**) simulation III , and (**b**) simulation IV, and their associated free energy differences (with respect to the lowest free energy state) calculated from a reweighted clustering analysis. Left column: cluster center (ϕ, ψ) dihedral angles (shown as red dots) overlaid on the Ramachandran FES. Center column: side view of the cluster center structure. Right column: top view of the cluster center structure.

2.3.3 Side Chain Orientation

Separate works by Weidner and Apte [63, 64, 66, 101] describe side chain orientation of LK peptides at hydrophobic and hydrophilic surfaces. There is a strong consensus that leucine residues will predominantly associate with hydrophobic surfaces, as lysine residues will with hydrophilic surfaces. LKα14 in its folded state is therefore expected to orient itself upon adsorption in such a way that a preferred hemicylinder (either leucine- or lysine-rich) comes into contact with the surface. The same reasoning can be applied to LKβ15, where only one kind of residue is expected to interact with the surface, leaving the remaining side chains to point into solution. In order to assess these orientation preferences from our own simulations, a new order

parameter had to be defined to describe side chain orientation with respect to the surface normal. The direction in which a side chain is oriented can be calculated from the coordinates of two atoms in a residue: one from the α-carbon, and the second from an atom on the side chain. For our purposes, the side chain atoms were selected to be leucine's γ-carbon and lysine's side chain nitrogen. The two atoms' positional information is substituted into the following equation:

$$S_O = \frac{(z_i - z_{C\alpha i})}{\sqrt{(x_i - x_{C\alpha i})^2 + (y_i - y_{C\alpha i})^2 + (z_i - z_{C\alpha i})^2}} \tag{2.4}$$

The orientation order parameter spans from -1 (oriented toward the surface) to +1 (oriented away from the surface). Leucine and lysine orientation values were calculated for each frame in the simulations' respective trajectories. Figure 2.7 shows free energy surfaces of peptide side chain orientation versus the surface separation distance that were calculated by reweighting the Metadynamics bias from simulations I through VI. The orientation free energy surfaces agree well with experimental observations in that there appears to be a clear preference for leucine and lysine residues to interact with hydrophobic and hydrophilic surfaces, respectively. Positively charged lysine side chains strongly prefer to interact with the negatively charged carboxyl-terminated SAM. Furthermore, when exposed to a methyl-terminated SAM, lysine residues are more inclined to associate with the solvent, whereas leucine residues prefer to interact with the surface. Farther away from the surface in each system, the orientation for both residues shifts toward an even distribution, indicating random side chain orientation. Past 26 Å, side chain orientation becomes completely random. At this point we consider the peptide to be beyond the influence of the surface, and therefore in the solvated state. This observed quality validates the assumption we make above where we define the solvated state as the region beyond a surface separation distance of 26 Å. The disoriented nature of the side chains in the solvated state also demonstrates that our simulations have explored peptide adsorption from a random distribution of orientations in the solution state.

While the distribution of LKα14 surface conformations differ between simulations II, V, and VI (see Appendix II), leucine residues are predominantly oriented toward the methyl-terminated SAM surface in each case. Figure 2.7 shows reweighted leucine and lysine orientations that qualitatively agree with experimental observations. Recently, Weidner and coworkers [101] used SFG and ssNMR to study a similar system consisting of the same LKα14 peptide adsorbed to a polystyrene surface. They reported tilt angles of each leucine's C_β-C_γ bond with respect to the surface normal. The same angles were measured and reweighted from simulation II, V, and VI trajectory data for structures that appeared in the adsorbed state. Although we are comparing leucine side chain orientations at different surfaces, we expect to see some similarity between the

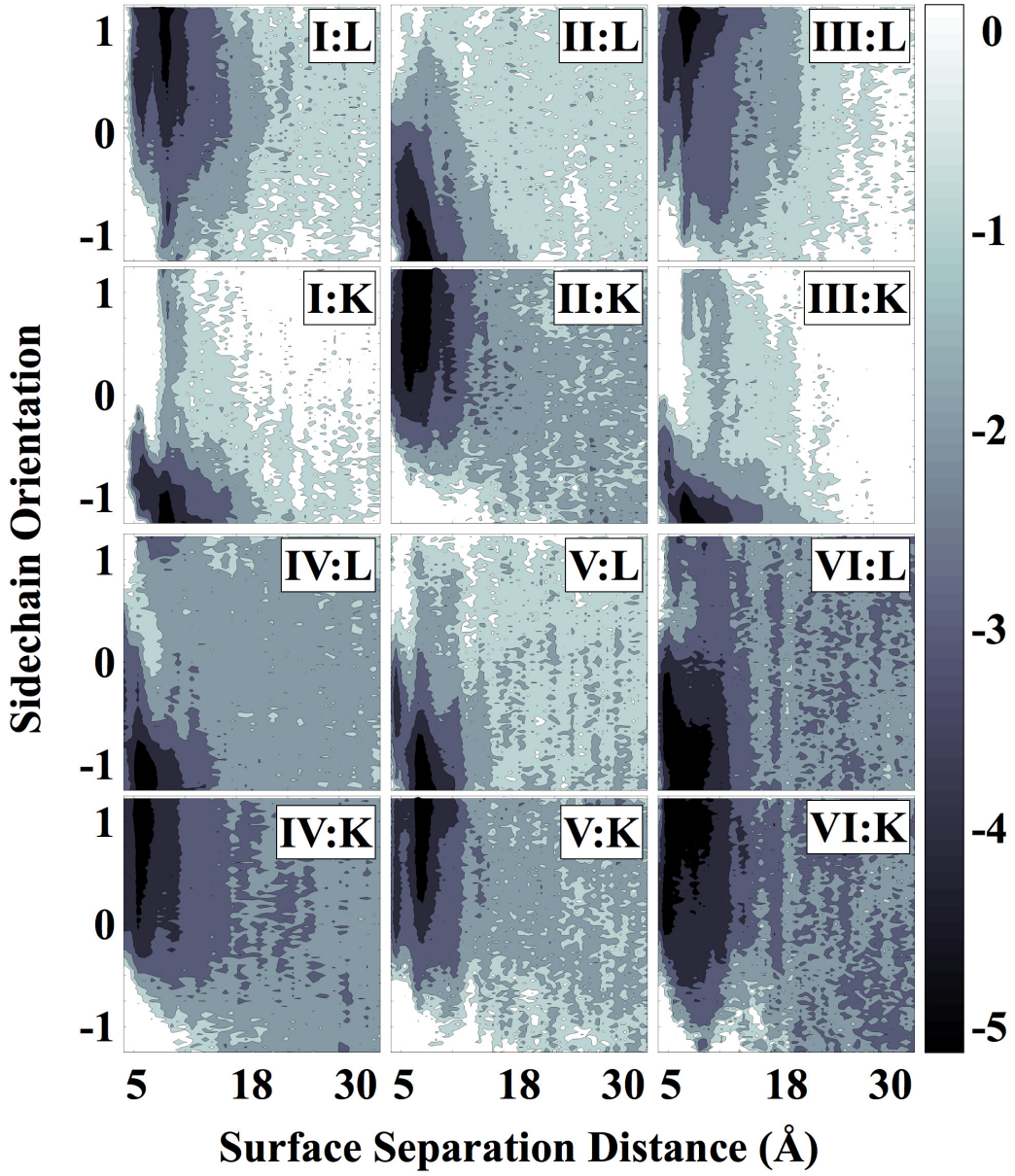


Figure 2.7. Free energy surfaces of leucine (L) and lysine (K) side chain orientations with respect to the peptide center of mass distance from the surface. Simulations shown here are (**I**) LKα14 on carboxyl SAM, (**II**) LKα14 on methyl SAM, (**III**) LKβ15 on carboxyl SAM, (**IV**) LKβ15 on methyl SAM, (**V**) LKα14 on methyl SAM with CHARMM22/CMAP, and (**VI**) LKα14 on methyl SAM with OPLS-AA/L. The colorbar is in units of k_BT with isolines spaced every k_BT.

two environments. Table 2.2 shows the experimentally derived angles alongside the reweighted angles from simulations II, V, and VI. The difference between experiment and simulation appears under the columns marked with a Δ. Leu1, Leu5, and Leu8 angles from each simulation deviate the least from experiment, whereas the average orientations of Leu7, Leu11, and Leu12 deviate the most from the experimental measurement. Of the three force fields, AMBER99SB deviates the least from experimental data. We note that

developing a strong link between theory and experiment is critical for characterizing complicated systems such as these. Although our simulation generally agrees with experiment, it does not completely validate the model. However, the comparison we make does give weight to the overall validity of our observations; namely, the adsorbed structures we observe and their orientations.

Table 2.2. Leucine Tilt Angles Derived from SFG Spectroscopy[a]; Corresponding Average Tilt Angles from Simulations II, V, and VI; and Their Differences (Δ).

Residue	Tilt Angle (deg)	II	Δ_{II}	V	Δ_{V}	VI	Δ_{VI}
Leu1	70	70.4	0.4	67.1	2.9	74.3	4.3
Leu4	47	53.2	6.2	56.4	9.4	81.9	34.9
Leu5	68	66.5	1.5	62.0	6.0	69.9	1.9
Leu7	39	64.4	25.4	77.9	38.9	80.8	41.8
Leu8	48	54.6	6.6	52.0	4.0	54.9	6.9
Leu11	30	69.6	39.6	65.7	35.7	85.8	55.8
Leu12	82	58.7	23.3	51.6	30.4	52.2	29.8
Leu14	74	74.9	0.9	40.3	33.7	49.8	24.2
Average Δ			**13**		**20**		**25**

[a] Experimental values come from work done by Weidner and coworkers. [101]

2.3.4 Comparing Methods

If we consider the simplified case of an ion or small molecule adsorbing to a surface, umbrella sampling would seem comparable (or perhaps preferable) to Metadynamics as the method of choice for an enhanced sampling simulation. However, if the molecule of interest is capable of undergoing complex conformational changes, then the computational effort required to reach convergence is an important factor to consider. US biases a system along one or more CVs like Metadynamics, but it samples in discrete windows that typically last for a nanosecond or less. This poses a serious challenge for studying the adsorption of flexible molecules such as proteins, which fold on the order of microseconds to seconds. The US method has been used by other groups to study the adsorption (but not folding) of biomolecules. For example, the adsorption of carbohydrate binding module to cellulose [78] and of flounder antifreeze protein to ice. [24]

We performed a series of US simulations whose results were compared to Metadynamics. For consistency, we chose to base our comparison off the system containing an LKα14 peptide and methyl-terminated SAM surface. Sampling was performed along the Z-distance collective variable from the adsorbed state (6 Å) into the solvated state (35 Å). Five US simulations were run, two of which were initiated from the folded adsorbed state and the remaining three were initiated from the unfolded adsorbed state. Particular details of

each simulation are as follows: (1) folded and restrained to the folded state; (2) folded and unrestrained; (3) unfolded and restrained to the starting state; (4 & 5) unfolded and unrestrained. The restraint implemented in simulations (1) and (3) was placed on the peptide's C_α radius of gyration. The purpose of the restraint was to investigate whether conformational changes in the peptide led to changes in the final PMF from the US calculations.

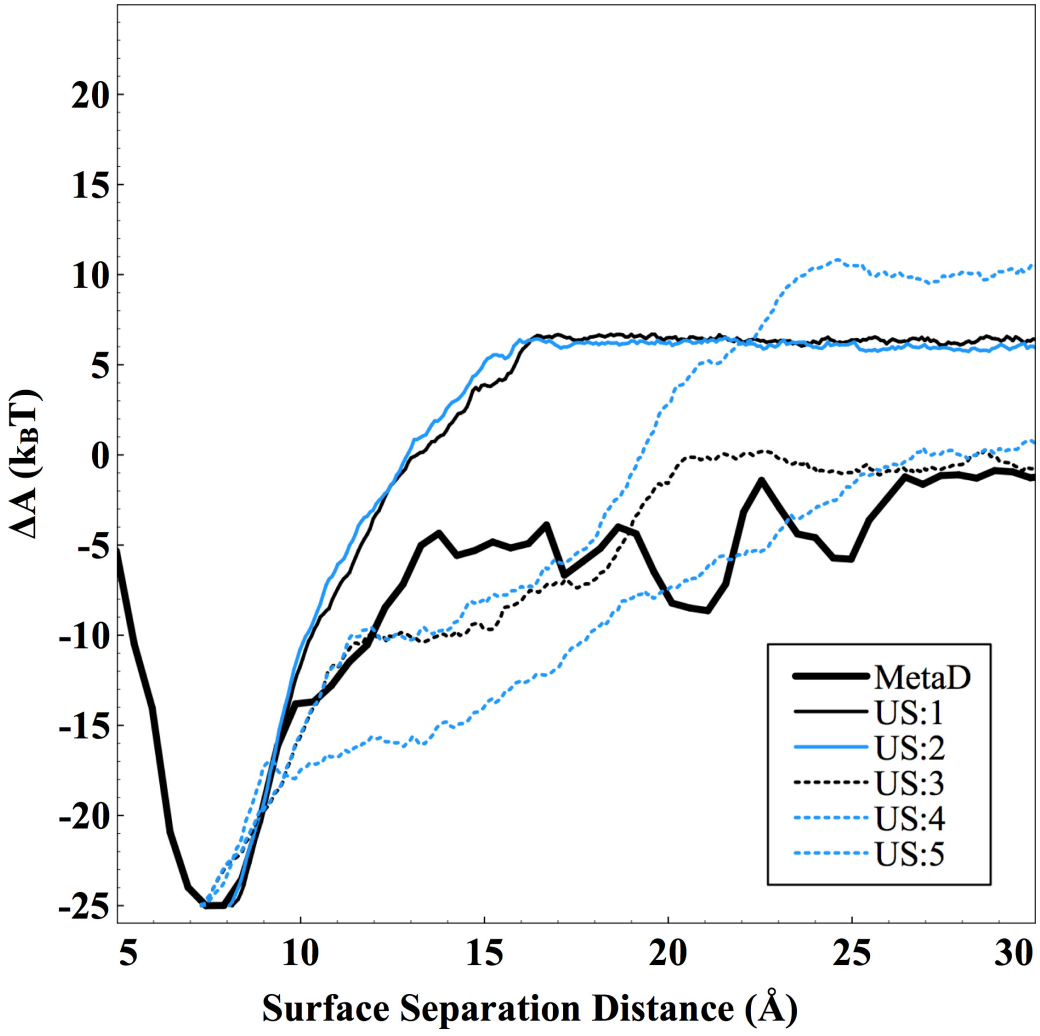


Figure 2.8. Binding free energy curves from Metadynamics (MetaD) and the Umbrella Sampling simulations where the peptide was: (**US:1**) folded and restrained, (**US:2**) folded and unrestrained, (**US:3**) unfolded and restrained, and (**US:4 & US:5**) unfolded and unrestrained.

Figure 2.8 shows the PMF profiles of the Z-distance CV for each US simulation compared to a reweighted Metadynamics free energy surface from simulation II. Simulations initiated from the folded state all exhibited very similar behavior, indicating that 75 ps was enough time for a US window to equilibrate and also that restraining the peptide's radius of gyration has little to no effect on sampling. In fact, although some unfolding occurred in both simulation (2) and the 300 ps long window simulaton, the helix reformed within

a few picoseconds, demonstrating that the peptide did not explore conformational space beyond the folded state. A greater degree of variability is evident in the simulations initiated from the unfolded state, however. Each of these simulations produced PMF curves that were significantly different from one another, both in slope and free energy difference between the adsorbed and solvated states. Simulation (3) produced a PMF that is somewhat similar to the Metadynamics free energy surface. To check the legitimacy of the simulation (3), a second restrained simulation, initiated from the unfolded state, produced a PMF that deviated significantly from the Metadynamics curve. A lack of strong agreement between the folded and unfolded simulations, as well as among the unfolded simulations themselves, is further evidence that US does not sample enough conformational space to produce a well-converged and reliable binding free energy profile. We note that if US were performed on an additional conformational CV, one would expect to reproduce the Metadynamics profile. However, such a simulation would have the same computational cost as one of our PTMetaD-WTE simulations, and would ultimately yield less information because it would only produce ensemble data for a single temperature.

Conclusions

In this study, the Parallel Tempering Metadynamics in the Well-Tempered Ensemble (PTMetaD-WTE) enhanced sampling algorithm facilitated the exploration of many adsorption and conformational states of two LK peptides. Previously shown to have effectively sampled the behavior of a fast-folding miniprotein [87], our algorithm has now produced ensemble data that is consistent with experimental observations. We emphasize that the well-tempered ensemble (WTE) greatly reduced the computational expense of our simulations, while enabling an exchange success probability of 40% between 25 replicas. In order to achieve the same exchange probability for a conventional parallel tempering (PT) simulation, the number of replicas would have to be quadrupled. We also demonstrate that while PTMetaD-WTE and umbrella sampling appear to be similar sampling techniques, the former is able to more effectively bias a system to explore a large number of conformational states at different temperatures.

We report strong qualitative similarities between low free energy structures extracted from simulation and experimental observations. In simulations I through IV, LKα14 predominantly adopts an α-helical conformation on the self-assembled monolayer (SAM) surfaces; whereas, LKβ15 tends to self-associate in a loop or coiled structure on a methyl-terminated surface but is more likely to adopt extended, less entropically favorable conformations at a carboxyl-terminated surface, where it is also capable of forming a β-sheet structure. Peptides that adsorbed to the hydrophobic methyl-terminated SAM were able to explore a greater number of surface conformations compared to those adsorbed to the hydrophilic carboxyl-terminated SAM.

We attribute the entropy difference between these two systems to Coulombic interactions, or a lack thereof. Coulombic forces restrict lateral movement for peptides adsorbed to the charged hydrophilic SAM, whereas their absence permits a greater degree of flexibility in peptides adsorbed to the hydrophobic SAM.

In three simulations the AMBER99SB, CHARMM22/CMAP, and OPLS-AA/L force fields were applied to an explicitly solvated system containing a LKα14 peptide and methyl-terminated SAM. Qualitatively, the AMBER99SB and CHARMM22 simulations produce low free energy peptide conformations in the adsorbed state that are more consistent with experimental observations than OPLS-AA/L (see Appendix II). Regarding CHARMM22, this agrees with a recent study done by Collier and coworkers. However, AMBER99SB yielded structural data that deviated the least from sum frequency generation (SFG) spectroscopy measurements of a similar system reported by Weidner and coworkers. [101] We expect this work will contribute to a growing body of literature on the use of empirical force fields for modeling protein/surface interfaces. If the predictive capability of these simulations is to improve, we must continue to explicate where and how classical force fields fall short. Effective and efficient sampling will play a critical role in this process, which PTMetaD-WTE has been shown to carry out in a computationally inexpensive manner. Future priorities for enhanced sampling simulations of peptide or protein adsorption include: 1) In general, to perform more simulations that exhaustively sample protein behavior at a variety of solid-liquid interfaces, 2) To run identical simulations with several different force fields, and most importantly 3) To model systems whose results can be quantitatively compared to specific experimental observables. [29]

Chapter 3

Eliciting a Detailed Picture of the SN15/Hydroxyapatite Interface with Metadynamics, Sum Frequency Generation Spectroscopy, and Existing Literature Data

Introduction

Biomineralization – the growth and development of mineral structures in and/or around an organism – is a process that is largely controlled by proteins. [102, 103] Hard tissues such as nacre, tooth enamel, and bone are examples of highly specialized materials that exist naturally due to underlying protein-mediated mechanisms. The morphology of these substances is determined by a complex orchestration of protein dynamics that ostensibly takes place at the solid/liquid (i.e., mineral/biofluid) interface. Understanding the behavior of proteins that govern biomineralization is of key interest to the fields of medicine and dentistry because of its relation to some osteological, cardiovascular, and neurological disorders (e.g., the calcification of soft tissues). [104, 105] Mechanistic details surrounding the structure-function relationship of these proteins, and how they interact with their environment, are also extremely relevant for designing new materials. Improved molecular-level knowledge of the biomineralization process could spur the development of new fabrication methods and inspire a range of novel materials that interface with biological environments in different ways. [106]

While several hypotheses exist that attempt describe biomineralization at the cellular level, little is known or understood about the process at the atomic scale. [107] Protein recognition, binding, and conformational change at the mineral surface are examples of complex events that require more attention. A simple system that has been extensively characterized is the salivary protein statherin in the presence of hy-

droxyapatite(HAP). [108–110] Statherin is a 43-residue enamel pellicle protein that is involved in lubricating tooth enamel, mediating bacterial adhesion, and preventing calcium phosphate from precipitating in saliva; it has the primary structure:

DpSpSEEKFLRRIGRFGYGYGPYQPVPEQPLYPQPYQPQYQQYTF

where "pS" denotes a phosphorylated serine residue. In the solvated state, statherin is not known to adopt any persistent secondary structure, but does undergo measurable conformational change in the presence of a hydroxyapatite surface. The surface-induced secondary structure of statherin, and the role it plays in surface recognition, binding, and folding, have been the foci of several surface analysis studies.

Early work on this system inferred that the first 15 N-terminal residues (SN15) bind and fold into an α-helical conformation at the HAP surface. [111] The current scheme describing this mechanism involves the first six charged residues (D1, pS2, pS3, E4, E5, and K6) strongly binding to the HAP surface. [112–114] Adsorption assays, solid-state nuclear magnetic resonance (ssNMR), sum frequency generation (SFG) spectroscopy, and near edge X-ray absorption fine structure (NEXAFS) spectroscopy experiments have also shown residues R9, R10, and F14 to have a weak affinity for a hydroxyapatite surface. [115, 116]

Folding ostensibly occurs during a recognition-and-binding event at the HAP surface, but a detailed explanation of the mechanism is still unknown. Solid-state NMR measurements of individual residues in the SN15 sequence suggest that the peptide assumes a loose α-helical conformation in the adsorbed state. [113, 117, 118] Dipolar recoupling with a windowless sequence (DRAWS) analysis of the phi dihedral angles of residues pS3, L8, and G12 suggest that the first few residues of SN15 are only slightly helical, whereas the C-terminus – which does not interact as strongly with the HAP surface – is thought to hold a more conclusive helical structure. Residues F7 and F14 have also been observed to orient in different directions, where F7 points into solution and F14 points toward the surface. [116, 119] This observation conflicts with the arrangement of an ideal α-helix, for which residues i and $i+7$ are aligned. Residues E4 and E5 have also been reported to be positioned farther from and closer to the HAP surface, respectively, which supports the distorted helix hypothesis. [120]

Currently, entire protein structures can only be resolved in solution with NMR or the crystal state with X-ray crystallography. Although experimental methods are capable of educing a number of molecular details that describe the biomolecule/surface interface, none can provide a complete picture of these systems. Similarly, while the structure of statherin – and, by association, the SN15 domain – at the HAP surface has been extensively characterized, its entire surface-bound structures remains unclear to a certain degree. The development of a viable method aimed at resolving the structure and behavior of an adsorbed biomolecule is an active research area – one that is challenging and for which there may not be a neat answer. Computer

simulation attempts to address this challenge by numerically iterating a microscopic system through phase space according to classical and/or quantum physics. Coarse Monte Carlo (MC) simulations have been performed on a statherin analog in the presence of model surfaces of varying functionalities. [121] Quantum mechanical simulations have also been used to study how different amino acids and small peptides interact with a hydroxyapatite surface. [122,123] The most revealing work regarding the structure of the protein at the statherin/HAP interface, however, has been a series of simulations that have employed the RosettaSurface algorithm. [23, 30, 70] RosettaSurface is a variation of the Rosetta [124] structure prediction software that seeks low energy structures according to van der Waals, implicit solvation, and hydrogen-bonding potentials in the presence of a solid surface. This computational technique has yielded statherin/HAP complexes that agree very well with experimental observations. The advent of RosettaSurface.NMR has refined the minimized structures even further by coupling the scoring function with reported ssNMR data. [30]

Herein we deliver results that conclusively demonstrate the descriptive power of enhanced sampling MD, coupled with experimental data. The accuracy of these simulations, and their ease of execution, should appeal to experts across disciplines in the natural sciences. Below, we focus on the SN15 domain of statherin, which has received much experimental attention. Our previous work focused on the application of the enhanced sampling method Parallel Tempering Metadynamics in the Well-Tempered Ensemble (PTMetaD-WTE), [87] which we geared toward efficiently sampling protein adsorption. [125] We note that the PTMetaD-WTE method has received attention from other groups and has been used effectively in computational research. [126] In the following sections we describe our methodology for simulating the SN15 binding domain in the presence of the [001] face of monoclinic hydroxyapatite, and discuss the results of free energy and clustering analyses. Additionally, we introduce an experimental fitness function that allowed us to pinpoint best-fit theoretical structures. The large experimental and theoretical dataset for the SN15/HAP system provided an excellent source of validation for our simulations.

