Discovery of Biomolecular Structure-Function Mechanisms with Computational Frameworks at the Nanoscale

Sarah Alamdari

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2021

Reading Committee:

Jim Pfaendtner, Chair

David Beck

Brandi Cossairt

Stéphanie Valleau

Program Authorized to Offer Degree:

Chemical Engineering

University of Washington

Abstract

Discovery of Biomolecular Structure-Function Mechanisms with Computational Frameworks at the Nanoscale

Sarah Alamdari

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

Biomolecular function is closely linked to the events that occur at the molecular level, which often takes place in a complex biological environment (i.e., at a complex interface). Approaches which can correlate the massive design space of biophysical and biochemical features at the nanoscale with their expressed macromolecular behavior are of fundamental interest to the field of bioinspired design. While experimental/AI approaches have successfully been applied to characterize the behavior of solution-phase proteins, there is a lack of methods which can probe interfacial phenomena of biomolecules at the same level of resolution. Increases in compute power point to simulation approaches as one avenue for advancing the frontier of biomolecular structure/function exploration at interfaces. In this dissertation, physics-based simulation frameworks were developed and applied to model interfacial peptide, protein, and peptoid systems. As a result, this work demonstrates the capability of computational molecular models to study different biological phenomena with high accuracy, providing insight to the behavior of these complex biomolecules within the areas of biomineralization, self-assembly, and enzyme catalysis.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

The rest of this dissertation will go on to describe my research contributions during this PhD, but it does not pay homage to all the people behind the scenes who made theses chapters possible. It would be impossible to try and thank everyone who has played a part in supporting me through grad school, so I want to first acknowledge and express my deepest gratitude for every friend, committee member, collaborator, co-author, and mentor who has been a part of this journey.

Next, I want to thank all the members of the Pfaendtner Research Group that I have had the absolute pleasure of working alongside the past 5 years. I think it is rare to have the chance to work with a group of people who often feel more like family and friends than coworkers. You all have absolutely set the highest bar when it comes to support, research, fun, and (most importantly) the average number of coffees drank per day. I want to extend a special thank you to; Orion Dollar – for always helping me work through my many (many) research problems and being my co-conspirator for our next side project. We never published anything, but we always had fun. Chris Fu – for teaching me everything about Metadynamics, then teaching me to question everything I know about Metadynamics. I became a better researcher because of it. Luke Gibson – for (maybe unwillingly at times) being my partner in crime at every step of "our" PhD. Having someone to celebrate and commiserate with at every step of the way made the highs and lows of grad school much more special. Jessica Kong – for keeping me well fed (albeit, with vegan food) and in positive spirits straight through the end of my PhD. I know to always count on you for a laugh and a smile. Janani Sampath – from helping me write my first paper to our cheese fries-filled research breaks I am so grateful to have had you as a mentor, friend, and role model to look up too.

Jim Pfaendtner, on a journey riddled with imposter syndrome and self-doubt, thank you for taking a chance on me when I joined your group with zero ability to code, and then continuing to believe in me when I didn't believe in myself. I cannot imagine that I would have finished this PhD without your mentorship. From teaching me how to ask scientific questions to pushing me to put out the best quality of work possible – I am no doubt the researcher I am today because of you. You have been a PI in ways that were not asked or required of you, an incredible mentor, and a great friend. From tear-filled conversations to empowering ones, you have inspired me to lead and have confidence in myself. For that, I am immensely grateful. I aspire to be half the mentor to someone one day that you have been to me.

To the dear friends I made in Seattle, especially – Claire Holcombe, Elena Pandres, Jaime Rodriguez, and Griffin Ruehl – thank you for filling the time between research with so many happy memories and adventures. I want to also thank all my amazing friends outside of grad school, especially – Molly Elkins, Sydney Files, Jack Harrington, Abby Looby, and McKenzie Pickett. Thank you for reminding me to use my vacation time, checking in on me when I was in the deepest of research rabbit holes, and overall being an amazing support system. It should go without saying, but I am extremely grateful to have you all in my life.

To my incredible family, your unconditional support has meant so much to me, and I am so proud to call you my mom, dad, sister, and brother. It is a privilege to obtain a college degree, let alone a PhD, and a privilege that I was afforded because of your support. Thank you for helping me navigate college when I didn't think I could do it, then supporting me on this PhD journey even though you didn't totally understand it. Mom – your ability to accomplish so much on your own and be headstrong in whatever you set your mind too has always inspired me. Dad – thank you for reminding me to stay centered and to not lose sight of the important things in life (one of those being to always have a plate of fruit in front of me). Kayla – you are unapologetically yourself, my best friend, and the person I always turn to and trust for advice. Thank you for being the little sister I look up too. Ryan – you have the biggest heart for being the littlest brother. Thank you for holding down the family fort in Arizona, it has been immensely easier to finish a PhD 1,000 miles away knowing everyone at home is in good hands. My last thank you is for my sweet dog Pizza - you have no idea what has been going on, but you have absolutely been the cutest part of the journey.

DEDICATION

For my family, all my accomplishments are as much yours as they are mine.

Chapter 1
Introduction

Over billions of years, biomolecules have evolved to be highly optimized and carry out specific functional tasks.[1] Taking inspiration from nature, scientists have long been interested in harnessing the power of biological molecules and applying them to solve problems across several fields including energy, biomedicine, drug delivery/design, biomineralization, and catalysis.[2–4] Optimizing and developing existing and novel biomolecular agents for applications other than those intended by nature, often requires the non-trivial modification or redesign of their existing structure. A large amount of effort has been directed at the design of stable and self-assembling biomolecular building blocks (e.g. proteins, peptides, and peptoids) which can mimic and achieve this high function.[5,6] Progress in this area has been driven by highly trained intuition and a bit of luck, as little is known about the fundamental driving forces that regulate these processes at the nanoscale.

Generally, it is hypothesized that the way in which these biological entities are arranged or assembled at the nanoscale, is what enables them to achieve such high function at the macroscale. In pursuit of rational and functional biomaterial design, there is a demand for methods that can link atomistic-level chemical features of biomolecules with their expressed function. Thus, the focus of this work has been aimed at answering the question; how is the sequence and structure of biomolecules related to their function? The breadth of this question stems from the fact that biomolecules are versatile, specific, and efficient building blocks. Consequently, biomolecules are highly diverse both in sequence and structure making up a massive design space which is difficult to navigate. Similarly, often these processes take place in the presence of complex interfaces (e.g., a solid surface interface, small molecules, or solvents). When these further complexities are introduced, there is often a substantial impact on their function, making these relationships between sequence, structure, and function increasingly more difficult to resolve.

Progress in solving these relationships for solution-phase and crystallographic structures has been impressive. Experimental techniques such as X-ray, and electron microscopy have given way to thousands of protein structures, many of which are all freely available on the protein data

bank (PDB). Similarly, advancements in machine learning and AI have accumulated in the development of AlphaFold, essentially resolving the sequence/structure relationships for any sequence of amino acids, otherwise known as the "protein folding problem".[7] However, even the world's most powerful experimental and computational techniques have not reached this level of resolution for biomolecules at an interface. In contrast to the ~ 100,000 resolved protein structures available on the protein data bank, very few structures have been solved for surface bound proteins. Several experimental methods have been applied to characterize the structure of proteins at aqueous or solid interfaces. To name a few; neutron reflectometry, atomic force microscopy (AFM), circular dichroism (CD), vibrational sum frequency generation (SFG), and solid-state nuclear magnetic resonance (NMR) have made strides in this area. Yet, there remains several limitations in both the environment in which they can be probed, the degree of resolution that can be achieved, and the type of information that can be extracted. Further, with the goal of smartly designing functional bioinspired systems, both a dynamic and mechanistic understanding of the interactions that regulate the behavior of these molecules at the nanoscale, and their behavior in the presence of complex interfaces is critical. Progress in AI and sheer compute power point to computational approaches as a promising avenue for advancing the frontier of protein structure/function exploration at interfaces.

Computational methods span large time and length scales, ranging from sub-atomic quantum mechanical descriptions (at short time/length scales) to mesoscopic or continuum (large time/length scale) descriptions of systems behavior. Such approaches can be used to complement experimental data and provide insight in cases where experiments might be difficult or expensive. Somewhere in the middle of this computational spectrum lies classical molecular dynamics (MD), a physics-based simulation approach for modeling the motion (or dynamics) of a system. MD simulations are powerful in that they can be used to sample configurational space, study the dynamic behavior of a system, and (in the limit of ergodic sampling) evaluate the thermodynamic properties of a system. As the power and volume of computational resources available increases, the ability to scale up atomistic models is increasing in a way that has not been feasible in the past. MD simulations can simulate biological systems on the order of nanoseconds to microseconds and with the most powerful computers has even hit milliseconds.[8]

Admittedly, while MD is powerful, it suffers in its ability to survey long timescales. Significant biological events like protein folding, aggregation, or self-assembly can oftentimes

occur on timescales of seconds, hours, even days; this is orders of magnitude larger than even the longest MD simulations. In simulations, these are often referred to as "rare events". There are several computational strategies that can be employed to overcome timescale barriers necessary to observe these rare events. In this work, a class of enhanced sampling simulation methods known as metadynamics (MetaD) is applied to many of the systems studied in this work.[9] In MetaD, a bias in the form of a gaussian potential is applied to a few slow degrees of freedom, discouraging the system from exploring the same region of phase space thereby encouraging sampling of the free energy landscape. These slow degrees of freedom are referred to as collective variables (CVs) and can be defined as any differentiable function describing the atomistic coordinates of a system. With a properly chosen CV, at long simulation times the full energy landscape is explored, and the deposited gaussians sum to the underlying free energy. A CV be anything as simple as a distance, angle or a dihedral, ranging to more complex ones such as coordination number (CN), or ring puckering coordinates. A good CV provides an estimate for the transition between two states of interest separated by some energetic penalty (e.g. distance from peptide to a surface would be a smart candidate for a CV if one was interested in studying protein adsorption onto a surface). CVs provide a useful metric for scaling the 3N-6 degrees of freedom in a system down to a few slow modes that can be monitored and explored. In an enhanced sampling biasing scheme, the CV choice is not trivial as they can be highly degenerate in describing the relevant metastable states. Additionally, this challenge becomes more difficult as the systems and rare events of interest become more complex by the addition of new solvents, solid surfaces, or biomolecules to the simulation system. This is because selection of CVs become less obvious to define in more complex systems. This is often one of the most difficult tasks in setting up a metadynamics simulation.

The development of different classical MD and enhanced sampling frameworks that can be broadly applied to model different biological phenomena is a common theme that unites the work presented throughout this dissertation. As a result, this work also expands on the current understanding of the behavior of some of these complex biomolecular systems. In the next six chapters, this applies to discussing and uncovering the mechanisms which underpin systems related to three main focus areas: (1) biomineralization, (2) self-assembly and (3) enzyme catalysis.

First, in Chapter 2, a variant of metadynamics is applied to study the structure/function relationships in biomineralization through studying the mechanisms of implant fouling by matrix

proteins in the body. Work in the field of tissue engineering suggests that specific surface properties of an exposed implant material are influential in guiding initial cellular events (*e.g.*, protein adsorption to the implant surface) that regulate the environment of the extracellular matrix, and result in osseointegration, or the integration of an implant with its surrounding tissue. Applying metadynamics to simulate a set of model systems, a description of these initial adhesion events is obtained and is used to compare mechanisms and thermodynamics driving forces of protein adsorption onto a mineral (or bone) and titania (or implant) surfaces. Through comparing the simulation results we were able to describe in detail the protein-surface protein-ion and surface-ion interactions that lead to protein denaturation at the implant surface, and by-proxy their possible implication in implant fouling. This level of detail provided a generalized set of material principles that can be used in the future design of implant materials that mimic the bone adhesion response.

Following this, in Chapter 3, a campaign of classical MD simulations was conducted to answer a fundamental question related to the self-assembly of a common model protein, Lysozyme, at an air-water interface. This system is one such example that highlights the difficulty in resolving protein conformation and orientation at the interface. Previously, different experimental methods had been used to contradictorily suggest Lysozyme is denatured or folded at an air-water interface. Further, there was additional discrepancy concerning whether the folded structure was in a head-on, axial, or side-on conformation. These challenges are likely attributed to difficulties in examining monolayer-level length scales at an interface. Using simulations combined with SFG we show with strong agreement that lysozyme adopts an axial conformation at pH 7. Further, we provide molecular-level insight as to how pH influences the binding domains of lysozyme, providing a plausible mechanistic explanation for prior contradictory conclusions in the literature.

Included in Chapter 4 is a review of the state of modeling and experimental approaches that have been used to study the self-assembly of biomolecules at interfaces in solution. We also include a perspective as to how machine learning and combined experimental/simulation approaches can help evolve this field of research. Chapter 5 is a review of parallel bias metadynamics, a simulation approach used to circumvent the high degrees of freedom related to studying complex systems with metadynamics. This chapter highlights different examples where this approach has previously been applied successfully.

In Chapter 6, a variant of metadynamics is used to elucidate the mechanisms that lead to product inhibition in an industrially relevant enzyme. A major limitation in the enzymatic production of renewable biofuels from cellulose is related to their cost. One reason for these can be attributed to cellulase enzymes high susceptibility to product inhibition, leaving them inactive and difficult to recycle. Thus, the rational design of enzymes that retain high substrate affinity and a reduced product affinity are desirable to reduce costs at a commercial scale. In this work, a mutant enzyme with reduced product inhibition and increased catalytic turnover is discovered. Funnel metadynamics is then used to investigate the effect of the amino acid substitutions on the mechanism of cellobiose binding (i.e. product inhibition). Atomistic level resolution of these structural and binding mechanism changes as a function of point mutations provides a framework for rationally designing cellulases with improved commercial activity.

In Chapter 7, a metadynamics approach is proposed to study the structure/function mechanisms that give rise to secondary structure assemblies in a peptidomimetic foldamer known as peptoids. Synthetic oligomers which can mimic the secondary structure of proteins and nucleic acids offer a route to potentially expand beyond the functions accessible to current molecular frameworks. One fundamental challenge in the application of these foldamers is in uncovering the rules that relate sequence and chemistry to folding behavior within these non-biological backbones. In this work we present a metadynamics framework that can successfully traverse the highly dimensional folding landscape of a peptoid backbone. Using three model systems with different sidechain chemistries, we demonstrate how this framework is generalizable to many peptoid chemistries. Through this, we uncover the underlying energetic and entropic driving forces that stabilize protein-like helices for a chiral bulky peptoid chemistry. This method provides a basis for peptoid-structure exploration which can be expanded upon to rationally design sidechains that enable more robust helix formation, or the design of peptoid sidechains that can assemble into other biomimetic secondary structure or higher order building blocks. Lastly, the final chapter then concludes with a discussion of the overall significance and future directions of this work.

