Understanding Self-Assembly in Solution and at Interfaces Using All-Atom Molecular Dynamics Simulations and Enhanced Sampling Methods

Arushi Prakash

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2018

Reading Committee:

Walter J. Pfaendtner, Chair
François Baneyx
Christopher J. Mundy

Program Authorized to Offer Degree:

Chemical Engineering

University of Washington

Abstract

Understanding Self-Assembly in Solution and at Interfaces Using All-Atom Molecular Dynamics Simulations and Enhanced Sampling Methods

Arushi Prakash

Chair of the Supervisory Committee:
Dr. Walter J. Pfaendtner
Chemical Engineering

Proteins and biopolymers self-assemble to form nanostructures in solution or at interfaces. Notable examples include, the formation of plaque during Alzheimer's disease, and the formation of protein-templates for biomineralization. Even though these assembly processes are well-studied, they remain poorly understood. The focus of this dissertation is to investigate early-stage assembly processes, including adsorption of particles on surfaces, and oligomer formation using classical molecular dynamics and enhanced sampling methods (like metadynamics and umbrella sampling). As these systems are explored, improved protocols for enhanced sampling of protein-adsorption simulations are developed to tackle the roles of ions in simulations, and to incorporate experimental data into the sampling using Bayesian inference. Further, a new sampling method is developed to make these enhanced sampling simulations more efficient.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Foremost, I would like to sincerely thank both my advisors Dr. Jim Pfaendtner and Dr. Christopher Mundy for encouraging me to pursue a doctoral degree, being patient with my progress, helping me in both personal and professional development, and always being available for questions. Besides my advisors, I would like to thank the rest of my thesis committee: Dr. Jim deYoreo and Dr. Francois Baneyx for their guidance in research and professional development during my Ph.D.

I thank Dr. Kayla Sprenger and Chris Fu, with whom I have collaborated closely with on exciting projects. Further, I thank research group members Josh Smith, Wesley Beckner, Khushmeen Sakhlot, Sarah Alamdari, Janani Sampath, Coco Mao, Luke Gibson, Chris Nyambura, Yeneneh Yimer, Vance Jaeger, Blake Hough, and Luiz Oliviera, for their constant support, insightful discussions and suggestions, and unending fun. I also thank Dr. Marcel Baer, Dr. Gregory Schenter, Dr. Jaehun Chun, and Dr. Evgenii Fetisov at the Pacific Northwest Laboratory (PNNL) for their advice and help during my graduate research.

I am also grateful to the graduate program advisor Allison Sherrill and DIRECT program manager Kelly Thornton for their help throughout my degree. Further, I thank the Department of Chemical Engineering for providing resources for professional development, a community of fellow graduate students, and access to different faculty mentors.

Further, I thank PNNL for the opportunity to work at their labs, access their computing clusters, and financial support through MS3 and NW IMPACT programs. I am also grateful to the UW IT department for their computing resources. Additionally, I want to express my gratitude to the NRT DIRECT program for my training in data science and software engineering methods.

Last but not the least, I would like to thank my parents, friends, and family without whose support my doctoral degree would not be a success.

DEDICATION

This thesis is dedicated to my grandfathers, Dr. Ramendra Sinha and Surendra Sinha, and my parents, Ekta Sinha and Om Prakash.

Chapter 1. INTRODUCTION

1.1 PROTEIN SELF-ASSEMBLY AT INTERFACES

Solid interfaces, for example the basal surface of mica, can promote the assembly of proteins.[1] It has been suggested that these interfaces drive the assembly process in several ways, either by affecting the morphology[2] by acting as templates or affecting the kinetics of aggregation.[3]

Several features of the surface and the environment can affect the formation of these nanostructures. For example, Zhang et al.[4] observed that the hydrophilicity of the surface determined whether the peptides formed upright or horizontal aggregates. Further, Whitehouse et al. observed that changing the solvent, which altered the binding energy of the protein to the surface, changed the nature of aggregates.[5] Unsurprisingly, pH[6] and electrolytes[3] also affects the assembly process.

Moreover, several features of the protein can alter ability to form nanostructures on the surface. It has been hypothesized that only proteins that are attracted to the surface and are able to diffuse along the surface after adsorption are able to form nanostructures.[1] Once adsorbed to the solid substrate, the protein may undergo conformational changes which may affect its ability to form these structures.[7] Additionally, the changing concentration of proteins can also alter the type of nanostructures formed.[2]

Understandably, these assembly processes are difficult to understand. Notably, a number of processes are happening at different length and time scales, each with different sensitivities to environmental variables. Therefore, it is difficult to obtain an overall understanding of an assembly process from one instrument alone. Oftentimes, all-atom molecular dynamics simulations are used to obtain a high-resolution, bottom-up picture of this process.

1.2 MOLECULAR DYNAMICS SIMULATIONS OF ASSEMBLING SYSTEMS

Classical molecular dynamics (MD) simulations is an important tool for understanding the structure and function of biomolecules. Since MD simulations provide molecular-level detail into the motions of particles, they can be used to address specific questions such as the role of side-chains in adsorption to a surface.[8] Moreover, with the rampant increase in computing power, it has become possible to run long MD simulations (ns to μs long) providing better data.[9]

In MD simulations, all particles are explicitly modeled with empirical potentials which are parameterized using experimental data or quantum/semi-empirical calculations. Commonly used potentials include AMBER[10,11] and CHARMM.[12] During the simulation, the atoms are propagated according to Newton's equations of motion using a timestep in femtoseconds.

MD simulations have been used to explore the assembly of peptides. For example, a simulation can be seeded with a pre-assembled structure to observe how the structure relaxes during the course of the simulation.[13] Many studies focus on the early stages of assembly by allowing biomolecules to naturally adsorb to surfaces, or self-assemble starting from a random state simulation.[14–17] These simulations can provide kinetic, thermodynamic, and structural information about the assembly process.

Importantly, the accuracy of the statistics calculated from MD simulations hinges on the requirement that the simulation follows the ergodic hypothesis. In other words, the simulation should sample all phase space in the finite time that we run the simulation.[18] However, these simulations are often stuck in kinetic traps or metastable states[19], thereby violating the ergodic hypothesis. Consequently, MD simulations are often combined with methods that can enhance sampling of phase space.

1.3 ENHANCED SAMPLING METHODS

There are several excellent reviews on enhanced sampling methods that are used for biomolecular simulations.[20,21] However, this section will focus on three types of methods that feature in this thesis – (1) collective variable (CV) biasing like umbrella sampling (US),[22] and metadynamics (MetaD; well-tempered (WTM),[23] parallel-bias (PBMetaD),[24] and parallel-bias with partitioned families (PBMetaDPF) [25], metadynamic metainference[26]), (2) tempering like parallel tempering (PT),[27] and (3) the combination of CV biasing and tempering like parallel-tempering metadynamics in the well-tempered ensemble (PTMetaD-WTE).[28] Here, a CV is defined as any function of the 3N-dimensional atomic configuration of the system. Additionally, CVs should be differentiable so that they can be biased.

Umbrella sampling, developed by Torrie and Valleau[29], is one of the oldest CV-based approaches. In this method, the CV is restrained at a target value using a bias potential. Several independent simulations (or windows) are run with different target values, such that the values span the CV space of interest. Finally, these windows are combined with the weighted histogram method (WHAM) to generate the free energy profile along the CV.[22]

On the other hand, metadynamics[30,31] seeks to sample all values of the CV in one simulation by building a history-dependent bias potential that retains information about previously visited phase space. The bias potential is created by the successive addition of small, repulsive Gaussian-shaped kernels centered at the CV value at each biasing step. This potential finally converges to the underlying true free energy surface. In fact, it was mathematically proven that the well-tempered metadynamics (WTM) variant, which scales the kernel height according to the explored phase space, converges asymptotically to the underlying free energy surface.[32] Notably, the Gaussian kernels in WTM are multivariate and their dimensionality is equal to the number of CVs being biased.

In contrast, parallel-bias metadynamics(PBMetaD)[33] constructs numerous single-dimensional bias potentials, where the number of potentials is equal to the number of CVs in the system. This method allows users to bias multiple variables, even 40-60 CVs,[34] without exponentially increasing the memory overhead. Further, parallel-bias metadynamics with partitioned families (PBMetaDPF) improves on PBMetaD by allowing degenerate variables to contribute to the same bias potential, along the lines of multiple walker simulations.[35]

On the other hand, parallel tempering (PT)[27] relies on increasing the temperature of the whole system, rather than a single CV, to enhance sampling. In PT, replicas of the system are simulated at higher temperatures, and configurations of the systems at nearby temperatures are exchanged using the Metropolis acceptance criteria. This allows low temperature replicas to sample configurations that are accessible at higher temperatures, instead of being trapped in local minima at lower temperatures.

1.4 ADDRESSING THE GAPS IN CURRENT KNOWLEDGE

This thesis includes studies of the adsorption of proteins at interfaces (precursor to assembly at interfaces) to highlight the most important factors influencing this process. Apart from drawing conclusions about the physical systems, this thesis will address several key issues regarding limitations of simulation protocols, and the limitations of sampling methods.

In Chapter 2 and 3, a model interfacial system, namely the basal surface of mica interacting with water and ions, is simulated to highlight key factors affecting the adsorption and assembly of any particle on a surface. In Chapter 3, an elongated peptide and peptoid are simulated at model hydrophilic and hydrophobic interfaces to understand how the conformation of molecules can be affected by the presence of an interface. On the other hand, Chapter 4 includes a real silica surface interacting with a model molecule to draw out the properties of the surface that affect the interaction of the molecule with an interface. Finally, experimentally relevant peptides and

interfaces are simulated in Chapters 5 and 7, to completely characterize how a peptide interacts with the interface, as a function of peptide structure, pH, and surface structure.

Further, Chapter 4 addresses important issues in sampling the adsorption of proteins to interfaces with regards to the choice CVs. Finally, Chapter 6 introduces improvements in sampling methods to address issues in biasing these complex systems.

Chapter 2. QUANTIFYING THE MOLECULAR-SCALE AQUEOUS RESPONSE TO THE MICA SURFACE

2.1 ABSTRACT

Modeling assembly at surfaces requires an understanding of the interactions between solutes, solvents, and surfaces at multiple scales. We investigated the solvent response (water structure and orientation) to a dielectric surface (mica) using density functional theory. A different water structure is engendered by replacing naturally-occurring surface ions (K) with H_3O^+. We also validate classical models for the mica surface (CLAYFF) against DFT predictions. The detailed microscopic response of water to mica can be used as input into continuum models for the total interactions between two mica surfaces supporting a strong correlation between physicochemical phenomena at different scales.

2.2 INTRODUCTION

Molecular features in the vicinity of charged dielectric surfaces drive phenomena central to electrochemistry, geology, and material synthesis. Thus, models of the interface require a departure from standard theoretical treatments that rely on a mean-field solution and ignore important molecular granularity.[36] Indeed, through surface force apparatus (SFA) experiments, it

is evident that microscopic details like ion speciation on the mica surface alter the interaction between surfaces contradicting the widely accepted, continuum Derjaguin–Landau–Verwey–Overbeek (DLVO) theory of colloidal stability.[37,38] Mica (or muscovite) is commonly used for SFA and other experimental studies of the solid–aqueous electrolyte interface[37,38] due to its atomistically smooth cleavage along the (001) plane. While mica is suitable for a variety of accurate surface measurements, the complexities associated with the placement and dynamics of ions and water near the interface are largely still a matter of scientific debate. It has been suggested that ions are non-uniformly distributed on the mica surface with varying substitutions of intrinsic ions with electrolyte ions.[39–41] X-ray reflectivity (XRR) measurements indeed show that water near mica is layered into three distinct regions—the adsorbate layer, the first hydration layer, and an oscillatory layer.[42] Different studies have described these water layers as either ice-like[43] or liquid-like[44] or both. [42] We have yet to achieve a molecular picture of the aqueous response to muscovite that is consistent with direct experimental observations and ab initio potential based quantum density functional theory (DFT) simulations.

Although classical or empirical interaction potentials may be good candidates for this purpose, their capability to accurately capture both short- and long-ranged responses of an aqueous solution to a broken symmetry, such as an interface, has recently been questioned.[45] For model interfaces such as the air–water interface, dominated by a broken hydrogen-bond network, classical models compare well to DFT models.[46] The correct solution response to a dielectric interface can serve as a useful mesoscale principle connecting molecular details to macroscopic phenomena. To this end, it was recently shown that Lifshitz theory can be modified to incorporate the molecular granularity at interfaces to account for macroscopic dispersion forces between two dielectric surfaces in aqueous solution.[47] This revised Lifshitz theory, for long-range (>3 nm)

interactions between two mica surfaces, displayed significantly stronger forces than standard Lifshitz theory and yielded a better agreement between the theory and SFA measurements.[47]

Herein, we simulate the aqueous response to a mica surface (see Figure 2-1A), that is either terminated with K^+ or H_3O^+ ions, using DFT (Figure 2-1). The aforementioned terminations represent two limiting cases of the surface − natural muscovite (or the muscovite surface equilibrated with K^+ electrolyte) and an acid-washed surface, respectively, in contact with pure water or dilute electrolyte. Using both ab initio and empirical classical interaction potentials, we calculate the X-ray reflectivity (XRR) from our trajectories and directly compare to experiment. We provide the molecular details of water structure and orientation in the vicinity of the interface that can be used to distinguish between representations of interaction in addition to a more quantitative understanding of surface hydration forces.

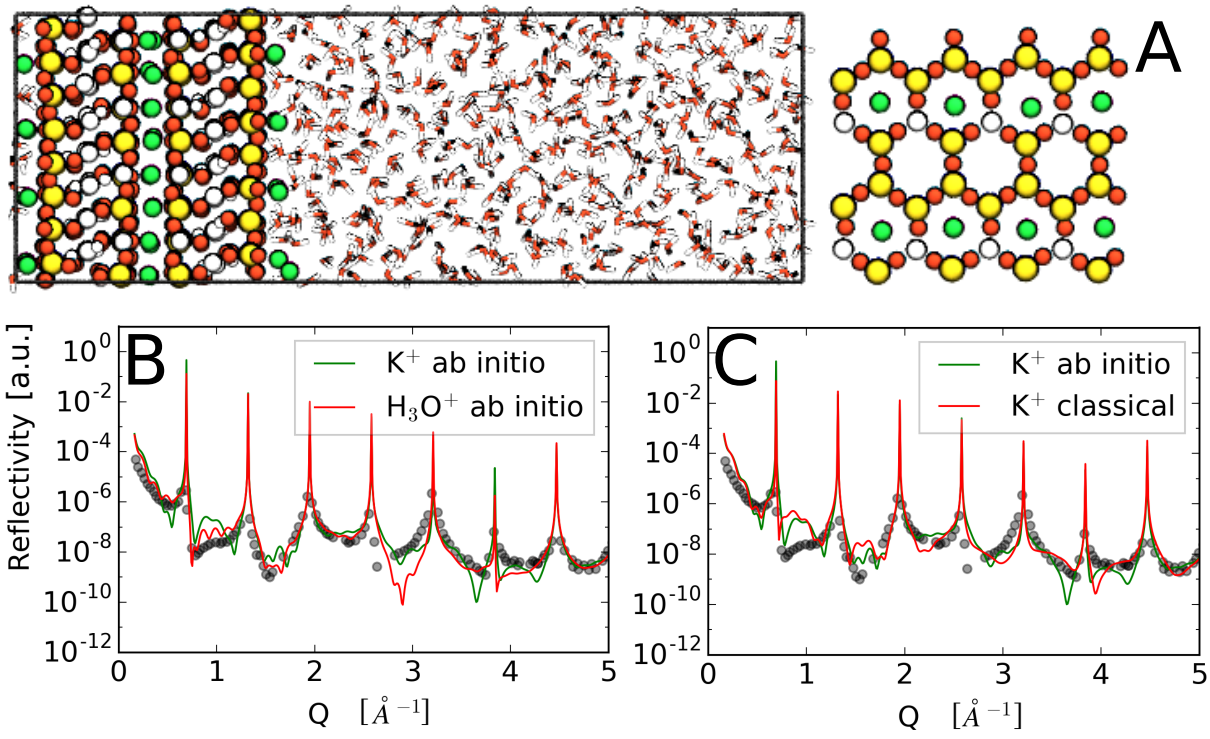


Figure 2-1 (A) (left) Simulation setup for the K+ covered-muscovite and water system for DFT and classical empirical interaction potentials with oxygen (red), silicon (yellow), aluminum (white), hydrogen (small, white), and potassium (green) (right) (001) surface of muscovite, in

contact with water, which consists of hexagonal cavities that are either empty or occupied by K^+ ions. Comparison of XRR curves: (B) Comparison of K+ covered-muscovite and H_3O^+covered-muscovite with experimental data taken from ref 7 (gray circles). (C) Comparison of K^+ covered-muscovite modeled using DFT and CLAYFF(13) classical potential with experimental data (black circles).

2.3 METHODS

CLASSICAL MOLECULAR DYNAMICS SIMULATIONS

An ideal mica (001) surface was created with K+ ions on surface hexagonal cavities and Al substitutions in accordance with Lowenstein's rule (to avoid Al–O–Al linkages). We used the CLAYFF potential[48] to model K^+-covered mica (672 atoms) and the SPC rigid water model (512 water molecules) to model water. The simulation box was 60 Å (20 Å mica) in depth and 20.756 Å × 17.990 Å in surface area (see Figure 1). Mica atoms were not constrained during the simulation. We used GROMACS 4.6.5[49] for all simulations. We started with a random setup of water molecules and minimized the energy of the system using the steepest descent method. Then, we simulated the system in the NPT ensemble using a Berendsen barostat[50] (tau = 2.0 ps, pressure = 1 bar) and a Donadio–Bussi–Parrinello thermostat[51] (tau = 0.1 ps, temperature = 300 K) for 2 ns. Finally, we simulated the system in the NVT ensemble, at 300 K, for 100 ns. LINCS[52] constraints were added to all bonds of hydrogen with other atoms to allow simulations with a time step of 2 fs. For all simulations, electrostatics were calculated using the particle mesh Ewald (PME)[53] summation method using a short-range cutoff of 0.8 nm. A van der Waals cutoff of 0.8 nm was used.

AB INITIO SIMULATIONS

Simulations of the system described in the classical simulation were also conducted with DFT. The starting structure for K^+-covered mica was obtained from structures equilibrated in the classical model. K^+ ions were substituted for H3O+ ions to generate the starting structure of H_3O^+ covered mica. We ran the simulations in the NVT ensemble using an electron density cutoff of 400 Ry using basis sets of double-ζ with polarization optimized to the condensed phase.[54] We performed 20 ps of equilibration and 25 ps of production run for each terminated mica surface. A time step of 0.5 fs was used. We employ generalized gradient corrected DFT at the level of Becke exchange[55] and Lee–Yang–Parr correlation[5656] (BLYP) in conjunction with Goedecker–Teter–Hutter (GTH) pseudopotentials[57] augmented with the empirical D2 dispersion term of Grimme.[58] The aforementioned level of theory is known to adequately describe the aqueous structure, mass density, and response to both the air–water and molecular interfaces.[45,59] Indeed, the importance of including dispersion corrections for describing such interfaces was stressed in a recent study by Feibelman.[60]

X-RAY REFLECTIVITY

XRR curves track the fraction of incident beam reflected or specular reflectivity signal (R(Q) or reflectivity) against the momentum transfer $Q = \frac{4\pi}{\lambda}\sin(\theta)$, where λ is the wavelength of incident beam and θ is the angle of incidence on the surface. From a recently developed formalism,[61] we can generate these curves from the laterally averaged nuclear density for different atom types ($\rho(z_i)$). The nuclear density profile is binned with a bin-size of 0.1 Å which was shown by Fenter et al. to remove artifacts from the XRR spectra.[62] The nuclear density of the system was calculated by averaging all frames of the simulation trajectory. Following ref 26, the spectra are calculated using the following equation

$$R(Q) = \frac{4\pi r_e}{A_{UC}Q^2}|F(Q)|^2 \qquad 2.1$$

where A_{UC} is the surface area of the muscovite unit cell and F(Q), the structure factor, is given by

$$F(Q) = exp\left(-\frac{1}{2}Q^2\sigma^2\right)\sum f_i(Q)\rho(z_i)\exp(iQz_i) \quad 2.2$$

for all i atoms in the simulated system. Here r_i is the position of atoms from the simulation, fi is the atomic scattering factor at a particular Q, $\rho(z_i)$). is the nuclear density for a particular atom type from the simulation, and σ is the Debye–Waller correction factor (with mean displacement of atoms assumed as 0.1 Å[62]). Different components of F(Q) are summed up as follows:

$$F(Q) = \frac{F_{mica,bulk}}{(1-\exp(iQz_i))} + F_{mica,interface} + F_{water,interface} + F_{water,bulk} \quad 2.3$$

The nuclear density for the XRR calculation is computed in the absence of periodic boundary conditions (PBCs) normal to the interface and takes full advantage of PBCs in the other directions of the simulation supercell.[61] The electron density (probed by XRR experiments) can be related to the simulated nuclear density of atoms by a simple convolution, so we utilize the Fourier transform of the nuclear density profile to generate the XRR spectra.

WATER ORIENTATION

The absolute orientation of a water molecule relative to a K^+ ion or the surface is defined by two angular coordinates: θ_μ and ϕ.[45,63] θ_μ is the angle formed by the vector from the water oxygen to the ion or surface, $\overrightarrow{r_{OS}}$, and the dipole moment vector of the water molecule, $\vec{\mu}$ (local z axis). ϕ is the angle made by the projection of $\overrightarrow{r_{OS}}$ onto the local xy-plane of the water molecule. The local x-axis is the normal to the H–O–H plane, and the local y-axis is the vector from one H to the other.

2.4 RESULTS AND DISCUSSION

X-RAY REFLECTIVITY COMPARISON

The simulated XRR spectra can be seen in Figure 2-1B with data points from the experimental XRR signal[42] from a deionized (DI) water–mica interface superimposed. Notably, both simulations produce the correct peak positions, indicating that our simulations reproduce the Bragg planes from experiments. We can also assess the curves with respect to two reference points—the surface fiducial (q = 0.85 $Å^{-1}$ or L = 2.7 Å) that is sensitive to changes on the mica surface and the bulk fiducial (q = 1.83 $Å^{-1}$ or L = 5.8 Å) that is sensitive to changes in the bulk of mica.[64] As expected, our results for both the H_3O^+ and K^+ terminated mica surface reproduce the bulk fiducial from experiments. Our results for the surface fiducial suggest minor differences between the two surfaces, but the H_3O^+ terminated surface provides a better overall fit. A detailed comparison between the two simulated surfaces is difficult because simulated spectra are prone to errors from (1) the discretization of the nuclear density, (2) significantly smaller linear dimension of the mica slab compared to the penetration depth of the X-rays, and (3) the finite size of the simulation box. [62]

Since simulated XRR curves from both surface terminations provide a similar quality of fit to the experimental data, we hypothesize that the natural mica surfaces may have a combination of H_3O^+ and K^+ ions, a result partially supported by NMR findings.[65] Because H_3O^+ is not directly detected, it is difficult to be confident about the dominant H_3O^+ speciation on the surface from XRR experiments. Similar concerns regarding the precise H_3O^+ and K^+ ion populations on the washed mica surface have been raised by a previous classical simulation study.[66] Nevertheless, because a significant body of work on soft matter self-assembly on a mica surface is ion-mediated,[67] it remains crucial to understand the precise surface termination of mica as a function

of the solution conditions and how the different terminations of the mica surface affect interfacial water structure.

SENSITIVITY ANALYSIS OF XRR RESULTS

It is a useful exercise to understand the sensitivity of XRR spectra to changes in the water structure and ion speciation in the vicinity of the mica surface. While it is understood that XRR spectra are extremely sensitive to ion speciation on the mica surface, the sensitivity of XRR experiments to water structure is not well-characterized.[64] Thus, to detect the XRR response to the water structure, we calculate the XRR spectra from four model electron densities corresponding to (1 = H_3O^+) the total nuclear density derived from the simulated H_3O^+ terminated mica as in Figure 1; (2 = no water) the nuclear density from (1) without the contribution from water; (3 = K^+ water) the mica nuclear density from (1) with water nuclear density determined from K^+ termination; and (4 = K^+ and water) the total nuclear density derived from K^+ terminated mica as in Figure 1. The cases (1), (2), and (3) have the same structure of mica but different water structure on the surface and test the sensitivity of MD-generated XRR spectra to water structure at the mica interface.

Figure 2-2 demonstrates that there are no significant differences in the XRR spectra on changing the water structure next to the mica interface (cases 1, 2, 3, and 4). However, the XRR spectra change with swapping of the terminating ions on the surface (case 4), namely, replacing H_3O^+ on the surface with K^+. This is not surprising in light of the study in ref 26 on the calcite surface in water. Although ref 26 finds a dramatic difference in the best fit and simulated water profiles, the overall fit to the raw experimental XRR data is superb in both cases. It is important to point out that the best fits of the XRR data in ref 26 use identical representations of the nuclear densities for calcite. Naively, the result in ref 26 in conjunction with our study could suggest that

the computed XRR spectrum is more sensitive to the precise nuclear density of the clay surface than to the solution response. This might be expected, since hydrogen and oxygen are light elements and have a smaller X-ray cross section, similar to lithium.[68] Unfortunately, the results presented herein, based on the computed XRR spectra, do not allow for a definitive statement about either the termination of mica or the response of water to the mica interface. Given the complexities regarding the precise termination of the mica surface, it would be useful to understand how the experimental XRR changes as a function of solution conditions to ascertain the possible effects of surface termination.

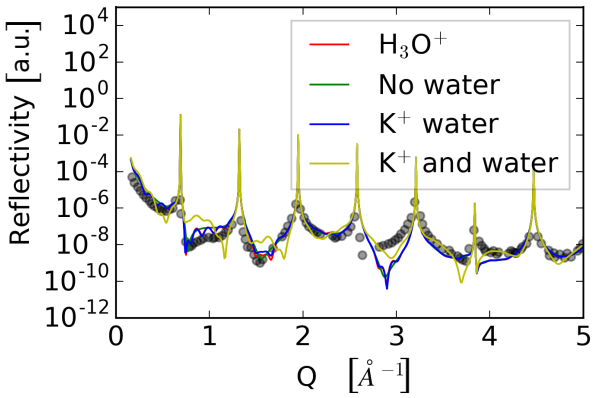


Figure 2-2 Analyzing the sensitivity of XRR curves, calculated from MD simulations trajectories, with changing water and ion structure.

COMPARISON OF WATER STRUCTURE

Due to the high computational cost of ab initio-based interaction potentials, most studies resort to modeling the clay surface with empirical potentials. A direct comparison of DFT studies to the empirical studies also provides an additional self-consistent check on the fidelity of fitted classical models. To this end, we modeled the natural muscovite surface, namely, K^+ terminated, using the CLAYFF[48] potential that has the exact same system size as in the DFT studies. The calculated XRR spectra of the natural muscovite surfaces (Figure 2-1C) are in good agreement with each

other with respect to the bulk and surface fiducials, and Bragg peaks. Nevertheless, there are significant deviations between the XRR spectra obtained from CLAYFF and both the experimental or DFT spectra that are due to the structural differences in the bulk region of the mica (Figure 2-3).

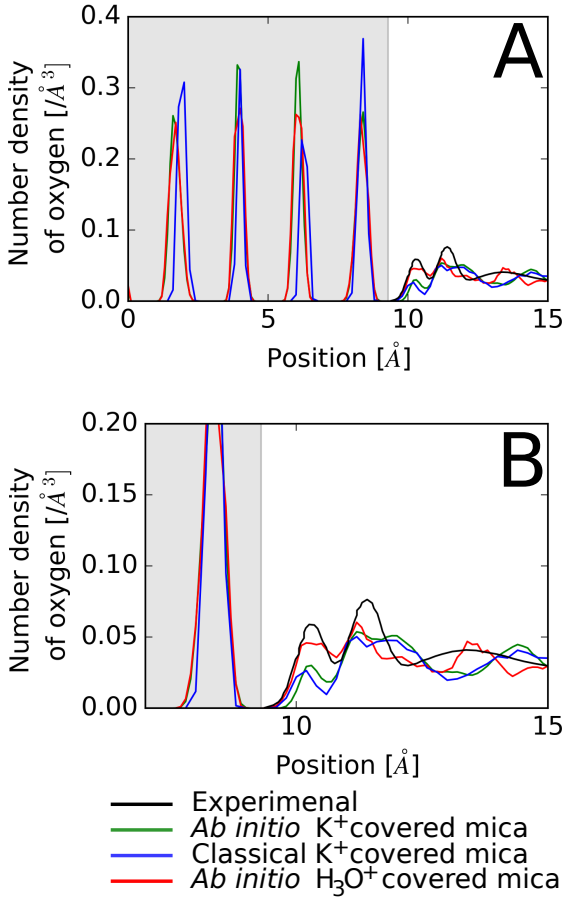


Figure 2-3 Comparison of the water structure on the surface of muscovite. (A) Comparison of the structuring of oxygen on the surface and bulk mica for the ab initio K^+ covered-muscovite (green), ab initio $H3O^+$ (red) covered-muscovite, classical K^+ covered-muscovite (blue), and experimental data points obtained from ref 7 (black). (B) Magnified water structure with the color coding the same as that in part A. The greyscale in both plots denotes the extent of the mica.

Although we have demonstrated that we are able to compute the XRR spectra from an ensemble of molecular dynamics trajectories, our results are not able to uniquely identify the differences in the response of water to mica surfaces with different ion terminations. However,

having access to the trajectories allows a detailed molecular level comparison between the different representations of interaction used in this study as it pertains to the structure of water in the vicinity of the mica–water interface.

We can directly examine the nuclear density of water oxygens (Figure 2-3 A and B) from our simulations. We find a three-tiered hydration layer,[69] with perturbations away from bulk water density, extends up to 8 Å from the mica surface. Beyond 8 Å, the mass density oscillates around 0.0325 waters/Å^3, in good agreement with the density of water under ambient conditions. Extended surfaces, both hydrophilic surfaces such as hydroxylated silica[70] and montmorillonite[71] and hydrophobic surfaces,[72–74] generally display multi-tiered water structuring with different relative heights of the peak indicative of the characteristics of the surface. A high first peak indicates that water "wets" the mica surface. The first peak of the aqueous response to mica is nearly twice the bulk density of water compared to the first peak of water pair-correlation function is 3 times the bulk density.[75] Although we cannot argue for a precise termination of mica using XRR alone, our water density profile is in excellent agreement with the indirect calculation of the water density profile in the XRR study by Lee et al.[68] that assumed a H_3O^+ terminated mica surface.

Although the computed XRR data in Figure 1 cannot distinguish the water response to different terminations of mica, the differences in water density profiles between both surface terminations of mica reveal quantitative differences. The K^+ terminated mica surface features a significantly smaller first peak than the H_3O^+ where it accommodates K^+ ions. The water density profiles for the K^+ terminated surface match the simulation of the same system performed by Wang et al.[66,76] Moreover, the water structures from both the DFT and CLAYFF interaction potentials of the K^+ terminated mica surfaces are in good agreement with each other and previous studies.[77] Unfortunately, there is no reliable classical model for H_3O^+ interacting with a clay surface to

systematically study the variety of different terminations that could be germane to the experimental conditions.

WATER ORIENTATION AROUND MICA

In addition to notable differences in water structure between different surface terminations, in Figure 2-4, we seek a more detailed picture of water on the surface of mica. We can assess the absolute orientation of water within 3.5 Å (or a single water molecule distance) from the surface by reducing the orientation into two angles θμ and ϕ.[63] The DFT results for the H_3O^+ terminated mica surface suggest that there is a population of water, at (20°, 0°), that lies at the surface with its dipole vector perpendicular to the basal plane and hydrogens pointed at the surface. The other prominent population, at (70°, 90°), indicates that water points a single hydrogen toward the surface and the other interacts with bulk water.

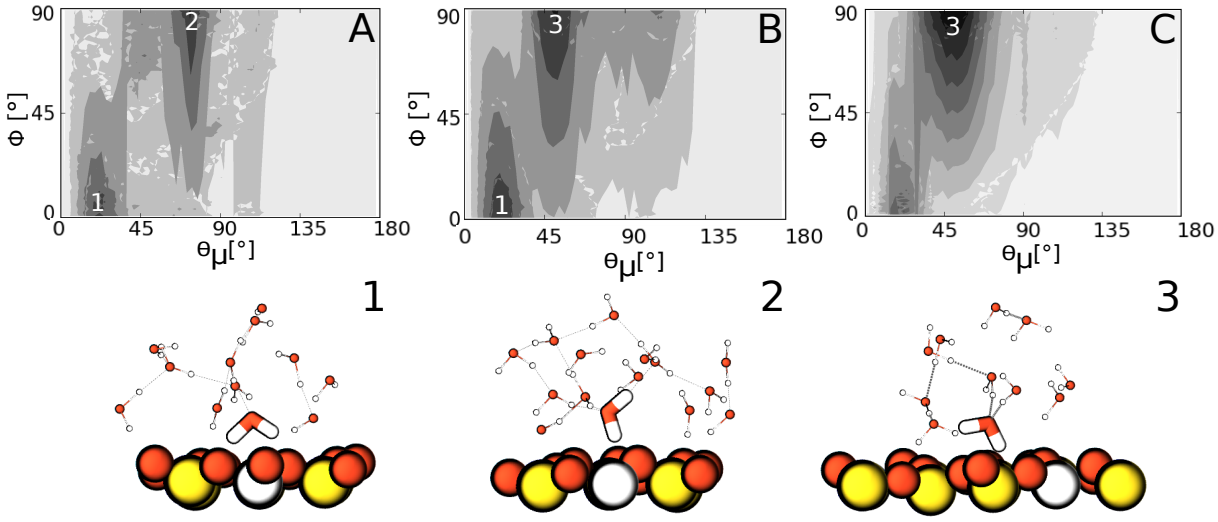


Figure 2-4 Comparison of water orientation using the analysis suggested by Jedlovszky et al.(28) Looking at water within 3.5 Å of the mica surface for (A) ab initio H_3O_+ covered-muscovite, (B) ab initio K^+ covered-muscovite, and (C) classical K^+ covered-muscovite. (Lower panel) Schematic representations of the most probable orientation orientation of different water with

respect to the basal plane. Lower-case labels “1”, “2”, and “3” provide a mapping from structure to absolute orientation.

A similar picture emerges for the DFT simulations of the K^+ terminated mica surface. There is a significant population at (70°, 90°), indicating a large directing effect of the surface ions on water. An additional population at (50°, 90°) suggests a small but statistically meaningful effect on the orientation of water in the first layer. Although oxygen density profiles are indistinguishable between the DFT and CLAYFF regarding the K^+ terminated mica surface (Figure 2-4), differences appear when examining the precise details of the orientation of water on the surface. The population at (20°, 0°) is significantly depleted for the CLAYFF model, suggesting that water is always hydrogen bonded to the water layer above the mica surface. Although difficult, the possibility of the detection of water in a specific orientation using surface sensitive vibrational spectroscopy is the only way to determine the relative accuracy of the interaction potentials to describe water in the vicinity of the natural mica surface. Our simulations also provide information regarding the orientation of H_3O^+ on the mica surface. Having a reactive defect, such as the proton, in the vicinity of an interface could lead to interesting speciation. However, our results suggest that the proton remains in the vicinity of the interface for the duration of our simulations. This is somewhat surprising given that the barriers to proton transfer in bulk water allow for proton diffusion on the picosecond time scale.[78,79] Interestingly, the terminating protons do not donate all three hydrogen bonds to the mica surface, forming an inert, hydrophobic Eigen-form.[80] Rather, as shown in Figure 2-5, the H_3O^+ donates two hydrogen bonds to the mica surface and the third is to a water comprising the liquid. Moreover, we do see fluctuations toward the Zundel transition state (e.g., a shared proton configuration shown in Figure 2-5B), but we observe no proton transfer events on the time scales of our simulation. The protons do not move further than 5 Å, from the

surface during the simulation, suggesting that they are strongly attracted to the negatively charged basal surface of mica.

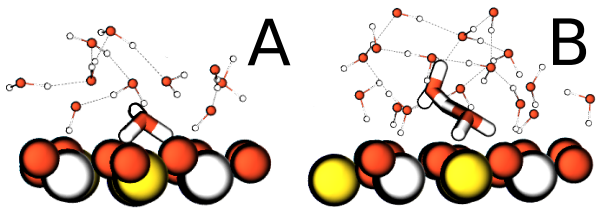


Figure 2-5 Snapshots from the simulation that show common H_3O^+ ion structures from ab initio simulations. (A) H_3O^+ donating a single hydrogen bond to water comprising the liquid and (B) H_3O^+ sharing a proton with a water comprising the liquid in the Zundel form.

CORRELATING MOLECULAR DETAILS TO CONTINUUM DESCRIPTIONS

One of the challenges in molecular simulations is to relate the molecular details, such as interfacial structure, to physicochemical phenomena at larger length scales and relevant boundary conditions. In order to illustrate the connection between different scales, we use the dilute limit of a monovalent aqueous electrolyte. Following the Gouy–Chapman–Stern model,[81] an adsorption of counterions on the surface leads to an absolute potential decrease that is the difference between the so-called surface and zeta potentials (i.e., an electrokinetic potential measured at the slipping). In general, this is influenced by the physicochemical nature of the solution including ion–surface hydration and interionic interactions as a function of ionic strength and type of electrolyte.[36,81]

For a flat surface geometry, the surface potential, ϕ_s, can be estimated from the surface charge density, σ_s, via the Grahame equation[82]

$$\sigma_s = 2(2\hat{\varepsilon}_b\hat{\varepsilon}_0\, k_b T\, n_b)^{\frac{1}{2}} \sinh\left(\frac{\phi_s}{2}\right) \qquad 2.4$$

where $\hat{\varepsilon}_b$, $\hat{\varepsilon}_0$, and k_B are the dielectric constant of liquid at bulk, the vacuum permittivity, and the Boltzmann constant. Here, nb and T are the bulk concentration of ions and system temperature. An equivalent surface charge density can be obtained from Figure 6B: ~0.04 e/(bohr^2) (≡2.289

C/m^2). Given that our simulation contains 512 water molecules, the introduction of a single ion would constitute roughly a 0.1 M solution. An equivalent bulk concentration of ions can be roughly estimated in the range of no ions (∼7 × 10–6 to ∼10–1 M). Using this assumption in Equation 2-4, the surface potential is estimated to be ∼ –248.28 mV as a lower limit (up to ∼ –495.61 mV). Given that an effective surface potential, obtained from a fitting of surface force measurements based on DLVO theory (i.e., the potential at the outer Helmholtz plane) under dilute conditions, is ∼ –150 mV,[83] the estimated surface potential from our simulations is comparable, considering possible uncertainties involved in determining ion concentrations in our simulations and the fitting based on DLVO theory. This comparison provides a simple correlation between molecular simulations and a boundary condition at the macroscopic scale, analogous to a work by Botan et al.[84]

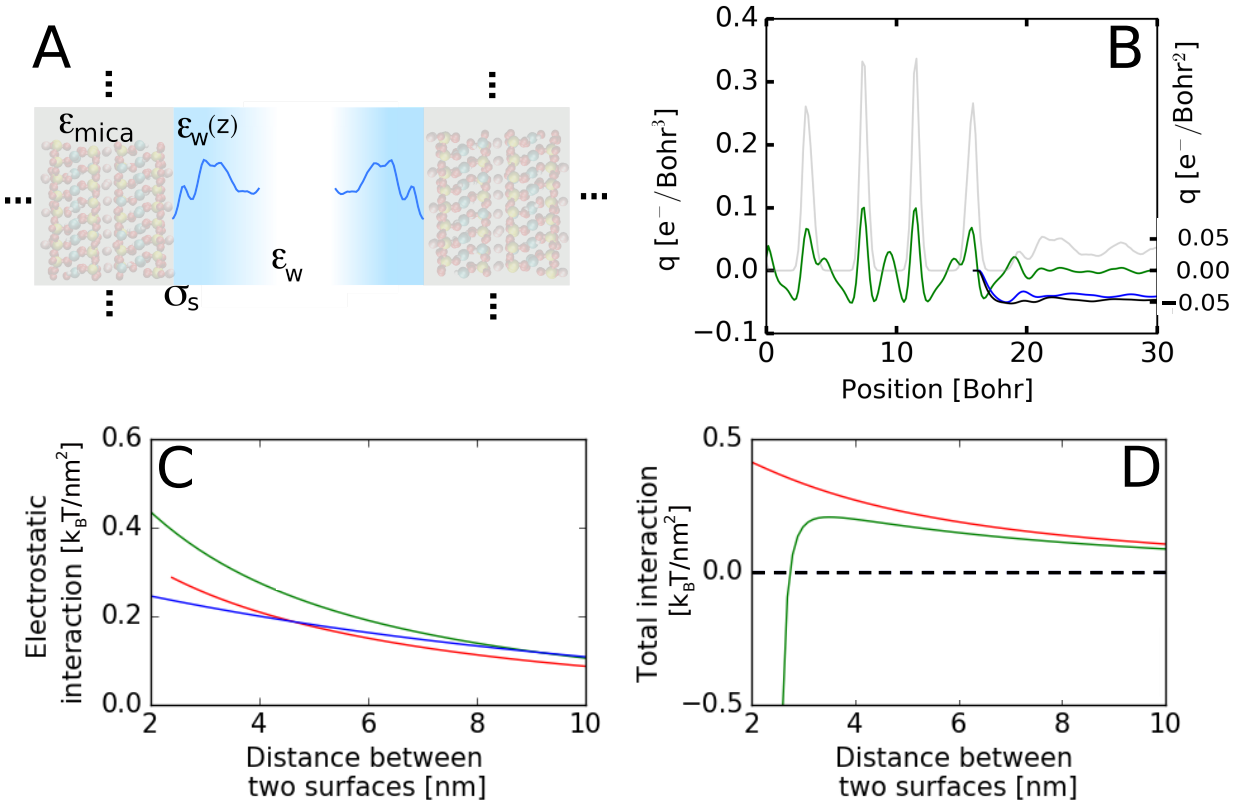


Figure 2-6 (A) Schematic of the mapping of the microscopic structure to a continuum framework for the calculation of the total interaction between two mica surfaces. The gray (with transparent mica structure) represents an infinite, nonperiodic mica slab in the directions of the ellipsis dots with dielectric constant, εmica. Between the mica surface are regions of bulk water denoted by dielectric constant, εw, and nonuniform water density represented by a spatial dielectric constant, εw(z). σs denotes the charge density of the surface that is computed from our trajectories in panel B. (B) Spatial charge density distribution (green) as a function of z for the K^+ terminated mica surface (H_3O^+ not shown) computed using DFT (units in leftmost ordinate). The running integral of the charge density starting at the terminating oxygen of the mica surface for both the K^+ (black) and H_3O^+ (blue) yielding σs (right ordinate). The gray scaled curve denotes the oxygen mass density profile for reference. (C) Electrostatic interactions between two mica surfaces based on full nonlinear PBE (FNLPB) with and without the nonuniform water density (red and green,

respectively) and superposition approximation (blue). (D) Total interactions between two mica surfaces based on DLVO theory (red) from FNLPB and conventional Lifshitz theories and DLVO theory incorporating the nonuniform water density (green).[47]

The structure of interfacial water in the vicinity of the mica is expected to influence both the electrostatic and dispersion interactions between two macroscopic surfaces. Water screens electrostatic potentials, thereby affecting their interactions (or free energies). Screening is also implicated for its effect on dispersion interactions. The fluctuations in water density in the vicinity of interfaces shown in Figure 2-3 introduces perturbations in the dielectric response near the interface, as studied by Chun et al.[47] In fact, understanding the origins of the macroscopic forces that drive nano-assembly in solution requires a molecular approach. The ingredients of a successful theory should require a mapping of solvent and charge responses into theories of particle interactions such as DLVO theory. Here we treat both the electrostatic and dispersion interactions between two mica surfaces at the same level by incorporating the response of water to the mica surface into macroscopic formulation of electrostatics and dispersion. This is schematically depicted in Figure 2-6A.

In general, the electrostatic potential, ϕ, near a flat surface can be estimated using a generalized Poisson–Boltzmann equation (PBE)

$$\hat{\varepsilon}\frac{d^2\phi}{dz^2} + \frac{d\hat{\varepsilon}}{dz}\frac{d\phi}{dz} = -e\sum_{k=1}^{N}\hat{z}^k\, n_b^k \exp\left(-\frac{e\hat{z}^k\phi}{k_BT}\right) \qquad 2.5$$

where n_b^k denotes the bulk concentration of the kth ionic species, $\hat{z}^k$ is the valence of the kth ionic species, N is the total number of ion species, $\hat{\varepsilon}(z)$ is the dielectric permittivity, and e is the unit electron charge. Here, z is a coordinate increasing from the surface. Note that $\frac{d\hat{\varepsilon}}{dz} = 0$ has been assumed so that the second term in the left-hand side of eq 2-5 is typically ignored in the PBE.[82,85]

Implementing $\hat{\varepsilon} = \hat{\varepsilon}_0\varepsilon(z)$ ($\hat{\varepsilon}_0$ is the vacuum permittivity, and ε(z) is a dielectric constant varying over z direction), eq 5 becomes

$$\varepsilon(z)\frac{d^2\phi}{dz^2} + \frac{d\varepsilon(z)}{dz}\frac{d\phi}{dz} = -\frac{e}{\hat{\varepsilon}_0}\sum_{k=1}^{N}\hat{z}^k\, n_b^k \exp\left(-\frac{e\hat{z}^k\phi}{k_BT}\right) \quad 2.6$$

To connect the local solvent response to dielectric constant ε(z), we invoke the Onsager model[86]

$$\frac{(\varepsilon-n^2)(2\varepsilon+n^2)}{\varepsilon(n^2+2)} = \frac{4\pi\rho\mu_0^2}{9k_BT} \quad 2.7$$

where n is the refractive index of the liquid in a long wavelength limit, ρ is the number density of liquid, and μ_0 is the dipole moment of liquid in a vacuum. Note that eq 2-7 is based on c.g.s. unit. Rewriting eq 7, one can obtain the dielectric constant of liquid ε(z) at a given density of liquid ρ(z)

$$\varepsilon(z) = \frac{B(z)+\sqrt{B(z)^2+8n^2}}{4} \quad 2.8$$

where

$$B(z) = n^2 + (n^2+2^2)\frac{4\pi\rho(z)\mu_0^2}{9k_BT} \quad 2.9$$

Equations 8 and 9 clearly correlate the density variation of solvent near interfaces with the variation of the dielectric constant of liquid. A previous study by Hill[87] demonstrated that the Onsager model is satisfactory for water within the uncertainty of ε_∞ if n^2 is replaced by ε_∞. This indicates that 4.5 would be the best value for ε∞ to obtain a known dipole moment of water at a vacuum (i.e., $\mu_0 = 1.83 \times 10^{-18}$ [stat C·cm]) using eq 2-7. Note that eq 2-7 gives rise to a known dielectric constant of water (i.e., ~80) with $n^2 = 4.5$, $\mu_0 = 1.83 \times 10^{-18}$ [stat C·cm], and the density of water at bulk (i.e., ρ = 0.0325 waters/Å^3). Consequently, eq 2-7 can be used to obtain ε(z) for water, along with such values for μ_0 and ε_∞ (=n^2), and water density profiles from aforementioned simulations.

The electrostatic pressure p (or electrostatic force per unit area in this case) between two flat identical surfaces at a given separation $\tilde{L}$ results from the excess osmotic pressure. For symmetric electrolytes, it can be described as[82]

$$p(\tilde{L}) = k_B T \sum_{k=1}^{N} n_b^k \left[e^{-\widetilde{\phi_m}(\tilde{L})} - 1\right] = 2k_B T n_b (\cosh\widetilde{\phi_m}(\tilde{L}) - 1) \quad 2.10$$

where $\tilde{\phi}_m$ is a normalized electrostatic potential at the midplane (i.e., $\widetilde{\phi_m} = \hat{z}_e \phi_m / k_B T$). Subsequently, an electrostatic interaction (or free energy) per area between two flat identical surfaces at a given separation $\tilde{L}$, $G_{elec}(\tilde{L})$, can be obtained by

$$G_{elec}(\tilde{L}) = \int_{\tilde{L}}^{\infty} p(\tilde{L}) d\tilde{L} \quad 2.11$$

We refer the reader to the Supporting Information regarding the details of calculating $\tilde{\phi}_m(\tilde{L})$.

Figure 2-6 shows illustrative calculations for interactions between mica surfaces by using representative physicochemical conditions corresponding to dilute electrolyte solutions: pH ~ 7 (where $\phi_s \sim -150$ mV[83]) and 1 mM of NaCl. It is important to note that our simulations (see Figure S4 in the Supporting Information for details) clearly demonstrate that there is no perturbation of the average water density profile due to the addition of an ion to a simulation cell of 512 waters (i.e., up to a notional concentration of 0.1 M). Subsequently, the presence of 1 mM NaCl will not disturb the average water profile that is used to connect the microscopic principle of solvent response to macroscopic forces between mica surfaces.

The electrostatic interaction between mica surfaces (Figure 2-6C) is based on three different approaches. A superposition approximation for the electrostatic interaction per unit area can be obtained by simply adding potentials from two isolated plates. This requires that a separation between surfaces is large compared with κ–1 (κ is an inverse Debye length) and can be described as[82]

$$\frac{G_{elec}}{k_B T} = 64 n_b \frac{1}{\kappa} \tanh\left(\frac{\widetilde{\Phi_s}}{4}\right)^2 e^{-\widetilde{\kappa L}} \quad 2.12$$

As per our results, the superposition approximation overestimates the electrostatic interaction as the separation decreases. More importantly, Figure 2-6C clearly demonstrates that the solvent structuring decreases the electrostatic repulsion and its effect becomes more critical as the separation decreases. Combined with dispersion interactions from our previous revised Lifshitz theory,[47] which also incorporates the same solvent response to the mica surface, Figure 2-6D demonstrates that the force becomes attractive at a relatively large separation (∼3 nm) under the solution conditions of pH ∼ 7 and 1 mM of NaCl. This is in contrast to the result that is obtained from DLVO theory that is a combination of full nonlinear PBE and conventional Lifshitz theories, predicting that a net attractive force occurs below ∼1 nm. Our research suggests that this difference is entirely due to the solvent response to the mica interface. SFA measurements between mica[88] have demonstrated that a jump-to-contact occurs at less than ∼5 nm in aqueous conditions spanning a range of pH values. More importantly, previous SFA measurements between two mica surfaces by Pashley[89] have observed that the jump-to-contact occurs at 2–3 nm of separation under physicochemical conditions (1.4 mM NaCl at pH 5.7) comparable to what is considered here. This exercise demonstrates that directly using information from a molecular simulation can inform continuum-based theory and produce a dramatically different picture of the forces that are responsible for particle assembly.

2.5 CONCLUSION

We have explored the solvent response to two limiting mica surfaces—H_3O_+ and K^+ terminated, using DFT and classical (K^+ terminated only) simulations. One of the main objectives of the study was to determine whether the use of high-quality interaction potentials based on DFT can produce better agreement with XRR measurements regarding the mass density response of water to the mica surface. Although the XRR spectra generated from our simulations show

reasonable agreement with experimental spectra, it is not straightforward to discern between representations of interaction or terminations of the mica surface. Further, we have shown that the XRR spectra are most sensitive to ion speciation on the surface, and significantly less sensitive to the water structure at the interface, in accordance with other research.[61] This might allude to a combination of H_3O^+ and K^+ ions on the mica surface under experimental conditions, suggesting that the surface termination of mica surfaces may be more complex than currently assumed. This finding also has implications for experiments that assume complete coverage by either H_3O^+ or K^+ ions under certain experimental conditions.[90] Although both the DFT and CLAYFF produce similar water density profiles for K^+ terminated mica, there are important details in the orientation of water with respect to the mica interface that are different and can possibly be resolved through surface sensitive spectroscopy such as sum-frequency generation.[91] Finally, we demonstrated that the water structure resolved from molecular simulations can be used to improve continuum-based descriptions of colloidal forces. Using a modified Poisson–Boltzmann equation and the number density obtained from simulations, it is possible to predict the jump-in-contact distance between two mica surfaces consistent with the SFA experiments. Such improvements in continuum-based theories are crucial for an accurate description of macroscopic forces that drive the assembly of nanostructured materials.[92]

2.6 ACKNOWLEDGEMENTS

The authors gratefully acknowledge conversations with Dr. Gregory Schenter, Dr. Marcel Baer, Dr. Sebastien Kerisit and Evgenii Fetisov. MD simulations and manuscript preparation were supported by the MS3 (Materials Synthesis and Simulation Across Scales) Initiative at Pacific Northwest National Laboratory (PNNL), a multi-program national laboratory operated by Battelle for the U.S. Department of Energy. Supporting MD simulations, detailed data analysis, and

manuscript preparation were supported by the U.S. Department of Energy (DOE) office of Basic Energy Sciences Grant No. BES DE-FG02-09ER46650. Theoretical analysis for macroscopic forces/interactions was supported by the U.S. DOE, Office of Basic Energy Sciences, Synthesis Science and Processing Program. This research used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. DOE under Contract No. DEAC02-05CH11231. Additional computing resources were generously allocated by PNNL's Institutional Computing program and the University of Washington's IT department.

Chapter 3. PEPTOID BACKBONE FLEXIBILITY DICTATES ITS INTERACTION WITH WATER AND SURFACES: A MOLECULAR DYNAMICS INVESTIGATION

Arushi Prakash[1], Marcel Baer[2], Christopher J. Mundy[1,2], Jim Pfaendtner[1,2].

[1]Department of Chemical Engineering, University of Washington, Seattle, USA.

[2]Physical Science Division, Pacific Northwest National Laboratory, P.O. Box 999, Richland Washington 99352, USA

3.1 ABSTRACT

Peptoids are peptide-mimetic biopolymers that are easy-to-synthesize and adaptable for use in drugs, chemical scaffolds, and coatings. However, there is insufficient information about their structural preferences and interactions with the environment in various applications. We conducted a study to understand the fundamental differences between peptides and peptoids using molecular dynamics simulations with semi-empirical (PM6) and empirical (AMBER) potentials, in

conjunction with metadynamics enhanced sampling. From studies of single molecules in water and on surfaces, we found that sarcosine (model peptoid) is much more flexible than alanine (model peptide) in different environments. However, the sarcosine and alanine interact similarly with a hydrophobic or a hydrophilic. Finally, this study highlights the conformational landscape of peptoids and the dominant interactions that drive peptoids towards these conformations.

3.2 INTRODUCTION

Biomimetic polymers are being developed to replace and improve upon naturally occurring biomaterials. Peptoid polymers are one such class of polymers that are thermally and enzymatically more stable than their naturally occurring counterparts (or peptides)[93]. In fact, researchers have already demonstrated their use as chemical scaffolds, drug carriers, antimicrobial coatings, anti-cancer compounds, and antibiotics[94–98]. These applications are possible due to the diversity of structures formed by self-assembling polypeptoids. For example, polypeptoids with alternating aromatic and charged side chains stack up as elongated chains with zig-zag backbones to form nanosheets at the air-water interface[99] or nanoribbons on a mica surface[90,100]. Di-block, amphiphilic peptoids with aromatic and charged side chains can form 2D self-repairing membranes with a similar conformation[101]. Finally, peptoids with α-chiral side chains form protein-like helical structures[102–105].

Peptoid polymers can form these structures due to several intra- and inter-molecular factors. Intra-molecular factors like chirality and hydrophobic side-chains allow peptoids to form helical structures.[103] The length of the peptoid also determines the stability of helices.[106] However, structures like bilayers, nanosheets, and nanoribbons are formed predominantly through inter-molecular and environmental factors. For example, nanosheet formation is highly dependent on inter-molecular electrostatic interactions.[99] Kudirka et al. showed that block co-polymer peptoids

formed more ordered and resilient nanosheets due to the stabilizing effect of nearby side chains.[99] The presence of an interface also stabilizes peptoid assemblies. The hydrophobic air-water interface promotes the formation of stable, free-floating peptoid nanosheets by retaining hydrophobic phenyl side-chains at the interface.[107] Similarly, the structured mica interface dictates peptoid assemblies by providing a template for growth.[100]

Consequently, it is essential to understand how peptoids interact with surfaces and solvents at a molecular level to better understand these peptoid architectures. Computational methods such as all-atom and coarse-grained molecular dynamics (MD) are excellent tools to investigate these systems at higher resolutions than those accessible through experiments. For example, Kudrika et al. used a simple charge model with Monte Carlo simulations to determine the ordering of peptoids in bilayers, which was useful to decipher why block peptoid bilayers were more crystalline than other peptoid bilayers.[99] On the other hand, Mannige *et al.* used all-atom MD simulations to establish the optimum peptoid length for the formation of robust nanosheets.[108] All-atom simulations were also used to understand the effect of hydrophobic tails in the self-assembly of peptoids in solution.[90]

While these simulations play a crucial role in understanding peptoid systems, the reliability of their conclusions ultimately hinges on the accuracy of the interactions represented in their model (or force field). Since peptoid force fields do not exist in ready-to-use fashion, researchers typically prepare a peptoid force field by modifying an analogous protein force field. Since peptoids and proteins have different backbone structures, one should not expect protein-optimized force fields to accurately describe peptoids. Indeed, Brandt *et al.* noted that common protein force fields like AMBER and CHARMM do not have the requisite parameters to describe peptoids.[109] Additionally, Voelz *et al.* noted that AMBER represented peptoids incorrectly in explicit water but correctly in implicit water simulations.[110] They suggested the addition of a 2 kcal/mol.rad^2

harmonic restraint on the ϕ dihedral angle (Figure 3-1A) to correct the force field. Similarly, Mukherjee *et al.* also proposed a 1.0 kcal/mol.rad^2 restraint on ϕ – with a 90° phase offset - in the GAFF force field.[111]

Herein, we validate and extend the force field corrections suggested by Mirijanian *et al.* [112] to produce a suitable representation of the peptoid Ramachandran free energy surface. Using the new parameters, we study the molecular driving forces that lead to changes in the adsorption, orientation and conformation of simple peptides and peptoids. Using molecular simulation, we study alanine (Figure 3-1A, top) and its peptoid analogue sarcosine (Figure 3-1A, bottom), as representative compounds, in pure water and near hydrophilic and hydrophobic self-assembled monolayers (SAMs). This allows us to understand how differences between the conformational flexibility of peptides and peptoids alters the equilibrium solution structure in diverse environments, namely the structure in both solution and near a surface. The results presented provide insight into the differences between peptoids and peptides in (a) conformation, (b) hydrophilicity and surface-binding properties, and (c) enthalpic and entropic contributions, thereby providing a gasp of the nanoscale driving forces that lead to aggregation, adsorption, and ultimately the formation of complex peptoid architectures.

3.3 METHODS

DISARCOSINE IN VACUUM

To obtain the free-energy surface (FES) of a canonical peptoid model, a disarcosine molecule was put in a vacuum box of size 3 nm x 3 nm x 3 nm. Charges for classical simulations of disarcosine were derived from the RESP[113] method using Antechamber[114] program. These charges were calculated after geometry optimization and energy calculations using Hartree–Fock (HF) level of theory with 6-31G(d)//6-31G(d) basis set in the Gaussian 09 program[115]. Classical simulations were performed on the GROMACS 5.1.2[116,117] simulation engine in the NVT

ensemble at 300 K temperature using a velocity rescaling thermostat (tau = 0.1 ps). A timestep of 2 fs was used since the hydrogen bonds were constrained with LINCS[52]. Van der Waals interactions were shifted to 0 at 1.1 nm. Electrostatic contributions were calculated explicitly for distances under 1.2 nm, and using particle-mesh Ewald (PME)[53] summations over 1.2 nm. Phi and psi angles were biased with well-tempered metadynamics[23,31,118]. The initial Gaussian height, Gaussian width and bias factor were 1.2 kJ/mol, 0.2 for phi and psi, 16, respectively. Hills were deposited every 1 ps. Simulations were considered converged after the 1-d free-energy surface, of phi and psi angles remained constant for 1 few ns (since it is a small system). After convergence, the hills were integrated to produce a 2-d FES. Metadynamics was implemented using the PLUMED[119] library in GROMACS.

This free energy calculation was also performed using semi-empirical PM6[120] level of theory in the CP2K simulation engine with PLUMED. The same method, as described as above, was used to construct the FES. The side-chain was changed to ethylcarboxyl or ethylphenylchloride for different calculations. The initial Gaussian height, Gaussian width and bias factor were 1.2 kJ/mol, 0.2 for phi and psi, 16, respectively. Hills were deposited every 4 ps using a timestep of 0.5 fs. Additionally, 2-d potential energy surfaces (PES) were generated using PM6 and MO62X/6-31G(d,p)[121] levels of theory from constrained geometry optimizations using Gaussian 09[122] and a grid spacing of 5°.

ALANINE AND SARCOSINE IN WATER AND NEAR SURFACES

Alanine and sarcosine undecamers (11-mers) were created in GaussView.[123] Point charges were determined as described above for disarcosine and dialanine. Each oligomer was solvated with SPC/E[124] water (5 nm x 5 nm x 5 nm). We conducted an energy minimization using the steepest descent algorithm with Van der Waals, neighbor list and coulomb cut-offs at 1 nm. The minimized configuration was equilibrated at 300 K and 1 atm pressure in the NPT ensemble, using

the Berendsen barostat[50] (tau = 1.0 ps) and a stochastic global thermostat[51] (tau = 0.1 ps), for 2 ns. Covalent bonds with hydrogen atoms were constrained using LINCS[52] to simulate with a timestep of 2 fs. To simulate the surface adsorption of oligomeric alanine and sarcosine, we used the same simulation protocol. Owing to the larger size of the SAM surfaces, these systems had larger box sizes (7.8 nm x 6.5 nm x 7.83 nm) which contained 9126 atoms of the SAM. The AMBER compatible force fields simulating model hydrophilic/hydrophobic surfaces (i.e., SH-$(CH_2)_{11}$-COOH and SH-$(CH_2)_{11}$-CH3 alkane thiols) are identical to those used by Sprenger *et al.*[125] All atoms of the SAM, except the two carbon atoms (and attached atoms) nearest the surface, were frozen to maintain the packed structure of the of the monolayer. The remaining volume was filled with SPC/E water molecules and one molecule of alanine or sarcosine.

The collective variables (CVs) selected for biasing with metadynamics were the radius of gyration, a cumulative dihedral parameter (colloquially called "alpha-beta"), and in the case of adsorption simulations[126], the orthogonal distance between a frozen atom in the SAM and the center-of-mass of the molecule. The radius of gyration (using Cα atoms only) is defined as $r_{Gyr} = \frac{\sum_{i=1}^{n} m_i |r_i - r_{COM}|}{\sum_{i=1}^{n} m_i}$. Alpha-beta is defined as $0.5 \sum_{i=1}^{n} \left(1 + \cos\left(\theta_i - \theta_{ref}\right)\right)$, which is a summation over 9 pairs of ϕ and ψ dihedral angles (thus, n=18) in undecamers. This CV can collectively bias all backbone angles without the need for multiple CVs for all the dihedral angles. θ_{ref} was chosen as 2.36 and -2.36 for all ϕ and all ψ angles, respectively, which represents an ideal alpha-helical conformation. Radius of gyration and alpha-beta were chosen so that molecules could explore all conformational degrees of freedom. The distance CV was chosen to obtain the binding free energy. CVs for conformation and distance have similarly been used in previous studies of the adsorption small peptides on surfaces.[125,127] Production simulations were performed with multiple walkers[35] metadynamics, with initial structures generated by using steered MD to position three starting structures in disparate regions of the CV space. The systems of molecules adsorbing to surfaces

were biased using a 3-replica multiple walker parallel-bias metadynamics (PBMetaD) scheme[24,35] to enable sampling of a high-dimensional free energy landscapes. All CVs had an initial Gaussian height of 2 kJ/mol and bias factor of 9. Hills were deposited every 2 ps. The production run was carried out in the NVT ensemble using the cut-offs mentioned for energy minimization. The simulations were considered converged if the free energy profile of the CV of interest did not change significantly during the last 25% of the production run (See Appendix). To quantify the error associated with each free energy profile, a block averaging analysis was conducted (See Appendix). Errors associated with all free energy profiles are less than 0.5 kcal/mol. The metadynamics parameters, and other specifics of the simulation are tabulated in Table 3-1.

Table 3-1 . Metadynamics parameters used in sarcosine (alanine) simulations.

	Solution	CH3-surface adsorption	COOH-surface adsorption
Total atoms	12140 (12158)	40877 (39332)	39893 (37790)
Simulation box size	4.9 x 4.9 x 4.9 nm^3 (4.9 x 4.9 x 4.9 nm^3)	7.8 x 6.5 x 8.1 nm^3 (7.8 x 6.5 x 7.8 nm^3)	7.8 x 6.5 x 7.8 nm^3 (7.8 x 6.5 x 7.4 nm^3)
Biased variables	Radius of gyration, Alpha-beta,	Radius of gyration, Alpha-beta, Distance from surface	
Sigma	0.03 nm, 0.4	0.03 nm, 0.4, 0.025 nm	
Biasfactor, Hill Height, Hill deposition rate, Number of walkers	9, 2 kJ/mol, 500 steps, 3		
Simulation length (per walker)	400 ns (400 ns)	400 ns (400 ns)	1000 ns (400 ns)

After completion of production runs, we processed these trajectories using the g_cluster algorithm (Cα atoms only, 0.2 nm cut-off, GROMOS method) in GROMACS, using every 25th frame. The clusters weights calculated using the technique of Tiwari *et al.*[128] which allowed us to estimate a time-independent offset bias (c(t)) as:

$$e^{\beta c(t)} = \int ds \left[e^{\frac{\gamma V(s,t+\Delta t)}{k\Delta T}} - e^{\frac{\gamma V(s,t)}{k\Delta T}} \right]$$

where β=1/kT, k is the Boltzmann constant, T is the temperature, ΔT is the so-called hills temperature for the well-tempered metadynamics simulation, $\gamma=\frac{T+\Delta T}{T}$ is the bias factor, Δt is the time between deposition of Gaussian hills and V(s,t) is the history-dependent bias potential for the system. The corrected weights of each occurrence of each cluster is calculated using $e^{\beta(V(s,t)-c(t))}$. To illustrate the flexibility of the different molecules, we also tracked the number of clusters over time, by processing different lengths of the trajectory from the initial point. The number of clusters with total weight (calculated from the method described above) greater than 99% were taken as the total number of clusters at that time point. For the PBMetaD simulations, we approximated the quantity *c(t)* using the relationship proposed by Tiwari *et al.*[128] and note that in this case the conclusions in the paper are unchanged irrespective of which of the three bias potentials are used.