Setup and Methods

3.2.1 Simulation Details

The simulation details for studying the behavior of the SN15 domain in explicit solvent and in the presence of the [001] face of hydroxyapatite are as follows. The primary structure of the SN15 peptide was constructed with VMD and the psfgen algorithm [90]; its two phosphoserines were constructed and parameterized according to data reported by several groups. [127–129] The HAP surface took the form of a continuous, monoclinic slab one unit cell thick (6.88 Å) and eight unit cells (75.4 Å) long along the a and b box vectors (box angles:

$\alpha = \beta = 90°$, $\gamma = 120°$). Atomic coordinates for a HAP unit cell were set according to experimental crystallographic data for monoclinic hydroxyapatite. [130, 131] The surface was parameterized with force field information made available by Hauptmann and coworkers. [132] We solvated the SN15 peptide with 13933 TIP3P [133] explicit water molecules and three sodium ions to maintain charge neutrality.

Simulations were performed with the Gromacs 4.5.5 molecular dynamics (MD) engine [51] and AMBER99SB-ILDN force field. [134] Free energy calculations were performed with the PLUMED 1.3 plugin for Gromacs. [53] A cutoff of 1 nm was placed on short-range, non-bonded interactions. Long-range electrostatic interactions were calculated with particle mesh Ewald (PME) summation. A time step of 2 fs was used in every simulation. Initially, a solvated SN15 system, sans surface, was energetically minimized for 10000 steps with a steepest decent algorithm. The minimized SN15 system was then equilibrated to a pressure of 1 bar and temperature of 310K with a Berendsen barostat [94] and stochastic velocity-rescaling thermostat [93] in a 1 ns simulation. Following the preceding equilibration step, the SN15 system was grafted above a HAP surface (box dimensions: $a = b = 75.4$ Å, $c = 96.3$ Å and $\alpha = \beta = 90°$, $\gamma = 120°$). This newly assembled system was then energetically minimized for 10000 steps and thermally equilibrated at 310K in a 1 ns simulation, where the HAP surface was frozen in place and would remain frozen for all subsequent simulations. An upper potential wall at Z=85Å was placed on the C_α center of mass in order to prevent any spurious interactions between the peptide and bottom of the HAP surface through the periodic boundary. The position of the wall provided enough space for the peptide to diffuse beyond the cutoff of the HAP surface and into solution.

3.2.2 Enhanced Sampling Method

PTMetaD-WTE is a robust enhanced sampling technique that efficiently samples phase space by: (1) Allowing identical replica systems to diffuse along an incremental temperature scale, overcome all relevant energy barriers, and explore many different configurational states;/citesugita.1999,sindhikara.2010 (2) Applying the Well-Tempered Metadynamics (WTM) algorithm to each replica, [41, 44] biasing specific slow degrees of freedom and accelerating the dynamics of the system; and (3) Requiring many fewer replicas for effective sampling through the use of a potential energy bias placed on each replica. [49] Furthermore, a powerful attribute of the WTM algorithm lies in the fact that the biased distribution of a converged simulation can be reweighted, producing the equilibrium Boltzmann distribution of any conceivable CV. [45] As we have recently shown, [125] this technique is capable of alleviating some of the challenges associated with simulating the adsorption of a biomolecule to a solid surface. Specifically, PTMetaD-WTE can simultaneously drive a peptide on and off a surface (avoiding the tendency for the peptide to become kinetically trapped), force the

peptide to explore relevant areas of conformational space, while providing a way for the system to overcome any "hidden" energy barriers that the WTM bias does not affect. In order to effectively carry out our enhanced sampling scheme, 24 identical replicas were generated from the system described in the simulation details. Each replica, initiated in a random unfolded state, was thermally equilibrated at a different temperature ranging from 310K to 450K. Temperatures were calculated according to the distribution developed by Prakash and coworkers, [95] which ensures an equal probability for exchange between adjacent replicas. A full list of temperatures is provided in Appendix III. Several short, preliminary runs established a reasonable Well-Tempered Ensemble (WTE) potential energy bias, which induced an exchange probability of about 30% between adjacent replicas. The WTE bias was generated and implemented in the same manner as in our past work. [87, 125] The parameter for scaling history-dependent bias energy, or bias factor, was 10 for all simulations. The production simulation ran until Metadynamics (MetaD hereafter) convergence criteria were satisfied, translating to 300 ns/replica, or 7.2 μs of total simulation time. We defined convergence as the point at which no measurable changes occurred in the free energy surface reweighted from the MetaD bias. One aspect of measuring convergence involved monitoring the MetaD bias decay over the course of the simulation. Convergence becomes more likely as the magnitude of each bias addition and the bias decay rate approach zero. We also measured how two respective free energy surfaces varied over the course of the simulation. Convergence was assumed once the free energy difference between the adsorbed and solvated states remained constant for a sufficient amount of time. In an *a posteriori* analysis, we determined that convergence occurred after approximately 225 ns (single replica basis), and extended the simulation 75 ns further to confirm our observation.

3.2.3 Collective Variables

After setting up the Well-Tempered Ensemble, two additional CVs were tuned for use in the production simulation. The first CV was the orthogonal distance between the C_α center of mass of the SN15 peptide and the surface. For a 15-residue peptide, biasing its center of mass is sufficient to ensure it comes into contact with the surface but never becomes kinetically trapped in the adsorbed state. Subtracting the surface thickness from the first CV yields the surface separation distance (SSD) with respect to the C_α center of mass. The second CV defined the pairwise carbonyl oxygen and amide hydrogen α-helical contacts along the backbone of the peptide. The contacts were calculated using a switching function that takes the form:

$$S_{Hbond}(R) = \sum_{i=1}^{N_{HB}} \begin{cases} 1 & r_i \leq 0 \\ \frac{1-(r_i/r_o)^n}{1-(r_i/r_o)^m} & r_i \geq 0 \end{cases} \tag{3.1}$$

As many others have noted, properly tuning parameters n, m, and r_o is an important step to ensure the MetaD simulation seamlessly explores conformational space. The values we selected for n, m, and r_o in this case were 8, 12, and 2.5 Å, respectively. During the PTMetaD-WTE simulation, bias was added at an initial deposition rate of 4 kJ·mol^{-1}·ps^{-1} using a hill widths of 0.1 nm for the first CV and 0.1 for the second.

3.2.4 Experimental Fitness Function

The RosettaSurface.NMR algorithm effectively incorporates experimental data as constraints for resolving the structures of adsorbed biomolecules. A structure prediction method like RosettaSurface.NMR can be thought of as a top-down simulation technique, whose convergence is contingent on optimizing a system according to a set of constraints. In the case of any pertinent molecule, the RosettaSurface.NMR solution structure automatically coincides with a structure that agrees well with experimental measurements. In contrast, MD (with or without enhanced sampling) can be thought of as a bottom-up technique, whose convergence depends on how extensively a system has explored its phase space. Therefore, it is not a trivial task to identify structures from MD simulations that agree well with experimental measurements. In this case, we require a means of measuring experimental fitness.

Over the past two decades the SN15 domain has been extensively characterized by various experimental techniques. This dataset is extremely advantageous to us, and any other researchers who study the statherin/hydroxyapatite interface, as it provides a basis for validating new measurements and calculations pertaining to the system. Experimental measurements of the SN15 domain are conveniently presented in the literature as geometric values. We drew from 22 different experimental measurements of SN15 in the adsorbed state for our comparison. These parameters are shown in Table 3.1.

Here, we introduce an experimental fitness function (EFF) designed to evaluate a simulation trajectory on a frame-by-frame basis, compare it to experimental measurements, and calculate fractional fitness values that indicate how well each structure agrees with experiment (0 to 1, worst to best). The entire experimental dataset that we considered comprises three flavors of information. The first is distance information. These data were taken from ssNMR measurements and consist of mean distances between two specific carbons within the peptide, or between a carbon and phosphorus at the HAP surface. The second flavor we define as angular information; specifically, backbone torsional angles, side chain tilt angles, and the overall backbone tilt angle taken with respect to the surface. This information can be obtained from ssNMR, SFG, and NEXAFS experiments. Finally, we consider limit information. Limits also come from ssNMR experiments, from which the radio labeled atoms did not produce a noticeable signal. A non-signal indicated that the atom was beyond the distance cutoff for detection in the experiment.

Table 3.1. Experimental Parameter Set and Sources

Residues	Measurement[a]	Exp. Value[b]	Method	Reference
pS3	ϕ dihedral angle	-60 ± 10	ssNMR	Long, 2001 [113]
pS3 – F7	Δ(C' – C')	4.3 ± 0.2	ssNMR	Long, 2001 [113]
E4	$\Delta(C_\delta$ – HAP)	> 6.5	ssNMR	Ndao, 2009 [120]
E5	$\Delta(C_\delta$ – HAP)	4.25 ± 0.09	ssNMR	Ndao, 2009 [120]
K6	$\Delta(N_\zeta$ – HAP)	3.5 ± 0.5	ssNMR	Gibson, 2005 [114]
F7	Δ(RING – HAP)	> 6.5	ssNMR	Gibson, 2006 [119]
F7	Θ ring axis tilt	11 ± 6	NEXAFS & SFG	Weidner, 2012 [116]
F7	Ψ ring twist	15 ± 6	NEXAFS & SFG	Weidner, 2012 [116]
F7	ρ ring normal tilt	76 ± 7	NEXAFS & SFG	Weidner, 2012 [116]
L8	ϕ dihdral angle	-60 ± 10	ssNMR	Shaw, 2000 [117]
L8	C_β – C_γ tilt	27 ± 12	SFG	Weidner, new data
L8 – G12	Δ(C' – C')	4.8 ± 0.4	ssNMR	Long, 2001 [113]
R9	$\Delta(C_\zeta$ – HAP)	4.62 ± 0.29	ssNMR	Ndao, 2010 [115]
R10	$\Delta(C_\zeta$ – HAP)	4.53 ± 0.16	ssNMR	Ndao, 2010 [115]
I11	C_α – C_β tilt	-56 ± 6	SFG	Weidner, new data
G12	ϕ dihedral angle	-73 ± 10	ssNMR	Shaw, 2000 [117]
R13	$\Delta(C_\zeta$ – HAP)	> 7.0	ssNMR	Ndao, 2010 [115]
F14	Δ(RING – HAP)	4.55 ± 0.25	ssNMR	Gibson, 2006 [119]
F14	Θ ring axis tilt	61 ± 7	NEXAFS & SFG	Weidner, 2012 [116]
F14	Ψ ring twist	34 ± 6	NEXAFS & SFG	Weidner, 2012 [116]
F14	ρ ring normal tilt	35 ± 6	NEXAFS & SFG	Weidner, 2012 [116]
SN15 peptide	average backbone tilt	85 ± 15	SFG	Weidner, new data

[a] Δ(X – X) indicates the distance between two objects. All tilt angles were measured with respect to the HAP surface normal.
[b] Angles are in degrees, distances are in Ångstroms.

Comparing such a diverse set of experimental data with simulation results requires a flexible approach – one that takes into account the different flavors of data, estimates how closely each simulation frame fits that data, and yields a single value indicative of an overall fitness. Calculating the sum of squared residuals is not satisfactory in our case because the data refers to a variety of inter- and intramolecular relationships, contains different units, and has inequalities. Therefore, a necessary attribute of the EFF must be that it can return a normalized fitness value from each comparison. Seeking an adequate method for experimental comparison, we first focus on distance and angular information. These experimental data are reported as average values with respective standard errors. In this case, a reduced chi-squared (χ^2_R) goodness of fit statistic can be employed. A χ^2_R statistic is a convenient for our purposes because it works with categorically different

data, takes experimental error into account, and normalizes regardless of model complexity. However, the presence of inequalities in the SN15 experimental dataset precludes us from using a common χ_R^2 statistic, which is simply the normalized summed squared deviation of an observation from its reference, divided by the variance of the reference. An alternative χ_R^2 model – our EFF – treats inequalities as step functions and squared-deviation-to-variance ratios as Gaussian functions. This approach ensures that all comparisons lie between zero and one. The EFF can be expressed as:

$$\Psi = \frac{1}{N}\sum_{i=1}^{N}\psi_i \tag{3.2}$$

where,

$$\psi = \begin{cases} \exp\left[-\frac{(O-E)^2}{2\sigma^2}\right] & \text{if distance or angle} \\ H(O,E) & \text{if limit} \end{cases} \tag{3.3}$$

N is equal to the number of observables, O represents an observable from simulation, whereas E and σ refer to an experimental measurement and its error, respectively. The expression $H(O,E)$ represents a step function that yields zero if O is less-than or equal to E, and one if O is greater than E. Conveniently, within the WTM framework, fitness can be treated as a CV, allowing us to extract its equilibrium Boltzmann distribution from our simulation results. [45] We present reweighted fitness free energy profiles and identify best-fit structures in the "Fitness to Experimental Measurements" section of the Results and Discussion.

Results and Discussion

Our analysis focuses primarily on the 310K replica, which is body temperature and therefore the most biologically relevant. We first focus on the general adsorption behavior of a SN15 peptide on hydroxyapatite. Following this, analyses for experimental fitness and clusters of surface-bound SN15 structures are presented. Finally, we discuss the results from several conventional MD simulations that were performed on selected structures of interest.

3.3.1 Free Energy Features of SN15 Adsorption

MetaD bias from the production run was reweighted (rejecting the first 100,000 hills) to produce a free energy surface (FES) of the biased CVs described above. By following the FES in Figure 3.1 along the SSD coordinate, along the x-axis, one can see a gradual free energy gradient between the adsorbed and solvated states. We have observed similar behavior in a small peptide adsorbing to a hydrophilic self-assembled

monolayer surface, whereas a much steeper gradient is evident for hydrophobic surfaces. The gradual free energy slope is indicative of a strong electrostatic interaction between the mineral surface and explicit waters, against which the SN15 peptide competes for space. Additionally, there appears to be a cleft-like formation running down the middle of the surface free energy trough. Simulation snapshots of this region show a dense monolayer of water molecules existing in close proximity to the mineral surface. The cleft itself may represent the acidic head of SN15 disrupting this layer of water, and establishing a strong electrostatic interaction with HAP surface.

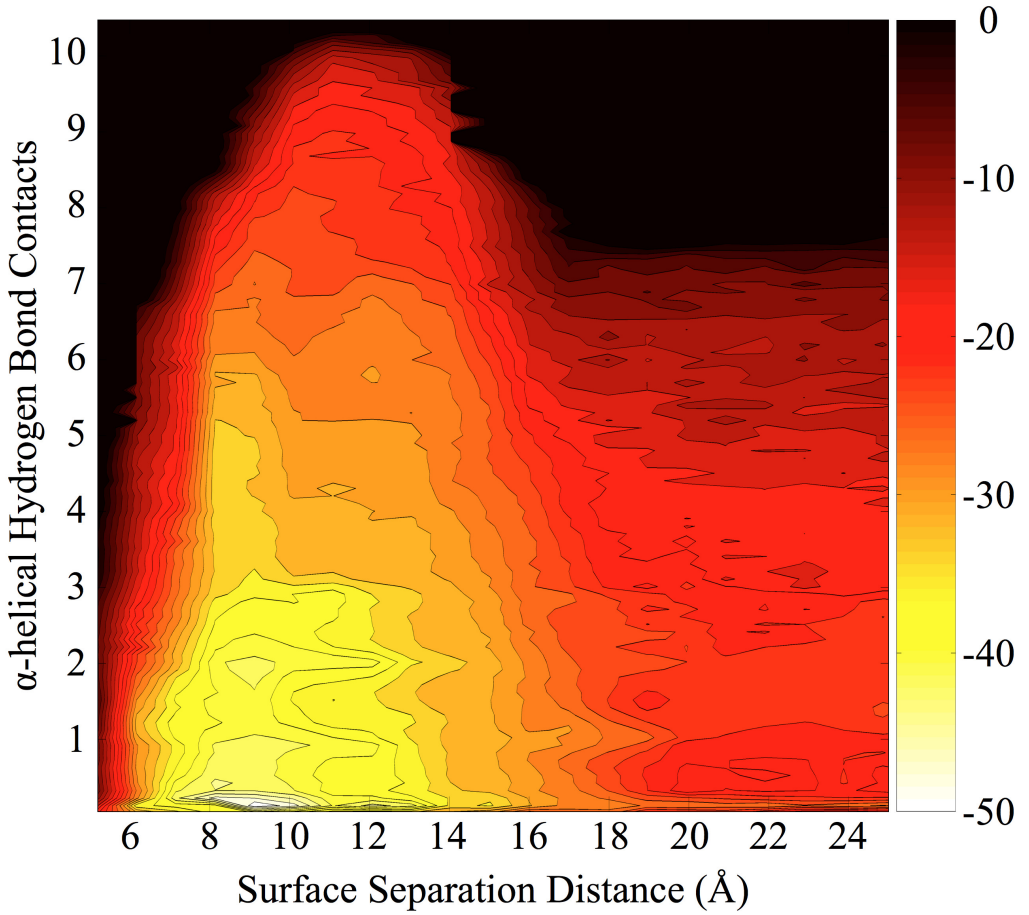


Figure 3.1. Free energy surface of SN15 center-of-mass surface separation distance in relation to α-helical content, represented by backbone hydrogen bond contacts. Isolines are spaced every 2kT, whereas the sidebar is in units of kT.

Figure 3.1 also shows that the SN15 peptide generally lacks strong secondary structural features in the solvated state. Conversely, the peptide is capable of adopting noticeable α-helical characteristics on the HAP surface, evidenced by the surface free energy trough extending into highly ordered conformational space. This feature of the FES could be indicative of surface-induced folding in the sense that the folded state is explored

more at the surface than in solution. However, the free energy difference between the unfolded state and folded state is about 20kT for both an adsorbed peptide and a solvated peptide. Moreover, we note that there is no stable macrostate in the folded region of the surface free energy trough. Instead, the trough slopes gradually from a region of high α-helicity down to a relatively disordered state. A solitary, α-helical SN15 does not appear to be stable in the adsorbed state.

As with any complex system whose information has been projected onto only a few select coordinates, the SN15/HAP system is not perfectly described by the FES generated from the MetaD CVs. To extract any further meaning from the Figure 3.1 would be ill advised due to the degeneracy inherent to any definite region of the FES. The reason for this lies in the combinatorial nature of the hydrogen bond CV, for which many different structures can yield the same value, providing little insight as to what structures are most prevalent at the HAP surface. Consequently, this condition requires us to analyze our dataset along different criteria. In the following section we introduce an analysis of how well our dataset compares to experimental observations. The analysis is performed primarily with our own experimental fitness function, which we explain in our Methods above.

3.3.2 Fitness to Experimental Measurements

In order to identify the surface structures that most closely fit experimental data, we applied our experimental fitness function (EFF) to each simulation frame that contained SN15 in the adsorbed state. Fitness scores range from zero to one, where zero is the poorest fit and one is the best. To demonstrate the capability for MD and PTMetaD-WTE to predict the structure of surface-bound peptides, we present the best-fit structure from a 13-parameter comparison in Figure 3.2. This particular structure and the RosettaSurface.NMR SN15 region were evaluated along the same 13 experimental parameters so they could be rationally compared. Quantitative details for each are provided in Table 3.2. One can see that the majority of the observables calculated from the best-fit structure from the 13-parameter set fall within error of both experimental and RosettaSurface.NMR results. Moreover, when we compare EFF scores between RosettaSurface.NMR results and the best-fit structure, both appear to match experimental results to a similar degree. Many individual residues from the high-scoring structure in Figure 3.2 occupy appropriate positions on and off the surface. Specifically, we see the first three acidic residues and K6 tightly bound to the surface; R9 and R10 are loosely bound, whereas R13 is more than 7Å away from the surface; and F7 points away from the surface, while F14 is oriented toward the surface. The only significant outlier is the position of E5, which is expected to be situated closer to the surface.