Chapter 2
Impact of glutamate carboxylation in the adsorption of the α-1 domain of osteocalcin to hydroxyapatite and titania[1]

2.1 Abstract

One proposed mechanism of implant fouling is attributed to the nonspecific adsorption of non-collagenous bone matrix proteins (NCPs) onto a newly implanted interface. With the goal of capturing the fundamental mechanistic and thermodynamic forces that govern changes in these NCP recognition domains as a function of γ-carboxyglutamic acid (Gla) post-translational modification and surface chemistry, we probe the adsorption process of the most commonly occurring NCP, osteocalcin, onto a mineral and metal oxide surface. Here, we apply two enhanced sampling methods to independently probe the effects of post-translational modification and peptide structure on adsorption. First, well-tempered metadynamics was used to capture the binding of acetyl and N-methylamide capped glutamic acid and Gla single amino acids onto crystalline hydroxyapatite and titania model surfaces at physiological pH. Following this, parallel tempering metadynamics in the well-tempered ensemble (PTMetaD-WTE) was used to study adsorption of the α-1 domain of osteocalcin onto hydroxyapatite and titania. Simulations were performed for the α-1 domain of osteocalcin in both its fully decarboxylated (dOC) and fully carboxylated (OC) form. Our simulations find that increased charge density due to carboxylation results in increased interactions at the interface, and stronger adsorption of the single amino acids to both surfaces. Interestingly, the role of Gla in promoting compact and helical structure in the α-1 domain resulted in disparate binding modes at the two surfaces, which is attributed to differences in interfacial water behavior. Overall, this work provides a benchmark for understanding the mechanisms that drive adsorption of Gla-containing mineralizing proteins onto different surface chemistries.

[1] This chapter was reproduced with permission from S. Alamdari and J. Pfaendtner. Impact of glutamate carboxylation in the adsorption of the α-1 domain of osteocalcin to hydroxyapatite and titania. Molecular Systems Design and Engineering, 5, 620-631 (2020). Copyright 2020. Royal Society of Chemistry.[259]

2.2 Introduction

When implants are introduced in the body (e.g. orthopedic or dental), the efficacy of the implanted material is hypothesized to be dependent upon its ability to mineralize or fuse properly with its surrounding tissue.[10–13] This phenomenon, known as osseointegration, takes place at the cell-implant or mineral-implant interface and can be promoted by the chemical or structurally specific design of the implant interface.[14–16] Work in the field of tissue engineering has revealed that specific surface properties of the exposed implant material are influential in guiding these initial cellular events (e.g., protein adsorption to the implant surface) and thus regulating the environment of the extracellular matrix.[13,17,18] In spite of these advances, much remains to be discovered regarding the specific molecular-level recognition mechanisms between bone extracellular matrix proteins (ECM) and both mineral and implant surfaces. This is particularly important for the goal of optimizing the molecular scale design of new materials for implant coatings.

Titanium and its alloys are commonly used as implant materials due to their high biocompatibility and mechanical (load bearing) properties. Upon exposure to oxygen, the highly reactive surface oxidizes and forms a layer of titania (TiO_2) between 2 nm and 5 nm thick.[19] It has previously been shown that implant treatment with grit-blasting, acid-etching, anodization, or various nanomodification techniques alter the surface roughness and mechanical surface properties resulting in improved osseointegration.[19–23] To further address implant biocompatibility, many groups have looked towards the adhesion of coatings to implanted materials which can improve cell recall to the implant surface.[11,17,24,25] For example, Baranowski et al. demonstrated a bone sialoprotein coating on titania implants first delayed the early stages of osteoblast formation, but ultimately positively influenced ECM mineralization when compared to uncoated titania.[26] Modifications of the implant surface have been shown to impact osseointegration in a variety of ways, which have been extensively summarized by Damiati et al.,[27] and Jemat et al.[28] Prior work concludes that the adsorption of cells to the implant surface is highly susceptible to both changes in morphology and chemistry of the surface. However, the fundamental mechanisms driving cell adsorption onto different materials cannot be captured with experiments alone.

The ECM contains both organic and inorganic components. The inorganic component of the ECM is mainly composed of hydroxyapatite (HA) crystals, a calcium phosphate mineral with formula $Ca_5(PO_4)_3(OH)$, and a major component of bone. Osteoblasts are responsible for the production of organic osteoid, which includes both collagen and non-collagenous proteins (NCPs). While collagen makes up approximately 90% of the ECM, NCPs are understood to be key regulators in mineralization.[29–31] NCPs undergo posttranslational modifications (i.e. carboxylation in osteocalcin and matrix Gla protein, or phosphorylation in osteopontin and bone sialoprotein) resulting in an increased negative charge density that has been attributed to the strong adsorption of these proteins onto mineral surfaces.[32] This led to the hypothesis that structurally specific adsorption of NCPs to the newly implanted surface is what allows NCPs to effectively carry out their mineralization role and integrate properly; inhibition of these processes are predicted to promote implant fouling.[14,33,34] Of these NCPs, Osteocalcin is the most abundantly occurring and is commonly used as a biomarker in bone health. Its structure is composed of three alpha helical domains with a single disulfide bond between the α−1 and α−2 domains, and this sequence is highly conserved across porcine, mice, and humans.[35,36] The calcium-binding affinity of osteocalcin is attributed to the presence of 3 vitamin-K dependent post-translationally modified γ-carboxyglutamic acid (Gla) residues located on the α−1 domain as shown in **Figure 2.1**.

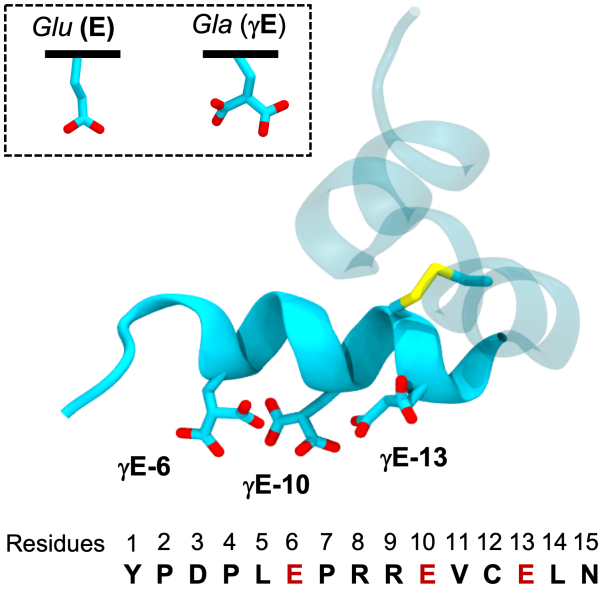


Figure 2.1. Representation of osteocalcin in its fully carboxylated form (PDB: 1QH8[36]), and the structure of Glu and Gla residues. The α-1 helix containing the first 15 residues is shown in cartoon representation and its sequence is listed below. Other protein residues are presented transparently in cyan. Residues that undergo a posttranslational modification from Glu to Gla are listed in red. Gla residues and residues C23 and C29 between the α−1 and α−2 helices forming a disulfide bond are pictured in licorice and colored by blue carbon, red oxygen, blue nitrogen, and yellow sulfur atoms.

The structure of osteocalcin and decarboxylated osteocalcin in the presence of calcium ions has been studied using circular dichroism, nuclear magnetic resonance (NMR) spectroscopy, and X-ray crystallography in solution concluding the important role calcium ions in producing a secondary structure of the protein in solution.[35–39] Only a handful of studies[40–43] have resolved the structure of biomineralizing proteins on mineral or implant surfaces, and even fewer for NCPs. Notably, in one study Scudeller, et al. used spectroscopy, spectrometry, and isotherm methods to reveal the orientation and conformation of the two forms of osteocalcin on phosphate and silica scaffolds, showing higher adsorption of all forms of osteocalcin on calcium phosphate when compared to silica, and greater helical denaturation on the silica surface.[44] This and another study[39] suggest that the mechanism of mineralization is related to the secondary structure of the adsorbed protein, and is partially dependent upon the surface chemistry of the substrate. To our knowledge there are no experimental studies using these same approaches to probe their structure on titania substrates. Rather, work in this area has focused on macroscopically characterizing osteoblast affinity and function.[45–47] These results, and lack of detail describing the mechanisms that govern these interactions in the literature have inspired additional molecular dynamics (MD) work in this

area with the goal of gaining a more microscopic understanding of the behavior of NCPs at mineral and implant interfaces. The role of sequence and post-translational modifications in the biorecognition mechanisms of biomineralizing proteins have previously been studied with success using MD simulations. These studies have been limited in scope, generally focused on understanding the effects of a common post translational modification phosphorylation[48–50], collagen protein[51], and another NCP[52]. Surprisingly, while there is a large interest in understanding phosphorylation, there are few studies that have looked at the effects of Gla post-translational modifications[53,54].

Many prior MD studies on HA have used generalized surface chemistries which corresponded to basic pH.[39,55–57] Older forcefields were modeled after bulk HA crystals lacking parameters for phosphate protonation on the water-exposed surface that occur in biological systems. The lack of parameterization of an interface at realistic conditions was resolved by Lin and Heinz who published a pH specific forcefield in 2016, where changes in pH are accounted for through protonation of surface exposed phosphate groups and the removal of calcium to maintain system charge neutrality.[58] Through this, they demonstrated pH to be influential in peptide adsorption on HA.[58] These limitations in HA simulations have also been discussed extensively by Walsh et al.[59] and continue to be addressed for this complex surface through more robust forcefield development.[53,60–62] Fortunately, this advancement has prompted the ability for new studies to begin to investigate peptide behavior on biologically relevant surfaces. Recently, an enhanced sampling MD study[63] quantified the energetic effects of phosphorylation on the mineralizing protein statherin binding onto HA, and a combined NMR/enhanced sampling MD study[64] uncovered the structure of a statherin mutant on HA, silica, and titania surfaces. These studies motivate using enhanced sampling MD to uncover molecular design rules, a protocol that has shown success in capturing both thermodynamic and structural fingerprints for large peptide-surface systems.

With respect to simulations on titania, there has been a large amount of work in the literature that has simulated the behavior of small molecules and unstructured titania-binding peptides onto titania surfaces[65–68] providing guidance towards understanding the sequence/structure/function relationships of NCPs. Defining the surface of a metal oxide in classical models has proven challenging due to the inability to dynamically represent water dissociation. To tackle this, Předota et al. developed a set of surface models that represent different

surface charge states and varying levels surface hydroxylation that describe four plausible molecular-level scenarios occurring at the titania interface in an aqueous environment.[69] With this in mind, we investigated the neutral hydroxylated interface, which allows us to directly probe the influence of surface composition when compared to the neutral hydroxyapatite hydroxylated surface model used in this study. The neutral hydroxylated model provides a model system for the case where the surface of titania is neutralized through the dissociation of water at the interface, forming terminal and bridging hydroxyl groups with exposed titania and oxygen molecules.

The aim of this work was to understand how structure, function, and chemistry of both the surface and peptide manifest in these unique peptide-surface interactions, and ultimately in different biorecognition mechanisms. A deeper understanding of these mechanisms, and the variables that influence them will guide the rational design of new implant coatings. Our approach applied MD simulations and the enhanced sampling method Parallel Tempering Metadynamics in the Well-Tempered Ensemble (PTMetaD-WTE) to extract energetic and structural information of the α-1 domain of carboxylated osteocalcin (OC) and the fully decarboxylated α-1 domain (dOC) adsorbed onto model basal HA (001) and neutral hydroxylated rutile (110) TiO2 surfaces at physiologically relevant pH. Lastly, we comment on how the results from these simulations can be generalized in the design of new implant coatings.

2.3 Methods

The PTMetaD-WTE enhanced sampling method was used for all simulations which has previously been shown to exhaustively sample peptides on surfaces.[70,71] All simulations were performed using GROMACS 2018.3[72] and enhanced sampling simulations were carried out using the Plumed 2.4.2 plugin.[73] The peptides were modeled using the CHARMM36 force field[74] and SPC/E water[75], which are compatible with both of the surface models we implemented. While the CHARMM36 forcefield was designed to be used with TIP3P water model, we chose to use the SPC/E water model which was used in development of the titania model parameters.[69] Previous work understanding the reliability of SPC in combination with CHARMM for biomolecular simulations on titania has shown this is a reasonable choice.[76] Additionally, The INTERFACE forcefield which was parameterized with CHARMM36 in mind, has been shown to capture biomolecular and polymer binding energies within 2-4% using both SPC/E and TIP3P forcefield water models supporting this choice for the HA simulations as well.[77] The Gla forcefield was built

by using existing CHARMM36 parameters for glutamic acid, and additional details and parameters used for Gla have been included in the **SI** (**Appendix A**). The INTERFACE forcefield[77] was used to model the HA (001) surface, and neutral hydroxylated Rutile (110) TiO_2 surface was modeled using parameters from Předota et al.[69], all near pH 7. The pH 10 HA surface given in the INTERFACE forcefield was modified to approximate pH 7 surface by adjusting the ratio of dihydrogen phosphate ($H_2PO_4^-$) and monohydrogen phosphate (HPO_4^{-2}) surface groups to be to be 69%, and 31% respectively (converting 16 dihydrogen phosphate groups to monohydrogen phosphate per face of the surface, 32 total), and adding 8 additional INTERFACE calcium ions per face (16 total). This ratio of phosphate groups is a close approximation of the chemistry at pH 7 (70%/30%). Calcium and chloride counterions were added to neutralize system charge. In recognition of the modified partial charges in the HA model, $Ca^{1.5+}$ ions were used in the HA/INTERFACE system whereas we used the standard Ca^{2+} from the CHARMM36 model for the TiO_2 system to maximize consistency within each model system. Sultan et al. previously verified that the addition of 2+ charge calcium ions can be appropriately used with the negative non-hydroxylated titania surface model, making this an appropriate choice.[78] **Table A-S1** includes information about the system charge and total ions added for each unique simulation.