As in our prior work[129], we reconstructed the free energy as a function of other system CVs using the reweighting method described above. In this case, we used reweighting to study the variation of the short-range contributions to the potential energy from Lennard-Jones and Coulombic terms. Formally, this is achieved by building a probability distribution P(*U,x*), where *U* represents one of the energies of interest and *x* is one of the biased CVs noted above. Proper weighted averages were then calculated at each distance slice to recover a profile of <*U*> vs *x,* allowing us to track the change in Lennard-Jones (or Coulombic) interaction with change in CV.

3.4 RESULTS AND DISCUSSION

MODEL REFINEMENT

Calculations by Mirijanian *et al.*[112] show that the Ramachandran FES for disarcosine (model peptoid) predicted by CHARMM22, a standard protein force field, is different than that predicted by quantum mechanical methods. Similarly, our results also show that the Ramachandran FES for disarcosine predicted by AMBER99SB-ILDN (Figure 3-1B), another commonly used protein force field, differs from that predicted by the PM6 semi-empirical potential (Figure 3-1C). We use PM6, in lieu of quantum mechanical methods, because these calculations are computationally cheaper. Moreover, the efficacy of PM6 is comparable to M062X, as evidenced by the similarity between the potential energy surfaces (PES) (Figure 3-2). Most discrepancies between PM6 and M062X occur in the higher energy regions. Although the shapes of the minima differ, the minima are approximately in the same location thereby predicting similar conformations. Therefore, we suggest that PM6 predictions are valid in this case. Moreover, the minima predicted by these PES were also corroborated by Baldauf *et al.*[130]

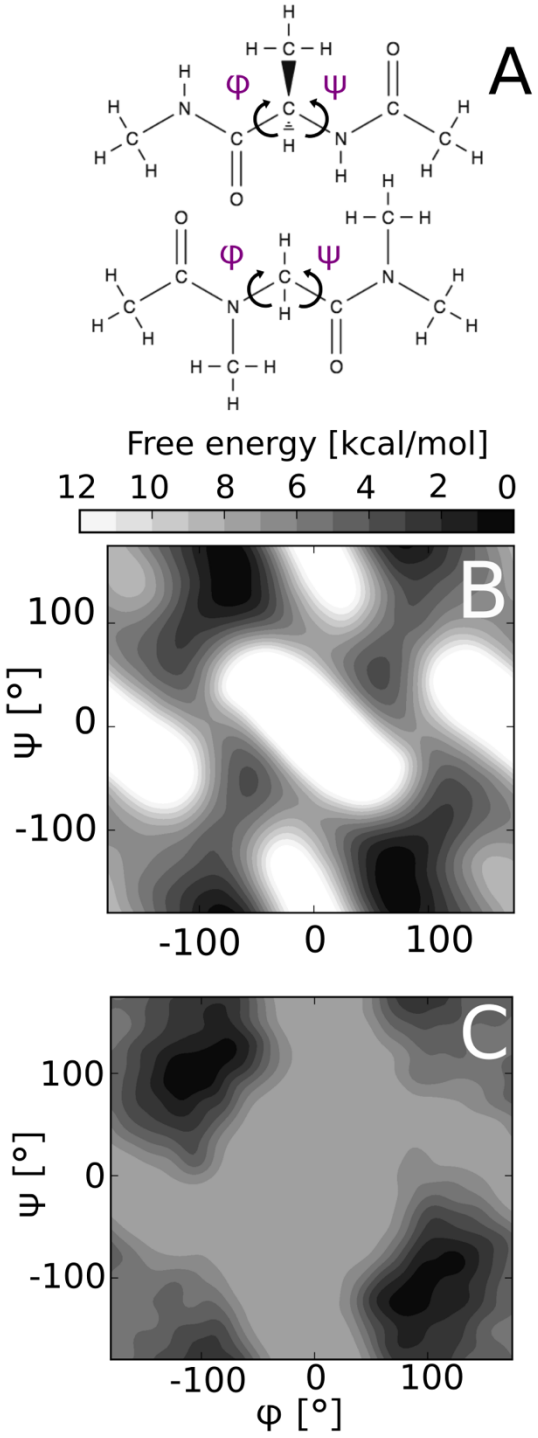


Figure 3-1 (A) Structure of alanine (top) and sarcosine (bottom). Free energy (kcal/mol) surfaces of (ϕ ,ψ) angles (degrees) for disarcosine in vacuum using (B) AMBER99SB-ILDN force field in GROMACS and (C) PM6 in CP2K

Further, we use the Ramachandran FES from PM6 to study the conformations of disarcosine and validate our peptoid model. We use the FES instead of the PES since the FES takes into account thermal fluctuations. The PM6 FES (Figure 3-1C) has two minima centered at (100°,-100°) and (-100°,100°) while the AMBER99SB-ILDN FES (Figure 3-1B) has two minima centered at (90°,-180°) and (-90°,180°). The only point of similarity is that these plots are center-symmetric, i.e. the energy of the point (ϕ ,ψ) is the same as the energy of (-ϕ ,-ψ), which is a result of the achirality of the peptoid backbone.[131] The AMBER99SB-ILDN result suggests that peptoids prefer forming a polyproline type I helix in cis form or a polyproline type II helix in trans form[131], the so-called cis α_D and trans α_D conformations, respectively[132]. Recent B3LYP/6-31+1G(2d,p)// HF/6-31G* calculations, by Butterfoss *et al.*[131], also corroborate these peptoid conformations. However, considering our PM6 calculations, we surmise that the peptoid structure might be different than that predicted by AMBER99SB-ILDN (Figure 3-1B) or Butterfoss *et al.* [131]

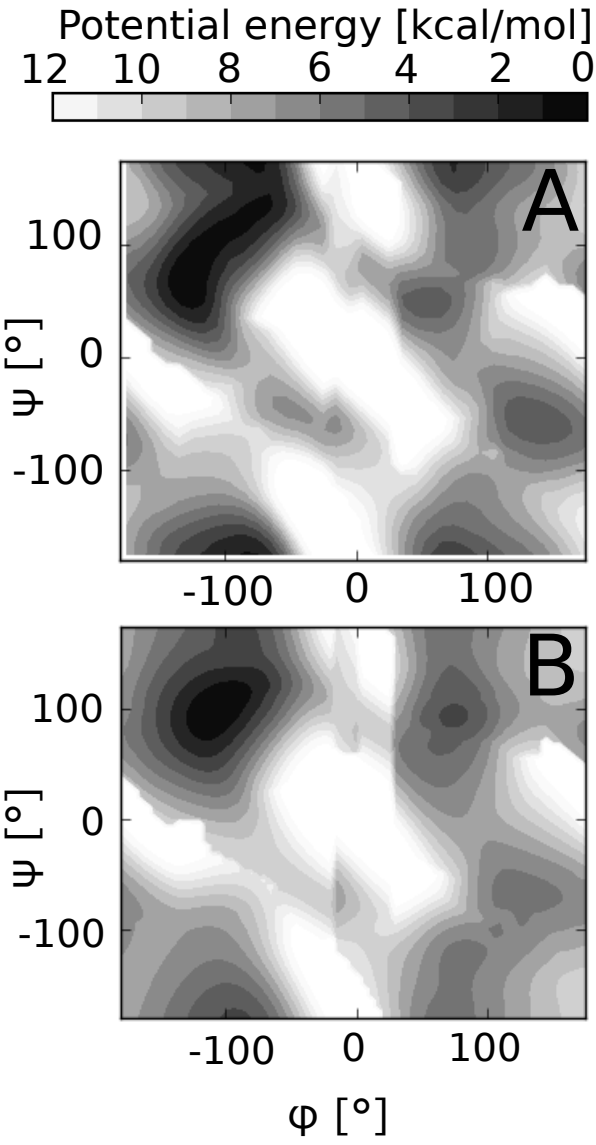


Figure 3-2 Potential energy (kcal/mol) surfaces of (ϕ ,ψ) angles (degrees) for disarcosine in vacuum using (A) M062X and (B) PM6 in Gaussian 09.

We argue that the disarcosine structure is reminiscent of glycine because peptoids are N-substituted versions of glycines. To this end, we found similarities to the glycine Ramachandran PES calculated by Head-Gordon *et al.*[132] using HF/3-21G, HF/6-31+G*, and MP2/6-31 G**//HF/6-31 G* levels of theory. The PES of glycine is also center-symmetric with two wide minima in the top-left and bottom-right quadrants. Using the naming conventions used by Head-

Gordon *et al.*, our conformations are close to $C7_{eq}$ and the saddle point between $C7_{eq}$ and C5. Indeed, Moehle and Hofmann also found that $C7_{eq}$ was a major minimum for disarcosine in their calculations using HF/6-31+G*, HF/6-31G**, HF/DZP, and HF/4-21G.[133] The differences between the protein force field AMBER99SB-ILDN and PM6, and previous studies[109–112] indicate that protein force field parameters may need revisions to accurately represent that peptoid Ramchandran FES. Thus, we modified the AMBER99SB-ILDN force field (Figure 3-3) by adapting the dihedral parameters of Mirijanian *et al.*[112] to obtain the minima predicted by PM6 (Figure 3-1C). All further calculations were done with this modified force field.

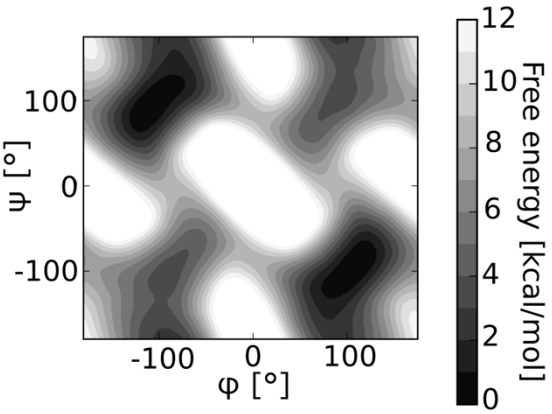


Figure 3-3 Free energy (kcal/mol) surface of the modified/corrected AMBER99SB-ILDN force field in GROMACS.

To ensure that the backbone parameters also represent peptoids available in literature[134,135], we calculated the FES with PM6 after substituting the methyl group in disarcosine with hydrophilic ethyl carboxyl or hydrophobic ethyl phenyl chloride. The Ramachandran FES for these peptoids were almost identical to that of disarcosine (see Appendix) which indicates that these peptoids share common backbone behavior and that the modified force field can represent peptoids with diverse backbones. Notably, our results can be improved by using PM6 with hydrogen bond

corrections[136] but in this system where hydrogen bonds are unlikely, PM6 serves as a good estimate.

SOLUTION PHASE CONFORMATIONS: PEPTOID VERSUS PEPTIDE

We used this modified force field to simulate poly-sarcosine and poly-alanine undecamers (11 residue chains or 11-mers) in water. The chosen CVs (radius of gyration and the alpha-beta function) allowed us to exhaustively sample the conformational phase space of these molecules. After convergence, the free energy profile for the radius of gyration was constructed to assess the most favored conformations in water.

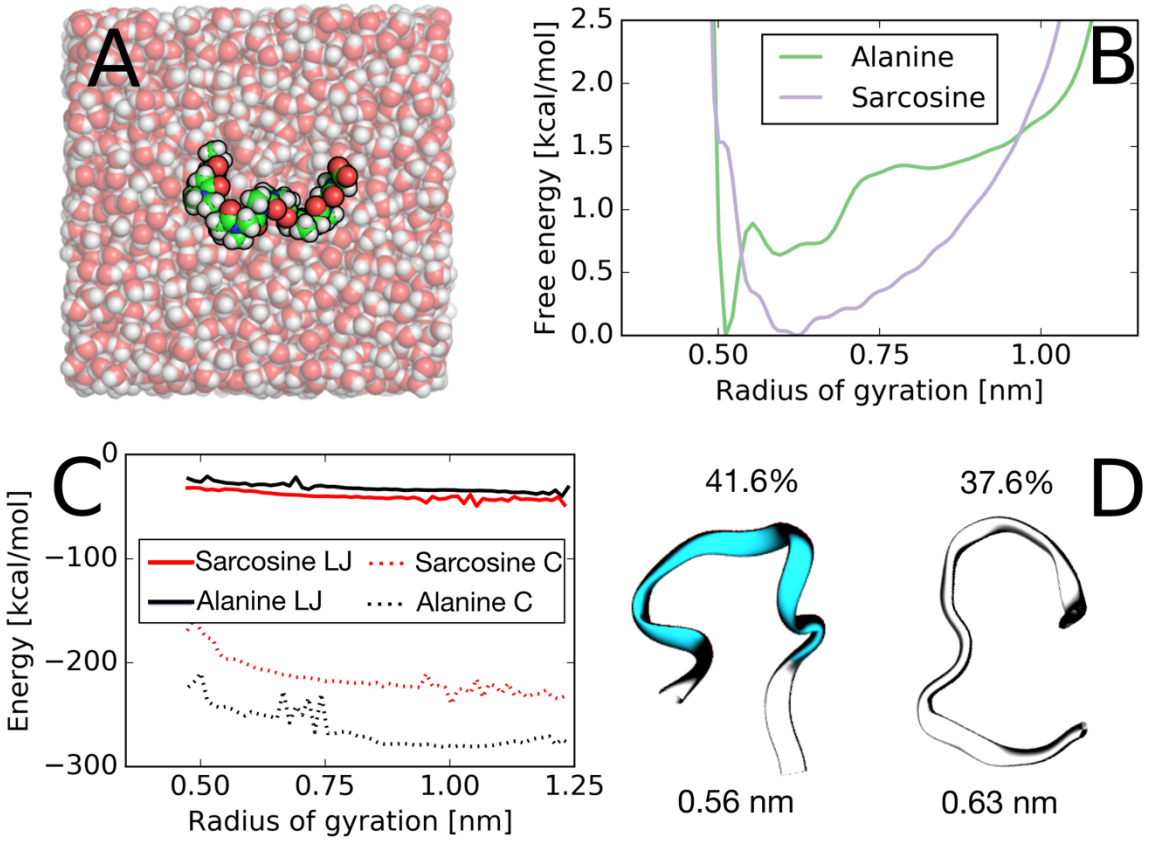


Figure 3-4 Simulation of sarcosine and alanine (11-mers) in water. (A) Simulation setup. (B) Free energy (kcal/mol) curve for radius of gyration in nm for alanine and sarcosine. (C) Energy contributions (kcal/mol) from short-range coulombic and Lennard-Jones interactions versus radius of gyration. (D) Top weighted structure from clustering analysis for alanine (left) and sarcosine (right) where turn and random coil structures are represented by cyan and white colors, respectively, and the radius of gyration is value is provided below.

The free energy profile of alanine (Figure 3-4B) has a sharp minimum at 0.5 nm which indicates that the structure has some preference towards a compact state. On the other hand, the free energy profile for sarcosine (Figure 3-4B) that has a deep well with a minimum at 0.6 nm. The representative structures for these minima, obtained after a clustering analysis combined with reweighting for the metadynamics bias, are provided in Figure 3-4D. Alanine primarily displays a turn/random-coil structure (Figure 3-4D, left). The same was predicted by Levy *et al.*[137], using MD simulations, where coil and beta conformations of alanine were ~ 20 kcal/mol more stable than alpha helical structures in aqueous solution. Completely helical structures are of extremely low weight (not shown here). On the other hand, sarcosine (Figure 3-4D, right) prefers a random coil structure.

We can also understand the energetic contributions to the free-energy by recovering short-range Lennard-Jones and coulombic interaction energies from our simulations (Figure 3-4C). We see that both molecules predominantly interact via electrostatic interactions. In our system, sarcosine interacts significantly stronger with water than alanine (Figure 3-4C), by 50 kcal/mol. Recent HPLC experiments have also shown that 20-mers of polysarcosine interact strongly with

water.[138,139] Hydrophobic molecules elute with less hydrophilic solvents from an HPLC column. But researchers saw that polysarcosine eluted with 81-78% water, indicating that sarcosine was interacting strongly with a solvent.[138] Further, contact angle measurements on sarcosine monolayers also suggest that sarcosine interacts strongly with water (it is comparable with the interaction of poly-ethylene glycol with water).[138] Recently, Arai et al. showed that ~ 4.5 waters were in contact with each sarcosine monomer.[140] Counting the number of waters within 0.35 nm of the heavy atoms of sarcosine, we find that 4.14 waters are in contact with each monomer (See Appendix), which is in good agreement with these experiments.

To further investigate the interaction of these molecules with water, we compute the distribution of water around backbone groups – nitrogen, alpha-carbon and carbonyl oxygen – by the use of radial distribution functions (Figure 3-5A, B). The most striking difference lies in the interaction of water with nitrogen (solid curves, Figure 3-5 A/B). Water could approach the nitrogen on alanine closer than that on sarcosine, an expected result since the nitrogen on sarcosine is shielded by the methyl side-chain. However, water hydrogens could approach alpha-carbons slightly closer for the case of sarcosine (dotted curve, Figure 3-5 A). These differences in water structure around sarcosine and alanine are manifestations of the placement of the side-chain on different backbone atoms (on nitrogen for sarcosine and on carbon for alanine). The water distribution around the carbonyl oxygen remains constant for both molecules since it is unmodified between the two molecules (dashed line, Figure 3-5 A/B). Further, to understand the range until which these differences in water structure persist, we integrated the radial distribution function for all interactions (Figure 3-5 C, D). We see that these differences quickly die out about 0.2 nm from the atoms of interest without contributing to the long-scale ordering of water. We hypothesize that the differences will be stronger and affect longer ranges for side-chains that are bulkier than methyl. However, for the model case of sarcosine and alanine, the differences are quite minor.

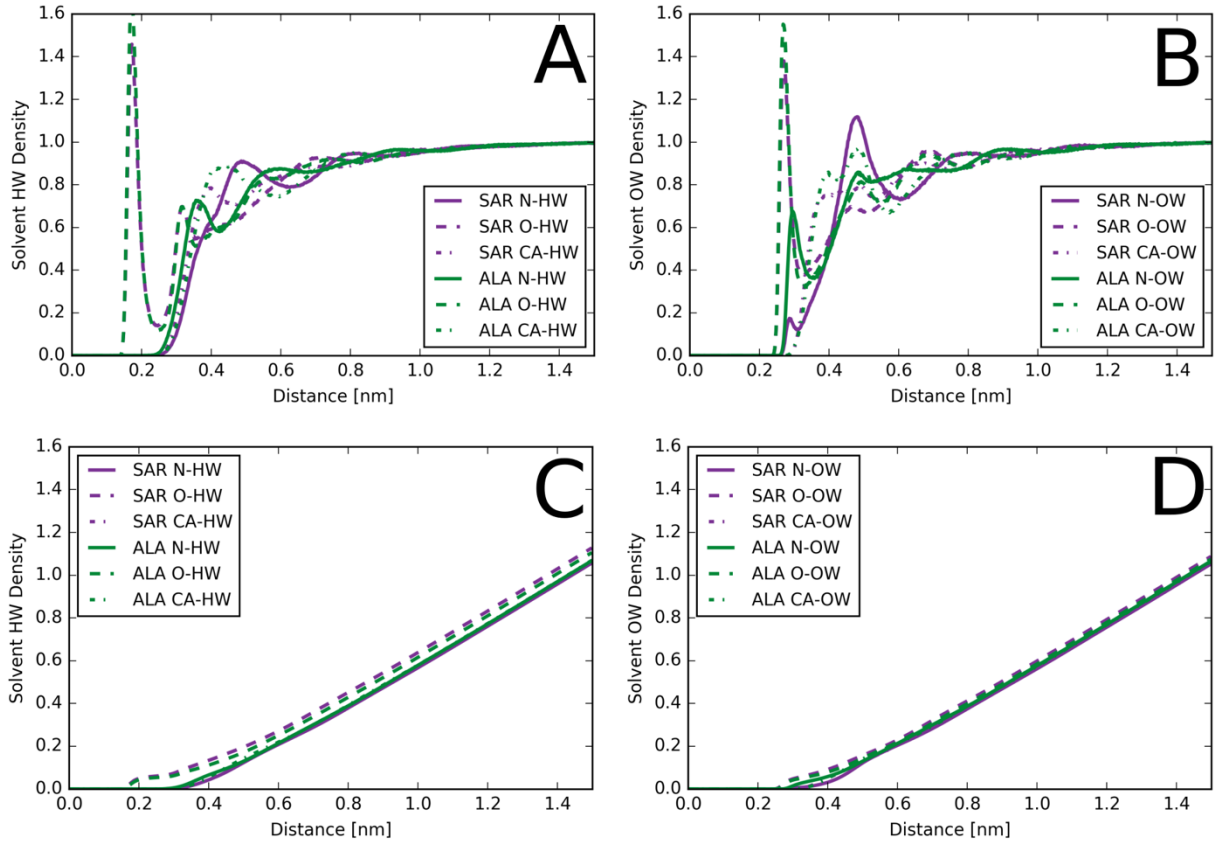


Figure 3-5 Radial distribution functions of water (A) hydrogens (HW) and (B) oxygens (OW) around heavy atoms (N, CA, O) in the backbone of alanine (ALA) and sarcosine (SAR). The integrals of respective distributions are shown in (C) and (D).

PEPTOID VERSUS PEPTIDE NEAR SURFACES

To explore the behavior of alanine and sarcosine at surfaces, we simulated them near SAMs (Figure 3-6A) of CH_3-terminated and COOH-terminated alkane-thiols, to model hydrophobic and hydrophilic surfaces, respectively. These simulations bias an additional CV (distance of the center-of-mass of the alanine/sarcosine from the surface) to ensure that binding/unbinding to the surface and folding/unfolding are simultaneously explored. The free energy curve of the new CV is used to calculate the binding energy of the molecule to the surface. From Figure 3-6B it is evident that both molecules bind comparably to each other at both surfaces. The interaction with a hydrophobic

surface is stronger (by ~ 4 kcal/mol) than with a hydrophilic surface. This difference in interaction between surface types will likely change if the methyl side-chain is replaced with a hydrophilic group.

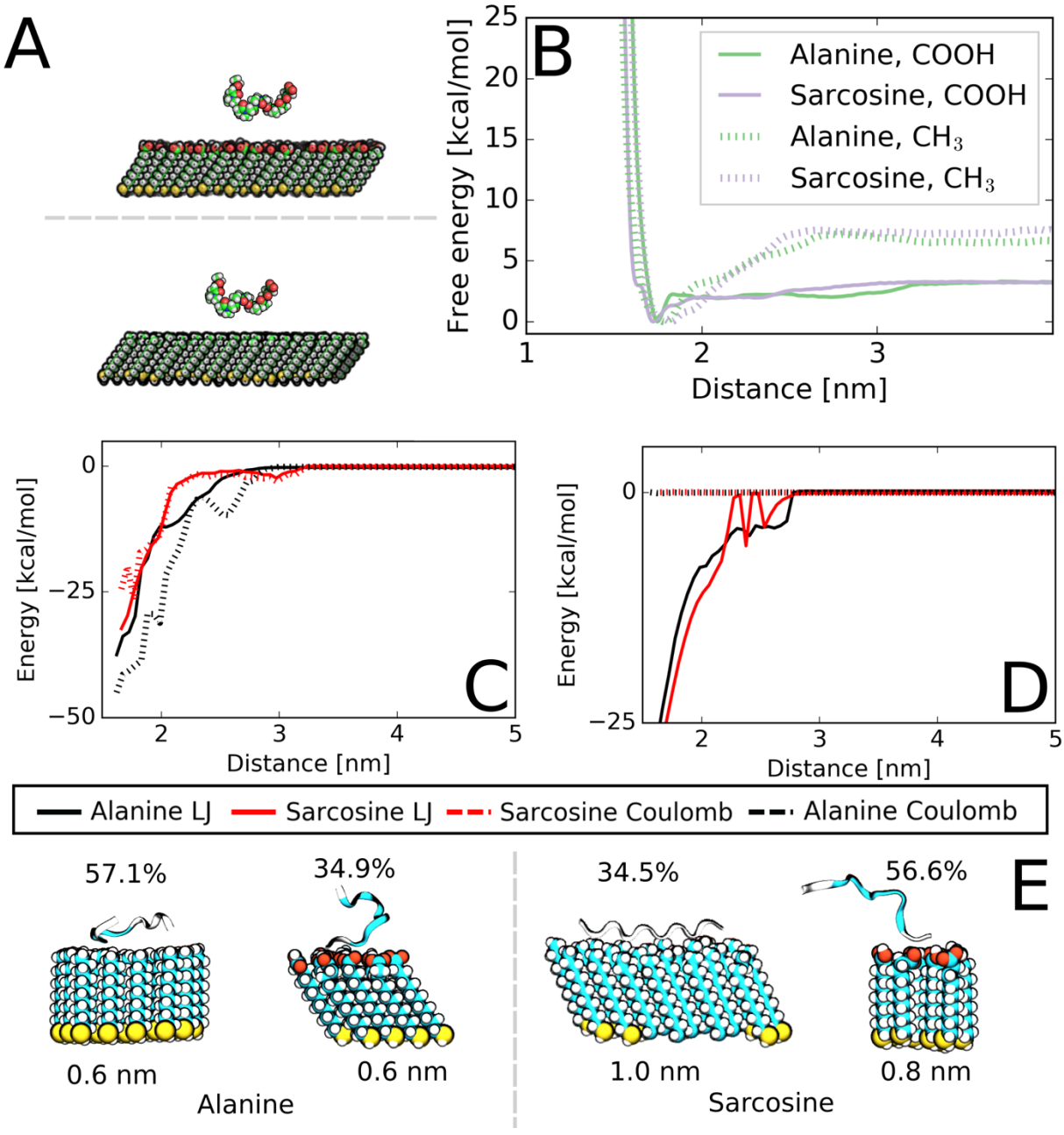


Figure 3-6 Simulation of sarcosine and alanine near hydrophobic (-CH3 terminated) and hydrophilic (-COOH terminated) SAMs. Simulation setup in (A) for hydrophilic surface

(bottom) and hydrophobic surface (top). B) Free energy (kcal/mol) curve for distance from surface in nm. Energy contributions (kcal/mol) from short-range coulombic and Lennard-Jones interactions for distance from the (C) hydrophilic and (D) hydrophobic surfaces. (E) Dominant configurations from clustering analysis for alanine (left) and sarcosine (right) on hydrophilic (left) and hydrophobic (right) surfaces.

Further, we can investigate the dominant energetic contributions that contribute to binding in our model system to understand how molecules can be designed to interact better with a specific surface. For this, the short-range interaction energies (Lennard-Jones and Coulomb), are recalculated from the simulation trajectories and reweighted to find their unbiased distributions. Figure 3-6 D shows the interaction energy of sarcosine and alanine on the hydrophobic surface. The interaction energies of both molecules are comparable; however, the Lennard-Jones energy dominates binding and the Columbic contributions are negligible. This is expected from molecules with hydrophobic side-chains interacting with hydrophobic surfaces. On the other hand, the hydrophilic surface (Figure 3-6 C) shows comparable energetic contributions from Lennard-Jones and Coulombic terms. For sarcosine, the contributions are nearly equal, whereas for alanine Coulombic interactions are larger.

These differences can further be explained by the top-weighted structures on the surface (Figure 3-6 E). On the hydrophilic surface, alanine seems to interact with 2-3 residues, whereas sarcosine interacts with 1-2 residues which might result in slightly higher electrostatic contributions to the binding energy. On the hydrophobic surface, all residues of alanine and sarcosine interact with the surface. Here, alanine maintains its turn structure, while sarcosine becomes extended. This extended structure of sarcosine is reminiscent of the zig-zag backbone structure observed for nanosheets.[108] Notably, this open, floppy structure is possible due to the co-

existence of trans and cis amide bonds in the backbone which results from the low energy difference between cis and trans amides.[141]

These differences in the dominant configuration on the hydrophilic and hydrophobic surfaces may also allude to a difference in binding mechanisms. To investigate their binding mechanisms, we studied the binding of sarcosine of length 2 and 5 on both surfaces, and compared their binding energies (Figure 3-7 A,B). For the hydrophobic surface (Figure 3-7 A), it is evident that the binding energy of sarcosine increases with the number of binding residues. In fact, the binding energy per residue obtained from dividing the number of monomers from the binding energy is nearly equal for all sarcosine polymers at the hydrophobic surface. On the hydrophilic surface (Figure 3-7 B), the binding energy of sarcosine polymers is constant, even on significantly increasing the length of the sarcosine chain. Coupling these results with the dominant structures, the difference in mechanism is evident. On the hydrophobic surface, all residues of sarcosine interact with the surface, presumably to decrease the hydrophobic surface area exposed to the water. Therefore, the binding energy scales with the number of monomers. On the hydrophilic surface, sarcosine interacts through 1-2 residues only, presumably due to comparable binding energy of water/surface and water/oligomer; therefore sarcosine polymers (>1 residues) are expected to have the same binding energy. This is also seen quantitatively when we reweight the simulation of sarcosine undecamer and determine the likelihood of a certain number of bound residues (Figure 3-7 C). Identical to the clustering analysis, the number of bound residues (bound = distance of N atom within 0.65 nm from the surface) is calculated for each configuration and weights (and c(t)) calculated from the simulations are used to recover the probability of binding with N number of residues. From Figure 3-7 C, it is seen that the hydrophilic surface has a strong preference for binding with 2 residues, while the hydrophobic surface is likely to bind with

numerous residues. Along the same vein, looking at the dominant structures of alanine (Figure 3-6 E), it is likely that it interacts with the same mechanism as sarcosine to both surfaces.

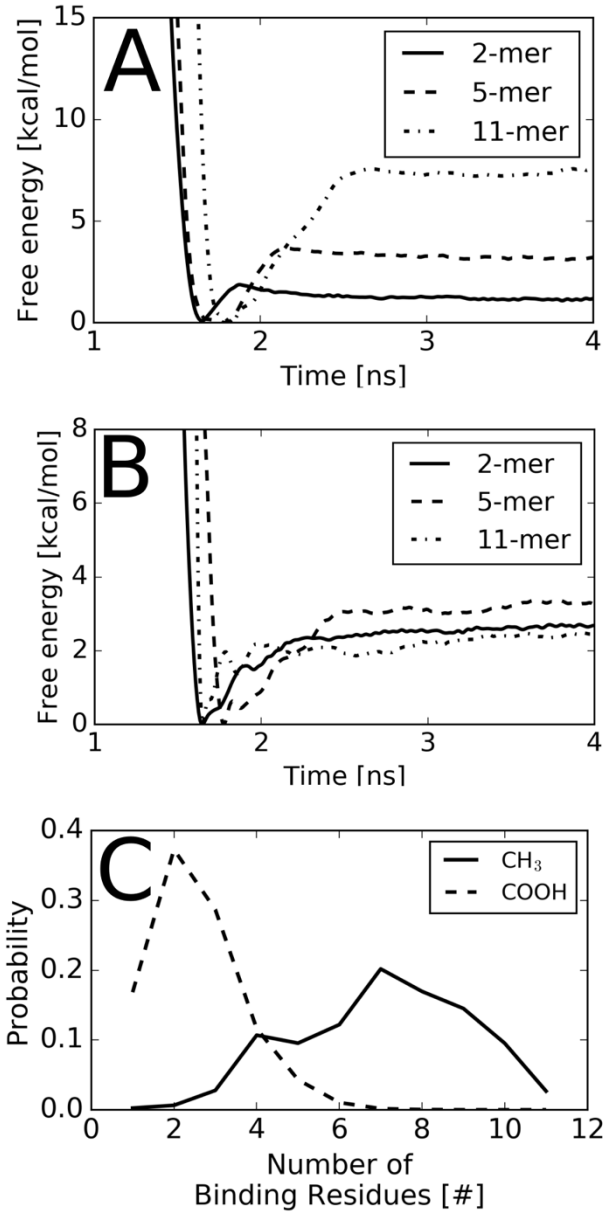


Figure 3-7 Sarcosine. (A) Binding energy (kcal/mol) versus distance from surface for the model hydrophobic surface. (B) Binding energy per molecule (kcal/mol) versus distance from the surface for the model hydrophilic surface (C) Probability versus number of binding residues of sarcosine to each surface.

Finally, to understand entropic contributions arising from conformational flexibility, we tracked the number of distinct structures in our trajectory over time (Figure 3-8). Clearly, sarcosine which exhibits greater number of structures is much more flexible than alanine in water. Sarcosine shows almost 2 times the number of clusters than alanine in water. On approaching a surface (hydrophobic or hydrophilic), both molecules are less flexible, although the decrease in flexibility is much greater for sarcosine. Nonetheless, sarcosine is more flexible than alanine at the surface. From these results, we conclude that sarcosine faces a greater entropic penalty than alanine to adsorb on surfaces (both hydrophobic and hydrophilic). The entropic penalty of adsorption is possibly compensated by more water molecules desorbing from the surface.

The retention of elongated structures at interfaces is evident in peptoids that form nanosheets at the air-water interface.[108,142] Indeed, they require an interface to assemble and do not assemble in solution, where it is likely that these peptoids are in aggregated form. Notably, the presence of hydrophobic side-chains also promote elongated structures.[143]

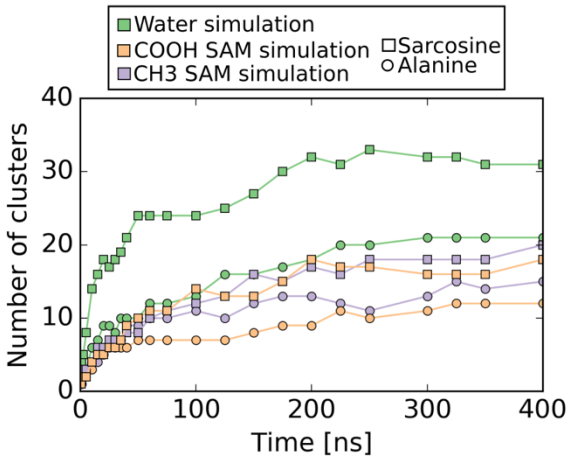


Figure 3-8 Number of distinct clusters of structures or conformations (identified by the clustering protocol in Methods) of alanine and sarcosine up till different time points in the trajectory.

3.5 CONCLUSION

In this study, we noted the failure of the protein force field AMBER99SB-ILDN to represent peptoids. To resolve this, we adapted dihedral parameters from literature[112], which were validated by PM6 calculations, for the simulation of peptoids. Consequently, the new model was used to predict the behavior of sarcosine, a model peptoid, in diverse environments – (1) pure water and near model (2) hydrophobic and (3) hydrophilic surfaces. We used classical molecular dynamics simulations with parallel bias metadynamics - using radius of gyration, alpha-beta and distance from surface (in surface simulations) - to comprehensively sample the conformational phase space of both molecules.

In all simulations, sarcosine displayed non-standard, random-coil conformations and a high degree of flexibility. This contrasted with alanine, that displayed more compact structures in solution and near surfaces. The contrast was most visible in their interaction with water where sarcosine interacted more strongly than alanine. This behavior has also been observed in experiments.[138,139] On the other hand, both molecules interacted similarly with model hydrophobic and hydrophilic surfaces. Notably, sarcosine was shown to face a larger entropic penalty than alanine when it interacts with surfaces.

These results may be useful to understand the energetic driving forces that allow some peptoids to form extended structures like self-repairing membranes[101] and nanosheets[108] and others to form threaded-loop structures[144] or helices[145]. Despite the structural heterogeneity displayed by peptoids, Mannige *et al.*[108] proposed that peptoids may share common backbone behavior, namely a zig-zag structure, that falls into the same (ϕ ,ψ) space as our proposed model. Therefore, our results can be used to make generalized predictions about peptoid behavior. Although extensions

to our work are needed to understand the specific effects of other environmental factors, like ions, inorganic interfaces and solvents, on the self-assembly of peptoids.

3.6 ACKNOWLEDGMENTS

MD simulations and manuscript preparation were supported by the MS[3] (Materials Synthesis and Simulation Across Scales) Initiative at Pacific Northwest National Laboratory (PNNL), a multi-program national laboratory operated by Battelle for the U.S. Department of Energy. CJM was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences Division of Chemical Sciences, Geosciences, and Biosciences. MDB was supported by the US Department of Energy, Office of Basic Energy Sciences, Biomolecular Materials Program at PNNL. Computing resources were generously allocated by University of Washington's IT department and PNNL's Institutional Computing program. The authors greatly acknowledge conversations with Dr. Kayla Sprenger, Josh Smith, and Dr. Yeneneh Yimer.

Chapter 4. ESSENTIAL SLOW DEGREES OF FREEDOM IN PROTEIN-SURFACE SIMULATIONS: A METADYNAMICS INVESTIGATION

Arushi Prakash[1+], K. G. Sprenger[1+], Jim Pfaendtner[1,]*

[1]Department of Chemical Engineering, University of Washington, Seattle, Washington 98105

4.1 ABSTRACT

Many proteins exhibit strong binding affinities to surfaces, with binding energies much greater than thermal fluctuations. When modelling these protein-surface systems with classical molecular dynamics (MD) simulations, the large forces that exist at the protein/surface interface generally

confine the system to a single minimum. Exploring the full conformational space of the protein, especially finding other stable structures, becomes prohibitively expensive. Coupling MD simulations with metadynamics (enhanced sampling) has fast become a common method for sampling the adsorption of such proteins. In this paper, we compare three different flavors of metadynamics, specifically well-tempered, parallel-bias and parallel-tempering in the well-tempered ensemble, to exhaustively sample the conformational surface-binding landscape of model peptide GGKGG. We investigate the effect of mobile ions and ion charge, as well as the choice of collective variable (CV), on the binding free energy of the peptide. We make the case for explicitly biasing ions to sample the true binding free energy of biomolecules when the ion concentration is high, and the binding free energies of the solute and ions are similar. We also make the case for choosing CVs that apply bias to all atoms of the solute to speed up calculations and obtain the maximum possible amount of information about the system.

4.2 INTRODUCTION

The free energy is an important thermodynamic quantity in biomolecular systems, containing information about relative stability between states, binding affinities, and solubilities[146]. Several enhanced sampling methods can provide reliable estimates of free energies of specific reaction coordinates or free energy differences between stable states from molecular simulations of the system. Recently, the metadynamics family of methods has gained widespread use in such calculations. It allows users to bias several slow degrees of freedom (or collective variables (CVs)) using a history-dependent, dynamic bias. Small biases (usually Gaussian-shaped hills) are deposited in CV phase space to force the system to visit new states, ultimately converging to the underlying free energy surface.[32] Several flavors of metadynamics have since been developed to overcome challenges related to convergence and scaling.[31] Some variations scale the Gaussian height[23] or width[147] according to its location in phase space to improve convergence. Others

introduce parallel replicas in phase[35] and temperature[127,148] space to increase sampling. And finally, some alter the way bias is deposited – by exchanging between replicas[149] or by depositing many low-dimensional bias potentials in parallel[24]– to tackle problems associated with biasing multiple CVs. (For an exhaustive review of metadynamics, the authors direct the readers to some excellent reviews.[21,31]

Some of these metadynamics methods have been used to calculate free energies of binding of molecules to proteins[150–153] and to surfaces[125,127,154–162], especially where strong binding prohibits exhaustive sampling with plain MD. For example, Hughes *et al.* used well-tempered metadynamics (WTM) to exhaustively compute the binding free energies of all amino acids to graphene[163]. Several studies have used parallel tempering metadynamics in the well-tempered ensemble (PTMetaD-WTE) to characterize binding[127,164,165]. Others have used replica exchange with solute tempering augmented with metadynamics (RESTMetaD) to calculate the binding of peptides to various gold crystal facets[156] and to silica[154,161]. Like other dimension-reducing algorithms, the efficacy of metadynamics is influenced by the choice of CV(s). The abovementioned studies used the center-of-mass distance of the molecule from the surface,[125,127,154,156,165] helicity of the peptide[161], hydrogen bonds[164,165], or the hydrophobicity of the protein core[165] as CVs. These CVs bias either intrinsic coordinates of the solute or the solute-surface distance.

However, commonly used CVs in biomolecular adsorption systems do not explicitly consider environmental variables that may affect binding, such as solvation[166,167], surface ion content[125], and other local order variables[168]. In fact, Wu et al. showed that ions in the local environment can affect the conformation of adsorbed peptides, stabilize different binding residues, and compete with peptide adsorption[169]. Therefore, it is essential to understand and account for the

effect of ions, either from the electrolyte or naturally occurring surface ions in metal oxides like silica, when studying biomolecular adsorption.

To address remaining gaps in knowledge about the best practices for simulating peptide adsorption in relation to charged interfaces and ions in solution, herein we study the interaction of model peptide GGKGG with a silica surface at pH 7.5. Mimicking neutral pH conditions, several surface sites are deprotonated and passivated by sodium ions. We compare the resulting peptide binding free energy profiles of three different variants of metadynamics– (1) well-tempered metadynamics, (2) parallel-tempering metadynamics in the well-tempered ensemble, and (3) parallel-bias metadynamics. We compare the sampling achieved by each of these methods and the results of using two different CVs, namely the center-of-mass and sidechain distance of the peptide from the surface. Finally, we compare peptide-surface binding in the presence of different electrolyte species. From our results, we hope to establish a general protocol that would allow researchers to comprehensively explore conformational phase space and obtain true energetics of inorganic surface-binding peptides. We especially stress thorough exploration of the role of ions in biomolecular adsorption.

4.3 METHODS

SIMULATION SETUP

All simulations, listed in Table 1, were conducted with the GROMACS 5.1.2[117] MD engine. The surface and peptide were modelled with the INTERFACE[170] and CHARMM36[12] force fields, respectively, and the SPC[171] model was used to represent water. The peptide, GGKGG, was generated with the psfgen plugin for Visual Molecular Dynamics (VMD[172]). It was capped with neutral ACE and NME groups to warrant the study of a single binding residue (i.e., positively charged lysine) to silica. To approximate the correct surface ionization state and terminal

chemistry for a silica surface at pH 7.5, the structure provided by Emami *et al.*[173] was used. Each simulation box consisted of a ~ 3.5 x 3.4 x 2.0 nm silica unit. This unit was placed next to 3.5 x 3.4 x 4.6 nm of solution containing a single peptide, water, counterions, and electrolyte (optional) (Table 4-1). For production runs, bulk atoms in the silica structure were frozen to prevent extensive deformation of the structure during the simulations. However, atoms within 0.5 nm of the surface remained free to capture the true interaction of silica with other species.[158] In each simulation, the peptide was placed near the surface in a random starting configuration. One Cl^- ion was added to each simulation box to neutralize the peptide. In some cases, electrolytes in the form of $Ca^{2+}(Cl^-)_2$, Na^+Cl^-, and $(Na^{0.5+})_2Cl^-$ were added to the solution (modelled with the CHARMM36 force field) to model varied experimental conditions. In the case of $(Na^{0.5+})_2Cl^-$, the charges of all sodium ions (including surface ions) were scaled to 0.5+ for consistency, creating a hypothetical micro-environment of like-charged counterions. The system was periodic in x, y, and z dimensions which allowed the peptide, water and ions to interact with both the top and bottom surfaces of silica.

All starting configurations were stabilized using a steepest descent method to minimize the energy of the system. The systems were then equilibrated in the NPT ensemble at 1 bar and 300 K for 1 ns using the Berendsen barostat[50] and Donadio-Bussi-Parrinello thermostat[51]. Finally, production runs were conducted in the NVT ensemble at 300 K, using the same thermostat, for varying simulation times as noted in Table 1. All simulations were performed with a 2 fs time step with bonds between hydrogen and heavy atoms constrained using the LINCS[52] algorithm. Short-range Coulomb and Van der Waals forces were explicitly calculated below a cutoff value of 1 nm. Particle Mesh Ewald (PME)[53] was used for calculating electrostatics with the short-range term cutoff at a value of 1 nm.

BIASING METHODS

Following NPT equilibration, different enhanced sampling schemes were employed to test their relative efficacy in determining the binding free energy of the peptide to the silica surface. Brief descriptions and relevant parameters for the three chosen flavors of metadynamics are given below. Complete protocols for employing these methods can be found elsewhere (WTM,[23] PTMetaD-WTE,[28,174] PBMetaD[24]).

Based on previous sampling procedures for using distance CVs, the initial Gaussian height was set to 2 kJ/mol, the bias factor to 10, and the Gaussian deposition rate to 1 hill/ps for all metadynamics simulations.[125] The σ value was calculated from an unbiased simulation of the system; the peptide-surface distance was tracked with PLUMED[175] and half the standard deviation of the equilibrium fluctuations was used as an estimate.[176] The value used for the simulations was 0.05 nm. During all metadynamics simulations, a harmonic restraint was placed on the distance of the peptide's center-of-mass (COM) from the surface (at ~ 2.2 nm from the top of the simulation box) to promote sampling in the region of interest and use computational time more efficiently. Free energy profiles recovered from the metadynamics simulations were considered converged when they retained their shape over the last 25% of the simulation; while the majority of the simulations converged prior to 1 μs of total sampling time, all simulations were nonetheless carried out to 1 μs for consistency.

WELL-TEMPERED METADYNAMICS (WTM)

Well-tempered metadynamics (WTM) simulations were performed by biasing either the orthogonal distance between the COM of all peptide atoms and the surface, or the distance between only the sidechain (SC) nitrogen atom of lysine and the surface.

A simple reweighting scheme[128] was used to recover the free energy profile of the SC-surface distance from a simulation where the COM was biased. The weight of each configuration was given by

$$e^{\beta(V(s,t)-c(t))} \tag{4.1}$$

where *V(s,t)* is the instantaneous value of the history-dependent bias potential for the system (*s* is the value of the CV at a given time *t*), and $\beta = 1/k_BT$, where k_B is the Boltzmann constant and T is temperature. The reweighting factor, c(t), is a time-dependent offset bias that was calculated on-the-fly using PLUMED[175] via the following equations,

$$c(t) = \frac{1}{\beta} log \frac{\int ds\, e^{-\beta F(s)}}{\int ds\, e^{-\beta(F(s)+V(s,t))}} \tag{4.2}$$

$$\beta F(s) = -\frac{\gamma V(s,t)}{k_B \Delta T} + log \int ds\, e^{\gamma V(s,t)/k_B \Delta T} \tag{4.3}$$

where ΔT is the so-called hills temperature for the WTM simulation, $\gamma = \frac{T+\Delta T}{T}$ is the bias factor, and Δt is the time between energy depositions of Gaussian hills. These weights were binned according to the SC-surface distance and the free energy was then obtained by taking the natural log of the sum of weights in each bin.

PARALLEL TEMPERING METADYNAMICS IN THE WELL-TEMPERED ENSEMBLE (PTMETAD-WTE)

Parallel tempering metadynamics in the well-tempered ensemble (PTMetaD-WTE) was employed to assess the use of temperature for overcoming hidden free energy barriers introduced by the presence of an interface. Six configurationally identical replicas were simulated across a temperature range of 300-450 K, spaced per the algorithm developed by Prakash et al.[177] In a 100 ps NVT parallel tempering simulation, the replicas were equilibrated at their respective temperatures with the Donadio-Bussi-Parrinello thermostat.[51] A 10 ns WTM simulation was then

performed to establish the well-tempered ensemble (WTE), biasing the potential energy of the system with a bias factor of 30. Gaussian hills were added to the potential energy every ps with a σ value of 150 kJ/mol and an initial height of 2.0 kJ/mol. After 10 ns, a constant exchange success probability of 21% was achieved between replicas.

Following the WTE setup, a second metadynamics simulation was performed with an additional one-dimensional bias potential on the orthogonal distance between the COM of all the atoms in the peptide and the silica surface. As in our previous study,[125] the WTE bias continued to be applied to the potential energy, though with a significantly reduced deposition pace of every 5 ps. Reweighting for the free energy as a function of the SC-surface distance was obtained as previously described for WTM simulations.

PARALLEL BIAS METADYNAMICS (PBMETAD)

Recently, PBMetaD was shown to be a flexible and easy method for sampling multi-dimensional free energy surfaces.[149] Instead of depositing a multi-dimensional bias (which scales poorly with the number of CVs), PBMetaD relies on depositing several low-dimensional biases in parallel in multiple CV spaces.

A 6-replica multiple walker PBMetaD simulation was conducted.[24,35] The distances of all ions (independently) and the COM of the peptide from the surface were biased, using the same parameters as previously described.

We used the technique of Tiwary *et a.,l*[128] to make an estimate of the unbiased probability distributions (free energies) of other CVs in PBMetaD (i.e., the SC-surface distance). This resulted in using the same approach described above to calculate weights using the c(t) quantity described above. As a technical note, the current version of PLUMED does not apply the reweighting method within PBMetaD, so we estimated (c(t)) as:

$$e^{\beta c(t)} = \int ds \left[e^{\frac{\gamma V(s,t+\Delta t)}{k\Delta T}} - e^{\frac{\gamma V(s,t)}{k\Delta T}} \right] \quad (4.4)$$

where the various components and variables were described in an earlier section.

Table 4-1 List of metadynamics parameters and collective variables used for different simulations of GGKGG on silica at pH 7.5.

Excess Electrolyte	# of Cations (Water)	Simulation Method	Collective Variable(s)	Total Simulation Time[b]
None[a]	20 (1705)	WTM WTM PBMetaD PTMetaD-WTE	SSD (Protein COM) SSD (Protein SC) SSD (Protein COM/Ions) SSD (Protein COM)	1,000 ns (1W) 1,000 ns (1W) 1,000 ns (3W) 1,000 ns (6W)
$Ca^{2+}(Cl^{-})_2$	45 (1575)	WTM PBMetaD	SSD (Protein COM) SSD (Protein COM/Ions)	1,000 ns (1W) 1,000 ns (6W)
$Na^{+}Cl^{-}$	45 (1600)	WTM PBMetaD	SSD (Protein COM) SSD (Protein COM/Ions)	1,000 ns (1W) 1,000 ns (6W)
$(Na^{0.5+})_2Cl^{-}$	46 (1652)	WTM PBMetaD	SSD (Protein COM) SSD (Protein COM/Ions)	1,000 ns (1W) 1,000 ns (6W)

[a]Refers to silica-neutralizing sodium ions (Na^{+} ions) modeled with the INTERFACE force field.
[b]Values in parentheses are the number of identical walkers (W) or replicas used in each simulation.

4.4 RESULTS AND DISCUSSION

CHOICE OF COLLECTIVE VARIABLE

The distance between the peptide and the surface can be biased by using the distance calculated from (a) the COM (a function of all atoms of the peptide, in our study), or (b) a specific sidechain atom. Here, both biasing schemes are implemented with WTM to compare their convergence speeds and ability to predict features of the peptide binding free energy profile. The binding free energy was calculated as the difference between the Boltzmann-averaged energy of the adsorbed (0.9 to 1.1 nm) and solvated (2 to 3 nm) states. Here, convergence was defined as the

point at which (1) the free energy difference between states stabilized to within 2.5 kJ/mol (k_BT at 300 K) (Figure 1, top right), and (2) minimal changes were observed in the free energy profile of the CV over the last 25% of the simulation. For simulations where the COM was biased (scheme (a)), the free energy profile along the SC-surface distance was calculated by reweighting (described in the Methods section).

We first compare the convergence of the SC and COM biasing schemes with WTM (Figure 4-1; top right). The results show the COM bias converges significantly faster than the SC bias. Indeed, the binding free energy of scheme (a) converges to within k_BT of the final value of ~ -8.6 kJ/mol (averaged across the three methods) nearly 400 ns before scheme (b).

In fact, even after 1 μs of total sampling the free energy profiles obtained from both schemes are markedly different (Figure 4-1; top left). While the binding free energy is identical for both sampling schemes, the minima have different features. Two preferred binding modes of the peptide can be resolved using the COM bias – (A) between z-distances of 0.9 and 1.0 nm, and (B) between 1.0 and 1.1 nm. Representative snapshots from the simulation (Figure 4-1; bottom) show that minimum A comprises structures that are flattened on the surface with multiple peptide/surface interaction points. However, minimum B comprises structures with one binding site only. These features are lost by biasing only the SC-surface distance; a single minimum is resolved where the peptide most commonly binds with one or two residues but overall remains at a distance from the surface (minimum C).

From the differences in the binding free energy profiles, it is evident that the profile obtained from biasing the COM incorporates more information about the adsorption process. It can differentiate between binding poses. Importantly, it incorporates changes to the whole peptide structure since the COM bias is a function of the coordinates of all peptide atoms. On biasing the sidechain atom(s) only, we lose information about the binding behavior of neighboring residues

that might affect the overall surface-bound conformations of the peptide. We hypothesize that these differences might be more appreciable for solutes with multiple binding residues. These findings suggest that metadynamics users should use CVs that apply bias to all atoms of the solute. On converging the free energy profile (and bias potential) of the original CV, the free energy profile of the true CV of interest can then be obtained via reweighting techniques.

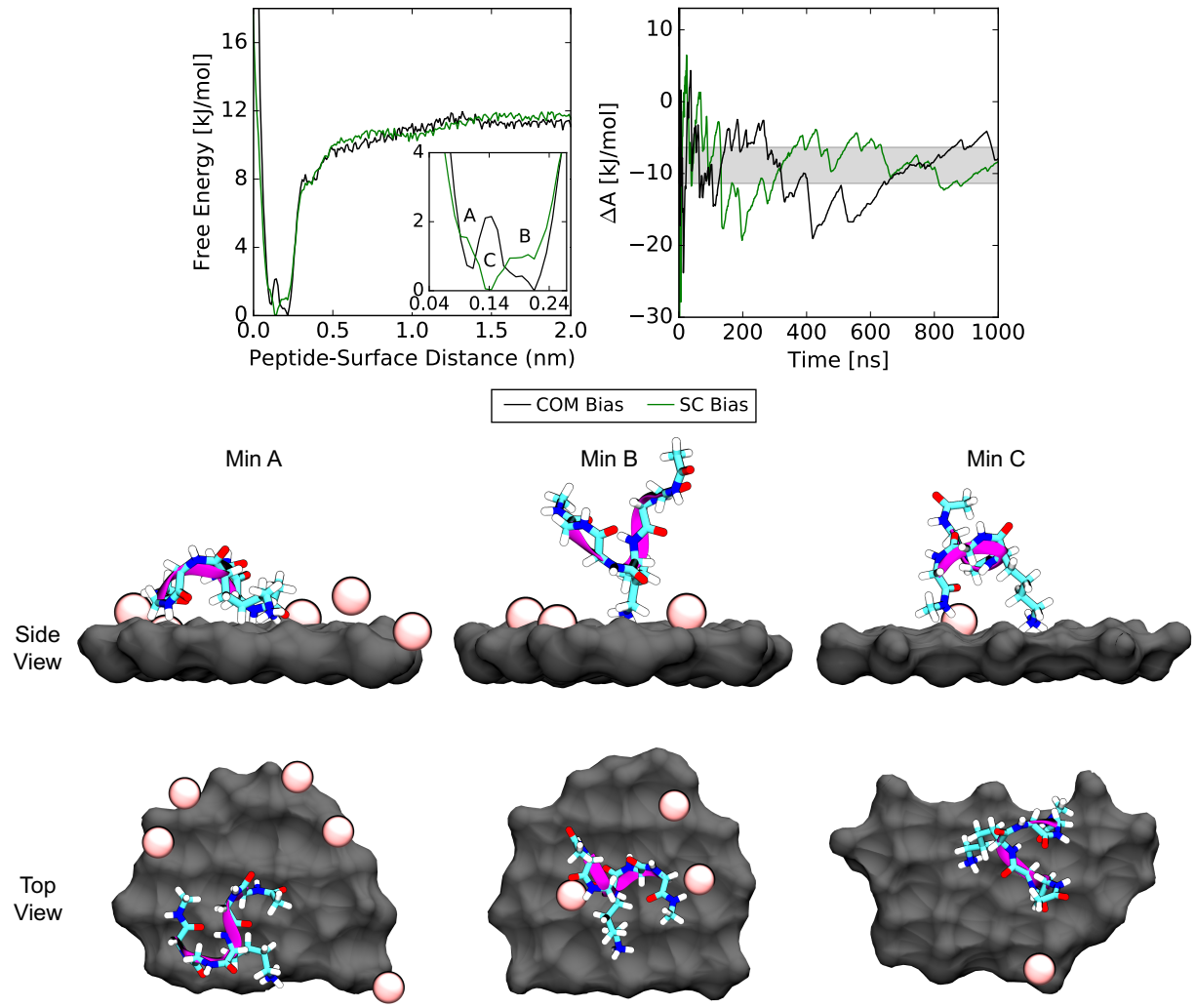


Figure 4-1 (top left) Free energy projected onto the distance between the sidechain (SC) nitrogen atom of lysine and silica for the two biasing schemes: bias added to all atoms in the peptide through the center-of-mass (COM), or bias applied to just the SC nitrogen atom. The black line is the result of reweighting. (top right) Change in Helmholtz energy between the solvated and adsorbed state of GGKGG on silica for the two different biasing schemes. A gray box highlights

the “convergence region” of +/- k_BT of the average of the final ΔA values of the three methods of ~ -8.6 kJ/mol. (bottom) Snapshots from each of the labeled minima in the top left plot. The surface and sodium ions are shown in gray and pink, respectively, and are restricted to within 1 nm of the peptide. The peptide backbone is shown in purple, and carbon, hydrogen, oxygen, and nitrogen atoms are shown in cyan, white, red, and blue, respectively. Water is not pictured for clarity.

COMPARISON OF SAMPLING SCHEMES

From the above simulations, it was concluded that applying bias to all atoms in a biomolecule results in faster calculations and provides greater information about the system. Another key issue that metadynamics users face is selecting the appropriate biasing scheme to calculate a variable of interest. For the calculation of peptide binding free energy, we compare three biasing methods - (1) WTM, (2) PTMetaD-WTE, and (3) PBMetaD. WTM is a relatively simple metadynamics sampling scheme where applying bias to more than 2 or 3 CVs becomes computationally expensive. In the PTMetaD-WTE framework, sampling is enhanced by a parallel tempering scheme that allows the system to sample several temperatures (by biasing the potential energy). However, on increasing the number of CVs beyond 2 or 3, this method also becomes computationally prohibitive as the scheme for building the bias potential is identical to WTM. In contrast, in the PBMetaD biasing scheme low-dimensional biases are deposited in parallel, instead of higher dimensional biases, making computations more efficient and allowing for the possibility of biasing many CVs. Thus, with PBMetaD the distance of each ion from the surface can be biased in addition to the peptide-surface distance. The free energy profile recovered from PBMetaD simulations is considered to be the most accurate among the three methods since it accounts for the most degrees of freedom in the system. The peptide COM distance from the surface is biased in all schemes, and the SC distance is reweighted for as described earlier.

The results (Figure 4-2) show that within the resolution of the calculations (i.e., $\sigma = 0.05$ nm and $k_BT = 2.5$ kJ/mol), PBMetaD and WTM recover nearly identical free energy profiles with

two free energy minima (A and B) resolved for the adsorbed state. These minima correspond to the minima and snapshots shown in Figure 1 for the WTM simulation.

However, the PTMetaD-WTE simulation does not recover the minimum closest to the surface, minimum A. We hypothesize that this discrepancy is either due to: 1) the behavior of ions at high temperature, since PTMetaD-WTE allows the system to visit temperatures from 300-450 K, or 2) incomplete sampling. In regards to the former hypothesis, it has been predicted that the binding affinity of sodium ions to a hydrophilic surface increases with increasing temperature.[125] Therefore, we expect most replicas at high temperatures to sample configurations with ions bound to the surface, rather than free in solution. As noted before, PTMetaD-WTE relies on the exchange of these high temperature configurations so that they eventually reach the lowest-temperature replica at 300 K. However, at 300 K, these ions are tightly bound beyond k_BT (Figure 4-3A), and the impact of the temperature alone on sampling the ion degrees of freedom is unclear. [125] Thus, it is possible that the 300 K trajectory primarily samples a surface with charged sites predominantly bound to ions and the peptide cannot adsorb at the free energy minimum located 0.1 nm from the surface (minimum A). Essentially, we propose that the temperature benefits of the enhanced sampling algorithm are reduced in this case. Examining the behavior of the MetaD transient throughout the PBMetaD simulation (results not shown), does show that the location of the minima fluctuates modestly in terms of location and energetics, albeit within the standard deviation of the CV (0.1 nm) and thermal noise (2.5 kJ/mol). In light of these results, we set out to explore the role of ions, ion binding, and the capability of a sampling algorithm to effectively tackle various ion-specific effects.

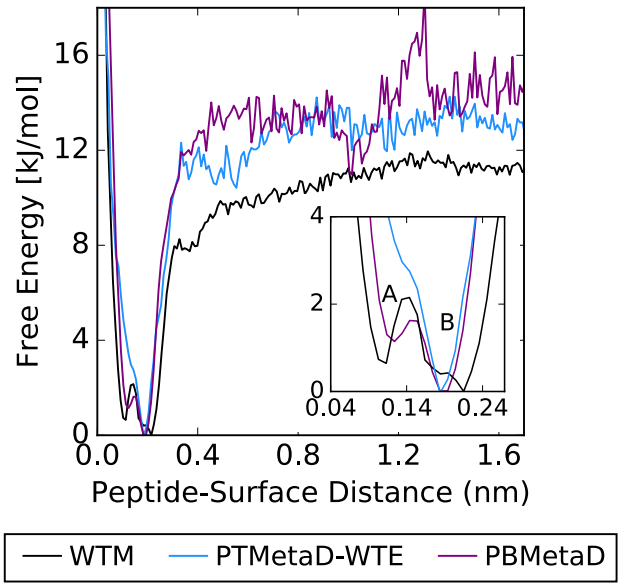


Figure 4-2 Free energy (kJ/mol) along the distance of the nitrogen atom of the lysine sidechain from the surface using enhanced sampling method WTM (black), PTMetaD-WTE (blue), and PBMetaD (purple).

ION-SURFACE DISTANCE AS A COLLECTIVE VARIABLE

The simulations discussed above were carried out at low ionic strengths. However, many proteins adsorb in conditions of high ionic strength where ions can promote aggregation or "salting-out" of proteins[178] or screening of protein-protein interactions or protein-surface interactions.[179] To understand the influence of ions on the adsorption behavior of our model peptide, additional metadynamics simulations were performed with different electrolytic conditions (i.e., monovalent and divalent cations). Simulations were carried out with additional cations (25 Ca^{2+}, 25 Na^{+}, or 26 $Na^{0.5+}$) and with requisite numbers of anions (Cl^{-}) to maintain system charge neutrality (Table 1). In all simulations, the added electrolytes were initially evenly distributed in the solution and allowed to adsorb spontaneously onto the surface. Both WTM and PBMetaD simulations were performed in each case, and the results (Figure 4-3) are discussed in detail below.

PROTEIN ADSORPTION WITHOUT EXCESS IONS

Without excess ions in the solution, the previous simulations represented the limit of a dilute electrolytic solution in contact with a silica surface. In those cases, the WTM and PBMetaD peptide binding free energies were shown to be identical (Figure 4-2, Figure 4-3A). Additionally, Figure 3A shows the binding free energy of the sodium ions, calculated from the PBMetaD simulation, was equal to that of the peptide. Thus, it is likely the ions competed with the peptides for binding sites; however, in the limit of a dilute electrolytic solution, 1 μs of total simulation time was found to be more than enough to converge to the correct peptide binding free energy profile with either method.