Additionally, in Table 3.2 we provide information detailing the best-fit model for a 22-parameter EFF

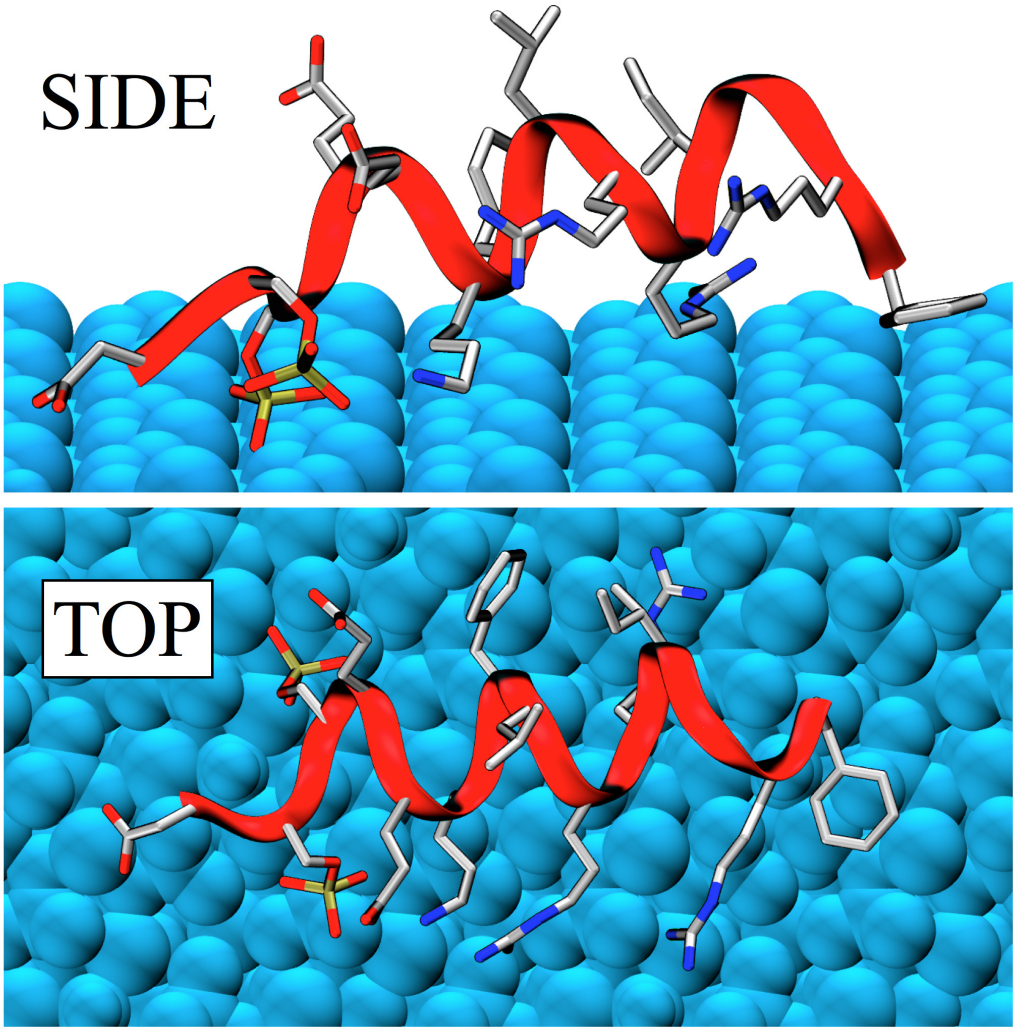


Figure 3.2. Side and top views of the best-fit SN15 structure from a 13-parameter experimental fitness calculation. Hydrogen atoms have been removed for clarity.

calculation. This calculation took into account the respective Eulerian angular orientations of the F7 and F14 phenyl rings with respect to the surface, [116] as well as the tilt angles of the backbone and residues L8 and I11 (new data). The best-fit structure from the 22-parameter calculation is very similar to the structure in Figure 3.2 – a snapshot is provided in Appendix III. We calculate their RMSD to be 2.1 Å. The most noticeable difference between the two best-fit structures is the position of the F14 residue at the C-terminus, which lies close to the surface for the 13-parameter best-fit and farther away for the 22-parameter best-fit. Regarding the angular orientations of the SN15 phenylalanine residues, both best-fit structures' F7 residues deviate nearly equally from experimental measurements. However, the 22-parameter best-fit more closely matches the SFG measurements for F14. Finally, while both best-fit models also deviate from the L8 and

Table 3.2. Comparison of Observables from Respective Datasets

Measurement[a]	Experiment[b]	RosettaSurface[b]	Best-Fit (13)	Best-Fit (22)
pS3 ϕ dihedral angle	-60 ± 10	-63.9 ± 12	-62.6	-64.2
pS3 – F7 Δ(C' – C')	4.3 ± 0.2	4.3 ± 0.3	4.2	4.2
E4 $\Delta(C_\delta$ – HAP)	> 6.5	11.7 ± 1.4	10.6	11.2
E5 $\Delta(C_\delta$ – HAP)	4.25 ± 0.09	5.0 ± 0.5	9.3	13.5
K6 $\Delta(N_\zeta$ – HAP)	3.5 ± 0.5	4.3 ± 0.8	3.0	2.0
F7 Δ(RING – HAP)	> 6.5	11.3 ± 0.4	6.7	9.8
L8 ϕ dihedral angle	-60 ± 10	-62.7 ± 6.2	-57.5	-59.3
L8 – G12 Δ(C' – C')	4.8 ± 0.4	4.3 ± 0.5	4.3	4.6
R9 $\Delta(C_\zeta$ – HAP)	4.62 ± 0.29	4.6 ± 0.4	6.9	10.5
R10 $\Delta(C_\zeta$ – HAP)	4.53 ± 0.16	4.8 ± 0.3	4.4	4.9
G12 ϕ dihedral angle	-73 ± 10	-71.4 ± 8.2	-70.1	-77.3
R13 $\Delta(C_\zeta$ – HAP)	> 7.0	7.6 ± 2.5	7.2	16.9
F14 Δ(RING – HAP)	4.55 ± 0.25	12.4 ± 2.8	4.5	13.3
F7 Θ ring axis tilt	11 ± 6	–	61.9	58.9
F7 Ψ ring twist	15 ± 6	–	14.6	15.9
F7 ρ ring normal tilt	76 ± 6	–	31.4	34.6
L8 C_β – C_γ tilt	27 ± 12	–	52.9	56.3
I11 C_α – C_β tilt	-56 ± 6	–	-18.0	-30.1
F14 Θ ring axis tilt	61 ± 6	–	84.7	34.2
F14 Ψ ring twist	34 ± 6	–	0.1	32.8
F14 ρ ring normal tilt	35 ± 6	–	5.3	61.8
average backbone tilt	85 ± 15	–	77.3	78.3
EFF Score (13 // 22)[c]	– // –	0.68 // –	0.72 // 0.55	0.60 // 0.62

[a] Δ(X – X) indicates the distance between two objects. All tilt angles were measured with respect to the HAP surface normal.
[b] Angles are in degrees, distances are in Ångstroms.
[c] Calculated from a 13- or 22-parameter comparison.

I11 tilt angle values, we note that the orientations of each sidechain agree with experiment.

Regarding the reported α-helical character of SN15, each of the backbone dihedral angles measured in our best-fit models match experimental values within error. These structures also match, within error, the predicted backbone tilt angle of the entire peptide (77.3° and 78.3° with respect to the surface normal, compared to 85° ± 15°). The backbone tilt angle was calculated by averaging the tilt angles of each backbone carbonyl C=O with respect to the surface, to mimic an SFG analysis of the backbone amide-I stretch. Keeping with this observation, we note the first three acidic residues have a strong affinity for the HAP

surface – they are able to effectively penetrate the layer of tightly bound surface waters mentioned in the preceding section. Residues past K6 have a weaker electrostatic affinity for the surface, and appear to rest on top of the water layer. Therefore, we propose that the slight tilt along the surface-bound SN15 peptide is a result of the displacement of the C-terminus by tightly bound waters at the HAP surface.

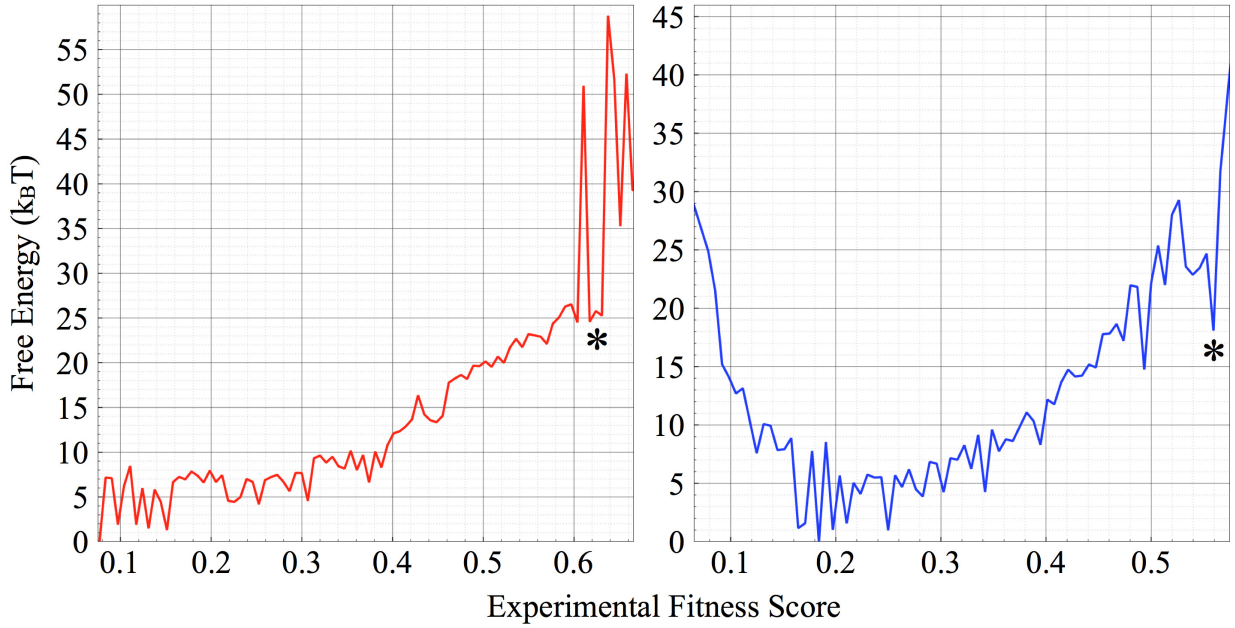


Figure 3.3. The reweighted free energy profiles of (a) a 13-parameter experimental fitness calculation and (b) a 22-parameter experimental fitness calculation. High-scoring wells are marked with asterisks.

Fitness values of all surface-bound structures were calculated, and subsequently reweighted (rejecting the first 35,500 hills in the surface set) to produce the equilibrium Boltzmann distribution of the set. We present two EFF free energy profiles. The first, shown in Figure 3.3a, is the fitness distribution based on the 13-parameter comparison we describe above. The best-fit structure from Figure 3.2 does not appear in Figure 3.3 because its free energy was too high to appear significant on the free energy profile. Figure 3.3b shows a similar distribution, but incorporates all 22 experimental parameters. In each profile, one can see two distinct minima that correspond to poorly fit structures and relatively closer fits, respectively. Both figures show that poorly fit structures make up the global free energy minimum, whereas structures that more closely match experimental data are markedly higher in free energy. The free energy difference between the two minima in Figure 3.3a is about 25kT, with deference to the poorly fit structures, and the difference in Figure 3.3b is somewhat lower at 18kT. A second free energy well can also be seen to the left of the 18kT well in Figure 3.3b around 15kT. Coincidentally, the difference in fitness score between the two wells is roughly $1/22$. The most plausible explanation for this difference is that that a single residue (E4, F7, or R13) is closer to the surface in the 15kT well than in the 18kT well.

These perplexing distributions appear to go against established experimental measurements. Based on our previous experience with the LK peptide on SAM systems, [125] we expected that our simulation would predominantly favor surface-bound structures that agree well with experimental observation. Even from a theoretical point of view we expected to see closer agreement, given the promising results from RosettaSurface.NMR simulations. What we see is that while structures exhibiting a high degree of fitness are sampled, they are also relatively high in free energy. To help resolve the apparent disagreement between our simulation and experimental evidence we provide a clustering analysis below.

3.3.3 Clustering Analysis of Adsorbed Structures

In order to avoid relatively high free energy structures in our analysis, we chose to take a representative sample based on an assessment of Figure 3.1 and a one-dimensional free energy profile of the surface separation distance (SSD). The region of space we examined was where the SN15 center of mass fell between 0.7 nm and 1.4 nm away from the surface. This region was wide enough to provide a sample size of roughly 1.6×10^5 individual surface-bound structures, while narrow enough to exclude structures in energetically unfavorable proximity to the surface, or in some transitional state between the surface and solution.

Our clustering analysis was performed according to a procedure developed in our previous work. [125] According to an RMSD cutoff of 3Å – an appropriate cutoff distance for a small biomolecule [135] – surface-bound structures were grouped into 232 individual, non-overlapping clusters. The one-to-one relationship existing between each cluster and its constituents permitted us to reweight the cluster set (rejecting the first 35,500 hills in the surface set), and determine the relative free energies of each individual cluster. An interesting dichotomy emerges from the cluster set: (1) unfolded, low-scoring structures typically have low free energies (structures I, II, and III of Figure 3.4), whereas (2) the two densest clusters in the set are folded, high-scoring structures that are comparably higher in free energy (structures IV and V of Figure 3.4). One might intuitively expect the denser clusters to be lower in free energy. However, we must note that the most significant factor in the reweighting algorithm is the production run MetaD bias. In other words, the MetaD weight inherent to a cluster trumps its population density. The unfolded, low-scoring clusters in Figure 3.4 received substantially more MetaD bias than the folded, high-scoring structures during the production run, and therefore are lower in free energy. This is also evidenced by the global free energy minimum appearing on the bottom left of Figure 3.1, where adsorbed structures exhibit low α-helical order.

Nevertheless, the presence of a few densely populated folded states shows that the HAP surface induces folding, but lacks a critical factor that stabilizes the SN15 folded state. Taking note of this observation, we see a consistent trend throughout each of our respective analyses. Recalling the surface free energy trough

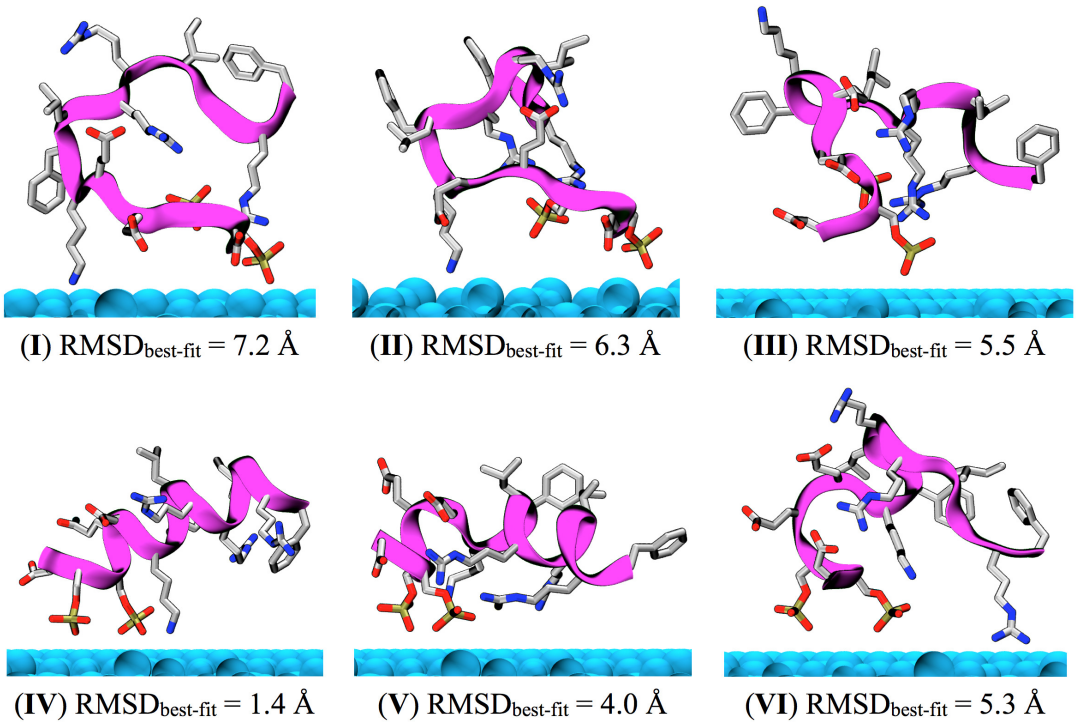


Figure 3.4. Representative structures extracted from a clustering analysis. Structures I, II, and III illustrate, in ascending order, the three lowest free energy surface-bound states. Structures IV, V, and VI illustrate, in descending order, the three most densely populated states. The RMSD value reported below each structure is with respect to the best-fit structure shown in Figure 3.2. Hydrogen atoms have been removed for clarity.

present in Figure 3.1, the SN15 domain folds more readily at the HAP surface compared to solution, but never achieves a stable folded conformation at the HAP surface. Instead, we see a relatively smooth and gradual slope extending downward from the folded state toward the unfolded state. Moreover, our results from experimental fitness reweighting in Figure 3.3 describe similar behavior, where a low-scoring, unfit state appears to be preferred over structures that agree with experimental observations.

Clearly, confinement to the HAP surface is only partially responsible for inducing and stabilizing the SN15 folded state. Based on the above analysis, we conclude that our system containing a single peptide, a HAP surface, and explicit water was inadequate for reproducing behavior measured with experimental techniques. We note that the pristine conditions typically assumed in molecular simulations rarely match experimental environments, which contain millimolar concentrations of protein, let alone media as heterogeneous as human saliva. Furthermore, recent work suggests complex dynamics occur through protein-protein interactions taking place at a solid surface. [136] Following this, we hypothesized that one or more unaccounted environmental factors, such as lateral protein-protein interactions, could be affecting the results.

The fact that saliva is relatively protein-rich, and how statherin has been characterized as a pellicle protein – meaning many adsorb to tooth enamel as a film – supports the idea that that protein-protein interactions could play a significant role in stabilizing the SN15 folded state. We carried out several conventional MD simulations as a first step to test this hypothesis.

3.3.4 MD Analysis of Selected Structures

Seven conventional MD simulations were run, whose initial configurations were as follows: (A) The most densely populated cluster, (B) The lowest free energy cluster, (C) The best-fit structure from Figure 3.2, (D) The structure with the lowest 13-parameter EFF score, (E) A representative structure from the low-scoring basin in Figure 3.3a, (F) A representative structure from the high-scoring basin in Figure 3.3a, assuming high degree of similarity between high-scoring structures in Figure 3.3a and Figure 3.3b. The seventh simulation (G) comprised the starting configuration from (F) surrounded by four other unfolded, surface-bound SN15 peptides. The purpose of (G) was to test the protein-protein interaction hypothesis by confining the folded SN15 peptide to a crowded microenvironment at the HAP surface. Each simulation was equilibrated and executed according to the setup procedure detailed in the Setup and Methods section above.

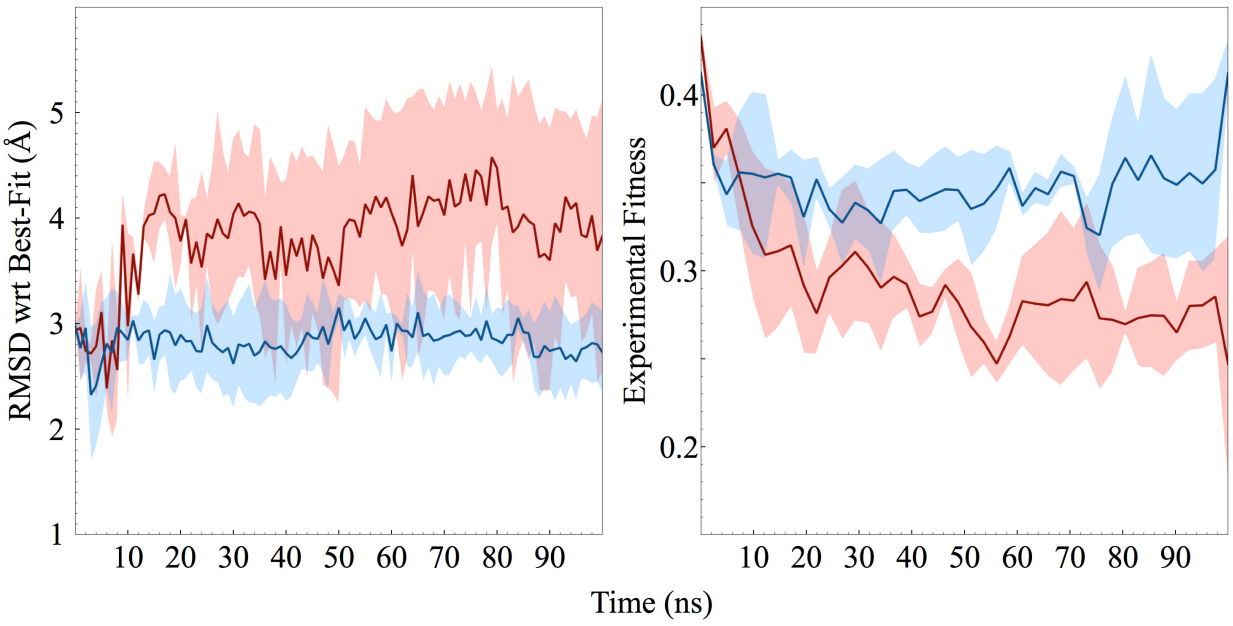


Figure 3.5. (a) RMSD trajectories of surface-bound, solitary (red) and crowded (blue) SN15 peptides taken with respect to the 22-parameter best-fit SN15 structure; and (b) fitness score trajectories produced from a 22-parameter EFF evaluation for surface-bound, solitary (red) and crowded (blue) SN15 peptides. Shaded regions represent the standard deviation of each trajectory.

Each of the first six simulations deviated significantly from its initial configuration after 25ns of simulation

time. Most notably, the systems initiated from the folded state exhibited a noticeable degree of unraveling from the C-terminus. However, the peptide within the seventh system maintained a folded conformation over the course of the entire simulation. To verify our observation, we extended simulations (F) and (G) to 100ns and ran each in triplicate. Figure 3.5a shows the averaged C_α RMSD trajectory (with respect to the best-fit structure from the 22-parameter calculation) of simulations (F) and (G) nested in shaded regions representing the standard deviation. After about 10ns the solitary peptide began to deviate farther from the best-fit conformation, later experiencing substantially larger RMSD fluctuations than its crowded counterpart. These comparatively large fluctuations indicate that the SN15 domain is free to explore a wide range of conformational space and unfold on very short timescales in the absence of stabilizing protein-protein interactions. Conversely, one can see the crowded peptide maintained its structure over the course of the entire 100ns, and also fell within the 3 Å cutoff limit we imposed in our clustering analysis; it also experienced comparatively diminished structural fluctuations when crowded by other peptides at the HAP surface.

Simulations (F) and (G) also differ in terms of experimental fitness. Figure 3.5b shows averaged fitness scores based on the 22-parameter evaluation for both systems over the 100ns duration. One can see each system experiences an initial drop in fitness, lasting about 25ns. However, the crowded system (G) maintains a higher fitness score than (F) and begins to trend upward after about 75ns, whereas the score for (F) consistently decreases over time. Therefore, we propose that lateral protein-protein interactions play a significant role in stabilizing the SN15 folded state on a hydroxyapatite surface.

Conclusions

In this report we show how we exhaustively sampled the adsorption behavior of the N-terminal binding domain of statherin (SN15) on the [001] face of monoclinic hydroxyapatite (HAP). Furthermore, we show that with MD and the enhanced sampling technique PTMetaD-WTE, it is possible to predict the surface-bound conformation of a domain known to fold only in the adsorbed state. Our predictions agree with results from the RosettaSurface.NMR algorithm, an alternative surface structure prediction technique. Moreover, best-fit structures generated from our simulation align well with 22 individual experimental measurements. Currently, no experimental technique exists that is fully capable of resolving the structure of a biomolecule at an interface. Barring future technological advances, computer simulations are now more than ever crucial for overcoming important scientific challenges. These results are evidence that a unique partnership between theory and experiment is not only possible, but also mutually beneficial to an exciting degree.

Throughout our analysis of the SN15/HAP system we noticed consistent behavior: the HAP surface

induces the SN15 domain to fold; however, any persistent structural stability is precarious at best. Several conventional MD simulations were performed to resolve the apparent conflict between our formal simulation results and experimental measurements reported in the literature. We found that a folded SN15 domain generally retains its conformation when crowded by other peptides at the HAP surface, whereas solitary peptides tend to unravel after a short period of simulation time. This result aligns well with statherin's biological role as a pellicle protein, which adsorb en masse to tooth enamel and create a film. Just as cellular proteins rely on a crowded, heterogeneous environment to fold properly, the SN15 domain likely requires both protein-surface and lateral protein-protein interactions to maintain its conformation. We emphasize the power inherent to these calculations. We have generated a statistical sample on the order of one hundred thousand individual surface-bound structures, made accurate of the peptide's behavior, and also are able to suggest environmental factors that could play a significant role in the dynamics of a real system. Enabled by current state-of-the-art computational resources, and powerful software (Gromacs and Plumed) that is available gratis, future studies should focus on decidedly larger and societally relevant systems.

Chapter 4

Simulating a Whole Statherin Protein on Hydroxyapatite

Simulating the adsorption of a biomolecule is a task that requires both detailed knowledge of the system at hand and a reliable method for sampling conformational space. The first criterion provides necessary experimental data that validates computational results. The second criterion ensures that a measurably sufficient amount of conformational space is explored, and establishes the computational results as the best-possible model of the physical system. For a statherin protein in the presence of a hydroxyapatite (HAP) surface, a large body of quantitative experimental research satisfies the former criterion. In our prior work, we successfully applied the Parallel Tempering Metadynamics in the Well-Tempered Ensemble (PTMetaD-WTE) method to simulate the adsorption of two comparably smaller peptides. Size is a primary concern for researchers studying the protein folding problem, as the number of possible conformations a protein can assume increases exponentially with chain length (Levinthal's Paradox). An extra degree of complexity is added to this problem when one considers conformational behavior at, or close to, an interface. Therefore, a thorough, logical approach must be taken when simulating the adsorption behavior of any protein larger than, say, a dozen residues. Herein, we present a simulation of the salivary protein statherin adsorbing to the [001] face of monoclinic hydroxyapatite. Statherin is a protein responsible for inhibiting the accretion of calcium phosphate in saliva. Statherin also naturally binds to tooth enamel with its acidic N-terminal residues, and acts as a lubricating agent in the adsorbed state. At 43 residues in length, statherin is the largest protein we have studied to date in the context of adsorption. While larger proteins on surfaces have been simulated elsewhere, those studies have not achieved the high degree of sampling we present below.