Each surface was built by replicating the unit cell in the x/y dimension and given a z-dimension of at least 10.5 nm in length to prevent self-interaction across the periodic boundaries. A peptide structure for the α-1 domain sequence shown in **Figure 2.1** was generated in Avogadro 1.2.0[79] and placed near the surface. Following this, the HA surface (7.6 nm × 6.6 nm × 11.0 nm) was solvated with an average of 12,100 water molecules and the TiO_2 surface (5.4 nm × 5.2 nm × 10.7 nm) was solvated with 8,200 water molecules. All systems were first minimized using a steepest descent algorithm to remove any unfavorable contacts over 10,000 steps. A timestep of 2 femtoseconds was used in all simulations, and hydrogen bonds were constrained using the LINCS[80] algorithm. The systems were then equilibrated to standard temperature and pressure, over 1 ns total, using a stochastic global thermostat[81] to couple temperature (tau=0.1 ps) and semi-isotropic pressure scaling using the Parrinello-Rahman barostat[82] (tau=10 ps) with x/y and z compressibility set to 4.5×10^{-15} bar^{-1}, and 4.5×10^{-5} bar^{-1}. Van der Waals interactions were calculated below a cutoff of 1.1 nm, and electrostatic interactions were calculated with particle-mesh Ewald[83] summations using a 1.2 nm cutoff. A 1 ns NVT simulation was then used to generate starting configurations for the enhanced sampling production runs, which were also carried out in

the NVT ensemble. During production simulations all phosphate atoms, and all bulk calcium ions in the HA system were held frozen. Similarly, all titanium and all bulk oxygen atoms were held frozen in the TiO_2 system. The bulk atoms are defined to be all the atoms in the surface excluding the first two layers on both the top and bottom interfaces. This was done to preserve surface integrity over the course of the long simulations, while simultaneously allowing interactions with the fluctuating surface atoms.

We first equilibrated 24, and 25 replicas for titania, and HA respectively over 100 ps to temperatures ranging from 300-450K in the NVT ensemble. An ensemble of initial peptide configurations was generated from a 500K NVT simulation of the peptide in water. Different peptide configurations from this trajectory were used as starting structures for each replica. This temperature range has been previously used for biasing large peptides with hidden free energy barriers that are otherwise difficult to sample in the MetaD scheme alone[84]. The specific temperature values for each replica in each system are provided in **Table A-S3**. The potential energy was then biased with metadynamics for 10ns, to establish the well-tempered ensemble[85,86] The potential energy was biased using an initial hill height of 2 kJ/mol, a bias factor of 10, and 1hill/1ps hill deposition rate, attempting replica exchange at the same pace. A sigma values of 290 and 400 kJ/mol were used for titania and HA, respectively. These values were calculated by using the half the standard deviation of the smallest equilibrium fluctuations in potential energy from the initial PT simulations, which were all calculated from the lowest temperature replica. The well-tempered ensemble was achieved by observing a constant replica exchange probability near 30% while remaining computationally feasible.[87] In the case of our systems we found 24 replicas were needed with the titania system for 37% exchange probability, and 25 replicas were needed with the HA system for 25% exchange probability. Once WTE was achieved, an additional 2D WTMetaD bias potential was introduced. The WTE deposition rate was reduced by a factor of 5 (1hill/5ps) which has been previously shown to maintain constant exchange, smoothing out any effects from the additional MetaD bias applied to the CVs.[88] This was done to enhance the fluctuations of those slow degrees of freedom to sampling peptide conformation and adsorption on and off the surface. The 2 CVs chosen for this study were; CV1 – the peptide radius of gyration using the alpha carbons, and CV2 – the z-distance between the all atom center of mass (COM) of the peptide and a reference surface atom. These CVs were biased using sigma values of 0.2nm, and 0.1nm, respectively, an initial hill height of 2 kJ/mol, a bias factor of 10, and 1hill/1ps hill

deposition rate in the 2D MetaD biasing scheme. In this step a harmonic restraint (k = 50,000 kJ/mol/nm^2) was introduced on the z-component of CV2 keeping the peptide within 4.5nm of the surface. This was done to limit sampling to one face of the surface and accelerate convergence. Production runs were extended up until convergence, which is further described in the **SI** (**Appendix A**). This window ranged from between 300 ns and 500 ns per replica for each subsystem. Considering that the total simulation time as the product of simulation time per replica and the number of replicas, this resulted in a total of 6-10 microseconds of computational time needed for each system to reach convergence.

Simulations of acetyl and N-methylamide capped Glu and Gla amino acids were done to characterize the impact of modification alone on adsorption to the surface. Each amino acid was simulated on both surfaces using 4 walkers in the multiple-walker well-tempered MetaD scheme. In the HA system, $Ca^{+1.5}$ INTERFACE ions were used as counterions, in the titania system sodium ions were used as to not introduce any competitive interactions with surface binding as to probe the impact of modification on surface binding energetics alone. Additional system setup details are included in **Table A-S2**. Systems were equilibrated using the same criteria as described above. A single 1D bias was added on the z-distance between the COM of all atoms in the peptide analogues.[51] All single amino acid simulations used a sigma value of 0.1nm, initial hill height of 2 kJ/mol, a bias factor of 10, and 1 hill/1ps deposition rate. A harmonic restraint (k = 50,000 kJ/mol/nm^2) was introduced on the z-distance to the surface, at 4.5 nm to limit sampling to one face of the surface. These simulations were carried out until convergence, following the same criteria for the other enhanced sampling simulations.

Standard reweighting procedures were used to obtain equilibrium probability distributions for unbiased CVs and is further described in the **SI** (**Appendix A**). All corresponding PLUMED input files required to reproduce the results reported in this paper are available on PLUMED-NEST (www.plumed-nest.org), the public repository of the PLUMED consortium (plumID: 19.079).[89]

2.4 Results and Discussion

To assess simulation convergence of all enhanced sampling simulations, free energy as a function of the biased CV was monitored over time. Binding free energies were calculated using a Boltzmann averaged difference of the Helmholtz free energy between the surface adsorbed and solution states. All systems were considered converged when the fluctuations in energy were less

than k_BT at 300K (~2.5 kJ/mol) for the final 30% of simulation time, during which the system continued to explore CV phase space. The binding free energy time series are included in **Figure A-S1** and discussed in greater depth in the **SI** (**Appendix A**).

Influence of Surface Composition

To date and to our knowledge, there are no experimental studies that have reported binding energies of dOC and OC on HA or titania to use as benchmarks for the thermodynamic quantities calculated from simulation. Furthermore, HA and TiO_2 were parameterized with different methods, thus it is important to stress extreme caution should be used in making direct quantitative comparisons between the binding free energies calculated from binding on HA and TiO_2. However, we recently showed in a combined NMR/MD study of statherin adsorption and folding on HA, titania, and silica (using the same forcefields) strong structural agreement between experimental and simulation results, indicating that the surface models are able to sufficiently capture the necessary physics to model these biomolecular-surface interactions.[64]

This study is motivated in estimating the structure of the surface adsorbed peptides and probing the thermodynamic and structural implications of the Gla mutation as a function of changing surface chemistry and charge. We began our investigation profiling the water density for the two surfaces (**Figure 2.2**), which match prior reported water densities.[58,69] These water densities are calculated from the nearest frozen heavy atom layer, for each surface respectively. The behavior of water on titania differs fundamentally from that of the hydroxyapatite surface. Most notably, there is a substantial heterogeneous charge distribution in titania resulting in highly ordered water behavior at the interface.[69]. The 001 face of HA, has been characterized as having the least affinity towards water, which is described by a water density profile that contains a peak at a slight distance from the surface.[58]

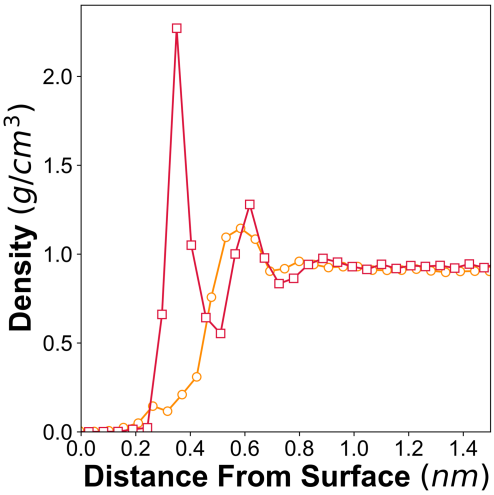


Figure 2.2. Water density profile on HA (red squares) and titania (orange circles) surfaces.

Similarly, ion affinity varies between the two surfaces. HA contains strong ordered layers of calcium ions within its crystal structure, and a driving force for calcium ions to bind directly to the surface.[58] The behavior of ions on titania has previously been explored by Walsh et. al. on a negative titania surface model where two binding modes were reported; 1) direct surface binding and, 2) water-mediated ion binding.[78] In the case of neutral titania only water-mediated ion binding occurs, and there is no direct affinity observed for the ions to adsorb to the surface, as expected.

Impact of γ-Carboxylation on Glutamate Binding to Surfaces

Binding free energies of capped Glu and Gla amino acids onto each surface were calculated to probe the influence of the modification alone on binding. In **Figure 2.3**, Glu and Gla binding profiles on HA were reweighted along the distance between the carboxylate COM and the surface, to gain insight into the minimum energy conformations at the surface. The free energy profiles (FEP) from the biased CVs are included in **Figure A-S2**. Gla binds approximately 2 k_BT stronger than Glu to HA, a direct result of increased electrostatic interactions with the surface, based on the COM distance to the surface in **Figure A-S2**. Interestingly, Glu is able to bind closer than Gla attributed to steric restrictions in the case of Gla, which has a larger sidechain.

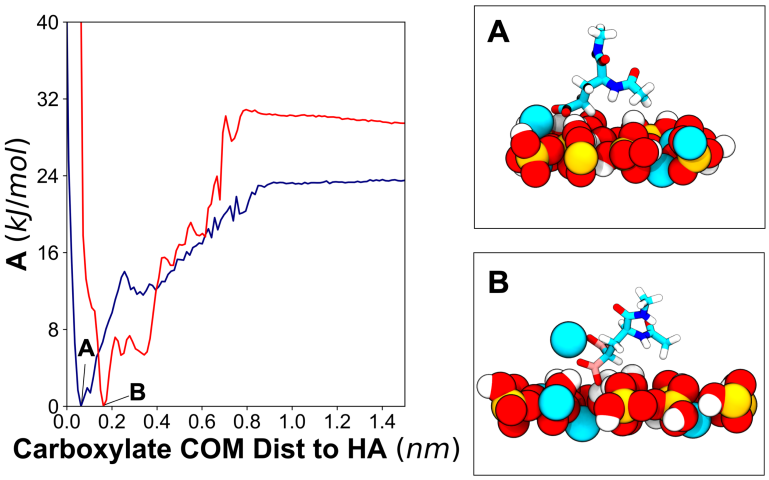


Figure 2.3. FEP for Glu (blue), and Gla (red) binding onto HA, reweighted along the distance between the carboxylate centers of mass and the surface. Representative structures corresponding to free energy minima are shown for (**A**) Glu and (**B**) Gla, respectively. Surface atoms are colored by orange phosphorous, red oxygen, white hydrogen, and cyan calcium. Amino acids are shown in licorice representation, corresponding to the same color scheme as **Figure 2.1**.

The TiO_2 binding profile in **Figure 2.4** contains many more features than HA, indicating adsorption behavior on the two surfaces is influenced by fundamental differences in the surface chemistry. Similar to HA, both Glu and Gla were found to bind to TiO_2 through direct sidechain interactions with surface atoms. In HA, this was largely attributed to interactions with calcium, in TiO_2 binding is achieved with surface hydroxyl groups. Gla binds approximately 5 k_BT stronger than Glu to the surface, based on COM binding energies reported in **Figure A-S1** which is also directly attributed to increased surface interactions. Similar to HA, Glu was able to bind much closer to the surface than Gla, likely due to steric reasons exemplified in **Figures 2.4A and 2.4D**, which represent then minimum energy structures bound closest to the surface. Interestingly, the strongly adsorbed water layer formed on titania was able to support water-mediated binding poses (**Figure 2.4C and 2.4F**) with a considerable binding affinity in the case of both Glu and Gla.

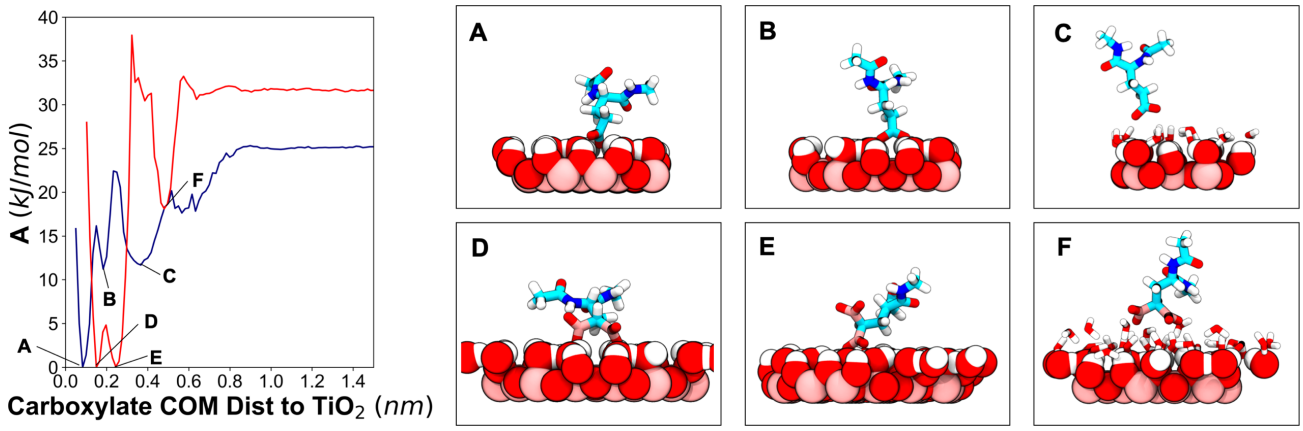


Figure 2.4. FEP for Glu (blue) and Gla (red) binding to TiO_2 reweighted along the distance between the carboxylate centers of mass and the surface. Representative structures for free energy minima associated with Glu binding are shown in (**A**), (**B**) and (**C**). Minima structures for Gla binding are shown in (**D**), (**E**), and (**F**), and labeled on the FEP. Surface atoms are shown in pink for titania, red for oxygen, and white for hydrogen.

The single amino acid binding profiles reveal site-specific adsorption to HA and TiO_2 for both wild type glutamate as well its carboxylated version. The presence of positively charged sites on both surfaces, with a favorable register against the spacing of negative charges on the amino acid sidechains leads to increased binding for the Gla residue in approximately the same amount on both surfaces. However, the contributing factors to binding are not solely due to the increased Coulombic interactions from the side chain owing to the fact that both HA and TiO_2 show roughly a 1/3 increase in binding energy upon carboxylation. Thus, other factors such as van der Waals interactions or Coulombic interactions from the backbone may also substantially contribute to binding.