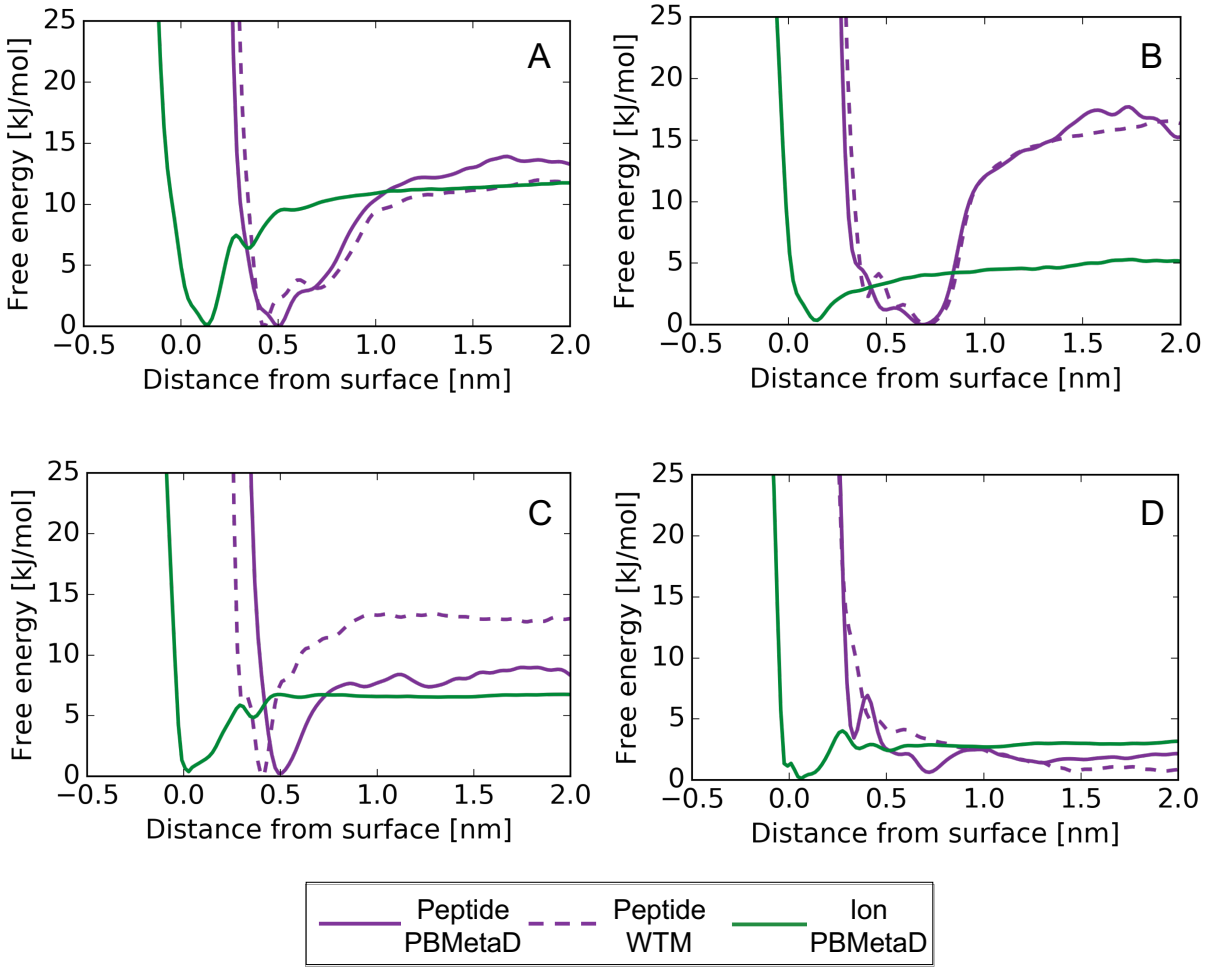


Figure 4-3 Free energy projected onto the COM-surface distance with: (A) no electrolyte, (B) excess $Na^{0.5+}$ ions, (C) excess Na^{+} ions, and (D) excess Ca^{2+} ions. Green lines indicate thermally-averaged ion binding profiles from PBMetaD simulations. Solid and dotted purple lines indicate peptide binding profiles from PBMetaD and WTM simulations, respectively.

EFFECT OF EXCESS $NA^{0.5+}$ IONS ON PEPTIDE ADSORPTION

On investigating peptide adsorption in the presence of excess $Na^{0.5+}$ ions from the PBMetaD simulation (Figure 4-3B), it is evident that the peptide binds much stronger to the surface (~ 20 kJ/mol) compared to the ions (~ 5 kJ/mol). This is a departure from the binding free energy predicted in previous simulations that did not have excess ions (~ 13 kJ/mol for both peptide and ions) but had the requisite monovalent sodium ions to neutralize the surface. When $Na^{0.5+}$ ions adsorb to the surface, they cannot completely compensate for the -1 charge of the surface sites.

Thus, an ion adsorbed on a surface site fails to completely screen the charge of the surface, promoting other solutes that can fully compensate for the charge, like the +1-charged lysine sidechain, to adsorb in its place. Additionally, the difference in charge (+0.5 versus +1) provides the lysine sidechain with a greater electrostatic driving force to adsorb on a surface site. These factors lead to an increase in the binding free energy of the peptide in the presence of $Na^{0.5+}$ ions compared to in the presence of Na^{+} ions.

Notably, both WTM and PBMetaD simulations again converge to the same free energy profile for the peptide SC-surface distance, even though ions were not explicitly biased in the WTM simulation. This results from the higher binding free energy of the peptide to the surface compared to the ions. Thus, a WTM simulation biasing a single CV to study peptide adsorption does not face sampling limitations from surface-bound ions. From these findings, we conclude that more intensive sampling methods like PBMetaD and PTMetaD-WTE are not required in this case to calculate the binding free energy of the peptide.

EFFECT OF EXCESS CA^{2+} IONS ON PEPTIDE ADSORPTION

With 25 additional Ca^{2+} ions and 50 Cl^{-} ions added to neutralize the system, this represents an electrolyte of high ionic strength with cations that have twice the charge of the peptide's lysine sidechain. The results of the PBMetaD simulation (Figure 4-3D) show that a driving force (~ 4 kJ/mol) exists for dissolved Ca^{2+} to adsorb to the surface. Conversely, both WTM and PBMetaD simulations predict that as the peptide approaches the surface, the peptide is repelled from it and thus prefers to remain ~0.7 nm away from the surface. Since Ca^{2+} ions preferentially bind to the surface adsorption sites and more than compensates for the surface charge of each site, the protein effectively feels a positive surface charge and is repelled. Visual analysis of the trajectory in VMD[172] confirms this; whenever the protein approaches the surface to bind (mainly driven by the influence of the bias potential), it is promptly repelled and remains in solution (Figure 4-4). We

note that while a small free energy minimum does appear for the PBMetaD simulation at a peptide-surface distance of ~0.4 nm, the peptide must cross a free energy barrier to fall into this minimum, presumably to displace the bound Ca^{2+}. We recommend future work to investigate this effect, perhaps performing simulations to explore the surface charge effects (e.g., number of charge sites and magnitude of charges) as well as extending the study to peptides with multiple charged side chains to promote different binding mechanisms and thermodynamics.

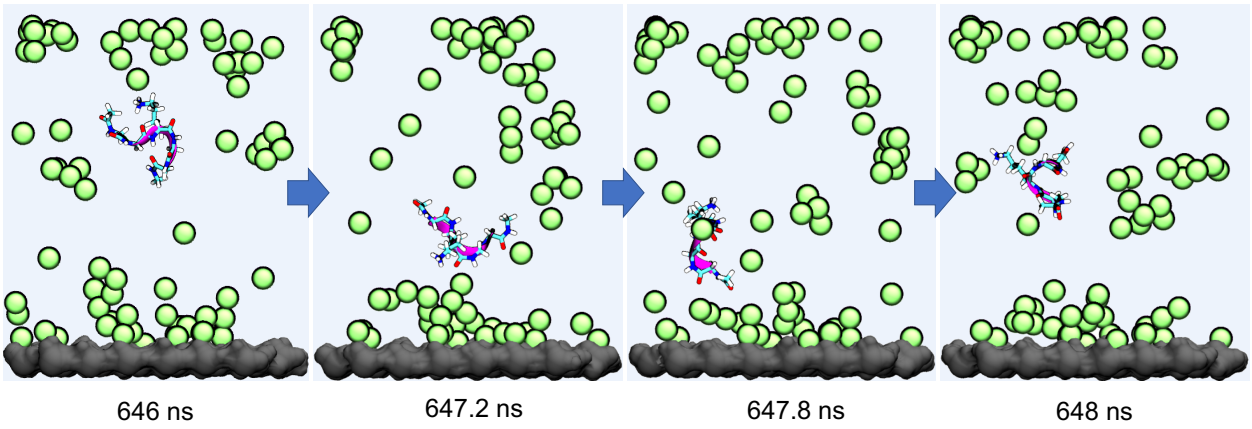


Figure 4-4 Snapshots from a WTM simulation with added electrolyte $Ca^{2+}(Cl^-)_2$. The surface and peptide are colored as previously described (Figure 1). Ca^{2+} ions are shown in green. Explicit water molecules and Cl^- ions are not pictured for clarity.

EFFECT OF EXCESS NA^+ IONS ON PEPTIDE ADSORPTION

Finally, in the case where the electrolyte contains excess Na^+ ions, a difference of ~ $2k_BT$ is observed in the peptide binding free energies of the WTM and PBMetaD simulations (Figure 4-3C). This is noteworthy, given that WTM and PBMetaD converged to the same peptide binding free energy in the absence of excess Na^+ ions (Figure 4-3A). In both cases (i.e., with or without excess Na^+ ions), the binding free energy of the peptide and ions are predicted to be the same with PBMetaD.

The excess Na^+ ions provide a crowded microenvironment where the peptide diffuses slowly in solution and ions quickly replace other ions on the surface. Consequently, when an ion

leaves a binding site, it is easier for another ion in the vicinity to take its place. Once an ion is bound to a surface site, it is a "rare event" for it to desorb spontaneously (in a 300 K simulation) and allow the peptide to adsorb in its place. Thus, in situations where an electrolyte and biomolecule have similar surface affinities, WTM simulations, which only provide energy to the degrees of freedom of the solute, are severely limited by the availability of surface binding sites when the peptide approaches the surface.

On the other hand, in PBMetaD bias is also provided to the ions to adsorb/desorb multiple times during the simulation. This leads to many opportunities for the peptide to adsorb to free surface sites and thus for the system to collect information about solute binding/unbinding. This is also evident in Figure 4-5, where we track the position of the lysine sidechain with respect to the surface. Although the WTM simulation (Figure 4-5; left) samples a few binding events, the peptide spends most of the time in solution. It draws near the surface several times, but likely due to inaccessible surface sites, quickly moves back into solution. In the PBMetaD simulation (Figure 4-5; right), the peptide spends time on the surface and in solution almost equally. Additionally, most approaches of the peptide to the surface result in a successful binding event. Thus, the use of PBMetaD results in many more binding/unbinding events collected during the simulation and converges much more rapidly than WTM. We hypothesize that the WTM simulation would require several more microseconds to sample enough peptide binding/unbinding events to converge to the same free energy profile as PBMetaD.

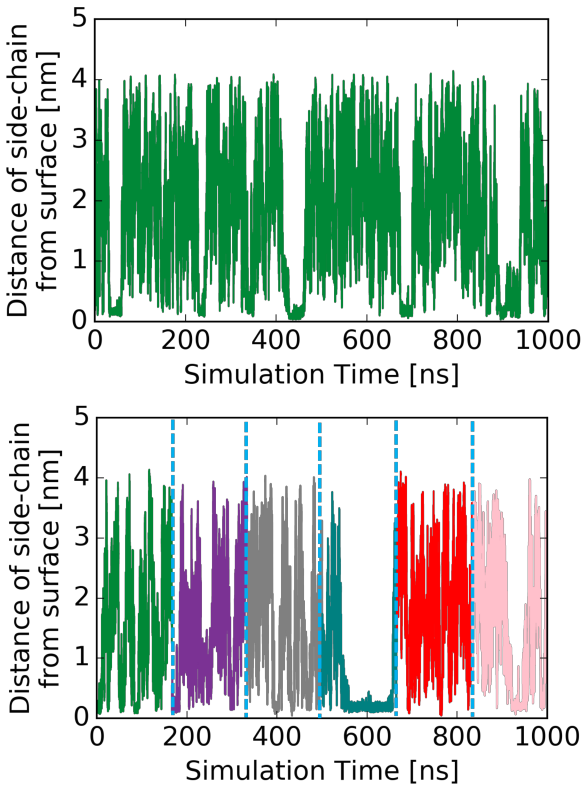


Figure 4-5 Distance of the sidechain nitrogen of lysine from the surface, as sampled by (left) WTM and (top) PBMetaD (bottom). Each walker in the PBMetaD simulation is colored separately, with the x-axis denoting cumulative simulation time. Each simulation starts from independent starting structures and evolves sharing the same bias potential for 166.67 ns each.

4.5 CONCLUSIONS

Three different metadynamics methods, namely WTM, PTMetaD-WTE, and PBMetaD, were employed to simulate the adsorption of model peptide GGKGG to silica. These methods were evaluated for their ability to converge to the "correct" peptide binding free energy profile in the presence of various electrolytic solutions. Here, PBMetaD was expected to provide the most accurate results due to its capacity to bias both solute (peptide) and solvent (ion) degrees of freedom. It was determined that competing slow degrees of freedom from solvent constituents can, in some cases, have a large impact on the resulting binding free energy of the solute. Thus, it is important to consider how surface-bound and electrolytic ions affect biomolecular adsorption.

Indeed, there are several cases where ions dictate how proteins interact with surfaces, for example in the polyamine-directed precipitation of silica nanospheres that only occurs in the presence of multivalent anions such as phosphate.[180]

Based on our findings, we recommend the following biasing protocols to recover "correct" binding free energy profiles when simulating biomolecular adsorption processes:

1) Dilute solutions – A simple metadynamics sampling scheme like WTM can be used. While competition for surface sites may exist between the solute and ions, the dilute electrolytic conditions allow for many solute binding/unbinding events to occur throughout the simulation. Thus, convergence to the correct solute binding free energy profile can be achieved in a shorter simulation timescale (within 1 μs).
2) Excess, competing electrolyte - It is necessary to use a method like PBMetaD that can bias both the solute and ion degrees of freedom. If the binding affinities of the constituents are not known *a priori*, it can be roughly assumed that ions with a charge equal to the charge of binding sidechain residues will have a similar binding affinity to the surface as the solute. We strongly emphasize the need to explicitly bias the ion degrees of freedom in these situations; it was determined that a method like PTMetaD-WTE that only indirectly biases the solvent degrees of freedom may result in an incorrect binding free energy profile of the solute.
3) Excess, weak-binding electrolyte - It is sufficient to use a sampling scheme like WTM. The protein sidechain can displace the surface ions upon adsorption, allowing for adequate sampling of binding/unbinding events to converge to the correct binding free energy profile of the solute.

4) Excess, strong-binding electrolyte – Since the ions bind considerably stronger, it leads to extremely long sampling times required to obtain the correct solute binding free energy. Thus, we recommend using a method like PBMetaD to speed up the time to convergence, and a larger bias factor for CVs.
5) Choice of solute CV - Conformational CVs should be chosen that apply bias to all atoms of the solute to obtain quicker convergence and maximum information about the solute conformation. On converging this free energy profile, the free energy along a different reaction coordinate can be calculated through simple reweighting procedures.

4.6 ACKNOWLEDGEMENTS

JP and KGS acknowledge the support of NSF award CBET-1264459. This work was facilitated using computational, storage, and networking infrastructure provided by the Hyak supercomputer system, supported in part by the University of Washington and the UW Student Technology Fee Proposal program (Award 2015-028).

Chapter 5. INVESTIGATING THE ROLE OF PHOSPHORYLATION IN THE BINDING OF SILAFFIN PEPTIDE R5 TO SILICA WITH MOLECULAR DYNAMICS SIMULATIONS

K. G. Sprenger[1+], Arushi Prakash[1+], Gary Drobny[2], Jim Pfaendtner[1,]*

[1]Department of Chemical Engineering, University of Washington, Seattle, Washington 98105

[2]Department of Chemistry, University of Washington, Seattle, Washington 98195

5.1 ABSTRACT

Biomimetic silica formation, a process that is largely driven by proteins, has garnered considerable interest in recent years due to its role in the development of new biotechnologies. However, much remains unknown of the molecular-scale mechanisms underlying the binding of proteins to biomineral surfaces such as silica, or even of the key residue-level interactions between such proteins and surfaces. In this study, we employ molecular dynamics (MD) simulations to study the binding of R5 – a 19-residue segment of a native silaffin peptide used for in vitro silica formation – to a silica surface. The metadynamics enhanced sampling method is used to converge the binding behavior of R5 on silica at both neutral (pH 7.5) and acidic (pH 5) conditions. The results show fundamental differences in the mechanism of binding between the two cases, providing unique insight into the pH-dependent ability of R5 and native silaffin to precipitate silica. We also study the effect of phosphorylation of serine residues in R5 on both the binding free energy to silica and the interfacial conformation of the peptide. Results indicate phosphorylation drastically decreases

the binding free energy and changes the structure of silica-adsorbed R5 through the introduction of charge and steric repulsion. New mechanistic insights from this work could inform rational design of new biomaterials and biotechnologies.

5.2 INTRODUCTION

Biosilicification is a complex process by which diatoms and sponges grow ornate silica exoskeletons. Biosilicification likely occurs through the self-assembly of proteins into a matrix, followed by the condensation of silicic acid to form silica nanostructures.[181] This process has been used to make silica nanostructures *in vitro* using not only silaffin proteins but long-chain polyamines, and specific silaffin domains.[181,182] In fact, the capacity of the 19-mer segment of silaffin named R5 (SSKKSGSYSGSKGSKRRIL) to precipitate silica and form nanostructures is well-documented.[181,183–186] It is highly desirable to obtain complete mechanistic control over this process for low-energy synthesis routes to produce nanomaterials of controlled morphology. To achieve this, it is important to determine the key interactions and driving forces governing biosilicification.

Several experimental studies have explored the conditions necessary for biosilicification to occur. Analysis of the silaffin sequence shows that posttranslational modifications (PTMs) - namely the modification of lysine to long-chain polyamines, and serine and trimethyl-hydroxyl-lysine residues to phosphorylated residues - allow native silaffins to precipitate silica.[180,182,183,187–189] Indeed, phosphorylation was found necessary for any biosilicification activity to occur since the negatively-charged phosphate groups countered the electrostatic repulsion between positively-charged amine groups from neighboring R5 peptides, permitting them to self-assemble.[190] However, complete phosphorylation of serine residues can drastically decrease the tendency of these peptides to self-assemble.[190] Notably, the effect of phosphorylating the sequence of the peptides can be mimicked by the addition of negatively-charged phosphate ions in the silicification

environment, which likely promotes self-assembly by bridging the positively-charged silaffin peptides together.[182,191,192] Other environmental variables, like pH, also affect biosilicification. Most strikingly, the pH for silica precipitation activity is different for synthetic R5 (pH 7) and native silaffin (pH 4-5). PTMs can also affect the pH sensitivity of R5.[193]

While these studies are primarily experimental, some computational investigations of the system have also been conducted. Some have investigated the large-scale assembly of peptides into matrices for silica condensation. For example, Lenoci *et al.* modelled silaffin peptides as coarse-grained beads that self-assembled via Brownian dynamics into a continuous matrix with 15-25 nm cavities – at which size silica would likely assemble into nanostructures – or seeds of proteins which would allow for the growth of silica.[192] Eby *et al.* also used coarse-grained models of peptides, in addition to phosphate ions, and observed the formation of a scaffold-like structure consisting of a mixture of peptides and phosphate ions.[194] However, these studies were agnostic about the atomic structure of the peptides in their models. Recently, Lutz *et al.* tried to decipher the molecular structure of R5 near silica aggregates at an air/water interface through fully atomistic molecular dynamics (MD) simulations.[195] They found that R5 loses its defined secondary structure upon interaction with extended silica sheets, and instead forms aggregates that interact with silica particles primarily through their C-terminal motifs (RRIL).[51] Despite these advances in obtaining a molecular understanding of biosilicification through simulations, significant questions remain. Specifically, the atomic level interactions of R5/silaffin peptides with ions, and other peptides, are not well-characterized. Also, it is unclear how peptide-surface, peptide-peptide, and peptide-ion interactions individually affect the different stages of biosilicification.

Herein, we aim to provide new, high-resolution insights into the effect of PTMs of R5 and of pH on peptide/silica binding through MD simulations combined with the metadynamics enhanced sampling method. We simulated a single R5 peptide (and its mutants) on silica at pH 7.5

and pH 5. R5 mutants were designed with varying degrees of serine phosphorylation in the R5 sequence (at both neutral and acidic conditions) following the experimental study of Lechner et al.[190] In each case, the binding free energy to silica and the conformation of R5 at the interface was obtained. We have simulated single-molecule systems, like other peptide-surface simulation studies,[156,160,196] to serve two purposes. First, exhaustive sampling of peptide aggregation at a solid/liquid surface is currently computationally prohibitive for the range of effects we consider in this work. Second, it highlights and isolates the effect of environmental conditions on peptide-surface interactions without the added complexity of multi-peptide interactions. These results can be interpreted as the baseline for a bottom-up understanding of the biosilicification process that seeks to resolve the dominant driving forces that contribute to differences in silica precipitation and morphology under varying conditions, at the molecular scale.

5.3 METHODS

The structure of R5 (SSKKSGSYSGSKGSKRRIL) capped with a neutral acetyl group on the N-terminus was constructed with Visual Molecular Dynamics (VMD[172]) and modeled with the CHARMM36[12] force field. Two additional R5 sequences were constructed: "local phosphoserine (pS) R5" in which S14 was phosphorylated (SSKKSGSYSGSKG*pS*KRRIL), and "global pS R5" in which every serine residue was replaced by a pS residue (*pSpS*KK*pS*G*pS*Y*pS*G*pS*KG*pS*KRRIL). At pH 7.5, the pS residues were parameterized to have a total charge of -2, such that overall charges were +6, +4, and -8 for the unphosphorylated ("no pS"), local pS, and global pS R5 peptides, respectively. At pH 5 (pS charge -1), the peptides had overall charges of +6, +5, and -1 for the no pS, local pS, and global pS R5 peptides, respectively. Charges from Steinbrecher *et al.*[10] were used for pS residues at pH 7.5. Charges for pS residues at pH 5 were obtained via the electronic structure program Gaussian,[122] with calculations performed at the Hartree-Fock (HF) level of theory with the 6-31G(d)//6-31G(d) basis set. Atomic point

charges were assigned with Antechamber[114] and the RESP[113] method. The final atomic point charges used for all pS residues are tabulated in the SI. For all pS residues, LJ and bond parameters were obtained from CHARMM36 while angle parameters were obtained from Steinbrecher *et al.*[10] The SPC/E[171] and INTERFACE[170] force fields were used to model water and silica surfaces, respectively. The silica surface at pH 7.5 was obtained from the repository provided by Emami e*t al.*[173] and represents a crystalline (100) quartz surface (or silica nanostructures >200 nm) with the surface ionization representing a surface stabilized at 0.1-0.3 M ionic strength. This surface was modified by substituting SiO-Na terminations with SiO-H terminations to construct the surface at pH 5. Further details about the structure of the silica surface are contained in the SI. The middle ~2.5 Å of each silica surface was frozen to prevent deformation during the simulations and to provide a stationary reference for biasing the peptide-surface separation distance as in our prior work.[127]

Six systems (2 pH values x 3 peptide sequences) were assembled with 11 nm of water and one R5 peptide placed in a random initial conformation above an approximately 6.9 x 6.9 x 1.9 nm silica surface, resulting in ~58,000 atoms in each case (see Appendix). The surface was constructed assuming a solution ionic strength of 0.1-0.3 M which mimic conditions at physiological pH.[173] Additional Na^+ or Cl^- ions (maximum 8 ions) were added to each system to achieve overall charge neutrality. However, given the number of waters in the system (~16900), the ions should have a negligible effect on the ionization state on the surface. The additional Na^+ ions (in global pS R5 at pH 5 and 7.5) were modeled with CHARMM36 LJ parameters, unlike surface Na^+ ions that were modeled with INTERFACE LJ parameters. More details about these setups are tabulated in the SI (Table S3) together with an analysis of the interactions between R5 and the two different "types" of Na^+ ions our simulations, the results of which imply the presence of both types of ions in our simulations likely had negligible effects on our results. Apart from the

above-mentioned simulations, we performed two additional simulations at pH 5 with local and global pS R5 exploring the effect of the sidechain charge state in phosphoserine (e.g., -1 vs -2), given that phosphoserine has an estimated pKa of 5.6.[197]

All simulations were conducted using the GROMACS 5.1.2[117] MD simulation engine and PLUMED[33] plugin. A steepest descent energy minimization was performed on all systems. Thereafter, all simulations utilized a 2 fs time step by constraining the bonds between hydrogen and other heavy atoms with the LINCS[52] algorithm. Electrostatic interactions were calculated with the particle mesh Ewald (PME[53]) summation method and a cutoff value of 1.0 nm. A van der Waals cutoff value of 1.0 nm was used along with the Donadio-Bussi-Parrinello thermostat[51] in all simulations.

The enhanced sampling method PTMetaD-WTE[127,174] was used to fully characterize the configurational ensemble of the surface-bound peptides, which is essential in properly simulating these types of complex interfacial systems.[198] First, 24 replicas spanning temperatures of 300 to 450 K[177] were equilibrated in 100 ps parallel tempering (PT) simulations in the NVT ensemble. Then, the systems were sampled in the well-tempered ensemble (WTE) by biasing the potential energy of each system. The σ value (Gaussian hill width) for each system was calculated by dividing in half the equilibrium fluctuations of the potential energy at 300 K (σ values ranged between 310 and 490 kJ/mol, depending on the pH and degree of phosphorylation). Other metadynamics (MetaD) parameters – the bias factor (=10), bias deposition pace (=1 hill/ps), and initial hill height (w_0=2.0 kJ/mol) – were constant across all simulations. WTE simulations were deemed converged within 10 ns, resulting in a constant exchange success probability of 34-38% between replicas. Following convergence of the WTE simulations, production MetaD simulations were performed biasing the following collective variables (CVs): (1) the radius of gyration of the peptides' Cα atoms, and (2) the orthogonal distance between the surface and the peptides' centers-

of-mass (COMs, calculated using all atoms). The bias factor, hill deposition pace, and initial hill height mentioned above were used. A σ value of 0.1 nm was used for both CVs in all production simulations. As in our previous work,[125] the potential energy was biased with a reduced pace (= 1 hill/5 ps) in the production runs. Additionally, a half harmonic restraint (i.e., a 'wall') was placed on the distance CV that began acting on the peptides at a surface separation distance of 8.0 nm to limit sampling to one of the two surfaces.

Production simulations were run until convergence was reached, which was considered as the point at which: (1) the free energy difference between the adsorbed and solvated states showed < kT change with simulation time (Figure S3), and (2) the profile of the free energy projected onto the peptide-surface separation distance also showed negligible change with simulation time (data not shown). The systems were considered converged within 250 ns/replica (6 μs total sampling) for simulations of unphosphorylated R5, and within 125 ns/replica (3 μs total sampling) for simulations of phosphorylated R5 peptides. Nonetheless, all simulations were carried out to either 150 or 250 ns/replica depending on the degree of peptide phosphorylation, for consistency. We do note that it is possible that the free energy surfaces could change after additional simulation time (on the order of μs); however, for the timescales simulated herein, these criteria were found to be sufficient for convergence. Additionally, a block averaging analysis was used to calculate the error in the free energy as a function of the CVs biased in the simulations. The results, displayed in Figure S4, show that in each case the error was within 1-2 kJ/mol, below the level of thermal fluctuations in the systems (~2.5 kJ/mol at 300 K).

The trajectories from the production runs were visualized using VMD, and the structure of R5 (and its variants) were clustered using a GOMACS tool. For clustering, the gromos[199] method was used with an RMSD cutoff of 0.10 nm between the radii of gyration of the peptide at each time step. An RMSD cut-off of 0.10 nm was chosen si

nce the radius of gyration free energy profile (see Appendix) showed several minima 0.15- 0.20 nm apart and a small cut-off was needed to differentiate between those structures. In the gromos clustering method, the data point with the largest number of neighbors (defined by a cut-off) is considered the first cluster medoid. Then, this data point and its neighbors are removed and the algorithm is repeated until all structures have been assigned to a cluster.[200] Thus, each frame is assigned a cluster number.

The unbiased cluster probabilities (denoted by the cluster number above) are obtained by the reweighting method described by Prakash *et al.*[25,34] The weight of each frame is given by:

$$e^{\beta(V(s,t)-c(t))} \quad (1)$$

where $V(s,t)$ is the instantaneous value of the history-dependent bias potential for the system (s is the value of the CV at a given time t), and $\beta = 1/k_BT$, where k_B is the Boltzmann constant and T is temperature. The reweighting factor, c(t), is a time-dependent offset bias that was calculated on-the-fly using PLUMED. Structures that fall within the particular region of 2D phase space that we are interested in are reweighted and the unbiased probability of the occurrence of a structure in that region of phase space is obtained. We note that the highest-probability structures and the probabilities associated with these structures are dependent on the clustering method (e.g., gromos clustering, 0.1 nm cut-off), the number of structures visited by the trajectory, and the convergence of the simulation.

5.4 RESULTS

EFFECT OF PHOSPHORYLATION ON R5/SILICA BINDING THERMODYNAMICS AND INTERFACIAL PEPTIDE STRUCTURE

During our PTMetaD-WTE simulations, different configurations of R5 and its variants were explored with respect to their distance from the surface and their radius of gyration. Figure 1 shows the free energy profile along the distance from the surface at pH 7.5. The binding free energy

of the peptide to the surface was obtained by subtracting the (thermal) average energy of the bulk region (~ 4-5 nm) from the (thermal) average energy of the minima (~ 0-1 nm). The binding free energy decreases from approximately -80 kJ/mol for R5 with no pS residues, to -55 kJ/mol for local pS R5, to -30 kJ/mol for global pS R5. Since the decrease in binding free energy is constant (= -25 kJ/mol) when the number of pS residues is increased from 0 to 1 and from 1 to 7, the change in binding free energy is clearly not proportional to the number of pS residues in the sequence of R5. Adding a single pS residue has a dramatic effect on the adsorption thermodynamics, a -25 kJ/mol decrease, as previously mentioned. However, adding additional pS residues does not cause an additional -25 kJ/mol decrease per pS residue, suggesting that the effect (on silica binding strength) of a pS modification is due not only to chemical effects in the sidechain, but also sequence effects (i.e., the neighboring residues matter).

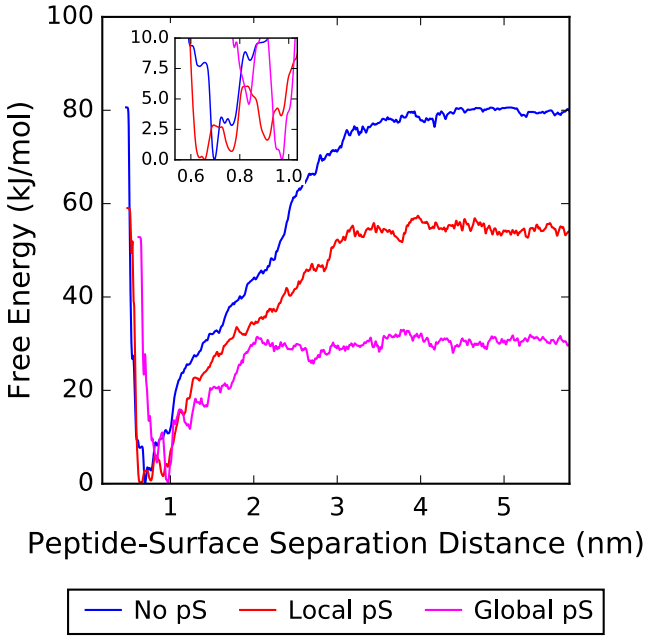


Figure 5-1 Free energy projected onto the distance between the silica surface at pH 7.5 and the center-of-mass of (blue) R5 without phosphorylation (“No pS”), (red) locally phosphorylated R5 (“Local pS”), and (purple) globally phosphorylated R5 (“Global pS”). A zoomed-in view of the energy minima is shown in the inset.

Apart from altering the binding free energy, the presence of pS residues also changes the preferred structure of R5 at the surface. Figure 5-1 (inset) highlights preferred locations of the peptides near the surface. The results show that R5, without phosphorylation, prefers being close to the surface at ~ 0.68 nm. With a single pS residue, R5 is still able to remain near the surface (minima at ~ 0.63 and 0.77 nm from the surface). However, the minimum further from the surface (at ~ 0.92 nm) grows larger. In contrast, global pS R5 sits away from the surface at ~ 0.88 nm. Thus, on increasing the degree of phosphorylation, R5 prefers to be further from the surface, resulting in decreased interactions with the surface and consequently decreases in binding free energy. On the other hand, the preferred radius of gyration of the peptides decreases with increasing degree of phosphorylation (Figure 5-2). Figure 5-2 shows the free energy of R5 along the two biased variables - the radius of gyration and distance from the surface.

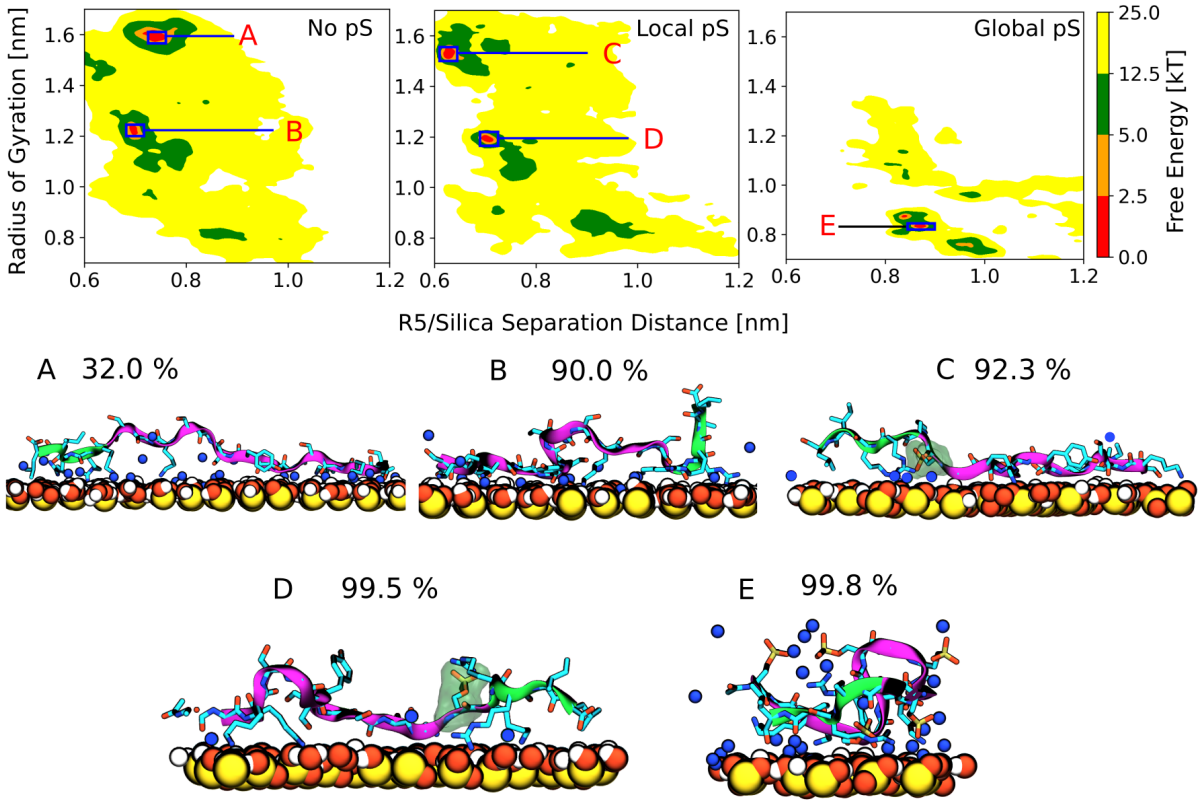


Figure 5-2 (Top) 2D plot of the free energy projected onto the radius of gyration (Cα atoms) and the distance between the silica surface at pH 7.5 and the center-of-mass of R5 without phosphorylation (left: "No pS"), locally phosphorylated (middle: "Local pS"), and globally phosphorylated (right: "Global pS"). The x-axis is cut off to highlight the surface-bound states only. (Bottom) Conformations of the peptides with the highest probabilities from the clustering analysis of the simulations at pH 7.5 in the regions of the above-shown minima, along with the corresponding cluster weights from reweighting. The backbone of the peptides is shown in purple. Cyan, red, yellow, blue, gold, and white coloring correspond to carbon, oxygen, silicon, nitrogen, phosphorus, and hydrogen atoms (peptide hydrogen atoms are not pictured for clarity). The surface is restricted to be within 1 nm of the peptide, and sodium ions are shown in blue. Water is not pictured for clarity. The RRIL motif of the peptides is shown as a green ribbon, and to help identify the lone pS residue in local pS R5, pS is surrounded by a transparent green shell.

To understand the preferred surface separation distances and radii of gyration of R5 and its variants, the trajectories were clustered and reweighted (see Methods section). Reweighting was

done in specific regions of interest in CV phase space (i.e., energetic minima on the surface). The top-weighted structures from each clustering analysis in the respective regions of phase space are presented here (Figure 5-2, bottom). For R5 without pS, the peptide lies close to the surface in extended structures (~ 1.2 nm and 1.6 nm radius of gyration). Electrostatic interactions between positively-charged sidechains (lysine and arginine) and negatively-charged adsorption sites on the silica surface allow the peptide to form and retain these extended structures on the surface. The RRIL motif at the C-terminus of R5 (green ribbon, Figure 5-2) also binds strongly to silica. This motif was shown to be crucial to silica precipitation using unphosphorylated R5 in experiments.[201] The RRIL motif is, however, absent in native silaffin and thus, it can be expected that it will have a negligible role in binding to the surface in our systems with phosphorylated R5 peptides.

Indeed, when a single pS residue is introduced (Figure 5-2 C, D, E), the RRIL motif interacts less strongly/frequently with the surface. This is demonstrated in Appendix, which shows the free energy profile as a function of the distance between the center-of-mass of the RRIL motif in each peptide and the silica surface (see Appendixfor pH 5 results). Instead, the RRIL motif in local pS R5 is observed to interact with the lone pS residue in the sequence; either two arginine residues, or one arginine residue and an Na^+ ion, bind to the pS residue to compensate for its -2 charge at pH 7.5. However, the N-terminal residues are free to interact with the surface through their positively-charged sidechains so that they extend and approach the surface closely (Figure 5-2 C, D). The C-terminus can also self-interact with other residues in the peptide sequence, leading to a compact structure (Figure 5-2 E). In Figure 5-2E, the hydrophobic leucine residue in the RRIL motif is shown to be interacting with the hydrophobic ring of a tyrosine residue near the middle of the peptide, holding the structure in a loop. In this case, the N-terminus is still able to bind to the silica surface.

Global pS R5 also prefers a compact structure, promoted by the interaction of intramolecular, charged species (lysine/arginine residues with pS residues), and shows a significant reduction in binding of the RRIL motif to silica (see Appendix). However, in the case of global pS R5, the pS residues outnumber the positively-charged amino acids in the sequence, resulting in an overall negative charge of the peptide. Consequently, some pS residues interact with 2 Na^+ ions in their vicinity to compensate for their -2 charge. Notably, the peptide remains tethered to the surface through pS residues that bind to surface-adsorbed Na^+ ions, albeit with a lesser binding free energy. It is important to note that the adsorption mechanism of global pS R5 to silica is different from unphosphorylated or local pS R5, since the peptide has an overall negative charge and would ordinarily be repelled from the like-charged silica surface. However, the strongly adsorbed surface ions allow the peptide to adsorb close to the surface.

EFFECT OF PH ON THE MECHANISM OF R5/SILICA BINDING

Since biosilicification is a pH-sensitive process, it is crucial to investigate the effects of pH on R5/silica binding. Figure 5-3 shows the free energy profile for adsorption of R5 (and its variants) to the silica surface at pH 5. The free energy profiles at pH 7.5 are also plotted to facilitate comparison. Markedly, all peptide sequences bind with lesser binding free energy at pH 5 than at pH 7.5. The decrease in binding free energy is greatest for R5 without phosphorylation. The magnitude of the decrease in binding free energy (= 23 kJ/mol) is comparable to the decrease upon introducing one pS residue into the peptide sequence (= 25 kJ/mol). Conversely, both local and global pS R5 show a much smaller absolute decrease in the binding free energy (= 0-10 kJ/mol).

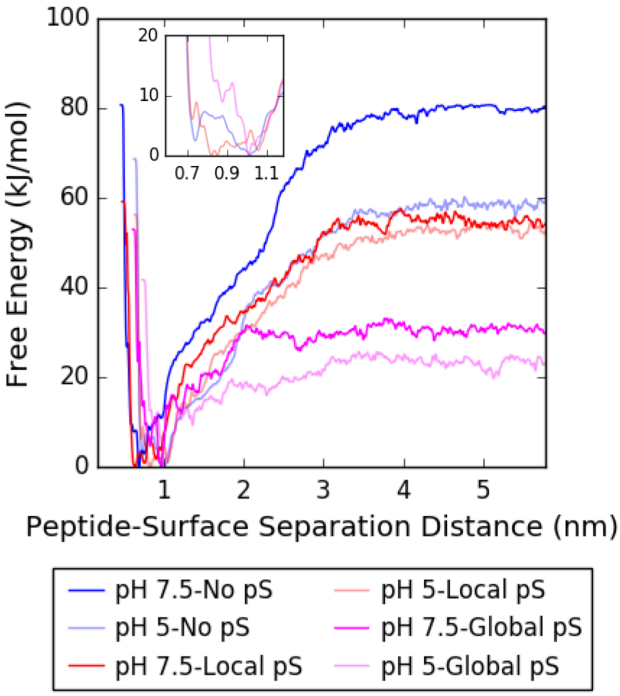


Figure 5-3 Free energy projected onto the distance between the silica surface at (dark colors) pH 7.5 and (light colors) pH 5 and the center-of-mass of (blue) R5 without phosphorylation ("No pS"), (red) locally phosphorylated R5 ("Local pS"), and (purple) globally phosphorylated R5 ("Global pS"). A zoomed-in view of the energy minima for the pH 5 simulations only is shown in the inset.

Looking closely at the free energy minima (Figure 5-3, inset), we find that as the degree of phosphorylation increases, the peptides preferentially adsorb at further distances from the surface. The same trend was evident from observing the free energy minima at pH 7.5 (Figure 5-1, inset). Like the case at pH 7.5, these increases in the surface separation distances of the peptides correlate with decreases in their binding free energies to silica at pH 5 (Figure 5-3). When R5 is unphosphorylated at pH 5, the peptide favorably resides in two minima at distances of ~ 0.73 nm and 1.0 nm (n.b., a single minimum was observed at ~ 0.68 nm at pH 7.5). This shift towards larger peptide-surface separation distances is likely the cause for the sizable decrease in binding free energy observed upon adsorption of unphosphorylated R5 to silica at pH 5 versus at pH 7.5. Local

pS R5 exhibits three minima at pH 5 at distances of ~ 0.80, 0.85, and 1.05 nm, which is a less dramatic departure from its behavior at pH 7.5 (n.b., three minima at ~ 0.63, 0.77, and 0.92 nm from the surface) compared to unphosphorylated R5, explaining the smaller decrease in binding free energy in moving from pH 7.5 to pH 5. Lastly, global pS R5 prefers to remain further from the surface at a distance of ~ 1.0 nm (like its behavior at pH 7.5 with a single minimum at ~ 0.88 nm from the surface).

From the 2D free energy profiles along the distance from the surface and radius of gyration (Figure 5-4), it is evident that the preferred radius of gyration decreases as the degree of phosphorylation increases, mimicking the behavior of R5 at pH 7.5. However, some differences do exist. For the case of the unphosphorylated peptide, the minimum at a radius of gyration of ~ 1.6 nm at pH 7.5 (Figure 5-2) is largely absent at pH 5 (Figure 5-4). Instead, the minimum at 1.3 nm is much larger. For local pS R5, the number of minima increases from pH 7.5 to pH 5, and the preference shifts towards a lower radius of gyration and greater distance from the surface. Moreover, the free energy surface at pH 5 is more diffuse and exhibits numerous small minima.

These findings necessitate an investigation of the structure of R5 and its variants, like the analysis for pH 7.5 (Figure 5-4, bottom). At pH 5, due to changes in the ionization state of the surface, fewer negatively-charged surface sites are available for adsorption of positively-charged sidechains in R5. With fewer interaction sites, R5 binds weakly to the surface – a trend that is evident in the overall decrease in binding free energy of all the peptides. These differences in the ionization state of the surfaces are directly reflected in the surface-bound configurations of R5 in Figure 5-4. Although all positively-charged sidechains bind to the surface at both pHs, other residues do not interact with the surface at pH 5, which prevents full extension of the peptide on the surface. At pH 5, the central residues hydrogen-bond with nearby water molecules instead of interacting with surface-bound Na^+ ions. The lack of Na^+ ions on the surface is directly correlated

with the lack of surface adsorption sites and the surface ionization state. Additionally, as seen at pH 7.5 (Figure 5-2), the introduction of pS residues forces the peptide into more compact structures due to intramolecular interactions of pS and lysine/arginine residues (Figure 5-4). In fact, R5 can even adopt a partially helical structure (Figure 5-4D.)

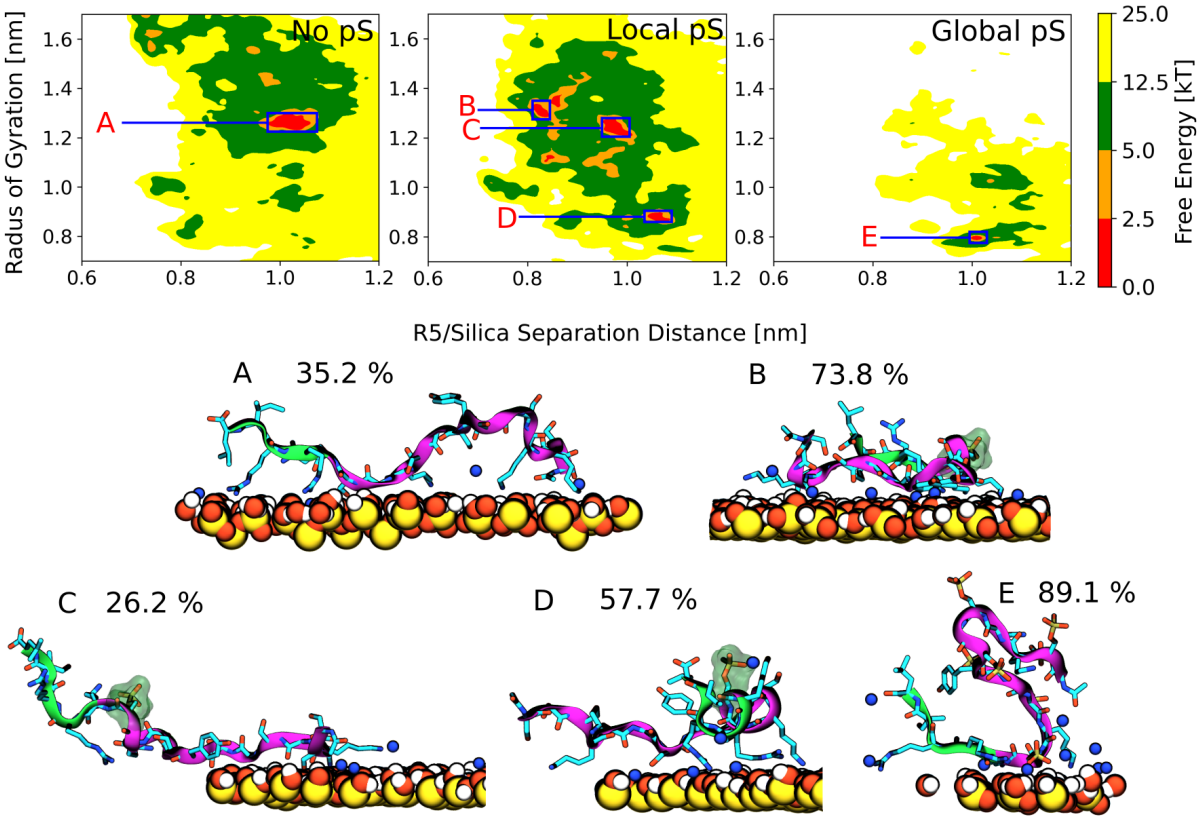


Figure 5-4 (Top) 2D plots of the free energy projected onto the distance between the silica surface at pH 5 and the center-of-mass of R5 without phosphorylation (left: "No pS"), locally phosphorylated (middle: "Local pS"), and globally phosphorylated (right: "Global pS"), as a function of the radius of gyration of the Cα atoms in R5. The x-axis is cut off to highlight the surface-bound states only. (Bottom) Conformations of the peptides with the highest probabilities from the clustering analysis of the simulations at pH 5 in the regions of the above-shown minima, along with the corresponding cluster weights from reweighting. Minima correspond to low free energy states indicated in the plots in the top row. Coloring is the same as described in Figure 2.

EFFECT OF THE PS SIDECHAIN PROTONATION STATE ON R5/SILICA BINDING

To understand how the observed differences in the binding free energies of the peptides at pH 5 and pH 7.5 are arising – i.e., to determine whether they are solely due to differences in the ionization state of the surfaces or pKa differences of the pS sidechain (-2 charge at pH 7.5 vs. -1 charge at pH 5) – two additional PTMetaD-WTE simulations were performed of local and global pS R5 at pH 5 with pS residues constructed to have a -2 overall charge. These simulations were carried out in the same manner, with the same parameters, and for the same length of time as previous simulations for consistency (see Methods section). Figure 5-6 shows a comparison of the change in free energy upon adsorption of local and global pS R5, with both sidechain protonation states, to the silica surface at pH 5.

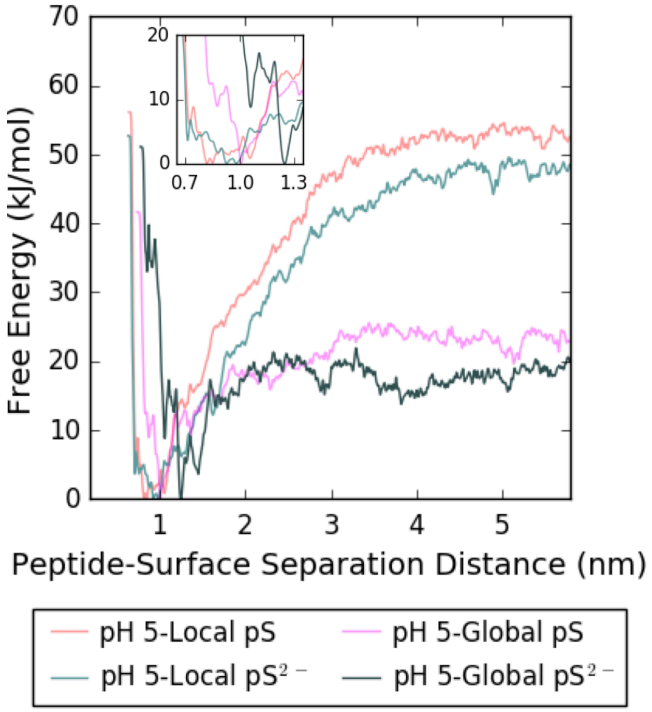


Figure 5-5 Free energy projected onto the distance between the silica surface at pH 5 and the center-of-mass of local pS R5 with a pS charge state of -1 (red) and -2 (light green), and global pS R5 with a pS charge state of -1 (purple) and -2 (dark green). A zoomed-in view of the energy minima is shown in the inset.

The results (Figure 5-6) show that the change in pS protonation state affects the binding free energy of local and global pS R5 to silica by ~3-5 kJ/mol, or by approximately 10% in either case. These findings suggest that, in addition to the surface chemistry effect, observed differences in the adsorption thermodynamics at pH 5 versus at pH 7.5 are also affected by the protonation state of the pS sidechain. On changing the protonation state of the pS sidechain so that it has a charge of -2 (state at pH 7.5) instead of -1 (state at pH 5), still provides a diffuse 2D free energy surface (like Figure 5-4, local pS). As shown in Figure 5-6, R5 shows only one accessible minimum which is close to the surface where it adopts an extended structure. However, the free energy surface has several other regions which might be accessible, within our free energy error. Unlike its behavior at -1 protonation state, the free energy surface does not have multiple low energy minima.

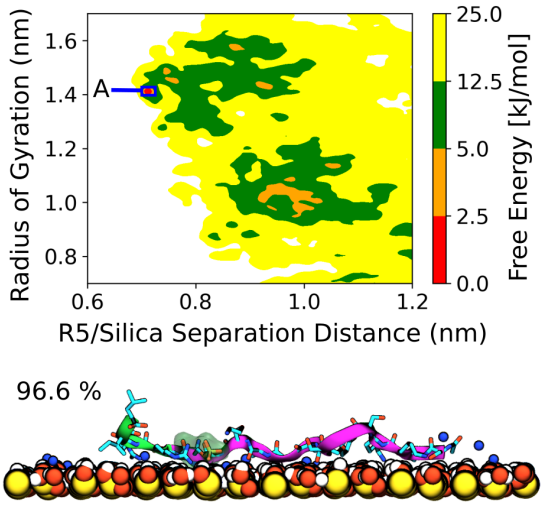


Figure 5-6 (Top row) 2D plot of the free energy projected onto the distance between the silica surface at pH 5 for local pS R5 (pS charge of -2). The x-axis is cut off to highlight the surface-bound states only. (Bottom rows) Top weighted surface-bound peptide conformations from a clustering analysis of the structures in the free energy minimum labeled as A in the plot above. Coloring is the same as described in Figure 2.

5.5 DISCUSSION

In peptide-mediated silica formation, cationic peptides like R5 are thought to aid the process at different stages. In the early stages of silica formation, these peptides promote condensation of silicic acid into polysilicic acid.[185] It is suggested that the peptides form templates where condensation can take place, but that sometimes the peptides themselves get encapsulated by silica.[183,201,202] Therefore, an alternative pathway for peptide-mediated biomineral formation may exist where precursors of the biomineral itself – or some other interface that is present (i.e., the air/water interface[195]) – bind silaffin-based peptides and facilitate peptide templating, which in turn leads to biomineral nanostructure formation.[203] Finally, cationic peptides are thought to induce flocculation of silica particles, by reducing the energy barrier for the growth of nuclei and thus promoting the growth of silica nanostructures.[185] To attain the desired yield and morphology of silica nanostructures from *in vitro* silicification, it is important to understand the forces that dominate every stage of this process. The interaction of peptides with silica is crucial to all stages of biosilicification.

In this paper, we investigated the interaction of silica with R5 – a commonly used peptide for *in vitro* silicification – with varying degrees of phosphorylation. It has been shown that unphosphorylated R5 precipitates silica at neutral pH but loses this ability at acidic pHs. Therefore, we analyzed the binding thermodynamics and structure of R5 at both pH 7.5 and pH 5. Our results show that unphosphorylated R5 binds much stronger to silica at pH 7.5 than pH 5, by approximately -23 kJ/mol. In fact, at pH 7.5, the peptide adopts extended structures on the surface, maximizing interactions of its positively-charged sidechains (i.e., lysine and arginine residues) with the surface. This also enables it to adopt a position close to the surface. On the other hand, unphosphorylated R5 assumes a more compact structure on the surface at pH 5. At acidic pH, it is unable to find the same number of charged surface sites for adsorption as at pH 7.5, forcing the

central residues to instead interact with nearby water molecules. The C-terminus RRIL motif that is implicated in the precipitation of silica nanostructures does not interact with the surface completely at pH 5. Therefore, we suggest that the inability of unphosphorylated R5 to assemble silica at pH 5 results from the changing ionization state of the surface that reduces the tendency of the peptide to interact with the silica surface. We note that the binding free energy of unphosphorylated R5 to silica at pH 5 is approximately the same as that of local pS R5 at pH 7.5, which was able to precipitate silica in experiments. Therefore, we further suggest that the inability of R5 to fully extend on the surface at pH 5 hinders or impairs the formation of peptide templates on the surface.

Another area of interest in R5/silaffin biosilicification has been to understand the effect of posttranslational modifications (PTMs) of the peptide sequence, such as phosphorylation or hydroxylation, on the silica precipitating activity of the peptides. Recently, Lechner *et al.*[190] phosphorylated the serine residues of R5 and measured the silica precipitating activity of a number of variants of R5 at neutral pH. R5, without modifications, precipitated silica nanospheres. However, on phosphorylating one and seven serine residues in R5, silica yields were drastically reduced to 38.8% and 0.6%, respectively, with singly-phosphorylated R5 yielding poorly-formed silica nanostructures. Our results highlight the effect of phosphorylation at a molecular-scale, hoping to shed light on the differences observed by Lechner *et al.* We find a clear trend of decreasing binding free energy to silica for an increasing degree of phosphorylation of the R5 peptide sequence. This trend remains the same at both neutral and acidic pHs. Global pS binds the least strongly of all three variants and is expected to yield the lowest amount of silica – confirmed by the results of Lechner *et al.* This is intuitive given the negative charges of both global pS R5 and the silica surface. However, in contrast to the hypothesis by Lechner *et al.* of a purely repulsive interaction between global pS R5 and silica, we see a minimum binding free energy of -30 kJ/mol,

predominantly due to interactions between global pS R5 and surface-bound ions. While the experimental and computational trends match in this case, it is not straightforward to assume that tight-binders to silica always precipitate the most silica. To this point, Patwardhan et al.[204] suggest that cationic peptides that bind strongly to silica perform poorly in early stages of silica precipitation. Limo and Perry[205] also argue that peptide-surface binding may not be the primary indicator of the biomineralization efficiency of a given peptide.

Nonetheless, details about the atomic-level interactions in these simulations can provide information useful to interpret these experiments. With a single pS residue introduced into the peptide sequence, the peptide remains cationic and we observe a binding free energy of -55 kJ/mol to silica at pH 7.5. Our results show that the binding free energy of local pS R5 is reduced compared to unphosphorylated R5 due to intramolecular interactions between nearby arginine and pS residues in the sequence of local pS R5. Local pS R5 mostly adopts compact structures compared to the extended structures of unphosphorylated R5 at pH 7.5, though extended structures of local pS R5 are also observed on the surface at pH 7.5. From our analysis of the surface-bound structures of R5, it is evident that it is primarily the binding of the C-terminus RRIL motif that is different between the unphosphorylated and locally phosphorylated peptides, suggesting a cause for the disruption in nanostructure formation and reduced silica yield observed with local pS R5 in the experiments by Lechner *et al.* Thus, our findings provide support for the previously proposed observation that the RRIL motif of R5 is important for *in vitro* silica precipitation.

5.6 CONCLUSION

Our study has highlighted some of the dominant driving forces that may contribute to the precipitation of silica and the formation of nanostructures in the biosilicification process. We find that the binding free energy of R5 to the silica surface decreases with decreasing pH and an increasing degree of phosphorylation. These differences are due to alterations in the electrostatic

interactions between the peptide and the surface or electrolyte ions, and to intramolecular hydrogen-bonding that dictates the binding between R5 (and its variants) with the silica surface at different pH conditions. At acidic pHs, the terminal chemistry of the surface dominates by forcing the peptide to interact with itself (independent of the degree of phosphorylation). However, at neutral pH, the degree of phosphorylation dominates the nature of R5 binding to silica, dictating the level of both self- and surface-interactions.

Our predictions of the nature of the interactions of R5 with silica highlights important molecular-scale features – the role of the C-terminal RRIL motif, the surface-repulsion of R5 at acidic pHs, the interaction of pS residues with the peptide/silica environment, and the facilitation of peptide-surface binding through surface-bound Na^+ ions. These results are crucial not only to understand previous experiments of the same nature (i.e., Lechner *et al.*) but also to predict how modifications of specific residues might affect the silica precipitating activity of the peptide. We hope our findings can serve as a guide for future experimental and theoretical efforts to direct bioinspired silicification towards the design of new biomaterials and biotechnologies.

5.7 ACKNOWLEDGEMENTS

JP and KGS acknowledge the support of NSF award CBET-1264459. JP and GD acknowledge support of NSF award MCB-1715123. This work was facilitated using computational, storage, and networking infrastructure provided by the Hyak supercomputer system, supported in part by the University of Washington and the UW Student Technology Fee Proposal program (award 2015-028) and NSF MRI program CHE-1624430.

Chapter 6. BIASING SMARTER, NOT HARDER, BY PARTITIONING COLLECTIVE VARIABLES INTO FAMILIES IN PARALLEL BIAS METADYNAMICS

Arushi Prakash[‡,1], Christopher D Fu[‡,1], Massimiliano Bonomi[2], Jim Pfaendtner[,1,3].*

**Corresponding author (jpfaendt@uw.edu), [‡]Equal contributions by authors*

[1]Department of Chemical Engineering, University of Washington, Seattle, Washington 98195, United States

[2]Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom

[3]Senior Scientist, Pacific Northwest National Laboratory, Richland, Washington, United States

6.1 ABSTRACT

Molecular simulations of systems with multiple copies of identical atoms or molecules may require the biasing of numerous, degenerate collective variables (CVs) to accelerate sampling. Recently, a variation of metadynamics (MetaD) named parallel bias metadynamics (PBMetaD) has been shown to make biasing of many CVs more tractable. We extended the PBMetaD scheme so that it partitions degenerate CVs into families that share the same bias potential, consequently expediting convergence of the free-energy landscape. We tested our method, named Parallel Bias MetaD with

Partitioned Families, on 3, 21, and 78 CV systems and obtained an approximately proportional increase in convergence speed compared to standard PBMetaD.

6.2 INTRODUCTION

Molecular simulations have immense potential to provide crucial molecular- and nano-scale details of physical, chemical, and biological processes. However, simulations of slow molecular transitions, like protein folding/unfolding and chemical reactions, continue to be stymied due to high free-energy barriers and rugged free-energy landscapes that limit sampling. Several enhanced sampling methods, like metadynamics (MetaD),[23,118] umbrella sampling,[29] hyperdynamics,[206] variationally enhanced sampling,[207] and adaptive force biasing,[208] have tried to alleviate this problem by applying a bias along predefined coarse-grained descriptors or collective variables (CVs) of the system to accelerate sampling. Since applying bias in a high-dimensional CV space is often inefficient, researchers have traditionally resorted to choosing a minimal set of CVs.[209]

Identifying a small set of CVs that can effectively differentiate between relevant states of a system is a primary challenge for many enhanced sampling approaches. While a variety of CV-selection methods, like time-lagged independent component analysis (TICA),[210] reconnaissance metadynamics,[211] and spectral gap optimization of order parameters (SGOOP),[212] have recently been developed to address this challenge, researchers primarily still rely on their physico-chemical intuition of the system to select optimal CVs. Frequently, more than one candidate CV is biased which presents the challenge of efficiently biasing them with limited computational resources. To this end, replica-based methods like multiple walker metadynamics,[35] altruistic metadynamics,[213] and the flying Gaussian method[214] have been developed to exploit parallel simulations that share the bias potential to accelerate sampling even further. However, these methods do not address the

problem of the high dimensionality of the bias potential, which requires extensive sampling, due to the larger phase space that needs to be explored, for convergence.

Some MetaD-based methods have been formulated to address the challenge of biasing a large number of CVs. In bias exchange[215] MetaD, a replica exchange approach is used, in which multiple replicas of the system, each biasing only one or few CVs, are simulated in parallel. Conformations of the system are periodically exchanged using a metropolis criterion. Other methods use several low dimensional bias potentials to bias individual CVs in lieu of a single high-dimensional bias potential.[33,209] Parallel bias Metadynamics (PBMetaD) is one such method that has been used to bias 4-40 CVs in the same simulation.[26,33,34,216,217]

This letter introduces PBMetaD with partitioned families (PBMetaDPF) for systems that require biasing multiple CVs that share identical properties. As an example, a simulation can contain multiple copies of a protein in a box. Since these proteins are identical subunits, they are expected to share identical properties. Thus, a CV describing property X of the n^{th} subunit would also describe property X of any other subunit of the system. Consequently, a system with N identical subunits would contain N CVs describing property X for each subunit. These CVs can be considered as indistinguishable or *degenerate,* as their equilibrium probability distributions are identical. In PBMetaDPF, the gain in efficiency over PBMetaD is achieved by grouping degenerate CVs into one family so that CVs in the same family share and contribute to the same bias potential. Finally, the free-energy profile for each CV family can be determined using the standard MetaD integration and standard reweighting techniques[147,218] can be used to study other degrees of freedom. The concept of a shared bias potential is inspired by multiple walker MetaD[35] where the potential of a CV evolves by contributions from *walkers* of the CV in parallel replicas of the

system. In contrast, in our method all contributions to this shared potential come from multiple CVs within one replica only.

6.3 THEORY

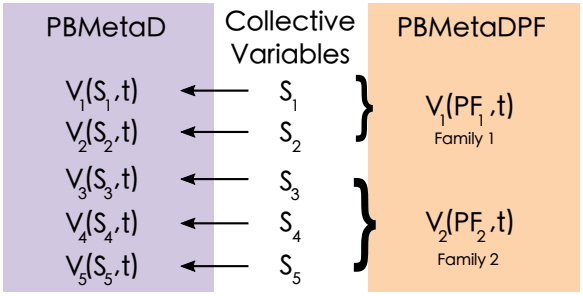
To highlight the differences between the PBMetaD and PBMetaDPF approaches, we will briefly review the theory of PBMetaD. In classical MetaD, or its well-tempered variant (WTMetaD),[23] a single multidimensional bias potential is constructed as a function of user-specified CVs, where the dimensionality of the potential is equal to the number of CVs. In contrast, PBMetaD constructs multiple monodimensional bias potentials applied along each individual CV. The method uses a new scheme to permit the instantaneous application of an arbitrary number of bias potentials, which was shown to converge (empirically) to the exact underlying free-energy surfaces.[33] To achieve this, the bias potential for the i^{th} CV(s_i), under the PBMetaD framework, is constructed through the following equation.

$$V_G(s_i,t) = \int_0^t dt'\, W * \exp\left(-\frac{V_G(s_i R(t'),t')}{k_B \Delta T}\right) * \exp\left(\frac{-(s_i(R(t)) - s_i(R(t')))^2}{2\sigma^2}\right) * W_{PB}(s_i,t') \quad 6.1$$

$$W_{PB}(s_i,t) = \frac{\exp\left(-\frac{V_G(s_i R(t),t)}{k_B T}\right)}{\sum_{j=1}^{n} \exp\left(-\frac{V_G(s_j R(t),t)}{k_B T}\right)} \quad 6.2$$

where W is the initial Gaussian height, σ_i the width of the Gaussian, dt' the pace of Gaussian deposition, k_B the Boltzmann constant, T the system temperature, ΔT is an input parameter with units of temperature which controls the rate at which Gaussians are scaled down, and n is the total number of CVs in the system. The first three terms in Eq. 6. 1 follow the algorithm of a typical WTMetaD simulation where the Gaussian height is reduced as bias accumulates, while the last term, shown in Eq. 2, is a conditional weight to account for the effect of the bias deposited along the other CVs due the correlations among CVs.[209] This conditional weight distributes a new

Gaussian across all CVs, with the CVs with lower bias at the time of deposition receiving a larger contribution. In this algorithm, each bias potential evolves independently, and the only interaction occurs via the conditional weight term.