As we mentioned in the preceding chapter, a substantial amount of experimental work has been done on the statherin/HAP system. [110, 137] There, we reviewed literature that focuses on the first 15 residues of statherin (SN15) – the putative binding domain. The C-terminal residues (Y16 to F43) have received less

attention due to their relatively disordered nature and comparatively weak affinity for hydroxyapatite, both of which complicate measurement. Nevertheless, some valuable information has been deduced regarding certain structural and functional attributes of the statherin C-terminus. In a solid-state nuclear magnetic resonance (ssNMR) experiment, Masica and coworkers showed that P28 does not interact with the HAP surface when statherin is in the adsorbed state. [30] Likewise, in a separate ssNMR study Ndao and coworkers showed that E26 also does not interact with a HAP surface. [120] This latter result is somewhat surprising, given the affinity acidic residues have for the mineral surface, and supports the tight packing density scheme proposed by R. Goobes and coworkers. [138]

These observations, however, do not negate C-terminal involvement with adsorption. Results from an *in situ* ellipsometry study, performed by Santos and coworkers, indicate the opposite is a more likely scenario. [139] Comparing affinity for a HAP surface, they found that statherin adsorbs in higher concentrations compared to a fragment comprising its first 21 residues (StN21), implicating the C-terminus in surface binding. With data obtained from ssNMR and the Rosetta structure prediction algorithm, G. Goobes and coworkers proposed that the P33–Y38 region of statherin adopts an α-helical conformation when statherin is adsorbed to HAP. [140] Furthermore, the hypothetical tertiary structure of the adsorbed protein could involve the P33–Y38 domain interacting with the proline-rich midsection (residues Y16–P28). Statherin C-terminal composition and structuring are also hypothesized to play a role in bacterial adhesion. Studies have shown the periodontopathogens *Porphyromonas gingivalis* and *Fusibacterium nucleatum* binding to the L29–F43 and P33–Q39 regions of statherin, respectively. [141, 142]

Below, we report on an all-atom molecular dynamics (MD) simulation of an explicitly solvated statherin molecule in the presence of a monoclinic HAP surface. The PTMetaD-WTE method was used to obtain a representative sample of conformational space on and off the HAP surface. We analyze the adsorption behavior of statherin, evaluate how well the simulation matched experimental results, and discuss the structural details of statherin in the adsorbed state. In the Sampling Details we discuss our rationale for balancing computational cost with comprehensive sampling, which must be seriously considered when simulating even moderately sized proteins. The protocol for determining simulation convergence is also discussed. These last two points are crucial when studying biomolecular adsorption, and even more so if the biomolecule of interest is capable of adopting a broad range of conformations.

Simulation Setup

The primary structure of statherin was built with the psfgen algorithm available in the Visual Molecular Dynamics (VMD) software package [90]; its two phosphoserines were constructed and parameterized according

to data reported by several groups. [127–129] The HAP surface took the form of a continuous, monoclinic slab, one unit cell thick (6.88 Å) and ten unit cells (94.2 Å) long along the a and b box vectors (box angles: $\alpha = \beta = 90°$, $\gamma = 120°$). Atomic coordinates for a HAP unit cell were set according to experimental crystallographic data for monoclinic hydroxyapatite. [130, 131] The surface was parameterized with force field information made available by Hauptmann and coworkers. [132] The solution environment consisted of 24190 TIP3P [133] explicit water molecules and four sodium ions to maintain charge neutrality.

Simulations were performed with the Gromacs 4.5.5 MD engine [51] and AMBER99SB-ILDN force field. [134] Free energy calculations were performed with the PLUMED 1.3 plugin for Gromacs. [53] A cutoff of 10 Å was placed on short-range, non-bonded interactions. Long-range electrostatic interactions were calculated with particle mesh Ewald (PME) summation. A time step of 2 fs was used in every simulation. Initially, a solvated statherin system, sans surface, was energetically minimized for 10000 steps with a steepest decent algorithm. The minimized statherin system was then equilibrated to a pressure of 1 bar and temperature of 310K with a Berendsen barostat [94] and stochastic velocity-rescaling thermostat [93] in a 1 ns simulation. Following the preceding equilibration step, the statherin system was grafted above a HAP surface (box dimensions: $a = b = 94.2$ Å, $c = 106$ Å and $\alpha = \beta = 90°$, $\gamma = 120°$). This newly assembled system was then energetically minimized for 10000 steps and thermally equilibrated at 310K in a 1 ns simulation, where the HAP surface was frozen in place and would remain frozen for all subsequent simulations.

The statherin molecule is significantly larger than the peptides in our prior studies. As such, it required a more elaborate restraint to prevent it from interacting with the bottom of the HAP surface through the periodic boundary. Thus, a three-part quartic restraint – Equation 4.1 – was placed on the C_α center of mass (COM) of three corresponding segments along the statherin backbone: the SN15 domain (residues 1 to 15), the following 15 residues, and the trailing 13 residues.

$$U(s) = K(s - L)^4 \tag{4.1}$$

The COM position is represented by s. The restraint was placed at L=85 Å, and used a spring constant K value of 1,000,000 kJ·mol^{-1}·Å^{-4} to prevent any spurious interactions between statherin and the bottom of the HAP surface. A sufficient amount of space remained for the protein to diffuse beyond the surface force field cutoff.

We used the PTMetaD-WTE technique to induce the exploration of a diverse range of states in our MD simulation. PTMetaD (Parallel Tempering, PT [42, 43] + Well-Tempered Metadynamics, WTM [41]) has been demonstrated to be effective for studying biomolecular phenomena, [44, 46, 85] including peptide adsorption [125] in the Well-Tempered Ensemble (WTE). [49] Results from a PTMetaD-WTE simulation

are amenable to reweighting, which produces the equilibrium Boltzmann distribution of any observable regardless of whether the observable was biased or not. [45] In the statherin simulation, bias energy was initially deposited at a rate of 4 kJ·mol^{-1}·ps^{-1}, and scaled over the course of the simulation by a bias factor of 10. The WTE was constructed in the same manner as our previous work with a bias factor of 15. A set of 24 replicas was generated from the system described in the setup above. Each replica was equilibrated at distinct temperature derived from a numerically generated distribution. [95]

We employ an augmented version of the experimental fitness function (EFF) introduced in the previous chapter. Experimental data describing the C-terminal residues were incorporated into the updated version. The full set includes the literature data shown in Table 3.1 in Chapter 3, and Table 4.2 below.

Table 4.1. Experimental Parameter Set and Sources

Residues	**Measurement**[a]	**Exp. Value**[b]	**Reference**
P23 – P33	Δ(H(4' ring position) – C')	8.8 ± 0.8	G. Goobes, 2006 [140]
		10.5 ± 1.0	
P23 – Y34	Δ(H(4' ring position) – C')	8.8 ± 0.8	G. Goobes, 2006 [140]
		10.5 ± 1.0	
E26	Δ(C_δ – HAP)	> 6.5	Ndao, 2009 [120]
P28	Δ(C_δ – HAP)	> 7.28	Masica, 2010 [30]
P33 – Y34	Δ(C' – C')	3.12 ± 0.13	G. Goobes, 2006 [140]
P33 – Y38	Δ(C' – N)	5.3 ± 0.5	G. Goobes, 2006 [140]
Y34	ϕ dihedral angle	-75 ± 15	G. Goobes, 2006 [140]
Y34	ψ dihedral angle	-40 ± 10	G. Goobes, 2006 [140]
Y34 – Y38	Δ(C' – N)	4.0 ± 0.5	G. Goobes, 2006 [140]

[a] Δ(X – X) indicates the distance between two objects. All tilt angles were measured with respect to the HAP surface normal.
[b] Angles are in degrees, distances are in Ångstroms.

Two values are reported for distance measurements between the 4' hydrogen of P23 (substituted as a fluorine in experiment) and the backbone carbonyl carbon of P33 and Y34 respectively. G. Goobes and coworkers report that a distinction could not be established between the two measurements; therefore, both values are reported together. To accommodate these data, the EFF was modified so that the scoring function incorporated two Gaussian peaks instead of one.

Sampling Details

When sampling a system with Metadynamics (MetaD), specific collective variables (CVs) representing inherently slow degrees of freedom are biased. Extensive and efficient exploration of phase space is dependent on which CVs are chosen. The process through which a biomolecule adsorbs and desorbs from a surface is often slow by simulation standards, taking place over seconds or even hours in some experimental procedures. If the attractive potential of a surface has a strong influence on a biomolecule, it is doubtful that thermal fluctuations alone would ever be sufficient to dislodge the biomolecule – at least on any computationally practical timescale. Therefore, it is necessary to bias a biomolecule's position normal to a surface in order to comprehensively sample a wide range of biomolecule-surface interactions. Accordingly, we biased the C_α COM of the statherin molecule along the Z-axis (normal to the HAP surface) of the simulation environment.

Computational expense (concomitant with convergence time) of a simulation increases with the number of biased CVs. In order to minimize computational expense, we chose to bias one additional CV for the purpose of sampling configurational space. Given that statherin is only known to adopt a persistent secondary structure in the adsorbed state – implying that confinement effects at the surface strongly influence the SN15 domain to fold – whereas the majority of the protein remains unstructured in both solvated and the adsorbed states, we assumed that the radius of gyration should be a sufficient descriptor. Moreover, biasing the radius of gyration of the entire statherin molecule should yield a more diverse sample of extended and collapsed states than if N-terminal hydrogen-bond contacts were biased instead.

Several auxiliary MD simulations of statherin on and off the HAP surface, at different temperatures, were carried out in order to test our hypothesis. At low temperatures, the statherin molecule preferentially assumed a globular conformation on and off the surface, and showed little deviation from the compact state. Conversely, at low and high temperatures, we observed the SN15 domain of a surface-bound statherin to repeatedly form and break α-helical hydrogen bonds. Large-scale structural changes occurred over much longer timescales. These preliminary simulations supported the notion that the extension and collapse of the statherin molecule occurs on a longer timescale than α-helical bond formation, especially in the adsorbed state. We proceeded with a production simulation wherein the statherin surface proximity and radius of gyration were biased. The MetaD sigma value for both CVs was 1 Å.

The simulation was run until our convergence criteria were satisfied, amounting to 500 ns/replica or 12μs of total simulation time. Generally, it is considered good practice to judge the convergence of a simulation along several metrics. Three metrics were used for this simulation. First, we tracked the sample evolution of biased and unbiased collective variables (CVs) over the course of the simulation. This convergence criterion depended on the observation that no new regions of CV space were being sampled after a sufficient

amount of simulation time. The second involved monitoring the free energy difference between two apparent macrostates. Convergence can be assumed once the free energy difference between the two states is observed to attain a constant value for a sufficient amount of time. We note this metric requires *a posteriori* knowledge of the system's macrostates, and therefore cannot be performed until a substantial amount of sampling has already taken place. We monitored the free energy difference between the solvated state and an arbitrary surface macrostate for several biased and unbiased CVs; namely, the whole protein COM, SN15 domain COM, protein midsection COM, protein tail COM, radius of gyration, and SN15 hydrogen bond contacts. Most of the CVs converged in less than 200ns (per replica basis), whereas the radius of gyration required nearly 400ns. The final convergence metric we used was related to the MetaD bias itself. This convergence criterion was met once both the magnitude of MetaD bias and its decay rate had asymptotically approached zero for a sufficient amount of time – where a trend showing hill height versus time is essentially flat and close to zero. We continued our simulation for 100ns/replica (20% of the total simulation time) past the point where all convergence criteria were satisfied in order to ensure the simulation was converged.

Results and Discussion

In this analysis, we examine the 310K replica as it is the most biologically relevant. Free energy results represent the equilibrium Boltzmann distributions of the respective CVs, which were extracted from the original MetaD bias with a reweighting algorithm. [45]

4.3.1 Free Energy Features of Statherin Adsorption

A reweighted free energy surface (rejecting the first 100,000 hills – the transient period) of the statherin surface separation distance (SSD) versus its radius of gyration is shown in Figure 4.1. As expected, statherin preferentially exists in a collapsed conformation in solution, and experiences a substantial degree of deformation upon adsorption. In contrast to the free energy surfaces in Chapters 2 and 3, we see three distinct macrostates take shape at the HAP surface (states I, II, and III). These three states represent conformations of varying elongation, whose local minima become more proximal to the surface with increasing extension. We also note two distinct free energy troughs extending from state I into solution. The first, running along the bottom of Figure 4.1, represents the case of a compact molecule approaching the surface. Upon adsorption, the protein orients itself in some thermodynamically preferred manner without much configurational disruption. The second trough, extending diagonally into solution from state I, most likely corresponds to a loose, globular state whose bulk is detached from the surface but still bound by some trailing domain. Further structural analysis of this region revealed the SN15 domain bound to the HAP surface with a glob-

ular C-terminus protruding into the bulk in most cases. A possible mechanism could exist where the SN15 domain hitches onto the HAP surface and, upon binding, draws the bulk of the protein closer to the surface as Figure 4.1 suggests.

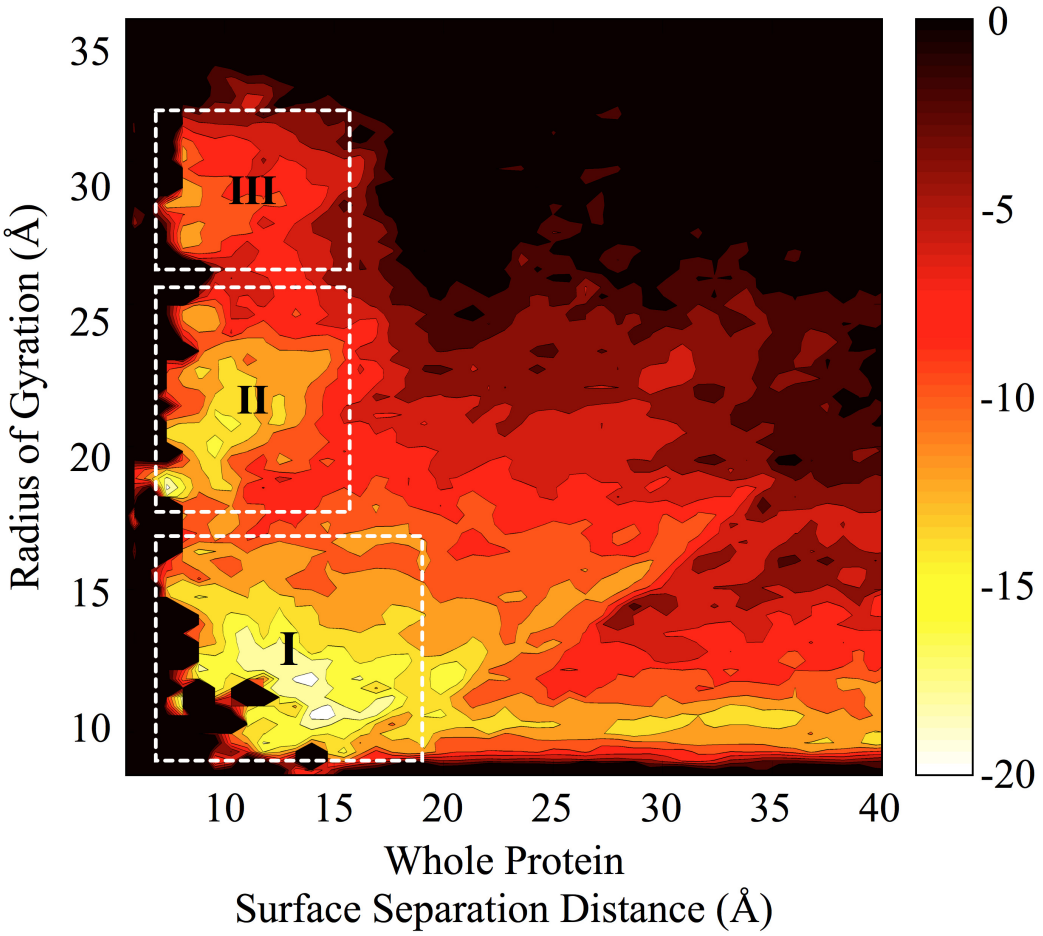


Figure 4.1. The reweighted free energy surface for the statherin radius of gyration versus its C_α COM SSD shows several features related to its adsorption behavior. The surface-bound, globular state (I) contains the global free energy minimum and suggests two possible pathways for statherin adsorption. States II and III represent extended surface structures whose globular forms are broken and stabilized by the attractive potential at the HAP surface. Isolines are spaced every 2kT whereas the color bar is in units of kT.

A reweighted, one-dimensional free energy profile of the C_α SSD (not shown) was calculated from the MetaD results; it depicted an adsorption free energy difference of -14.12 ± 0.04 kJ/mol, which differs from the experimental literature value – -42.3 ± 2.1 kJ/mol [143] – by less than a factor of three. The fact that the calculated adsorption free energy is the same order of magnitude as the experimental value gives weight to the accuracy of our simulation. The statherin adsorption free energy we calculated is much less than that of a lone SN15 fragment. This is most likely because of the C-terminal statherin residues interfering with the formation of close contacts between the N-terminal residues and the HAP surface. The fact that statherin

binds to HAP more strongly in experimental studies suggests that the protein experiences a cooperative binding effect in the presence of other proteins.

The free energy gradient between the adsorbed and solvated states appears to steepen with increasing protein extension. Several factors most likely contribute to this behavior. As we note in our prior studies, a protein is likely to compete with water for surface contact if the surface is very hydrophilic. Other factors to consider are charged residues with high electrostatic affinity for the HAP surface, and hydrodynamic behavior (e.g., the hydrophobic effect), which balance to influence adsorption behavior.

Surface-bound structures were clustered and reweighted in order to give context to our interpretation of statherin adsorption. Low free energy structures for states I, II, and III from Figure 4.1 are illustrated in Figure 4.2. Structures in state I appear to preferentially situate themselves with their charged residues – mostly at the N-terminus – oriented toward the surface. Notably, structure Ia assumes a slab-like conformation but rests precariously on its edge. This case is an example of a stable conformation arising from the balance between charged residues forming favorable surface contacts and C-terminal residues self-associating into a hydrophobic bundle. The same behavior is evident in structures Ia, Ib, and Ic. Proline and tyrosine residues compose much of the C-terminus, and are not only more likely to configure themselves in a hydrophobic bundle but also lack the electrostatic potential necessary to pierce the layer of tightly bound surface waters. Residue E26 stands out in snapshots Ia and Ib where it appears bound to the surface, in disagreement with experiment. [120] In a physical system, surface crowding by other biomolecules would prevent this interaction. [138] States II and III display a variety of extended, disordered conformations. In the examples we provide, it is visually apparent that the N-terminal binding domain strongly associates with both the surface and itself, in several cases forming a distorted α-helix. The globular C-terminus, seen in state I, does not exist in states II and III. Rather, the disrupted hydrophobic residues only loosely associate with each other and are drawn in closer to the surface by electrostatic and van der Waals forces. However, even in these extended states, few residues are able to pierce the layer of water bound by the HAP surface.

Additionally, we reweighted the hydrogen-bond contacts with respect to the SSD of the SN15 domain, and compare the resulting FES to the case of an SN15 fragment in Figure 4.3. We see a greater free energy difference between the solvated and bound states for an SN15 fragment, suggesting that the overall size of the adsorbing molecule plays a role in how tightly it binds to a surface. In the case of statherin, the global free energy minimum at the surface shows greater α-helical content than the SN15 fragment, which could be due to some feature of the C-terminus stabilizing the fold.

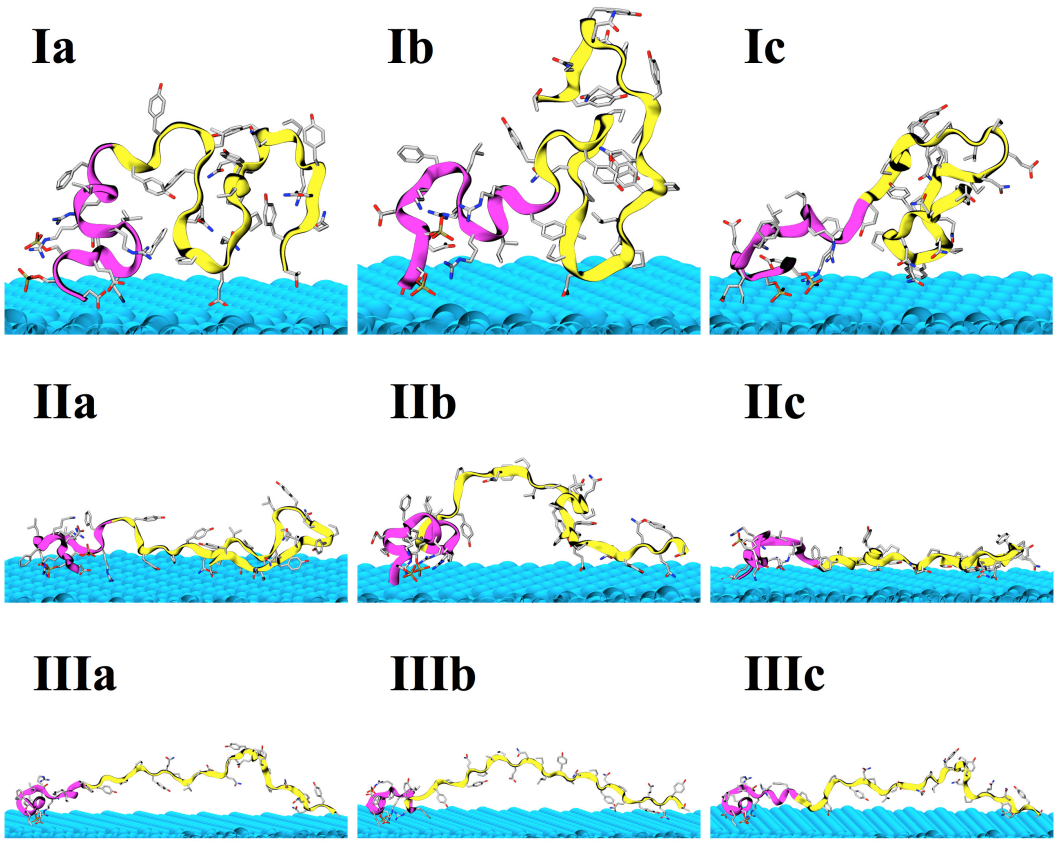


Figure 4.2. Low free energy examples of the structures that populate each state shown in Figure 4.1. In each case the N-terminus (pink) appears tightly bound to the HAP surface, while the remaining residues (yellow) exist in various states of self-association through hydrophobic interactions.

4.3.2 Experimental Fitness

The statherin molecule has been extensively studied over the past few decades. In addition to the attention paid to the SN15 binding domain, the C-terminus has been characterized to some extent through ssNMR experimentation. [30,120,140] We incorporated data on the C-terminus into an experimental fitness function, and performed a frame-by-frame analysis of each adsorbed structure whose COM fell below a SSD of 25 Å. The cutoff of 25 Å encompassed a range of collapsed and extended structures of various orientations, and embodied a sufficient sample for our purposes. The surface set was reweighted (rejecting the first 35,000 – the transient period for the surface set) in order to see how well our simulation replicated experimental findings.