Glutamate Carboxylation Influences Peptide Binding Affinity through Interfacial Structure

To explore the impact of modification on the magnitude of binding to HA and TiO_2 binding affinities for both systems were calculated for the fully unmodified (dOC – containing all Glu residues) and fully modified (OC – containing all Gla residues) forms of the α-1 helix and shown in **Figure 2.5** (n.b., here we follow consistent nomenclature from the literature adopting dOC and OC in discussing osteocalcin posttranslational modifications). In the case of HA, OC binds with over twice the affinity, and slightly closer to the surface than dOC. The OC sequence is more negatively charged than dOC, indicating adsorption to HA is driven by electrostatic interactions,

and subsequently the overall charge density of the system agreeing with what was predicted from the single amino acid binding profiles. Electrostatic driven binding has also been reported for phosphorylated peptides to metal oxide surfaces[90], LK peptides on negative self-assembled monolayers[88], and large globular proteins lysozyme and α-lactalbumin at charged interfaces[91]. In contrast, on the titania surface there was almost a tenfold decrease in the binding affinity with increasing charge density; dOC bound much stronger to the surface than OC, and OC was also bound much further to the surface than dOC. These non-linear binding trends prompted further exploration into the role of peptide structure on the impact of binding.

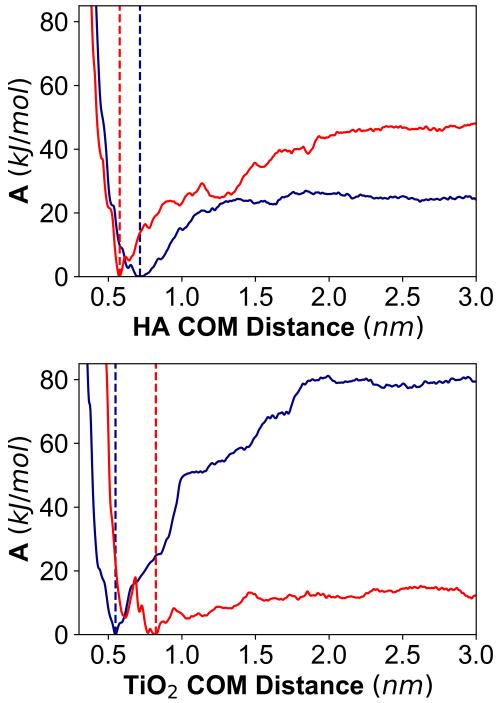


Figure 2.5. FEP for dOC (blue) and OC (red) for the all-atom peptide COM distance to HA (**top**) and TiO_2 (**bottom**). Dashed lines indicate location of free energy minima on each FEP.

Structure, configurational entropy, and the behavior of water on the two surfaces are discussed as dominating driving forces causing differences in adsorption behavior. Free energy surfaces (FES) describing the peptide distance to the surface, and radius of gyration were used to characterize surface-bound and solution peptide structure of each system. The radius of gyration was reweighted based on solution (distance > 3.0 nm) and surface (distance <= 3.0 nm) distance from the surface, to clearly understand the impact of modification on structure. The 2D FES are

provided in **Figure A-S9** for reference. Clustering techniques were then used to describe the peptide structure at minimum energy regions on the FES and to further quantify entropic driving forces. And lastly, reweighted FES were used to quantify contacts between the peptide modification sites and the surface which are further discussed in detail in the **SI** (**Appendix A**).

Figure 2.6 shows the structural analysis for the surface bound states of dOC and OC adsorbed to HA (i.e. a COM distance <= 3.0 nm). The free-energy profile for the peptide radius of gyration shows that dOC adopts an extended conformation at the interface, given by the free energy minima at a radius of gyration value of 1.2 nm. This extended conformation persisted in all 3 of the top structures from clustering the structures at the surface, representing over 50% of the surface bound conformations (**Figure A-S3**). A similar analysis for configurational energy was done for the peptide in solution (i.e. at COM distances > 3.0 nm from the surface) and is given in **Figure A-S5**, with the top structures from clustering given in **A-S7**. In solution, dOC adopts a slightly extended conformation, but the energetic differences between collapsed and extended states are significantly smaller than the adsorbed state, indicating the unmodified peptide has no conformational preference in solution.

In contrast, the carboxylated peptide OC shows a stabilized collapsed or compact structure on the surface. Additionally, OC demonstrated helical adsorption onto HA, which persists in the top clusters (**Figure A-S4**). This helical structure captured in **Figure 2.6B** and **C**, is slightly different from the helical structure proposed by the solution studies in **Figure 2.1** where helices are presumed to be stabilized due to strong Gla/divalent interactions in solution. In the surface adsorbed structure of OC, a kinked tight helix with two turns appears in two of the top clusters, representing 79% of all adsorbed conformations. This kinked helix was found to be due to strong stabilized amino acid/ion interactions between Asp-3 and Gla-6 and the surface, and a proline residue (Pro-5) induced kink in the backbone. Two different hydrogen-bonding modes forming an irregular helix were found to occur in the backbone and have been characterized further in **Figure A-S12**. No helical structure is observed past residue Arg-9, in the second half of the peptide. In the full osteocalcin sequence, a disulfide bond at position 12 acts as a stabilizing bridge between the α-1 and α-2 domains, suggesting the presence of the second helical domain might be critical in stabilizing structure in the second half of the α-1 domain. Scudeller et. al. proposed that the α-2 domain also adsorbed onto HA, prompting future investigation in the role of this second domain and its stabilizing disulfide bond.[44] Unlike dOC, OC takes on a distinctly more compact form in

solution (**Figure A-S5 and A-S8**). This indicates that the combination of calcium and carboxylation lead to a distinct solution structure. This agrees with previous experimental work which has reported OC to be unstructured in solution without the presence of Ca^{2+} ions, and supports the importance of the Gla modification in promoting structure of the full protein.[92]

To further characterize increased binding due to increased interactions at the surface, the simulations were reweighted to quantify the number of interactions between Glu or Gla residues and the surface. The reweighted FES for dOC on HA shows that the adsorbed peptide has an average of 1 O-H interaction, and 2 O-Ca^{1+5} interactions with Glu (**Figure A-S10**) indicating the extended surface adsorption is a function of many residues on dOC interacting with the surface. As shown in **Figure A-S10**, OC adopts a number of energetic minima at the surface. The deepest free energy minima accounts for an average of 6 O-H interactions, and 13 O-Ca^{1+5} interactions with Gla. In this case, carboxylate side groups are stabilized by both protein/surface and protein/ion interactions. The modification resulted in a larger binding free energy, which can directly be described through increased surface contacts, and closer binding to the surface. This indicates enthalpic contributions act as a dominant driving force in HA-adsorption.

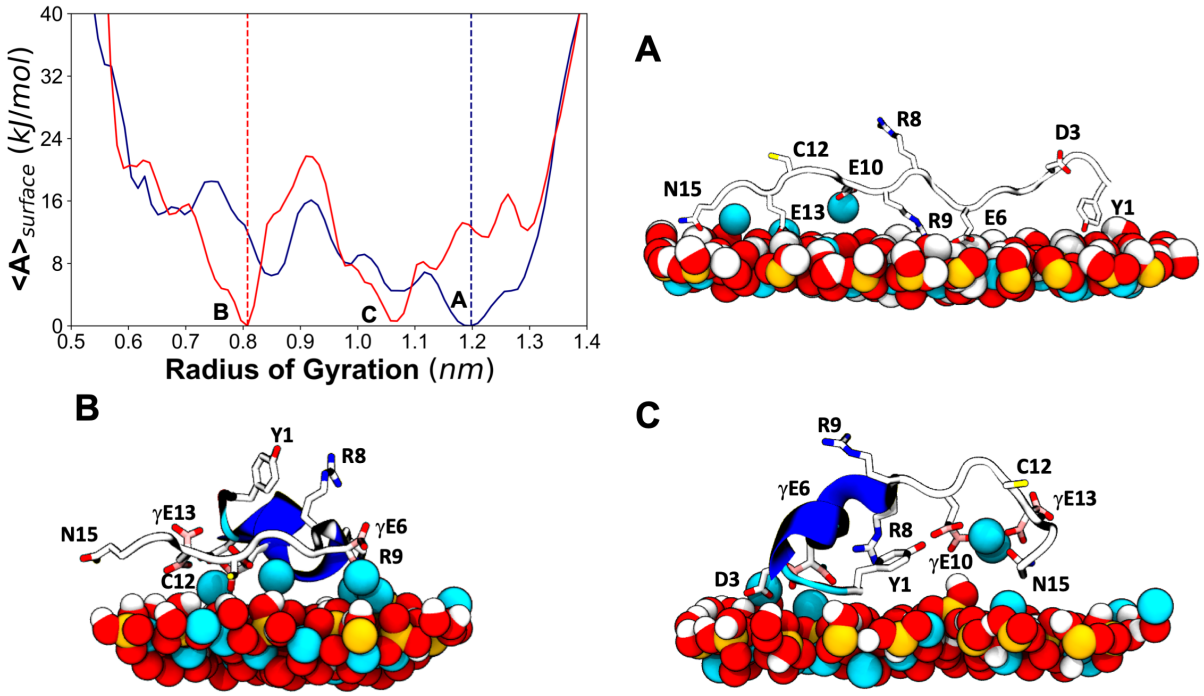


Figure 2.6. Reweighted 1D FEP for dOC (blue) and OC (red) adsorbed on HA along the radius of gyration calculated from c-alpha atoms. (**A**) Representative snapshot of the lowest free energy minima found for dOC, representing 38% of the surface adsorbed structures. Representative snapshots of the free energy minima for OC at positions (**B**) and (**C**) on the FES. OC structures represent the ensemble of configurations described by cluster 1 and 2 (**Figure A-S4**) which in total accounts for 79% of the surface adsorbed structures. Amino acids are labeled and represented in licorice, not including hydrogens, carbons are shown in white. Nonpolar non-interacting amino acids were omitted from the representation for clarity. The structure of the peptide is represented by the new cartoon drawing scheme colored by structure, where a 1-4 helix is shown in purple, irregular helices are given in royal blue, turns are represented by cyan, and coils are shown in white.

As noted above, the adsorption thermodynamics on neutral TiO_2 were found to be dramatically different compared to HA for both peptides. The unmodified peptide (dOC) was shown to have a much larger binding affinity towards the surface than the carboxylated form (OC), opposite the single amino acid predictions. This indicates the role of structure is a critical consideration in these systems. Interestingly, although demonstrating different binding affinities, the both HA and TiO_2 showed similar structural behavior at the interface and in solution. As shown in **Figure 2.7** dOC is extended at the surface, and this persists in all of the top clusters, as shown in **Figure A-S3**. This mechanism of interaction is similar to a recent study showing that extended phosphorylated R5 peptides derived from silaffin strongly bind quartz surfaces compared to collapsed and more globular peptides.[48] Similarly, OC adopts a compact form at the surface, with some structure in the top clusters shown in **Figure A-S8**. A similar kinked helix as described in **Figure A-S12**, is seen in the top structures for TiO_2, but less prominently than upon adsorption to

HA. This is attributed to differences in the crystalline structure of the two surfaces, which do not contain bound calcium ions, and the alternate modes by which the peptide interacts with the surface.

It is important to note that while we can predict that the $Ca^{1.5+}$ interactions could be underestimating the energetic impact on structure especially in solution, we see the same conformational trends in solution (albeit to a lesser degree in the case of the $Ca^{1.5+}$ containing system) independent of the charge on calcium used; dOC shows little structural preference in solution whereas OC adopts a significantly compact conformation in solution (**Figure A-S6**). Future work should look to investigate the impact of the HA force field's use of $Ca^{1.5+}$ and the energetic impacts on Glu and Gla binding, perhaps with the use of DFT simulations to accurately capture solvation free-energies and enthalpic contributions to the binding free-energy.

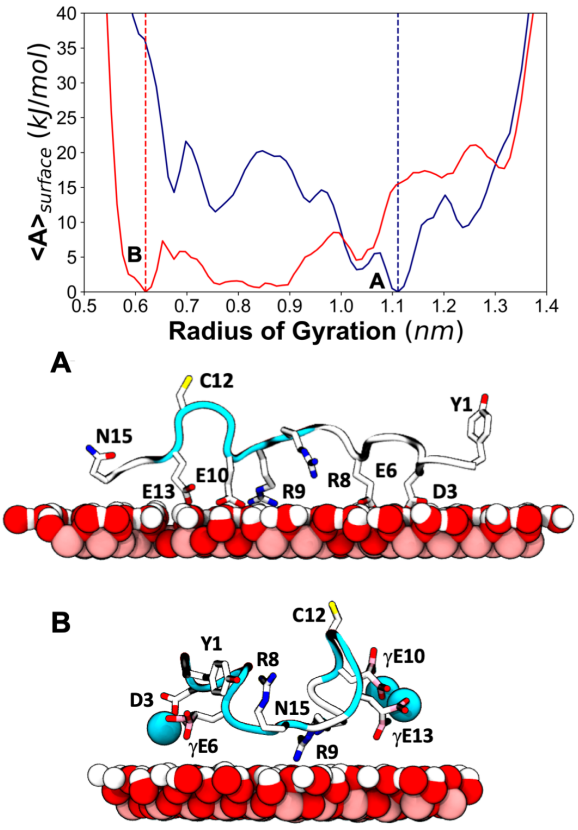


Figure 2.7. Reweighted 1D FEP for dOC (blue) and OC (red) on TiO_2 along the radius of gyration calculated from c-alpha atoms. (**A**) Representative snapshot of the lowest free energy minima found for dOC, representing 41% of the surface adsorbed structures. Representative snapshots of the free energy minima for OC at positions (**B**) representing 21% of the surface adsorbed structures. Follows coloring scheme from **Figures 2.4** and **2.6**.

The FES for dOC shown in **Figure A-S11**, demonstrates that the Glu residues are highly coordinated with hydroxyl groups on the titania surface with an average of 7 O-H interactions, and 3 O-Ca^{2+} interactions. In contrast, the minimum energy OC adsorbed structure has no O-H Gla interactions (no interactions with the surface), and highly interacts with calcium over an average of 8.5 O-Ca^{2+} Gla interactions. This indicates that the negative modified Gla residues strongly prefer to coordinate with calcium over the surface hydroxyls. In the neutral surface, there is no electrostatic driving force for positive calcium ions to adsorb directly to the neutral surface. The surface bound structure is instead stabilized by few water-mediated interactions of both the ions and arginine residues, giving rise to weakly bound interactions. This minimum energy binding pose, and high affinity to interact with calcium, explains the over tenfold reduction in binding energy that was reported in **Figure 2.5**. Similar solvent-mediate contact has been reported in titania binding peptides, when calcium ions were introduced to the system.[78] Peptide/calcium affinities compete with direct peptide/surface binding in the case where there are no ion/surface driven interactions.