Schematic 1. Diagrammatic view of the differences between PBMetaD and PBMetaDPF sampling schemes. Under the PBMetaD biasing scheme, an individual bias potential is evolved for each CV and the CV only acts under its own potential. In contrast, the PBMetaDPF schemes allows for all of the members of a given family to contribute to the formation of a single bias potential that, in turn, acts on all of the members of a particular family.

While PBMetaD offers a more scalable way to perform high-dimensional sampling, converging numerous bias potentials independently can still require lengthy simulation times, as evidenced by Prakash et al.[34] In PBMetaDPF, we expedite the convergence of these energy landscapes, which comprise indistinguishable particles, by partitioning degenerate CVs into families (PFs). For book-keeping purposes, we will refer to CVs as s_{PFf-k}, where *f* refers to the PF it belongs to and *k* refers to the CV index in that PF (e.g. s_{PF1-2} is the second CV belonging to PF one). CVs partitioned into the same PF deposit bias similar to the multiple walkers framework,[35] where the Gaussians deposited along the different individual CVs of a particular PF all contribute to the formation of a single bias potential that acts on all the CVs belonging to that PF. The bias potential for any CV belonging to partitioned family PF1 (s_{PF1-x}; where x is any member of family 1), which has m members, is recovered through:

$$V_G(s_{PF1-x},t) = \sum_{k=1}^{m}\int_0^t dt'\, W * \exp\left(-\frac{V_G(s_{PF1-k}R(t'),t')}{k_B\Delta T}\right) * \exp\left(\frac{-(s_{PF1-x}(R(t))-s_{PF1-k}(R(t')))^2}{2\sigma^2}\right) * W_{PB}(s_{PF1-k},t') \qquad 6.3$$

Here, the bias potential of each family is constructed by the contributions of every member. In other words, the bias potential of the family acts on each member of the family. Consequently, only one free-energy profile is recovered per PF instead of recovering one free-energy profile for each CV. Note that the denominator of the conditional weight term still sums over all the CVs biased in a system, as is done in regular PBMetaD.

6.4 RESULTS

3-PARTICLE LENNARD-JONES SYSTEM

To assess the accuracy and efficiency of our approach, we used a simple three-particle Lennard-Jones (LJ) system (σ = 0.39 nm, ε = 20 kJ/mol). For both PBMetaD and PBMetaDPF, we biased all three interatomic distances with initial Gaussian heights of 2.0 kJ/mol, Gaussian widths of 0.01 nm, a bias factor of 10, and a deposition pace of 1 ps. Sixteen independent biased simulations were run in the NVT ensemble for 2 μs. Further, to confirm that the PBMetaD framework is suitable for describing such systems, we performed parallel tempering (PT) simulations to provide an independent reference free-energy profile (see SI for details).

We monitored the root mean squared deviation (RMSD) of the free-energy profiles recovered from PBMetaD and PBMetaDPF relative to that obtained with PT to assess both convergence speed and accuracy. The RMSD between two profiles was defined as:[219]

$$RMDS\left(F_{ref}, F\right) = \sqrt{\frac{1}{\Omega}\int dS[\left(F_{ref}(S) - \bar{F}_{ref}\right) - (F(S) - \bar{F})]^2} \qquad 6.4$$

where S is the CV value, $F_{ref}(S)$ and $F(S)$ are the two free-energy profiles being compared, $\bar{F}$ and $\bar{F}_{ref}$ are the free-energy averaged over the region Ω. For the three-particle and 13-particle systems, the region of interest was defined as the CV space within 30 kJ/mol of the global minimum of the reference PT profile, while for the seven-particle system the region of interest was defined as the CV space within 100 kJ/mol of the global minimum of the reweighted interatomic

distance profile from WTMetaD. For each PBMetaD simulation, we recovered three free-energy profiles, one for each interatomic distance. In the case of PBMetaDPF, a single free-energy profile for a given simulation is naturally recovered because the method groups the three CVs into the same family.

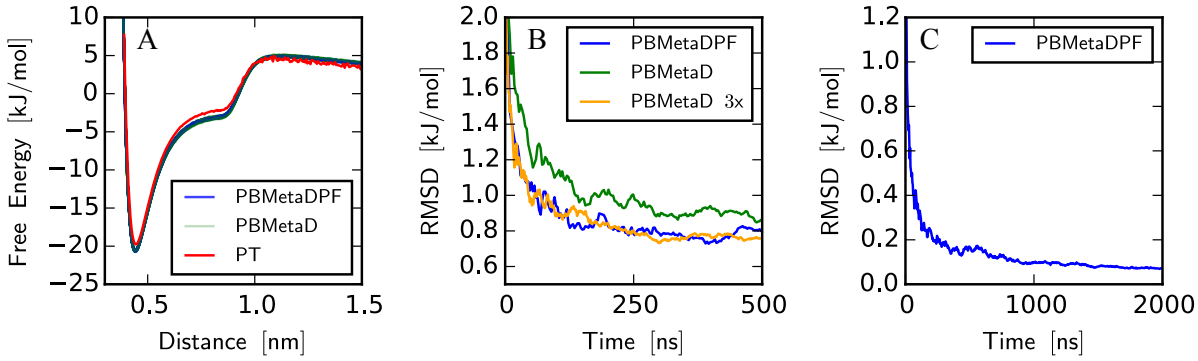


Figure 6-1 (A) Mean-aligned free-energy profiles of the interatomic distance between LJ particles. In total, the 16 PBMetaDPF profiles, the 48 PBMetaD profiles, and one parallel tempering (PT) profile are plotted. (B) The average RMSD of PBMetaDPF profiles (blue), PBMetaD profiles (green), and of PBMetaD with a projected convergence rate of three times faster (orange), all RMSD calculations are relative to the reference PT profile. (C) The average RMSD of PBMetaDPF relative to the converged PBMetaD profile over the course of the simulation.

As shown in Figure 6-1A, both PBMetaD and PBMetaDPF accurately reproduced the free-energy profile along the interatomic distance obtained with PT. The RMSD of the free-energy profile recovered from PBMetaDPF simulations is well-within k_bT (~ 2.5 kJ/mol at 300 K) when compared to PT (Figure 6-1B) and PBMetaD (Figure 6-1C) profiles. The error primarily stems from differences in the higher free-energy regions (see Appendix) since it is difficult for the parallel tempering run to converge in that area. However, the small value of free energy difference shows that partitioning CVs into families does not introduce additional errors, for this system. In fact, on average, PBMetaDPF converges to the reference PT profile approximately three times

faster than PBMetaD, as shown by the overlap between the red and green lines in Figure 6-1B. A three-fold acceleration in convergence is attributed to the fact that the bias potential in PBMetaDPF is constructed by three CVs as opposed to a single CV in PBMetaD.

13-PARTICLE LENNARD-JONES SYSTEM

To further demonstrate the accelerated convergence offered by the PBMetaDPF, we simulated a 13-particle LJ system (σ = 0.39 nm, ε = 11 kJ/mol). A lower ε value was chosen so that PT simulations could converge in a reasonable amount computational time. In PBMetaDPF, all interatomic distances (78 CVs) were biased and grouped into the same family. Again, we performed 16 independent simulations using the same bias parameters as the three-particle system. We also performed a PT simulation of the system (see details in Appendix) to provide an independent reference for the free-energy profiles.

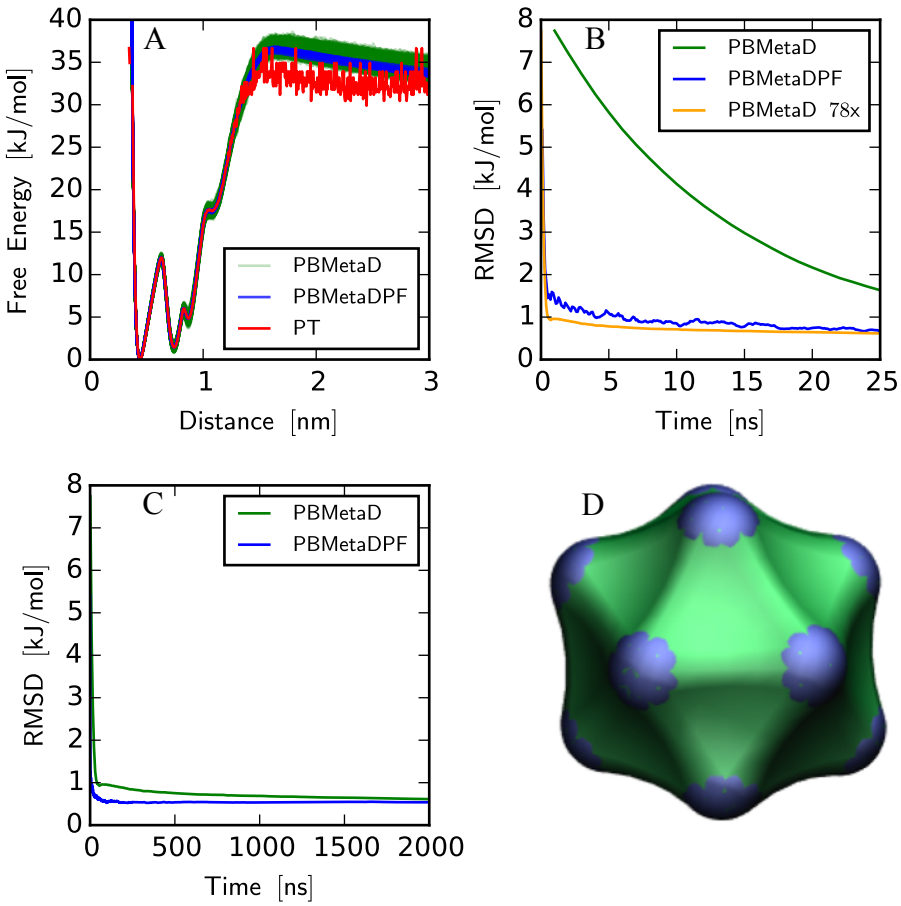


Figure 6-2 (A) All of the mean-aligned free-energy profiles for PBMetaD after 4 μs (78 profiles x 16 trials) and PBMetaDPF after 4 μs (16 trials) and one profile from parallel tempering (PT). (B) The average RMSD, with respect to the converged PT profile, of PBMetaDPF profiles (blue), PBMetaD profiles (green), and of PBMetaD with a projected convergence rate of 78 times faster (orange) over the course of the simulation. (C) The average RMSD of PBMetaD profiles (green) and PBMetaDPF profiles (blue), all RMSD calculations are relative to the converged PT profile over the course of the simulation. (D) Structure corresponding to the global free-energy minimum.

PBMetaD (all 78 CVs) and PBMetadPF converged to the same free-energy profiles (Figure 6-2A). However, the PBMetaD profiles in Figure 6-2A exhibit more fluctuations between free-

energy profiles than the single PBMetaDPF profile. This highlights the challenge of converging 78 independent free-energy profiles, which necessitates that each atom pair explore the entire CV space. However, we do observe that in the long-time limit (> 2 μs), both PBMetaD and PBMetaDPF give approximately equivalent average RMSD values (Figure 2B). Similar to the LJ3 system, it is shown that partitioning the CVs into a single family allows accelerated convergence of the free-energy profile to (Figure 6-2C). Most notably, the speed in convergence is again proportional to the number of CVs in the family, 78 in this case.

In addition to recovering the free-energy with respect to the biased CVs (interatomic distances), the coordinates from the trajectory and metadynamics biases can be utilized to obtain stable structures of the LJ particle system. To find the most stable structure, the frames of the PBMetaD and PBMetaDPF trajectories were clustered (see Sppendix Methods). The clusters were reweighted (described in detail in the SI Methods) using the PBMetaD and PBMetaDPF bias and the method of Torrie and Valleau[29] to find the most stable structure – an icosahedron (Figure 6-2D), which had highest probability of all the structures (~ 100%). This result is consistent with previous analyses of the system where the icosahedron structure of LJ_{13} was shown to be at least 2.85ε (= 31.35 kJ/mol for this system) more stable than other structures.[220,221]

Before this, a MetaD approach that biases 78 CVs for exploring the structural minima of aggregating systems such as the one above would have been intractable. But by partitioning the CVs into a single family, scaling this approach to larger systems is now possible.

7-PARTICLE LENNARD JONES SYSTEM

Lastly, to demonstrate the effectiveness of PBMetaDPF to explore and describe systems with multiple, metastable states, we applied it to a 7-particle LJ system constrained to two dimensions. This system is well-studied using a variety of methods and is known to have four

stable structural minima and 19 transition states.[222–224] Here, we apply PBMetaD and PBMetaDPF to explore the potential energy landscape by biasing the 21 interatomic distances, running each simulation for 2 μs. As in previous treatments of interatomic distances, we partition them into a single PF for PBMetaDPF. For comparison, we also carried out unbiased MD simulations as well as a WTMetaD simulation where the second and third moments of the coordination numbers were biased, following the work of Nava et al.[222] Both PBMetaD and PBMetaDPF yielded identical free-energy profiles for the interatomic distances. These profiles were identical to the free-energy profiles recovered from reweighting the WTMetaD onto the interatomic distances (Figure 6-3A). In this case too, PBMetaDPF could converge the free-energy surface for the interatomic distances faster (Figure 6-3B), and the degree of acceleration is commensurate with the reduction of profiles (21 times faster). Unsurprisingly, the unbiased MD simulation (~ 2 μs) did not generate the correct free-energy surface (see Appendix) for this system emphasizing the need for enhanced sampling techniques.

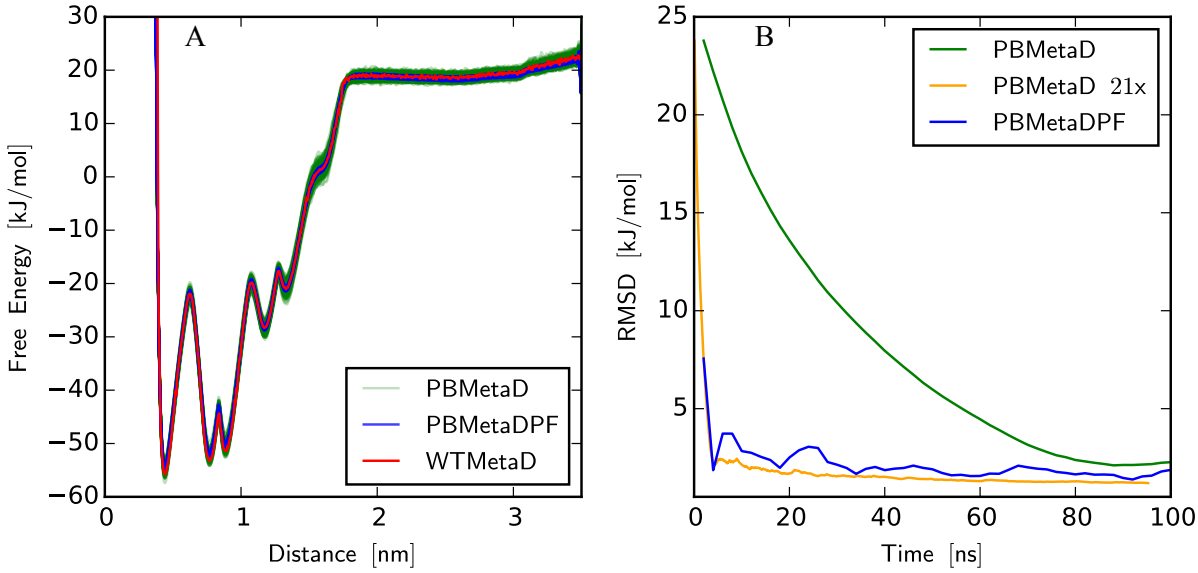


Figure 6-3 (A) Mean-aligned free-energy profiles of the interatomic distance between LJ particles. In total, the 16 PBMetaDPF profiles, the 336 PBMetaD profiles, and one WTMetaD profile (reweighted) are plotted. (B) The average RMSD of PBMetaDPF profiles (blue), PBMetaD profiles (green), and average RMSD of PBMetaD projected to converge 21 times faster (orange) relative to the converged WTMetaD profile over the course of the simulation. The area of interest was restricted to be 100 kJ/mol of the minimum of the reweighted WTMetaD profile.

In previous MetaD studies of this system, the second and third moments of the coordination number were selected and biased, as these CVs were able to differentiate between the four stable structural minima.[222,223] The limit to two CVs was also more amenable to methods that were unable to bias more than a few CVs at the same time. However, we chose to bias the interatomic distances using the PBMetaD and PBMetaDPF frameworks since it is a far more general approach and the CVs are more interpretable. After the simulations were converged, we reweighted the second and third moments of the coordination number to demonstrate the consistency of our results with prior work.[222,223] As shown in Figure 6-4A-C, the correct FESs were recovered with reweighting using both PBMetaD and PBMetaDPF. The FESs recovered using the two methods are identical in shape

(see Appendix) and exhibit identical approach to convergence over the course of the simulation (Figure 6-4 B and C).

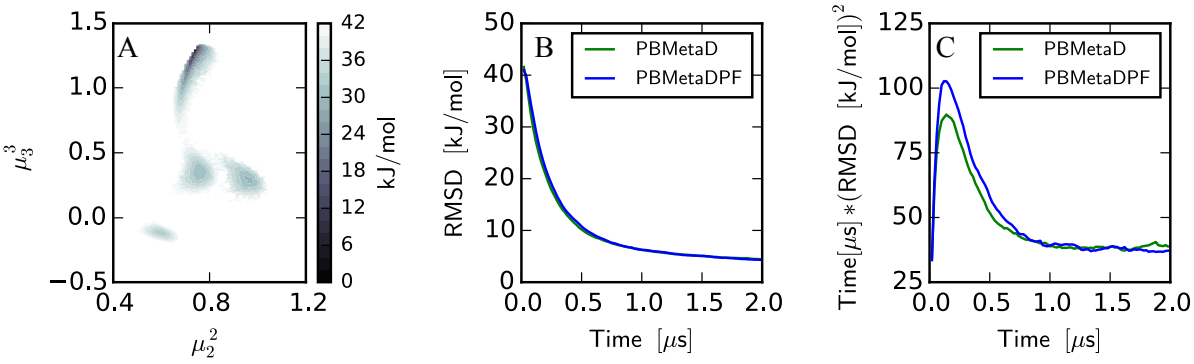


Figure 6-4 (A) Free-energy surface for the 7-particle LJ system reweighted for the second and third moments of coordination numbers using PBMetaDPF. (B) The average RMSD from 16 PBMetaD and PBMetaDPF simulations (each) reweighted for the second and third moments of coordination numbers with respect to a WTMetaD simulation biasing those same CVs. (C) A demonstration of the absence of systematic error in reweighting both PBMetaD and PBMetaDPF. The area of interest was restricted to be 40 kJ/mol of the minimum of the reweighted WTMetaD surface.

Further, we analyzed the trajectories using the clustering and reweighting method described in the SI to find the most stable structural minima. We were able to obtain the minima obtained in previous investigations of the system (Figure 6-5). Further, the probabilities of the structural minima predicted using PBMetaD and PBMetaDPF were within 0.5 % of each other (see Apendix).

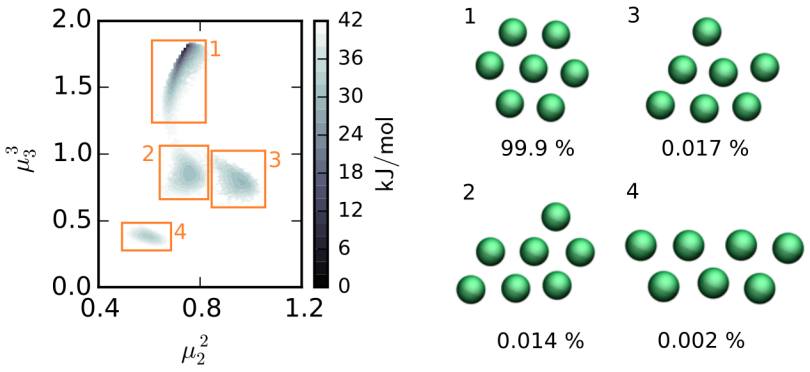


Figure 6-5 (left) Free-energy surface recovered from PBMetaDPF simulation of the 2D 7-particle LJ system after reweighting for second and third moments of the coordination number. (right) Representative structures for the regions highlighted in orange on the free-energy surface along with the probability of occurrence of each structure in the 2D phase space plotted on the right.

This study highlights that PBMetaDPF can be used to effectively sample the simple yet challenging structural phase space of the 7-particle LJ system. Furthermore, biasing the interatomic distances is a far more general, and intuitive approach for studying these complex structures. This eliminates the use of unintuitive CVs like moments of coordination numbers. Moreover, unlike PBMetaD, PBMetaDPF offers a more scalable approach by leveraging the excessive number of degenerate CVs to expedite converging an individual FES, rather than converging 21 profiles independently (as also proved in the LJ_{13} system). Lastly, the free-energy surface with respect to other CVs of interest can always be reweighted and constructed after converging the FES for the distance CVs.

OTHER POTENTIAL APPLICATIONS

In this study, we have applied PBMetaDPF to simple LJ systems. However, we can predict several other applications for this method where researchers were restricted to using a single CV because previous methods were not amenable to multiple, degenerate CVs. In a recent study of

ion-channel dynamics,[215] the authors treated degenerate ion distances as separate CVs in bias-exchange metadynamics and noted that the result was the same. Conceivably, the PBMetaDPF scheme could be used to bias these degenerate variables. A similar use case could be obtaining the potential of mean force (PMF) profiles of ion-pairing systems, using distances between ions as their CV.[225] Most ion PMFs are calculated with only a single ion-pair in water which requires an additional correction term to account for differences between simulated (one pair) and experimental ion concentrations.[226] PBMetaDPF can be used in simulations to bias the distances with multiple ions pairings, treating unique ion-pair distances as PFs, which would allow the recovery of free-energy estimates at realistic concentrations, potentially providing efficient routes to describe ion solvation/desolvation.[227]

Another potential area of application is in aggregation studies, where the free-energy of association of a particle is computed to predict the behavior of aggregates at higher concentrations for experiments. For example, the dimerization free-energy of peptoids in the dilute limit was used to predict how peptoids at higher concentrations aggregate in solution.[90] Similarly, the dimerization free of two MgO nanoparticles in vacuum was calculated to understand the crystal growth that occurs in saturated solutions.[228] We suggest that more crowded environments should be simulated and all the inter-particle distances should be biased to recover better estimates of the free energies of association at realistic, experiment level concentrations. While the concentration levels would typically require biasing too many CVs, partitioning the families of indistinguishable particle interactions should allow for this type of approach to be feasible. Many aggregation studies simulate multiple particles in a box, for example when looking at amyloid peptide aggregation with MD simulations.[229,230] We propose that using PBMetaDPF to bias inter-particle distances, and radius of gyration of all residues exhaustively sample the system and reduce the time required to sample all structural minima.

6.5 CONCLUSION

In this letter, we presented PBMetaDPF, an extension of the PBMetaD approach to expedite convergence for a special sub-class of enhanced sampling problems. The method partition degenerate CVs into families, such that they contribute to the same bias potential, in the spirit of the multiple walker scheme. For 3-, 7-, and 13-particle LJ systems, we demonstrated that partitioning CVs into one family led to accelerating the convergence of free-energy profiles. Further, in the case of the 7-particle LJ system, we show that standard reweighting techniques can be used to calculate the free-energy as a function of other CVs not directly biased and detect stable structural minima. Thus, even after partitioning CVs into families, the method maintains its compatibility with commonly-used reweighting protocols. Finally, and most importantly, grouping the CVs into one family led to an increase in the convergence speed of the free-energy profiles that was proportional to the number of CVs grouped into the family. Lastly, PBMetaDPF has been implemented in the open-source PLUMED[175] library (developers' version soon to be made available to the public), allowing for this approach to not only be readily accessible, but also be easily applied to variety of systems and problems.

6.6 ACKNOWLEDGEMENTS

This work was accomplished with support of AFOSR Grant FA9550-16-1-0051 monitored by Dr. Chiping Li. This work is supported in part by an NSF IGERT grant DGE-1258485. This work was also supported by the MS3 (Materials Synthesis and Simulation Across Scales) Initiative at Pacific Northwest National Laboratory (PNNL), a multi-program national laboratory operated by Battelle for the U.S. Department of Energy. This work was facilitated through the use of computational, storage, and networking infrastructure provided by the Hyak super-computer system, supported in part by the University of Washington.

Chapter 7. A SOLID-STATE NMR AND MD STUDY OF THE STRUCTURE OF THE STATHERIN MUTANT SNA15 AS A FUNCTION OF MINERAL SURFACE

Erika L. Buckle[1+], Arushi Prakash[2+], Massimiliano Bonomi[3], Jim Pfaendtner[2,], and Gary P. Drobny[1,*]*

[1]Department of Chemistry, University of Washington, Box 351700, Seattle, Washington 98195, United States

[2]Department of Chemical Engineering, University of Washington, Seattle, Washington 98195, United States

[3]Structural Bioinformatics Unit, Institut Pasteur, CNRS UMR 3528, 75015 Paris, France

[+]Authors contributed equally to this work

7.1 ABSTRACT

Elucidation of the structure and interactions of proteins at native mineral interfaces is key to understanding how biological systems regulate the formation of hard tissue structures. In addition, understanding how these same proteins interact with non-native mineral surfaces has important implications for the design of medical and dental implants, chromatographic supports, diagnostic tools, and a host of other applications. Here, we combine solid-state NMR spectroscopy, isotherm measurements, and molecular dynamics simulations to study how SNa15, a peptide derived from the hydroxyapatite (HAP) recognition domain of the biomineralization protein statherin, interacts with HAP, silica (SiO_2) and titania (TiO_2) mineral surfaces. Adsorption isotherms are used to characterize the binding affinity of SNa15 to HAP, SiO_2, and TiO_2. We also apply 1D ^{13}C CP MAS, 1D ^{15}N CP MAS, and 2D ^{13}C-^{13}C DARR experiments to SNa15 samples with uniformly

^{13}C- and ^{15}N- enriched residues to determine backbone and side-chain chemical shifts. Different computational tools, namely TALOS-N and molecular dynamics simulations, are used to deduce secondary structure from backbone and side-chain chemical shift data. Our results show that SNa15 adopts an α-helical conformation when adsorbed to HAP and TiO_2, but the helix largely unravels upon adsorption to SiO_2. Interactions with HAP are mediated in general by acidic and some basic amino acids, although the specific amino acids involved in direct surface interaction vary with surface. The integrated experimental and computational approach used in this study is able to provide high-resolution insights into adsorption of proteins on interfaces.

7.2 INTRODUCTION

Several organisms produce hard structures such as bone, teeth, shells, and exoskeletons from inorganic materials (i.e. calcite, silica, and hydroxyapatite) through a process known as biomineralization. These organisms use specialized proteins that interact with these inorganic materials thereby accelerating[231] or inhibiting[232,233] the growth of mineral solids , or even altering mineral morphology.[180,187,188,193,234–236] Detailed atomic-level insight of how of proteins interact with these minerals is required to better understand the role that proteins play in biomineralization and to thereafter exploit it for materials synthesis. This requires the resolution of the secondary and tertiary structure of proteins adsorbed on surfaces, the identification of both the residues that are crucial for protein-surface interaction and the nature of protein-surface interactions.

To date, only a handful of experimentally confirmed structural models of proteins interacting with their native mineral surfaces have been reported.[237–239] These protein-surface systems include either small extracellular matrix (ECM) proteins or ECM protein domains that are adsorbed onto hydroxyapatite (HAP) crystals.[162,240–250] These studies suggest that protein secondary structure can change when adsorbed to surfaces. They also hypothesize that the interaction of amino acid side-chains with inorganic surfaces are dependent on the secondary

structures of adsorbed proteins.[241,244–247,249] Thus, all components of protein-surface interaction can be determined only when the structure of the adsorbed protein is resolved.

Salivary statherin is a 43-residue ECM protein (DpSpSEEKFLRRIGRFGYGYGPYQPVPEQPLYPQPYQPQYQQYTF) that regulates HAP nucleation at enamel surfaces. The structure of statherin adsorbed onto HAP has been resolved.[250–253] Statherin has been studied by calorimetric methods[12,25], adsorption isotherms,[251,254] solution NMR,[255,256] solid-state NMR (ssNMR),[241–243,245,246,248–250,252,253,257,258] and circular dichroism.[251,259] These studies indicate that statherin is unstructured in solution but structures when adsorbed onto HAP surfaces. Notably, the 15-residue long N-terminus adopts a distorted α-helical conformation and the C-terminus also becomes partly helical upon adsorption to HAP.[253,260,261]

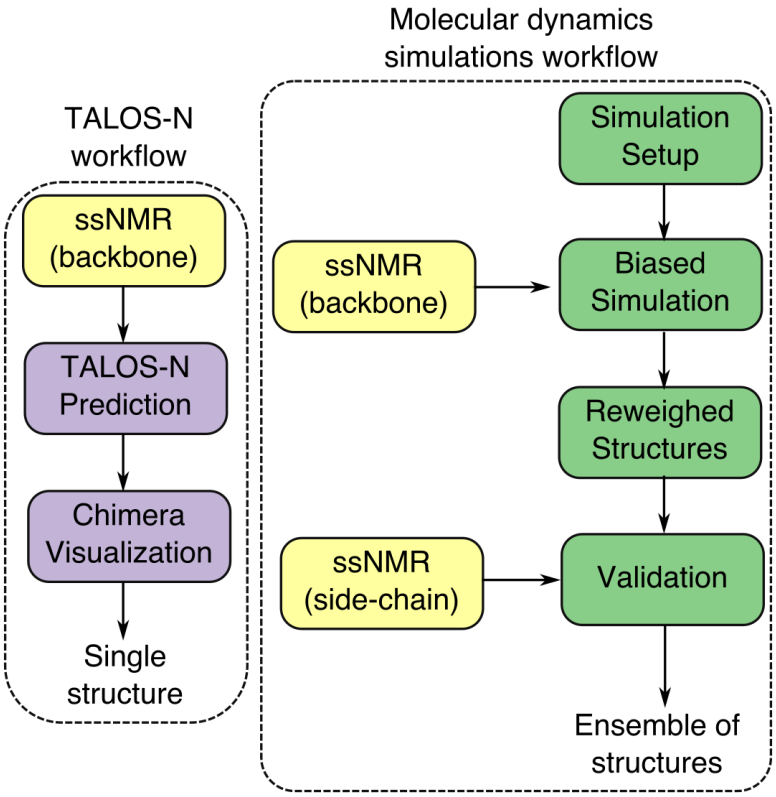
These studies also identify specific residues in statherin that interact with HAP. For example, acidic side-chains exhibit strong affinity for calcium phosphate surfaces.[251] Consequently, the N-terminus of statherin, which containing acidic residues like phosphoserine (pS), aspartic acid (D), and glutamic acid (E), has been shown to be essential for HAP binding.[251] Similarly, the interaction of basic residues with HAP have also been studied. When Goobes et al. mutated selected basic residues to alanine, the protein displayed lower the binding affinity to the surface but did not affect the surface coverage or the adsorption enthalpy.[250] In contrast, when all the basic residues (K6, R9, R10, and R13) were replaced with alanine, there was a significant change in the adsorption enthalpy and surface coverage. Therefore, they hypothesized that the nature of basic residues is not as important to binding as the presence of a number of basic residues.[250] In contrast to acidic and basic residues, the role of non-polar residues in protein-surface interactions is less clear. The HAP-binding domain of statherin has four types of nonpolar residues (L, I, G, and F) but only F14 was found to be in proximity with the HAP surface.[20,23]

There are limited studies that model the interaction of statherin and HAP. Notably, Rosetta and RosettaSurface modeling programs have been used for this.[244,252,262] However, these all-atom models of the protein include simplifying assumptions that make calculations faster, like freezing the surface, restraining the protein backbone, and not accounting for waters and ions. Thus, they might exclude important interactions that affect protein structures. MD simulations, which can explicitly model the surface, protein, and solvent, include these interactions. In fact, recent advances in the use of advanced MD-based simulation methods like metadynamics[31,118] have been shown to be highly effective in simulating the structure and binding of peptides at surfaces[127]. Furthermore, owing to the ease with which NMR chemical shifts can be directly estimated from classical simulations,[263,264] the metadynamic metainference approach[26,265] has been developed as part of a growing integrative structural biological toolkit[266] for pairing enhanced sampling MD simulations with experimental structural restraints. Given the comparatively small number of constraints/restraints that arise in a ssNMR structural study, this approach is particularly attractive for resolving the structure of adsorbed proteins and peptides on different surfaces.

In this study, we make use of the N-terminal domain of statherin, SN15 (D*pSpS*EEKFLRRIGRFG, where pS indicates a phosphorylated serine), which is often used as a model for statherin since it has nearly the same binding affinity for HAP as statherin[26]. Importantly, the binding of SN15 to HAP is greatly diminished if the serines are not phosphorylated. Remarkably, this binding affinity can be restored when the serine residues are replaced by aspartic acid[251], leading to the model peptide SNa15 (D*DD*EEKFLRRIGRFG). We study the adsorption of SNa15 to different inorganic surfaces, namely HAP, silica (SiO_2), and titania (TiO_2). We explore the surface-bound structure of the peptide, identify the binding residues, and investigate the changes in binding upon changing the mineral substrate to which the peptide binds. For this, we use an integrative experimental and simulation approach, similar to recent studies,[240,243,267],

especially using NMR data as structural restraints in the molecular dynamics step to predict binding motifs of peptides on surfaces.[26,268]

To characterize binding experimentally, we use both adsorption isotherms and 1D ^{13}C CP MAS, 1D ^{15}N CP MAS, and 2D ^{13}C-^{13}C DARR experiments. These experiments provide chemical shifts for the backbone and side-chain which can resolve the peptide structure with high fidelity. Further, we use two modeling approaches – TALOS-N and molecular dynamics (MD) simulations. TALOS-N utilizes a protein database to predict the most likely structure of the peptide given the database and the experimentally-determined chemical shift data.[45] However, since peptides often present an equilibrium ensemble of many structures with similar free-energies, even when bound to a surface, a single snapshot derived from TALOS-N might not provide a complete picture.[265,269] Thus, we use MD simulations to generate an ensemble of surface-bound conformations consistent with experimental data. We use a recently proposed method named metadynamic metainference[26,265] which uses Bayesian inference to incorporate structural restraints derived from experimental data, while accounting for errors in both the experimental data and simulation models. This integrated approach allows us to resolve the peptide structure with higher fidelity than possible with a single structure prediction, allows us to zoom in on the specificity of the side-chains to the surface. Measured side-chain chemical shifts are not used in the molecular simulation step but kept aside for validation as illustrated in Schematic 1.



Scheme 1. Schematic of the modeling and simulation workflows used in this study

7.3 METHODS

MOLECULAR DYNAMICS SIMULATIONS

SYSTEM SETUP

The structure of SNa15 peptide when adsorbed onto HAP, SiO_2, and TiO_2 surfaces was probed using classical all-atom molecular dynamics simulations. The peptide was modeled using the AMBER99SB-ILDN force field.[270] Water was modeled using SPC/E water model.[271] HAP and SiO_2 were modeled using the INTERFACE force field,[170] while the TiO_2 was modeled using the force field by Matsui and Akaogi.[272] (details about surface construction in SI). These models were chosen according to previously published studies of peptide and surfaces.[34,126,273–275] Both HAP

and TiO_2 surfaces were neutral, whereas the negatively-charged SiO_2 surface was neutralized by Na^+ ions. The peptide had a net negative charge of -1 which was neutralized by a single Na^+ ion.

For each peptide-surface system, a slab of the inorganic surface was created, and the peptide was placed near the surface. This system was then solvated with water. Unfavorable contacts during this packing were removed by using a steepest descent algorithm that minimized the energy of the system. The pressure of the system was equilibrated to 1 bar by propagating the system in the NPT ensemble (Temperature = 300 K; Pressure = 1 bar) for 1 ns. The semi-isotropic version of Berendsen barostat[276] (tau =1.0 ps) was used so that only the z-axis of the box changed to correct the pressure of the system. The Bussi-Donadio-Parrinello thermostat[51] (tau = 0.1 ps) was used to maintain the temperature at 300 K. For the production run, the system was propagated in the NVT ensemble (300 K) for a total of 2 microseconds with 4 replicas (500ns/replica). For all simulations, long-range electrostatics were treated with particle-mesh Ewald summation.[53] A cut-off of 1.0 nm was used for Lennard-jones and coulombic interactions. All bonds between hydrogens and heavy atoms were constrained using the LINCS algorithm[52] allowing for a simulation time step of 2 fs. All simulations were conducted using GROMACS 2016.[117]

BIASING SCHEME

We used the parallel bias metadynamics[33] scheme which allows for efficient sampling of such high-dimensional free energy landscapes. To ensure that all binding conformations of the peptide were explored, we biased 4 collective variable – 3 for structural elements of the peptide (radius of gyration, alpha-helical and 3_{10}-helix coordination numbers) and 1 for the distance of the peptide from the surface. The radius of gyration was calculated using all alpha carbon atoms. Coordination numbers were calculated using carbonyl oxygen and amide hydrogen contacts within 0.30 nm of the i^{th} and $i+4^{th}$ residues for the alpha-helix and i^{th} and $i+3^{th}$ residues for the 3_{10}-helix. The distance was calculated by using the z-distance between the center of mass of all alpha-carbon

atoms and a reference surface atom. The peptide was restrained to explore only distances below 2 nm so that the peptide only explores surface-bound structures. The parameters for the bias were – initial hill height of 1.2552 kJ/mol, hill deposition pace of 1 hill/500 steps, biasfactor of 10, and temperature 300 K. The collective variables had hill widths of 0.5 nm (radius of gyration and distance), and 0.1 (alpha-helical and 3_{10}-helix coordination numbers).

The combination of classical MD with an enhanced sampling scheme promotes the exploration of multiple conformation states of the peptide. However, the prevalence of specific conformations can be over-predicted or under-predicted according to the accuracy of the force field used.[11] To correct for these inaccuracies and ensure that experimentally relevant conformations are correctly represented in the simulation, we combined the parallel bias enhanced sampling method with metainference.[26,265] Using Bayesian inference, this approach allows for the NMR chemical shifts to be used as restraints on the predicted chemical shifts averaged over an ensemble of peptide conformations. Chemical shifts were calculated on-the-fly using CAMSHIFT[264] (as implemented in the PLUMED 2.4[175] library). Our criteria for assessing convergence, and demonstration of convergence of the calculations are discussed in the Appendix.

VALIDATION OF PEPTIDE STRUCTURES

After the production simulations are completed, the simulation trajectories are analyzed to find the most representative structures of the surface-bound peptide. Since the trajectories are large, every 50th frame is used for analysis, resulting in 20000 frames. Further, some frames were excluded from the trajectory as mentioned in Section 1 in the SI. Finally, the chemical shifts calculated at every frame. Backbone chemical shifts were calculated on the fly using CAMSHIFT[264] as implemented in PLUMED, whereas side-chain chemical shifts were calculated using SHIFTX[263] since CAMSHIFT does not provide side chain chemical shifts.

At the end of the simulation, we can use the accumulated bias potential as weights to calculate the free energy with respect to another variable (separate from the variables biased in the original simulation).[25] A weight (=$e^{\frac{V}{kT}}$, where kT = 2.5 kJ/mol at 300 K, V is the bias at each frame) is assigned to each frame using the bias potential as a quasi-static bias potential. These weights were used to calculate the weighted average chemical shift from the trajectory, following the protocol of Torrie and Valleau.[29] The root-mean-squared error of the backbone and side-chain chemical shifts, with reference to the experimentally determined values, were calculated using the formula:

$$RMSE = \left(\sum_{i=1}^{N} \left(\mathrm{CS}_{i,predicted} - \mathrm{CS}_{i,exp}\right)^2\right)^{0.5} \quad 7.1$$

DETECTION OF TOP-WEIGHTED PEPTIDE CONFORMATIONS

The structures of the peptide in the trajectory are identified and clustered using the gromos[199] method (RMSD calculation using alpha-carbon atoms with cut-off of 0.20 nm). Then, the cluster number assigned to each frame of the trajectory is reweighted with the above-mentioned weights.

DETECTION OF TOP-WEIGHTED PEPTIDE ORIENTATIONS

To study relaxation of the sidechains and orientation of the peptides on the surfaces, the top three weighted conformations for each surface were extracted and simulated for 50 ns in the NVT ensemble (300 K). The backbone was held rigid during the simulation using a restraint on the RMSD of the C-alpha carbons using a harmonic restraining potential (Value = 0.15 nm; Kappa = 10000.0). No other bias was added during these simulations. The trajectories were viewed in VMD to identify the residues that interact with each surface (see Appendix). For a quantitative analysis of contacting residues, the normal surface distance of side-chain atoms (Cγ, Cδ, Cε, and Cζ) from a reference surface atom was calculated for the whole trajectory. The probability of the

occurrence of the side-chain atom at a certain distance from the surface was calculated using the kernel density estimation technique (shown in Figure S7).

7.4 RESULTS

SNA15 SECONDARY STRUCTURE PREDICTIONS OF ADSORBED PEPTIDES FROM MD SIMULATIONS

Further, MD is used to generate an ensemble of conformations of the peptide consistent with the experimentally measured chemical shifts. The backbone ^{13}CO, $^{13}C\alpha$, and $^{13}C\beta$ chemical shift data for SNa15-HAP, SNa15-SiO_2, and SNa15-TiO_2 were used to restrain an ensemble of configurations using the metadynamic metainference approach[26] (See Methods). The restraints are applied to the average chemical shifts across many simulation replicas, thereby limiting the average deviation of the backbone chemical shift to be within the error predicted by experiments. The calculated side-chain chemical shifts are not directly used in the simulations or analysis of the clusters and provide *a posteriori* validation of our approach. Root mean squared errors (RMSE) of the backbone (training error) and side chain (validation error) for the surface adsorbed peptides are provided in Table 7-1. We see that the error in both the backbone and side chain chemical shifts are between 1.48 - 2.25 ppm, which denotes agreement with experimental data (experimental error ~2 ppm), providing strong evidence that the ensemble of peptide conformations generated by the metadynamic metainference simulation are consistent with the experimental data.

Table 7-1 The root-mean squared error of the backbone and side-chain

Simulation	RMSE of backbone	RMSE of side-chains
SNa15-HAP	2.17	1.59
SNa15-SiO2	2.25	1.75
SNa15-TiO2	1.91	1.48

We next determined the most likely conformations of the peptide generated by the MD

simulation by performing a clustering (See Methods). The center of the top three most likely clusters of peptide backbone conformations (and their weights determined from post-processing of the metadynamics simulations) along with the representative structure from TALOS-N are shown in Figure 10. For the case of SiO_2, all conformations are random coils with a helical middle section. Noticeably, the structure predicted by TALOS-N has the same structural motif. The comparisons of SNa15 on HAP generally show similar amounts of secondary structure, with more variance in the center number of helical turns. In contrast, the top 3 clusters of conformations predicted from MD for SNa15 TiO_2 predict fewer helical turns, and a similar structure to SiO2. In both TiO2 and HAP, the TALOS-N predicted structure displays a greater level of helicity (with unstructured ends). These discrepancies, when viewed in light of the fact that all MD predictions show ~ 2 ppm agreement within the backbone chemical shifts, demonstrates the need to view SNa15 on mineral and oxide surfaces as an 'ensemble' and not a single structure. We note that in all cases, the exact structure predicted by TALOS-N can be found within the biased MD simulations, although it does not appear in the top three reweighted cluster centers shown here.

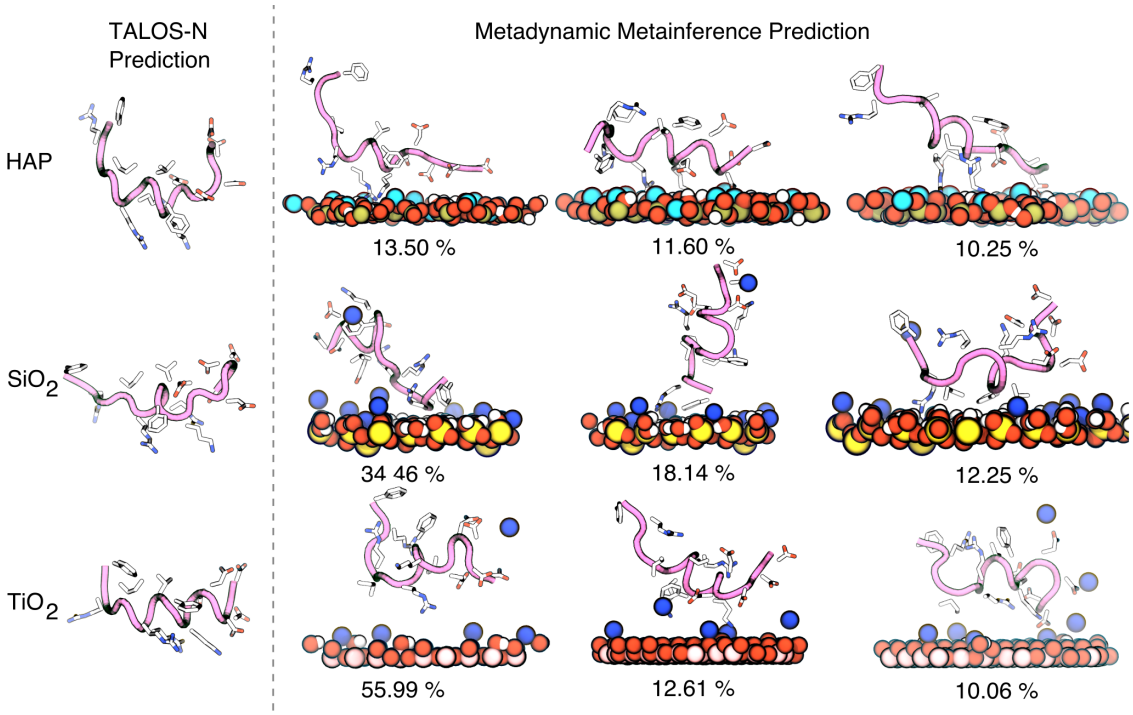


Figure 7-1 TALOS-N prediction and metadynamics metainference predictions of the structures of SNa15 on HAP (top), SiO_2 (center), and TiO_2 (bottom) surfaces. The peptide backbone is shown in mauve with tube structure. Surface atoms are colored blue for sodium ions, light blue for calcium, red for oxygen, yellow for silicon. The heavy side-chain atoms on the protein are rendered with licorice and colored blue for nitrogen, red for oxygen, and white for carbon. The percentages of the MD configurations refer to the weight of that cluster within the structural ensemble.

Further, to ascertain which residues are interacting with the surface, the top 3 clusters were placed near the respective surface and simulated for 50 ns (with the backbone restrained using an RMSD restraint on the Cα atoms) (see Appendix). For the case of the HAP surface, SNa15 is anchored to the interface by a mixture of basic (R9, K6, and R10) and acidic (D1, D2, and E5) residues. Since the surface is neutral overall, and has local pockets of positive and negative charge, it is understandable that the peptide needs both types of binding residues to bind to the surface. The TiO_2 surface is also neutral overall, such that both basic (R10, R13, and K6) and acidic (D2 and D3) residues are involved in binding SNa15 to the surface (see Appendix). F14 residues are

close to the surface. In contrast, the silica surface is negatively charged with mobile, displaceable sodium ions that compensate for the surface charge. In this case, SNa15 is anchored with basic residues only (R13 and R9) (see Appendix). Notably, neutral, hydrophobic residues like F14, F7, and I11 also approach the surface. On approach, F14 and F7 lay flat on the surface. Several residues show a bimodal distribution of contacts at the surface (D2, D3, K6, F7, R9, and F14) (see Appendix).

7.5 DISCUSSION

Both the structure of the protein and the interaction of its side-chains with the surface are integral to the process of recognition of biomineral surfaces by proteins. Expectedly, the foundational goals in this field are centered around the determination of protein secondary structure, the evaluation of changes in structure upon adsorption, and the estimation of the proximity or orientation of amino acids to the surface. To further these goals, we have chosen the model peptide SNa15 and elucidated its structure upon binding to three naturally occurring mineral surfaces (HAP, SiO_2, and TiO_2). In fact, we have demonstrated an integrated approach that pairs experimental data with MD simulations to make more robust predictions about the ensemble of surface adsorbed structures. Experimentally, we used 1D and 2D ssNMR techniques to determine the secondary structure of SNa15 in the pure, i.e. unadsorbed solid state, and when adsorbed to HAP, SiO_2, and TiO_2 surfaces. On the modeling front, we used the structure prediction software TALOS-N, and classical all-atom MD simulations (with metadynamic metainference) to generate an ensemble of conformations of the peptide on the surface, which are consistent with the experimental data.

Overall, the backbone (^{13}CO, $^{13}C\alpha$, and $^{13}C\beta$) chemical shift data and subsequent TALOS-N[267] predictions indicate that SNa15 is helical (with unraveling at the termini) in water, and when adsorbed to HAP and TiO_2. In contrast, the peptide is predicted to be largely unfolded on the SiO_2

surface. The picture, when viewed from the perspective of integrated MD+ssNMR is more nuanced, with a need to explore in more detail the full conformational ensemble.

Further, we explore the interaction of various amino-acid side-chains with the surface via the ΔCS for the side-chain ^{13}C spins. Notably, large downfield perturbations in the ΔCS of E4, E5, and F14 (in all cases) suggest that they are interacting with each surface. There are large downfield perturbations for the side-chain $^{13}C\gamma$ and $^{13}C\delta$ spins of E4 and E5. In fact, this CS perturbation is attributed to the decreased shielding of the carboxyl group due to its interaction with the surface.[277] Fernandez et al.[277] also noted similar downfield shifts in $^{13}C\delta$ spins (~ 2 ppm) in their study of poly-L-glutamic acid adsorbed on HAP and SiO_2. In our study, these downfield shifts range from 3-5.3 ppm indicating a stronger effect than that seen by Fernandez et al. Moreover, since the perturbations are roughly the same magnitude within error for both residues, this suggests a similar degree of contact with the surface. This contrasts with the predicted behavior of SN15 and statherin on HAP, where only E5 was proposed to interact with HAP.[245] In fact, MD simulations predict that E4 and E5 have a much higher degree of conformational flexibility (see Appendix) on HAP compared to SiO2 and TiO2. Notably, NMR-derived side-chain constraints were not included in the MD simulation, therefore some deviations in the side-chain behavior is expected.

Amongst the aromatic groups, we observe that F14 is also interacting significantly with the surface. The role of phenylalanine residues in the binding of statherin to HAP has been studied extensively.[241,246,249] Gibson and coworkers[241] studied the role of F7 and F14 in SN15 adsorption to HAP through isotropic chemical shift perturbations, $^{13}C\{^{31}P\}$ REDOR, and $T_{1\rho}$ relaxation measurements. They found that when compared to F7, F14 is closer to the surface and is more dynamically constrained. In fact, F7 is observed to be oriented away from the HAP surface.[241] The study by Weidner and coworkers[246] used NEXAFS and SFG also corroborates this observation. Correspondingly, in this study, we note large ΔCS for F14 and minimal ΔCS for F7. MD

simulations also predict the interaction of F14 with SiO_2 and TiO_2. In fact, the ^{13}CO and $^{13}C\alpha$ chemical shift data also show that F14 behaves differently in HAP than the other two surfaces.

In the case of basic residues, there are negligible perturbations for most of the side-chain ^{13}C spins for K6, R9, and R10. MD simulations also show that other basics residues (K6, R9, and R10) are away from the surface (see Appendix). This suggests that these residues do not specifically adsorb to or identify any of these surfaces. In fact, previous studies where the 4 basic residues (K6, R9, R10, R13) of statherin were mutated to alanine have shown that individually, none of these residues are crucial to binding.[250] However, as a group they act to decrease the overall charge of statherin, thus reducing repulsive protein-protein interactions and promoting a higher surface affinity and coverage.[250]

In contrast, the R13 $^{13}C\gamma$ and $^{13}C\delta$ spins experience significant perturbations in the range of -0.9 to -2 ppm. This suggests that R13 interacts with the surfaces while the remaining basic residues do not. From the end-view of TALOS-N predicted structures, it is evident that the R13 side-chain is oriented on the same face of SNa15 as E4, E5, and F14 (the residues that are hypothesized to interact with the surfaces). The behavior of R13 is corroborated with MD simulations in the case SiO_2, and TiO_2. However, in the case of HAP, where the C-terminus of the peptide is away from the surface, R13 is seen to have more conformational flexibility and generally further from the surface.

Finally, we observe that D2 and D3 are also crucial to the recognition of surfaces. In the case of HAP, D2 and D3 are oriented along the hypothesized binding face of SNa15. These residues experience a downfield shift of 1.2 and 1.3 ppm, respectively, upon adsorption. This is also true for adsorption on SiO_2, since D2 and D3 both experience a downfield shift of 2 ppm. However, in the case of TiO_2, only D2 is oriented along the binding face (predicted by a significant ΔCS in D2 only). From MD simulations, we see that D2 and D3 contribute to the binding of SNa15

to HAP and TiO_2. However, they do not bind to the silica surface. On silica, D2 and D3 have a bimodal (what does this mean?) contact preference.

Based on the information above, we hypothesize that the binding domain of SNa15 consists of some residues from N-terminus (D2-E5) and C-terminus (R13 and F14) residues for binding to HAP and SiO_2. For the case of TiO_2, D3-E5, R13, and F14 residues contribute to binding. MD simulations predict that SNa15 binds from the N-terminus on the HAP surface, from the C-terminus on the SiO_2, and from both termini on the TiO_2 surface. The residues composing the hypothesized binding domain are all oriented on the same face of SNa15 and all experience significant chemical shift perturbations upon adsorption. In fact, the lack of α-helical character in the N-terminus allows consecutive acidic residues to be positioned along the same face and interact completely with the surface.

Since SNa15 is comparable to other model peptides (SN15) and its parent protein statherin, it is possible to compare the behavior of SNa15 with previously reported observation about the other peptides. In fact, prior studies of statherin and SN15 on HAP were based on distance measurements obtained from dipolar couplings between ^{13}C and ^{15}N spins within the peptides and between ^{13}C spins in peptide side chains and ^{31}P spins in the HAP surface. Since these studies and the current studies results in predictions of the peptide structure, it is useful to compare these results, albeit qualitatively. For example, the distorted helical structure obtained for SNa15 on HAP, obtained by analysis of the ^{13}CO, $^{13}C\alpha$, and $^{13}C\beta$ chemical shifts via TALOS and MD simulations, is in agreement with earlier dipolar coupling-based studies of the structure of SN15 and statherin on HAP.[15,31,37] Our present study of SNa15 and prior studies of statherin and SN15 find that the acidic N-termini of SNa15 and SN15 interact strongly with the HAP surface, but details of exactly which side chains are involved in surface interactions differ. Downfield changes to the chemical shifts of the γ and δ carboxyl ^{13}C spins in SNa15's $D_2D_3E_4E_5$ moiety observed

upon adsorption to HAP indicate interactions with the HAP surface. Relaxation and chemical shift line shape studies of pS2 and pS3 in SN15 on HAP similarly find that these N-terminal amino acids are anchored to the surface[36], but a $^{13}C\{^{31}P\}$ REDOR study finds that the $^{13}C\delta$ spin of E5 is much closer to the HAP surface than the carboxyl group of E4.[19] Based on $^{13}C\zeta$ upfield chemical shift changes upon adsorption to HAP, only R13 is assumed to interact with the HAP surface in SNa15, while a $^{13}C\{^{31}P\}$ REDOR study of statherin on HAP finds that the$^{13}C\varepsilon$ spin of K6 and $^{13}C\zeta$ spins of R9 and R10 are close to the HAP surface. Some of these differences may result from variations in local secondary structure, where for example the substitution of D for pS at positions 2 and 3 in SNa15 may perturb the local helical structure making it possible to simultaneously expose E4 and E5 in SNa15 to the surface. The apparent differences between exposure of basic amino acids in SN15 and SNa15 to the HAP surface may similarly result from structural changes in the acidic portion of the N-terminus that perturb the surface orientation of the basic amino acids.

Another factor to consider is the effect that the orientation of the side chain has on ΔCS values. In a study of SK rich peptides in SiO_2 and TiO_2 composites, density functional theory (DFT) calculations showed that the chemical shift of $^{13}C\varepsilon$ spins in lysine side chains has a large upfield ΔCS when the side chain is oriented perpendicular to the surface, but the ΔCS is greatly diminished for $^{13}C\varepsilon$ when the side chain is oriented parallel to the surface.[278] Also, for lysine $^{13}C\varepsilon$ spins ΔCS values drop off rapidly as a function of the distance between the amine group and the surface. For example, when the distance from the lysine side chain amine to the SiO_2 surface exceeds 3 Angstroms ΔCS perturbations are comparable to experimental error.

7.6 Conclusion

This study of the adsorption of SNa15 onto HAP, SiO_2, and TiO_2 surfaces demonstrates an integrated approach that pairs ssNMR with advanced MD simulations as a means to predict

ensembles of surface-bound peptide structures. Measured adsorption isotherms and companion structure predictions from the software TALOS-N provide additional characterization of the systems. In contrast to using TALOS-N to predict a single set of dihedral angles, the integrated approach allows for structure prediction of an entire conformational ensemble with average backbone chemical shifts constrained to the measured experimental values. This approach is thus a promising way to estimate not only the conformational ensemble, but also the key binding residues to the surface and orientation of the adsorbed biomolecule.

Although this method can be applied to many peptide/surface systems in the future to study biomineralization processes, the approach has potential areas for future investigation. First, there is a continued need to improve surface potentials and representation of accurate surface chemistries for use in MD simulations of peptide adsorption, even with metadynamic metainference approach to help overcome systematic errors. Second, DFT can be used to provide accurate estimates of the surface response of amino acids such as those shown here and to help provide an assessment of the accuracy of using the SHIFTX and CAMSHIFT methods for peptides on surfaces.

7.7 ACKNOWLEDGEMENTS

GPD acknowledges National Institutes of Health Grants RO1-GM109417, GPD and JP acknowledge grants NIH R21 A126113 and National Science Foundation Grant MCB-1715123. This work was facilitated using computational, storage, and networking infrastructure provided by the Hyak supercomputer system, supported in part by the University of Washington and the UW Student Technology Fee Proposal program and NSF MRI program CHE-1624430. GPD also acknowledges support from sub-contract ANSK-0119-16 and helpful conversations with Professors Havard J. Haugen and S. Petter Lyngstadaas of the Department of Biomaterials, Institute for Clinical Chemistry University of Oslo.

Chapter 8. SPECIATION OF IONS AT THE MUSCOVITE INTERFACE IN DILUTE AND CONCENTRATED ELECTROLYTES

Arushi Prakash[1], Jim Pfaendtner,[1,2], Christopher J. Mundy[,1,2]*

[1]Department of Chemical Engineering, University of Washington, Seattle, WA. [2]Physical and Computational Sciences Directorate, Pacific Northwest National Laboratory, Richland, Washington, United States.

8.1 ABSTRACT

Molecular-scale insights into the adsorption of ions on the surface of muscovite is obtained using molecular dynamics simulations and metadynamics enhanced sampling. The preferences of different ions (Na^{1+}, K^{1+}, and Ca^{2+}) for different adsorption sites are explained in terms of ion size, hydration, and entropy. Moreover, significant differences are observed in the binding of ions to mica at different concentrations. The free energy profiles of the binding of ions to the surface, resolved using metadynamics, reveal complex interdependencies between concentration, ion-ion correlation, and hydration.

8.2 INTRODUCTION

Electric double layers (EDLs) are important in diverse fields like colloids science, geochemistry, electrochemistry, and biophysics.[279] Broadly, the formation of an EDL refers to a separation of charge in a system. EDLs are formed in the presence of a charged surface. For example, when mica is immersed in water, the intrinsic surface ions get hydrated but remain

associated with the negatively-charged surface due to electrostatic interactions.[280] This charged surface and associated layers of ions form the electric double layer. The structure, and dynamics of this layer significantly effects how two interfaces interact with each other.

Consequently, several studies have tried to elucidate the structure of the EDL. In the case of mica, the EDL has been imaged using atomic force microscopy (AFM).[281] Further, X-ray reflectivity (XRR) experiments combined with modelling and simulations, have been used to resolve the adsorbed states of different ions on the surface of mica.[282] These studies propose that ion speciation relies heavily on the interplay between ion and surface hydration.[283–285] This leads to three types of surface adsorbate species - inner-sphere surface complexes (ISSC), outer-sphere surface complexes (OSSC), and diffuse swarm (DS) species.[286] Further, although electrostatic interactions keep ions associated at certain distances from the surface, the configuration in each layer is hypothesized to be governed by the hydration of ions.[281,287–290]. In fact, hydrated ions seem to be correlated which causes them to accumulate in patches on the surface.[281] Notably, all these studies are phenomenological, calling for theory to provide a unifying picture. However, theoretical treatments of the EDL, using Derjaguin, Landau, Verwey, and Overbeek (DLVO) theory, have not been able to account for such effects due to oversimplifying assumptions.[285] Consequently, one looks to modeling and simulation efforts for predictions about the EDL.

Most modeling studies of the electrical double layer of mica accompany XRR studies, where the electron density derived from experiments is fit to a model.[291–294] These studies help ascertain the adsorption heights of different ions from the surface. Further, molecular dynamics (MD) simulations of the mica interface are also common. In fact, MD simulations can provide lateral resolution which is usually averaged during experiments.[281] However, MD simulation efforts have largely focused on the fate of water on the surface.[69,295–298] Additionally, most

simulations that focus on the behavior of ions at the mica surface only treat mica in contact with pure water. [287–290,299]

In nature, clay minerals are usually in contact with electrolytes at finite concentrations, with charges concentrations ranging between 0.07 – 4.1 M.[286] The EDL is different at these concentrations and simulations of mica near pure water may not be a good approximation at other electrolyte concentrations.[83] Noting this, Adapa and Malani exhaustively characterized the adsorption of counterions on mica at finite electrolyte concentrations using MD simulations.[300] They observed differences in adsorption in response to the size, hydration, and concentration of ions.

Similarly, this study uses MD simulations to address the critical question of how ion adsorption on the surface of mica is affected by changes in the concentration of electrolytes in contact with the surface. Metadynamics enhanced sampling is used to provide free energy estimates for ion adsorption. In the dilute case, we elucidate how ion hydration, and the identity of surface sites affect adsorption of single ions on the surface of mica. Moving to finite electrolyte concentrations (1m, 2m, and 3 m), we examine how ions form wider EDLs and explore the major factors governing the structure of these EDLs.

8.3 METHODS

SYSTEM SETUP

Similar to the study by Prakash et al.,[298] an ideal mica (001) surface was created with K^{1+} ions on surface cavities. Si were substituted with Al in accordance with Lowenstein's rule (to avoid Al–O–Al linkages). This resulted in a surface of size 20.756 Å × 17.990 Å, with a 20 Å layer of mica. A 30 Å thick water layer was added which included 384 water molecules. This represented the infinitely dilute case. Water ions were replaced by 9 K^{1+} and 9 Cl^{1-} ions, or 17

K^{1+} and 17 Cl^{1-} ions, or 25 K^{1+} and 25 Cl^{1-} ions to represent 1 m, or 2 m, or 3 m electrolyte solution of monovalent ions in contact with the mica surface. The clay atoms and ions were modelled using the CLAYFF[48] force field. The water was modelled using the SPC[171] water model. These systems were charge neutral.

For simulations where a single ion of type Na^{1+} or Ca^{2+} was biased from the mica surface, a K^{1+} ion on the mica surface was substituted for the respective ion. When K^{1+} was substituted with Ca^{2+}, and additional Cl^{1-} ion was added to the solution to ensure that the system remained charge neutral.

MOLECULAR DYNAMICS SIMULATION

GROMACS 2016 was used for all simulations. After generating a simulation box as described above, the energy of system was minimized using the steepest descent method. Then, we stabilized the pressure of the system in the NPT ensemble using a Berendsen barostat[50](15) (tau = 1.0 ps; pressure = 1 bar; semiisotropic pressure coupling allowing only the z-component of the box to relax) and a Donadio–Bussi–Parrinello thermostat[51] (tau = 0.1 ps, temperature = 300 K) for 1 ns. Finally, for the production run, we simulated the system in the NVT ensemble using the Donadio–Bussi–Parrinello thermostat[51], as mentioned above, for 500 ns. All mica atoms, except the ions on the surface, were frozen during the production run.

During the simulation, all bonds between hydrogen and heavy atoms were constrained using the LINCS[52] algorithm, which allowed us to use a timestep of 2 fs. For all simulations, electrostatics were calculated using the particle mesh Ewald (PME)[53] summation method using a short-range cutoff of 0.8 nm. A van der Waals cutoff of 0.8 nm was used.

To generate the velocity autocorrelation function, the system representing infinite dilution was propagated in the NVE ensemble for 10 ns. Frames were written out every 20 fs.

BIASING SCHEMES

PLUMED 2.4.1[175] was used for all biasing schemes. To assess the differences in ion binding to the surface, the effect of temperature, and water coordination, a single ion was biased on and off the basal surface of mica. Well-tempered metadynamics[23] was used to bias the z-component of the distance of the ion from the surface. The biasing parameters were – biasfactor 5, temperature 300 K, initial hill height 1.25 kJ/mol, hill deposition pace 1 hill/ps, and sigma 0.01 nm. To test binding at different temperatures, the temperature for biasing was changed to represent the temperature of the system.