We performed two separate calculations: the first (A) only considered data that described the SN15 domain; the second (B) took into account data for the whole protein. The free energy profiles from each of these calculations are shown in Figure 4.4. We note the similarity between these two profiles and Figure 3.3

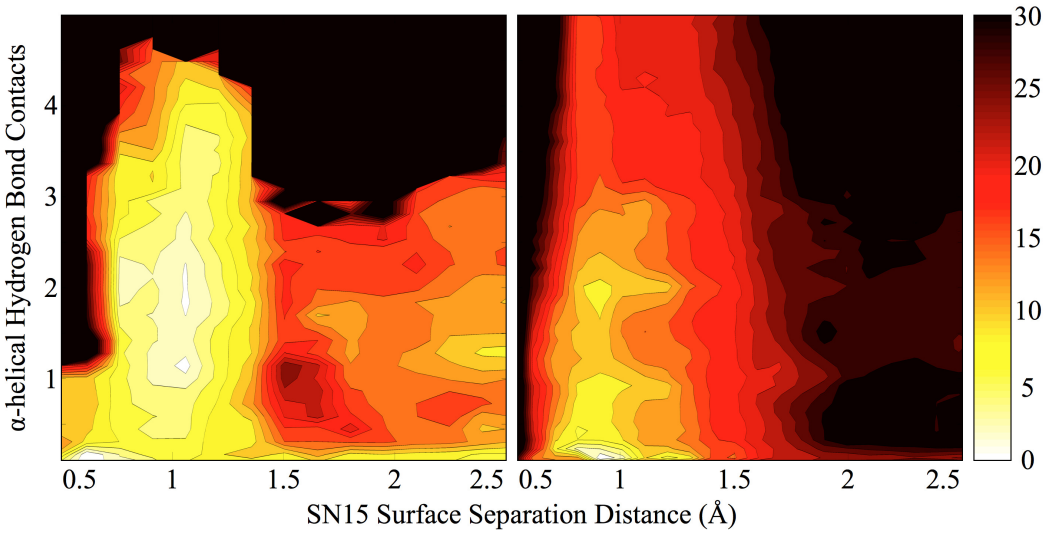


Figure 4.3. The reweighted free energy surface for α-helical hydrogen bonds versus the SN15 domain COM SSD for the statherin system (left) and a system containing a solitary SN15 domain (right). The N-terminal residues of a whole statherin molecule appear slightly more helical than the SN15 domain. Both systems show a similar degree of α-helical content in solution. Isolines are spaced every 2kT whereas the color bar is in units of kT.

in Chapter 3. Specifically, structures of low fitness appear more energetically favorable, whereas a few metastable states appear at the jagged, high-scoring extremes in either profile. As a comparison, we thought it useful to see how well the structures scored for general structural features. Therefore, two supplementary reweighting calculations (C and D) were performed, which omitted side chain tilt angles – specifically, L8 and I11, as well as the F7 and F14 ring tilts. Calculation C paralleled A, using only SN15 data, whereas D paralleled B and used whole protein data.

In the cases of C and D in Figure 4.5, both profiles depict even more pronounced high-scoring metastable states that closely resemble those seen in Figure 3.3. This could suggest an error in the molecular model, the sampling, or the simulation environment. We believe the last option is the most likely culprit as discussed in the previous chapter. The reason for the noticeable disagreement between our simulation and experimental findings could be due to the fact that we are only sampling a single peptide on a pristine [001] HAP surface. Surface protein-protein interactions probably play a significant role in stabilizing the surface-bound conformation of statherin.

Best-fit structures were extracted from each calculation. Quantitative structural information for each best-fit conformation is compared with experimental data and RosettaSurface.NMR results in Table **??**. Structures from calculations A and B closely resemble each other (RMSD = 0.644 Å), so we present B in Figure 4.6. The best-fit structure resembles a constituent of state I in Figure 4.1 – a compact, globular structure whose loosely α-helical N-terminus is tightly bound to the HAP surface. The C-terminus forms a

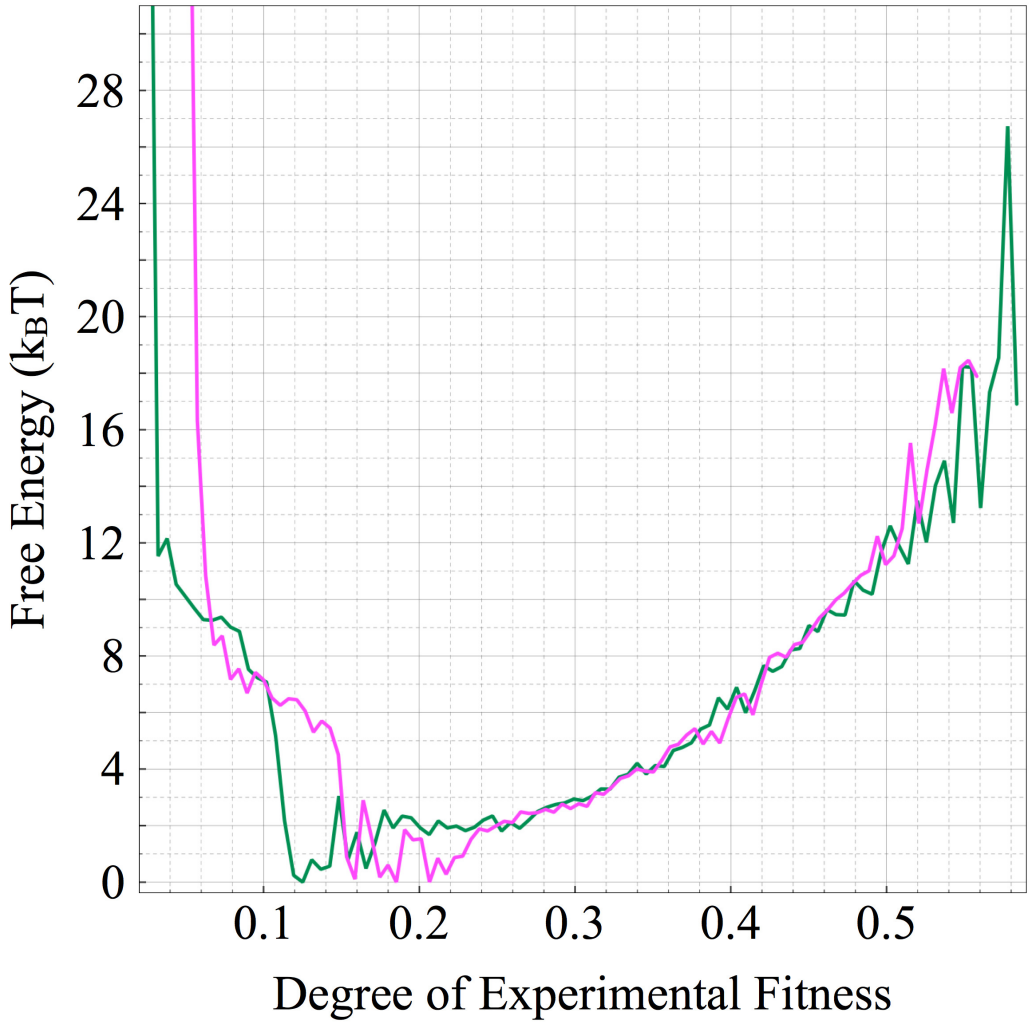


Figure 4.4. Reweighted free energy profiles for experimental fitness calculated from a set of surface-bound structures. The calculation using only SN15 domain data (pink) and the calculation that used whole statherin molecule data (green) show that the statherin molecule weakly fits to experimental values in the simulated adsorbed state.

self-associated hydrophobic core, whose marginal residues interact weakly with the surface. This supports the experimental finding that the trailing residues interact with the proline-rich midsection of statherin. However, no apparent C-terminal α-helicity is noticeable. The best-fit structures from calculations C and D in Figure 4.7 possess N-termini that are more loosely bound to the surface, and C-termini that are more elongated than the best-fit structures from A and B. Such a diverse set of surface conformations may suggest that there may not be a single "precise" adsorption state because of the inherently disordered nature of the statherin molecule. Rather, the protein could exist in a variety of functional states at the HAP surface. A functional state may not need to be any more specific than one whose N-terminus facilitates recognition and binding, leaving the C-terminus free to lubricate, interact with other surface-bound entities, and bind to

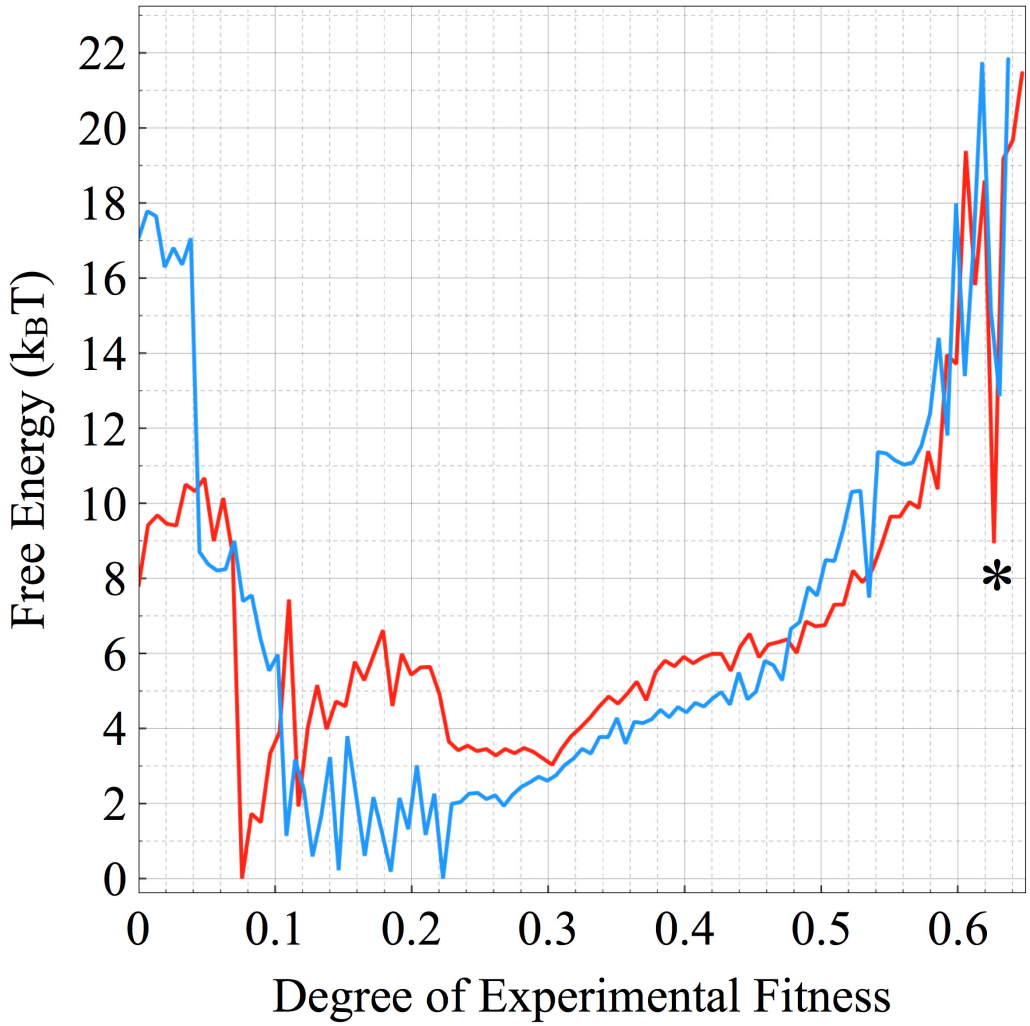


Figure 4.5. Reweighted free energy profiles for experimental fitness calculated from a set of surface-bound structures without using data for sidechain tilt angles. The calculation using only SN15 domain data (red) and the calculation that used whole statherin molecule data (blue) show a similar trend as Figure 4.4, favoring states that weakly fit to experimental data. However, the pronounced metastable state (asterisk) is reminiscent of the fitness profile from Chapter 3, suggesting that high-scoring states are being visited but the simulation lacks a crucial feature necessary for experimental agreement.

exogenous life forms. This scheme is flexible and adaptable to various environmental features, which would be ideal for a salivary protein whose milieu is subject to fluxes in pH and molecular concentrations, not to mention irregular tooth morphology and chipped teeth.

Conclusions

In this final chapter, we examined the adsorption behavior of a whole statherin molecule on hydroxyapatite. The extensive sample of surface-bound structures was reweighted to show three distinct states of varying

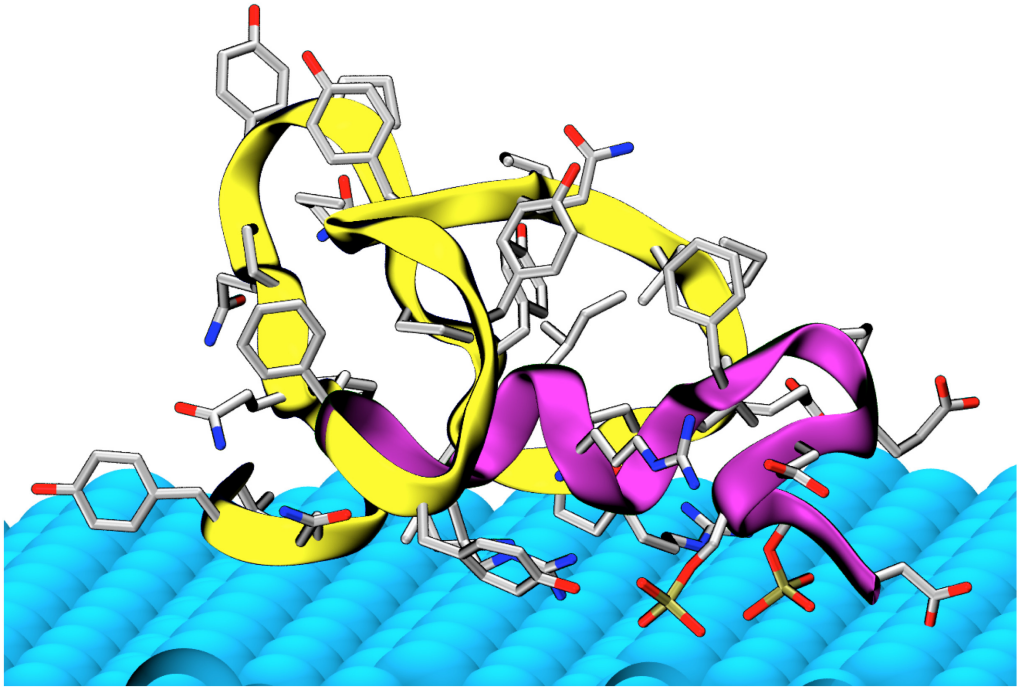


Figure 4.6. The best-fit structure from a 31-parameter fitting calculation. N-terminal residues (purple) are in close contact with the HAP surface; its basic residues appear to fold and assist with surface binding. The trailing C-terminus (yellow) forms a compact hydrophobic globule consisting of tyrosine and proline residues; its polar glutamine residues point away from the globule into solution.

elongation at the HAP surface, the most favorable of which is a compact globular conformation. We highlight two pathways for statherin recognizing and binding to a HAP surface: one involves a compact globule coming to rest on the surface without undergoing substantial distortion; whereas the second suggests the N-terminus first hitches onto the surface, from which point the entire protein succumbs to the attractive surface potential. Surface-bound structures generally show low to moderate agreement with experimental data, suggesting that a more heterogeneous simulation environment is required to reproduce experimental conditions.

The statherin/HAP system is the largest system we have studied to date with PTMetaD-WTE method for the purpose of studying adsorption behavior. Furthermore, to the best of our knowledge, statherin is the largest protein ever simulated in the presence of a surface with the Metadynamics algorithm, offering new evidence that such a complex system can be rationally modeled and converged. We show in this work that biasing a CV that describes general structural behavior (i.e. the radius of gyration) may be advantageous compared to biasing several specific degrees of freedom (e.g. hydrogen bonds), especially when the surface is known to induce folding. The most notable reason for this is the SN15 H-bond contact FES comparison shown in Figure 4.3. A similar degree of low free energy α-helical conformational space was sampled while statherin extended and contracted at the surface under the influence of the MetaD bias. Specific challenges for future

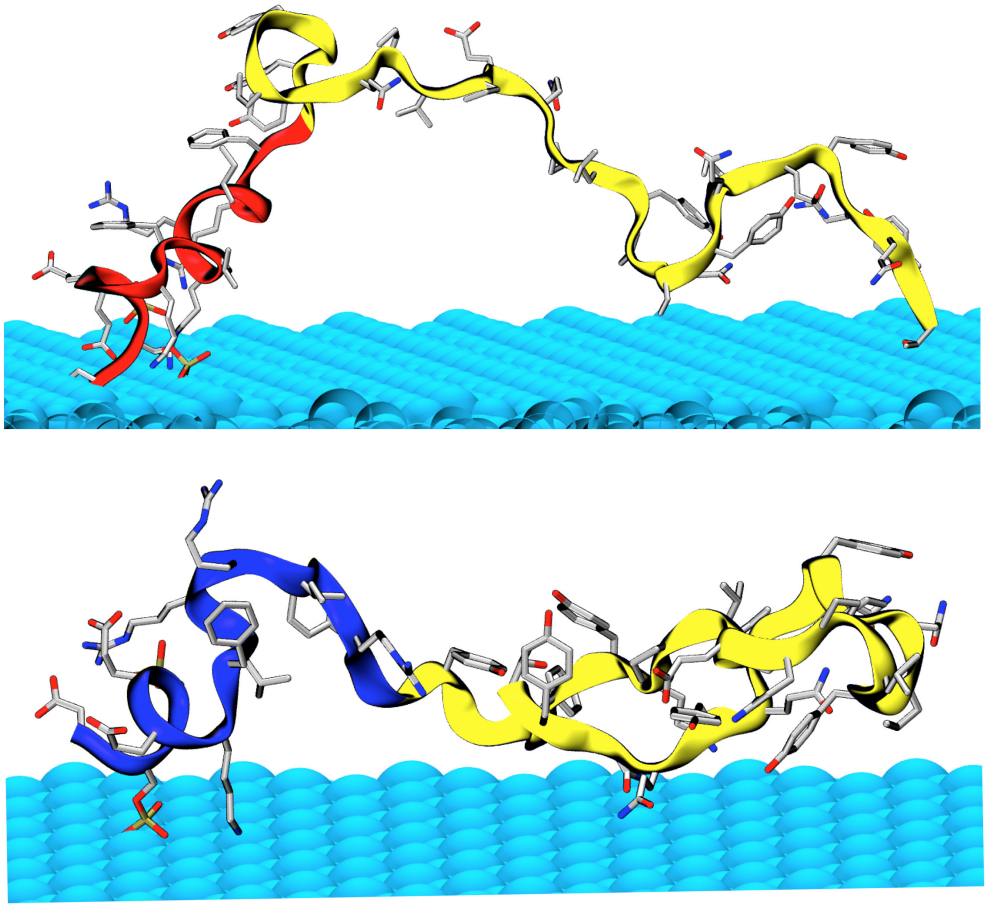


Figure 4.7. Best-fit structures from EFF calculations performed without considering side chain tilt angles. The best-fit structure from a 13-parameter calculation (red N-terminus) shows a high degree of α-helicity in its N-terminus, but lacks C-terminal detail seen in experiments. The best-fit structure from a 22-parameter calculation (blue N-terminus) shows a more distorted N-terminus. The C-terminal proline and tyrosine residues interact in this case, but no α-helical character is apparent in the C-terminus.

work on large proteins will concern proper selection of CVs, ascertaining convergence, and mimicking an experimental or living environment. We emphasize that none of this would be as relevant if not for extensive experimental characterization. Experimental techniques like ssNMR and SFG, coupled with simulations such as the one presented here, stand to offer a dramatically enhanced picture of adsorption phenomena in the coming years.

Table 4.2. Comparison Between Experiment, RosettaSurface.NMR, and Best-Fit Structures

Measurement[a]	Experiment[b]	RosettaSurface.NMR[b]	A[bc]	B[bc]	C[bc]	D[bc]
pS3 ϕ dihedral angle	-60 ± 10	-63.9 ± 12	**-63.2**	**-75.6**	**-59.0**	**-73.0**
pS3 – F7 Δ(C' – C')	4.3 ± 0.2	4.3 ± 0.3	**4.2**	**4.4**	**4.2**	**4.1**
E4 $\Delta(C_\delta$ – HAP)	> 6.5	11.7 ± 1.4	**7.0**	**7.5**	**9.3**	**13.5**
E5 $\Delta(C_\delta$ – HAP)	4.25 ± 0.09	5.0 ± 0.5	**7.4**	**9.3**	**4.3**	**8.0**
K6 $\Delta(N_\zeta$ – HAP)	3.5 ± 0.5	4.3 ± 0.8	**11.3**	**10.8**	**2.8**	**2.2**
F7 Δ(RING – HAP)	> 6.5	11.3 ± 0.4	**13.6**	**13.4**	**12.5**	**13.8**
L8 ϕ dihedral angle	-60 ± 10	-62.7 ± 6.2	**-59.9**	**-69.7**	**-56.1**	**-60.7**
L8 – G12 Δ(C' – C')	4.8 ± 0.4	4.3 ± 0.5	**4.5**	**4.5**	**5.0**	**5.0**
R9 $\Delta(C_\zeta$ – HAP)	4.62 ± 0.29	4.6 ± 0.4	**4.5**	**4.7**	**3.6**	**12.3**
R10 $\Delta(C_\zeta$ – HAP)	4.53 ± 0.16	4.8 ± 0.3	**8.8**	**8.0**	**14.1**	**20.7**
G12 ϕ dihedral angle	-73 ± 10	-71.4 ± 8.2	**-70.2**	**-75.2**	**-71.8**	**-74.1**
R13 $\Delta(C_\zeta$ – HAP)	> 7.0	7.6 ± 2.5	**4.0**	**3.9**	**9.2**	**10.9**
F14 Δ(RING – HAP)	4.55 ± 0.25	12.4 ± 2.8	**10.1**	**10.6**	**18.5**	**14.7**
average backbone tilt	85 ± 15	–	**75.5**	**76.5**	**69.3**	**73.2**
P23 – P33 Δ(4'H – C')	8.8 ± 0.8 10.5 ± 1.0	12.8 ± 2.5	10.2	**8.9**	18.7	**2.9**
P23 – Y34 Δ(4'H – C')	8.8 ± 0.8 10.5 ± 1.0	13 ± 2.6	12.1	**10.9**	18.5	**2.6**
E26 $\Delta(C_\delta$ – HAP)	> 6.5	21.3 ± 8.5	9.3	**8.7**	25.0	**10.9**
P28 $\Delta(C_\delta$ – HAP)	> 7.28	21.7 ± 8.7	6.1	**5.2**	17.5	**13.8**
P33 – Y34 Δ(C' – C')	3.12 ± 0.13	3.1 ± 0.1	3.6	**3.3**	3.3	**3.1**
P33 – Y38 Δ(C' – N)	5.3 ± 0.5	6.4 ± 0.4	10.2	**9.4**	13.9	**8.8**
Y34 ϕ dihedral angle	-75 ± 15	-67.2 ± 8.5	-118.1	**-74.8**	-94.2	**-55.2**
Y34 ψ dihedral angle	-40 ± 10	-36.8 ± 8.2	2.7	**-94.7**	175.2	**-32.3**
Y34 – Y38 Δ(C' – N)	4.0 ± 0.5	4.3 ± 0.6	8.5	**8.1**	10.8	**8.9**
F7 Θ ring axis tilt	11 ± 6	–	**8.1**	**12.0**	90.5	37.6
F7 Ψ ring twist	15 ± 6	–	**5.7**	**6.3**	15.3	28.0
F7 ρ ring normal tilt	76 ± 6	–	**82.0**	**78.1**	15.3	57.4
L8 C_β – C_γ tilt	27 ± 12	–	**0.0**	**-17.2**	14.8	32.1
I11 C_α – C_β tilt	-56 ± 6	–	**-38.6**	**-35.7**	-17.9	28.8
F14 Θ ring axis tilt	61 ± 6	–	**33.2**	**39.6**	90.7	53.7
F14 Ψ ring twist	34 ± 6	–	**17.6**	**25.5**	31.0	35.1
F14 ρ ring normal tilt	35 ± 6	–	**58.4**	**54.9**	31.0	48.8

[a] Δ(X – X) indicates the distance between two objects. All tilt angles were measured with respect to the HAP surface normal.
[b] Angles are in degrees, distances are in Ångsrtoms.
[c] Values used in EFF calculation are in boldface.