Configurational Entropy

To further understand the trends in binding across the different surface chemistries we used a configurational entropy analysis on the peptide structures. A clustering analysis on surface and solution phase structures was used to assess the probability of each of the microstates (p_i). The change in conformational entropy upon adsorption ($\Delta S_{ads,config}$) was obtained by difference between the solution ($S_{sol,config}$) and absorbed states ($S_{surf,config}$) using **Equations 2.1** and **2.2**.

$$TS_{config} = -k_B T \sum_{0}^{i} p_i \ln(p_i) \qquad 2.1$$

$$T\Delta S_{ads,config} = TS_{surf,config} - TS_{sol,config} \qquad 2.2$$

Convergence of this analysis is demonstrated in **Figure A-S13**. This method has previously been used to estimate LK-peptides adsorbed onto SAMs[88] and does not include translational or rotational degrees of freedom, which are assumed to be less significant in this case. The entropy calculations are given in **Table 2-1**. All systems, except the adsorption of OC onto neutral titania demonstrate an entropic penalty upon adsorption with the solution phase showing more conformational flexibility. As noted in **Table A-S2**, each system contains a unique number of ions

due to the protocol of charge neutralization that was used for setting up each simulation. Therefore, great care should be taken in comparison between the different solution phase entropy values only, although the ΔS values should be comparable across all systems.

Table 2-1. Change in configurational entropy at 300K

Surface	Peptide	$TS_{surf,\,config}$ (kJ/mol)	$TS_{sol,\,config}$ (kJ/mol)	$T\Delta S_{ads,\,config}$ (kJ/mol)
HA	dOC	1.48	6.37	-4.88
HA	OC	0.91	3.17	-2.26
TiO_2	dOC	0.69	5.09	-4.41
TiO_2	OC	2.50	1.19	+1.31

A larger entropic penalty was found for dOC adsorption onto HA, and for dOC adsorption onto TiO_2. As described above, both peptides adsorb in a similar extended and rigid confirmation. OC adsorption experienced less of an entropic penalty, compared to dOC in both systems. In addition to configurational entropy, previous work on peptide adsorption[59,65,88] report that the entropic gains from tightly bound water layers can often be used to explain the driving forces behind peptide binding affinities on metal oxides and strongly hydrophobic surfaces. And similarly, large binding affinities have also been reported for extended peptides binding onto other surfaces.[48] As noted above although care should be taken when comparing energetics across different force fields, we expect the entropic gains due to water when the peptide binds on TiO_2 to be much larger than when binding on HA from the water density profiles given in **Figure 2.2**. This can partially rationalize the large difference in binding free energetics between two structurally similar systems, which experience different driving forces.

In contrast to dOC, binding of OC shows different entropic trends between the two surfaces. While they were structurally similar, binding on HA was mediated by ion interactions at the surface, and binding on TiO_2 was driven by water-mediated binding of ions and the peptide sidechains. Calcium ions have previously been shown to influence the structure of titania binding peptides thereby influencing conformational flexibility when adsorbed on titania.[65,78] The increase of binding entropy on OC adsorption to TiO_2 is consistent with the other energetic and structural analyses, namely that this peptide binds the weakest (**Figure 2.5**) and furthest from the surface (in

a water-stabilized binding mode) and shows a wide range of stable Rg values on the surface (**Figure 2.7**).

2.5 Conclusions

This paper presents a comprehensive study of the behavior of the first domain of Osteocalcin interacting with different model surfaces. Using single amino acid studies as a starting point, we described the effect of a common Gla post-translational modification on binding to the two surfaces. In both systems, the modification increased overall binding to the surface, however we observe disparities in the binding mechanisms which can be attributed to differences in the behavior of water and ions on each surface. A strongly adsorbed water layer on TiO_2 promotes competitive binding modes that can occur at the surface. To further explore the impact of glutamate carboxylation on structure and energetics, we used enhanced sampling to capture the behavior of the α-1 helix on both surfaces.

Simulations showed that the Gla-modification provides stable binding with surface adsorbed calcium ions, giving way to helical structure. These peptide/surface interactions result in uniquely kinked helix that occurs in the first 7 residues due to strong interactions between aspartic acid, Glu, and Gla with surface ions and the presence of prolines in the peptide backbone. We predict that a fully helical structure of α-1, which was not captured in these simulations, may be partially dependent on adsorption and stability of the α-2 helix, which was also previously predicted to bind to HA experimentally.[44] A single disulfide bond between the α-1 and α-2 helices would provide a significant amount of stability to the N-termini that was seen to be flexible and random coil in all simulations, prompting further investigation by simulation into more complex systems.

dOC was found to behave similarly on HA and TiO_2, adsorbing in an extended and unstructured conformation. Differences in binding affinities of the similar structures, while not definitive, are likely explained by the differences in water behavior on the two surfaces, and differences in competing interactions at the interface (i.e. entropic over enthalpic binding affinities). Gla induced structure in the case of OC, resulted in further differences in the binding modes on the two surfaces. Gla was able to promote both structure and binding on HA, due to calcium ions bound to the HA surface. Alternatively, while Gla promoted structure in the case of

TiO_2, a lack of ions bound directly at the titania interface gave rise to enthalpically driven solvent-mediated binding of structured OC on TiO_2.

Osteocalcin adsorption appears to not only be a function of the specific protein/surface interactions as a function of post translational modifications but is highly dependent upon the behavior of water and ions surrounding the surface. This environment surrounding titania resulted in mechanistically different adsorption when compared to HA. Increased peptide/ion interactions appear to be important in eliciting helical structure, which are stabilized by the post translational carboxylation modification, and was found to be a dominating driving force, enhancing peptide adsorption to HA, and disrupting binding to TiO_2.

Our study helps suggest design principles for surface coatings that could help promote binding similar to the natural HA surface. Namely, design of implant coatings should consider two aspects found to be critical in replicating adsorption to HA; 1) high enthalpic (electrostatic) driving must be designed into the system (i.e. some inherent affinity of the surface for calcium), and 2) surfaces should minimize strong ordering of water. Effects like surface etching and modification that disrupt the highly crystalline nature of a titania surface and improve osseointegration may indirectly impact the effect of tightly bound waters at the surface, prompting further consideration and investigation from both an experimental and simulation standpoint. Similarly, the use of negative or non-hydroxylated surface titania surfaces could positively influence more direct calcium-surface interactions, prompting further investigation into additional model simulations.

Nonetheless, the ability to gain some insight into the driving forces of these protein-surface interactions using simulations alone is exciting. These simulations are a first look at understanding the mechanisms that govern protein binding interactions at mineral and implant surfaces in the ECM and prompt future studies.

2.6 Acknowledgements

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1762114 and in part by National Institute of Health award R21-A126113. Part of the simulations in this work were facilitated though the use of advanced computational, storage, and networking infrastructure provided by the Hyak supercomputer system and funded by the STF at the University of Washington. The authors would

like to thank Rachel Gebhart, Gary Drobny, Janani Sampath, and Orion Dollar for their helpful discussions and insight.

Chapter 3
Orientation and Conformation of Proteins at the Air-Water Interface Determined from Integrative Molecular Dynamics Simulations and Sum Frequency Generation Spectroscopy[2]

3.1 Abstract

Understanding the assembly of proteins at the air water interface (AWI) informs the formation of protein films, emulsion properties, and protein aggregation. Determination of protein conformation and orientation at an interface is difficult to resolve with a single experimental or simulation technique alone. To date, the interfacial structure of even one of the most widely studied proteins, lysozyme, at the AWI remains unresolved. In this study, MD simulations, are used to determine if the protein adopts a side-on, head-on, or axial orientation at the AWI with two different forcefields GROMOS-53a6+SPC/E and a99SB-*disp*+TIP4P-D. Vibrational SFG spectroscopy experiments and spectral SFG calculations validate consistency between the structure determined from MD and experiments. Overall, we show with strong agreement, that lysozyme adopts an axial conformation at pH 7. Further, we provide molecular-level insight as to how pH influences the binding domains of lysozyme resulting in side-on adsorption near the isoelectric point of lysozyme.

[2] This chapter was reproduced with permission from S. Alamdari, S. J. Roeters, T. W. Golbek, L. Schmüser, T. Weidner, and J. Pfaendtner. Orientation and Conformation of Proteins at the Air–Water Interface Determined from Integrative Molecular Dynamics Simulations and Sum Frequency Generation Spectroscopy. Langmuir, 36, 40, 11855–11865 (2020). [522]

3.2 Introduction

Protein adsorption is important for the stabilization of interfaces.[93] Proteins have markedly complex and diverse structure due to a high variability in size, shape, amphiphilicity, and charge needed to carry out their biological function. The interfacial structure of adsorbed proteins governs macroscopic properties such as foaming and stabilization, relevant to applications in pharmacy, biotechnology, and the food industry.[94–97] However, there is no single experimental or simulation technique alone that can determine the interfacial structure of proteins at a high level of resolution. One massive challenge of linking the activity of proteins with their interfacial function lies in uncovering the mechanisms which govern assembly at interface, and inherently the orientation and conformation of the structures themselves. Lysozyme, a globular protein, has shown high antibacterial and food preservation properties desirable to applications in food technology.[98,99] Hen egg white lysozyme (HEWL) is also a widely used model protein, whose crystal structure was first captured in 1965, and has been used regularly for decades in the development of novel approaches to probe orientation, conformation and function of proteins.[100] However, despite being one of the most broadly studied proteins for interfacial behavior, there remains disagreement around the orientation and folding of lysozyme at the air water interface (AWI).

Experimentally, studies of lysozyme conformation at the AWI have suggested both folded[101–106] or denatured[107–110] adsorption. Spectral methods such as circular dichroism[107], and FTIR[109] as well as x-ray reflectometry[108,110,111] were used to study the secondary structure of lysozyme in solution and at the AWI. In these studies, a shift from mainly helical secondary-structure in solution to unstructured or beta-sheet content at the interface indicated surface-induced denaturation. It has even been proposed that this denaturation occurs to the extent of lysozyme fully refolding into an antiparallel beta-sheet conformation at the AWI.[110] In contrast to this, lysozyme is known to be highly robust against denaturation even in the presence of non-aqueous solvents (like urea) and is characterized as having a highly stable structure due to the presence of four disulfide bridges.[105,108,112] Many of the early experiments which proposed folded adsorption did so by matching the thickness of the adsorbed protein layers with at least one possible dimension of the folded protein (4.5 nm × 3.0 nm × 3.0 nm), and kinetic modeling.[101–103,105] Later work from Delahaije et al.[104] used relationships between surface elastic modulus and surface pressure to demonstrate the adsorbed proteins remain in their native state, and a previous vibrational sum-

frequency generation (SFG) study[106] concluded some structured organization of lysozyme at the AWI.

As described above, the dimensionality of the interfacial protein layer, determined by neutron reflectometry[101,105,113,114] or x-ray reflectometry[103] techniques, has been previously used as an indicator of protein orientation based on the thickness of the protein layer, using the length scales of the ellipsoidal globular form of lysozyme as a guideline. A layer 4.5 nm thick was determined as being adsorbed head-on, 3.0 nm thick was classified as side-on, and a value in-between 3.0 and 4.5 nm was classified as axial. Any reported thicknesses shorter than any dimension of the protein would be concluded as being denatured. However, even among studies using same techniques, there remained disagreement in the obtained layer thickness leaving unresolved as to whether the protein lies axial, side-on, or head-on. Interestingly, a handful of studies[101,102,115] have separately explored the effects of concentration on adsorption, proposing change from side-on to head-on at high concentration at pH 7. Lu et al.[114] also explored the impact of pH on adsorption, reporting a side-on conformation at both high and low concentrations when a pH of 11 was used (near the isoelectric point (pI) of lysozyme which is 11.35). There have also been a number of proposed kinetic and thermodynamic adsorption models which include terms accounting for changes in the adsorbed monolayers – both in terms of conformation and orientation (i.e. side-on to head-on, or unfolding).[102,104,116–119]

Experiments, simulations, and combined studies have previously been used to propose the structure and conformation of lysozyme at a variety of hydrophobic[113,120–124], hydrophilic[113,125–127], and charged[128–132] interfaces. Notably, in the case of hydrophobic self-assembled monolayers both disordered and folded lysozyme structures have been reported.[113,120,121,127] Molecular simulation can provide atomistic-level resolution into the assembly at interfaces, complementing experimental results nicely. This approach of using MD to validate experimental predictions has previously been used to provide insight into similar lysozyme-centric adsorption debates in the past. For example, Kubiak et. al. suggested that ionic strength and structure of the surface could provide a molecular understanding for the discrepancies in experimental studies on the structure of lysozyme at charged silica interfaces.[132] Similarly, simulations have been long used to understand the conformation of AWI-adsorbed proteins such as enzymes, antifreeze proteins, and hydrophobins, and a number of other globular proteins (i.e. β-Lactoglobulin, or β-Casein).[133–141] However, there have been limited simulation studies that have looked at the assembly of lysozyme

at the AWI. Recently, Cieplak et. al presented a coarse-grained model for modeling proteins at the AWI by relating amino acid hydropathy to its preferred orientation at the interface.[139] This was followed by a single atomistic MD study aiming to understand the differences in behavior of globular proteins (including lysozyme) at air-water compared to water-oil interfaces.[133]

Vibrational SFG is a nonlinear optical method with the ability to capture the behavior of proteins at interfaces *in situ*.[142,143] Spectra in the amide I region can provide information about orientation, folding and dynamics.[144–147] At the air-water interface, SFG can probe proteins by frequency mixing of narrowband visible and femtosecond broadband infrared laser pulses. For the resulting vibrational spectra, SFG selection rules dictate that signal is only generated by proteins adsorbed to an interface.[148] For complex proteins such as lysozyme, SFG signal photons originating from different protein sites will interfere in complex ways making direct spectral interpretation difficult. Therefor the interpretation of protein SFG spectra has often been aided through the use of complementary MD simulations.[149,150]

In this study, we combine MD, vibrational SFG, and spectral calculations to determine the orientation and conformation of lysozyme at the AWI, as outlined in **Figure 3.1**. First, using two forcefields, a number of candidate orientations are proposed. Then, the structures determined from MD are validated by comparing experimental SFG spectra to spectra calculated from simulation trajectories. The calculated spectra are averaged over several thousand structural states along the trajectory and therefore represent the ensemble of structural poses present at the interface. Overall, there was strong agreement between the most likely pose predicted by MD and experiment. Lastly, additional simulations were conducted at pH 11 to gain atomistic-level insight into disagreements over the proposed mechanisms of assembly at the AWI. In addition to providing additional concrete evidence explaining the dominant lysozyme structure at the AWI while also resolving previous discrepancies, our work provides a template for future studies of proteins at interfaces to make maximum use of integrated computational and experimental approaches.