To assess the differences in binding due to different electrolyte concentrations, all ions (intrinsic ions and electrolyte ions) were biased. Parallel-bias metadynamics with partitioned families[25] was used to bias the z-component of all ions from the surface and the number of waters around the ion. All distance variables contributed to the first family, and all water coordination variables contributed to the second family. The biasing parameters were – biasfactor 5, temperature 300 K, initial hill height 1.25 kJ/mol, hill deposition pace 1 hill/ps, and sigma 0.01 nm for distance and 0.1 for coordination number. The coordination number for the i^{th} ion was calculated using the formula:

$$s_i = \sum_{j=1}^{N_{water}} \frac{1 - \left(\frac{r_{ij} - d_0}{r_0}\right)^n}{1 - \left(\frac{r_{ij} - d_0}{r_0}\right)^m}$$

Where N_{water} represents the total number of waters in the system, n =12, m = 24, d_0 = 0.5 nm, and r_0 = 0.315 nm. To lower the computational cost, neighborlist cutoffs were set (NL_CUTOFF = 0.55 nm).

ANALYSIS

The velocity autocorrelation function was calculated, from NVE simulations, using each component of the velocity of oxygens in water molecules. Only water molecules that resided within a region for at least 60 ps were retained for the calculation. Once a molecule exited the

region, it was treated as a new molecule upon re-entry. The velocity autocorrelation was calculated separately for each molecule using the in-built *analyze* tool in GROMACS. The profiles generated from this calculation were averaged to generate the final auto-correlation function for each region.

To calculate a free energy estimate with respect to another variable, a weight for each frame was calculated using the Torrie-Valleau[29] method. These weights were calculated using the formula $W = \exp(\beta V)$ where, V = bias from the metadynamics potential, $\beta = 1/k_B T$ (T = temperature, k_B = Boltzmann's constant), W= weight of each frame. To calculate V, the simulation was rerun using the GROMACS command *mdrun rerun* and bias deposited during the production run is treated as a quasi-static potential. The Gaussian deposition pace was set to 1000000 during rerun to prevent the deposition of new Gaussian hills, which would change free-energy estimates. The value of the new observable was also calculated for each frame. A weighted histogram of the variable was then constructed using the above-mentioned weights. The histogram was normalized to obtain a probability distribution, which was converted to a free energy estimate using the formula $F = - k_B T \ln p$, where p is the probability of that value of the observable. This reweighting method was used since it has been shown to exactly reproduce the underlying free energy surface.[25]

8.4 Results and discussion

Ion-surface binding in the dilute electrolyte limit

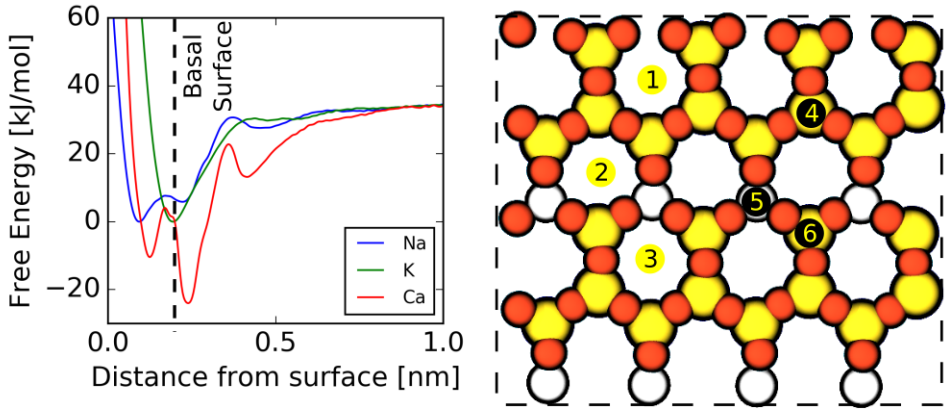


Figure 8-1 (left) Free energy [kJ/mol] of ions as a function of their distance from the basal surface of mica (right) Adsorption spots on the surface of mica for K^{1+}/Na^{1+} ions (yellow, numbered circles) and Ca^{2+} ions (black, numbered circle).

To explore the binding characteristics of different ions on the surface of mica, the distance of a single ion from the surface was biased using well-tempered metadynamics resulting in the free energy profiles in Figure 8-1. Notably, the three ions prefer to adsorb at different distances from the surface and on different locations on the surface. Their preference for the distance from the surface is evaluated from the free energy profile in Figure 8-1(left). Due to the small size of Na^{1+} and Ca^{2+}, both of these ions can intercalate inside the surface, thereby exhibiting adsorption minima below the basal surface. Owing to the slightly smaller size of the Na^{1+} ion, when compared to the Ca^{2+} ion, it intercalates further into the surface. In this position, both ions are almost completely dehydrated, and their charge is being compensated by the oxygen atoms of the clay. In fact, AFM measurements by Nishimura et al.[301] also showed that small ions (Li^{+} and Mg^{2+}) were likely intercalated, therefore fixed on the surface, while K^{+} ions were mobile and could not be fixed on the surface.

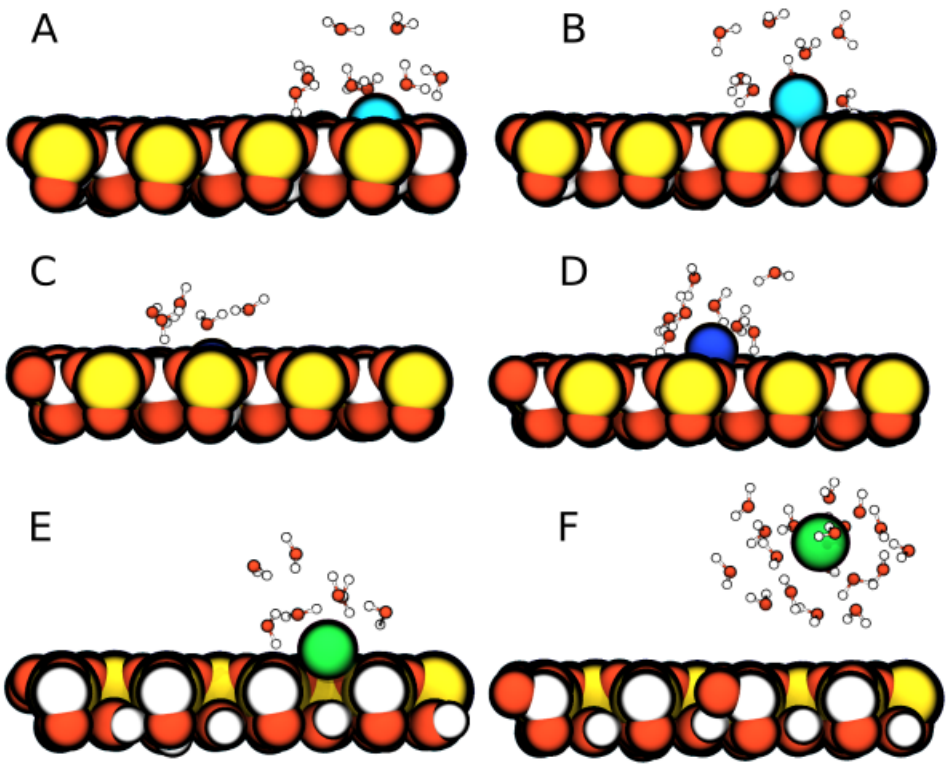


Figure 8-2 Snapshots of ions – Ca^{2+} (A and B), Na^{1+} (C and D), and K^{1+} (E and F) – adsorbed onto the basal surface of mica. $Ca^{2+}$and Na^{1+} are shown in two adsorbed states, while K^{1+} is shown in one adsorbed state, and solution state. Atoms are colored red for oxygen, light blue for calcium, dark blue for sodium, green for potassium, white for hydrogen, yellow for silicon, and white for aluminum.

Further, Na^{1+} and K^{1+} ions show preference for adsorption onto the ditrigonal cavities at the basal surface (points marked 1,2,3 in Figure 1(right) and Figure 8-2). At this site, they are partially hydrated and the oxygens on the basal surface complete the other part of their solvation shell. In fact, Na^{1+} sits slightly further from the surface than K^{1+}, to be closer to waters that form its partial hydration shell. Notably, Ca^{2+} sits even further from the surface (Figure 8-2 B), atop a triad of oxygens. Here, the substitution of Si by Al creates a negative site that attracts the highly charged Ca^{2+} ions. Further, for both Na^{1+} and Ca^{2+} ions, adsorption to the surface from the bulk proceeds via an activation barrier. This energetic barrier results from losing waters from part of its solvation shell to incorporate oxygen atoms from the basal surface of mica. Ca^{2+} faces a larger

barrier (12.5 kJ/mol) than Na^{1+} (3 kJ/mol) since it has a higher hydration free energy. In fact, the experimentally determined hydration energies for these ions (–364.97 kJ/mol for Na^{1+} and –1504.98 kJ/mol for Ca^{2+})[302] have the same ratio as these barrier heights, providing more proof that this barrier originates from the loss of water.

Overall, from these free energy profiles, it appears that Na^{1+} ions prefer intercalating into mica, K^{1+} ions prefer adsorbing onto ditrigonal cavities at the basal surface and Ca^{2+} ions prefer adsorbing onto oxygen triads on the surface of mica. Molecular dynamics simulations by Adapa and Malani also showed the same trend.[300]

The free energy profiles can also be used to characterize the strength of adsorption of ions. It is seen that Ca^{2+} very strongly adsorbs to the interface, as evidenced from the downward curve (0.7 nm from the origin in Figure 8-1(left)) in the free energy profile much before the other ions start showing attraction to the surface. Furthermore, Ca^{2+} binds strongest to the surface, while Na^{1+} and K^{1+} bind with almost equal binding free energies. The strong binding of Ca^{2+} is also corroborated by other studies.[303] In fact, the binding of Ca^{2+} can make the mica surface more hydrophilic or increase the wettability, compared to Na^{+} or K^{+} ions on the surface.[304] Furthermore, K^{1+} is more strongly bound to the ditrigonal cavity than Na^{1+}, as also seen in other studies.[305,306] Adapa and Malani also observe that Na^{1+} and divalent ions have two free energy minima, while K^{1+} has a single free energy minimum.[300] Notably, ionic polarizabilities have also been implicated in the ion-specificity.[285] However, these simulations do not account for that.

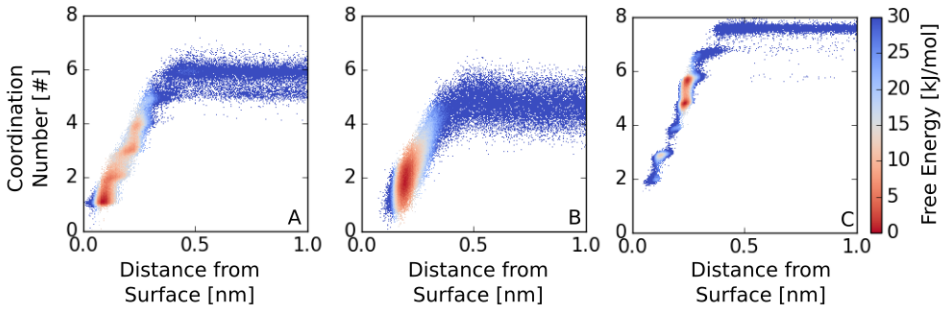


Figure 8-3 Free energy [kJ/mol] as a function of the number of waters coordinated around ions and the distance of ions from the basal surface of mica for (A) Na^{1+} (B) K^{1+}, and (C) Ca^{2+} ions.

These differences can be explored further by tracking how the waters are removed from the coordination shell of these ions as they adsorb onto the surface from the bulk solution (Figure 8-3). It is seen that water coordination numbers of Ca^{2+} fluctuate in a much narrower range than Na^{1+} or K^{1+}. This clearly highlights that the waters are more tightly bound to Ca^{2+} than Na^{1+} and very loosely bound to K^{1+}, mimicking experimentally known hydration energy trends.[302]

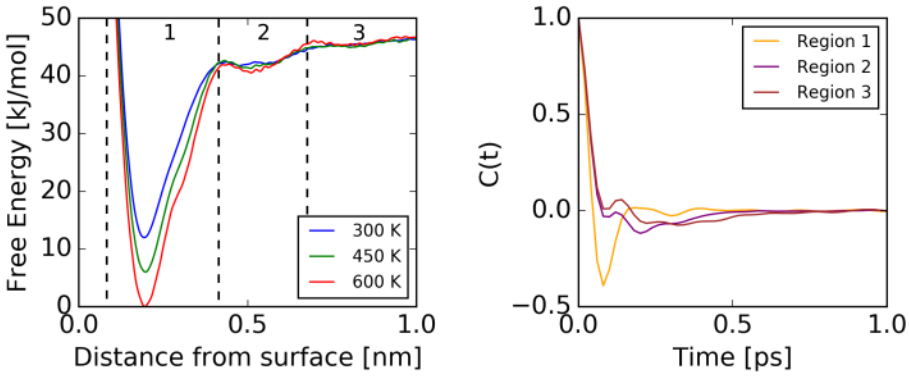


Figure 8-4 (left) Free energy [kJ/mol] of K^{1+} ion as a function of its distance from the basal surface of mica. (right) velocity autocorrelation function of the z component ($VACF_z$) of water in different regions near the surface, as shown in the left plot.

Apart from ions affecting the water, the mica surface also has an effect on water, which can in turn affect ion adsorption onto the surface of mica. Like any ion, the mica surface also considerably affects the motion of water in its first solvation shell (waters within 0.35 nm from the surface) as evident from Figure 8-4(right). Figure 8-4 (right) shows the auto-correlation of the z-component of the velocity ($VACF_z$) of water (considering oxygen atoms of water only). The deep minimum in the $VACF_z$ for Region 1 indicates rigidity of water molecules due to their strong interaction with the mica surface. This ice-like rigidity of water next to the mica surface is well-documented in experiments and simulations.[69,295–298] Region 3 corresponds to bulk water behavior. Region 2 has the same shape as Region 3, indicating that the fastest motions are taking place at the same timescale. However, the slight dip in the curve for Region 2 shows that waters are slightly more correlated. In fact, autocorrelation curves calculated by Lee and Rossky[70] showed identical results for water between silica surfaces.

Commonly, the contribution of entropy to ionic hydration is smaller than the contribution of enthalpy.[307] However, it is still interesting to note how entropy might play a role in the adsorption process. In Figure 8-4 (left), we see the free energy curves for the adsorption of K^{1+} onto the mica surface at different temperatures. We see that Region 1 shows the greatest change in free energy between temperatures, when compared to other regions. This is also the region where water has lost significant translational freedom, as shown in the autocorrelation plots. This leads us to make the hypothesis that the entropy of water contributes to the free energy of binding and this contribution is a function of the distance from the surface. Ion binding in Region 2 and 3 are almost unperturbed by temperature. The same trend is seen in the free energy profiles of Na^{1+} and Ca^{2+} (see Appendix).

ION-SURFACE BINDING IN FINITE CONCENTRATION ELECTROLYTES

The previous calculations were conducted in the limit of infinite dilution (mica in contact with pure water). However, most real-world applications involve finite electrolytes and the behavior at infinite dilution may not hold. To understand how this binding behavior changes at different concentrations, the system is simulated with at different electrolyte concentrations, and ions are biased from the surface using metadynamics (PBMetaDPF method). The free energy profile of the binding ions at each concentration is plotted in Figure 8-5A.

First, at 0 m, the K^{1+} ions are strongly attracted to the surface such that they feel this attraction much further in the bulk (0.9 nm). The adsorption to the surface is barrier-less, and the predicted binding free energy to the surface is -33.6 kJ/mol (the adsorbed state is considered at 0.20 nm, and the bulk state at 1.2 nm). On the other hand, binding energies predicted at 1 m, 2 m, and 3 m are -16.3 kJ/mol, 2m -7.78 kJ/mol, and 3m -11.3 kJ/mol, respectively. In all cases, the ions are predicted to bind to the surface. However, the binding energy of K^{1+} to the mica surface in the dilute limit is much higher (by a factor of 2 or more) because there are no excess ions to screen the negative charge of the basal surface of mica, leading to greater attraction with the surface. Predictably, with the presence of more ions than the intrinsic charge of the mica surface, the binding energy decreases.

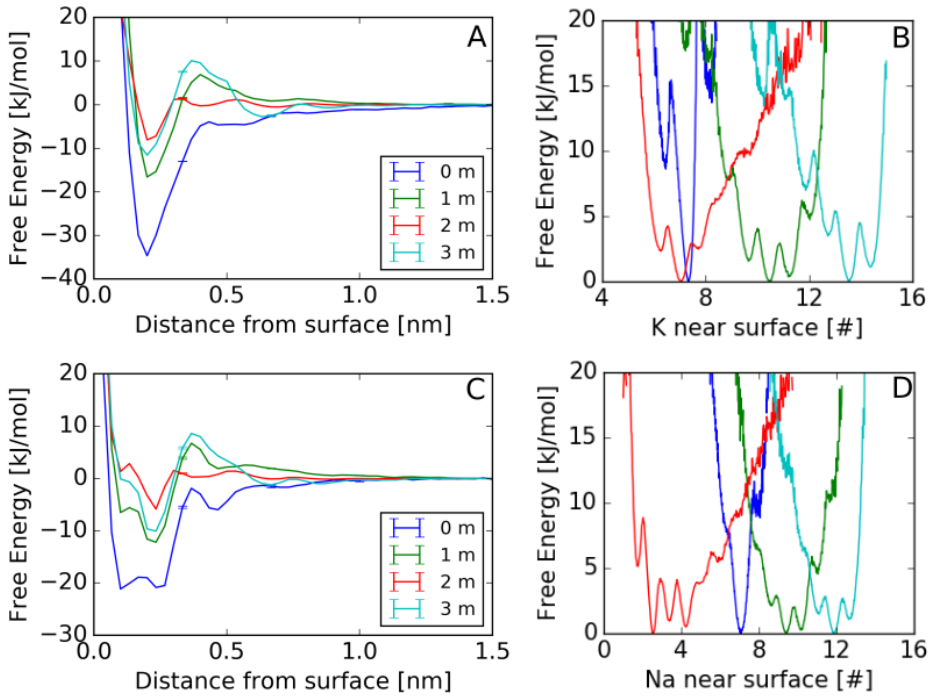


Figure 8-5 (A) Free energy (kJ/mol) of ion binding from the mica surface for different electrolyte concentrations. (B) Number of ions within 0.35 nm from the reference atom on the surface. The top panel represents K^{1+} ions, and the bottom panel represents Na^{1+} ions.

Interestingly, there are some unintuitive trends in the binding energy at finite concentrations. The free energy curves do not transition smoothly from the shape at 0 m to shape at 3 m. Specifically, at 0 m and 2 m, the free energy profiles look similar, since both these curves display barrier-free adsorption to the surface. On the other hand, the free energy profiles at 1 m and 3 m look similar since they display a significant barrier to adsorption of K^{1+} ions at 0.4 nm. Moreover, the binding curves after the activation barrier are almost identical.

We propose that these differences in the shapes of the free energy profiles are indicative of ion-ion correlations in the ions absorbed to the surface of mica. On calculating the number of ions within 0.35 nm of the reference surface atom, it is seen that at 0 m and 2 m, the surface accommodates ~ 7.5 atoms, which compensates for the intrinsic surface charge (-8) for our

model mica surface. On the other hand, at 1 m and 3 m, the surface accommodates 10.5-11.5 ions, and 13.5-14.5 atoms, respectively. When the surface accommodates only enough ions as the intrinsic charge of the surface, there is no activation barrier faced by the ions to adsorb on the surface. However, to accommodate more than 8 K^{1+} ions, the surface atoms need to be reconfigured to accommodate more ions, which possibly leads to an activation barrier in the free energy profile at 0.4 nm. At 3 m, the surface accommodates more charges, leading to a greater activation barrier to adsorption than at 1 m. A recent study also alluded to ion-ion correlation at the mica surface.[281]

Subsequently, due to the accommodation of more than 8 K^{1+} ions on the surface, the mica surface now has excess positive charges. This leads to a charge reversal of the surface. Chloride ions now piggy-back on the surface adsorbed K^{1+} ions to compensate for this charge, leading to a second layer of adsorbed ions on the surface. Since the second layer has negatively charged ions on the surface, there is a possibility to form a third layer of positively charged ions. This third layer is evident in the free energy profile at 3 m, which exhibits a free energy minimum at 0.70 nm. Notably, the surface at 1 m will also have chloride ions in the second adsorption shell but since the simulation does not have more K^{1+} ions, there is a no energy minimum at 0.70 nm to form a third adsorption layer.

A similar trend is seen in the case of Na^{1+} ions (Figure 8-5; bottom panel). Similar to the previous case, the free energy profiles at 1 m and 3 m follow each other closely and accommodate higher number of ions on the surface. In this case, at 0 m, the surface still accommodates close to 8 ions. However, at 2 m, the distribution is fairly different, but the free energy profile has more similarities to 0 m, than the other concentrations.

8.5 CONCLUSION

In this this study, the adsorption of ions (Na^{1+}, K^{1+}, and Ca^{2+}) onto the basal surface of mica is explored in different concentrations using MD simulations and metadynamics enhanced sampling. Several factors that affect ion adsorption are highlighted to demonstrate the complexity of the process.

In the case of single ions adsorbing to the surface in the limit of infinite dilution, it was shown that the size, charge, and hydration of the ions play important roles in determining the distance of adsorption from the surface and the location of adsorption on the mica surface. Ca^{2+}, which has more charge, exhibits a greater binding energy. Moreover, due to their small sizes, Ca^{2+} and Na^{1+} ions intercalate into the surface. Since both Ca^{2+} and Na^{1+} ions are strongly hydrated, they adsorb further from the surface (when adsorbed on surface spots) than K^{1+} ions. Further, it was seen that K^{1+} and Na^{1+} ions adsorb on ditrigonal cavities on the surface, while Ca^{2+} ions adsorbs on oxygen triads. Additionally, it was shown that water is retarded near the surface of mica, and this might affect the binding energy of the ions through entropic contributions. This trend was observed for all ions.

Finally, mica was simulated in contact with electrolytes at finite concentrations. The resulting free energy profiles of ions binding to the surface were different from that at infinite dilution. This result showed that some aspects of adsorption at infinite dilute may not translate to other concentrations. Moreover, the changes in the free energy profiles were not directly correlated to the concentration. We hypothesized that the configurations of surface-adsorbed played an important role in determining how ions bind to the surface. To that end, it was seen that similar number of ions are accommodated on the surface at 0 m and 2m, and at 1m and 3 m, leading to similarities in free energy profiles at these concentrations. Moreover, at higher

concentrations (2 m and 3 m), ions overcharge the suface such that chloride ions are able to form a second layer and eventually positive ions form a third layer at 3 m.

Through single ion and multiple-ion biasing strategies using metadynamics and MD simulations, this study was able to highlight how factors like the size, charge, and hydration of ions, the concentrations of electrolytes, and the adsorptions sites on the surface can affect adsorption and monolayer formation on the surface of mica. These results provide a fundamental understanding of how electrical double layers are formed, which is central to many fields like geochemistry and biochemistry.

8.6 ACKNOWLEDGEMENTS

This work was supported by the MS3 (Materials Synthesis and Simulation Across Scales) Initiative at Pacific Northwest National Laboratory (PNNL), a multi-program national laboratory operated by Battelle for the U.S. Department of Energy. This work was facilitated through the use of computational, storage, and networking infrastructure provided by the Hyak super-computer system, supported in part by the University of Washington.

Chapter 9. ELUCIDATING THE EFFECT OF IONS ON THE CLUSTERING OF AMPHIPHILIC BIOPOLYMERS

Arushi Prakash[1], Christopher J. Mundy[1,2], Jim Pfaendtner[1,2,]*

[1]Department of Chemical Engineering, University of Washington, Seattle, USA.

[2]Physical Science Division, Pacific Northwest National Laboratory, P.O. Box 999, Richland Washington 99352, USA

9.1 ABSTRACT

The effect of ions on protein clustering is well-documented experimentally, leading to their characterization as kosmotropes and chaotropes. However, the molecular underpinnings of such an effect remains elusive. To this end, we systematically studied the behavior of amphiphilic biopolymers (peptoids) in different electrolytes (+1, +1.5, +2 charges), representing different charge and charge-screening conditions, using classical molecular dynamics simulations. In the case of an isolated peptoid, ions have minimal effect on the structure beyond what is dictated by the side-chains. In the case of dimerization, ions affect the free energy of association if the association is electrostatically driven. Finally, in the clustering of several peptoids in a box, the effect of ions is visible although the description of states becomes increasingly complex with greater number of peptoids.

9.2 INTRODUCTION

Protein aggregation is important in the etiology of disease, functioning of enzymes, and formulation of drug products.[308] Consequently, this is a well-studied phenomenon with several excellent reviews on the topic.[308,309] In the area of modeling protein aggregation, traditionally, colloid theories, like the Derjaguin–Landau–Verwey–Overbeek (DLVO) theory, are adapted for this purpose. But such continuum, isotropic models fail to correctly represent crowded systems where proteins aggregate and self-assemble.[310]

Coarse-grained models improve upon these models by treating proteins as patchy particles or rigid-spheres, some with surface binding sites. However, it was recently shown that in cases where proteins undergo conformational changes during aggregation, such models cannot capture important behavior.[311] On the other hand, all-atom models used in molecular dynamics

(MD) can provide high-resolution into the aggregation process. These all-atom models can provide vital insight into changes in conformation, inter-molecular contacts, and orientation of the protein. They can also explicitly account for the interaction of the protein with water and co-solutes in the simulations, which is lacking in continuum and coarse-grained model.[16]

Since all-atom MD simulations are computationally expensive, they are commonly used to study early-stage cluster formation which require the simulation of a few proteins only. Frequently, these simulations have been performed by placing multiple peptides in a box and allowing them to aggregate spontaneously during the course of the simulation. Subsequently, information about the structure, and the kinetics of aggregate formation are accrued using Markov State Models[312] or Kinetic Transition Networks.[313] Although such MD simulations are prone to finite size effects,[314] they have been able to provide some valuable insights into the aggregation process. Atomistic simulations of amyloidogenic peptides, implicated in Alzheimer's disease, were used to highlight a new pathway for oligomer formation by a two-stage lock-and-dock mechanism.[315] Pathways for the formation of beta-sheet rich oligomers of insulin peptide segments and the commonly occupied structural ensembles were also detected using MD simulations.[316]

Notably, these simulations were focused on the structure and kinetics of protein aggregates, while the role of ions was ignored. However, ions play a central role in experimental studies of protein aggregation[90] To this end, when Batoulis et al.[317] simulated the oligomerization of a segment of SNAP25 protein in varying calcium concentrations, they observed that oligomerization was favored until a particular calcium concentration after which the oligomers dispersed due to overcharging effects. Therefore, ions cannot be treated as mere neutralizing agents in simulations that are used to study protein aggregation.

Herein, the effect of ion charge and concentration on the early-stage cluster formation of peptoid polymers is studied using classical MD simulations and enhanced sampling methods. Since peptoids are essentially poly-N-substituted glycines, studying their cluster formation behavior may also provide insights into protein aggregation. This paper presents a multi-step approach, by simulating monomers, dimers, and oligomers, to highlight how a charged environment affects early-stage clustering process. Peptoids are studies in different types of ion environments are studied, i.e. potassium (+1 charge) and calcium (+1.5, and +2) ions with variations in concentration. This study intends to highlight the non-negligible role that ions in simulations of clustering biomolecules. Further, the study seeks to highlight how analyzing the behavior of the same biomolecules in different ionic environments can reveal their clustering mechanism.

9.3 METHODS

FORCEFIELD PARAMETERS. Forcefield parameters proposed by Mirijanian et al.,[112] and later used by Prakash et al.[318] were used to simulate peptoids. Charges for all peptoids were derived from the RESP[113] method using Antechamber[114] program. These charges were calculated after geometry optimization and energy calculations using Hartree–Fock (HF) level of theory and 6-31G(d)//6-31G(d) basis set in Gaussian 09 program[115]. Charges for Calcium ions were either +2+ or +1.5 (75% of +2) according to the scaling program proposed by Daily et al.[319] In the scaled monomer simulations, the charges for Cl were -0.75, while all other simulations did not use chlorine with scaling. Where noted, the charges on the carboxylate side-chain (gamma and delta carbons and hydrogens, and oxygen atoms) of peptoids were also scaled following Daily et al.[319] so that the charge on the total peptoid was scaled from +6 to +4.5 (75% of +6). Additionally, we calculated the Mulliken charges for carboxylate side-chain and calcium systems in vacuum and implicit solvent (water) using geometry optimization in Gaussian 09 using Hartree–Fock (HF)

level of theory and 6-31G(d)//6-31G(d) basis set. The average charge for calcium in both solvent systems were +1.46, and close to the proposed +1.5 charge proposed above. The SPC water model was used in all simulations.[124]

Classical simulations were performed on the GROMACS 5.1.2[116,117] simulation engine. Production runs were conducted in the NVT ensemble at 300 K temperature using the Bussi-Donadio-Parrinello thermostat[51] (tau = 0.1 ps). NPT runs were also conducted at 300 K and 1 atm, with the above-mentioned themostat and Berendsen barostat[50] (tau = 1.0 ps) with isotropic scaling. A timestep of 2 fs was used since the hydrogen bonds were constrained with LINCS[52]. Van der Waals interactions were shifted to 0 at 1.0 nm. Electrostatic contributions were calculated explicitly for distances under 1.0 nm, and using particle-mesh Ewald (PME)[53] summations over 1.0 nm. Energy minimization was conducted using the steepest descent algorithm.

After each simulation system was setup, we conducted energy minimization and NPT equilibration (1 ns). The production runs were conducted in the NVT ensemble (300 K) after equilibration are described below for different systems.

MONOMER SIMULATIONS. The systems were setup by constructing a box of size 10 nm x 10 nm x 10 nm with a single peptoid, solvent molecules, and ions. For production runs, each system was simulated in the NVT ensemble for 600 ns. Sampling was considered adequate after 600 ns of simulation time since the RMSD of the Cα atoms of the backbone converged to a certain value (see SI). The simulated peptoids and ion concentrations are summarized in Table 1. For analysis the radius of gyration was calculated using the Cα atoms of the backbone. To identify the most stable conformations, the structures in the simulation trajectory were clustered using the gromos[199] clustering algorithm (0.30 nm cut-off).

Table 9-1 Simulation of peptoid monomers in electrolyte solutions

Peptoid	**Electrolyte**
Sarcosine 12-mer OR Sar12	No ions 0.2 M KCl, 1 M KCl 0.1 M $CaCl_2$, 0.2 M $CaCl_2$
(N-ethylcarboxyl-N-ethyl)$_6$ OR (Nce-Net)$_6$	0.1 M $CaCl_2$, 0.2 M $CaCl_2$
(N-ethylcarboxyl-N-ethylisobutyl)$_6$ OR (Nce-Nib)$_6$	0.1 M $CaCl_2$, 0.2 M $CaCl_2$
(N-ethylcarboxyl-N-ethylphenylchloride)$_6$ OR (Nce-Ncp)$_6$	0.1 M $CaCl_2$, 0.2 M $CaCl_2$

*For Ca, the charges are Ca^{+2} (Cl^{-1}) or $Ca^{+1.5}$ ($Cl^{-0.75}$)

OLIGOMER SIMULATIONS. The systems were setup by constructing a box of size 10 nm x 10 nm x 10 nm with 4 peptoids of type NH_2-(Nce-Ncp)$_6$-H or NH_2-(Nce-Ncp)$_6$-H, solvent molecules, and ions. There were three types of electrolyte solutions – only K+ ions for charge neutralization, 89 Ca ions, and 164 Ca ions. After energy minimization and NPT equilibration, each system was simulation in the NVT ensemble (150 ns).

By 50 ns, the peptoids had assembled into aggregates and their sizes were classified for this early time period using an in-house Python script. The distance between alpha-carbon atoms of peptoids were calculated. Two peptoids were considered in the same aggregate if their alpha-carbon atoms (at least 6) were within 1.10 nm of each other. The number of dimers, trimers, and higher order oligomers were thus calculated for each frame. These labels for structures were verified by visualizing it on VMD, which in-turn verified the cut-off choices for choosing two peptoids in one aggregate. Monomers were calculated by subtracting the number of peptoids involved in oligomers from the total number of peptoids in the box.

Several descriptors for peptoid aggregation were calculated from these simulations (75001 values for each descriptor were obtained from every simulation). These included, the sum of radii

of gyration, and the sum of distances between the COMs of the peptoids (calculated using Cα atoms), sum of all phi dihedral angles, sum of all psi dihedral angles, sum of all omega dihedral angles, which were calculated using PLUMED. Number of contacts between atoms using CONTACTMAP (using SUM to add all contacts) function in PLUMED. We found contacts between chlorine atoms of the peptoid side-chain (SWITCH={RATIONAL R_0=0.50}), carboxyl atom from carboxylate side-chain of the peptoid side-chain (SWITCH={RATIONAL R_0=0.65}), and Cα atoms (SWITCH={RATIONAL R_0=0.65}). The number of contacts between calcium ions and carboxylate group oxygens were calculated by using COORDINATION (R_0=0.3 NLIST NL_CUTOFF=0.5).

Further, short-range and long-range coulomb and Lennard-Jones energies between peptoids, ions, and peptoids and ions were obtained from the GROMACS *energy* function. Nematic order parameters $\overline{P_1}$ and $\overline{P_2}$ were calculated using[320]

$$\overline{P_2} = \frac{1}{N}\sum_{i=1}^{N}\frac{3}{2}\left(\hat{z}_i.\hat{d}\right)^2 - \frac{1}{2}$$

$$\overline{P_1} = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{z}_i.\hat{d}\right)$$

Where $\hat{d}$ is taken as z axis, and $\hat{z}_i$ is the vector from the first nitrogen atom of the peptoid to the end nitrogen, and N is the number of peptoid molecules in the box.

Since the peptoids and their clusters are floppy and unstructured, it is difficult to determine their behavior visually and identify reaction coordinates to describe their aggregation. Thus, to identify potential reaction coordinates that captured the structures of aggregates independent of the ionic environment, we conducted a principal component analysis (PCA) on the above-

mentioned descriptors that depend on the peptoid only. Descriptors, as mentioned above, were calculated for all concentrations, and the data was collected in the same file to conduct PCA. More information about the PCA analysis are provided in the SI.

DIMERIZATION SIMULATIONS. Peptoids of type NH_2-(Nce-Ncp)$_6$-H or NH_2-(Nce-Ncp)$_6$-H (those in oligomer simulations) were used in these simulations. The PLUMED 2 plugin[119] and the parallel-bias metadynamics method were used for biasing variables. The radius of gyration (defined by alpha-carbon atoms) of both peptoids and the distance between the centers-of-mass (COM; also defined by alpha-carbon atoms) were biased using the parallel-bias metadynamics framework.[24] The initial Gaussian height, Gaussian width and bias factor were 2.0 kJ/mol, 0.01 nm for radius of gyration and 0.05 nm for COM distance, 10, respectively. Hills were deposited every 1 ps or 500 simulation steps. Simulations were considered converged after the 1-d free energy the COM distances did not change significantly over the last 20% of simulation time (see SI).

In order to estimate the error at each point in the free energy profile, 5 free energy profiles were generated between 60-100 % of the simulation time. To align these free energy profiles, the average of all points within 25 kJ/mol was calculated for and subtracted from each profile. Thereafter, an average value of the free energy was calculated. The error was calculated using $\frac{\sigma}{\sqrt{5}}$.

9.4 RESULTS AND DISCUSSION

MONOMER ION INTERACTIONS

To understand how ions may affect the folding of a single peptoid, we simulated sarcosine, (Nce-Net)$_6$, (Nce-Nib)$_6$, and (Nce-Ncp)$_6$ in different electrolytes and scaling conditions (See Methods). First, we look at sarcosine, which is a model peptoid and is characterized by open,

unstructured conformations in water.[318] In Figure 1, it is shown that Ca^{+2} induces folded structures but scaling the charge on calcium allows sarcosine to adopt open conformations. Notably, sarcosine is floppier in $Ca^{+1.5}$ than in Ca^{+2} as evidenced noisier RMSD curves (see Figure SI X). Consequently, the top-weighted structures have lesser weights than structures in Ca^{+2}.

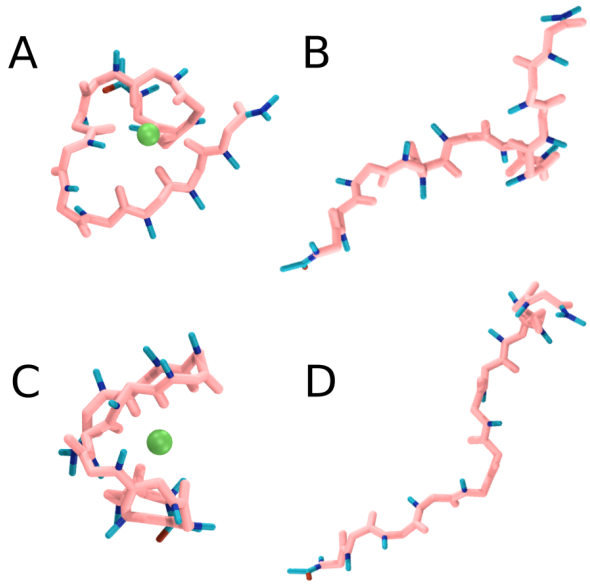


Figure 9-1 Top weighted structure of SAR12 in the presence of (A) 6 Ca^{+2} (6.13 % weight amongst structures identified by the clustering algorithm), (B) 6 $Ca^{+1.5}$ (2.93 %), (C) 32 Ca^{+2} (11.6 %), and (D) 32 $Ca^{+1.5}$ (3.19 %) ions. The backbone is colored in pink, while other atoms are represented in cyan (carbon), red (oxygen), blue (nitrogen), and green (calcium).

Finally, we explore the behavior of amphiphilic peptoids, which are most commonly found in experiments.[90,321] The $(Nce\text{-}Net)_6$ contains carboxylic side-chains (Figure 2), which are expected to preferentially interact with positive ions in the solution. In the case of Ca^{+2}, the carboxylic side-chains interact with the ions, but this interaction is reduced when the charge of calcium is scaled, evidenced by the absence of calcium ions within 0.5 nm of carboxylic side chain. Similarly, for $(Nce\text{-}Nib)_6$ and $(Nce\text{-}Ncp)_6$ peptoids (Figure 3), Ca^{+2} interacts with the carboxylic side chains and the interaction is reduced on scaling charges.

Notably, the latter peptoids have bulkier side chains, such as isobutyl and chlorophenyl side-chains than ethyl or methyl. These bulky side chains force the peptoids to adopt a collapsed structure. However, when the ions and side-chain charges are not scaled, the repulsion by like charges forces the peptoid to form more open structures (Figure 2 A and C, Figure 3 A and C). When the charges are scaled, the repulsion decreases, causing the structure-directing effect of bulky-hydrophobic groups to dominate (Figure 2 B and D, Figure 3 B and D).

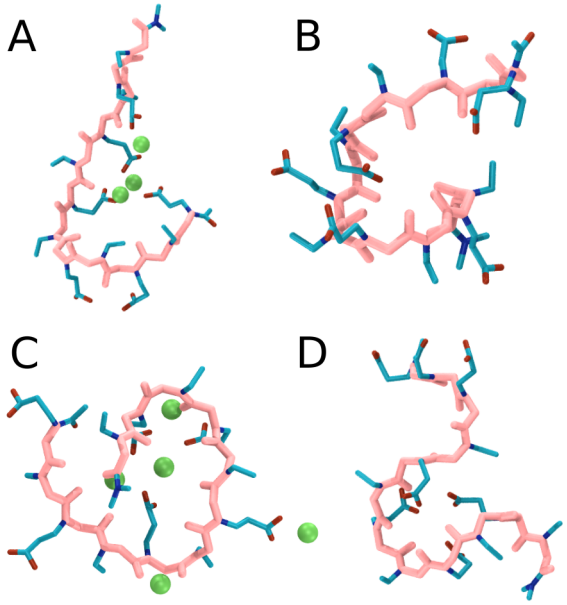


Figure 9-2 Top-weighted structure of $(Nce\text{-}Net)_6$ in the presence of (A) 14 Ca^{+2} (48.2 % weight amongst structures identified by the clustering algorithm), (B) 14 $Ca^{+1.5}$ (17.4 %), (C) 57 Ca^{+2} (43.6 %), and (D) 57 $Ca^{+1.5}$ (14.5 %) ions. The backbone is colored in pink, while other atoms are represented in cyan (carbon), red (oxygen), blue (nitrogen), and green (calcium).

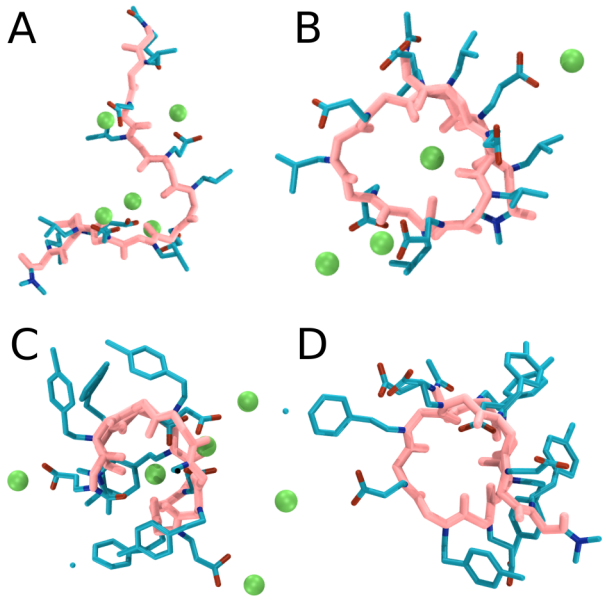


Figure 9-3 Top weighted structure of $(Nce\text{-}Nib)_6$ in the presence of (A) 60 Ca^{+2} (80.1 % weight amongst structures identified by the clustering algorithm), (B) 60 $Ca^{+1.5}$ (44.6 %) ions. Top weighted structure of $(Nce\text{-}Ncp)_6$ in the presence of (C) 52 Ca^{+2} (58.2 %), and (D) 52 $Ca^{+1.5}$ (53.3 %) ions. The backbone is colored in pink, while other atoms are represented in cyan (carbon), red (oxygen), blue (nitrogen), and green (calcium).

Similar structures, and differences in structure in changing ionic environments were also observed by Daily et al.[319] in their simulations of these peptoids using a different forcefield.[319] In fact, we also simulated these peptoids using the force field used by Daily et al.[319] and obtained comparable results (Results are described in the Supplementary Material; Figure SI2, SI3, SI4).

Thus, at a monomer level, the structure of the peptoid is dictated by the interplay of bulky hydrophobic side-chains that promoted collapsed structures and charge repulsion between like-charged groups that promote extended structures. When groups have full charge, the extended structures dominate. In contrast, when the charges are scaled, the structure-directing property of hydrophobic side-chains start to dominate. Finally, sarcosine, which does not have either charged

or bulky side-chains forms the same structures independent of the ionic environment that it is present in.

DIMER ION INTERACTIONS

Next, we investigated how the peptoids of type NH_2-(Nce-Ncp)$_6$-H and NH_2-$(CH_2)_6$-(Nce-Ncp)$_6$-H associated into dimers in different electrolyte solutions. Previously, Ma et al.[90] studied the dimerization of these peptoids in potassium and calcium electrolytes and concluded that dimerization studies can provide clues to larger-scale processes like crystallization and nucleation.[90] This study extends those dimerization calculations to explore the effect of ion concentrations and further highlight the effect of the addition of a hydrophobic tail.

From Figure 4A, we see that in the case of NH_2-(Nce-Ncp)$_6$-H (peptoid without a hydrophobic tail), the peptoids do not associate in the presence of K^{+1}, matching the prediction of Ma et al.[90] We hypothesize that K^{+1} ions are unable to form stable salt-bridges between peptoids, thereby the peptoids repel each other and stay in monomeric state in the solution. On the other hand, the presence of calcium ions allows them to associate via salt bridges, again matching predictions by Ma et al.[90] In fact, the presence of more calcium ions forces them to associate more strongly, as evidenced by a steep minimum in Figure A(green). In contrast, the association of NH_2-$(CH_2)_6$-(Nce-Ncp)$_6$-H (peptoid with a hydrophobic tail) is not sensitive to the concentration of calcium ions, as shown by similar binding free energies at different calcium concentrations in Figure 4B. Notably, the presence of a hydrophobic tail (-$(CH_2)_6$) seems to change this association characteristics of the peptoid.

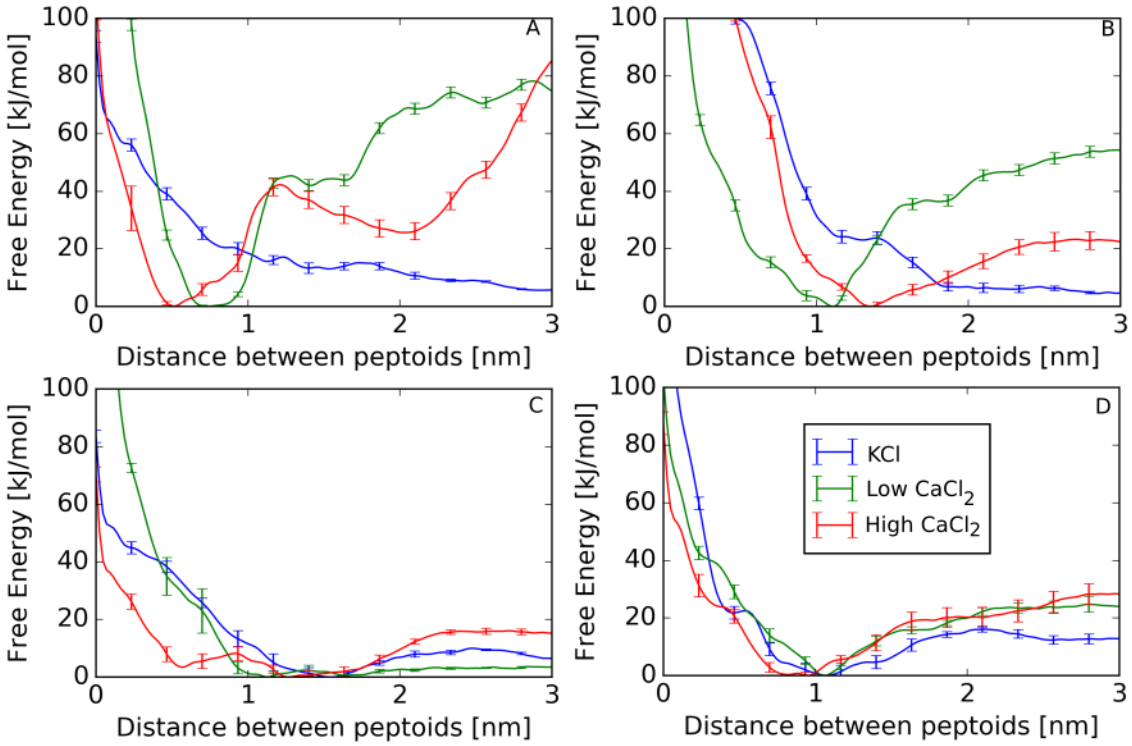


Figure 9-4 Free energy (kJ/mol) of association as a function of distance between peptoids (nm). Peptoid NH_2-(Nce-Ncp)$_6$-H (A and C) and peptoid NH_2-$(CH_2)_6$-(Nce-Ncp)$_6$-H (B and D) with (blue) potassium ions, (green) low calcium ions, and (red) high calcium ions. Calcium charges are 2+ (A and B) and 1.5+ (C and D)

Further, we calculate the free energy of binding of the peptoids when the charges of calcium ions are scaled to 1.5 from 2 (Figure 4 C and D) (the charge of the carboxylate group is also scaled from -1 to -0.75). In the case of the peptoid without a tail (Figure 4C) in calcium electrolytes, the binding energy between peptoids decreases drastically (a 55 kJ/mol decrease in the case of low calcium, and a 25 kJ/mol decrease in the case of high calcium). In contrast, the binding energy for the peptoids with a hydrophobic tail does not decrease as much (a 30 kJ/mol decrease in the case of low calcium, and a ~ 0 kJ/mol decrease in the case of high calcium). This alludes to differences in the association mechanisms of peptoid changes on adding a hydrophobic tail. We hypothesize

that the association of the latter is likely through hydrophobic association and not primarily through salt-bridges. Consequently, binding energy does not change much on changing the ionic environment.

Finally, the peptoids with scaled charges on the carboxylate group now associate in the K^{+1} solution. This is due to the decreased electrostatic repulsion between the peptoid groups, and the ability of K^{+1} to make salt bridges with these groups of lower charge.

Therefore, the dimerization study was able to show the differences in the clustering mechanisms between NH_2-(Nce-Ncp)$_6$-H and NH_2-$(CH_2)_6$-(Nce-Ncp)$_6$-H by characterizing their response to changing electrolyte types and concentrations. These results verify the differences in mechanism proposed in their paper.[90] However, the propensity for dimerization gives us limited insight into the formation and structure of higher order structures, like those observed in experiments by Ma et al.[90]

OLIGOMER ION INTERACTIONS

Four peptoids of type NH_2-(Nce-Ncp)$_6$-H or NH_2-$(CH_2)_6$-(Nce-Ncp)$_6$-H were simulated at different ion concentrations to understand how they equilibrated into clusters in diverse environments and study the effect of adding a hydrophobic tail. In Figure 5, we plot the probability of observing a structure at different time points in the simulation. The probabilities are computed over the first 50 ns to capture early aggregation behavior, the calculations are done using 10 ns windows. Overall, we see that the peptoids only form monomers or dimers in a potassium solution, whereas they form higher order oligomers in calcium solutions.

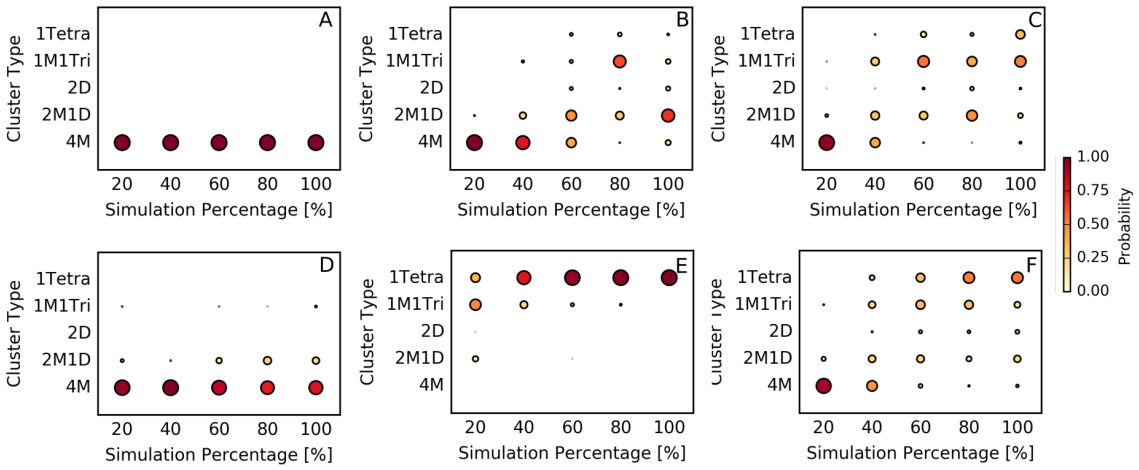


Figure 9-5 Types of aggregates formed during 50 ns of MD simulations of 4 peptoids of type (top panel) NH_2-(Nce-Ncp)$_6$-H and (bottom panel) NH_2-$(CH_2)_6$-(Nce-Ncp)$_6$-H in electrolytes with potassium ions (A, D), low concentrations of calcium ions (B, E), and high concentration of calcium ions (C, F). The color of the circle and size demonstrates the probability of being in an aggregated state.

Looking closely at the behavior of the peptoids in potassium electrolyte (Figure 5 A, D), we see that the peptoid with a hydrophobic tail begins to associate into oligomers in the first 10 ns of the simulation. However, the peptoid without a tail remains in the dissociated state. When the potassium ions are replaced with calcium (Figure 5 B, E), the peptoids associate in oligomer states early in the simulation. In fact, the peptoid with a tail associates more readily. The same trend is observed when the concentration of calcium is increased. At higher calcium concentrations, the peptoids without a tail associate into larger oligomers. Similar to this, Batoulis et al.[317] also saw an increase in oligomer formation in solution on increasing Ca^{2+} concentration.

Notably, the dimerization free energy profile also predicted that the peptoids were unlikely to associate in potassium electrolyte, since they did not display an energy minimum for association. Moreover, the dimerization predictions show that peptoids associate in calcium solution which is also observed in these oligomer simulations. However, by simulating 4 peptoids together, we are

able to observe the formation of double dimers, trimers, and tetramers which was not possible in dimerization simulations.

9.5 CONCLUSION

We studied the effect of both ion type and concentration on (1) the structure of peptoid monomers, (2) the dimerization free energy of NH_2-(Nce-Ncp)$_6$-H and NH_2-$(CH_2)_6$-(Nce-Ncp)$_6$-H peptoids, and (3) the oligomerization of peptoids in (2) using classical molecular dynamics, employing the forcefield suggested by Prakash et al.,[318] and metadynamics for enhanced sampling. Through this study we were able to systematically study how ions affect each process and can hypothesize which features are most important in this process.

Results show that the structure of a peptoid monomer is affected by side chains. Large, hydrophobic side-chains can make the peptoid collapse in itself in aqueous electrolyte. In contrast, charge side-chains repel each other make extended strands of peptoids. Ions mediate this structure minimally but associate with oppositely charged side-chains during the process.

In contrast, ion type and concentration can affect the dimerization free energy of peptoids. We observe that the NH_2-(Nce-Ncp)$_6$-H peptoid associates through the formation of salt-bridges. Consequently, scaling the charges of ions and carboxylate groups decreases their attraction, resulting in a smaller binding free energy. On the other hand, NH_2-$(CH_2)_6$-(Nce-Ncp)$_6$-H associates through its hydrophobic tail and the dimerization free energy profile is unaffected by the change in calcium concentration or the charge of calcium ions.

Further, we studied the formation of peptoid oligomers for the above-mentioned peptoids. Similar to dimerization, the oligomer formation capacity of NH_2-$(CH_2)_6$-(Nce-Ncp)$_6$-H is unaffected by ion type and concentration. Both peptoids form trimer and tetramers in calcium solution, however, the peptoid without a hydrophobic tail is in the tetrameric state at higher calcium concentration.

Thus, ions do not affect monomeric states but affect association of peptoids. However, this affect is present in peptoids that associate predominantly through electrostatic interactions and the effect is diminished when the main form of association is through hydrophobic tails. Therefore, simulations of cluster formation should take ions into account explicitly when dealing with systems that associate via electrostatic interactions, like colloidal particle and charged proteins.

9.6 ACKNOWLEDGEMENTS

This work was supported by the MS3 (Materials Synthesis and Simulation Across Scales) Initiative at Pacific Northwest National Laboratory (PNNL), a multi-program national laboratory operated by Battelle for the U.S. Department of Energy. This work was facilitated through the use of computational, storage, and networking infrastructure provided by the Hyak super-computer system, supported in part by the University of Washington.

Chapter 10. CONCLUSION

In this thesis, we aimed to understand how proteins and other particles assemble at interfaces owing to the importance of these assembly processes to many areas like disease etiology, industrial food and enzyme production, and biomineralization. All-atom MD simulations with various enhanced sampling methods were applied to obtain molecular-level insights into the early-stages of the assembly process like adsorption of species to an interface, and dimerization.

In Chapter 2, a model mica system is simulated using classical MD simulations to probe the structure of water layers adjacent to the basal surface of mica. This water structure is then used in a modified Poisson-Boltzmann equation to calculate forces between two mica plates. These force predictions are superior to the predictions made using standard colloidal theories, like

DLVO. This work is extended in Chapter 8, where the structure of ion layers adsorbed to the basal surface of mica is probed at different concentration using MD simulations. Here, the enhanced sampling method parallel-bias metadynamics with partitioned families is used to bias all ions in the system. This study shows how the structure of the adsorbed layer, and the configuration of ions in the first layer evolves with increasing electrolyte concentrations. Similar to chapter 2, these estimates of the structure of the ionic layer can be used to predict forces between mica plates. These studies put forth a protocol for the prediction of forces between nanoparticles through high-resolution, low-cost MD simulations of a model surface.

In Chapter 3, peptoid biopolymers are simulated in model environments (in pure water and near model surfaces) to detect stable conformations of the molecule using a forcefield validated in this study. The comparison of a model peptoid (sarcosine) with a model peptide (alanine) reveals differences in preferred structure in water and when adsorbed to model hydrophilic and hydrophobic surfaces. Sarcosine is demonstrated to be more flexible than its peptide counterpart, alanine, as also endorsed by simulations by Settani et al.[322] This study is extended in Chapter 9, where the structure of peptoids is explored in more realistic environments – in electrolyte solutions and in the presence of other peptoids. The structure of peptoids with charged side-chains are affected by ions, but large hydrophobic groups can counter this effect. Further, the interaction between two peptoids was shown to be repulsive in monovalent electrolytes, but attractive in divalent electrolytes due to the formation of stable salt bridges between peptoids. Finally, the study explores oligomer formation when four peptoids are put in a box, and their behavior is similar to the predictions by dimerization free energy calculations.

In Chapter 3, the adsorption of a model peptide (GGKGG) to a silica surface is investigated using classical MD simulations and metadynamics enhanced sampling method. The study shows that ions that are adsorbed strongly to the surface can frustrate sampling of peptide adsorption.

Therefore, ions should also be biased during the simulation to obtain correct free energy estimates of peptide adsorption. In Chapter 6, a method, namely parallel bias metadynamics with partitioned families, is introduced to efficiently bias multiple, degenerate variables, like all ions in Chapter 3, in one simulation. The method demonstrates linear increase in the speed of convergence of free energy profiles for systems of 3, 21, and 78 collective variables. The development of this method is crucial to sampling high density systems which often have identical copies of the same particles leading to several, degenerate collective variables. This method provides a smart way to address this sampling issue and has been adopted in the Pfaendtner Research Group, and other research groups.

In Chapter 5, the structure of R5 segment of the silaffin peptide is probed on the surface of silica using MD simulations enhanced by parallel tempering metadynamics in the well-tempered ensemble. The study illustrates how the strength of binding of the peptide changes as a function of the number of phosphorylated residues in the peptide, and the pH of the environment. R5 binds lesser with decreasing pH and increasing phosphorylation. This study also shows that as the degree of phosphorylation increases, the peptide prefers to be in more compact structures. In Chapter 7, the study probes the structure of another peptide, the SNA15 segment from statherin peptide, when adsorbed to the surface of minerals like hydroxyapatite, silica, and titania. In contrast to Chapter 5, where only MD simulations were used for predictions of peptide structure, this study incorporates chemical shifts from NMR experiments using a Bayesian inference framework to predict the most stable structures of the peptide. The simulation-predicted structures of the peptide are close to the predictions by standard tools, namely TALOS-N, used by NMR scientists. With the success of these predictions, we propose this integrated simulation and experimental protocol as a standard for high fidelity predictions of protein structure.

In summary, studies in this thesis have been instrumental in predicting how particles (water, ions, peptoid biopolymers, model peptides, and protein fragments) adsorb to interfaces and how their interaction with the surface changes as a function of variables in the environment. These studies demonstrate the power of MD to provide molecular-level insights, like free energy estimates for adsorption, ensemble of stable structures, and the structure of adsorbed monolayer, that are often hard to probe experimentally. These studies are a crucial baseline for simulations of more complex experimental systems, like peptide monolayer formation on mineral interfaces.

BIBLIOGRAPHY

(1) Hamon, L.; Panda, D.; Savarin, P.; Joshi, V.; Bernhard, J.; Mucher, E.; Mechulam, A.; Curmi, P. A.; Pastré, D. Mica Surface Promotes the Assembly of Cytoskeletal Proteins. *Langmuir* **2009**, *25* (6), 3331–3335.

(2) Kowalewski, T.; Holtzman, D. M.; Huynh, T.; Lei, H.; Castelli, M.; Hu, J.; Zhang, Y.; Zhou, R. In Situ Atomic Force Microscopy Study of Alzheimer's Beta-Amyloid Peptide on Different Substrates: New Insights into Mechanism of Beta-Sheet Formation. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96* (7), 3688–3693.

(3) Dai, B.; Kang, S. -g.; Huynh, T.; Lei, H.; Castelli, M.; Hu, J.; Zhang, Y.; Zhou, R. Salts Drive Controllable Multilayered Upright Assembly of Amyloid-like Peptides at Mica/Water Interface. *Proc. Natl. Acad. Sci.* **2013**, *110* (21), 8543–8548.

(4) Zhang, F.; Du, H.-N.; Zhang, Z.-X.; Ji, L.-N.; Li, H.-T.; Tang, L.; Wang, H.-B.; Fan, C.-H.; Xu, H.-J.; Zhang, Y.; et al. Epitaxial Growth of Peptide Nanofilaments on Inorganic Surfaces: Effects of Interfacial Hydrophobicity/Hydrophilicity. *Angew. Chemie Int. Ed.* **2006**, *45* (22), 3611–3613.

(5) Whitehouse, C.; Fang, J.; Aggeli, A.; Bell, M.; Brydson, R.; Fishwick, C. W. G.; Henderson, J. R.; Knobler, C. M.; Owens, R. W.; Thomson, N. H.; et al. Adsorption and Self-Assembly of Peptides on Mica Substrates. *Angew. Chemie Int. Ed.* **2005**, *44* (13), 1965–1968.

(6) Hoyer, W.; Cherny, D.; Subramaniam, V.; Jovin, T. M. Rapid Self-Assembly of α-Synuclein Observed by In Situ Atomic Force Microscopy. *J. Mol. Biol.* **2004**, *340* (1), 127–139.

(7) Giacomelli, C. E.; Norde, W. Conformational Changes of the Amyloidβ-Peptide (1-40) Adsorbed on Solid Surfaces. *Macromol. Biosci.* **2005**, *5* (5), 401–407.

(8) Karplus, M.; McCammon, J. A. Molecular Dynamics Simulations of Biomolecules. *Nat. Struct. Biol.* **2002**, *9* (9), 646–652.

(9) Klepeis, J. L.; Lindorff-Larsen, K.; Dror, R. O.; Shaw, D. E. Long-Timescale Molecular Dynamics Simulations of Protein Structure and Function. *Curr. Opin. Struct. Biol.* **2009**, *19* (2), 120–127.

(10) Steinbrecher, T.; Latzer, J.; Case, D. A. Revised AMBER Parameters for Bioorganic Phosphates. *J. Chem. Theory Comput.* **2012**, *8* (11), 4405–4412.