Coda

At the onset of the work presented in this thesis, specific challenges were facing the surface science community from both experimental and theoretical points of view. While experimentalists had at their disposal numerous sophisticated surface analysis methods for studying biomolecules at interfaces, they lacked a unifying factor that could bring their individual measurements into sharp clarity. Computational scientists sought to provide this unifying factor, and molecular simulation appeared to be a sensible solution. However, early adsorption simulations used rudimentary sampling techniques, which provided some detail but were ineffectual at addressing the onerous sampling challenge inherent to simulating three interfaces (protein/liquid, protein/surface, liquid/surface). As we have shown, the robust sampling power of the Parallel Tempering Metadynamics (PTMetaD) scheme, coupled with the reduced computational cost that the Well-Tempered Ensemble (WTE) allows, improves upon past strategies in a quantifiable manner. The accomplishments and contributions made by this work can be summarized as the following:

- Extended the applicability of the PTMetaD enhanced sampling method to study protein adsorption; specifically, by designing, testing, and implementing appropriate collective variables for simulation and analysis.
- Showed that the WTE reduces the computational cost of large-scale all-atom PTMetaD simulations.
- Showed that results from a PTMetaD-WTE adsorption simulation can be compared to, and validated by, experimental measurements.
- Provided predictions for low free energy surface-bound conformations that agree with experimental findings and other structure prediction algorithms.
- Evaluated the efficacy of modern adsorption simulation protocol; specifically, in regards to choice of force field and simulation environment.
- Gained attention in the simulation and experimental communities through conversations, publications, and talks at national conferences.

Impact

The findings reported in this work should appeal generally to anyone interested in simulating large-scale all-atom simulations, and specifically to those who study the behavior of complex molecules at interfaces. We note that several domestic and international research groups have taken notice of our contributions. The PTMetaD-WTE scheme for all-atom simulations has been employed for the purpose of studying: the dimerization and trimerization of a homooligomeric peptide, [144] the free energy landscape of a small peptide folding, [145] structural and mechanistic details of allostery in a transcription factor, [146] and structural changes in a protein with oncogenic mutations. [126] The results from each of these studies aligned well with experimental findings, highlighting the robustness of the PTMetaD-WTE method. Furthermore, the effectiveness of the PTMetaD-WTE method has been cited in fundamental thermodynamic research that examined phase transitions. [147] Our work has also gained notoriety and validity by being cited by several groups interested in simulating adsorption phenomena. [148–151] One other group has also begun using PTMetaD-WTE for their adsorption simulations. [31] The applicability of PTMetaD-WTE is clear and broad – as evidenced by these published studies – and likely to remain relevant far into the future.

Future Work

A particularly exciting aspect of this work is the fact that it provides a strong basis for future projects. The large-scale simulation work that the PTMetaD-WTE method facilitates has the potential to enhance many fields of study. Much work is yet to be done, however, in terms of simulating large molecular systems while keeping computational costs low. In addition to the impact our work has had on other groups, the Pfaendtner Research Group has used this enhanced sampling method to study structural changes in human fibrin protein, [152] and how ionic liquids affect the conformational free energy of glucose. [153] For protein adsorption simulation, this work has created more questions than could have possibly been answered in a single thesis. To begin, both LK peptides and statherin are thought to aggregate in solution. Undoubtedly, this multibody behavior affects adsorption mechanics and the free energy landscape of the surface-bound state. Future simulations should take into account heterogeneous mixtures, the affect of molecular crowding in solution and at an interface. Moreover, many adsorption simulations currently assume ideal surface and solvent conditions. In reality, the two phases are in flux, and may affect how solute molecules behave at the interface (e.g., nucleated mineral growth, pitting, or diffusion). This is related to an active area in adsorption simulation research, focused primarily on ascertaining the accuracy of current models and force fields, with a long-term goal of producing one or more interfacial force fields. [28, 81, 154, 155] Additionally, temperature

appears to affect protein adsorption in different ways, depending on surface functionality. With increasing temperature, preliminary data shows adsorption free energies for hydrophilic and hydrophobic surfaces to increase and decrease, respectively. A full thermodynamic analysis of this phenomenon should be carried out, as it may greatly enhance our current understanding of adsorption.

Finally, there is room for development in regards to the PTMetaD-WTE method itself in two ways. Firstly, a known issue in the simulation community is that Metadynamics results do not lend themselves to kinetic interpretation. The rich dataset that PTMetaD-WTE simulations provide require a specialized tool that would extract mechanistic details surrounding adsorption and conformational change at interfaces. Fortunately, some work is already underway with this in mind. [135,156,157] Secondly, where comprehensive sampling was an issue addressed by this work, sampling efficiency will be a future challenge. In concrete terms, sampling efficiency can be measured in terms of convergence times (for Metadynamics) and round-trip times (for Parallel Tempering) – future systematic algorithms would work to minimize both. Improving Metadynamics convergence times would be a difficult exercise. On the other hand, a method similar to Metadynamics or Wang-Landau sampling could conceivably act on a set of replicas in PT to induce efficient diffusion through temperature space. Regardless, important work has been, is, and will continue to be done to further the validity and legitimacy of molecular simulation in surface science. As with any theoretical model, scientific progress will continue to push the limits of the model until it breaks, then fix it, and continue moving forward.

Appendix I

Further Details of MD and Metadynamics Simulations

The simulation box was constructed as follows: First, a conjugate gradient energy minimization was performed for approximately 5000 steps. Next, a 1 ns NPT simulation was performed with a stochastic velocity rescale thermostat and Berendsen barostat at 300K and 1 atm in order to equilibrate the box dimensions. [93, 94] We obtained starting structures for the unfolded state through thermal annealing, where the system was heated to 600 K for 1 ns, allowing it to equilibrate to an unfolded state. Before either the PTMetaD or PTMetaD-WTE simulations were initiated, short 500 ps simulations were performed on each individual replica so that they could equilibrate to their specific temperatures, ensuring that the replicas were not at identical points in phase spaceat the beginning of the formal simulations. The C_α RMSD (compared to the lowest energy NMR structure) of the replicas that started in the unfolded state ranged from 4 to 9 Å. We performed the PTMetaD-WTE simulations using a two-step process. We first performed a 1 ns PTMetaD simulation biasing only the potential energy (PE) CV. The cumulative bias potential from this simulation was then implemented as a static bias potential (no further Gaussians were added to the PE CV) while the other folding CVs were biased according to the well-tempered metadynamics algorithm. The parameters we set for the work reported in Chapter 1 are as follows. After initial testing, we selected values of 12 and 24 for γ. The Gaussian width for the PE CV was set to 500 kcal/mol.

Functional Form of Collective Variables and Associated Parameters for PTMetaD

Two collective variables were biased during the simulations: 1) backbone hydrogen bonds and 2) hydrophobic core contacts. Mathematically, the CVs are described by well-tuned switching function shown here:

$$s_{ij}(t) = \frac{1 - \left(r_{ij}(t)/r_o\right)^n}{1 - \left(r_{ij}(t)/r_o\right)^m}$$

where r_{ij} is the distance between two CV atoms at time t, r_o defines a fully formed contact (r_o (Hbond) = 2.5 Å, r_o (Hcore) = 5.0 Å), and n and m are exponential factors that correspond the variable the bond length. For both CVs, the value of n and m were 8 and 12, respectively. Hydrogen bonds were counted as individual pairs, whereas each hydrophobic atom combination was counted to define the folded core. [44,158,159] These two particular CVs can be expressed as:

$$S_{Hbond}(t) = \sum_{i=j=1}^{HB} s_{ij}(t)$$

$$S_{Hcore}(t) = \sum_{i=1}^{HC} \sum_{j=1}^{HC} s_{ij}(t)$$

where $S(t)$ is the value of the CV at time t, HB is the number of α-helical hydrogen bonds in the native state, HC is the number of C_γ side-chain atoms in the hydrophobic core, and $s_{ij}(t)$ is the expression in equation 1. The i and j sets for hydrogen bonds contain the relevant hydrogen and oxygen indices, respectively. The i and j sets for the hydrophobic core CV contain identical sets indices of the atoms that describe the core.

To construct the MetaD bias potential used to study tryptophan cage folding, a Gaussian deposition rate of 2 hills/ps was chosen, σ values for the hydrogen bond and hydrophobic contact CVs were both 0.4, and the initial Gaussian hill height was 0.6 kcal/mol with the separate Well-Tempered Metadynamics scaling parameter γ_{WTM} set at 8. For clarity we note that the PTMetaD-WTE algorithm can have two different γ parameters: one sets the magnitude of the PE fluctuations and the other sets the magnitude of the bias potential for the additional CVs. Exchanges were attempted between replicas every 0.4 ps with the PTMetaD algorithm implemented in PLUMED/GROMACS. [51, 53]

The Quantitative Comparison of PTMetaD-WTE and PTMetaD Free Energy Surfaces

To quantify the differences between the PTMetaD-WTE and the PTMetaD FESs, we used the RMSD of the PTMetaD-WTE FES aligned to that of the reference (simulation I). This has been shown to be effective in comparing biased MD simulations. [41, 160] Once the estimate of the FES (F_{REF}) from the reference simulation and the FES (F) from the PTMetaD-WTE simulations are aligned to their mean values, the RMSD between the FESs after a given amount of simulation time is defined as:

$$RMSD_{FES} = \beta \cdot \sqrt{\int_{\Omega} \left(F_{REF}(s) - F(s)\right)^2 ds}$$

The region Ω is defined as the set of points on the reference FES with free energy within $8k_BT$ of the global minimum.

The Analysis of the RTT and Calculation of Folding Events

The round trip times in Table 1.1 were calculated by averaging the time each replica took to traverse temperature space defined as: diffusing from its original temperature, to both the hottest and coldest temperatures, and back to start. A folding/unfolding event was described as any instance where the C_α RMSD, compared to the NMR structure, of trp-cage passed from less than 2 Å to greater than 8 Å, or the reverse. This metric has been successfully used to characterize folding of the villin headpiece subdomain. [161] The error estimates in Table 1.1 (for RTT and folding time) are the estimated standard error of the mean for the entire sample of round trips and folding event times.

Free Energy Difference Versus Time for Reference Simulation I

Since it is extremely challenging for this system to achieve a perfectly converged reference FES, we assess the quality of our reference simulation by evaluating the convergence of an experimental observable as a function of the simulation time. An experimental quantity with which we can directly evaluate our reference simulation is the reported folding free energy difference of trp-cage (-0.5 kcal/mol) at 300K. [162] Figure A.2 was obtained by first projecting the FES only onto the Hbond collective variable since the folded and unfolded states are well separated in that projection. Next the free energy difference was obtained by comparing the minima in the folded and unfolded states. Comparing the final 25 ns/replica of the simulation, the mean

difference between our simulation and experiment is about 0.35 kcal/mol, which is within the range that others obtained using the same family of force fields from brute force simulations of an explicitly solvated villin headpiece. [161]

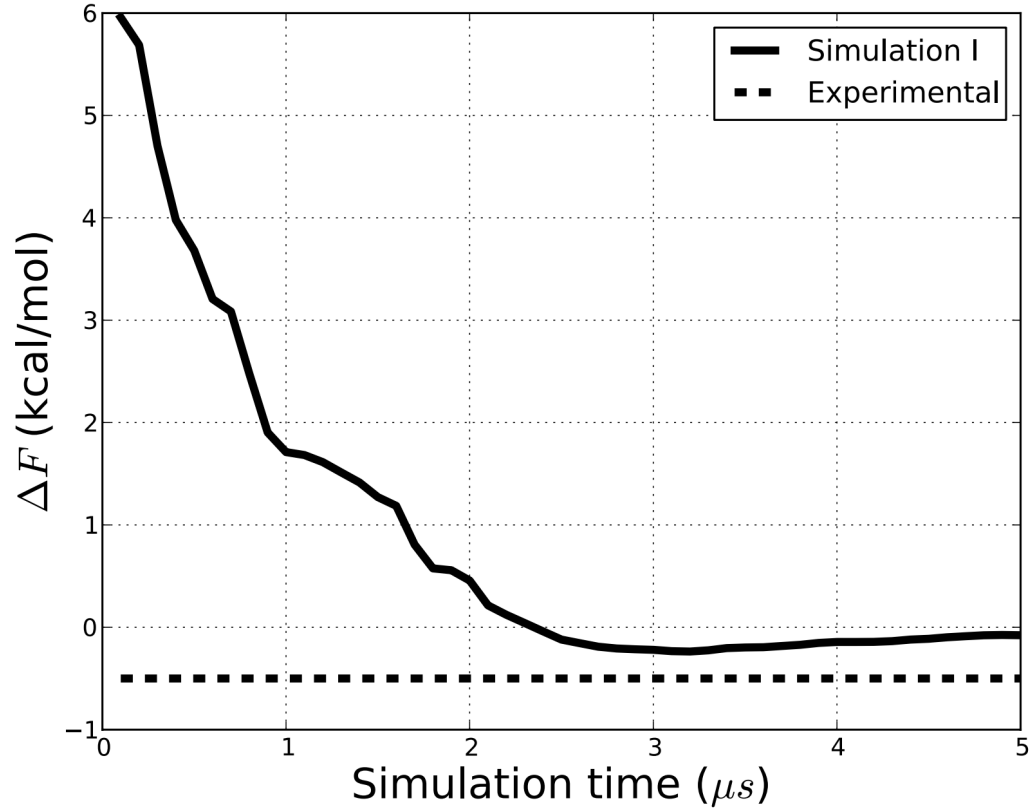


Figure A.1. Assessing the quality of the reference FES: The free energy difference between folded and unfolded states versus simulation time per replica for the reference PTMetaD simulation (solid line), and the experimental value [162] (-0.5 kcal/mol) at 300K (dashed line).

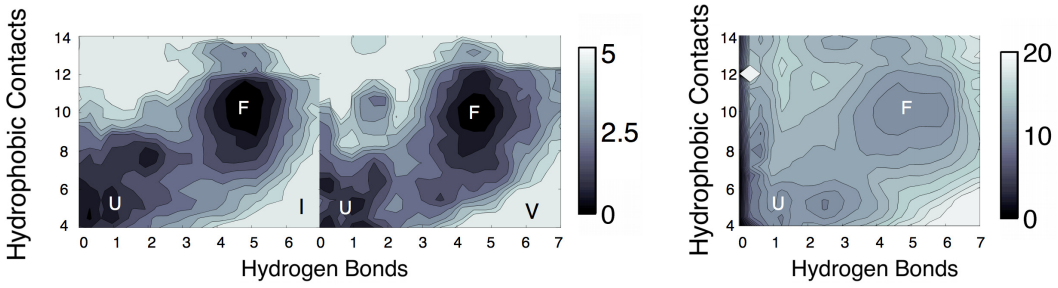


Figure A.2. To demonstrate the presence of hidden energy barriers(and their energetic nature), the reference simulation (left, I) is compared with the FES obtained after a 50 ns/replica simulation (middle, V) and contrasted with a single replica 500 ns long well-tempered metadynamics simulation initiated from the unfolded state at 300K (right). Note the vastly different energy scales: the left side ranges from 0 to 5 kcal/mol compared with 0 to 20 kcal/mol on the right side. This is a result of the significant overfilling of the unfolded region due to hidden degrees of freedom that nontrivial features of the FES, but are not accounted for by the metadynamics CVs.

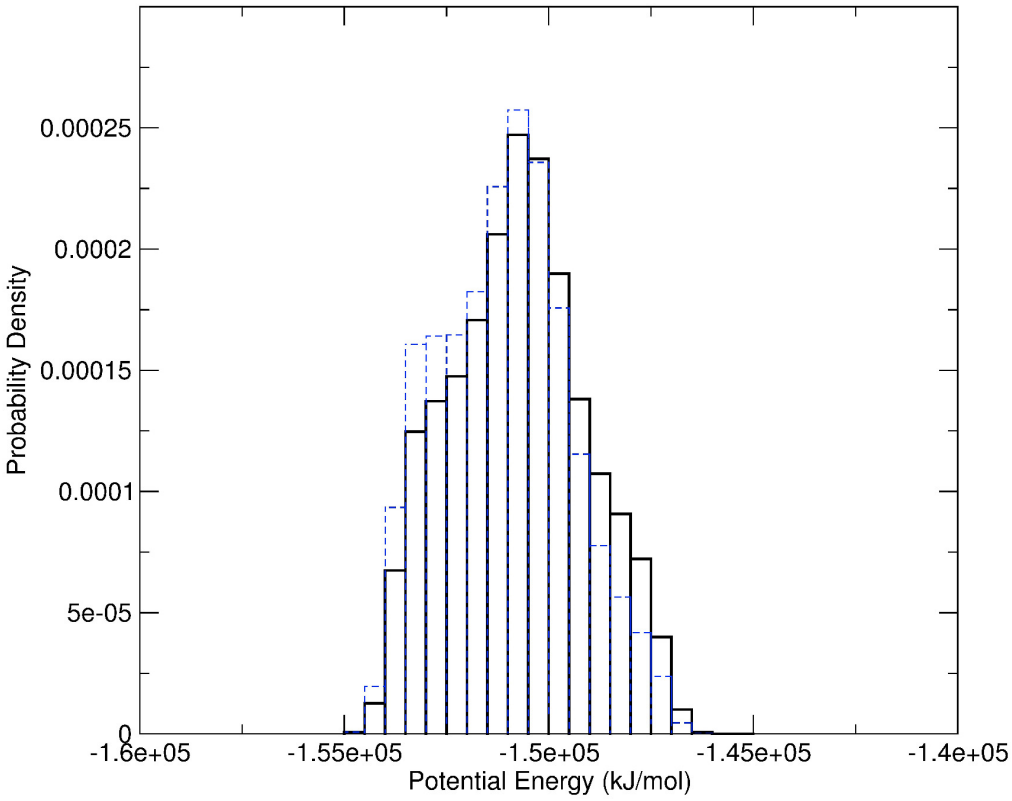


Figure A.3. Comparison of the potential energy distributions from the first 50 ns (dashed line) and last 50 ns (solid line) of simulation IV. The bin width is 500 kJ/mol and the potential energy fluctuations of the simulation average to approximately 1600 kJ/mol.

Appendix II

Force Field Comparison

Since MD simulation of peptide adsorption is still a relatively new area of research, questioning the influence of the force field is of utmost importance. To date, solid-liquid interfaces have not received much consideration in force field development. Prior simulations have been carried out that used generalized force field values to parameterize their surfaces. Others, including our own work, assigned amino acid force field parameters to solvent-exposed head groups, which corresponded to residues with identical chemical functionality (e.g., alanine and glutamic acid). Recently the Latour group performed a series of PT simulations on two LK peptides adsorbed to methyl- and carboxyl-terminated SAM surfaces. In their study, they compared three force fields: AMBER94, CHARMM22 with CMAP, and OPLS-AA. [81] A qualitative comparison of their simulation results with experimental findings suggested that of the three force fields, CHARMM22 produces the most accurate adsorption model.

To contribute to the ongoing discussion on modeling protein adsorption, we provide our own force field comparison using converged free energy calculations. PTMetaD-WTE simulations were set up and carried out – in the same manner as we describe in Chapter 2 – on a system containing a single LKα14 peptide and methyl-terminated SAM surface. The force fields we considered were: AMBER99SB (II), CHARMM22 with CMAP (V), and OPLS-AA/L (VI). Simulations V and VI each consisted of 25 replicas and ran for 50 ns/replica, converging after roughly the same amount of sampling as simulation II. Figure 2.2 (main text) clearly shows that while the AMBER99SB and CHARMM22 simulations converge to within less than $2k_{\mathrm{B}}T$ of each other, the OPLS-AA/L simulation differs from each by a substantial margin. These energetic similarities and differences are further exemplified in Figure A.4. Again, AMBER99SB and CHARMM22 appear to show similar behavior. Both have a strong preference for the adsorbed folded state in the upper left of the FES, as well as similar wells in the adsorbed, unfolded region in the lower left. Additionally, the solvated regions in both II and V are more or less evenly distributed along the entirety of the Hbond CV, with simulation V showing a slightly greater preference for the folded state than simulation II (not shown).

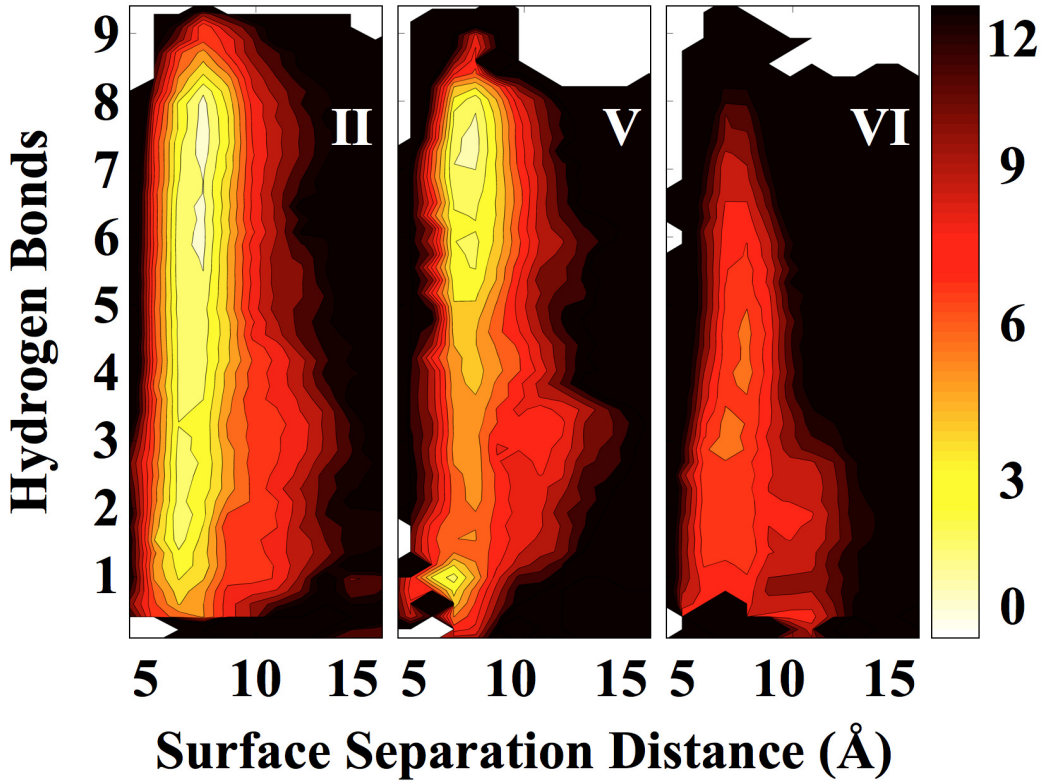


Figure A.4. Free energy surfaces of the biased Metadynamics collective variables for simulations (**II**) LKα14 on methyl SAM, (**V**) LKα14 on methyl SAM with CHARMM22/CMAP, and (**VI**) LKα14 on methyl SAM with OPLS-AA/L. The colorbar is in units of $k_{\mathrm{B}}T$ with isolines spaced every $2k_{\mathrm{B}}T$.