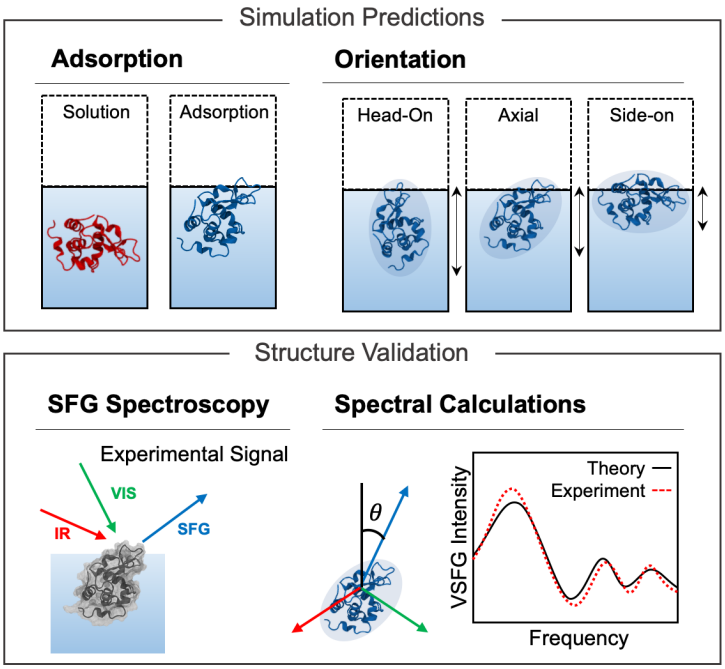


Figure 3.1. Protocol for determining interfacial structures and orientation using a combined MD+SFG approach. First simulations are used to capture the interfacial structure. SFG spectroscopy is used to measure the signal from experimentally derived structure. Spectral calculations on the structure determined from simulation, are then used to determine conformation and orientation consistent between experiment and simulation.

3.3 Methods

MD simulations

The sequence chosen for this study is derived from tetragonal hen egg white lysozyme protein PDB code 1IEE.[151] The protein was simulated in both a vacuum/water simulation box, using two different forcefields and water model combinations. The forcefields used in this study included GROMOS-53A6[152] in combination with the SPC/E[75] water model, and a99SB-*disp*[153] with TIP4P-D[154] water model. Two commonly used forcefields were chosen as a means to compare the role of potentials in studying biomolecular adsorption at the AWI. All simulations were performed using GROMACS 5.1.2.[72] Ten trials were performed for each forcefield, initialized with the protein in a random orientation (a total of 20 simulations). Four disulfide bonds were defined between cystine residues 75 and 93, 63 and 79, 114 and 29, and 98 and 126. Each simulation box was solvated with between 11614 to 11866 water molecules, and 8 chloride counterions to create a system of neutral charge in a 7.2 nm x 7.2 nm x 7.2 nm simulation box.

To ensure the solvent and ions were well equilibrated, the water and ions were heated to 500K then quenched back down to 300K over 3 ns using a stochastic velocity-rescaling thermostat (τ = 0.1 ps), while the entire protein was held frozen.[81] The velocity-rescaling thermostat (with τ = 0.1 ps), was used for temperature control in all subsequent simulations. Following this, the entire system was rapidly equilibrated to 300 K over and 1 bar using a Berendsen barostat (τ = 1.0 ps) over 2ns, followed by another NPT run using the Parrinello-Rahman barostat (τ = 1.0 ps) over 10ns.[82,155] Van der Waals interactions were set to switch at 1.0 nm using a 1.2 nm cutoff, and electrostatic interactions were calculated with particle-mesh Ewald electrostatics using a 1.2 nm cutoff. Periodic boundaries were applied in all directions.[83] All simulations were run using a timestep of 2 fs, and hydrogen bonds were constrained in all simulations by the LINCS algorithm.[80] After the water in the system was properly equilibrated to standard temperature and pressure, the z-dimension of the box was doubled (to 14.2 nm) to create a vacuum water interface (VWI), which is used interchangeably with AWI in this paper. With periodic boundaries, this resulted in two VWI (top and bottom) present in the simulations. After equilibration, the protein was placed within 2.5nm of the interface to facilitate rapid adsorption to the AWI over 100 ns. This is one approach for generating randomized trials, a number of methods can be implemented to simulate proteins at the AWI.[133,138,150] Production runs were carried out in the NVT ensemble for 100ns. One trial of each interfacial pose was carried out to 500ns to assess stability, orientation, and conformation of the adsorbed proteins at the interface (**Figures B-S1, B-S4,** and **B-S5**).

The adsorbed structures from the a99SB-*disp*+TIP4P-D trials were used to simulate lysozyme at pH 11. The lysine residues were deprotonated to -NH2, reducing the overall protein (and thus system) charge from +8 to +2. The excess chloride counterions were removed from the system to keep the system neutral. The systems were again relaxed using a steepest descent energy minimization, then simulated using the same protocol described for the other production simulations. Raw data for adsorption and orientation vs. time are included in **Figure B-S6**.

Simulation Analysis

Trajectories were analyzed using the PLUMED 2.3.0-dev library.[73] To monitor the orientation of lysozyme at the interface we used a single collective variable (CV), which described the vector ($\vec{v}$) along the proteins long axis between the center of mass (COM) of Threonine (T-69) and the COM of Cystine (C-127), as shown in **Figure B-S2,** and used the cosine of the angle

(θ) between the vector and its normal ($\overrightarrow{v_Z}$) to describe orientation. A second CV was used to track the minimum distance between the COM of the protein and all the oxygens in water molecules to approximate its distance from the AVI. This was monitored using the MAXDISTANCE (beta=0.1) colvar in PLUMED. Visualizations and surface area calculations were done using VMD 1.9.3.[156]

Experimental Methods

A Teflon trough at room temperature (22°C) containing approximately 20 mL of phosphate buffer saline (PBS; 0.01 M phosphate buffer, 0.0027 M KCl, and 0.137 M NaCl, pD 7.8, Sigma – Aldrich), prepared in D_2O (Eurisotop) was used. Throughout all experiments, the water level was held constant by a syringe pump (New Era Pump Systems Inc.) with the cannula submerged at the bottom of the trough. A powder of lysozyme from hen egg-white (HEWL, molecular weight 14 300 Da, Sigma-Aldrich) was dissolved in D_2O PBS and injected into the trough subphase to reach a final concentration of 0.1 mg/ml.

SFG Spectroscopy

The SFG setup has been described previously.[157] Briefly, the setup is based on a 7 W, 35 femtosecond laser system (Astrella, Coherent) with pulses centered at 800 nm and a repetition rate of 1 kHz. One part of the output was used to pump an optical parametric generator (OPG)/ optical parametric amplifier (OPA) with a noncollinear difference frequency generation (NDFG) extension (TOPAS Prime, Light Conversion) to generate broadband (300 cm^{-1}) IR pulses tunable at 6.1 μm. A narrow band (15 cm^{-1}) visible beam was generated by guiding 1 mJ of the fundamental through a Fabry-Perot etalon. The visible and IR beams were spatially and temporally overlapped on the surface. The SFG signal was focused into a spectrograph (303i, Shamrock) and detected by an EMCCD camera (Newton 971, Andor). All SFG spectra were recorded in SSP (S-SFG, S-visible, P-IR) and PPP polarization combination**.** The sample area and IR path were flushed with nitrogen to avoid artifacts due to IR light adsorption by water vapor**.** In the present study, SFG measurements were recorded in the amide I region 1600-1700 cm^{-1}. D_2O was used to avoid overlap between amide I band and H_2O bending vibrational modes and to reduce H_2O contamination at the sample surface. All spectra were normalized using a reference spectra obtained from gold.

Spectral Calculations

The spectral calculations are performed with a FORTRAN90 implementation of the amide-I one-exciton Hamiltonian formalism which has previously been described.[143,158] Briefly, a one-exciton Hamiltonian is constructed for the amide-I mode of the backbone amide groups of lysozyme, with couplings that are estimated differently for nearest- and nonnearest-neighbor amide groups. The nearest-neighbor interactions, which are dominated by through-bond effects, are computed using a parameterized map of an *ab initio* calculation with the 6-31G+(d) basis set and the B3LYP functional, thus providing the coupling as a function of the dihedral angle.[159,160] The nonnearest-neighbor interactions, dominated by through-space effects, are estimated with the transition-dipole coupling model.[161] The local-mode frequencies ν_{local} are shifted with the same model employed by Lu et al.[162], based on the *ab initio* study of Ham et al.[163] that finds a linear relationship between the C=O length of the amide group, and the presence and strength of each of the hydrogen bonds that the amide group can participate in, leading to the relation in **Equation 3.1**.

$$\nu_{local} = \epsilon_0 - \alpha * \left(r_{C=O} - r_{C=O\ equi.}\right) \qquad 3.1$$

The optimal values for the gas-phase frequency ϵ_0, the scaling factor α, and the number of frames over which the C=O length is averaged, was found to be 1675 cm^{-1}, 1500 cm^{-1}/Å and 10, respectively. The equilibrium bond distance $r_{C=O\ equi.}$ of C=O, was set to 1.229 Å for amide groups with secondary amines, which corresponds to the equilibrium value of a C-O bond in the AMBER99SB-ILDN force field, and to 1.232 Å for amide groups with tertiary amines (prolines), for which the local-mode frequency is also redshifted by 26.3 cm^{-1} to account for the redshift due to the larger carbon mass as compared to the hydrogen mass bound to the amide N atom in other amino acids.

The Hamiltonian is then diagonalized to obtain the amide-I eigenvalues and eigenvectors, from which the SFG hyperpolarizability is calculated by multiplication of the eigenmodes' transition-dipole moments and Raman polarizabilities. The thus-derived orientation-dependent SFG spectra are obtained by convoluting the response with Lorentzians with a half-width-at-half-max of 10 cm^{-1}. Because of the intrinsically unknown interfacial refractive index, which heavily affects the magnitude of the PPP polarization combination.[158,164], we chose to only compare the spectral shape of the calculations with the experiments, by normalizing the calculated spectra to

each of the associated experimental polarization combination spectra The spectra were convoluted with a Gaussian with a σ of 13 cm^{-1} to account for the spectral width of the visible beam. The residual-sum-of-squares was determined by interpolation of the experimental data on a 1600-1700 cm^{-1} grid with a spacing of 1 cm^{-1} and subtracting it from the calculated data before taking the squaring at each frequency point and summing these squared residuals.

3.4 Results

Simulation Predictions of Lysozyme Assembly at the AWI

In **Figure 3.2,** the dynamics of the ten trials for GROMOS-53a6[152]+SPC/E[75] and a99SB-*disp*[153]+TIP4P-D[154] forcefield combinations are analyzed. Over 100 ns and 10 trials we captured both the tendency to observe adsorption to the AWI, as well as forcefield specific effects on the orientational and conformational dynamics of the protein.

Interfacial dynamics were classified as being either not-adsorbed, transiently bound, or strongly bound (**Figure 3.2a and 3.2b**). Trials that rapidly diffused between the bulk and interface (total adsorption time < 20 ns) were considered as not-adsorbed, trials where the protein remained adsorbed for at least 20 ns of the production run were categorized as transiently bound, and trials where the protein remained adsorbed for at least 80 ns were classified as strongly bound. Overall, the GROMOS-53a6+SPC/E trials demonstrated strongly bound interfacial behavior (as depicted in **Figure 3.2b**) and defined by the criteria above, indicating this forcefield is highly sensitive for conducting simulations at the AWI. The GROMOS-53a6 forcefield was optimized to reproduce the hydration free enthalpy of amino acids in water.[152] A cumulative effect of these amino acid free enthalpies could be amplified in a protein model, explaining the strong interfacial adsorption at the AWI. With a99sb-disp+TIP4P-D the interfacial interactions were more dynamic and transient (**Figure 3.2a**). This model is a recently developed Amber forcefield and water model with a dispersion correction which has been shown to accurately reproduce experimental data for 21 different proteins (including both ordered and disordered protein and peptides).[153] To validate the choice of 100 ns long trials, and to verify dynamics at longer length scales, we extended several trials within each set of simulations that uniquely adsorbed to the interface to 500 ns. In **Figure B-S1** we demonstrate that at long time-scales adsorption trends remained consistent with trends observed in the first 100ns.

A vector along the protein was defined and then used to track orientation (with respect to the z-axis) for each trial (**Figure B-S2**). The angle between the protein vector and the z-axis was transformed using cosine into a discrete space to monitor the orientation of the protein over the course of the simulation (**Figure 3.2c and 3.2d**). Time-series data was then used to probe orientation and stability of this orientation at the interface, and a probability density analysis was used to identify trends across trials.

In the case of GROMOS-53A6+SPC/E, while more trials adsorbed to the interface across the set of ten simulations, the forcefield produced no self-consistent interfacial orientations, whereas trials from a99SB-disp+TIP4P-D distinctly adsorbed in two orientations (which will be referred to as the axial (I) and head-on (II) poses, respectively), as determined by the mean values of the orientation distributions plotted in **Figure 3.2c** and **3.2d**. The GROMOS-53A6 forcefield was parameterized against amino acid free energy of solvation, so these results could be interpreted as being overcompensated in a protein model, due to an amplified cumulative effect of these solvation energies at hydrophobic interfaces. The GROMOS-53A6/SPC/E trials that adsorbed, on average, did so over wider distributions than when compared to a99SB-disp/TIP4P-D, indicating high orientational flexibility at the interface. Alternatively, the a99SB-disp+TIP4P-D trials that adsorbed had stronger specific interfacial interactions given by strongly peaked distributions about the mean values. Structures corresponding to the mean orientations of the adsorbed proteins for both forcefields are given in **Figure 3.2e** and **Figure B-S3**. The orientations of the trials that did not adsorb were also analyzed in **Figure 3.2c** and **3.2d**. As expected, there was no significant conformational preference of the non-adsorbing trials, suggesting the “bulk” of our simulation environment is not affected by the presence of the interface.