(11) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone

Parameters. *Proteins Struct. Funct. Bioinforma.* **2006**, *65* (3), 712–725.
(12) Huang, J.; MacKerell, A. D. CHARMM36 All-Atom Additive Protein Force Field: Validation Based on Comparison to NMR Data. *J. Comput. Chem.* **2013**, *34* (25), 2135–2145.
(13) Lee, O.-S.; Stupp, S. I.; Schatz, G. C. Atomistic Molecular Dynamics Simulations of Peptide Amphiphile Self-Assembly into Cylindrical Nanofibers. *J. Am. Chem. Soc.* **2011**, *133* (10), 3677–3683.
(14) Cheon, M.; Chang, I.; Mohanty, S.; Luheshi, L. M.; Dobson, C. M.; Vendruscolo, M.; Favrin, G. Structural Reorganisation and Potential Toxicity of Oligomeric Species Formed during the Assembly of Amyloid Fibrils. *PLoS Comput. Biol.* **2007**, *3* (9), e173.
(15) Broglia, R. A.; Tiana, G.; Pasquali, S.; Roman, H. E.; Vigezzi, E. Folding and Aggregation of Designed Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95* (22), 12930–12933.
(16) Matthes, D.; Gapsys, V.; De Groot, B. L. Driving Forces and Structural Determinants of Steric Zipper Peptide Oligomer Formation Elucidated by Atomistic Simulations. *J. Mol. Biol.* **2012**, *421* (2–3), 390–416.
(17) Benzinger, T. L. S.; Gregory, D. M.; Burkoth, T. S.; Miller-Auer, H.; Lynn, D. G.; Botto, R. E.; Meredith, S. C. Propagating Structure of Alzheimer's -Amyloid(10-35) Is Parallel -Sheet with Residues in Exact Register. *Proc. Natl. Acad. Sci.* **1998**, *95* (23), 13407–13412.
(18) Abrams, C.; Bussi, G. Enhanced Sampling in Molecular Dynamics Using Metadynamics, Replica-Exchange, and Temperature-Acceleration. *Entropy* **2014**, *16* (1), 163–199.
(19) Kelley, N. W.; Vishal, V.; Krafft, G. A.; Pande, V. S. Simulating Oligomerization at Experimental Concentrations and Long Timescales: A Markov State Model Approach. *J. Chem. Phys.* **2008**, *129* (21), 214707.
(20) Bernardi, R. C.; Melo, M. C. R.; Schulten, K. Enhanced Sampling Techniques in Molecular Dynamics Simulations of Biological Systems. *Biochim. Biophys. Acta - Gen. Subj.* **2015**, *1850* (5), 872–877.
(21) Barducci, A.; Bonomi, M.; Parrinello, M. Metadynamics. *Wiley Interdisciplinary Reviews: Computational Molecular Science*. Wiley/Blackwell (10.1111) September 1, 2011, pp 826–843.
(22) Kästner, J. Umbrella Sampling. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1* (6), 932–942.
(23) Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **2008**, *100* (2), 020603.
(24) Pfaendtner, J.; Bonomi, M. Efficient Sampling of High-Dimensional Free-Energy Landscapes with Parallel Bias Metadynamics. *J. Chem. Theory Comput.* **2015**, *11* (11), 5062–5067.
(25) Prakash, A.; Fu, C. D.; Bonomi, M.; Pfaendtner, J. Biasing Smarter, Not Harder, By Partitioning Collective Variables Into Families. *J. Chem. Theory Comput.* **2018**, acs.jctc.8b00448.
(26) Bonomi, M.; Camilloni, C.; Vendruscolo, M. Metadynamic Metainference: Enhanced Sampling of the Metainference Ensemble Using Metadynamics. *Sci. Rep.* **2016**, *6* (1), 31232.
(27) Earl, D. J.; Deem, M. W. Parallel Tempering: Theory, Applications, and New Perspectives. *Phys. Chem. Chem. Phys.* **2005**, *7* (23), 3910.
(28) Deighan, M.; Bonomi, M.; Pfaendtner, J. Efficient Simulation of Explicitly Solvated Proteins in the Well-Tempered Ensemble. *J. Chem. Theory Comput.* **2012**, *8* (7), 2189–

2192.
(29) Torrie, G. M.; Valleau, J. P. Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *J. Comput. Phys.* **1977**, *23* (2), 187–199.
(30) Barducci, A.; Bonomi, M.; Parrinello, M. Metadynamics. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1* (5), 826–843.
(31) Valsson, O.; Tiwary, P.; Parrinello, M. Enhancing Important Fluctuations: Rare Events and Metadynamics from a Conceptual Viewpoint. *Annu. Rev. Phys. Chem.* **2016**, *67* (1), 159–184.
(32) Dama, J. F.; Parrinello, M.; Voth, G. A. Well-Tempered Metadynamics Converges Asymptotically. *Phys. Rev. Lett.* **2014**, *112* (24), 240602.
(33) Pfaendtner, J.; Bonomi, M. Efficient Sampling of High-Dimensional Free-Energy Landscapes with Parallel Bias Metadynamics. *J. Chem. Theory Comput.* **2015**, *11* (11), 5062–5067.
(34) Prakash, A.; Sprenger, K. G.; Pfaendtner, J. Essential Slow Degrees of Freedom in Protein-Surface Simulations: A Metadynamics Investigation. *Biochem. Biophys. Res. Commun.* **2017**, *498* (2), 274–281.
(35) Raiteri, P.; Laio, A.; Gervasio, F. L.; Micheletti, C.; Parrinello, M. Efficient Reconstruction of Complex Free Energy Landscapes by Multiple Walkers Metadynamics. *J. Phys. Chem. B* **2006**, *110* (8), 3533–3539.
(36) Brown, M. A.; Bossa, G. V.; May, S. Emergence of a Stern Layer from the Incorporation of Hydration Interactions into the Gouy-Chapman Model of the Electrical Double Layer. *Langmuir* **2015**, *31* (42), 11477–11483.
(37) Lyons, J. S.; Furlong, D. N.; Healy, T. W. The Electrical Double-Layer Properties of the Mica (Muscovite)-Aqueous Electrolyte Interface. *Aust. J. Chem.* **1981**, *34* (6), 1177–1187.
(38) Pashley, R. M. DLVO and Hydration Forces between Mica Surfaces in Li+, Na+, K+, and Cs+electrolyte Solutions: A Correlation of Double-Layer and Hydration Forces with Surface Cation Exchange Properties. *J. Colloid Interface Sci.* **1981**, *83* (2), 531–546.
(39) Miklavic, S. J.; Ninham, B. W. Competitiion for Adsorption Sites by Hydrated Ions. *J. Colloid Interface Sci.* **1990**, *134* (2), 305–311.
(40) Raviv, U.; Laurat, P.; Klein, J. Time Dependence of Forces between Mica Surfaces in Water and Its Relation to the Release of Surface Ions. *J. Chem. Phys.* **2002**, *116* (12), 5167.
(41) Kobayashi, K.; Oyabu, N.; Kimura, K.; Ido, S.; Suzuki, K.; Imai, T.; Tagami, K.; Tsukada, M.; Yamada, H. Visualization of Hydration Layers on Muscovite Mica in Aqueous Solution by Frequency-Modulation Atomic Force Microscopy. *J. Chem. Phys.* **2013**, *138* (18), 184704.
(42) Cheng, L.; Fenter, P.; Nagy, K. L.; Schlegel, M. L.; Sturchio, N. C. Molecular-Scale Density Oscillations in Water Adjacent to a Mica Surface. *Phys. Rev. Lett.* **2001**, *87* (15), 156103.
(43) Odelius, M.; Bernasconi, M.; Parrinello, M. Two Dimensional Ice Adsorbed on Mica Surface. *Phys. Rev. Lett.* **1997**, *78* (14), 2855–2858.
(44) Swenson, J.; Bergman, R.; Howells, W. S. Quasielastic Neutron Scattering of Two-Dimensional Water in a Vermiculite Clay. *J. Chem. Phys.* **2000**, *113* (7), 2873–2879.
(45) Remsing, R. C.; Baer, M. D.; Schenter, G. K.; Mundy, C. J.; Weeks, J. D. The Role of Broken Symmetry in Solvation of a Spherical Cavity in Classical and Quantum Water Models. *J. Phys. Chem. Lett.* **2014**, *5* (16), 2767–2774.
(46) Eggimann, B. L.; Siepmann, J. I. Size Effects on the Solvation of Anions at the Aqueous Liquid-Vapor Interface. *J. Phys. Chem. C* **2008**, *112* (1), 210–218.

(47) Chun, J.; Mundy, C. J.; Schenter, G. K. The Role of Solvent Heterogeneity in Determining the Dispersion Interaction between Nanoassemblies. *J. Phys. Chem. B* **2015**, *119* (18), 5873–5881.
(48) Cygan, R. T.; Liang, J.-J.; Kalinichev, A. G. Molecular Models of Hydroxide, Oxyhydroxide, and Clay Phases and the Development of a General Force Field. *J. Phys. Chem. B* **2004**, *108* (4), 1255–1266.
(49) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, Flexible, and Free. *Journal of Computational Chemistry*. 2005, pp 1701–1718.
(50) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81* (8), 3684–3690.
(51) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126* (1), 014101.
(52) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18* (12), 1463–1472.
(53) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An $N \cdot \log(N)$ Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98* (12), 10089–10092.
(54) VandeVondele, J.; Hutter, J. Gaussian Basis Sets for Accurate Calculations on Molecular Systems in Gas and Condensed Phases. *J. Chem. Phys.* **2007**, *127* (11), 114105.
(55) Becke, A. D. Density-Functional Exchange-Energy Approximation with Correct Asymptotic Behavior. *Phys. Rev. A* **1988**, *38* (6), 3098–3100.
(56) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B* **1988**, *37* (2), 785–789.
(57) Goedecker, S.; Teter, M. Separable Dual-Space Gaussian Pseudopotentials. *Phys. Rev. B - Condens. Matter Mater. Phys.* **1996**, *54* (3), 1703–1710.
(58) Grimme, S. Semiempirical GGA-Type Density Functional Constructed with a Long-Range Dispersion Correction. *J. Comput. Chem.* **2006**, *27* (15), 1787–1799.
(59) Baer, M. D.; Mundy, C. J.; McGrath, M. J.; Kuo, I. F. W.; Siepmann, J. I.; Tobias, D. J. Re-Examining the Properties of the Aqueous Vapor-Liquid Interface Using Dispersion Corrected Density Functional Theory. *J. Chem. Phys.* **2011**, *135* (12).
(60) Feibelman, P. J. K+-Hydration in a Low-Energy Two-Dimensional Wetting Layer on the Basal Surface of Muscovite. *J. Chem. Phys.* **2013**, *139* (7), 074705.
(61) Fenter, P.; Kerisit, S.; Raiteri, P.; Gale, J. D. Is the Calcite–Water Interface Understood? Direct Comparisons of Molecular Dynamics Simulations with Specular X-Ray Reflectivity Data. *J. Phys. Chem. C* **2013**, *117* (10), 5028–5042.
(62) Fenter, P.; Lee, S. S.; Skelton, A. A.; Cummings, P. T. Direct and Quantitative Comparison of Pixelated Density Profiles with High-Resolution X-Ray Reflectivity Data. *J. Synchrotron Radiat.* **2011**, *18* (2), 257–265.
(63) Jedlovszky, P.; Předota, M.; Nezbeda, I. Hydration of Apolar Solutes of Varying Size: A Systematic Study. *Mol. Phys.* **2006**, *104* (15), 2465–2476.
(64) Lee, S. S.; Nagy, K. L.; Fenter, P. Distribution of Barium and Fulvic Acid at the Mica-Solution Interface Using in-Situ X-Ray Reflectivity. *Geochim. Cosmochim. Acta* **2007**, *71* (23), 5763–5781.
(65) Bowers, G. M.; Bish, D. L.; Kirkpatrick, R. J. Cation Exchange at the Mineral−Water Interface: H_3O^+/K^+ Competition at the Surface of Nano-Muscovite. *Langmuir* **2008**, *24* (18), 10240–10244.

(66) Wang, J.; Kalinichev, A. G.; Kirkpatrick, R. J.; Cygan, R. T. Structure, Energetics, and Dynamics of Water Adsorbed on the Muscovite (001) Surface: A Molecular Dynamics Simulation. *J. Phys. Chem. B* **2005**, *109* (33), 15893–15905.
(67) Leow, W. W.; Hwang, W. Epitaxially Guided Assembly of Collagen Layers on Mica Surfaces. *Langmuir* **2011**, *27* (17), 10907–10913.
(68) Lee, S. S.; Fenter, P.; Nagy, K. L.; Sturchio, N. C. Monovalent Ion Adsorption at the Muscovite (001)–Solution Interface: Relationships among Ion Coverage and Speciation, Interfacial Water Structure, and Substrate Relaxation. *Langmuir* **2012**, *28* (23), 8637–8650.
(69) Malani, A.; Ayappa, K. G. Relaxation and Jump Dynamics of Water at the Mica Interface. *J. Chem. Phys.* **2012**, *136* (19), 194701.
(70) Lee, S. H.; Rossky, P. J. A Comparison of the Structure and Dynamics of Liquid Water at Hydrophobic and Hydrophilic Surfaces—a Molecular Dynamics Simulation Study. *J. Chem. Phys.* **1994**, *100* (4), 3334–3345.
(71) Marry, V.; Rotenberg, B.; Turq, P. Structure and Dynamics of Water at a Clay Surface from Molecular Dynamics Simulation. *Phys. Chem. Chem. Phys.* **2008**, *10* (32), 4802–4813.
(72) Lee, C. Y.; McCammon, J. A.; Rossky, P. J. The Structure of Liquid Water at an Extended Hydrophobic Surface. *J. Chem. Phys.* **1984**, *80* (9), 4448–4455.
(73) Patel, A. J.; Varilly, P.; Chandler, D. Fluctuations of Water near Extended Hydrophobic and Hydrophilic Surfaces. *J. Phys. Chem. B* **2010**, *114* (4), 1632–1637.
(74) von Rudorff, G. F.; Jakobsen, R.; Rosso, K. M.; Blumberger, J. Hematite(001)-Liquid Water Interface from Hybrid Density Functional-Based Molecular Dynamics. *J. Phys. Condens. Matter* **2016**, *28* (39), 394001.
(75) Stillinger, F. H.; Rahman, A. Improved Simulation of Liquid Water by Molecular Dynamics. *J. Chem. Phys.* **1974**, *60* (4), 1545–1557.
(76) Wang, J.; Kalinichev, A. G.; Kirkpatrick, R. J. Asymmetrie Hydrogen Bonding and Orientational Ordering of Water at Hydrophobic and Hydrophilic Surfaces: A Comparison of Water/Vapor, Water/Talc, and Water/Mica Interfaces. *J. Phys. Chem. C* **2009**, *113* (25), 11077–11085.
(77) Kerisit, S.; Okumura, M.; Rosso, K. M.; Machida, M. Molecular Simulation of Cesium Adsorption at the Basal Surface of Phyllosilicate Minerals. *Clays Clay Miner.* **2016**.
(78) Marx, D.; Tuckerman, M. E.; Hutter, J.; Parrinello, M. The Nature of the Hydrated Excess Proton in Water. *Nature* **1999**, *397* (6720), 601–604.
(79) Berkelbach, T. C.; Lee, H. S.; Tuckerman, M. E. Concerted Hydrogen-Bond Dynamics in the Transport Mechanism of the Hydrated Proton: A First-Principles Molecular Dynamics Study. *Phys. Rev. Lett.* **2009**, *103* (23), 238302.
(80) Jagoda-Cwiklik, B.; Cwiklik, L.; Jungwirth, P. Behavior of the Eigen Form of Hydronium at the Air/Water Interface. *J. Phys. Chem. A* **2011**, *115* (23), 5881–5886.
(81) Brown, M. A.; Goel, A.; Abbas, Z. Effect of Electrolyte Concentration on the Stern Layer Thickness at a Charged Interface. *Angew. Chemie Int. Ed.* **2016**, *55* (11), 3790–3794.
(82) Robins, M.; Fillery-Travis, A. Colloidal Dispersions. *J. Chem. Technol. Biotechnol.* **2007**, *54* (2), 201–202.
(83) Pashley, R. . Hydration Forces between Mica Surfaces in Aqueous Electrolyte Solutions. *J. Colloid Interface Sci.* **1981**, *80* (1), 153–162.
(84) Boţan, A.; Marry, V.; Rotenberg, B.; Turq, P.; Noetinger, B. How Electrostatics Influences Hydrodynamic Boundary Conditions: Poiseuille and Electro-Osmostic Flows in Clay Nanopores. *J. Phys. Chem. C* **2013**, *117* (2), 978–985.

(85) Blees, M. H. Foundations of Colloid Science. *Colloids Surfaces A Physicochem. Eng. Asp.* **2002**, *210* (1), 125.
(86) Onsagbr, L. Electric Moments of Molecules in Liquids. *J. Am. Chem. Soc.* **1936**, *58* (8), 1486–1493.
(87) Hill, N. E. Interpretation of the Dielectric Properties of Water. *Trans. Faraday Soc.* **1963**, *59* (0), 344.
(88) Tan, Q.; Zhao, G.; Qiu, Y.; Kan, Y.; Ni, Z.; Chen, Y. Experimental Observation of the Ion-Ion Correlation Effects on Charge Inversion and Strong Adhesion between Mica Surfaces in Aqueous Electrolyte Solutions. *Langmuir* **2014**, *30* (36), 10845–10854.
(89) Pashley, R. . DLVO and Hydration Forces between Mica Surfaces in Li+, Na+, K+, and Cs+ Electrolyte Solutions: A Correlation of Double-Layer and Hydration Forces with Surface Cation Exchange Properties. *J. Colloid Interface Sci.* **1981**, *83* (2), 531–546.
(90) Ma, X.; Zhang, S.; Jiao, F.; Newcomb, C. J.; Zhang, Y.; Prakash, A.; Liao, Z.; Baer, M. D.; Mundy, C. J.; Pfaendtner, J.; et al. Tuning Crystallization Pathways through Sequence Engineering of Biomimetic Polymers. *Nat. Mater.* **2017**, *16* (7), 767–774.
(91) Du, Q.; Freysz, E.; Shen, Y. R. Vibrational Spectra of Water Molecules at Quartz/Water Interfaces. *Phys. Rev. Lett.* **1994**, *72* (2), 238–241.
(92) Bishop, K. J. M.; Wilmer, C. E.; Soh, S.; Grzybowski, B. A. Nanoscale Forces and Their Uses in Self-Assembly. *Small* **2009**, *5* (14), 1600–1630.
(93) Ganesh, S. D.; Saha, N.; Zandraa, O.; Zuckermann, R. N.; Sáha, P. Peptoids and Polypeptoids: Biomimetic and Bioinspired Materials for Biomedical Applications. *Polym. Bull.* **2017**, 1–12.
(94) Knight, A. S.; Zhou, E. Y.; Francis, M. B.; Zuckermann, R. N. Sequence Programmable Peptoid Polymers for Diverse Materials Applications. *Adv. Mater.* **2015**, *27* (38), 5665–5691.
(95) Zuckermann, R. N. Peptoid Origins. *Biopolymers* **2011**, *96* (5), 545–555.
(96) Sun, J.; Zuckermann, R. N. Peptoid Polymers: A Highly Designable Bioinspired Material. *ACS Nano* **2013**, *7* (6), 4715–4732.
(97) T. Dohm, M.; Kapoor, R.; E. Barron, A. Peptoids: Bio-Inspired Polymers as Potential Pharmaceuticals. *Curr. Pharm. Des.* **2011**, *17* (25), 2732–2747.
(98) Jang, H.; Fafarman, A.; Holub, J. M.; Kirshenbaum, K. Click to Fit: Versatile Polyvalent Display on a Peptidomimetic Scaffold. *Org. Lett.* **2005**, *7* (10), 1951–1954.
(99) Kudirka, R.; Tran, H.; Sanii, B.; Nam, K. T.; Choi, P. H.; Venkateswaran, N.; Chen, R.; Whitelam, S.; Zuckermann, R. N. Folding of a Single-Chain, Information-Rich Polypeptoid Sequence into a Highly Ordered Nanosheet. *Biopolymers* **2011**, *96* (5), 586–595.
(100) Chen, C.-L.; Zuckermann, R. N.; DeYoreo, J. J. Surface-Directed Assembly of Sequence-Defined Synthetic Polymers into Networks of Hexagonally Patterned Nanoribbons with Controlled Functionalities. *ACS Nano* **2016**, *10* (5), 5314–5320.
(101) Jin, H.; Jiao, F.; Daily, M. D.; Chen, Y.; Yan, F.; Ding, Y.-H.; Zhang, X.; Robertson, E. J.; Baer, M. D.; Chen, C.-L. Highly Stable and Self-Repairing Membrane-Mimetic 2D Nanomaterials Assembled from Lipid-like Peptoids. *Nat. Commun.* **2016**, *7*, 12252.
(102) Wu, C. W.; Sanborn, T. J.; Zuckermann, R. N.; Barron, A. E. Peptoid Oligomers with α-Chiral, Aromatic Side Chains: Effects of Chain Length on Secondary Structure. *J. Am. Chem. Soc.* **2001**, *123* (13), 2958–2963.
(103) Wu, C. W.; Kirshenbaum, K.; Sanborn, T. J.; Patch, J. A.; Huang, K.; Dill, K. A.; Zuckermann, R. N.; Barron, A. E. Structural and Spectroscopic Studies of Peptoid Oligomers with ??-Chiral Aliphatic Side Chains. *J. Am. Chem. Soc.* **2003**, *125* (44),

13525–13530.
(104) Seo, J.; Barron, A. E.; Zuckermann, R. N. Novel Peptoid Building Blocks: Synthesis of Functionalized Aromatic Helix-Inducing Submonomers. *Org. Lett.* **2010**, *12* (3), 492–495.
(105) Armand, P.; Kirshenbaum, K.; Goldsmith, R. A.; Farr-Jones, S.; Barron, A. E.; Truong, K. T.; Dill, K. A.; Mierke, D. F.; Cohen, F. E.; Zuckermann, R. N.; et al. NMR Determination of the Major Solution Conformation of a Peptoid Pentamer with Chiral Side Chains. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95* (8), 4309–4314.
(106) Rosales, A. M.; Murnen, H. K.; Kline, S. R.; Zuckermann, R. N.; Segalman, R. A. Determination of the Persistence Length of Helical and Non-Helical Polypeptoids in Solution. *Soft Matter* **2012**, *8* (13), 3673–3680.
(107) Sanii, B.; Kudirka, R.; Cho, A.; Venkateswaran, N.; Olivier, G. K.; Olson, A. M.; Tran, H.; Harada, R. M.; Tan, L.; Zuckermann, R. N. Shaken, Not Stirred: Collapsing a Peptoid Monolayer To Produce Free-Floating, Stable Nanosheets. *J. Am. Chem. Soc.* **2011**, *133* (51), 20808–20815.
(108) Mannige, R. V.; Haxton, T. K.; Proulx, C.; Robertson, E. J.; Battigelli, A.; Butterfoss, G. L.; Zuckermann, R. N.; Whitelam, S. Peptoid Nanosheets Exhibit a New Secondary-Structure Motif. *Nature* **2015**, *526* (7573), 415–420.
(109) Brandt, W.; Herberg, T.; Wessjohann, L. Systematic Conformational Investigations of Peptoids and Peptoid–Peptide Chimeras. *Pept. Sci.* **2011**, *96* (5), 651–668.
(110) Voelz, V. A.; Dill, K. A.; Chorny, I. Peptoid Conformational Free Energy Landscapes from Implicit-Solvent Molecular Simulations in AMBER. *Biopolymers* **2011**, *96* (5), 639–650.
(111) Mukherjee, S.; Zhou, G.; Michel, C.; Voelz, V. A. Insights into Peptoid Helix Folding Cooperativity from an Improved Backbone Potential. *J. Phys. Chem. B* **2015**, *119* (50), 15407–15417.
(112) Mirijanian, D. T.; Mannige, R. V.; Zuckermann, R. N.; Whitelam, S. Development and Use of an Atomistic CHARMM-Based Forcefield for Peptoid Simulation. *J. Comput. Chem.* **2014**, *35* (5), 360–370.
(113) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollmann, P. A. Application of RESP Charges to Calculate Conformational Energies, Hydrogen Bond Energies, and Free Energies of Solvation. *J. Am. Chem. Soc.* **1993**, *115* (21), 9620–9631.
(114) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graph. Model.* **2006**, *25* (2), 247–260.
(115) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; et al. Gaussian 09 Revision E.01.
(116) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. GROMACS: A Message-Passing Parallel Molecular Dynamics Implementation. *Comput. Phys. Commun.* **1995**, *91* (1), 43–56.
(117) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindah, E. Gromacs: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25.
(118) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (20), 12562–12566.
(119) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A.; et al. PLUMED: A Portable Plugin for Free-Energy Calculations with Molecular Dynamics. *Comput. Phys. Commun.* **2009**, *180*

(10), 1961–1972.
(120) Stewart, J. J. P. Optimization of Parameters for Semiempirical Methods V: Modification of NDDO Approximations and Application to 70 Elements. *J. Mol. Model.* **2007**, *13* (12), 1173–1213.
(121) Zhao, Y.; Truhlar, D. G. The M06 Suite of Density Functionals for Main Group Thermochemistry, Thermochemical Kinetics, Noncovalent Interactions, Excited States, and Transition Elements: Two New Functionals and Systematic Testing of Four M06-Class Functionals and 12 Other Function. *Theor. Chem. Acc.* **2008**, *120* (1–3), 215–241.
(122) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenb, D. J. Gaussian 09, Revision E.01.
(123) Dennington, R.; Keith, T.; Millam, J. GaussView, Version 5. 2009.
(124) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* **1987**, *91* (24), 6269–6271.
(125) Sprenger, K. G.; Pfaendtner, J. Strong Electrostatic Interactions Lead to Entropically Favorable Binding of Peptides to Charged Surfaces. *Langmuir* **2016**, *32* (22), 5690–5701.
(126) Sprenger, K. G.; Prakash, A.; Drobny, G.; Pfaendtner, J. Investigating the Role of Phosphorylation in the Binding of Silaffin Peptide R5 to Silica with Molecular Dynamics Simulations. *Langmuir* **2018**, *34* (3), 1199–1207.
(127) Deighan, M.; Pfaendtner, J. Exhaustively Sampling Peptide Adsorption with Metadynamics. *Langmuir* **2013**, *29* (25), 7999–8009.
(128) Tiwary, P.; Parrinello, M. A Time-Independent Free Energy Estimator for Metadynamics. *J. Phys. Chem. B* **2015**, *119* (3), 736–742.
(129) Fleming, K. L.; Tiwary, P.; Pfaendtner, J. New Approach for Investigating Reaction Dynamics and Rates with Ab Initio Calculations. *J. Phys. Chem. A* **2016**, *120* (2), 299–305.
(130) Baldauf, C.; Günther, R.; Hofmann, H.-J. Helices in Peptoids of α- and β-Peptides. *Phys. Biol.* **2006**, *3* (1), S1–S9.
(131) Butterfoss, G. L.; Renfrew, P. D.; Kuhlman, B.; Kirshenbaum, K.; Bonneau, R. A Preliminary Survey of the Peptoid Folding Landscape. *J. Am. Chem. Soc.* **2009**, *131* (46), 16798–16807.
(132) Head-Gordon, T.; Head-Gordon, M.; Frisch, M. J.; Brooks, C. L.; Pople, J. A. Theoretical Study of Blocked Glycine and Alanine Peptide Analogs. *J. Am. Chem. Soc.* **1991**, *113* (16), 5989–5997.
(133) Moehle, K.; Hofmann, H.-J. Peptides and Peptoids-A Quantum Chemical Structure Comparison. *Biopolymers* **1996**, *38* (6), 781–790.
(134) Chen, C.; Qi, J.; Tao, J.; Zuckermann, R. N.; Deyoreo, J. J. Tuning Calcite Morphology and Growth Highly Stable Protein-Mimetics. *Sci. Rep.* **2014**, *4*, 1–11.
(135) Chen, C.-L.; Qi, J.; Zuckermann, R. N.; DeYoreo, J. J. Engineered Biomimetic Polymers as Tunable Agents for Controlling $CaCO_3$ Mineralization. *J. Am. Chem. Soc.* **2011**, *133* (14), 5214–5217.
(136) Řezáč, J.; Fanfrlík, J.; Salahub, D.; Hobza, P. Semiempirical Quantum Chemical PM6 Method Augmented by Dispersion and H-Bonding Correction Terms Reliably Describes Various Types of Noncovalent Complexes. *J. Chem. Theory Comput.* **2009**, *5* (7), 1749–1760.
(137) Levy, Y.; Jortner, J.; Becker, O. M. Solvent Effects on the Energy Landscapes and Folding Kinetics of Polyalanine. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98* (5), 2188–2193.

(138) Lau, K. H. A.; Ren, C.; Sileika, T. S.; Park, S. H.; Szleifer, I.; Messersmith, P. B. Surface-Grafted Polysarcosine as a Peptoid Antifouling Polymer Brush. *Langmuir* **2012**, *28* (46), 16099–16107.
(139) Tang, Y.-C.; Deber, C. M. Hydrophobicity and Helicity of Membrane-Interactive Peptides Containing Peptoid Residues. *Biopolymers* **2002**, *65* (4), 254–262.
(140) Arai, K.; Sagawa, N.; Shikata, T.; Sternhagen, G. L.; Li, X.; Guo, L.; Do, C.; Zhang, D. Pronounced Dielectric and Hydration/Dehydration Behaviors of Monopolar Poly(*N* -Alkylglycine)s in Aqueous Solution. *J. Phys. Chem. B* **2016**, *120* (37), 9978–9986.
(141) Sisido, M.; Imanishi, Y.; Higashimura, T. Static and Dynamic Studies on the End-to-End Intrachain Energy Transfer on a Polysarcosine Chain. *Macromolecules* **1979**, *12* (5), 975–980.
(142) Robertson, E. J.; Battigelli, A.; Proulx, C.; Mannige, R. V.; Haxton, T. K.; Yun, L.; Whitelam, S.; Zuckermann, R. N. Design, Synthesis, Assembly, and Engineering of Peptoid Nanosheets. *Acc. Chem. Res.* **2016**, *49* (3), 379–389.
(143) Olivier, G. K.; Cho, A.; Sanii, B.; Connolly, M. D.; Tran, H.; Zuckermann, R. N. Antibody-Mimetic Peptoid Nanosheets for Molecular Recognition. *ACS Nano* **2013**, *7* (10), 9276–9286.
(144) Huang, K.; Wu, C. W.; Sanborn, T. J.; Patch, J. A.; Kirshenbaum, K.; Zuckermann, R. N.; Barron, A. E.; Radhakrishnan, I. A Threaded Loop Conformation Adopted by a Family of Peptoid Nonamers. *J. Am. Chem. Soc.* **2006**, *128* (5), 1733–1738.
(145) Fuller, A. A.; Yurash, B. A.; Schaumann, E. N.; Seidl, F. J. Self-Association of Water-Soluble Peptoids Comprising (S)-N-1- (Naphthylethyl)Glycine Residues. *Org. Lett.* **2013**, *15* (19), 5118–5121.
(146) Kollman, P. Free Energy Calculations: Applications to Chemical and Biochemical Phenomena. *Chem. Rev.* **1993**, *93* (7), 2395–2417.
(147) Branduardi, D.; Bussi, G.; Parrinello, M. Metadynamics with Adaptive Gaussians. *J. Chem. Theory Comput.* **2012**, *8* (7), 2247–2254.
(148) Bussi, G.; Gervasio, F. L.; Laio, A.; Parrinello, M. Free-Energy Landscape for β Hairpin Folding from Combined Parallel Tempering and Metadynamics. *J. Am. Chem. Soc.* **2006**, *128* (41), 13435–13441.
(149) Piana, S.; Laio, A. A Bias-Exchange Approach to Protein Folding. *J. Phys. Chem. B* **2007**, *111* (17), 4553–4559.
(150) Limongelli, V.; Bonomi, M.; Parrinello, M. Funnel Metadynamics as Accurate Binding Free-Energy Method. *Proc. Natl. Acad. Sci.* **2013**, *110* (16), 6358–6363.
(151) Soderhjelm, P.; Tribello, G. A.; Parrinello, M. Locating Binding Poses in Protein-Ligand Systems Using Reconnaissance Metadynamics. *Proc. Natl. Acad. Sci.* **2012**, *109* (14), 5170–5175.
(152) Gervasio, F. L.; Laio, A.; Parrinello, M. Flexible Docking in Solution Using Metadynamics. *J. Am. Chem. Soc.* **2005**, *127* (8), 2600–2607.
(153) Biarnés, X.; Bongarzone, S.; Vargiu, A. V.; Carloni, P.; Ruggerone, P. Molecular Motions in Drug Design: The Coming Age of the Metadynamics Method. *J. Comput. Aided. Mol. Des.* **2011**, *25* (5), 395–402.
(154) Meißner, R. H.; Wei, G.; Ciacchi, L. C. Estimation of the Free Energy of Adsorption of a Polypeptide on Amorphous SiO<inf>2</Inf> from Molecular Dynamics Simulations and Force Spectroscopy Experiments. *Soft Matter* **2015**, *11* (31), 6254–6265.
(155) Schneider, J.; Colombi Ciacchi, L. Specific Material Recognition by Small Peptides Mediated by the Interfacial Solvent Structure. *J. Am. Chem. Soc.* **2012**, *134* (4), 2407–2413.

(156) Wright, L. B.; Palafox-Hernandez, J. P.; Rodger, P. M.; Corni, S.; Walsh, T. R. Facet Selectivity in Gold Binding Peptides: Exploiting Interfacial Water Structure. *Chem. Sci.* **2015**, *6* (9), 5204–5214.
(157) Palafox-Hernandez, J. P.; Tang, Z.; Hughes, Z. E.; Li, Y.; Swihart, M. T.; Prasad, P. N.; Walsh, T. R.; Knecht, M. R. Comparative Study of Materials-Binding Peptide Interactions with Gold and Silver Surfaces and Nanostructures: A Thermodynamic Basis for Biological Selectivity of Inorganic Materials. *Chem. Mater.* **2014**, *26* (17), 4960–4969.
(158) Wright, L. B.; Freeman, C. L.; Walsh, T. R. Benzene Adsorption at the Aqueous (0 1 1) α-Quartz Interface: Is Surface Flexibility Important? *Mol. Simul.* **2013**, *39* (13), 1093–1102.
(159) Sultan, A. M.; Hughes, Z. E.; Walsh, T. R. Binding Affinities of Amino Acid Analogues at the Charged Aqueous Titania Interface: Implications for Titania-Binding Peptides. *Langmuir* **2014**, *30* (44), 13321–13329.
(160) Sultan, A. M.; Westcott, Z. C.; Hughes, Z. E.; Palafox-Hernandez, J. P.; Giesa, T.; Puddu, V.; Buehler, M. J.; Perry, C. C.; Walsh, T. R. Aqueous Peptide-TiO2Interfaces: Isoenergetic Binding via Either Entropically or Enthalpically Driven Mechanisms. *ACS Appl. Mater. Interfaces* **2016**, *8* (28), 18620–18630.
(161) Meißner, R. H.; Schneider, J.; Schiffels, P.; Colombi Ciacchi, L. Computational Prediction of Circular Dichroism Spectra and Quantification of Helicity Loss upon Peptide Adsorption on Silica. *Langmuir* **2014**, *30* (12), 3487–3494.
(162) Brandt, E. G.; Lyubartsev, A. P. Molecular Dynamics Simulations of Adsorption of Amino Acid Side Chain Analogues and a Titanium Binding Peptide on the TiO 2 (100) Surface. *J. Phys. Chem. C* **2015**, *119* (32), 18126–18139.
(163) Hughes, Z. E.; Walsh, T. R. What Makes a Good Graphene-Binding Peptide? Adsorption of Amino Acids and Peptides at Aqueous Graphene Interfaces. *J. Mater. Chem. B* **2015**, *3* (16), 3211–3221.
(164) Elder, R. M.; Pfaendtner, J.; Jayaraman, A. Effect of Hydrophobic and Hydrophilic Surfaces on the Stability of Double-Stranded DNA. *Biomacromolecules* **2015**, *16* (6), 1862–1869.
(165) Levine, Z. A.; Fischer, S. A.; Shea, J.-E.; Pfaendtner, J. Trp-Cage Folding on Organic Surfaces. *J. Phys. Chem. B* **2015**, *119* (33), 10417–10425.
(166) Wu, P.; Hu, X.; Yang, W. λ-Metadynamics Approach To Compute Absolute Solvation Free Energy. *J. Phys. Chem. Lett.* **2011**, *2* (17), 2099–2103.
(167) Mori, T.; Hamers, R. J.; Pedersen, J. A.; Cui, Q. An Explicit Consideration of Desolvation Is Critical to Binding Free Energy Calculations of Charged Molecules at Ionic Surfaces. *J. Chem. Theory Comput.* **2013**, *9* (11), 5059–5069.
(168) Quigley, D.; Freeman, C. L.; Harding, J. H.; Rodger, P. M. Sampling the Structure of Calcium Carbonate Nanoparticles with Metadynamics. *J. Chem. Phys.* **2011**, *134* (4), 044703.
(169) Wu, C.; Skelton, A. A.; Chen, M.; Vlček, L.; Cummings, P. T. Modeling the Interaction between Integrin-Binding Peptide (RGD) and Rutile Surface: The Effect of Cation Mediation on Asp Adsorption. *Langmuir* **2012**, *28* (5), 2799–2811.
(170) Heinz, H.; Lin, T.-J.; Kishore Mishra, R.; Emami, F. S. Thermodynamically Consistent Force Fields for the Assembly of Inorganic, Organic, and Biological Nanostructures: The INTERFACE Force Field. *Langmuir* **2013**, *29* (6), 1754–1765.
(171) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. Interaction Models for Water in Relation to Protein Hydration. *Intermol. Forces* **1981**, 331–342.
(172) Humphrey, W.; Dalke, A.; Schulten, K. VMD - Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38.

(173) Emami, F. S.; Puddu, V.; Berry, R. J.; Varshney, V.; Patwardhan, S. V.; Perry, C. C.; Heinz, H. Force Field and a Surface Model Database for Silica to Simulate Interfacial Properties in Atomic Resolution. *Chem. Mater.* **2014**, *26* (8), 2647–2658.
(174) Bonomi, M.; Parrinello, M. Enhanced Sampling in the Well-Tempered Ensemble. *Phys. Rev. Lett.* **2010**, *104* (19), 190601.
(175) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New Feathers for an Old Bird. *Comput. Phys. Commun.* **2014**, *185* (2), 604–613.
(176) Bussi, G.; Branduardi, D. Free-Energy Calculations with Metadynamics: Theory and Practice; Wiley-Blackwell, 2015; pp 1–49.
(177) Prakash, M. K.; Barducci, A.; Parrinello, M. Replica Temperatures for Uniform Exchange and Efficient Roundtrip Times in Explicit Solvent Parallel Tempering Simulations. *J. Chem. Theory Comput.* **2011**, *7* (7), 2025–2027.
(178) COLLINS, K.; WASHABAUGH, M. The Hofmeister Effect and the Behavior of Water At Interfaces. *Q. Rev. Biophys.* **1985**, *18* (4), 323–422.
(179) Israelachvili, J. N. *Intermolecular and Surface Forces*; Academic Press, 2011.
(180) Sumper, M.; Kröger, N. Silica Formation in Diatoms: The Function of Long-Chain Polyamines and Silaffins. *Journal of Materials Chemistry*. 2004, pp 2059–2065.
(181) Lechner, C. C.; Becker, C. F. W. Modified Silaffin R5 Peptides Enable Encapsulation and Release of Cargo Molecules from Biomimetic Silica Particles. *Bioorganic Med. Chem.* **2013**, *21* (12), 3533–3541.
(182) Otzen, D. The Role of Proteins in Biosilicification. *Scientifica (Cairo).* **2012**, *2012*, 1–22.
(183) Knecht, M. R.; Wright, D. W. Functional Analysis of the Biomimetic Silica Precipitating Activity of the R5 Peptide from Cylindrotheca Fusiformis. *Chem. Commun.* **2003**, *0* (24), 3038.
(184) Ndao, M.; Goobes, G.; Emani, P. S.; Drobny, G. P. A REDOR SsNMR Investigation of the Role of an N-Terminus Lysine in R5 Silica Recognition. *Langmuir* **2016**, acs.langmuir.5b04114.
(185) Senior, L.; Crump, M. P.; Williams, C.; Booth, P. J.; Mann, S.; Perriman, A. W.; Curnow, P. Structure and Function of the Silicifying Peptide R5. *J. Mater. Chem. B* **2015**, *3* (13), 2607–2614.
(186) Roehrich, A.; Ash, J.; Zane, A.; Masica, D. L.; Gray, J. J.; Goobes, G.; Drobny, G. Solid-State NMR Studies of Biomineralization Peptides and Proteins. In *ACS Symposium Series*; 2012; Vol. 1120, pp 77–96.
(187) Kröger, N.; Lorenz, S.; Brunner, E.; Sumper, M. Self-Assembly of Highly Phosphorylated Silaffins and Their Function in Biosilica Morphogenesis. *Science (80-.).* **2002**, *298* (5593), 584–586.
(188) Kroger, N.; Deutzmann, R.; Bergsdorf, C.; Sumper, M. Species-Specific Polyamines from Diatoms Control Silica Morphology. *Proc. Natl. Acad. Sci.* **2000**, *97* (26), 14133–14138.
(189) Poulsen, N.; Kröger, N. Silica Morphogenesis by Alternative Processing of Silaffins in the Diatom Thalassiosira Pseudonana. *J. Biol. Chem.* **2004**, *279* (41), 42993–42999.
(190) Lechner, C. C.; Becker, C. F. W. Exploring the Effect of Native and Artificial Peptide Modifications on Silaffin Induced Silica Precipitation. *Chem. Sci.* **2012**, *3* (12), 3500.
(191) Francisco Rodríguez, †,‡; Diana D. Glawe, §; Rajesh R. Naik, †; Kevin P. Hallinan, ‡ and; Morley O. Stone*, †. Study of the Chemical and Physical Influences upon in Vitro Peptide-Mediated Silica Formation. **2004**.
(192) and, L. L.; Camp*, P. J. Self-Assembly of Peptide Scaffolds in Biosilica Formation: Computer Simulations of a Coarse-Grained Model. **2006**.
(193) Kröger, N.; Deutzmann, R.; Sumper, M. Polycationic Peptides from Diatom Biosilica

That Direct Silica Nanosphere Formation. *Science (80-.).* **1999**, *286* (5442), 1129–1132.
(194) Eby, D. M.; Johnson, G. R.; Farmer, B. L.; Pandey, R. B. Supramolecular Assembly of a Biomineralizing Antimicrobial Peptide in Coarse-Grained Monte Carlo Simulations. *Phys. Chem. Chem. Phys.* **2011**, *13* (3), 1123–1130.
(195) Lutz, H.; Jaeger, V.; Schmüser, L.; Bonn, M.; Pfaendtner, J.; Weidner, T. The Structure of the Diatom Silaffin Peptide R5 within Freestanding Two-Dimensional Biosilica Sheets. *Angew. Chemie Int. Ed.* **2017**, *56* (28), 8277–8280.
(196) Schwierz, N.; Horinek, D.; Liese, S.; Pirzer, T.; Balzer, B. N.; Hugel, T.; Netz, R. R. On the Relationship between Peptide Adsorption Resistance and Surface Contact Angle: A Combined Experimental and Simulation Single-Molecule Study. *J. Am. Chem. Soc.* **2012**, *134* (48), 19628–19638.
(197) Xie, Y.; Jiang, Y.; Ben-Amotz, D. Detection of Amino Acid and Peptide Phosphate Protonation Using Raman Spectroscopy. *Anal. Biochem.* **2005**, *343* (2), 223–230.
(198) Walsh, T. R. Pathways to Structure–Property Relationships of Peptide–Materials Interfaces: Challenges in Predicting Molecular Structures. *Acc. Chem. Res.* **2017**, *50* (7), 1617–1624.
(199) Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E. Peptide Folding: When Simulation Meets Experiment. *Angew. Chemie Int. Ed.* **1999**, *38* (1–2), 236–240.
(200) Fraccalvieri, D.; Pandini, A.; Stella, F.; Bonati, L. Conformational and Functional Analysis of Molecular Dynamics Trajectories by Self-Organising Maps. *BMC Bioinformatics* **2011**, *12* (1), 158.
(201) Lechner, C. C.; Becker, C. F. W. A Sequence-Function Analysis of the Silica Precipitating Silaffin R5 Peptide. *J. Pept. Sci.* **2014**, *20* (2), 152–158.
(202) Ravera, E.; Cerofolini, L.; Martelli, T.; Louka, A.; Fragai, M.; Luchinat, C. 1H-Detected Solid-State NMR of Proteins Entrapped in Bioinspired Silica: A New Tool for Biomaterials Characterization. *Sci. Rep.* **2016**, *6* (1), 27851.
(203) Kharlampieva, E.; Jung, C. M.; Kozlovskaya, V.; Tsukruk, V. V. Secondary Structure of Silaffin at Interfaces and Titania Formation. *J. Mater. Chem.* **2010**, *20* (25), 5242.
(204) Patwardhan, S. V.; Emami, F. S.; Berry, R. J.; Jones, S. E.; Naik, R. R.; Deschaume, O.; Heinz, H.; Perry, C. C. Chemistry of Aqueous Silica Nanoparticle Surfaces and the Mechanism of Selective Peptide Adsorption. *J. Am. Chem. Soc.* **2012**, *134* (14), 6244–6256.
(205) Limo, M. J.; Perry, C. C. Thermodynamic Study of Interactions Between ZnO and ZnO Binding Peptides Using Isothermal Titration Calorimetry. *Langmuir* **2015**, *31* (24), 6814–6822.
(206) Voter, A. F. Hyperdynamics: Accelerated Molecular Dynamics of Infrequent Events. *Phys. Rev. Lett.* **1997**, *78* (20), 3908–3911.
(207) Valsson, O.; Parrinello, M. Variational Approach to Enhanced Sampling and Free Energy Calculations. *Phys. Rev. Lett.* **2014**, *113* (9), 1–5.
(208) Comer, J.; Gumbart, J. C.; Hénin, J.; Lelièvre, T.; Pohorille, A.; Chipot, C. The Adaptive Biasing Force Method: Everything You Always Wanted To Know but Were Afraid To Ask. *J. Phys. Chem. B* **2015**, *119* (3), 1129–1151.
(209) Gil-Ley, A.; Bussi, G. Enhanced Conformational Sampling Using Replica Exchange with Collective-Variable Tempering. *J. Chem. Theory Comput.* **2015**, *11* (3), 1077–1085.
(210) Noé, F.; Clementi, C. Kinetic Distance and Kinetic Maps from Molecular Dynamics Simulation. *J. Chem. Theory Comput.* **2015**, *11* (10), 5002–5011.
(211) Tribello, G. A.; Cuny, J.; Eshet, H.; Parrinello, M. Exploring the Free Energy Surfaces of

Clusters Using Reconnaissance Metadynamics. *J. Chem. Phys.* **2011**, *135* (11), 114109.
(212) Tiwary, P.; Berne, B. J. Spectral Gap Optimization of Order Parameters for Sampling Complex Molecular Systems. **2015**, *113* (11), 2839–2844.
(213) Hošek, P.; Toulcová, D.; Bortolato, A.; Spiwok, V. Altruistic Metadynamics: Multisystem Biased Simulation. *J. Phys. Chem. B* **2016**, *120* (9), 2209–2215.
(214) Šućur, Z.; Spiwok, V. Sampling Enhancement and Free Energy Prediction by the Flying Gaussian Method. *J. Chem. Theory Comput.* **2016**, *12* (9), 4644–4650.
(215) Domene, C.; Barbini, P.; Furini, S. Bias-Exchange Metadynamics Simulations: An Efficient Strategy for the Analysis of Conduction and Selectivity in Ion Channels. *J. Chem. Theory Comput.* **2015**, *11* (4), 1896–1906.
(216) Heller, G. T.; Aprile, F. A.; Bonomi, M.; Camilloni, C.; De Simone, A.; Vendruscolo, M. Sequence Specificity in the Entropy-Driven Binding of a Small Molecule and a Disordered Peptide. *J. Mol. Biol.* **2017**, *429* (18), 2772–2779.
(217) Fu, C. D.; Pfaendtner, J. Lifting the Curse of Dimensionality on Enhanced Sampling of Reaction Networks with Parallel Bias Metadynamics. *J. Chem. Theory Comput.* **2018**, *14*, 2516–2525.
(218) Tiwary, P.; Parrinello, M. A Time-Independent Free Energy Estimator for Metadynamics. *J. Phys. Chem. B* **2015**, *119* (3), 736–742.
(219) Laio, A.; Rodriguez-Fortea, A.; Gervasio, F. L.; Ceccarelli, M.; Parrinello, M. Assessing the Accuracy of Metadynamics. *J. Phys. Chem. B* **2005**, *109* (14), 6714–6721.
(220) Doye, J. P. K.; Miller, M. A.; Wales, D. J. Evolution of the Potential Energy Surface with Size for Lennard-Jones Clusters. *J. Chem. Phys.* **1999**, *111* (18), 8417–8428.
(221) Pietrucci, F.; Andreoni, W. Graph Theory Meets Ab Initio Molecular Dynamics: Atomic Structures and Transformations at the Nanoscale. *Phys. Rev. Lett.* **2011**, *107* (8), 085504.
(222) Nava, M.; Palazzesi, F.; Perego, C.; Parrinello, M. Dimer Metadynamics. *J. Chem. Theory Comput.* **2017**, *13* (2), 425–430.
(223) Tribello, G. A.; Ceriotti, M.; Parrinello, M. A Self-Learning Algorithm for Biased Molecular Dynamics. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107* (41), 17509–17514.
(224) Wales, D. J. Discrete Path Sampling. *Mol. Phys.* **2002**, *100* (20), 3285–3305.
(225) Fennell, C. J.; Bizjak, A.; Vlachy, V.; Dill, K. A. Ion Pairing in Molecular Simulations of Aqueous Alkali Halide Solutions. *J. Phys. Chem. B* **2009**, *113* (19), 6782–6791.
(226) General, I. J. A Note on the Standard State's Binding Free Energy. *J. Chem. Theory Comput.* **2010**, *6* (8), 2520–2524.
(227) Laage, D.; Hynes, J. T. On the Residence Time for Water in a Solute Hydration Shell: Application to Aqueous Halide Solutions. *J. Phys. Chem. B* **2008**, *112* (26), 7697–7701.
(228) Spagnoli, D.; Banfield, J. F.; Parker, S. C. Free Energy Change of Aggregation of Nanoparticles. *J. Phys. Chem. C* **2008**, *112* (38), 14731–14736.
(229) Mo, Y.; Lu, Y.; Wei, G.; Derreumaux, P. Structural Diversity of the Soluble Trimers of the Human Amylin(20-29) Peptide Revealed by Molecular Dynamics Simulations. *J. Chem. Phys.* **2009**, *130* (12), 125101–212413.
(230) Buchanan, L. E.; Dunkelberger, E. B.; Tran, H. Q.; Cheng, P.-N.; Chiu, C.-C.; Cao, P.; Raleigh, D. P.; de Pablo, J. J.; Nowick, J. S.; Zanni, M. T. Mechanism of IAPP Amyloid Fibril Formation Involves an Intermediate with a Transient -Sheet. *Proc. Natl. Acad. Sci.* **2013**, *110* (48), 19285–19290.
(231) Elhadj, S.; De Yoreo, J. J.; Hoyer, J. R.; Dove, P. M. Role of Molecular Charge and Hydrophilicity in Regulating the Kinetics of Crystal Growth. *Proc. Natl. Acad. Sci.* **2006**, *103* (51), 19237–19242.
(232) Boskey, A. L.; Maresca, M.; Ullrich, W.; Doty, S. B.; Butler, W. T.; Prince, C. W.

Osteopontin-Hydroxyapatite Interactions in Vitro: Inhibition of Hydroxyapatite Formation and Growth in a Gelatin-Gel. *Bone Miner.* **1993**, *22* (2), 147–159.

(233) Shiraga, H.; Min, W.; VanDusen, W. J.; Clayman, M. D.; Miner, D.; Terrell, C. H.; Sherbotie, J. R.; Foreman, J. W.; Przysiecki, C.; Neilson, E. G. Inhibition of Calcium Oxalate Crystal Growth in Vitro by Uropontin: Another Member of the Aspartic Acid-Rich Protein Superfamily. *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89* (1), 426–430.

(234) Naka, K.; Chujo, Y. Control of Crystal Nucleation and Growth of Calcium Carbonate by Synthetic Substrates. *Chemistry of Materials*. American Chemical Society 2001, pp 3245–3259.

(235) Söllner, C.; Burghammer, M.; Busch-Nentwich, E.; Berger, J.; Schwarz, H.; Riekel, C.; Nicolson, T. Control of Crystal Size and Lattice Formation by Starmaker in Otolith Biomineratization. *Science (80-.).* **2003**, *302* (5643), 282–286.

(236) Kröger, N.; Poulsen, N. Diatoms—From Cell Wall Biogenesis to Nanotechnology. *Annu. Rev. Genet.* **2008**, *42* (1), 83–107.

(237) Weiner, S.; Addadi, L. Design Strategies in Mineralized Biological Materials. *J. Mater. Chem* **1997**, *7* (5), 689–702.

(238) Goobes, G.; Goobes, R.; Shaw, W. J.; Gibson, J. M.; Long, J. R.; Raghunathan, V.; Schueler-Furman, O.; Popham, J. M.; Baker, D.; Campbell, C. T.; et al. The Structure, Dynamics, and Energetics of Protein Adsorption - Lessons Learned from Adsorption of Statherin to Hydroxyapatite. *Magnetic Resonance in Chemistry*. John Wiley & Sons, Ltd. December 2007, pp S32–S47.

(239) DeGrado, W. F.; Lear, J. D. Induction of Peptide Conformation at Apolar Water Interfaces. 1. A Study with Model Peptides of Defined Hydrophobic Periodicity. *J. Am. Chem. Soc.* **1985**, *107* (25), 7684–7689.

(240) Masica, D. L.; Gray, J. J.; Shaw, W. J. Partial High-Resolution Structure of Phosphorylated and Non-Phosphorylated Leucine-Rich Amelogenin Protein Adsorbed to Hydroxyapatite. *J. Phys. Chem. C* **2011**, *115* (28), 13775–13785.

(241) Gibson, J. M.; Popham, J. M.; Raghunathan, V.; Stayton, P. S.; Drobny, G. P. 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.* **2006**, *128* (16), 5364–5370.

(242) Gibson, J. M.; Raghunathan, V.; Popham, J. M.; Stayton, P. S.; Drobny, G. P. A REDOR NMR Study of a Phosphorylated Statherin Fragment Bound to Hydroxyapatite Crystals. *J. Am. Chem. Soc.* **2005**, *127* (26), 9350–9351.

(243) Masica, D. L.; Ash, J. T.; Ndao, M.; Drobny, G. P.; Gray, J. J. Toward a Structure Determination Method for Biomineral-Associated Protein Using Combined Solid- State NMR and Computational Structure Prediction. *Structure* **2010**, *18* (12), 1678–1687.

(244) Makrodimitris, K.; Masica, D. L.; Kim, E. T.; Gray, J. J. Structure Prediction of Protein-Solid Surface Interactions Reveals a Molecular Recognition Motif of Statherin for Hydroxyapatite. *J. Am. Chem. Soc.* **2007**, *129* (44), 13713–13722.

(245) Ndao, M.; Ash, J. T.; Breen, N. F.; Goobes, G.; Stayton, P. S.; Drobny, G. P. A 13C{ 31P} REDOR NMR Investigation of the Role of Glutamic Acid Residues in Statherin-Hydroxyapatite Recognition. *Langmuir* **2009**, *25* (20), 12136–12143.

(246) Weidner, T.; Dubey, M.; Breen, N. F.; Ash, J.; Baio, J. E.; Jaye, C.; Fischer, D. A.; Drobny, G. P.; Castner, D. G. Direct Observation of Phenylalanine Orientations in Statherin Bound to Hydroxyapatite Surfaces. *J. Am. Chem. Soc.* **2012**, *134* (21), 8750–8753.

(247) Weidner, T.; Breen, N. F.; Li, K.; Drobny, G. P.; Castner, D. G. Sum Frequency

Generation and Solid-State NMR Study of the Structure, Orientation, and Dynamics of Polystyrene-Adsorbed Peptides. *Proc. Natl. Acad. Sci.* **2010**, *107* (30), 13288–13293.
(248) Ndao, M.; Ash, J. T.; Stayton, P. S.; Drobny, G. P. The Role of Basic Amino Acids in the Molecular Recognition of Hydroxyapatite by Statherin Using Solid State NMR. *Surf. Sci.* **2010**, *604* (15–16), L39–L42.
(249) Li, K.; Emani, P. S.; Ash, J.; Groves, M.; Drobny, G. P. A Study of Phenylalanine Side-Chain Dynamics in Surface-Adsorbed Peptides Using Solid-State Deuterium NMR and Rotamer Library Statistics. *J. Am. Chem. Soc.* **2014**, *136* (32), 11402–11411.
(250) Goobes, R.; Goobes, G.; Shaw, W. J.; Drobny, G. P.; Campbell, C. T.; Stayton, P. S. Thermodynamic Roles of Basic Amino Acids in Statherin Recognition of Hydroxyapatite. *Biochemistry* **2007**, *46* (16), 4725–4733.
(251) Raj, P. A.; Johnsson, M.; Levine, M. J.; Nancollas, G. H. Salivary Statherin: Dependence on Sequence, Charge, Hydrogen Bonding Potency, and Helical Conformation for Adsorption to Hydroxyapatite and Inhibition of Mineralization. *J. Biol. Chem.* **1992**, *267* (9), 5968–5976.
(252) Goobes, G.; Goobes, R.; Schueler-Furman, O.; Baker, D.; Stayton, P. S.; Drobny, G. P. Folding of the C-Terminal Bacterial Binding Domain in Statherin upon Adsorption onto Hydroxyapatite Crystals. *Proc. Natl. Acad. Sci.* **2006**, *103* (44), 16083–16088.
(253) Long, J. R.; Shaw, W. J.; Stayton, P. S.; Drobny, G. P. Structure and Dynamics of Hydrated Statherin on Hydroxyapatite as Determined by Solid-State NMR. *Biochemistry* **2001**, *40* (51), 15451–15455.
(254) Moreno, E. C.; Kresak, M.; Hay, D. I. Adsorption of Molecules of Biological Interest onto Hydroxyapatite. *Calcif. Tissue Int.* **1984**, *36* (1), 48–59.
(255) Elgavish, G. A.; Hay, D. I.; Schlesinger, D. H. 1H and 31P Nuclear Magnetic Resonance Studies of Human Salivary Statherin. *Int. J. Pept. Protein Res.* **1984**, *23* (3), 230–234.
(256) Naganagowda, G. A.; Gururaja, T. L.; Levine, M. J. Delineation of Conformational Preferences in Human Salivary Statherin By1H,31P NMR and CD Studies: Sequential Assignment and Structure-Function Correlations. *J. Biomol. Struct. Dyn.* **1998**, *16* (1), 91–107.
(257) Roehrich, A.; Drobny, G. Solid-State NMR Studies of Biomineralization Peptides and Proteins. *Acc. Chem. Res.* **2013**, *46* (9), 2136–2144.
(258) Raghunathan, V.; Gibson, J. M.; Goobes, G.; Popham, J. M.; Louie, E. A.; Stayton, P. S.; Drobny, G. P. Homonuclear and Heteronuclear NMR Studies of a Statherin Fragment Bound to Hydroxyapatite Crystals. *J. Phys. Chem. B* **2006**, *110* (18), 9324–9332.
(259) Douglas, W. H.; Reeh, E. S.; Ramasubbu, N.; Raj, P. A.; Bhandary, K. K.; Levine, M. J. Statherin: A Major Boundary Lubricant of Human Saliva. *Biochem. Biophys. Res. Commun.* **1991**, *180* (1), 91–97.
(260) Shaw, W. J.; Long, J. R.; Campbell, A. A.; Stayton, P. S.; Drobny, G. P. A Solid State NMR Study of Dynamics in a Hydrated Salivary Peptide Adsorbed to Hydroxyapatite [3]. *Journal of the American Chemical Society*. 2000, pp 7118–7119.
(261) Shaw, W. J.; Long, J. R.; Dindot, J. L.; Campbell, A. A.; Stayton, P. S.; Drobny, G. P. Determination of Statherin N-Terminal Peptide Conformation on Hydroxyapatite Crystals. *J. Am. Chem. Soc.* **2000**, *122* (8), 1709–1716.
(262) Masica, D. L.; Gray, J. J. Solution- and Adsorbed-State Structural Ensembles Predicted for the Statherin-Hydroxyapatite System. *Biophys. J.* **2009**, *96* (8), 3082–3091.
(263) Neal, S.; Nip, A. M.; Zhang, H.; Wishart, D. S. Rapid and Accurate Calculation of Protein 1H, 13C and 15N Chemical Shifts. *J. Biomol. NMR* **2003**, *26* (3), 215–240.
(264) Kohlhoff, K. J.; Robustelli, P.; Cavalli, A.; Salvatella, X.; Vendruscolo, M. Fast and

Accurate Predictions of Protein NMR Chemical Shifts from Interatomic Distances. *J. Am. Chem. Soc.* **2009**, *131* (39), 13894–13895.
(265) Bonomi, M.; Camilloni, C.; Cavalli, A.; Vendruscolo, M. Metainference: A Bayesian Inference Method for Heterogeneous Systems. *Sci. Adv.* **2016**, *2* (1), e1501177.
(266) Bonomi, M.; Camilloni, C. Integrative Structural and Dynamical Biology with PLUMED-ISDB. *Bioinformatics* **2017**, *33* (24), 3999–4000.
(267) Shen, Y.; Bax, A. Protein Backbone and Sidechain Torsion Angles Predicted from NMR Chemical Shifts Using Artificial Neural Networks. *J. Biomol. NMR* **2013**, *56* (3), 227–241.
(268) Robustelli, P.; Kohlhoff, K.; Cavalli, A.; Vendruscolo, M. Using NMR Chemical Shifts as Structural Restraints in Molecular Dynamics Simulations of Proteins. *Structure* **2010**, *18* (8), 923–933.
(269) Boehr, D. D.; Nussinov, R.; Wright, P. E. The Role of Dynamic Conformational Ensembles in Biomolecular Recognition. *Nature Chemical Biology*. Nature Publishing Group November 10, 2009, pp 789–796.
(270) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved Side-Chain Torsion Potentials for the Amber Ff99SB Protein Force Field. *Proteins Struct. Funct. Bioinforma.* **2010**, *78* (8), NA-NA.
(271) Perera, L.; Berkowitz, M. L. Many-Body Effects in Molecular Dynamics Simulations of Na+(H2O)n and Cl−(H2O)n Clusters. *J. Chem. Phys.* **1991**, *95* (3), 1954.
(272) Matsui, M.; Akaogi, M. Molecular Dynamics Simulation of the Structural and Physical Properties of the Four Polymorphs of Tio2. *Mol. Simul.* **1991**, *6* (4–6), 239–244.
(273) Emami, F. S.; Puddu, V.; Berry, R. J.; Varshney, V.; Patwardhan, S. V.; Perry, C. C.; Heinz, H. Prediction of Specific Biomolecule Adsorption on Silica Surfaces as a Function of PH and Particle Size. *Chem. Mater.* **2014**, *26* (19), 5725–5734.
(274) Heinz, H. Adsorption of Biomolecules and Polymers on Silicates, Glasses, and Oxides: Mechanisms, Predictions, and Opportunities by Molecular Simulation. *Curr. Opin. Chem. Eng.* **2016**, *11*, 34–41.
(275) Wu, C.; Chen, M.; Xing, C. Molecular Understanding of Conformational Dynamics of a Fibronectin Module on Rutile (110) Surface. *Langmuir* **2010**, *26* (20), 15972–15981.
(276) Berendsen, H. J. C.; Postma, J. P. M.; Van Gunsteren, W. F.; Dinola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81* (8), 3684–3690.
(277) Fernandez, V. L.; Reimer, J. A.; Denn, M. M. Magnetic Resonance Studies of Polypeptides Adsorbed on Silica and Hydroxyapatite Surface. *J. Am. Chem. Soc.* **1992**, *114*, 9634–9642.
(278) Buckle, E. L.; Lum, J. S.; Roehrich, A.; Stote, R. E.; Dracinsky, M.; Filocamo, S. F.; Drobny, G. P. Serine-Lysine Peptides as Mediators for the Production of Titanium Dioxide : Effects of Primary Structure and Secondary Structure Investigated Using Solid-State NMR Spectroscopy and DFT Calculations. *J. Phys. Chem. B* **2018**.
(279) Carnie, S. L.; Torrie, G. M. The Statistical Mechanics of the Electrical Double Layer. *Adv. Chem. Phys.* **2007**, 141–253.
(280) Espinosa-Marzal, R. M.; Drobek, T.; Balmer, T.; Heuberger, M. P. Hydrated-Ion Ordering in Electrical Double Layers. *Phys. Chem. Chem. Phys.* **2012**, *14* (17), 6085.
(281) Ricci, M.; Spijker, P.; Voïtchovsky, K. Water-Induced Correlation between Single Ions Imaged at the Solid-Liquid Interface. *Nat. Commun.* **2014**, *5*, 4400.
(282) Xu, M.; Spinney, R.; Allen, H. C. Water Structure at the Air− Aqueous Interface of Divalent Cation and Nitrate Solutions. *J. Phys. Chem. B* **2009**.

(283) Park, C.; Fenter, P. A.; Nagy, K. L.; Sturchio, N. C. Hydration and Distribution of Ions at the Mica-Water Interface. *Phys. Rev. Lett.* **2006**, *97* (1), 016101.
(284) Lee, S. S.; Fenter, P.; Park, C.; Sturchio, N. C.; Nagy, K. L. Hydrated Cation Speciation at the Muscovite (001)−Water Interface. *Langmuir* **2010**, *26* (22), 16647–16651.
(285) Parsons, D. F.; Boström, M.; Nostro, P. Lo; Ninham, B. W. Hofmeister Effects: Interplay of Hydration, Nonelectrostatic Potentials, and Ion Size. *Phys. Chem. Chem. Phys.* **2011**, *13* (27), 12352.
(286) Bourg, I. C.; Sposito, G. Molecular Dynamics Simulations of the Electrical Double Layer on Smectite Surfaces Contacting Concentrated Mixed Electrolyte (NaCl-CaCl2) Solutions. *J. Colloid Interface Sci.* **2011**, *360* (2), 701–715.
(287) Meleshyn, A. Aqueous Solution Structure at the Cleaved Mica Surface: Influence of K +, H3O+, and Cs+ Adsorption. *J. Phys. Chem. C* **2008**, *112* (50), 20018–20026.
(288) Meleshyn, A. Potential of Mean Force for K^+ in Thin Water Films on Cleaved Mica. *Langmuir* **2010**, *26* (16), 13081–13085.
(289) Meleshyn, A. Adsorption of Sr2+ and Ba2+ at the Cleaved Mica–Water Interface: Free Energy Profiles and Interfacial Structure. *Geochim. Cosmochim. Acta* **2010**, *74* (5), 1485–1497.
(290) Meleshyn, A. Potential of Mean Force for Ca2+ at the Cleaved Mica-Water Interface. *J. Phys. Chem. C* **2009**, *113* (41), 17604–17607.
(291) Schlegel, M. L.; Nagy, K. L.; Fenter, P.; Cheng, L.; Sturchio, N. C.; Jacobsen, S. D. Cation Sorption on the Muscovite (0 0 1) Surface in Chloride Solutions Using High-Resolution X-Ray Reflectivity. *Geochim. Cosmochim. Acta* **2006**, *70* (14), 3549–3565.
(292) Lee, S. S.; Fenter, P.; Nagy, K. L.; Sturchio, N. C. Changes in Adsorption Free Energy and Speciation during Competitive Adsorption between Monovalent Cations at the Muscovite (001)-Water Interface. *Geochim. Cosmochim. Acta* **2013**, *123*, 416–426.
(293) Fenter, P.; Lee, S. S.; Skelton, A. A.; Cummings, P. T. Direct and Quantitative Comparison of Pixelated Density Profiles with High-Resolution X-Ray Reflectivity Data. *J. Synchrotron Radiat.* **2011**, *18* (2), 257–265.
(294) Bourg, I. C.; Lee, S. S.; Fenter, P.; Tournassat, C. Stern Layer Structure and Energetics at Mica–Water Interfaces. *J. Phys. Chem. C* **2017**, *121* (17), 9402–9412.
(295) Vatti, A. K.; Todorova, M.; Neugebauer, J. Ab Initio Determined Phase Diagram of Clean and Solvated Muscovite Mica Surfaces. *Langmuir* **2016**, *32* (4), 1027–1033.
(296) Ou, X.; Wang, X.; Lin, Z.; Li, J. Heterogeneous Condensation of Water on the Mica (001) Surface: A Molecular Dynamics Simulation Work. *J. Phys. Chem. C* **2017**, *121* (12), 6813–6819.
(297) Wang, J.; Kalinichev, A. G.; Kirkpatrick, R. J.; Cygan, R. T. Structure, Energetics, and Dynamics of Water Adsorbed on the Muscovite (001) Surface: A Molecular Dynamics Simulation. *J. Phys. Chem. B* **2005**, *109* (33), 15893–15905.
(298) Prakash, A.; Pfaendtner, J.; Chun, J.; Mundy, C. J. Quantifying the Molecular-Scale Aqueous Response to the Mica Surface. *J. Phys. Chem. C* **2017**, *121* (34), 18496–18504.
(299) Sakuma, H.; Kawamura, K. Structure and Dynamics of Water on Li+-, Na+-, K+-, Cs+-, H3O+-Exchanged Muscovite Surfaces: A Molecular Dynamics Study. *Geochim. Cosmochim. Acta* **2011**, *75* (1), 63–81.
(300) Adapa, S.; Malani, A. Role of Hydration Energy and Co-Ions Association on Monovalent and Divalent Cations Adsorption at Mica-Aqueous Interface. *Sci. Rep.* **2018**, *8* (1), 12198.
(301) Nishimura, S.; Biggs, S.; Scales, P. J.; Healy, T. W.; Tsunematsu, K.; Tateyama, T. Molecular-Scale Structure of the Cation Modified Muscovite Mica Basal Plane. *Langmuir* **1994**, *10* (12), 4554–4559.