The preferred LKα14 adsorption state in the OPLS-AA/L simulation, on the other hand, ranges from the unfolded state to a slightly α-helical conformation, and also strongly prefers the unfolded state in solution (not shown). The energetic differences between the FESs in Figure A.4 most likely result from differing dihedral potentials since the non-bonded parameters for leucine and lysine residues do not vary by much between force fields. This point is emphasized by Figure A.5, which shows reweighted Ramachandran free energy surfaces for simulations II, V, and VI. Each surface is clearly different from its neighbors and indicative of the structures it describes. We performed a clustering analysis on the adsorbed peptides found in simulations V and VI, and reweighted the cluster centers. Figure A.6 follows the same organization as Figure 2.5 and Figure 2.6 in Chapter 2, showing the five lowest free energy surface conformations for simulations V and VI. Although CHARMM22 strongly favors a compact α-helix at the surface, AMBER99SB produces similar low free energy structures despite its comparatively more degenerate Ramachandran surface. In Figure A.6 OPLS-AA/L is the outlier again, echoing what could be inferred from Figure A.4 and Figure A.5: that its surface structures consist of coils and α-helices of varying shapes, and exhibit some β-sheet character as well.

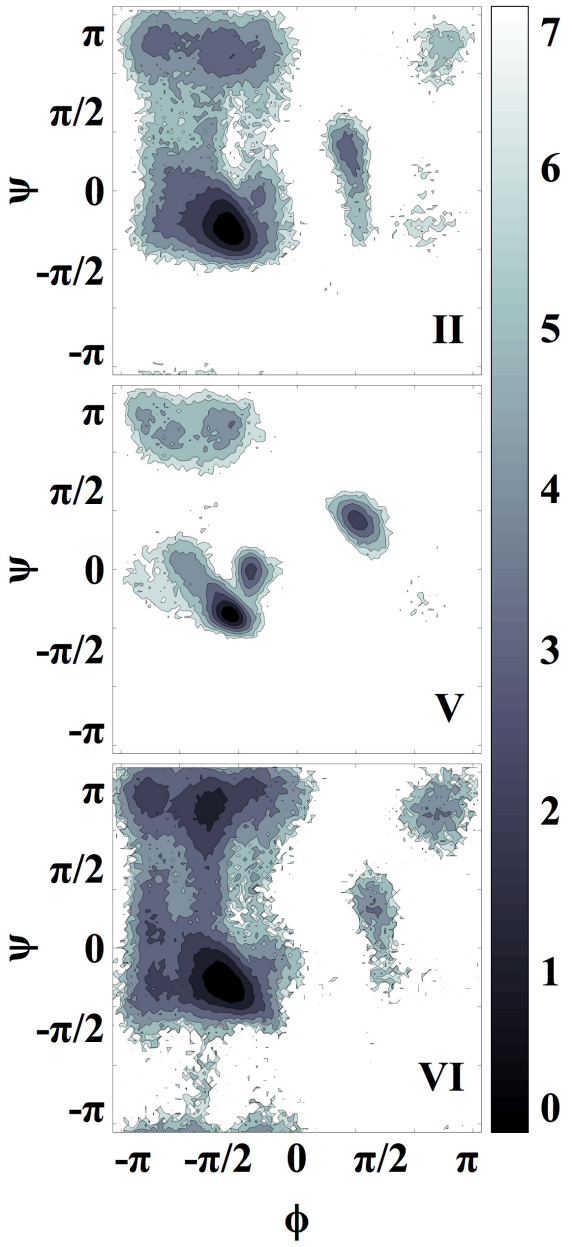
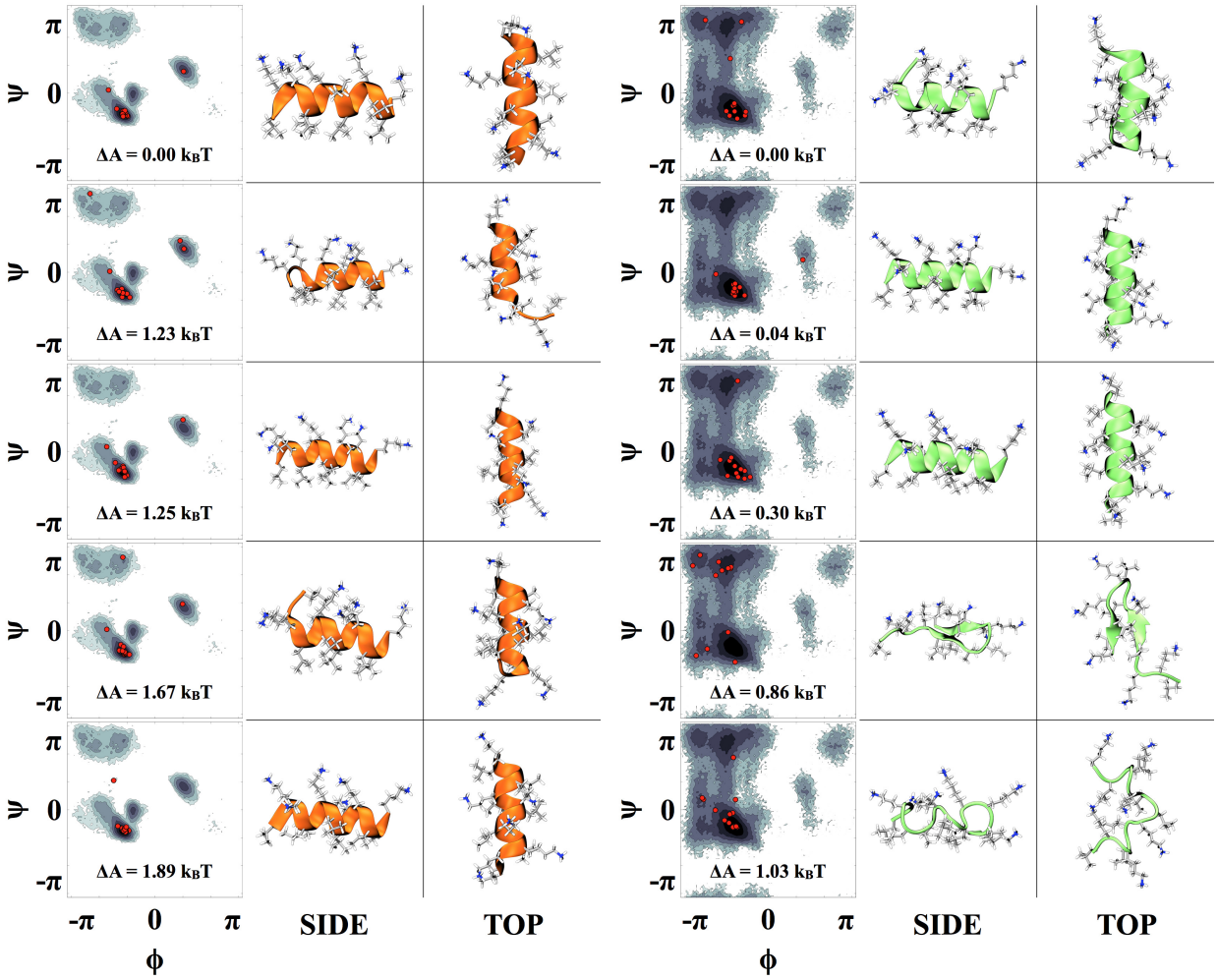


Figure A.5. Reweighted Ramachandran free energy surfaces for adsorbed structures from simulations (**II**) LKα14 on methyl SAM, (**V**) LKα14 on methyl SAM with CHARMM22/CMAP, and (**VI**) LKα14 on methyl SAM with OPLS-AA/L. The colorbar is in units of $k_{\mathrm{B}}T$ with isolines spaced every $k_{\mathrm{B}}T$.



(a) LKα14 on a CH_3-terminated SAM (CHARMM22 with CMAP)

(b) LKα14 on a CH_3-terminated SAM (OPLS-AA/L)

Figure A.6. The most probable adsorbed conformations of LKα14 found in (**a**) simulation V, and (**b**) simulation VI, and their associated free energy differences (with respect to the lowest free energy state) calculated from a reweighted clustering analysis. Left column: cluster center (ϕ, ψ) dihedral angles (shown as red dots) overlaid on the Ramachandran FES. Center column: side view of the cluster center structure. Right column: top view of the cluster center structure.

Appendix III

Temperature Distribution Used in Simulation (Kelvin)

310
314.772
319.524
324.428
329.092
334.199
337.71
343.44
351.153
354.653
360.391
367.877
374.48
378.083
387.637
391.122
397.011
407.101
414.209
419.345
428.953
434.475
443.38
450

Snapshot of Best-Fit Structure from a 22-Parameter EFF Calculation

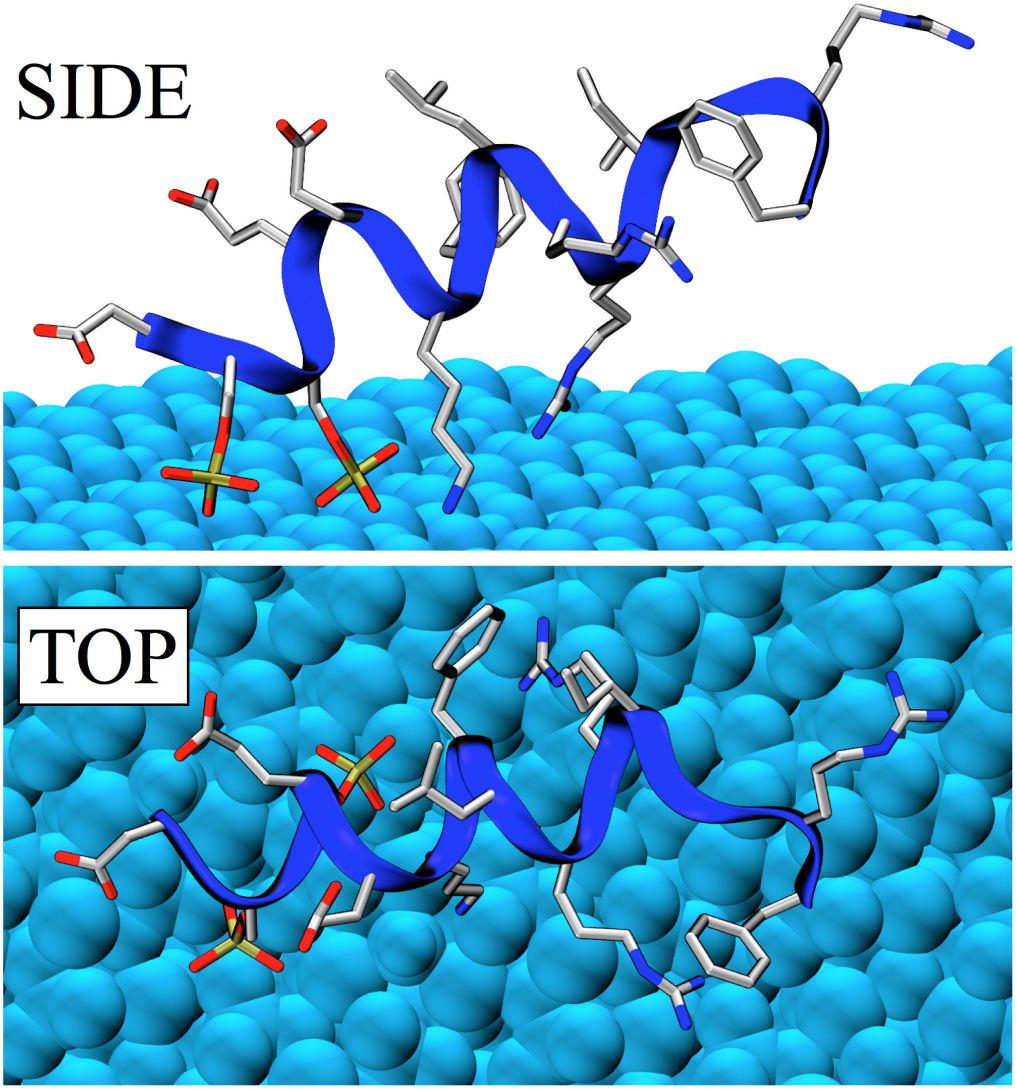


Figure A.7. Side and top views of the best-fit SN15 structure from a 22-parameter experimental fitness calculation. Hydrogen atoms have been removed for clarity.

Bibliography

[1] D. X. Fu, A. Libson, L. J. W. Miercke, C. Weitzman, P. Nollert, J. Krucinski, and R. M. Stroud. Structure of a glycerol-conducting channel and the basis for its selectivity. *Science*, 290:481–486, 2000.

[2] V. Santori, M. Molloy, and T. Rabilloud. Membrane proteins and proteomics: Un amour impossible? *Electrophoresis*, 21:1054–1070, 2000.

[3] I. Banerjee, R. C. Pangule, and R. S. Kane. Antifouling coatings: Recent developments in the design of surfaces that prevent fouling by proteins, bacteria, and marine organisms. *Adv. Mater.*, 23:690–718, 2011.

[4] A. K. Gupta and M. Gupta. Synthesis and surface engineering of iron oxide nanoparticles for biomedical applications. *Biomaterials*, 26:3995–4021, 2005.

[5] S. Tombelli, A. Minunni, and A. Mascini. Analytical applications of aptamers. *Biosens. Bioelectron.*, 20:2424–2434, 2005.

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

[7] F. Ganazzoli and G. Raffaini. Computer simulation of polypeptide adsorption on model biomaterials. *Phys. Chem. Chem. Phys.*, 7:3651–3663, 2005.

[8] R. A. Latour. Molecular simulation of protein-surface interactions: Benefits, problems, solutions, and future directions (review). *Biointerphases*, 3:FC2–FC12, 2008.

[9] M. Rabe, D. Verdes, and S. Seeger. Understanding protein adsorption phenomena at solid surfaces. *Adv. Colloid Interfac.*, 162:87–106, 2011.

[10] L. M. Szott and T. A. Horbett. Protein interactions with surfaces: Computational approaches and repellency. *Curr. Opin. Chem. Bio.*, 15:683–689, 2011.

[11] S. J. Lee and K. Park. Protein-interaction with surfaces – separation distance-dependent interaction energies. *J. Vac. Sci. Technol. A*, 12:2949–2955, 1994.

[12] S. M. McDonald, A. White, P. Clancy, and J. W. Brady. Binding of an antifreeze polypeptide to an ice water interface via computer-simulation. *AIChE J.*, 41:959–973, 1995.

[13] A. H. Juffer, P. Argos, and J. DeVlieg. Adsorption of proteins onto charged surfaces: A Monte Carlo approach with explicit ions. *J. Comput. Chem.*, 17:1783–1803, 1996.

[14] A. L. Cheng and K. M. Merz. Ice-binding mechanism of winter flounder antifreeze proteins. *Biophys. J.*, 73:2851–2873, 1997.

[15] A. M. Bujnowski and W. G. Pitt. Water structure around enkephalin near a PE surface: A molecular dynamics study. *J. Colloid Interf. Sci.*, 203:47–58, 1998.

[16] C. M. Wijmans and E. Dickenson. Brownian dynamics simulation of a bonded network of reversibly adsorbed particles: Towards a model of protein adsorbed layers. *Phys. Chem. Chem. Phys.*, 1:2141–2147, 1999.

[17] R. E. Anderson, V. S. Pande, and C. J. Radke. Dynamic lattic Monte Carlo simulation of a model protein at an oil/water interface. *J. Chem. Phys.*, 112:9167–9185, 2000.

[18] C. M. Shepherd, K. A. Schaus, H. J. Vogel, and A. H. Juffer. Molecular dynamics study of peptide-bilayer adsorption. *Biophys. J.*, 80:579–596, 2001.

[19] S. Mantero, D. Piuri, F. M. Montevecchi, S. Vesentini, F. Ganazzoli, and G. Raffaini. Albumin adsorption onto pyrolytic carbon: A molecular mechanics approach. *J. Biomed. Mater. Res.*, 59:329–339, 2002.

[20] D. M. Basalyga and R. A. Latour. Theoretical analysis of adsorption thermodynamics for charged peptide residues on SAM surfaces of varying functionality. *J. Biomed. Mater. Res. A*, 64A:120–130, 2003.

[21] G. Raffaini and F. Ganazzoli. Molecular dynamics simulation of the adsorption of a fibronectin module on a graphite surface. *Langmuir*, 20:3371–3378, 2004.

[22] J. Zheng, L. Li, H.-K. Tsao, Y. J. Sheng, S. Chen, and S. Jiang. Strong repulsive forces between protein and oligo (ethylene glycol) self-assembled monolayers: A molecular simulation study. *Biophys. J.*, 89:158–166, 2005.

[23] K. Makrodimitris, D. L. Masica, E. T. Kim, and J. J. Gray. Structure prediction of protein-solid surface interactions reveal a molecular recognition motif of statherin for hydroxyapatite. *J. Am. Chem. Soc.*, 129:13713–13722, 2007.

[24] K. Battle, E. A. Salter, R. W. Edmunds, and A. Wierzbicki. Potential of mean force calculation of the free energy of adsorption of type I winter flounder antifreeze protein on ice. *J. Cryst. Growth*, 312:1257–1261, 2010.

[25] C. P. O'Brien, S. J. Stuart, D. A. Bruce, and R. A. Latour. Modeling of peptide adsorption interactions with a poly(lactic acid) surface. *Langmuir*, 24:14115–14124, 2008.

[26] J.-W. Shen, T. Wu, Q. Wang, and H.-H. Pan. Molecular simulation of protein adsorption and desorption on hydroxyapatite surfaces. *Biomaterials*, 29:513–532, 2008.

[27] D. M. Zuckerman. Equilibrium sampling in biomolecular simulations. *Annu. Rev. Biophys.*, 40:41–62, 2011.

[28] N. A. Vellore, J. A. Yancey, G. Collier, R. A. Latour, and S. J. Stuart. Assessment of the transferability of a protein force field for the simulation of peptide-surface interactions. *Langmuir*, 26:7396–7404, 2010.

[29] A. Barducci, M. Bonomi, and M. Parrinello. Linking well-tempered metadynamics simulations with experiments. *Biophys. J.*, 98:L44–L46, 2010.

[30] D. L. Masica, J. T. Ash, M. Ndao, G. P. Drobny, and J. J. Gray. Toward a structure determination method for biomineral-associated protein using combined solid-state NMR and computational structure prediction. *Structure*, 18:1678–1687, 2010.

[31] R. H. Meißner, J. Schneider, P. Schiffels, and L. C. Ciacchi. Computational prediction of circular dichroism spectra and quantification of helicity loss upon peptide adsorption on silica. *Langmuir*, 30:3487–3494, 2014.

[32] H. D. Espinosa, J. E. Rim, F. Barthelat, and M. J. Buehler. Merger of structure and material in nacre and bone – perspectives on *de novo* biomimetic materials. *Prog. Mater. Sci.*, 54:1059–1100, 2009.

[33] S. J. Marrink, A. H. de Vries, and D. P. Tieleman. Lipids on the move: Simulations of membrane pores, domains, stalks and curves. *Biochim. Biophys. Acta*, 1788:149–168, 2009.

[34] J. E. Stone, J. C Phillips, P. L. Freddolino, D. J. Hardy, L. G. Trabuco, and K. Schulten. Accelerating molecular modeling applications with graphics processors. *J. Comput. Chem.*, 28:2618–2640, 2007.

[35] D. E. Shaw, M. M. Deneroff, R. O. Dror, J. S. Kuskin, R. H. Larson, J. K. Salmon, C. Young, B. Batson, K. J. Bowers, J. C. Chao, M. P. Eastwood, J. Gagliardo, J. P. Grossman, C. R. Ho, D. J. Ierardi, I. Kolossvári, J. L. Klepeis, T. Layman, C. McLeavey, M. A. Moraes, R. Mueller, E. C. Priest, Y. Shan, J. Spengler, M. Theobald, B. Towles, and S. C. Wang. Anton: A special-purpose machine for molecular dynamics simulation. In *Proceedings of the 34th Annual International Symposium on Computer Architecture (ISCA '07)*, New York, NY, 2007. ACM.

[36] M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids*. Clarendon Press, Oxford, 2009.

[37] H. Q. Ding, N. Karasawa, and W. A. Goddard. Atomic level simulations on a million particles – the cell multipole method for coulomb and london nonbond interactions. *J. Chem. Phys.*, 97:4309–4315, 1992.

[38] P. L. Freddolino, A. S. Arkhipov, S. B. Larson, A. McPherson, and K. Schulten. Molecular dynamics simulations of the complete satellite tobacco mosaic virus. *Structure*, 14:437–449, 2006.

[39] J. Pfaendtner, N. Volkmann, D. Hanein, P. Dalhaimer, T. D. Pollard, and G. A. Voth. Key structural features of the actin filament arp2/3 complex branch junction revealed by molecular simulation. *J. Mol. Bio.*, 416:148–161, 2012.

[40] A. Laio and M. Parrinello. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U. S. A.*, 99:12562–12566, 2002.

[41] A. Barducci, G. Bussi, and M. Parrinello. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys. Rev. Lett.*, 100:020603, 2008.

[42] U. H. E. Hannsman. Parallel tempering algorithm for conformational studies of biological molecules. *Chem. Phys. Lett.*, 281:140–150, 1997.

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

[44] G. Bussi, F. L. Gervasio, A. Laio, and M. Parrinello. Free-energy landscape for β hairpin folding from combined parallel tempering and metadynamics. *J. Am. Chem. Soc.*, 128:13435–13441, 2006.

[45] M. Bonomi, A. Barducci, and M. Parrinello. Reconstructing the equilibrium Boltzmann distribution from well-tempered metadynamics. *J. Comput. Chem.*, 30:1615–1621, 2009.

[46] A. Barducci, M. Bonomi, and M. Parrinello. Metadynamics. *WIREs Comput. Mo. Sci.*, 1:826–843, 2011.

[47] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *J. Chem. Phys.*, 21:1087–1092, 1953.

[48] D. J. Earl and M. W. Deem. Parallel tempering: Theory, applications, and new perspectives. *Phys. Chem. Chem. Phys.*, 7:3910–3916, 2005.

[49] M. Bonomi and M. Parrinello. Enhanced sampling in the well-tempered ensemble. *Phys. Rev. Lett.*, 104:190601, 2010.

[50] M. Ceriotti, G. A. R. Brain, O. Riordan, and D. E. Manolopoulos. The inefficiency of re-weighted sampling and the curse of system size in higher-order path integration. *Proc. R. Soc. London, Ser. A*, 468:2–17, 2012.

[51] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindhal. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.*, 4:435–447, 2008.

[52] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Protein*, 65:712–725, 2006.

[53] M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri, D. Donadio, F. Marinelli, F. Pietrucci, R. A. Broglia, and M. Parrinello. PLUMED: A portable plugin for free-energy calculations with molecular dynamics. *Comput. Phys. Commun.*, 180:1961–1972, 2009.

[54] J. W. Neidigh, R. M. Fesinmeyer, and N. H. Anderson. Designing a 20-residue protein. *Nat. Struct. Biol.*, 9:425–430, 2002.

[55] R. Day, D. Paschek, and A. E. Garcia. Microsecond simulations of the folding/unfolding thermodynamics of the trp-cage miniprotein. *Proteins: Struct., Funct., Bioinf.*, 78:1889–1899, 2010.

[56] H. Nymeyer. How efficient is replica exchange molecular dynamics? an analytic approach. *J. Chem. Theory Comput.*, 4:626–636, 2008.

[57] E. Rosta and G. Hummer. Error and efficiency of replica exchange molecular dynamics simulations. *J. Chem. Phys.*, 131:165102, 2009.

[58] W. F. DeGrado and J. D. Lear. Induction of peptide conformation at apolar/water interfaces. 1. a study with model peptides of defined hydrophobic periodicity. *J. Am. Chem. Soc.*, 107:7684–7689, 1985.