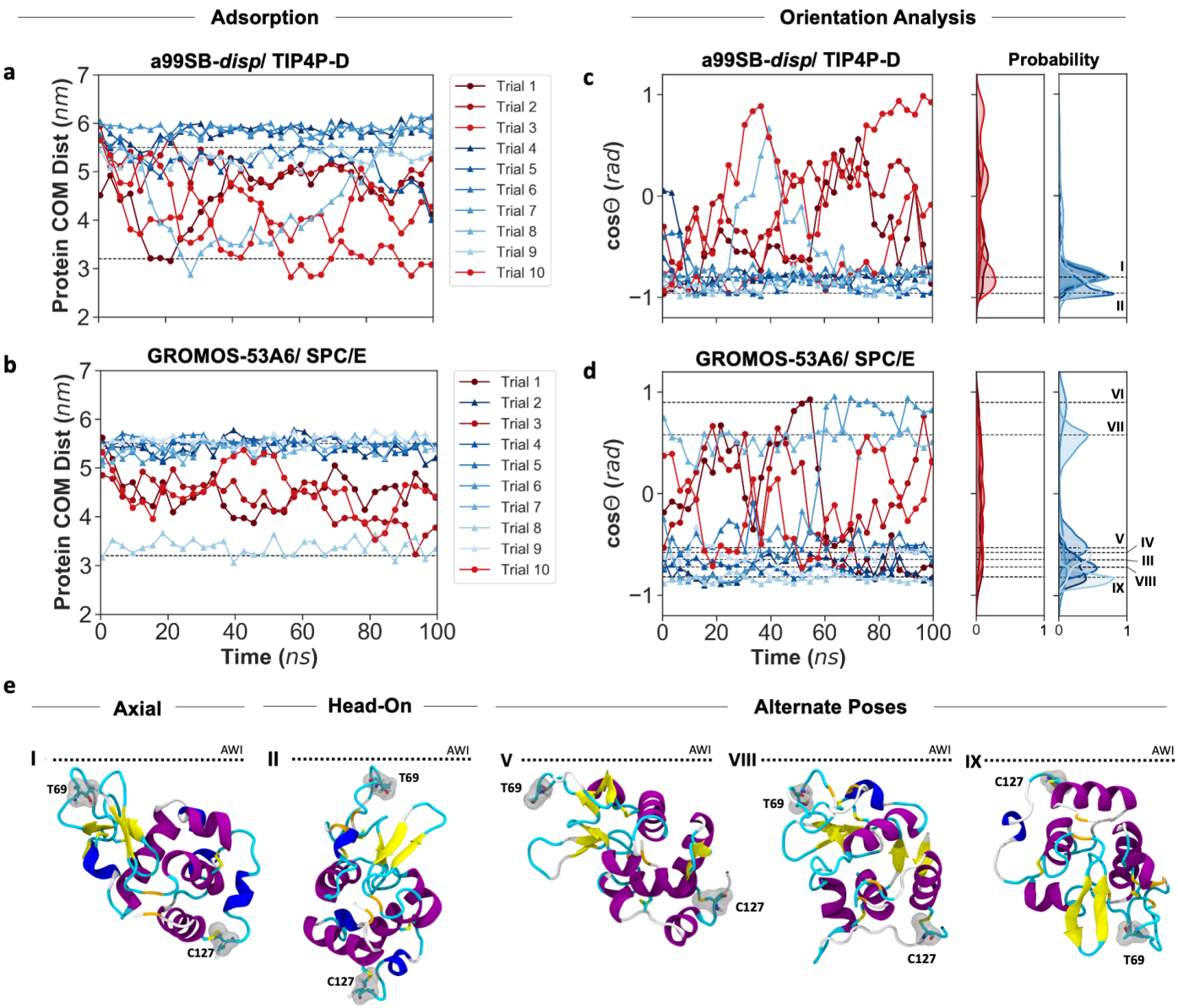


Figure 3.2. **Simulation predictions of lysozyme orientation at the AWI.** Time series data for all ten trials of simulations monitoring protein distance to the interface for (**a**) a99SB-*disp*+TIP4P-D and (**b**) GROMOS-53A6+SPC/E. Simulations that adsorbed to the interface (given at 5.5nm for the top interface, or 3.2 nm for the bottom interface, due to periodic boundaries) for at least 20ns are colored by blue triangles, and simulations that did not are colored in red circles, shading was used to differentiate data between different trials. Orientation analysis of each lysozyme trial measured over 100ns for (**c**) a99SB-*disp*+TIP4P-D and (**d**) GROMOS-53A6+SPC/E. Probability density plots of the 100ns are given for the trials that did not adsorb. Adsorption and orientation data were averaged every 1 ns and is plotted every 2.5 ns for clarity. Structures corresponding to unique adsorbed orientations are labeled **I** through **IX** and provided in (**e**) for (**I**) axial and (**II**) head-on orientations determined by a99SB-*disp*+TIP4P-D as well as exemplary poses (**V**) (**VIII**) and (**IX**) predicted by GROMOS-53A6+SPC/E. A dotted line is used to indicate the direction of the AWI, and waters were omitted for clarity. Additional structures corresponding to structures not shown for GROMOS-53A6+SPC/E can be found in **Figure B-S3**.

Characterization of the Lysozyme Binding Domain

In a99SB-disp+TIP4P-D, trials adsorbed to an axial or head on orientation, but both of these poses are dominated by the same coil and beta regions interacting at the interface, which we define in **Figure 3.3** as the primary binding domain. Interestingly, Proline 70 and Glycine 71 located in this primary domain, have both been previously reported as facing the AWI by previous reports from Perriman et al.[103] Both poses captured by this forcefield are also characterized as being adsorbed with beta-content facing the direction of the interface, which supports a number of studies that have previously reported the presence of beta-sheets at the AWI.[107–109] There is some transience between the two head-on and axial poses observed over the course of 100 ns (quantified by the slight shoulders present in the distribution of trial 9), indicating weak molecular interactions assist in binding of a secondary domain leading to a head-on to axial shift at the interface. Here, we define a secondary binding domain in F**igure 3.3,** located in the helical region facing towards the interface. This domain contains charged and hydrophobic residues, which likely results in competing interactions driving an axial pose at the interface. The interaction of this secondary binding domain with the interface, exposes the hydrophobic core of the protein to the interface providing additional stability, and a driving force for assembly. The presence of positively charged residues (arginine and lysine), sensitive to deprotonation at high pH indicates some tunability of the charge in the secondary binding domain, alluding to the ability to retain some control over an axial to side-on orientational shift at the interface which is explored later in this work.

Each GROMOS-53A6+SPC/E trial adsorbed in a unique conformation (**Figure 3.2d, 3.2e** and **Figure B-S3)**. Pose III and IV both adsorbed through the same primary and secondary domains as described in the a99SB-*disp*+TIP4P-D trials, indicating some ability to capture consistent interfacial poses between forcefields. Pose V and VI were bound to the interface through the primary binding domain but were oriented with the second binding domain facing opposite the interface. Lastly, poses VII, VIII, and IX were bound through regions outside of the domains determined in the a99SB-*disp*+TIP4P-D trials, and markedly bind with the primary binding domain (beta-containing domain) facing away from the interface.

To monitor the interfacial stability (in terms of long-scale dynamics), and backbone stability (unfolding) of these adsorbed poses, we looked again at the trials extended out to 500 ns. A root means square deviation (RMSD) of the α-carbons was used to show insignificant unfolding at the interface compared to the crystal structure, predictably due to stability provided from the

four disulfide bonds (**Figure B-S4**). This indicates little change in the secondary structure of lysozyme over long time scales. Orientational dynamics over 500 ns showed trends that persisted from the first 100 ns. GROMOS-53A6+SPC/E trials underwent large fluctuations in interfacial orientation while remaining adsorbed over 500 ns, indicating no preference towards a single binding pose. In the a99SB-*disp* trials the axial orientation demonstrated the most interfacial stability (trials 4, 6 and 8, **Figure B-S5**) remaining adsorbed for over 400 ns in one case, and the head-on orientation showed some transience at the interface (trial 9, **Figure B-S5**), remaining adsorbed for 120 ns.

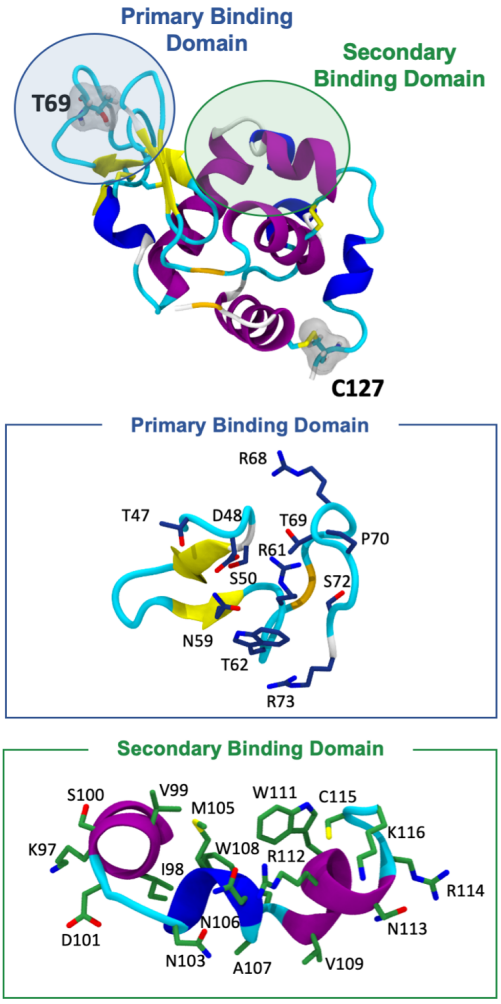


Figure 3.3. Atomistic description of the primary binding domains involved in lysozyme binding at the AWI. Using the axial pose as reference, the primary (blue) and secondary (green) domains are highlighted. A top-down view of each binding domain is shown below the protein to display the residues in each binding domain involved in binding to the interface. Hydrogens and backbone carbon and nitrogen atoms are omitted from this representation (and therefore glycine residues), for clarity. All other atoms are presented in a licorice representation and colored in red for oxygen, blue for nitrogen, and yellow for cystine, and carbons are colored to the corresponding binding domain colors.

Overall, consistency across the a99SB-*disp*+TIP4P-D trials and at long time scales indicated that lysozyme is likely interacting with the air-water interface in a head-on or axial pose through primary and secondary binding domains that place the beta-content near the interface, and expose the hydrophobic core to the AWI. This is supported through some consistency across forcefields where similar binding domains were observed in two the adsorbed trials from GROMOS-53A6+SPC/E simulations.

SFG and Spectral Calculations

To validate our prediction of the head-on or axial orientation at the interface, we used vibrational SFG spectra calculations to compare these results to the experimentally determined SFG signal, and used additional structures predicted from GROMOS-53A6+SPC/E as controls. Experimental SFG spectra of lysozyme at the AWI in SSP and PPP (s-polarized SFG and p-polarized infrared) polarization are displayed **in Figure 3.4**. The spectra show broad resonances between 1640 cm^{-1} and 1680 cm^{-1}, as expected for a convoluted signal that has contributions from the β-sheet, α-helix and turn motifs within the lysozyme secondary structure. Since the SFG selection rules dictate that only ordered species at interfaces are detected, the observation of SFG signal alone confirms that lysozyme binds the AWI and has a preferential orientation in its bound state.

Chen et al. have developed strategies to determine the orientation of proteins at interfaces using SFG peak ratios for β-sheet and α-helix-related modes within the spectra[165,166] In the present case such an approach to determine the binding pose of lysozyme would lead to significant uncertainties because of the strongly overlapping and interfering modes observed within the spectra. There is also a considerable amount of undefined secondary structure present in the lysozyme structure, which can also contribute to the SFG spectra but would remain unaccounted for in a direct analysis of the experimental spectra. Calculating theoretical SFG spectra from the MD trajectories directly is not limited by the accuracy of spectral fitting and peak assignment, makes full use of the complexity of SFG data and is therefore a very direct way to validate the simulation results. In addition, by averaging over the MD trajectories, the SFG calculations can include the dynamics and structural variations present in the adsorbed state of lysozyme. **Figure 3.4d** contains a residual sum of squares (RSS) estimate for the match of experimental and theoretical SSP and PPP spectra for each of the trials that demonstrated adsorption to the interface, where a low RSS value indicates the best match to experimental data. Overall, the structures that adsorbed to pose I and II from a99SB-*disp*+TIP4P-D resulted in excellent agreement with the experimental data and gave the lowest RSS values, as expected. Both models resulted in similar interfacial structures bound by similar domains. The best fit SFG calculation to the experimental signal is given in **Figure 3.4a,** for the lowest RSS trial adsorbed to an axial-bound pose with a mean RSS of 0.919 +/- 0.012 (across the three trials that were classified as adsorbing to an axial pose). The second-best overall RSS fit was for the head-on pose (II), as determined across multiple

trials (**Figure 3.4b)**. The head-on pose gave a mean RSS 0.971 with higher deviation (0.110) between the two calculated spectra compared to the axial pose which gave similar estimates across three MD trials.

The RSS values for III and IV from GROMOS-53A6+SPC/E were similarly low, as expected since both of these trials were bound by the same two binding domains as determined from the a99SB-*disp*+TIP4P-D trials. Poses V and VI, which were bound by the primary binding domain only, but found the secondary binding domain to be facing the interface was captured in having the second-best SFG signal as judged by the RSS value. And the other GROMOS-53A6-determined poses (pose VII, VIII, and IX) which contained no similar binding domains and captured the β-sheets to be oriented away from the interface, resulted in the highest RSS values. An SFG calculation for pose IX (the control pose) is also included in **Figure 3.4c**. Pose IX was the single trial that adsorbed in an orientation completely opposite the other predicted orientations, indicating it is unlikely that lysozyme binds to the interface through this mechanism.

The fact that two models resulted in a good agreement with experiment and that both models predict very similar binding poses underlines the robustness of the general approach. It should also be noted that the models used only simulate single proteins at the interface. In the dense protein films used in the experiment there is a balance of protein-protein as well as protein-surface interactions, which can impact protein orientation and even folding. The high degree to which the models capture the experimental spectra indicates that these effects do not play a significant role for the behavior of lysozyme and that, within the accuracy of the theoretical models used, simulating single proteins interacting with the surface describes the interfacial processes well.