(302) Sun, D.; Lakkaraju, S. K.; Jo, S.; MacKerell, A. D. Determination of Ionic Hydration Free Energies with Grand Canonical Monte Carlo/Molecular Dynamics Simulations in Explicit Water. *J. Chem. Theory Comput.* **2018**, acs.jctc.8b00604.
(303) Xu, L.; Salmeron, M. An XPS and Scanning Polarization Force Microscopy Study of the Exchange and Mobility of Surface Ions on Mica. *Langmuir* **1998**, *14* (20), 5841–5844.
(304) Mugele, F.; Bera, B.; Cavalli, A.; Siretanu, I.; Maestro, A.; Duits, M.; Cohen-Stuart, M.; van den Ende, D.; Stocker, I.; Collins, I. Ion Adsorption-Induced Wetting Transition in Oil-Water-Mineral Systems. *Sci. Rep.* **2015**, *5* (1), 10519.
(305) Osman, M. A.; Moor, C.; Caseri, W. R.; Suter, U. W. Alkali Metals Ion Exchange on Muscovite Mica. *J. Colloid Interface Sci.* **1999**, *209* (1), 232–239.
(306) Scales, P. J.; Grieser, F.; Healy, T. W. Electrokinetics of the Muscovite Mica-Aqueous Solution Interface. *Langmuir* **1990**, *6* (3), 582–589.
(307) Beck, T. L. A Local Entropic Signature of Specific Ion Hydration. *J. Phys. Chem. B* **2011**, *115* (32), 9776–9781.
(308) Morris, A. M.; Watzky, M. A.; Finke, R. G. Protein Aggregation Kinetics, Mechanism, and Curve-Fitting: A Review of the Literature. *Biochim. Biophys. Acta - Proteins Proteomics* **2009**, *1794* (3), 375–397.
(309) Roberts, C. J. Non-Native Protein Aggregation Kinetics. *Biotechnol. Bioeng.* **2007**, *98* (5), 927–938.
(310) Petsev, D. N.; Vekilov, P. G. Evidence for Non-DLVO Hydration Interactions in Solutions of the Protein Apoferritin. *Phys. Rev. Lett.* **2000**, *84* (6), 1339–1342.
(311) Prausnitz, J. The Fallacy of Misplaced Concreteness. *Biophysical Journal*. 2015, pp 453–454.
(312) Kuroda, Y.; Suenaga, A.; Sato, Y.; Kosuda, S.; Taiji, M. All-Atom Molecular Dynamics Analysis of Multi-Peptide Systems Reproduces Peptide Solubility in Line with Experimental Observations. *Sci. Rep.* **2016**, *6* (1), 19479.
(313) Barz, B.; Wales, D. J.; Strodel, B. A Kinetic Approach to the Sequence–Aggregation Relationship in Disease-Related Protein Assembly. *J. Phys. Chem. B* **2014**, *118* (4), 1003–1011.
(314) Singh, G.; Brovchenko, I.; Oleinikova, A.; Winter, R. Peptide Aggregation in Finite Systems. *Biophys. J.* **2008**, *95* (7), 3208–3221.
(315) Nguyen, P. H.; Li, M. S.; Stock, G.; Straub, J. E.; Thirumalai, D. Monomer Adds to Preformed Structured Oligomers of Abeta-Peptides by a Two-Stage Dock-Lock Mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (1), 111–116.
(316) Matthes, D.; Gapsys, V.; Daebel, V.; de Groot, B. L. Mapping the Conformational Dynamics and Pathways of Spontaneous Steric Zipper Peptide Oligomerization. *PLoS One* **2011**, *6* (5), e19129.
(317) Batoulis, H.; Schmidt, T. H.; Weber, P.; Schloetel, J.-G.; Kandt, C.; Lang, T. Concentration Dependent Ion-Protein Interaction Patterns Underlying Protein Oligomerization Behaviours. *Sci. Rep.* **2016**, *6* (1), 24131.
(318) Prakash, A.; Baer, M. D.; Mundy, C. J.; Pfaendtner, J. Peptoid Backbone Flexibilility Dictates Its Interaction with Water and Surfaces: A Molecular Dynamics Investigation. *Biomacromolecules* **2018**, *19* (3), 1006–1015.
(319) Daily, M. D.; Baer, M. D.; Mundy, C. J. Divalent Ion Parameterization Strongly Affects Conformation and Interactions of an Anionic Biomimetic Polymer. *J. Phys. Chem. B* **2016**, *120* (9), 2198–2208.
(320) Cecchini, M.; Rao, F.; Seeber, M.; Caflisch, A. Replica Exchange Molecular Dynamics Simulations of Amyloid Peptide Aggregation. *J. Chem. Phys.* **2004**, *121* (21), 10748–

10756.
(321) Haxton, T. K.; Zuckermann, R. N.; Whitelam, S. Implicit-Solvent Coarse-Grained Simulation with a Fluctuating Interface Reveals a Molecular Mechanism for Peptoid Monolayer Buckling. *J. Chem. Theory Comput.* **2016**, *12* (1), 345–352.
(322) Settanni, G.; Schäfer, T.; Muhl, C.; Barz, M.; Schmid, F. Poly-Sarcosine and Poly(Ethylene-Glycol) Interactions with Proteins Investigated Using Molecular Dynamics Simulations. *Comput. Struct. Biotechnol. J.* **2018**.
(323) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. Numerical Recipes in FORTRAN (Cambridge. Cambridge Univ. Press 1992.
(324) Chan, D. Y. C.; Pashley, R. M.; White, L. R. A Simple Algorithm for the Calculation of the Electrostatic Repulsion between Identical Charged Surfaces in Electrolyte. *J Colloid Interface Sci.* **1980**, *77* (1), 283.
(325) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New Feathers for an Old Bird. *Comput. Phys. Commun.* **2014**, *185* (2), 604–613.
(326) Hansen, K.; Biegler, F.; Ramakrishnan, R.; Pronobis, W.; Von Lilienfeld, O. A.; Müller, K. R.; Tkatchenko, A. Machine Learning Predictions of Molecular Properties: Accurate Many-Body Potentials and Nonlocality in Chemical Space. *J. Phys. Chem. Lett.* **2015**, *6* (12), 2326–2331.
(327) Brandt, E. G.; Lyubartsev, A. P. Molecular Dynamics Simulations of Adsorption of Amino Acid Side Chain Analogs and a Titanium Binding Peptide on the TiO2 (100) Surface. *J. Phys. Chem. C* **2015**, *119* (100), 18126–18139.
(328) Healy, K. E.; Ducheyne, P. Hydration and Preferential Molecular Adsorption on Titanium in Vitro. *Biomaterials* **1992**, *13* (8), 553–561.
(329) Merkys, A.; Vaitkus, A.; Butkus, J.; Okulič-Kazarinas, M.; Kairys, V.; Gražulis, S. COD::CIF::Parser: An Error-Correcting CIF Parser for the Perl Language. *J. Appl. Crystallogr.* **2016**, *49* (Pt 1), 292–301.
(330) Gražulis, S.; Merkys, A.; Vaitkus, A.; Okulič-Kazarinas, M.; IUCr. Computing Stoichiometric Molecular Composition from Crystal Structures. *J. Appl. Crystallogr.* **2015**, *48* (1), 85–91.
(331) Gražulis, S.; Chateigner, D.; Downs, R. T.; Yokochi, A. F. T.; Quirós, M.; Lutterotti, L.; Manakova, E.; Butkus, J.; Moeck, P.; Le Bail, A. Crystallography Open Database – an Open-Access Collection of Crystal Structures. *J. Appl. Crystallogr.* **2009**, *42* (4), 726–729.
(332) Gražulis, S.; Daškevič, A.; Merkys, A.; Chateigner, D.; Lutterotti, L.; Quirós, M.; Serebryanaya, N. R.; Moeck, P.; Downs, R. T.; Le Bail, A. Crystallography Open Database (COD): An Open-Access Collection of Crystal Structures and Platform for World-Wide Collaboration. *Nucleic Acids Res.* **2012**, *40* (D1), D420–D427.
(333) Pietrucci, F.; Laio, A. A Collective Variable for the Efficient Exploration of Protein Beta-Sheet Structures: Application to SH3 and GB1. *Journal of Chemical Theory and Computation*. American Chemical Society September 8, 2009, pp 2197–2201.
(334) Caricato, M. Absorption and Emission Spectra of Solvated Molecules with the EOM–CCSD–PCM Method. *J. Chem. Theory Comput.* **2012**, *8* (11), 4494–4502.

APPENDIX A

SUPPLEMENTARY INFORMATION FOR CHAPTER 2

WATER ORIENTATION AROUND K ON MICA

We assessed the orientation of water with respect to the surface ions for K-terminated surface, as shown in Figure A-1 A and B for *ab initio* and CLAYFF interaction potentials, respectively. Here, both distributions have a maximum at (125,0) that represent water with its oxygen atom pointed towards K and one hydrogen towards bulk water and the other towards the surface as shown schematically in Figure A-1.

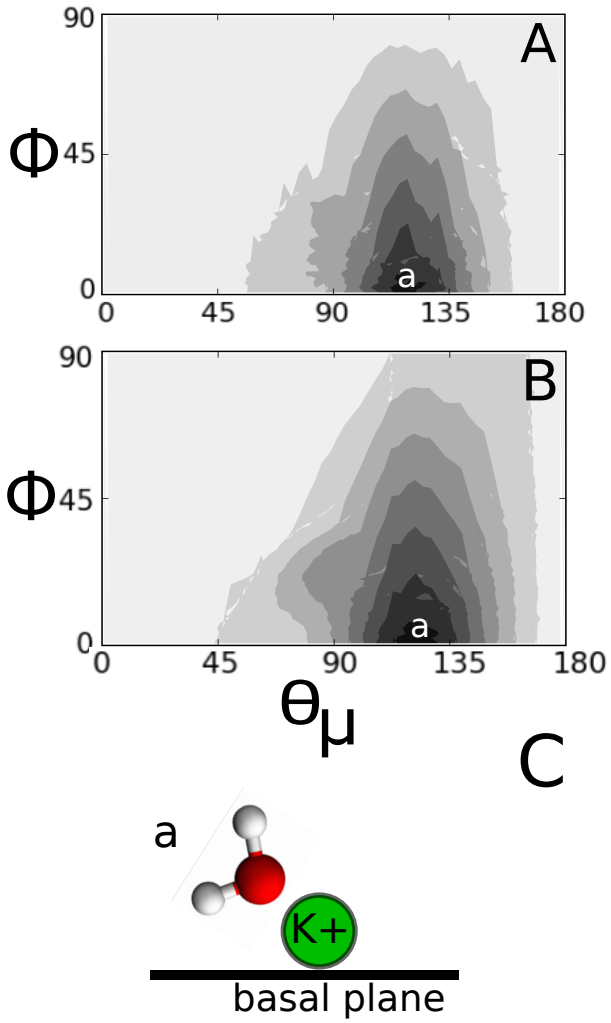


Figure 10-1 Comparison of the orientation of water within 2 Å of K ion on K terminated-muscovite modeled using (A) DFT and (B) CLAYFF classical potential. (C) Schematic representing the orientation for population at (125°,0°). The lower-case label, "a" provides a mapping from the structure to absolute orientation.

ANALYSIS OF CLASSICAL SIMULATION PROTOCOL

To fully test our understanding of how the water structure is affected by the classical simulation protocol, we modeled mica in two different ways using the CLAYFF and SPC potentials for mica and water, respectively. As shown in Figure A-2, the water structure with unconstrained or "free" mica atoms closely resembles the water structure from the simulation where mica atoms are "frozen" or constrained to their original location. Constraining the atoms of mica decreases

the excluded volume effect, thereby allowing the waters to approach the surface more closely. Although, the effect is not pronounced for our system, we simulated classical systems with unconstrained mica atoms.

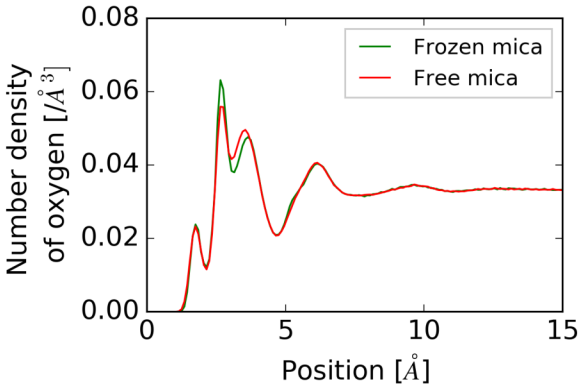


Figure 10-2 Comparison of the structuring of water at the basal surface of mica in simulations where mica atoms are constrained (green) and unconstrained (red)

EFFECT OF ALUMINUM SUBSTITUTIONS

Although the *ab initio* and classical simulations were conducted using a perfectly ordered mica crystal, we calculated the water structure with different aluminum substitutions to characterize the effect of these substitutions. As shown in Figure A-3, these substitutions do not have an effect on the overall water structure. This provides more confidence in our simulation protocol. Interestingly, they do affect the placement of oxygens on the surface and in the bulk.

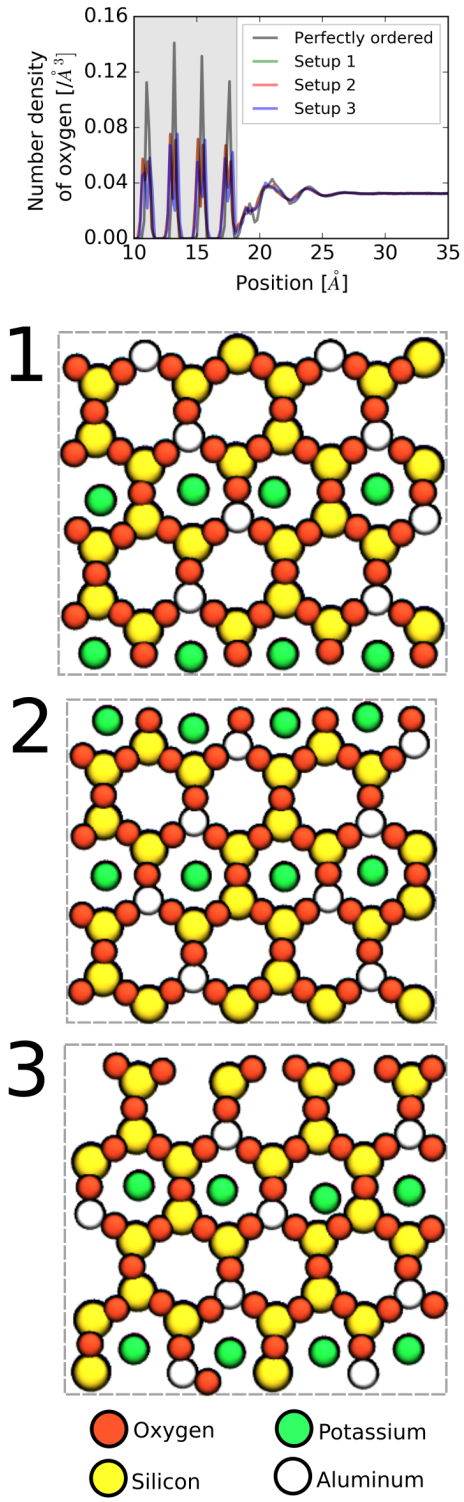


Figure 10-3 (top)Comparison of the water structure on the basal surface of mica with randomized aluminum substitutions in the basal layer of mica (1,2,3)

EFFECT OF SURFACE ION SUBSTITUTIONS

To quantify the effect of ion substitutions on the water structure we simulated the mica surface with one surface K ion substituted with either sodium, calcium, or chlorine. The starting positions of these substitutions were chosen as the minima of their potentials of mean force (calculated as stated below). As shown in Figure A-4, single substitutions do not have a noticeable effect on the structure of water or mica. This result shows the uncertainty in interpreting ions speciation on the by water structuring at the surface.

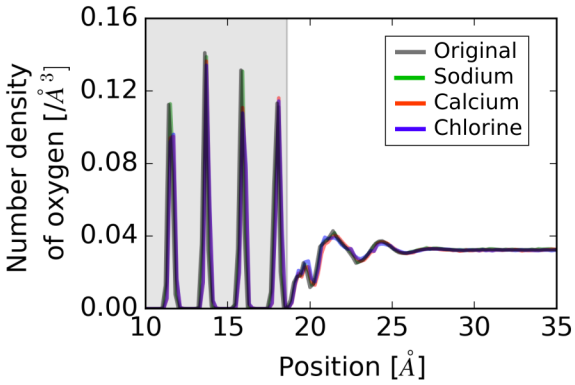


Figure 10-4 Comparison of the water structure on the basal surface of mica with natural muscovite structure (gray), one surface K ion substituted with sodium (green), calcium (red), and chlorine (blue)

POTENTIALS OF MEAN FORCE OF IONS

To further understand how ions behave on the surface of mica, we performed an umbrella sampling to generate the potentials of mean force. Simulation parameters were identical to the NVT simulation described in the Methods section. Umbrella sampling windows were place every 0.5 Å to sample the distance from the surface thoroughly. Each window was sampled for 2 ns of simulation time. After discarding the first 200 ps of simulation, the potential of mean force curve was calculated using the weighted histogram method (WHAM) in GROMACS. As shown in Figure A-5, monovalent ions have a relatively simple potential of mean force. Chlorine

repelled from the surface while potassium ions are attracted to the surface. Calcium shows some rather interesting behavior on the surface. It binds stronger than potassium which is expected from a divalent ion. However, it displays two minima where the barrier is likely associated with shedding it solvation shell to transform from an outer-sphere complex to an inner-sphere complex. These graphs certainly need further study and could prove invaluable in understanding the electrical double layer on the surface.

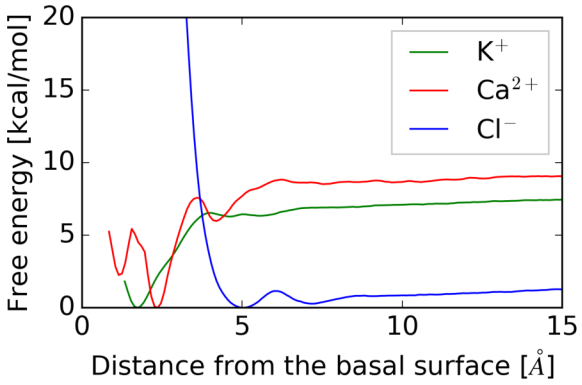


Figure 10-5 Potentials of mean force (kcal/mol) of ions from the basal surface of mica for ions K (green), Ca (red) and Cl (blue).

CALCULATION OF $\widetilde{\phi_m}(\tilde{L})$

In order to calculate $\widetilde{\phi_m}(\tilde{L})$ with the spatial variation of dielectric constant, a simple shooting method[323] was applied. A simple transformation of Eq.(6) (main text) using $\tilde{\phi} = y_1$ ($\tilde{\phi} = \frac{\hat{z}e\phi}{k_B T}$), $\frac{d\tilde{\phi}}{d\tilde{z}} = y_2$, and $\tilde{z} = \kappa z$ (κ is in inverse Debye length) leads to two coupled first order differential equations:

$$\frac{dy_1}{d\tilde{z}} = y_2 \qquad \text{A.10.1}$$

and

$$\frac{dy_2}{d\tilde{z}} = -\frac{1}{\tilde{\varepsilon}(\tilde{z})}\frac{d\tilde{\varepsilon}(\tilde{z})}{d\tilde{z}}y_2 + \frac{1}{\tilde{\varepsilon}(\tilde{z})}\sinh y_1 \qquad \text{A.10.2}$$

Similar to the algorithm by Chan *et al.*[324], the shooting scheme starts at a specified electrostatic potential at midplane ($y_1 = \widetilde{\phi_m}$), with the boundary condition at midplane ($y_2 = 0$), and iterates until a calculated matches with a scaled surface potential (*i.e.*, $y_1 = \widetilde{\phi_s}$ at the surface). Note the calculations take place within a half of separations because of the symmetry of the problem. As suggested in Chun *et al.*[47] [Chun2015], the spatial variation of dielectric constant becomes appreciable within $\widetilde{\lambda_d}$, a scaled characteristic length associated with the structuring owing to the solvent response to the surface, so that the second term in the right hand side of Eq. A.2 is not effective until such separation.

APPENDIX B

SUPPLEMENTARY INFORMATION FOR CHAPTER 3

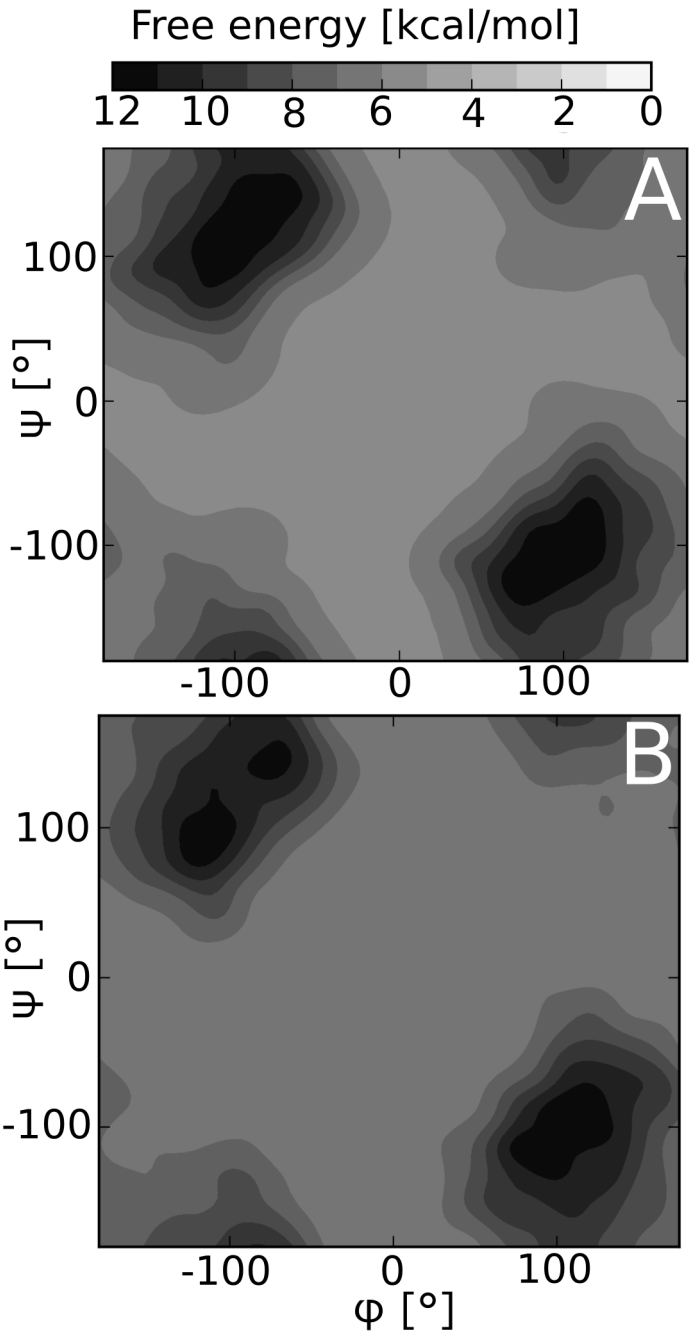


Figure 10-6 Ramachandran free energy surface (kcal/mol) of disarcosine with CH_3 side chains replaced by (A) ethylcarboxyl and (B) ethylphenyl.

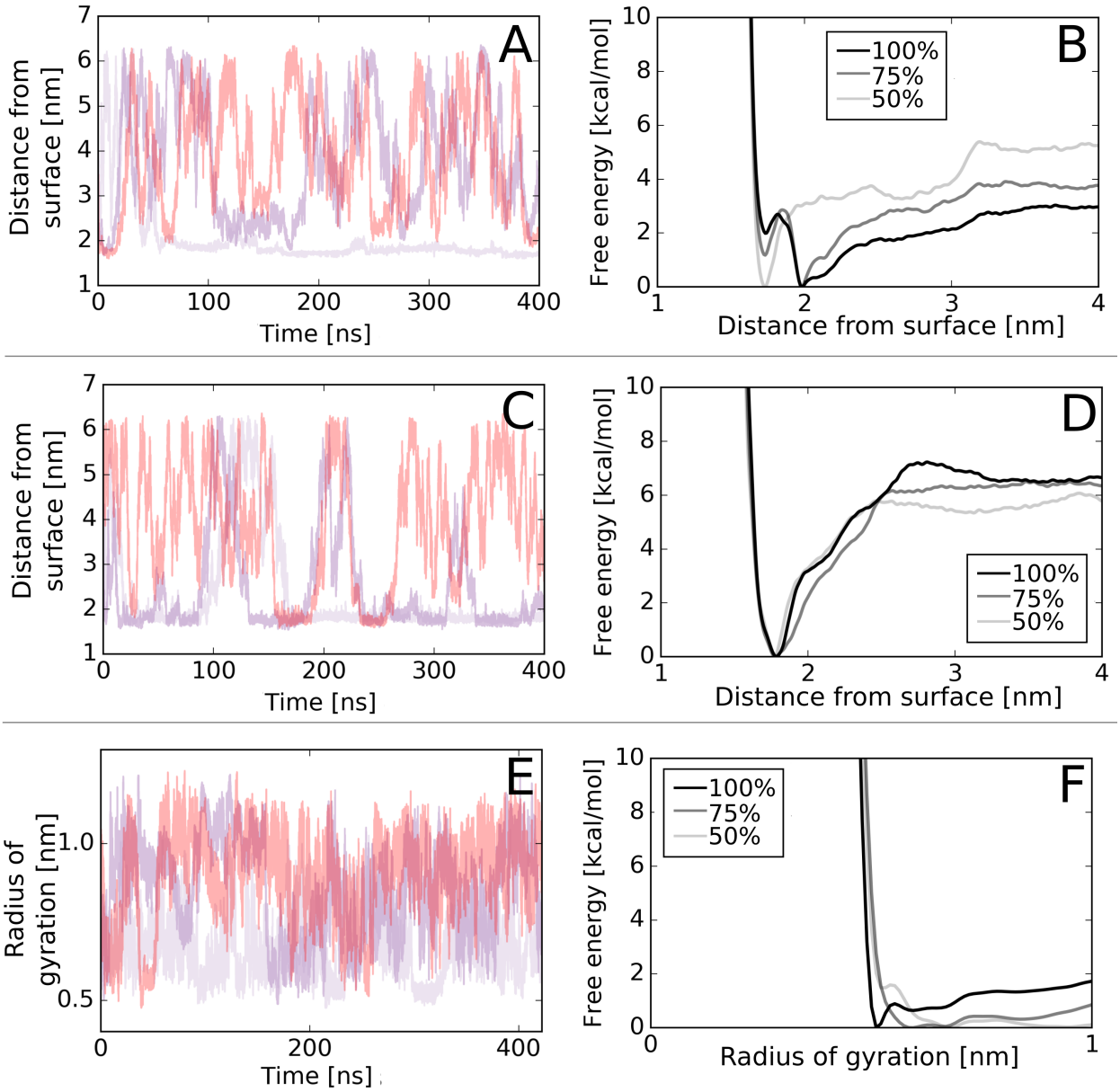


Figure 10-7 Analysis of the convergence of metadynamics simulations of alanine near a COOH-SAM (A and B), near a CH3-SAM (C and D) and in water (E and F). Plots A, C and E shows phase explored by the collective variable (3 walkers) during the simulation. Plots B, D and F show the development of the free energy surface at 50% (red), 75% (blue) and 100% (black) of the simulation time.

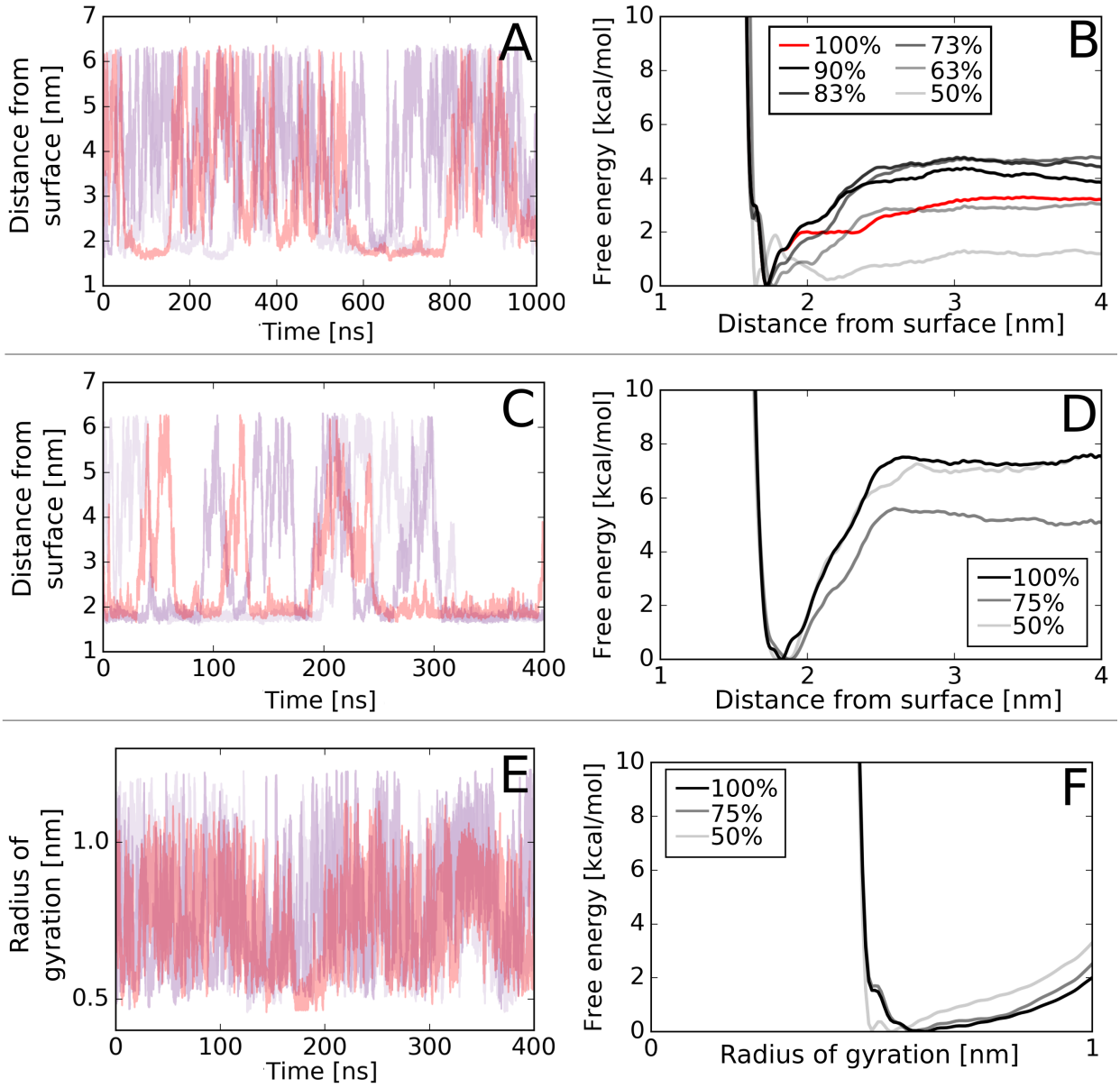


Figure 10-8 Analysis of the convergence of metadynamics simulations of sarcosine near a COOH-SAM (A and B), near a CH3-SAM (C and D) and in water (E and F). Plots A, C and E shows phase explored by the collective variable (3 walkers) during the simulation. Plots B, D and F show the development of the free energy surface at 50% (red), 75% (blue) and 100% (black) of the simulation time.

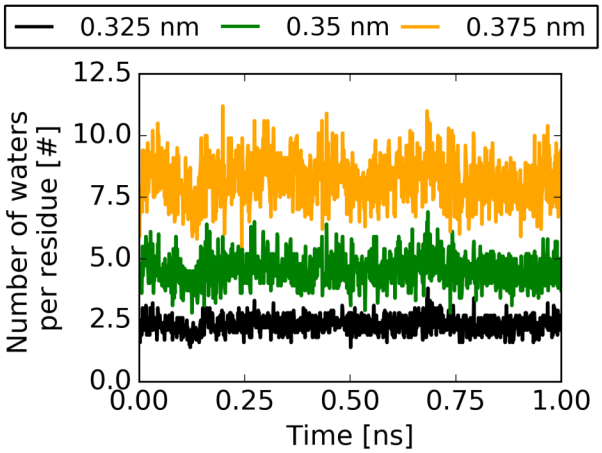


Figure 10-9 Number of water residues in contact with sarcosine (heavy atoms only), during a 1 ns NVT (300 K) simulation, using cut-offs (black) 0.325 nm (average 2.12 waters), (green) 0.35 nm (average 4.14 waters), and (orange) 0.375 nm (average 7.44 waters).

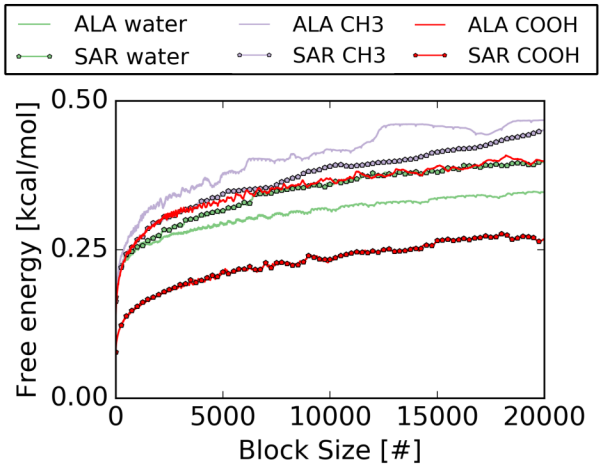


Figure 10-10 Results of a block averaging analysis for sarcosine 11-mers showing the error in the free energy for the radius of gyration in the case of water systems (green), and distance from the surface in the case of (red) hydrophilic and (purple) hydrophobic surface systems from production simulations.

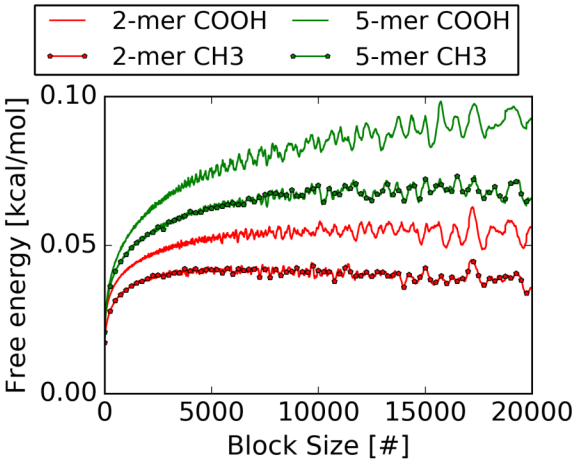


Figure 10-11 Results of a block averaging analysis for sarcosine showing the error in the free energy for distance from the surface in the case of (red) 2-mer and (purple) 5-mer systems from production simulations.

APPENDIX C

SUPPLEMENTARY INFORMATION FOR CHAPTER 4

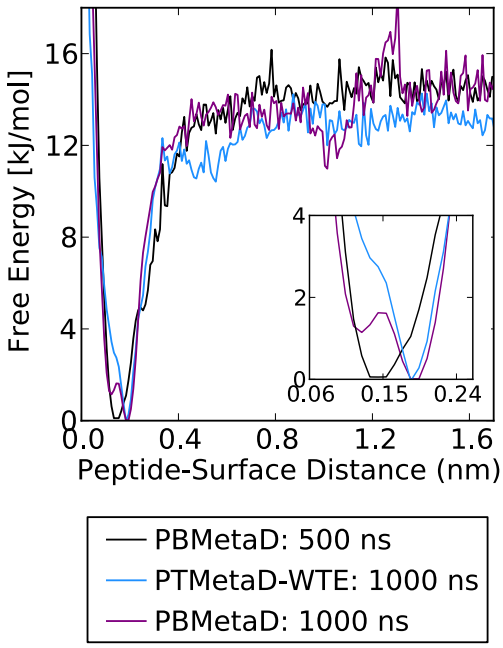


Figure 10-12 Free energy profile of the protein COM from the surface at different stages of the PBMetaD simulation.

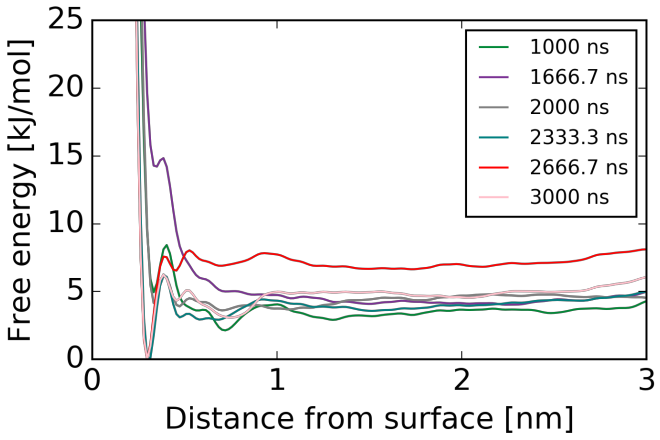


Figure 10-13 Free energy profile of the protein COM from the surface at different stages of the PBMetaD simulation.

APPENDIX D

SUPPLEMENTARY INFORMATION FOR CHAPTER 5

PHOSPHOSERINE PARAMETERIZATION

As noted in the main text, serine (S) residues were replaced by phosphoserine (pS) residues to understand the effect of posttranslational modifications (PTMs) on the binding thermodynamics of R5. We have used 2 different types of phosphoserine residues – pS^{2-} and pS^{1-} – representing the ionization states at pH 7.5 and pH 5, respectively. The bonded parameters were obtained from the CHARMM forcefield. The partial charges used for the pS residues are provided below:

Table 10-1 Charge parameters for pS^{2-} (Total charge -2)

CHARMM Atom Type	Atomic Charge
OR	-0.43034
P	1.23348
OP	-0.98295
OP	-0.98295
OP	-0.98295

Table 10-2 Charge parameters for pS^{1-} (Total charge -1)

CHARMM Atom Type	Atomic Charge
OR	-0.441559
P	1.292900
OP	-0.794264
OP	-0.794264
OX	-0.711228

SILICA SURFACE DESCRIPTION

The initial silica surface at pH 7.5 was obtained from the website (maintained by Interface FF developers): https://bionanostructures.com/interface-md/. The models for the silica surface from release 1.5 were used. From the recommendations in the above-mentioned files, the surface with 9.4 SiO(H,Na) per nm^2 and 0.84 SiO-Na^+ per nm^2 was chosen as the representative surface for pH 7.5 (named: silica_Q2_9_4OH_9pct_ion.pdb). These model surfaces were equilibrated with Na^+ ions considering a solution ionic strength of 0.1-0.3 M. With an ionization of 0.84 SiO-Na^+ per nm^2, the system had 80 Na^+ ions. For pH 5, the ionization was reduced by half, following the trend of other silica surfaces provided by the developers. This resulted in an ionization of 0.42 SiO-Na^+ per nm^2 with a system of 40 Na^+ ions. These surface structures are said to best represent crystalline quartz or silica nanoparticles >200 nm.

INITIAL COORDINATES

Given below are snapshots of the initial coordinates of the systems.

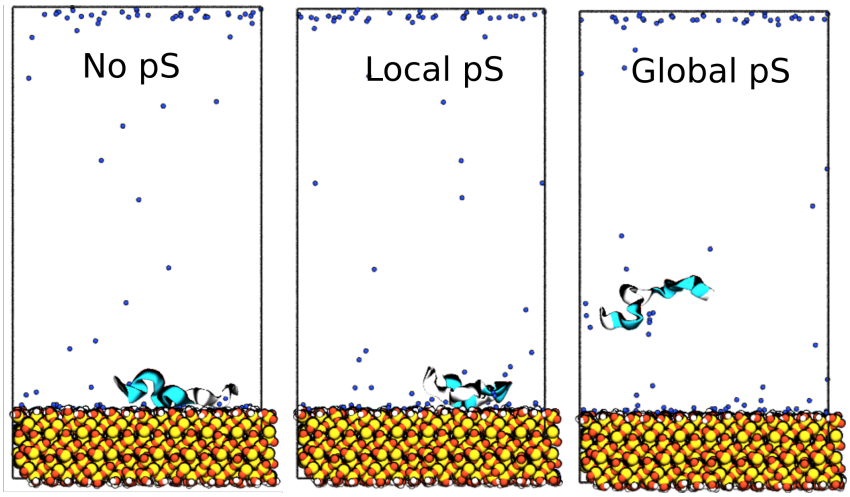


Figure 10-14 Initial configuration for simulations at pH 7.5 for No pS, Local pS and Global pS. Water is not pictured for clarity. The peptide is shown in cyan and white. Red, yellow, and white coloring correspond to oxygen, silicon, and hydrogen atoms, respectively, and sodium ions are shown in blue.

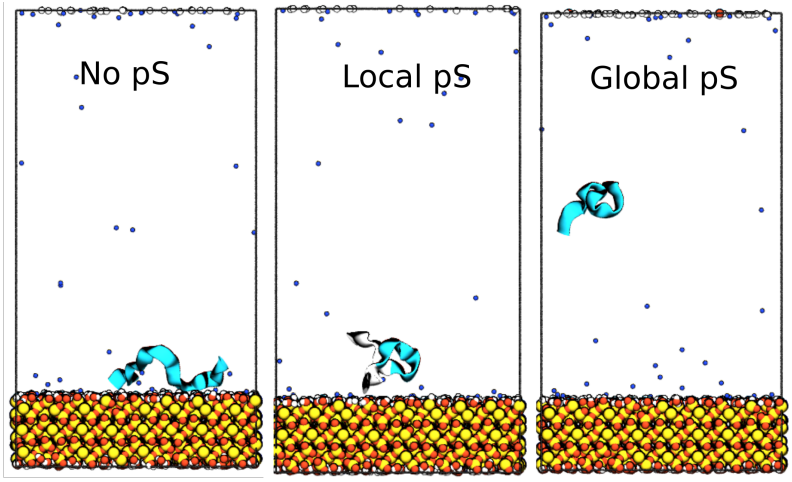


Figure 10-15 Initial configuration for simulations at pH 5 for No pS, Local pS and Global pS. Colors are as described in Figure S1. Some surface atoms appear at the top of the simulation box due to periodic boundary conditions (see Methods, main text).

SIMULATED SYSTEMS PARAMETERS

The exact parameters of the systems are outlined below for clarity and reproducibility.

Table 10-3 Setup of PTMetaD-WTE simulations

System	pH	Free Ions	Water Molecules	Peptide Charge
No pS R5 (pS^{1-})	7.5	80 $^{a}Na^{+}$, 6 $^{b}Cl^{-}$	16,829	+6
Local pS R5 (pS^{1-})	7.5	80 $^{a}Na^{+}$, 4 $^{b}Cl^{-}$	16,835	+4
Global pS R5 (pS^{1-})	7.5	80 $^{a}Na^{+}$, 8 $^{b}Na^{+}$	16,810	-8
No pS R5 (pS^{1-})	5	40 $^{a}Na^{+}$, 6 $^{b}Cl^{-}$	16823	+6
Local pS R5 (pS^{1-})	5	40 $^{a}Na^{+}$, 5 $^{b}Cl^{-}$	16,909	+5
Global pS R5 (pS^{1-})	5	40 $^{a}Na^{+}$, 1 $^{b}Na^{+}$	16,910	-1
Local pS R5 (pS^{2-})	5	40 $^{a}Na^{+}$, 4 $^{b}Cl^{-}$	16,909	+4
Global pS R5 (pS^{2-})	5	40 $^{a}Na^{+}$, 8 $^{b}Na^{+}$	16,910	-8

^{a}INTERFACE FF, ^{b}CHARMM FF

ION-PEPTIDE INTERACTION ANALYSIS

The following analysis was performed to assess whether the use of different "types" of Na^{+} ions in our simulations (i.e., those modeled with the INTERFACE force field versus those modeled with the CHARMM force field), had any impact on our results due to possible differences in ion-peptide interactions. Calculations were performed for the simulation of global pS R5 at pH 7.5, which had the highest numbers of both types of ions. The total number of frames for which an ion of either type was in contact with the peptide – defined as being within 4 Å of any peptide atom – was divided by the total number of frames to calculate a "percentage occupancy" for each type of ion in the simulation. The results show that the peptide was occupied by at least one ion modeled by the INTERFACE force field for ~77.0% of the 125-ns simulation, and by at least one ion modeled by the CHARMM force field for ~70.0% of the simulation. We note that while these results have not been explicitly reweighted, it is expected that the reweighted ensemble averages would remain close to the unweighted ensemble averages, given the closeness in the calculated

occupancy values for the two ion types and the fact that metadynamics weights are the same for each frame of the simulation for all quantities of interest.

CONVERGENCE ASSESSMENT AND ERROR ANALYSIS

Given below are plots assessing: 1) the convergence of the PTMetaD-WTE simulations by measuring the change in free energy upon peptide-surface binding as a function of simulation time; and 2) the error in the free energy projected onto either collective variable biased in the PTMetaD-WTE simulations as determined by a block averaging analysis.

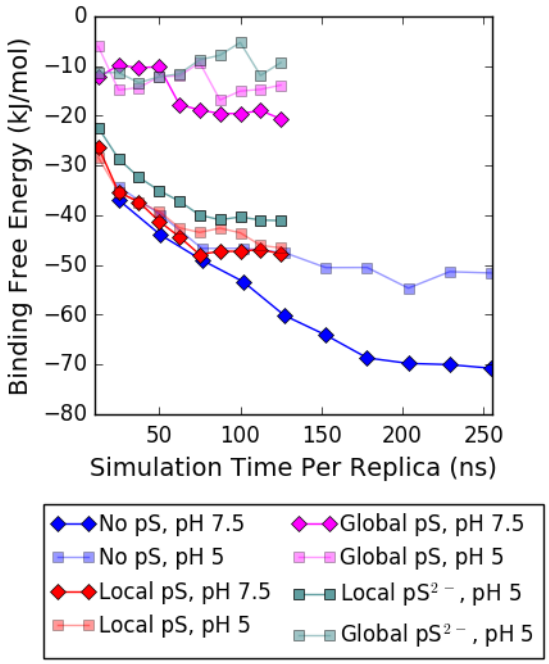


Figure 10-16 Convergence of the change in free energy upon binding for all PTMetaD-WTE simulations, projected onto the distance between the silica surface and the center-of-mass of (blue) R5 without phosphorylation (“No pS”), (red) locally phosphorylated R5 (“Local pS”), (purple) globally phosphorylated R5 (“Global pS”). pH 7.5 results are shown in darker colors

and with diamond markers; pH 5 results are shown in lighter colors and with square markers. Results are also shown for local/global pS R5 with a pS charge state of -2 ("Local pS^{2-}" and "Global pS^{2-}", respectively).

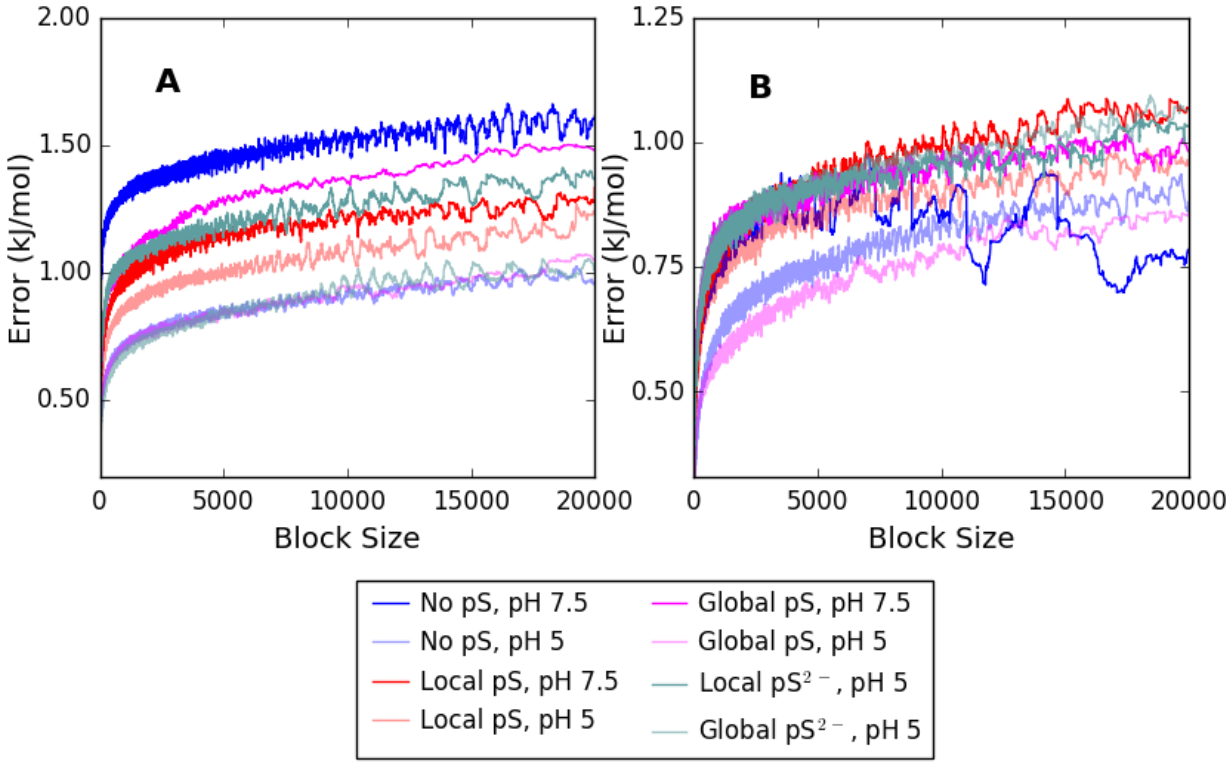


Figure 10-17 Results of a block averaging analysis showing the error in the free energy projected onto (A) the orthogonal distance between the peptide and surface, and (B) the radius of gyration of the peptide, as a function of block size, for all PTMetaD-WTE simulations.

RRIL BINDING

The binding of the tail RRIL sequence was obtained from the biased simulations by calculating the distance from the surface and reweighting it using the metadynamics bias, like the procedure followed for reweighting clusters (see Methods section, main text). The free energy profiles recovered from this method are noisy and may not give a very accurate estimate of the

binding energy. However, the readers can make an estimate of the trend of RRIL binding at different pHs and degrees of phosphorylation.

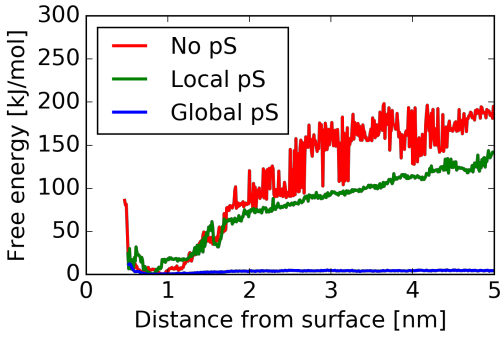


Figure 10-18 Free energy profiles for the distance of the RRIL motif from the surface for no pS (red), local pS (green), and global pS (blue) at pH 7.5.

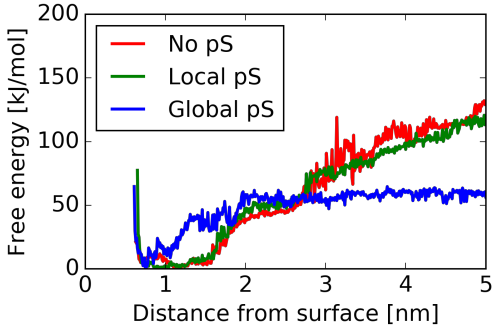


Figure 10-19 Free energy profiles for the distance of the RRIL motif from the surface for no pS (red), local pS (green), and global pS (blue) at pH 5.

Appendix E

SUPPLEMENTARY INFORMATION FOR CHAPTER 6

METHODS

PBMETAD AND PBMETADPF SIMULATIONS

All simulations were performed using GROMACS[49] 2016 simulation engine and PLUMED 2 (Developer's version).[119,325] All systems with Lennard-Jones (LJ) particles were set up in vacuum in a periodic cubic box of side equal to 20 nm, with the requisite number of atoms. There were no charges in the system. Van der Waals cut-off was set to 9.5 nm. The energy of the

system was minimized using the steepest descent algorithm. For production run, the system was simulated in the NVT ensemble (T = 300 K) using the Bussi-Donadio-Parrinello thermostat[51] (temperature coupling time constant, τ = 0.1 ps). A 2 fs timestep was used.

The OPLS-AA force field was used for all simulations. The Lennard-Jones interactions were calculated using the following form:

$$V(r) = 4\varepsilon\left(\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right)$$

For the 3- and 7-particle systems, the σ and ε were set to 0.393 nm and 30 kJ/mol, respectively. For the 13-particle system, the σ and ε were set to 0.393 nm and 11 kJ/mol, respectively. A smaller ε was used for the 13-particle system so that it was possible to converge higher energy values (~100 kJ/mol) within accessible computational power.

PARALLEL TEMPERING SIMULATION (LJ_3 and LJ_{13})

Parallel tempering simulations of the 3-particle LJ system were set up with 16 replicas, and temperatures for the NVT simulation ranging from 300-1000 K. The temperature spacing (300.0, 318.85, 339.54, 362.33, 387.51, 415.42, 446.48, 481.18, 520.12, 564.03, 613.78, 670.48, 735.48, 810.50, 897.73, 1000.00) was calculated using the formula proposed by Prakash et al.[177]

$$\frac{1}{T_i} = \frac{1}{T_{i-1}} - \sqrt{\frac{k}{T_{i-1}}}$$

Where T_i = temperature of the current replica, T_{i-1} = temperature of the previous replica, k = constant optimized for a specific starting temperature, end temperature and number of replicas. The authors derived it for biomolecular systems where the heat capacity changes with temperature. Since the LJ systems also exhibit varying heat capacities with temperatures, this scheme was

chosen in favor of a geometric temperature distribution (which is useful for systems with constant heat capacities).[177]

Exchanges were attempted every 250 simulation steps. Other details of the NVT simulation remained the same as the metadynamics runs. A converged profile was obtained after ~ 2 μs/replica simulation time.

Parallel tempering simulations of the 13-particle LJ system were set up with 32 replicas, and temperatures for the NVT simulation ranging from 300-5000 K. The temperature spacing (300, 302.98, 306.02, 309.10, 312.22, 315.40, 318.62, 321.89, 325.22, 328.59, 332.02, 335.50, 339.04, 342.63, 346.29, 350.0, 383.88, 422.98, 468.43, 521.70, 584.71, 660.0, 751.0, 862.57, 1001.31, 1177.04, 1404.35, 1705.97, 2118.86, 2706.67, 3587.15, 5000.00) was calculated using the method of Prakash et al.[177] Exchanges were attempted every 125 simulation steps. Other details of the NVT simulation remained the same as the metadynamics runs. A converged profile was obtained after ~ 2 μs/replica simulation time.

CLUSTERING

The trajectories for LJ systems (13 particles and 7 particles) were clustered using a 2-step method. First, using the GROMACS tool *gmx cluster* (method gromos[199], cut-off 0.05 nm), cluster numbers were assigned to frames. Since this tool considers all atoms as distinguishable particle when calculating the RMSD of structures, another clustering method was required to cluster the resulting structures. Second, using the Bag-of-Bonds method,[326] where the inter-atomic distances are used as features (78 features for the 13-particle system, and 21 features for the 7-particle system), and the average linkage method of hierarchical clustering, and limiting the clustering to structures of all atoms only, the most similar clusters from the first step were combined into one.

Thus, new cluster numbers were assigned to every frame. The entire simulation trajectory was used.

REWEIGHTING

After assigning cluster numbers, the weight for each cluster was calculated using the Torrie-Valleau[29] method. The weights were calculated using the formula:

$$W = \sum \exp(\beta V)$$

where, V = bias from the PBMetaDPF potential, $\beta = 1/k_BT$ (T = temperature, k_B = Boltzmann's constant), W= weight of each cluster which is the sum of all the weights for the frames where the cluster was identified. For the calculation of V, the simulation was rerun using the GROMACS command *mdrun rerun* and the information from the Gaussian hills deposited during the production run. This provided an estimate of V at each frame of the trajectory. The Gaussian deposition pace was set to 1000000 during rerun to prevent the deposition of new Gaussian hills and changing free-energy estimates.

The weights (Ws) were then normalized to obtain a percentage representation of weights. At this step, only frames corresponding to the regions of phase space that we were interested in (red boxes in Figure 5) were reweighted for. As mentioned earlier, clustering was done using the entire trajectory, however reweighting was done on a subset of frames only.

WELL-TEMPERED METADYNAMICS SIMULATION (LJ_7 SYSTEM)

In order to create an independent reference for the LJ_7 system we carried out a well-tempered metadynamics (WTMetaD) simulation biasing the second and third moments of coordination, as was done by Nava et al.[222] These simulations had an initial Gaussian height of 2

kJ/mol, Gaussian width of 0.01 along each CV, bias deposition pace of 500 steps, and a bias factor of 10. For consistency, we applied the same restraints on interatomic distances placed at 3.5 nm. The coordination number was calculated using the formula

$$s = \frac{1 - \left(\frac{r}{r_0}\right)^n}{1 - \left(\frac{r}{r_0}\right)^m}$$

where, r_0 = 0.6 nm, n = 8, and m = 16, for this system.[223]

Additionally, these simulation results were also reweighted to capture a free-energy profile for the interatomic distance. We followed the reweighting procedure described above, except we accounted for contributions from all 21 interatomic distances to create the free-energy profile.

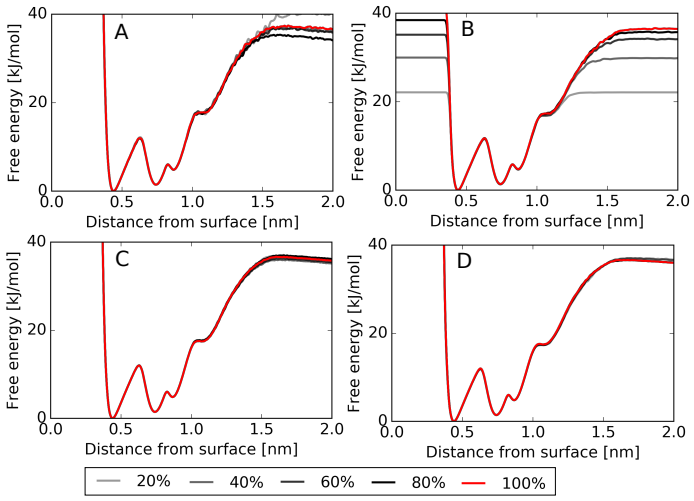


Figure 10-20 For the 13-particle LJ system, evolution of the free-energy profiles for (A & C) PBMetaDPF and (B & D) PBMetaD (averaged over 78 profiles) for the first (A & B) 100 ns and (C & D) total simulation time of 2 μs.

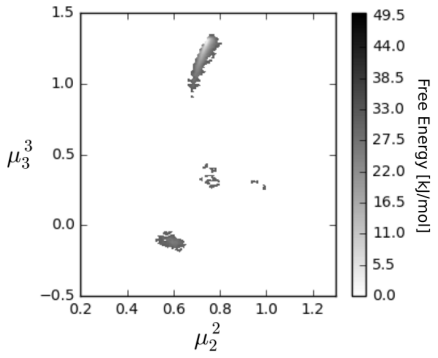


Figure 10-21 (left) Free-energy surface recovered from MD simulations without enhanced sampling of the 2D 7-particle LJ system for ~ 2 μs at 300 K.

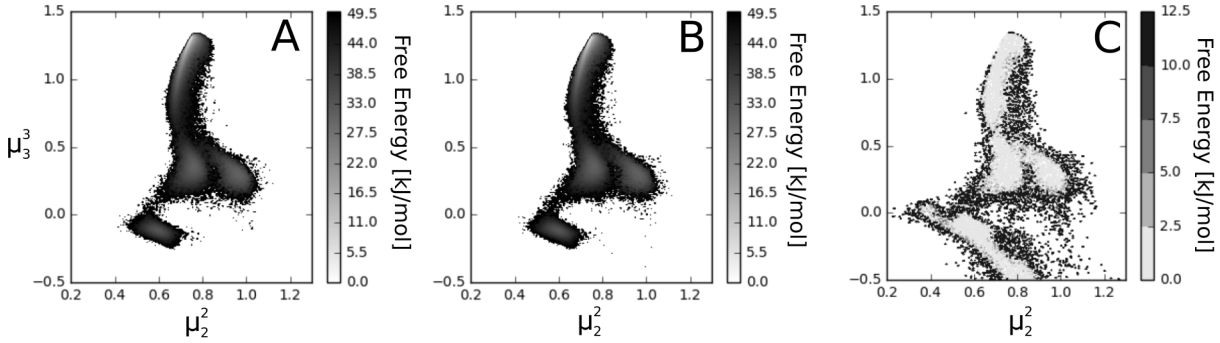


Figure 10-22 Free-energy surface for the 7-particle LJ system reweighted for the second and third moments of coordination numbers using (A) PBMetaDPF and (B) PBMetaD. (C) Difference in free-energy between PBMetaD and PBMetaDPF free-energy surfaces.

Table 10-4 Weights calculated from biased trajectories of PBMetaD and PBMetaDPF simulations.

Weight of top cluster/ Region	**Region 1**	**Region 2**	**Region 3**	**Region 4**
PBMetaDPF trajectory, biased weight	99.93%	92.5%	95.9%	97.9%
PBMetaDPF trajectory, unbiased weight	94.50%	80.89%	86.67%	91.62%
PBMetaD trajectory, biased weight	99.72 %	98.07%	98.50%	98.19%
PBMetaD trajectory, biased weight	94.50 %	97.28%	96.98%	87.83%

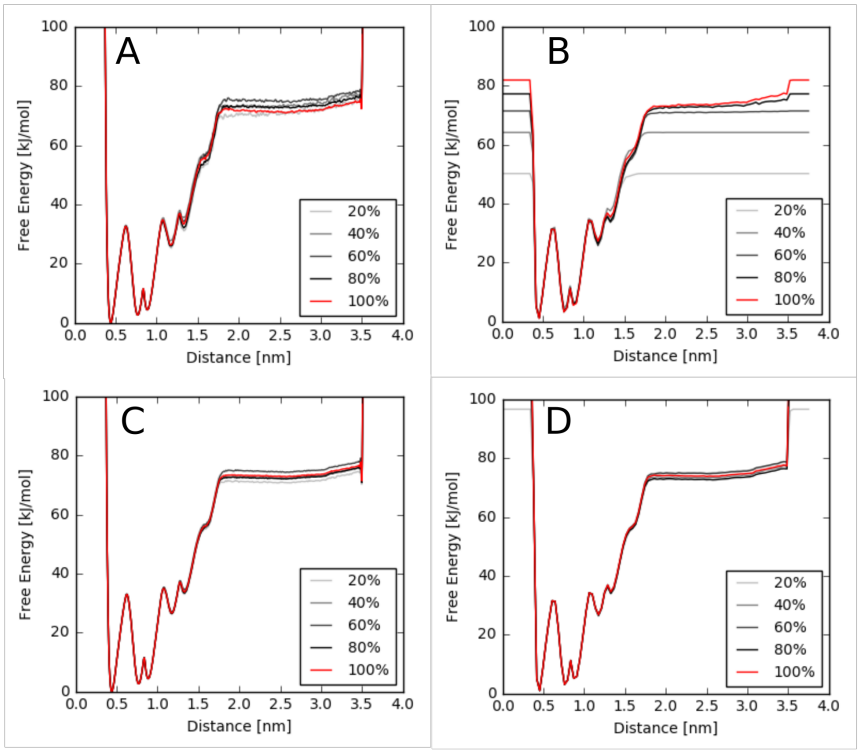


Figure 10-23 For the 7-particle LJ system, evolution of the free-energy profiles for (A & C) PBMetaDPF and (B & D) PBMetaD (averaged over 21 profiles) for the first (A & B) 125 ns and (C & D) total simulation time.