[59] O. Mermut, D. C. Phillips, R. L. York, K. R. McCrea, R. S. Ward, and G. A. Somorjai. In situ adsorption studies of a 14-amino acid leucine-lysine peptide onto hydrophobic polystyrene and hydrophilic silica surfaces using quartz crystal microbalance, atomic force microscopy, and sum frequency generation vibrational spectroscopy. *J. Am. Chem. Soc.*, 128:3598–3607, 2006.

[60] R. L. York, W. K. Browne, P. L. Geissler, and G. A. Somorjai. Peptides adsorbed on hydrophobic surfaces – a sum frequency generation vibrational spectroscopy and modeling study. *Isr. J. Chem.*, 47:51–58, 2007.

[61] R. L. York, O. Mermut, D. C. Phillips, K. R. McCrea, R. S. Ward, and G. A. Somorjai. Influence of ionic strength on the adsorption of a model peptide on hydrophilic silica and hydrophobic polystyrene surfaces: Insight from SFG vibrational spectroscopy. *J. Phys. Chem.*, 111:8866–8871, 2007.

[62] D. C. Phillips, R. L. York, O. Mermut, K. R. McCrea, R. S. Ward, and G. A. Somorjai. Side chain, chain length, and sequence effects on amphiphilic peptide adsorption at hydrophobic and hydrophilic surfaces studied by sum-frequency generation vibrational spectroscopy and quartz crystal microbalance. *J. Phys. Chem.*, 111:255–261, 2007.

[63] T. Weidner, N. T. Samuel, K. R. McCrea, L. J. Gamble, R. S. Ward, and D. G. Castner. Assembly and structure of α-helical peptide films on hydrophobic fluorocarbon surfaces. *Biointerphases*, 5:9–16, 2010.

[64] T. Weidner, J. S. Apte, L. J. Gamble, and D. G. Castner. Probing the orientation and conformation of α-helix and β-strand model peptides on self-assembled monolayers using sum frequency generation and nexafs spectroscopy. *Langmuir*, 26:3433–3440, 2010.

[65] J. R. Long, N. Oyler, G. P. Drobny, and P. S. Stayton. Assembly of α-helical peptide coatings on hydrophobic surfaces. *J. Am. Chem. Soc.*, 124:6297–6303, 2002.

[66] J. S. Apte, G. Collier, R. A. Latour, L. J. Gamble, and D. G. Castner. XPS and ToF-SIMS investigation of α-helical and β-strand peptide adsorption onto SAMs. *Langmuir*, 26:3423–3432, 2010.

[67] J. S. Apte, L. J. Gamble, D. G. Castner, and C. T. Campbell. Kinetics of leucine-lysine peptide adsorption and desorption at –CH3 and –COOH terminated alkylthiolate monolayers. *Biointerphases*, 5:97–104, 2010.

[68] Q. Wang, N. Shah, J. Zhao, C. Wang, C. Zhao, L. Liu, F. Zhao, and J. Zheng. Structural, morphological, and kinetic studies of β-amyloid peptide aggregation on self-assembled monolayers. *Phys. Chem. Chem. Phys.*, 13:15200–15210, 2011.

[69] J. Zhao, Q. Wang, G. Liang, and J. Zheng. Molecular dynamics simulations of low-ordered alzheimer β-amyloid oligomers from dimer to hexamer on self-assembled monolayers. *Langmuir*, 27:14876–14887, 2011.

[70] D. L. Masica and J. J. Gray. Solution- and adsorbed-state structural ensembles predicted for the statherin-hydroxyapatite system. *Biophys. J.*, 96:3082–3091, 2009.

[71] Y. He, J. Hower, S. Chen, M. T. Bernards, Y. Chang, and S. Jiang. Molecular simulation studies of protein interactions with zwitterionic phosphorylcholine self-assembled monolayers in the presence of water. *Langmuir*, 24:10358–10364, 2008.

[72] Y. Xie, M. Liu, and J. Zhou. Molecular dynamics simulations of peptide adsorption on self-assembled monolayers. *Appl. Surf. Sci.*, 258:8153–8159, 2012.

[73] L. Ou, Y. Luo, and G. Wei. Atomic-level study of adsorption, conformational change, and dimerization of an α-helical peptide at graphene surface. *J. Phys. Chem. B*, 115:9813–9822, 2011.

[74] Y. Liu, M. Wu, X. Feng, X. Shao, and W. Cai. Adsorption behavior of hydrophobin proteins on polydimethylsiloxane substrates. *J. Phys. Chem. B*, 116:12227–12234, 2012.

[75] W. Soliman, S. Bhattacharjee, and K. Kaur. Adsorption of an antimicrobial peptide on self-assembled monolayers by molecular dynamics simulation. *J. Phys. Chem. B*, 114:11292–11302, 2010.

[76] T. Wei, M. A. Carignano, and I. Szleifer. Lysozyme adsorption on polyethylene surfaces: Why are long simulations needed? *Langmuir*, 27:12074–12081, 2011.

[77] K. A. Marino and P. G. Bolhuis. Confinement-induced states in the folding landscape of the trp-cage miniprotein. *J. Phys. Chem. B*, 116:11872–11880, 2012.

[78] M. R. Nimlos, G. T. Beckham, J. F. Matthews, L. Bu, M. E. Himmel, and M. F. Crowley. Binding preferences, surface attachment, diffusivity, and orientation of a family 1 carbohydrate-binding module on cellulose. *J. Biol. Chem.*, 287:20603–20612, 2012.

[79] V. P. Raut, M. A. Agashe, S. J. Stuart, and R. A. Latour. Molecular dynamics simulations of peptide-surface interactions. *Langmuir*, 21:1629–1639, 2005.

[80] S. De Miranda and T. R. Walsh. Atomistic modelling of the interaction between peptides and carbon nanotubes. *Mol. Phys.*, 105:221–229, 2007.

[81] G. Collier, N. A. Vellore, J. A. Yancey, S. J. Stuart, and R. A. Latour. Comparison between empiriacal force fields for the simulation of the adsorption behavior of structured LK peptides on functionalized surfaces. *Biointerphases*, 7:1–19, 2012.

[82] S.-W. Hung, P.-Y. Hsiao, and C.-C. Chieng. Dynamic information for cardiotoxin protein desorption from a methyl-terminated self-assembled monolayer using steered molecular dynamics simulation. *J. Chem. Phys.*, 134:194705, 2011.

[83] M. Mijajlovic, M. J. Penna, and M. J. Biggs. Free energy of adsorption for a peptide at a liquid/solid interface via nonequilibrium molecular dynamics. *Langmuir*, 29:2919–2926, 2013.

[84] C. D. Christ, A. E. Mark, and W. F. van Gunsteren. Basic ingredients of free energy calculations: A review. *J. Comput. Chem.*, 31:1569–1582, 2010.

[85] A. Laio and F. L. Gervasio. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry, and material science. *Rep. Prog. Phys.*, 71:126601, 2008.

[86] V. Leone, F. Marinelli, P. Carloni, and M. Parrinello. Targeting biomolecular flexibility with metadynamics. *Curr. Opin. Struc. Biol.*, 20:148–154, 2010.

[87] M. Deighan, M. Bonomi, and J. Pfaendtner. Efficient simulation of explicitly solvated proteins in the well-tempered ensemble. *J. Chem. Theory Comput.*, 8:2189–2192, 2012.

[88] A. D. Mackerell, M. Geig, and C. L. Brooks. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.*, 25:1400–1415, 2004.

[89] G. A. Kaminski, R. A. Friesner, J. RTirado-Rives, and W. L. Jorgensen. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B*, 105:6474–6487, 2001.

[90] W. Humphrey, A. Dalke, and K. Schulten. VMD: Visual molecular dynamics. *J. Mol. Graphics*, 14:33–38, 1996.

[91] A. Ulman, J. E. Eiler, and N. Tillman. Packing and molecular orientation of alkanethiol monolayers on gold surfaces. *Langmuir*, 5:1147–1152, 1989.

[92] K. P. Fears, S. E. Creager, and R. A. Latour. Determination of the surface pK of carboxylic- and amine-terminated alkanethiols using surface plasmon resonance spectroscopy. *Langmuir*, 24:837–843, 2008.

[93] G. Bussi, D. Donadio, and M. Parrinello. Canonical sampling through velocity rescaling. *J. Chem. Phys.*, 126:014101, 2007.

[94] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNota, and J. R. Haak. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81:3684–3690, 1984.

[95] M. K. Prakash, A. Barducci, and M. Parrinello. Replica temperatures for uniform exchange and efficient roundtrip times in explicit solvent parallel tempering simulations. *J. Chem. Theory Comput.*, 7:2025–2027, 2011.

[96] D. J. Sindhikara, D. J. Emerson, and A. E. Roitberg. Exchange often and properly in replica exchange molecular dynamics. *J. Chem. Theory Comput.*, 6:2804–2808, 2010.

[97] Grossfield, alan, "WHAM: the weighted histogram analysis method", version 2.0.6, http://membrane.urmc.rochester.edu/content/wham.

[98] W. Kabsch. A solution for the best rotation to relate two sets of vectors. *Acta Cryst.*, A32:922–923, 1976.

[99] X. Daura, K. Gademann, B. Jaun, D. Seeback, W. F. van Gunsteren, and A. E. Mark. Peptide folding: When simulation meets experiment. *Agnew. Chem. Int. Ed.*, 38:236–240, 1999.

[100] J. Pfaendtner, D. Branduardi, M. Parrinello, T. D. Pollard, and G. A. Voth. Nucleotide-dependent conformational states of actin. *Proc. Natl. Acad. Sci. U. S. A.*, 106:12723–12728, 2009.

[101] T. Weidner, N. F. Breen, K. Li, G. P. Drobny, and D. G. Castner. Sum frequency generation and solid-state NMR study of the structure, orientation, and dynamics of polystyrene-adsorbed peptides. *Proc. Natl. Acad. Sci. U. S. A.*, 107:13288–13293, 2010.

[102] S. Mann. Molecular recognition in biomineralization. *Nature*, 332:119–124, 1988.

[103] L. Addadi and S. Weiner. Control and design principles in biological mineralization. *Angew. Chem. Int. Ed. Engl.*, 31:153–169, 1992.

[104] R. J. Lund, M. R. Davies, and K. A. Hruska. Bone morphogenic protein-7: an anti-fibrotic morphogenic protein with therapeutic importance to renal disease. *Curr. Opin. Nephrol. Hy.*, 11:31–36, 2002.

[105] P. Collin-Osdoby. Regulation of vascular calcification by osteoclst regulatory factors RANKL and osteoprotegerin. *Circ. Res.*, 95:1046–1057, 2004.

[106] F. Nudelman and A. J. M. Sommerdijk. Biomineralization as an inspiration for materials chemistry. *Angew. Chem. Int. Ed.*, 51:6582–6596, 2012.

[107] S. Weiner and L. Addadi. Crystallization pathways in biomineralization. *Annu. Rev. Mater. Res.*, 41:21–40, 2011.

[108] N. Ramasubbu, L. M. Thomas, K. K. Bhandary, and M. J. Levine. Structural characteristics of human salivary statherin: A model for boundary lubrication at the enamel surface. *Crit. Rev. Oral Biol. M.*, 4:363–370, 1993.

[109] P. S. Stayton, G. P. Drobny, W. J. Shaw, J. R. Long, and M. Gilbert. Molecular recognition at the protein-hydroxyapatite interface. *Crit. Rev. Oral Biol. M.*, 14:370–376, 2003.

[110] G. Goobes, R. Goobes, W. J. Shaw, J. M. Gibson, J. R. Long, V. Raghunathan, O. Schueler-Furman, J. M. Popham, D. Baker, C. T. Campbell, P. S. Stayton, and G. P. Drobny. The structure, dynamics, and energetics of protein adsorption – lessons learned from adsorption of statherin to hydroxyapatite. *Magn. Reson. Chem.*, 45:S32–S47, 2007.

[111] P. A. Raj, M. Johnsson, M. J. Levine, and G. H. Nancollas. Salivary statherin. *J. Biol. Chem.*, 267:5968–5976, 1992.

[112] J. R. Long, J. L. Dindot, H. Zebroski, S. Kiihne, R. H. Clark, A. A. Campbell, P. S. Stayton, and G. P. Drobny. A peptide that inhibits hydroxyapatite growth is in an extended conformation on the crystal surface. *Proc. Natl. Acad. Sci. U. S. A.*, 95:12083–12087, 1998.

[113] J. R. Long, W. J. Shaw, P. S. Stayton, and G. P. Drobny. Structure and dynamics of hydrated statherin on hydroxyapatite as determined by solid-state NMR. *Biochemistry-US*, 40:15451–15455, 2001.

[114] J. M. Gibson, V. Raghunathan, J. M. Popham, P. S. Stayton, and G. P. Drobny. A REDOR NMR study of a phosphorylated statherin fragment bound to hydroxyapatite crystals. *J. Am. Chem. Soc.*, 127:9350–9351, 2005.

[115] M. Ndao, J. T. Ash, P. S. Stayton, and G. P. Drobny. The role of basic amino acids in the molecular recognition of hydroxyapatite by statherin using solid state NMR. *Surf. Sci.*, 604:L39–L42, 2010.

[116] T. Weidner, M. Dubey, N. F. Breen, J. Ash, J. E. Baio, C. Jaye, D. A. Fischer, G. P. Drobny, and D. G. Castner. Direct observation of phenylalanine orientations in statherin bound to hydroxyapatite surfaces. *J. Am. Chem. Soc.*, 134:8750–8753, 2012.

[117] W. J. Shaw, J. R. Long, J. L. Dindot, A. A. Campbell, P. S. Stayton, and G. P. Drobny. Determination of statherin N-terminal peptide conformation on hydroxyapatite crystals. *J. Am. Chem. Soc.*, 122:1709–1716, 2000.

[118] W. J. Shaw, J. R. Long, A. A. Campbell, P. S. Stayton, and G. P. Drobny. A solid state NMR study of dynamics in a hydrated salivary peptide adsorbed to hydroxyapatite. *J. Am. Chem. Soc.*, 122:7118–7119, 2000.

[119] J. M. Gibson, J. M. Popham, V. Raghunathan, P. S. Stayton, and G. P. Drobny. A solid-state NMR study of the dynamics and interactions of phenylalanine rings in a statherin fragment bound to hydroxyapatite crystals. *J. Am. Chem. Soc.*, 128:5364–5370, 2006.

[120] M. Ndao, J. T. Ash, N. F. Breen, G. Goobes, P. S. Stayton, and G. P. Drobny. A $^{13}C\{^{31}P\}$ REDOR NMR investigation of the role of glutamic acid residues in statherin-hydroxyapatite recognition. *Langmuir*, 25:12136–12143, 2009.

[121] M. Skepö. Model simulations of the adsorption of statherin to solid surfaces: Effects of surface charge and hydrophobicity. *J. Chem. Phys.*, 129:185101, 2008.

[122] A. Rimola, M. Aschi, R. Orlando, and P. Ugliengo. Does adsorption at hydroxyapatite surfaces induce peptide folding? insights from large-scale B3LYP calculations. *J. Am. Chem. Soc.*, 134:10899–10910, 2012.

[123] A. Rimola, M. Corno, J. Garza, and P. Ugliengo. *Ab initio* modelling of protein-biomaterial interactions: influence of amino acid polar side chains on adsorption at hydroxyapatite surfaces. *Phil. Trans. R. Soc. A*, 370:1478–1498, 2012.

[124] O. Schueler-Furman, C. Wang, P. Bradley, K. Misura, and D. Baker. Progress in modeling of protein structures and interactions. *Science*, 310:638–642, 2005.

[125] M. Deighan and J. Pfaendtner. Exhaustively sampling peptide adsorption with metadynamics. *Langmuir*, 29:7999–8009, 2013.

[126] L. Sutto and F. L. Gervasio. Effects of oncogenic mutations on the conformational free-energy landscape of EGFR kinase. *Proc. Natl. Acad. Sci. U. S. A.*, 2013.

[127] M. Zachariou, I. Traverso, L. Spiccia, and M. T. W. Hearn. Potentiometric investigations into the acid-base and metal ion binding properties of immobilized metal ion affinity chromatographic (imac) adsorbents. *J. Phys. Chem.*, 100:12680–12690, 1996.

[128] N. Homeyer, A. H. C. Horn, H. Lanig, and H. Sticht. AMBER force-field parameters for phosphorylated amino acids in different protonation states: phosphoserine, phosphothreonine, phosphotyrosine, and phosphohistidine. *J. Mol. Model*, 12:281–289, 2006.

[129] T. Steinbrecher, J. Latzer, and D. A. Case. Revised AMBER parameters for bioorganic phosphates. *J. Chem. Theory Comput.*, 8:4405–4412, 2012.

[130] R. M. Wilson, J. C. Elliot, and S. E. P. Dowker. Rietveld refinement of the crystallographic structure of human dental enamel apatites. *Am. Mineral*, 84:1406–1414, 1999.

[131] T. Ikoma, A. Yamazaki, S. Nakamura, and M. Akao. Preparation and structure refinement of monoclinic hydroxyapatite. *J. Solid State Chem.*, 144:272–276, 1999.

[132] S. Hauptmann, H. Dufner, J. Brickmann, S. M. Kast, and R. S. Berry. Potential energy function for apatites. *Phys. Chem. Chem. Phys.*, 5:635–639, 2003.

[133] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Imprey, and M. L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79:926–935, 1983.

[134] K. Lindorf-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, and D. E. Shaw. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Struct., Funct., Bioinf.*, 78:1950–1958, 2010.

[135] G. R. Bowman, X. Huang, and V. S. Pande. Using generalized ensemble simulations and Markov state models to identify conformal states. *Methods*, 49:197–201, 2009.

[136] B. B. Langdon, M. Kastantin, R. Walder, and D. K. Schwartz. Interfacial protein-protein associations. *Biomacromolecules*, 15:66–74, 2013.

[137] A. Roehrich and G. P. Drobny. Solid-state NMR studies of biomineralization peptides and proteins. *Accounts Chem. Res.*, 46(9):2136–2144, 2013.

[138] R. Goobes, G. Goobes, W. J. Shaw, G. P. Drobny, C. T. Campbell, and P. S. Stayton. Thermodynamic roles of basic amino acids in statherin recognition of hydroxyapatite. *Biochemistry-US*, 46:4725–4733, 2007.

[139] O. Santos, J. Kosoric, M. P. Hector, P. Anderson, and L. Lindh. Adsorption behavior of statherin and a statherin peptide onto hydroxyapatite and silica surfaces by *in situ* ellipsometry. *J. Colloid Interf. Sci.*, 318:175–182, 2008.

[140] G. Goobes, R. Goobes, O. Schueler-Furman, D. Baker, P. S. Stayton, and G. P. Drobny. Folding of the C-terminal bacterial binding domain in statherin upon adsorption onto hydroxyapatite crystals. *Proc. Natl. Acad. Sci. U. S. A.*, 103:16083–16088, 2006.

[141] A. Amano, K. Kataoka, P. A. Raj, R. J. Genco, and S. Shizukuishi. Binding sites of salivary statherin for *Porphyromonas gingivalis* recombinant fimbrillin. *Infect Immun.*, 64(10):4249–4254, 1996.

[142] S. Sekine, K. Kataoka, M. Tanaka, H. Nagata, T. Kawakami, K. Akaji, S. Aimoto, and S. Shizukuishi. Active domains of salivary statherin on apatitic surfaces for binding to *Fusobacterium nucleatum* cells. *Microbiology*, 150:2373–2379, 2004.

[143] R. Goobes, G. Goobes, C. T. Campbell, and P. S. Stayton. Thermodynamics of statherin adsorption on hydroxyapatite. *Biochemistry*, 45:5576–5586, 2006.

[144] A. Barducci, M. Bonomi, M. K. Prakash, and M. Parrinello. Free-energy landscape of protein oligomerization from atomistic simulations. *Proc. Natl. Acad. Sci. U. S. A.*, pages E4708–E4713, 2013.

[145] C. Camilloni, A. Cavalli, and M. Vendruscolo. Replica-averaged metadynamics. *J. Chem. Theory Comput.*, 9:5610–5617, 2013.

[146] F. Palazzesi, A. Barducci, M. Tollinger, and M. Parrinello. The allosteric communication pathways in KIX domain of CBP. *Proc. Natl. Acad. Sci. U. S. A.*, 110:14237–14242, 2013.

[147] O. Valsson and M. Parrinello. Thermodynamical description of a quasi-first-order phase transition from the well-tempered ensemble. *J. Chem. Theory Comput.*, 9:5267–5276, 2013.

[148] S. Roy, T. L. Naka, and D. K. Hore. Enhanced understanding of amphipathic peptide adsorbed structure by modeling of the nonlinear vibrational response. *J. Phys. Chem. C*, 117:24955–24966, 2013.

[149] R. M. Elder and A. Jayaraman. Simulation study of the effects of surface chemistry and temperature on the conformations of ssdna oligomers near hydrophili and hydrophobic surfaces. *J. Chem. Phys.*, 140:155103, 2014.

[150] E. D. Zhong and M. R. Shirts. Thermodynamics of coupled protein adsorption and stability using hybrid Monte Carlo simulations. *Langmuir*, 30:4952–4961, 2014.

[151] S. A. Hall, K. C. Jena, P. A. Covert, S. Roy, T. G. Trudeau, and D. K. Hore. Molecular-level surface structure from nonlinear vibrational spectroscopy combined with simulations. *J. Phys. Chem. B*, 2014.

[152] P. R. Burney, N. White, and J. Pfaendtner. Structural effects of methionine oxidation on isolated subdomains of human fibrin D and αC regions. *PLOS ONE*, 9:e86981, 2014.

[153] Z. Jarin and J. Pfaendtner. Ionic liquids can selectively change the conformational free-energy landscape of sugar rings. *J. Chem. Theory Comput.*, 10:507–510, 2014.

[154] J. A. Snyder, T. Abramyan, J. A. Yancey, A. A. Thyparambil, Y. Wei, S. J. Stuart, and R. A. Latour. Development of a tuned interfacial force field parameter set for the simulation of protein adsorption to silica glass. *Biointerphases*, 7(56), 2012.

[155] P. K. Biswas, N. A. Vellore, J. A. Yancey, T. G. Kucukkal, G. Collier, B. R. Brooks, S. J. Stuart, and R. A. Latour. Simulation of multiphase systems utilizing independent force fields to control intraphase and interphase behavior. *J. Comput. Chem.*, 33:1458–1466, 2012.

[156] P. Tiwary and M. Parrinello. From metadynamics to dynamics. *Phys. Rev. Lett.*, 111:230602, 2013.

[157] M. Salvalaglio, P. Tiwary, and M. Parrinello. Assessing the reliability of the dynamics reconstructed from metadynamics. *J. Chem. Theory Comput.*, 10:1420–1425, 2014.

[158] I. Yarovsky, N. Todorovna, F. Marinelli, and S. Piana. Exploring the folding free energy landscape of insulin using bias exchange metadynamics. *J. Phys. Chem. B*, 113:3556–3564, 2009.

[159] A. Barducci, R. Chelli, P. Procacci, V. Schettino, F. L. Gervasio, and M. Parrinello. Metadynamics simulation of prion protein β-structure stability and the early stages of misfolding. *J. Am. Chem. Soc.*, 128:2705–2710, 2006.

[160] A. Laio, A. Rodriguez-Fortea, F. L. Gervasio, M. Ceccarelli, and M. Parrinello. Assessing the accuracy of metadynamics. *J. Phys. Chem. B*, 109:6714–6721, 2005.

[161] S. Piana, K. Lindoff-Larsen, and D. E. Shaw. How robust are protein folding simulations with respect to force field parameterization. *Biophys. J.*, 100:L47–L49, 2011.

[162] L. Qiu, S. A. Pabit, A. E. Roitberg, and S. J. Hagen. Smaller and faster: The 20-residue trp-cage protein folds in 4 μs. *J. Am. Chem. Soc.*, 124:12952–12953, 2002.