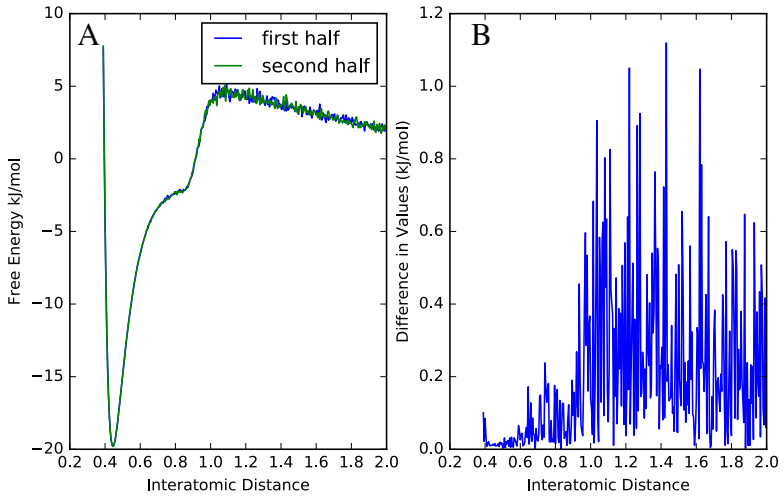


Figure 10-24 3-particle LJ system. (A) Free-energy (kJ/mol) profiles as a function of inter-atomic distance (nm) recovered, aligned by mean values, for the 3-particle LJ system using data

from the first and second half of the simulation (1 μs each) of the parallel tempering simulation. (B) Difference (kJ/mol) between the PBMetaDPF free energy values and parallel tempering free energy values across the range of interatomic distances (nm) sampled (B).

Appendix F

SUPPLEMENTARY INFORMATION FOR CHAPTER 7

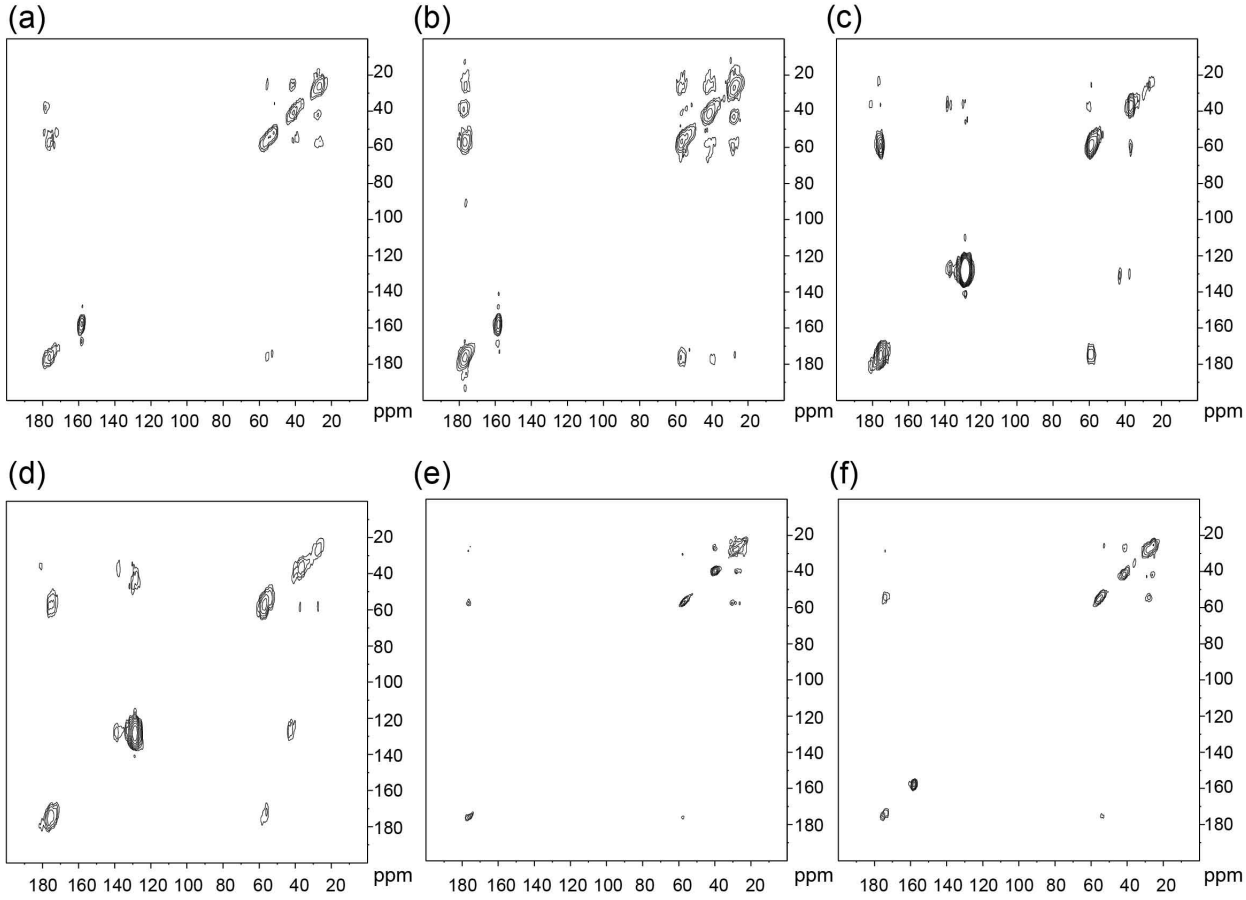


Figure 10-25 ^{13}C-^{13}C DARR spectra of HAP-adsorbed (a) D2R9, (b) D3R10, (c) E4F7, (d) E5F14, (e) K6, and (f) R13.

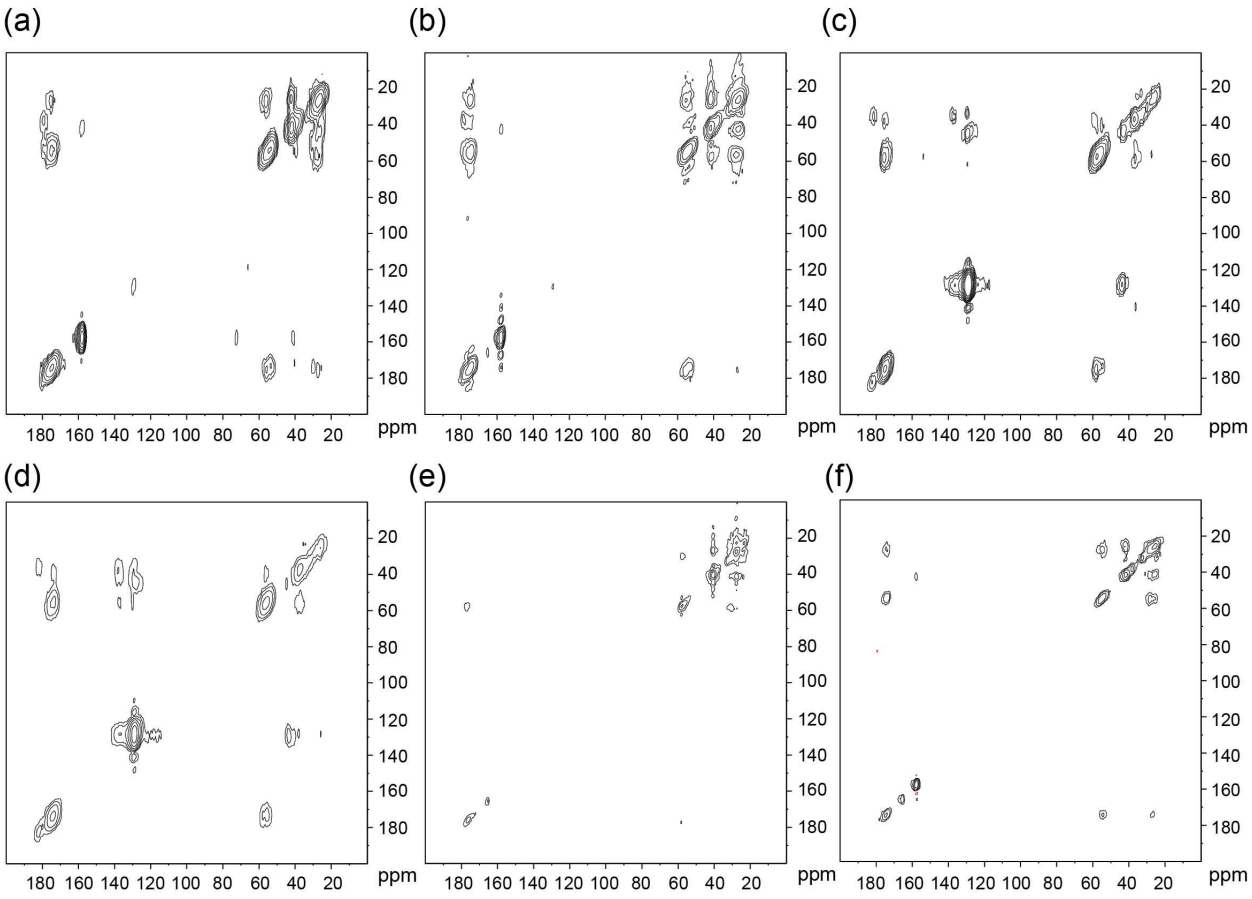


Figure 10-26 ^{13}C-^{13}C DARR spectra of SiO_2-adsorbed (a) D2R9, (b) D3R10, (c) E4F7, (d) E5F14, (e) K6, and (f) R13.

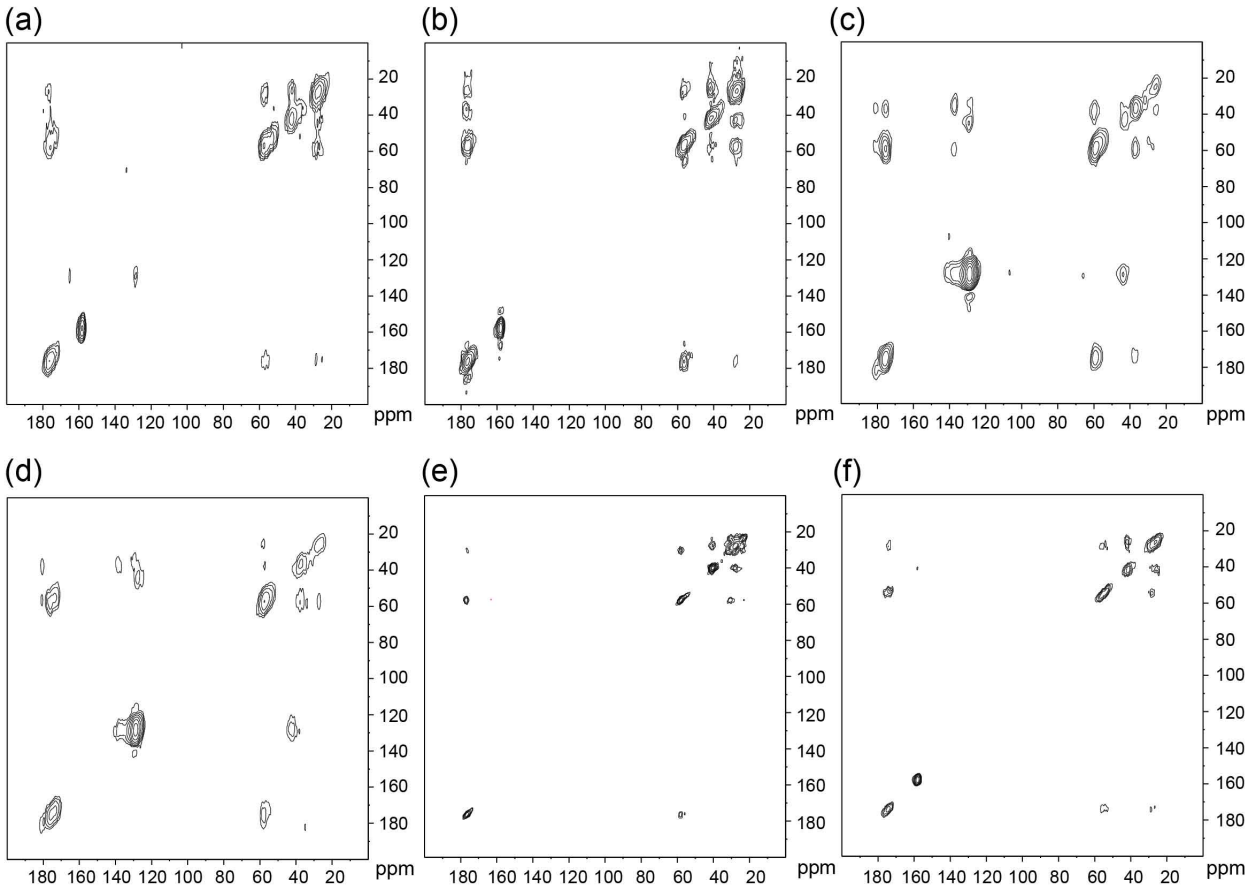


Figure 10-27 ^{13}C-^{13}C DARR spectra of TiO_2-adsorbed (a) D2R9, (b) D3R10, (c) E4F7, (d) E5F14, (e) K6, and (f) R13.

Table 10-5 ^{13}C chemical shift assignments for backbone residues in the neat SNa15 peptide.

Residue	CO (ppm)	Cα (ppm)	Cβ (ppm)
D2	174.1	51.1	36.1
D3	175.1	51	35.9
E4	176.4	58.4	26.8
E5	176.3	57.2	25.9
K6	175.8	58.4	31.3
F7	175.3	60.4	37.4
R9	176.4	57.9	28.1
R10	176.5	58.9	28.4
G12	168.4	39.3	n/a
R13	174.8	56.4	28.6
F14	174	55.8	37.3

Table 10-6 ^{13}C chemical shift assignments for backbone residues in the HAP-adsorbed SNa15 peptide.

Residue	CO (ppm)	Cα (ppm)	Cβ (ppm)
D2	176.4	52.1	40.6
D3	177.6	54	39.1
E4	175.6	59.6	25.3
E5	176.2	59.2	27.8
K6	176	57.6	28.6
F7	175.3	59.6	37.4
R9	174.6	54.2	27.5
R10	177.2	57.2	27.7
G12	172.9	43.2	n/a
R13	174.6	55.2	27.8
F14	178.6	58.6	37.4

Table 10-7 ^{13}C chemical shift assignments for backbone residues in the SiO_2-adsorbed SNa15 peptide.

Residue	CO (ppm)	Cα (ppm)	Cβ (ppm)
D2	174.9	53.7	38.3
D3	176.4	51.2	37.8
E4	175.4	58	27.1
E5	174.5	55.8	26.1
K6	176.2	57.7	30.9
F7	177.9	60.4	34.6
R9	175.4	56.3	28.4
R10	175.9	55.1	26.9
G12	173	43.1	n/a
R13	174.4	54.7	28.5
F14	173.8	55	37.2

Table 10-8 ^{13}C chemical shift assignments for backbone residues in the TiO_2-adsorbed SNa15 peptide.

Residue	CO (ppm)	Cα (ppm)	Cβ (ppm)
D2	172.8	53.5	39
D3	175.3	56.8	39.9
E4	175	59.5	26.8
E5	175.5	57.6	27.4
K6	175.7	58.2	30.9
F7	175.7	59.7	35.6
R9	176.4	57.9	28.5
R10	176.6	58.4	27.6
G12	173.2	43.4	n/a
R13	174.3	55.4	28.5
F14	174.3	54.2	37.1

Table 10-9 ^{13}C chemical shift assignments for side-chain residues in the neat SNa15 peptide.

Residue	Cγ (ppm)	Cδ (ppm)	Cε (ppm)	Cζ (ppm)
D2	177.2	n/a	n/a	n/a
D3	176.7	n/a	n/a	n/a
E4	30.5	177.3	n/a	n/a
E5	30.9	176.8	n/a	n/a
K6	24.2	28.1	40.7	n/a
F7	136.8	129.7	129.5	129.3
R9	27.3	42.3	n/a	158
R10	26.8	42.4	n/a	157.6
R13	27.5	42.5	n/a	157.8
F14	137.8	129.7	129.3	129.3

Table 10-10 ^{13}C chemical shift assignments for side-chain residues in the HAP-adsorbed SNa15 peptide.

Residue	Cγ (ppm)	Cδ (ppm)	Cε (ppm)	Cζ (ppm)
D2	178.4	n/a	n/a	n/a
D3	178	n/a	n/a	n/a
E4	37.1	180.4	n/a	n/a
E5	37.1	181.1	n/a	n/a
K6	23.9	27.4	39.7	n/a
F7	137.4	129.3	128.5	128.5
R9	26.1	42	n/a	158.3
R10	26.1	43.3	n/a	158.1
R13	25.8	41.6	n/a	158.1
F14	139.4	129.5	129.5	129.5

Table 10-11 ^{13}C chemical shift assignments for side-chain residues in the SiO_2-adsorbed SNa15 peptide.

Residue	Cγ (ppm)	Cδ (ppm)	Cε (ppm)	Cζ (ppm)
D2	179.2	n/a	n/a	n/a
D3	178.7	n/a	n/a	n/a
E4	36.2	181.8	n/a	n/a
E5	38.1	182.1	n/a	n/a
K6	22.7	27.5	40.6	n/a
F7	136.7	129.5	129.3	128.8
R9	27.2	42.7	n/a	158.1
R10	26	41.7	n/a	157.5
R13	25.5	41.6	n/a	157.5
F14	136.9	129.6	128.3	128.1

Table 10-12 ^{13}C chemical shift assignments for side-chain residues in the TiO_2-adsorbed SNa15 peptide.

Residue	Cγ (ppm)	Cδ (ppm)	Cε (ppm)	Cζ (ppm)
D2	177.9	n/a	n/a	n/a
D3	177.1	n/a	n/a	n/a
E4	37.1	181.3	n/a	n/a
E5	37.7	180.6	n/a	n/a
K6	24.7	27.9	40.3	n/a
F7	137.5	129.8	129	129
R9	26.7	42.1	n/a	158.4
R10	26.1	43	n/a	158.2
R13	25.6	41.3	n/a	158.1
F14	138.7	130.1	129.3	128.2

Table 10-13 ^{15}N chemical shift assignments for side-chain residues in the neat SNa15 peptide.

Residue	Nε (ppm)	NH (ppm)	Nζ (ppm)
K6	n/a	n/a	33.7
R9	85.1	72.9	n/a
R10	85.7	73.1	n/a
R13	85.4	72.9	n/a

Table 10-14 ^{15}N chemical shift assignments for side-chain residues in the HAP-adsorbed SNa15 peptide.

Residue	Nε (ppm)	NH (ppm)	Nζ (ppm)
K6	n/a	n/a	34.6
R9	85	75.2	n/a
R10	83.6	73.7	n/a
R13	85.7	73.1	n/a

Table 10-15 ^{15}N chemical shift assignments for side-chain residues in the SiO_2-adsorbed SNa15 peptide.

Residue	Nε (ppm)	NH (ppm)	Nζ (ppm)
K6	n/a	n/a	33.9
R9	85.9	74.2	n/a
R10	85.3	72.6	n/a
R13	85.4	72.2	n/a

Table 10-16 ^{15}N chemical shift assignments for side-chain residues in the TiO_2-adsorbed SNa15 peptide.

Residue	Nε (ppm)	NH (ppm)	Nζ (ppm)
K6	n/a	n/a	33.9
R9	85.3	76.4	n/a
R10	84.7	74.5	n/a
R13	85	73.6	n/a

Table 10-17 TALOS-N[38] –generated torsion angles for the neat SNa15 peptide obtained from experimental chemical shifts.

Residue	**ϕ (deg.)**	**Ψ (deg.)**
D3	-99±14.9	61.7±31.2
E4	-59.7±4	-41.9±7.1
E5	-62.4±4.9	-39.1±6.7
K6	-63.8±9.3	-40.5±8.6
F7	-63.9±7.1	-40.4±8.7
L8	-66±5.6	-36.8±8.1
R9	-63.1±5.4	-34.6±5.4
R10	-69.3±7.5	-28.1±8.6
I11	-90.2±16.3	-20.8±15.6
G12	88.9±15.8	-168±22.6
R13	-82.8±27.5	137.1±10.3
F14	-84±10.4	123±12.9

Table 10-18 TALOS-N[38] –generated torsion angles for the HAP-adsorbed SNa15 peptide obtained from experimental chemical shifts.

Residue	**ϕ (deg.)**	**Ψ (deg.)**
D3	-53.2±61.2	162.2±33.6
E4	62.8±8.8	27.5±12.4
E5	-57±23	-34±17
K6	-62.4±6	-38.1±9
F7	-69.3±6.7	-33.5±8.7
L8	-82.1±15.2	-12.6±16.6
R9	-77.2±14.7	-7.8±37.9
R10	-90.3±24.2	-19.1±24.1
I11	-75.7±14.1	-26.9±20
G12	92.4±19.2	-164.7±20.9
R13	-78.7±13.4	110.4±21.3
F14	-65.3±7.8	138.6±10.2

Table 10-19 TALOS-N[38] –generated torsion angles for the SiO_2-adsorbed SNa15 peptide obtained from experimental chemical shifts.

Residue	**ϕ (deg.)**	**Ψ (deg.)**
D3	-98.5±9.2	16.2±17.8
E4	57.7±6.7	39.8±8.4
E5	-95.1±19.6	81.1±23.3
K6	-62.4±7	-35.2±8.3
F7	-61.7±4.6	-30.6±7.9
L8	-66.2±5.7	-27.2±7.9
R9	-81.1±12.9	-5.8±11.6
R10	-87.3±15.9	53.6±65.2
I11	-103.1±22.1	139.5±13.2
G12	-114.1±20.4	159.9±19.9
R13	-82.9±13.7	129±13.7
F14	-88.8±8.6	118.6±12.3

Table 10-20 TALOS-N[38] –generated torsion angles for the TiO_2-adsorbed SNa15 peptide obtained from experimental chemical shifts.

Residue	**ϕ (deg.)**	**Ψ (deg.)**
D3	-73.7±17	-22.8±29.2
E4	-69.3±13	-34.3±19.5
E5	-69±13	-25.2±45.1
K6	-60±5.6	-35±6.2
F7	-68.4±7.7	-33.5±11.8
L8	-64.9±7.9	-35.1±9.4
R9	-63.3±6.3	-30.6±9.2
R10	-74.9±11.8	-21.9±16.9
I11	-82.9±19.3	-21.8±19.6
G12	88.2±15.8	-166.7±24.9
R13	-75±12.3	131.8±10.1
F14	-90.7±9.1	115.5±15.4

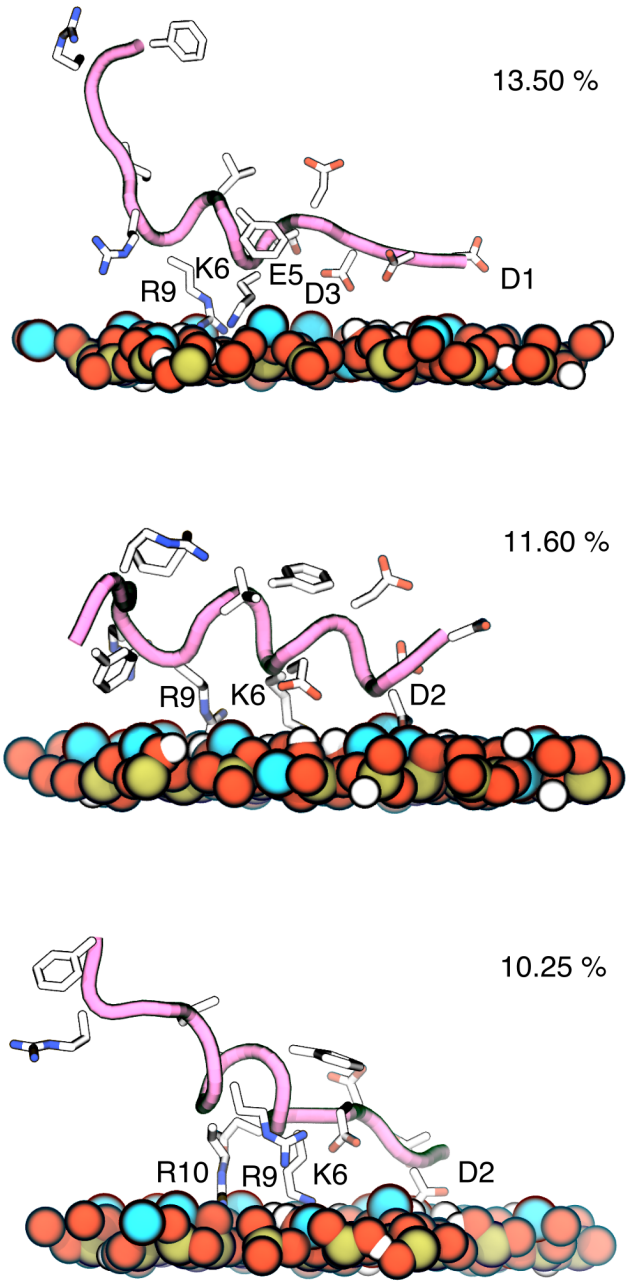


Figure 10-28 Binding poses and residues for SNa15 on the HAP surface predicted from metadynamics metainference simulation trajectories.

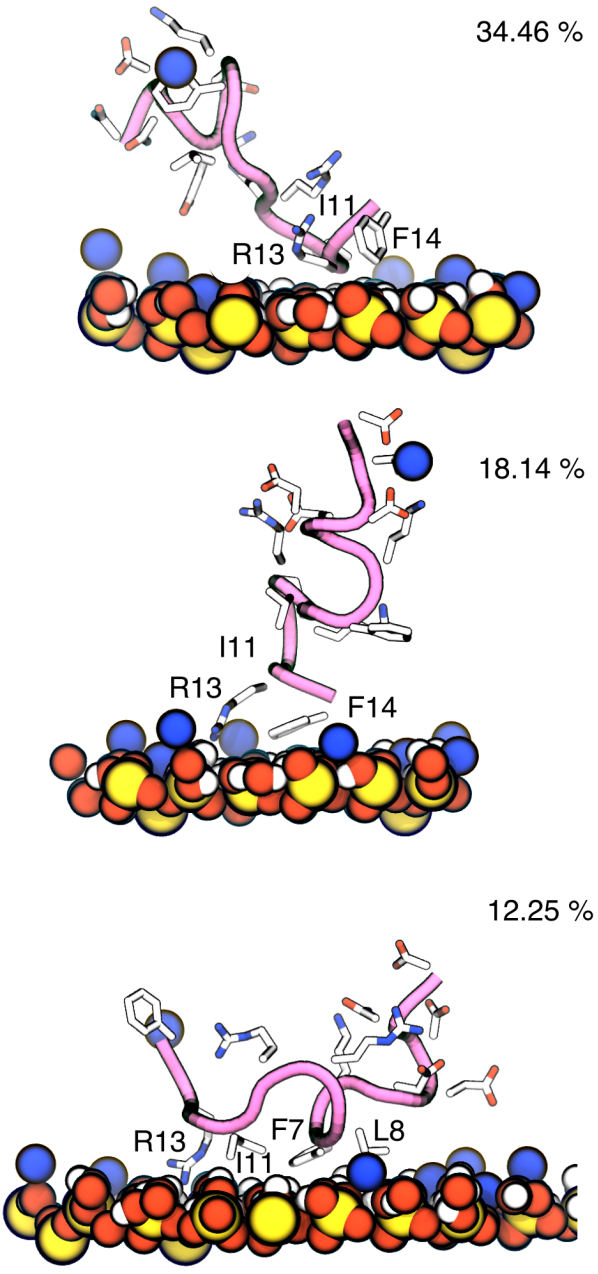


Figure 10-29 Binding poses and residues for Sna15 on the SiO_2 surface predicted from metadynamics metainference simulation trajectories.

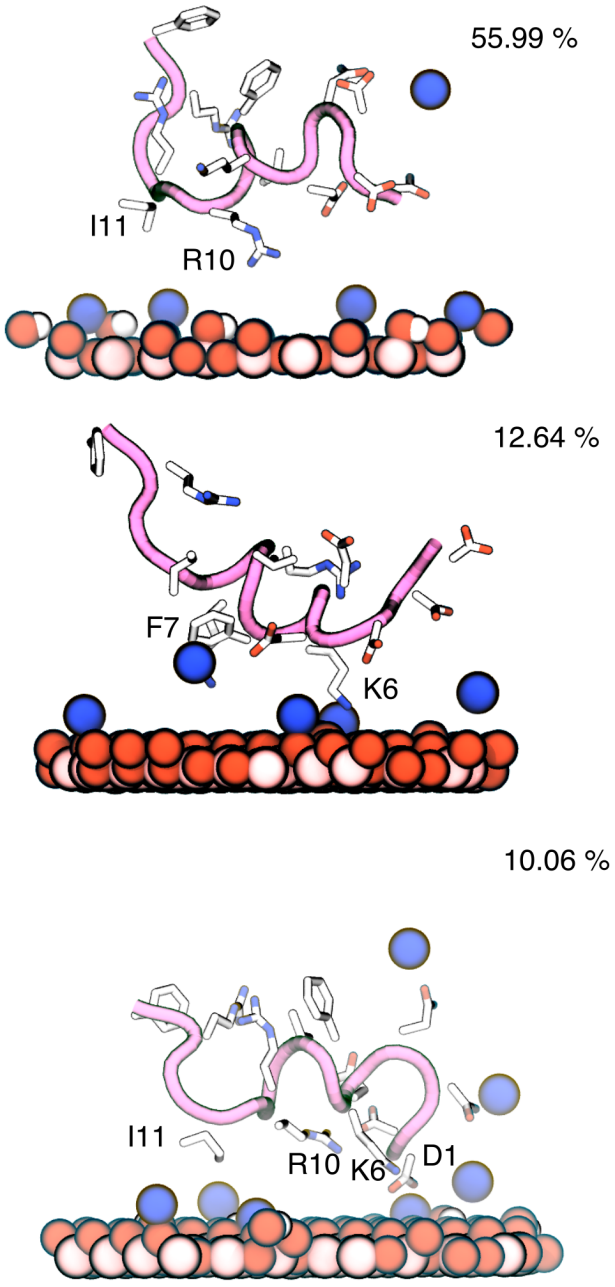


Figure 10-30 Binding poses and residues for SNa15 on the TiO_2 surface predicted from metadynamics metainference simulation trajectories.

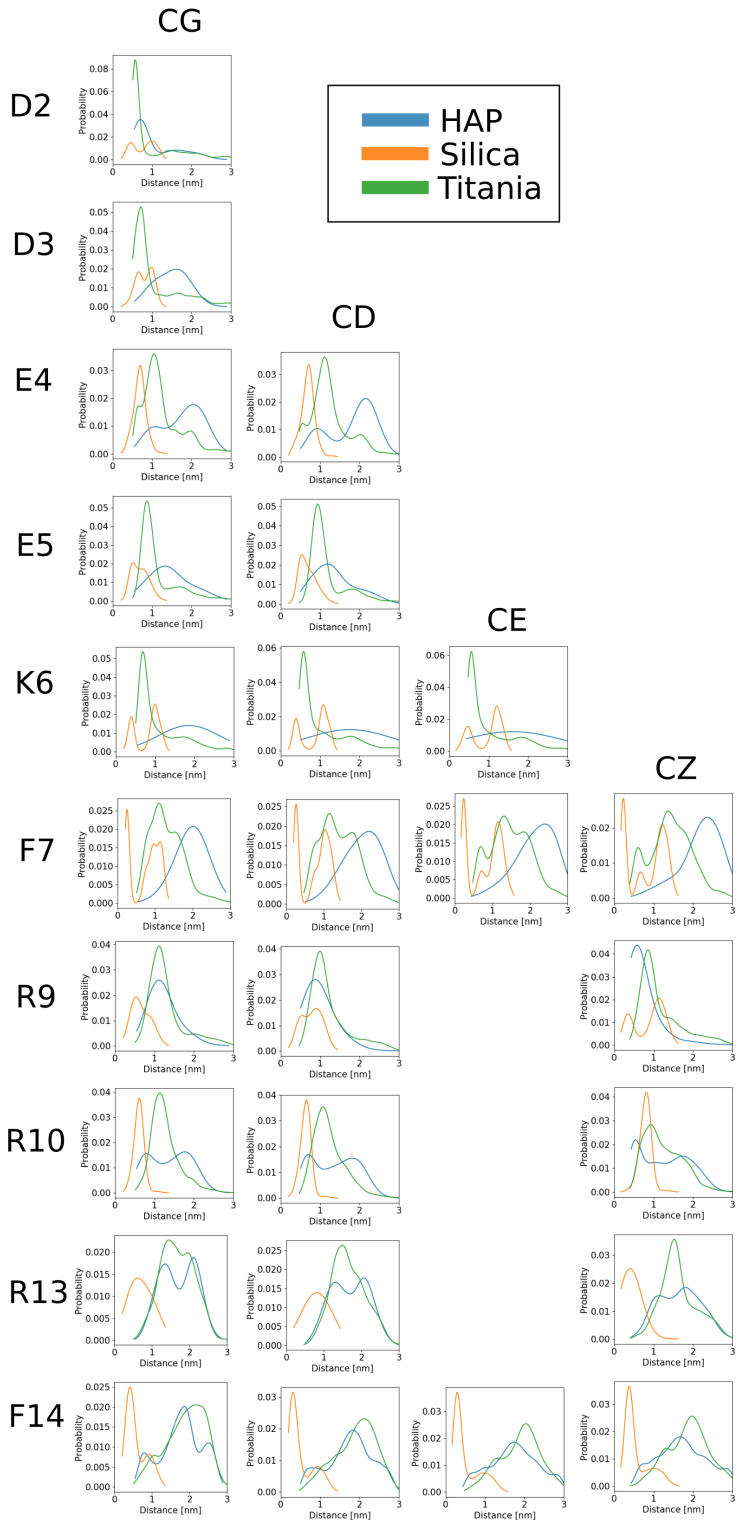


Figure 10-31 Quantitative analysis of binding atoms on surfaces using kernel density estimation for calculating the probability of a side-chain atom from the surface

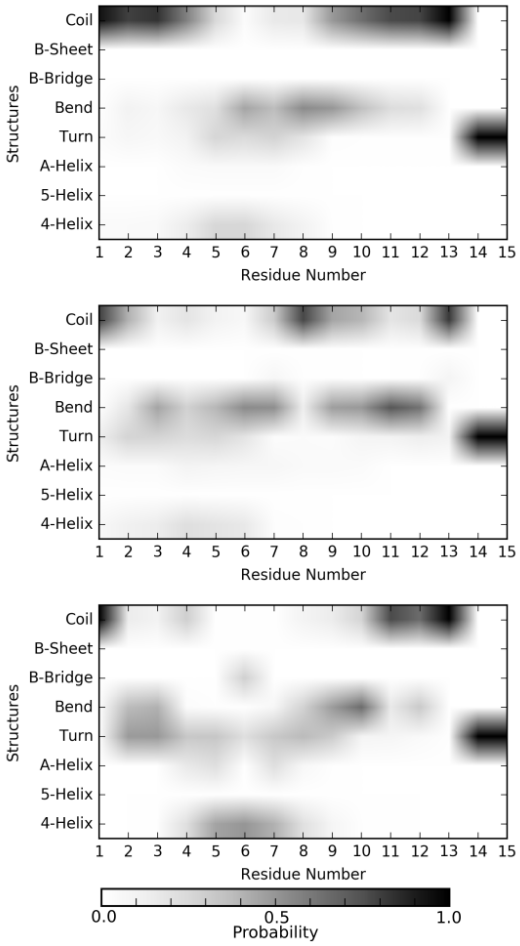


Figure 10-32 Probability of forming different secondary structures as a function of the residue number of SNa15 on surfaces (top) HAP, (middle) SiO_2, and (bottom) TiO_2

CONVERGENCE CRITERIA AND DEMONSTRATION OF SIMULATION CONVERGENCE

For the biased simulations, the free energy profiles of the biased variables were tracked during the simulation. The simulations were considered converged if the free energy surface did not show large (>> kT = 2.5 kJ/mol) changes in free energy in the last 10-15 % of the simulations. Free energy profiles for variables for each simulation are plotted below.

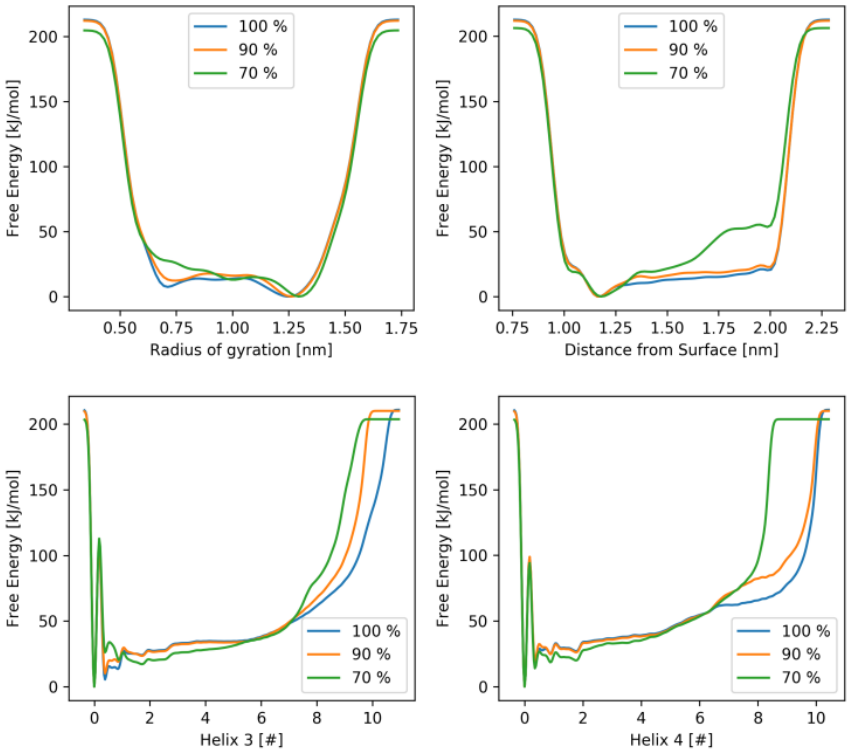


Figure 10-33 Development of free energy profiles in the final stages of the biased simulations for SNA15 on HAP

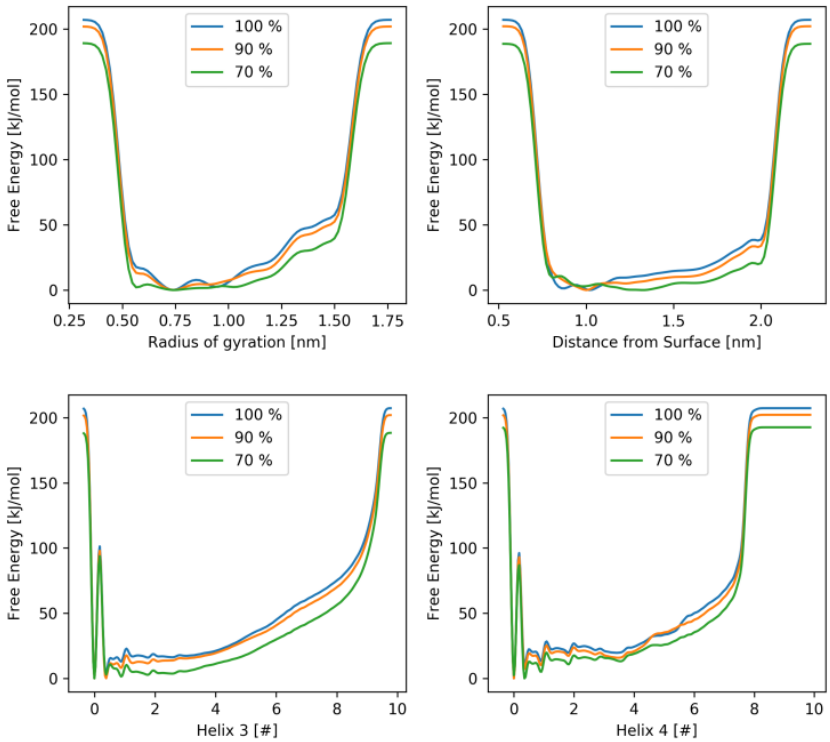


Figure 10-34 Development of free energy profiles in the final stages of the biased simulations for SNA15 on SiO_2

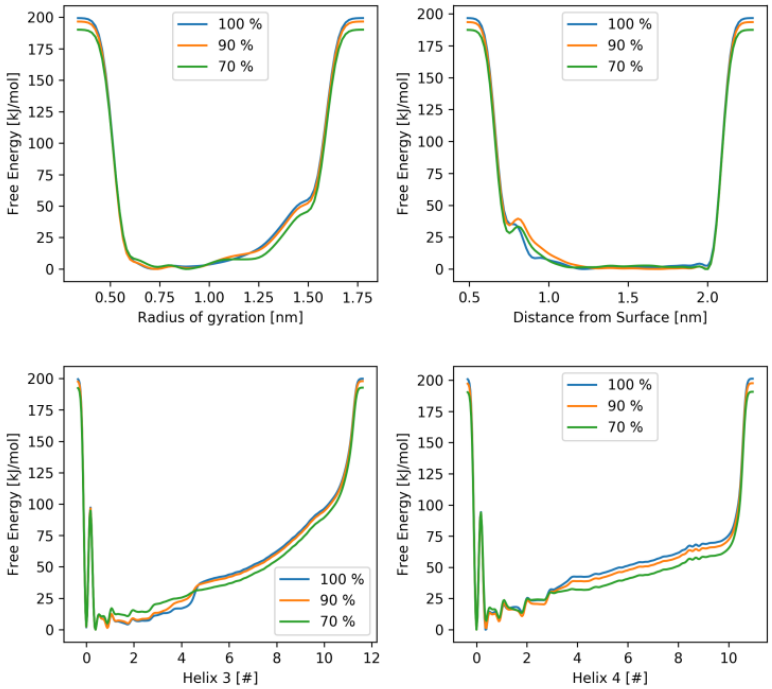


Figure 10-35 Development of free energy profiles in the final stages of the biased simulations for SNA15 on TiO_2

SURFACE MODELS FOR HAP, SIO_2, AND TIO_2

The HAP and SiO_2 surfaces were modelled using the INTERFACE forcefield. They represented pH 7 conditions, as dictated by the rules by Heinz et al.[170] The HAP surface had a surface of ~ 7.5 nm x 6.5 nm, and height of 2 nm. A model of the surface at pH 5 (fully hydroxylated surface) was obtained from the database provided by Emami et. al,[173]which only contained models for pH 5 and pH 10. The surface hydroxylation was decreased by half to better represent the surface at pH 7 (since at pH 10, the surface is dehydroxylated). The charges for the atoms were modified according to the rules mentioned by Emami et. al.[173]

The SiO_2 surface (6.9 nm x 6.8 nm x 2 nm) obtained from a 3.5 x 3.4 x 2.0 nm model of the α-quartz (001) crystal provided by Emami et. al[173]. Surface silicon atoms were hydroxylated with two silanol groups ($=Si(OH)_2$), resulting in 9.4 silanol groups/nm^2. Approximately, 1.0 silanol group/nm^2 was ionized to SiO^- and neutralized with requisite sodium ions. This model is assumed to represent the surface chemistry of quartz at physiological conditions.[34]

Following the protocol used in past studies[327,328], the (100) surface of rutile TiO_2 was simulated with terminal oxygen atoms (i.e., without OH termination) to mimic a titanium surface in aqueous media at a neutral pH. The surface orientation and degree of cleavage were chosen to be consistent with recent simulations of TiO_2 conducted by Brandt et al.[327], with tri- and doubly-coordinated oxygen atoms, and with five- and six-coordinated titanium atoms in the top/bottom and bulk layers of the surface, respectively. TiO_2 unit cell structure was downloaded from the Crystallography Open Database.[329–332] A TiO_2 supercell of dimensions 4 x 6 x 9 nm was built from unit cell belonging to space group P42/mnm, with cell parameters of a = b = 4.594 Å and c = 2.959 Å. The final cell size used in the MD simulations was ~2.7 x 2.8 x 1.7 nm.

MULTI-REPLICA SIMULATIONS OF PEPTIDES

Four, different starting structures were generated for each replica of the simulation. Each peptide-surface system (SNA15-Hap, SNA15-SiO2, SNA15-TiO2) was simulated in an NVT ensemble (600 K) for 3 ns. During this simulation, the center of mass of the peptide was restrained within 1.75 nm of the surface using PLUMED (UPPER_WALLS). Four frames which represented visually different structures were selected from this trajectory, using VMD, and used as starting configurations for each replica.

FILTERING CONFORMATIONS FROM MOLECULAR SIMULATION TRAJECTORIES

We observed during the simulations that the biasing scheme we applied would occasionally force a total dehydration of the surface in totally extended configurations. At this point, the large charge centers present on the surface force the peptide to bind very strongly and completely unfold on the surface and provide a high degree of energetic stabilization. The peptide is then in random coil structure and loses any helical character, at variance with all known experimental measurements of SNa15 and related peptides, which we consider this to unphysical at room temperature and pressure conditions. Accordingly, we systematically filtered out these charge-stabilized unfolded structures according to the following two criteria – degree of alpha-helicity and contacts on the surface. The degree of helicity was calculated using the ALPHARMSD[333] collective variable in PLUMED which denotes the degree of alpha-helical character of the peptide. The z-distances of heavy atoms of the side chains with a reference surface atom were calculated. Distances within 0.75 nm were counted contacts and a total count was calculated for each frame of the trajectory. Thus, frames with less than 0.50 values for the ALPHARMSD CV (completely random coil) were removed along with frames that had > 20 contacts were deleted from the reweighting analysis.

APPENDIX G

SUPPLEMENTARY INFORMATION FOR CHAPTER 8

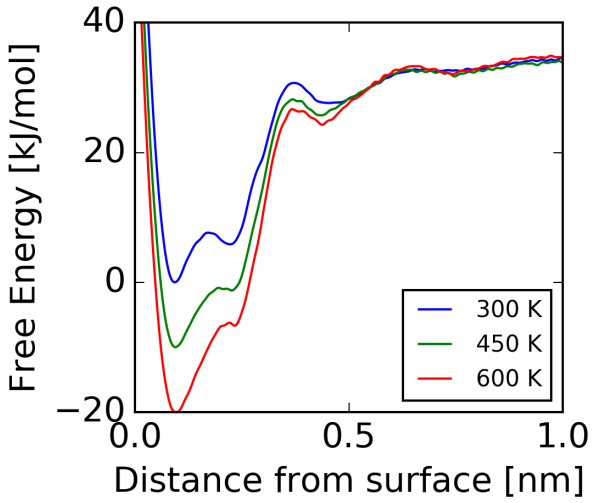


Figure 10-36 Free energy [kJ/mol] of Na^{1+} from the surface of mica as a function of temperature

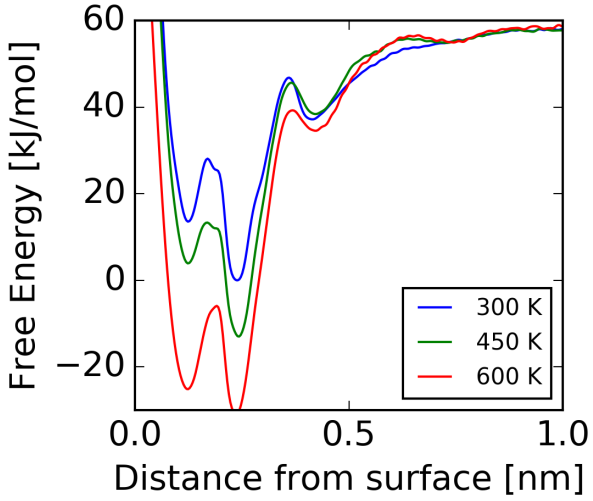


Figure 10-37 Free energy [kJ/mol] of Ca^{2+} from the surface of mica as a function of temperature

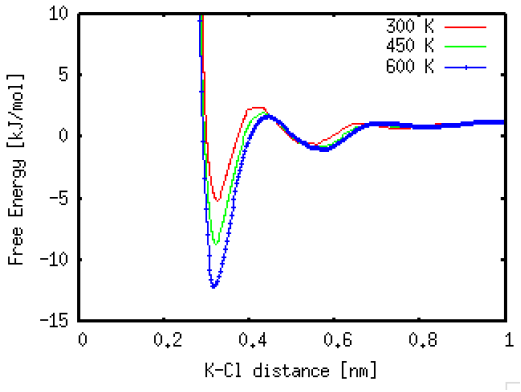


Figure 10-38 Free energy [kJ/mol] of K-Cl binding as a function of temperature

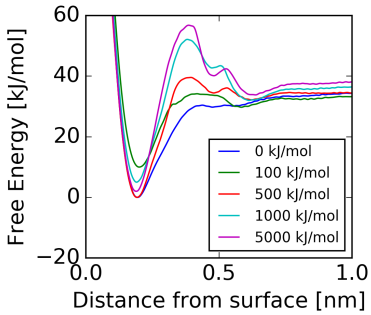


Figure 10-39 Free energy [kJ/mol] of K^{1+} from the surface of mica as a function of the strength of x-y restraints. The shape of the free energy profile changes as the restraints are strengthened.

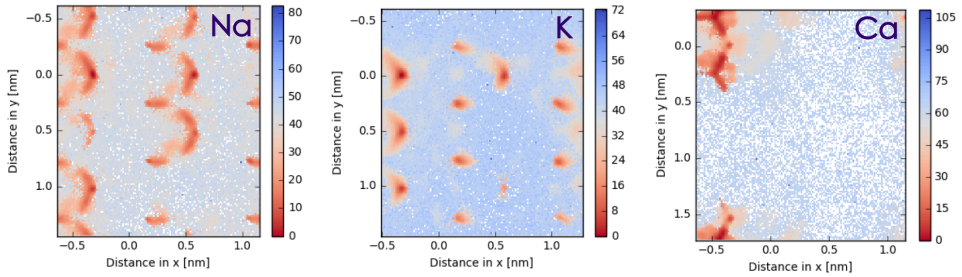


Figure 10-40 Free energy [kJ/mol] of ions on the surface of mica

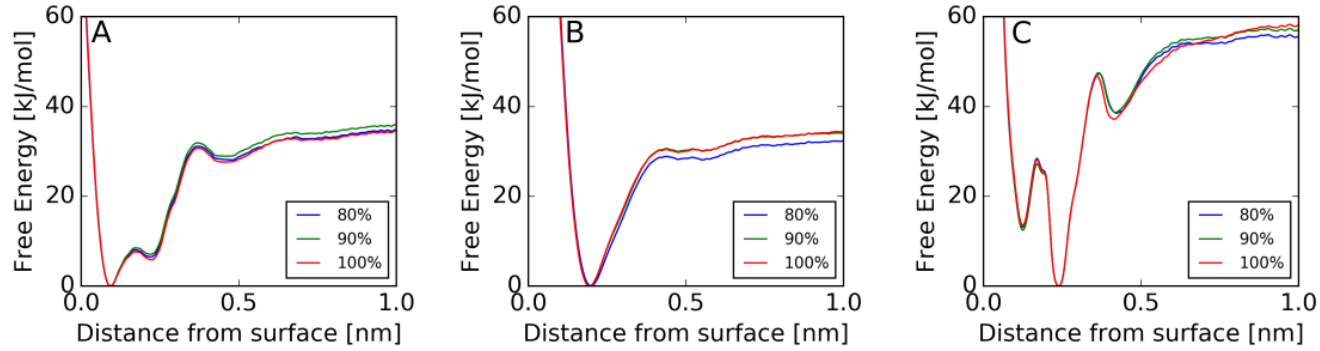


Figure 10-41 Convergence plots. Free energy profiles over the last 30% of the simulation for (A) sodium (B) potassium (C) calcium for the FES presented in Figure 1 (main text).

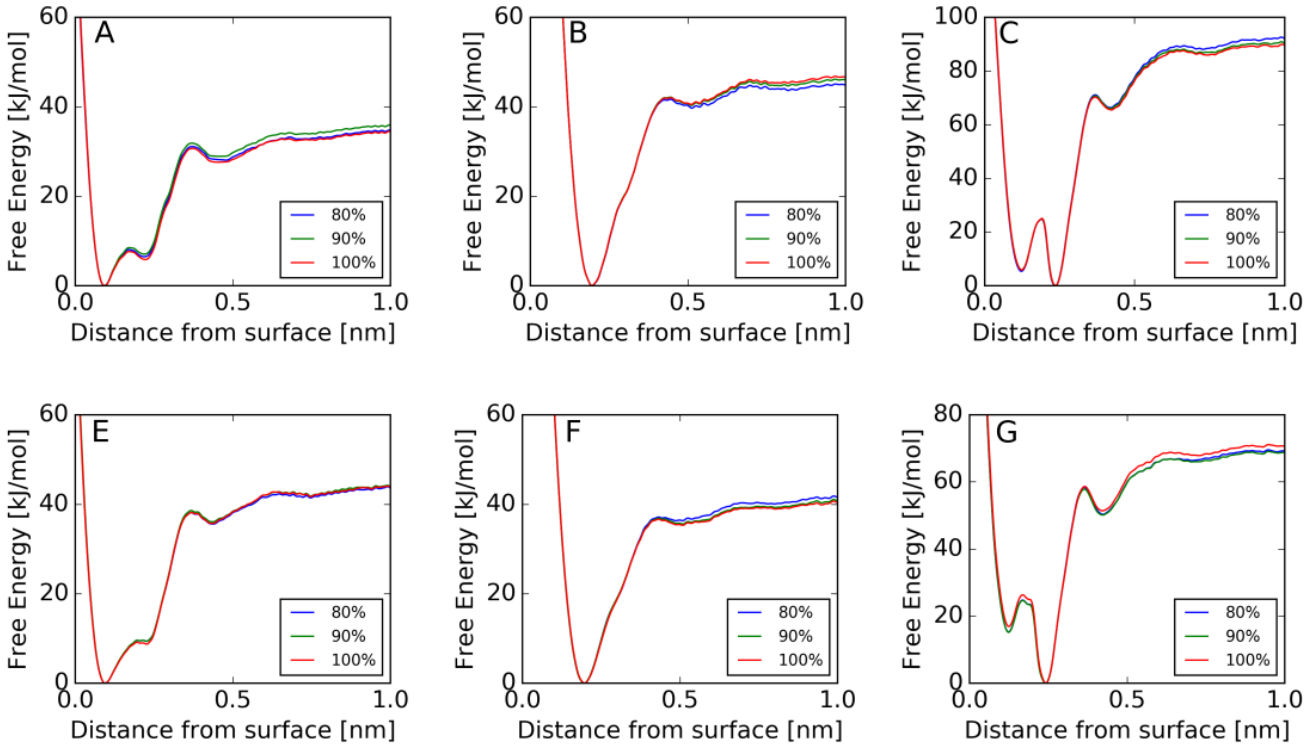


Figure 10-42 Convergence plots. Free energy profiles over the last 30% of the simulation for (A, E) sodium (B, F) potassium (C, G) calcium for temperatures 450K (E, F, G) and 600 K (A, B, C) the FES presented in Figure 3 (left) (main text).

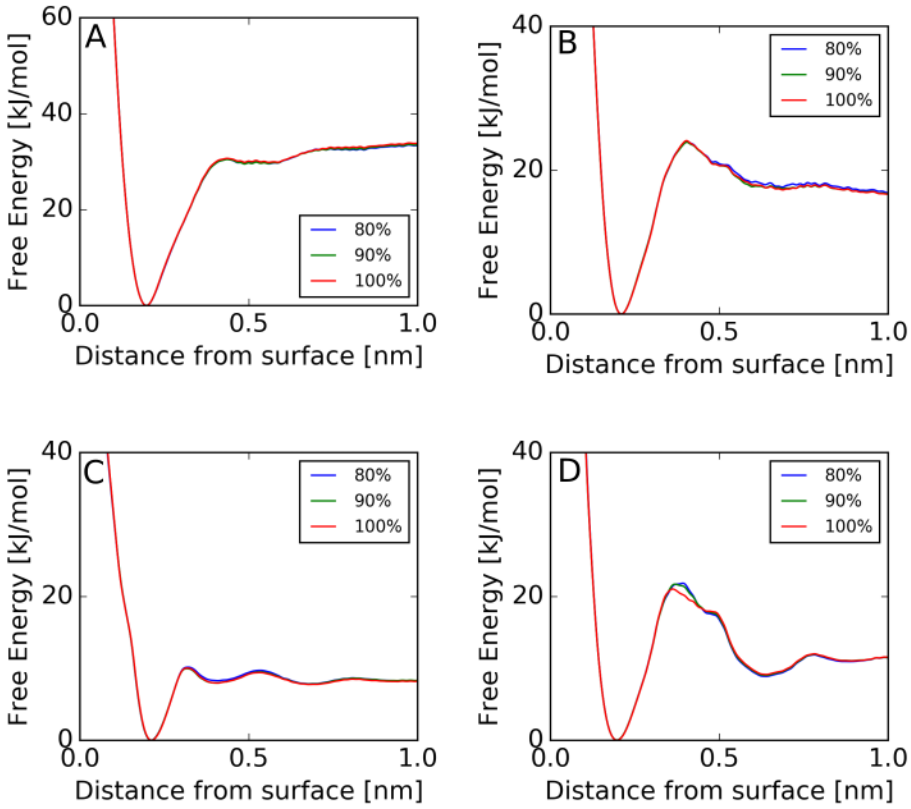


Figure 10-43 Convergence plots. Free energy profiles over the last 30% of the simulation for concentrations (A) 0 m, (B) 1 m, (C) 2 m, and (D) 3 m or FES presented in Figure 4 (left) (main text).

APPENDIX H

SUPPLEMENTARY INFORMATION FOR CHAPTER 9

MULLIKEN CHARGES. To further confirm the charge scaling scheme adapted by Daily et al.[319], we calculated the Mulliken charges for calcium ion in the presence of carboxylate. These charges were calculated after geometry optimization calculations using Hartree–Fock (HF) level of theory and 6-31G(d)//6-31G(d) basis set in Gaussian 09 program[115]. We used two model molecules to mimic the carboxylate side chain – CH_3-CH_2-COO and CH_3-COO. These were minimized in vacuum and implicit solvent (modeled using the Polarizable Continuum Model[334]). The configuration and charges of the COO motif and calcium ion are provided below.

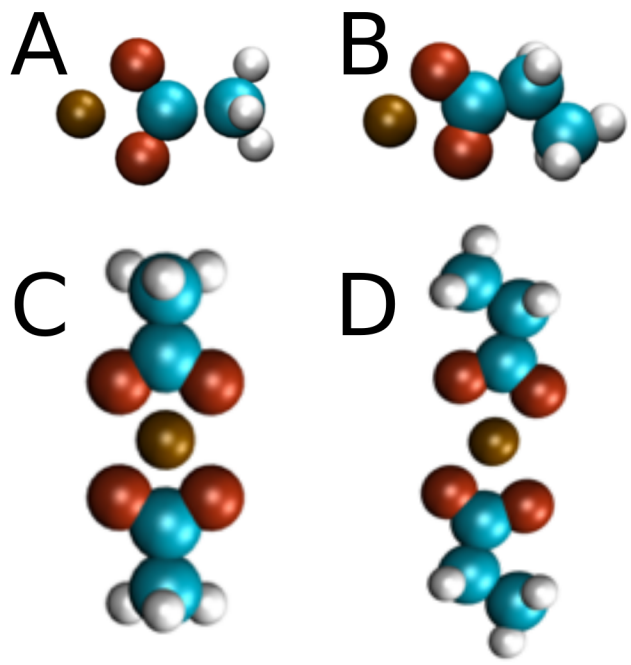


Figure 10-44 Charges were calculated for the above configurations in vacuum and implicit solvent. Atoms are represented in red (oxygen), brown (calcium), cyan (carbon), and white (hydrogen).

Table 10-21 Charges for COO and calcium ion calculated from the configurations above

Motif	Solvent	COO charge	Ca charge
A	None	-0.582193	1.416104
A	Implicit Water	-0.688046	1.639507
B	None	-0.540497	1.410102
B	Implicit Water	-0.64716	1.640207
C	None	-0.612724	1.125308
C	Implicit Water	-0.676391	1.28633

D	None	-0.571573	1.123474
D	Implicit Water	-0.629897	1.282657

SECOND PEPTOID MODEL. To corroborate our results with those of Daily et al.[319], we did the same simulations using their forcefield. Figures SI2, SI3, and SI4 represent the top-weighted structures from simulation in this force field, analogous to those in Figure 1, 2, and 3.

The amphiphilic peptoids (Figure SI2, and SI3) show similar behavior, when compared to our results. These peptoids also form more compact structures on scaling the charge of calcium ions in the simulation. However, the behavior of sarcosine is different in both forcefields. In the force field by Daily et al.[319] sarcosine readily forms loops with both calcium forms. In our force field, sarcosine adopts open conformations with $Ca^{+1.5}$. Since sarcosine does not have bulky side-chains that might affect its structures, these differences in structure are a direct consequence of the backbone dihedral parameters in both these models. Daily et al.[319] modified AMBER03 force field by removing a dihedral constraint, which makes their structures floppier.

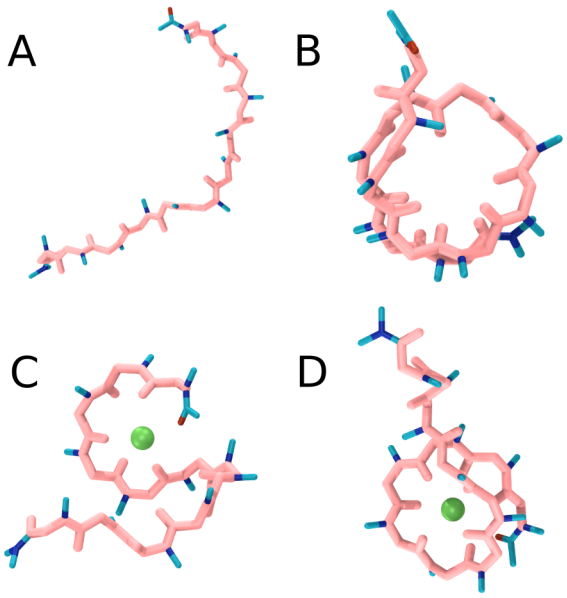


Figure 10-45 Effect of the presence of (A) 6 Ca^{+2} (4.93 %), (B) 6 $Ca^{+1.5}$ (7.15 %), (C) 32 Ca^{+2} (4.90 %), and (D) 32 $Ca^{+1.5}$ (8.35 %) on the structure of sar12. The backbone is colored in pink, while other atoms are represented in cyan (carbon), red (oxygen), blue (nitrogen), and green (calcium).

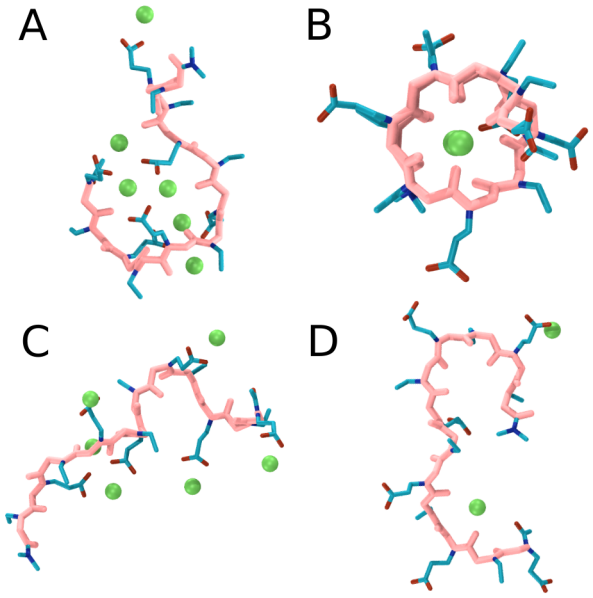


Figure 10-46 Effect of the presence of (A) 14 Ca^{+2} (55.6 %), (B) 14 $Ca^{+1.5}$ (38.1 %), (C) 57 Ca^{+2} (35.4 %), and (D) 57 $Ca^{+1.5}$ (24.3 %) on the structure of $(Nce\text{-}Net)_6$. The backbone is colored in pink, while other atoms are represented in cyan (carbon), red (oxygen), blue (nitrogen), and green (calcium).

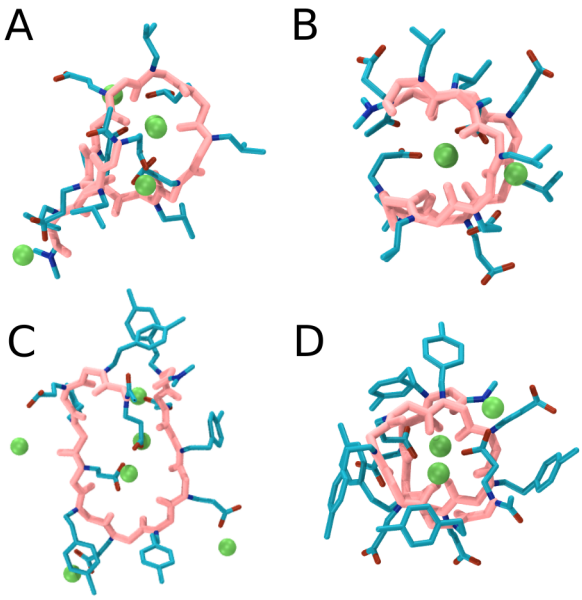


Figure 10-47 Effect of the presence of (A) 60 Ca^{+2} (81.4 %), (B) 60 $Ca^{+1.5}$ (98.9 %) on the structure of $(Nce\text{-}Nib)_6$ and (C) 52 Ca^{+2} (75.6 %), and (D) 52 $Ca^{+1.5}$ (77.9 %) on the structure of $(Nce\text{-}Ncp)_6$. The backbone is colored in pink, while other atoms are represented in cyan (carbon), red (oxygen), blue (nitrogen), and green (calcium).

VITA

Arushi Prakash obtained her undergraduate degree in Chemical Engineering from the Birla Institute of Technology & Science, Pilani, India in 2014. Following this, she came to the University of Washington, Seattle, for her graduate degrees (MS and PhD) in Chemical Engineering. During the course of the graduate program, she has authors seven journal publications (plus two in draft stage) and communicated her research at several conferences worldwide. She has also won numerous awards, including AIChE's Computational Molecular Engineering & Science Forum's Graduate Student Award, SWE Outstanding Female Award in Chemical Engineering, nomination for Student Teaching Award from the College of Engineering, and the AIChE Women's Initiative Committee Travel Award.