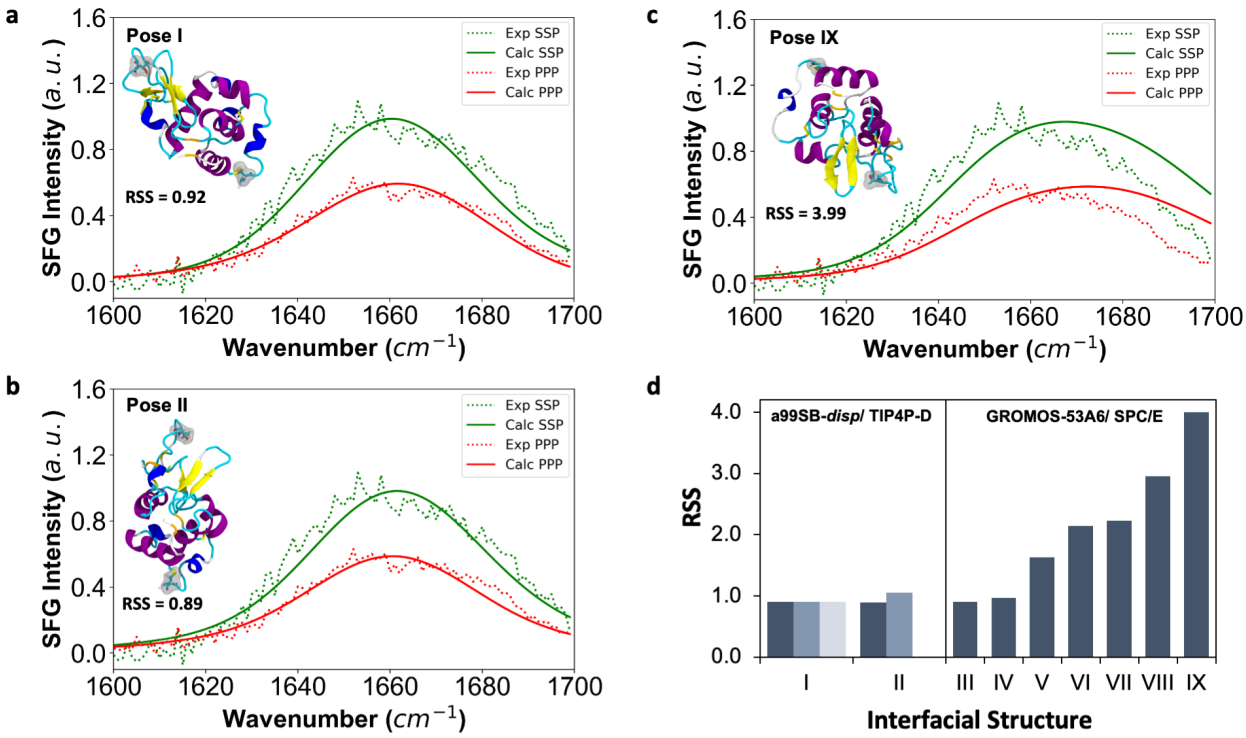


Figure 3.4. Experimental SFG signals for the best and worst fits determined by RSS between experimental and calculated spectra. SFG intensity is given for (**a**) a99SB-disp+TIP4D trial 4 which adsorbs in Pose I (axial) **(b)** a99SB-disp+TIP4D trial 5 which adsorbs in Pose II (head-on), and (**c**) GROMOS-53A6+SPC/E trial 8 which adsorbs in Pose IX (axial). The experimental data is given by a green dotted line for SSP and a red dotted line for PPP, the spectral calculations of simulation structures are provided in solid lines. Interfacial structures corresponding to each trial, and their respective RSS values are provided in plot for clarity. (**d**) RSS comparing the normalized SSP and PPP of each pose determined from simulation to the experimental signals for every interfacial structure determined from MD. In cases where multiple simulations adsorbed to the same interfacial poses, additional shaded bars are provided. Additional spectra are included in **Figure B-S7** and **Figure B-S8**.

Sensitivity to pH

A number of experimental studies have reported orientation of the bound protein in its folded state, however there has still not been agreement over whether the protein retains head-on and side-on orientation.[101–103,105,114] In this study we find a head-on or axial pose to be the most likely conformation at pH 7, likely due to the chemistry of the secondary binding domain which contains hydrophobic and charged residues. Lysozyme has a uniquely high pI of 11.35. At high pH, we expect the positively charged residues to deprotonate, reducing their polar (or hydrophilic) nature, and near the pI we expect all lysine (with a sidechain pKa of 10.55) residues to become deprotonated. Lysozyme contains 6 lysine residues, all of which are solvent-exposed, and two of which are contained in the secondary binding domain defined in **Figure 3.3 and 3.5**, suggesting these are influential in the pH-dependent surface behavior observed in lysozyme.

To probe the effects of pH, the six adsorbed a99SB-*disp*+TIP4D trials were simulated at pH 11 by deprotonating the lysine residues (see **SI (Appendix B)** for additional simulation details). Of the six trials, four of these simulations remained adsorbed to the interface over 100ns. Of those 5 trials all proteins were in a slightly shifted axial pose which frequently bind in a side-on conformation over the course of 100ns (as demonstrated by the wider distributions biased away from the peak A mean distribution in **Figure 3.5c**), and none of the proteins remained adsorbed over 100 ns in a head-on pose, indicating some instability of this orientation at high pH (**Figure B-S6**). A depiction of the shifted axial pose is shown in **Figure 3.5d**, where the deprotonated lysine (K*) residues are depicted showing assisted binding behavior at the interface, further exposing the hydrophobic core. The additional binding of this secondary domain validates some of the early mechanisms proposed by Lu et al.[114] and Hunter et al.[102] Additionally, Thakur et. al[167] also previously proposed pH 11 produced optimal conditions for classic Langmuir type adsorption. However, these mechanistic changes have also been attributed to concentration effects.[101,115] In these simulations we can ascribe these changes in assembly to higher driving forces for side-on assembly at the interface, versus nearly equal propensity driving the assembly of two protein orientations seen at pH 7.

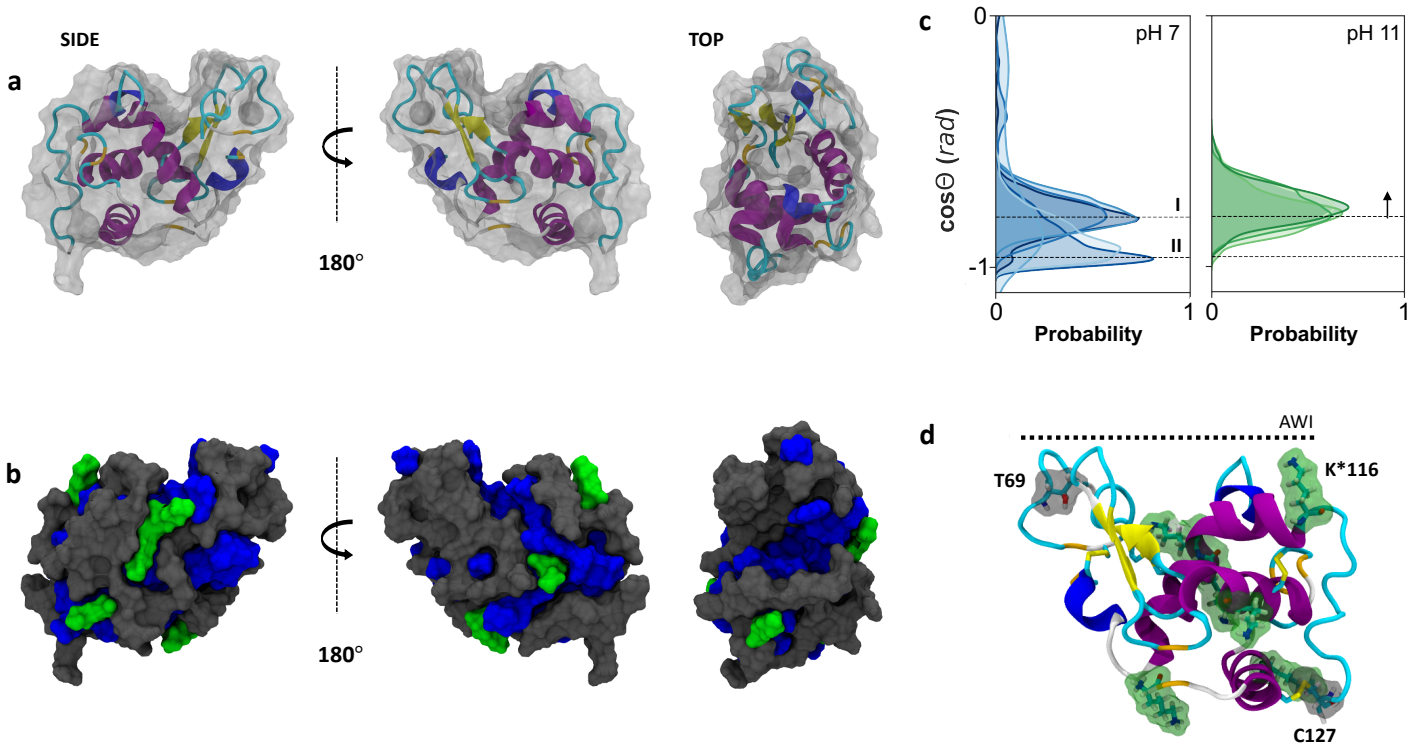


Figure 3.5. Orientation shifts due to changes in pH. (**a**) reference schematic for images in (**b**) which shows the exposed hydrophobic residues in blue (Ala, Val, Leu, Ile, Met, Phe, Trp, Pro), and lysine (Lys) residues in green from a side-on view (**left**) and a top down view (**right**). (**c**) On the left are the original pH 7 simulations which adsorb in a distinct axial (**I**) and head-on (**II**) conformation (replotted here for clarity), and the pH 11 simulations are given on the right with an arrow indicating a shift from the axial pose at pH 7 into a side-on conformation (**d**) structure for shifted axial/side-on pose found in pH 11 simulations following previous coloring schemes, deprotonated lysine (K*) are also demonstrated in green for clarity.

3.5 Discussion

Developing and validating approaches to determine the conformation and orientation of proteins at interfaces is critical for uncovering the function of proteins and still remains a challenge today. The interfacial structure of even one of the most widely studied proteins in the literature, lysozyme, has been the topic of a long-standing discussion in the literature since the early 90s, and yet its structure at the air-water interface had yet to be resolved without dispute.

Through a cross-validated SFG-MD approach, we found lysozyme to adopt a head-on or axial conformation at the interface with the beta-content facing the interface consistent between simulation and experiment. This was supported by the SFG calculations showing strongest agreement with simulation structures that had both the primary and secondary domains facing the interface, some agreement when the primary domain was bound to the interface, and the least

agreement for oppositely oriented structures. This is partially consistent with previous studies which have reported beta-content at the interface, but in our simulations it is never observed that this occurs to the extent of the protein restructuring at the interface, but instead we observe a very stable backbone, consistent with lysozyme being characterized as highly stable.[105,108,112] Lastly, we provide atomistic detail which can help explain how changes in pH could promote side-on adsorption. By paying close attention to the chemistry of the solvent-exposed residues we find that changes in pH near the pI of lysozyme result in the deprotonation of positively charged (hydrophilic) residues, which influence its interfacial orientation, supporting a number of previous proposed mechanisms of AWI adsorption.

Through SFG calculations we were able to show that the a99SB-*disp*+TIP4D forcefield can predict interfacial structure of proteins without a clear hydrophobic binding domain with high consistency compared to experiment. While the GROMOS-53A6 forcefield was sometimes able to capture the correct interfacial structure, a lack of self-consistency within the orientation indicates some caution should be used when applying this forcefield to study interfacial structure in the absence of additional experimental or simulation data. With the additional simulation results and experimental data, we show correlation to some structures obtained from the GROMOS-53A6+SPC/E simulations, showing some consistency across forcefields.

3.6 Conclusion

Here we have presented the structure and orientation of lysozyme at the AWI, with high consistency between both experiment and simulation obtained structures, through the use of spectroscopy and spectral calculations. We have also presented a streamlined approach from a simulation perspective for evaluating the structure of complex proteins at interfaces and identified a forcefield sensitive to interfacial simulations. With atomistic-level detail about the binding domains of lysozyme we are able to make mechanistic predictions about the adsorption and equilibrium behavior of this system and provide new information to help resolve long-standing discrepancies among prior studies. Future work stemming from this study could focus on applying simulation techniques to further explore the effects of pH and concentration, providing additional mechanistic insight into different adsorption models that have been proposed for lysozyme and other proteins. Lysozyme contains a number of arginine residues on the solvent-exposed surface of the protein, which would predictably result in additional mechanistic changes at higher pH.

Similarly, studying the influence of disulfide bonds and protein stability on the adsorbed orientation and conformation is another interesting direction. A system where disulfide bonds have been reduced would greatly impact the proteins flexibility and therefore, which regions are exposed and demonstrate an affinity to the AWI. A greater understanding of these assembly mechanisms would result in high control over assembly at the AWI, and in their application. Advancement of simulation techniques to probe both the thermodynamics and kinetics of unfolding mechanisms at the interface would also be critical in understanding the behavior of not only lysozyme but a number of protein behaviors. This might involve the application of enhanced sampling methods to probe these more complex processes.

3.7 Acknowledgements

The research design and simulations in this project were supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, as part of the Energy Frontier Research Centers program: CSSAS--The Center for the Science of Synthesis Across Scales under Award Number DE-SC0019288. TW acknowledges funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 819039 F-BioIce) in support of the experimental measurements and theoretical SFG analyses in this work. SA was supported by a National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1762114 to support this research. SJR was supported by the Lundbeck Foundation for a postdoctoral fellowship. The authors gratefully acknowledge Stefano Piana-Agostinetti for supplying the amber forcefield files used in this study.

Chapter 4
Closing the Gap Between Modeling and Experiments in the Self-assembly of Biomolecules at Interfaces and in Solution [3]

4.1 Abstract

Molecular self-assembly is a powerful tool in materials design, wherein non-covalent interactions like electrostatic, hydrophobic, hydrogen bonding, and van der Waals can be exploited to produce supramolecular nanostructures that are functional and highly tunable. Biomolecules are attractive building blocks, as they are biocompatible, biodegradable and adopt a wide array of higher order structures. Moreover, naturally occurring protein systems display a manifold of structures and interactions that can be replicated in synthetic biomolecules. In this perspective, we highlight advances in multiscale simulation techniques across broad spatiotemporal scales that can aid in characterizing self-assembly of hybrid and hierarchical bionanomaterial systems, with an emphasis on physics-based simulation approaches currently employed to study biomolecules at mineral interfaces. The power of these approaches is highlighted across a few recent areas where molecular simulations have advanced our understanding of self-assembly spanning peptides to protein self-assembly. Looking forward, we discuss how in the near future emerging methods in statistical and machine learning will advance this research field in all areas from expanding the capabilities of physics-based simulation methods to enabling new analyses of high throughput experiments. These advances will pave the way for understanding the molecular recognition patterns in systems that are dictated by self-assembly - biomineralizing peptides, hierarchical peptoids, and large protein assemblies, and will aid in the development of a new synthesis science for achieving precise molecular control in materials design.

[3] This chapter was reproduced with permission from J. Sampath, S. Alamdari, and J. Pfaendtner. losing the Gap Between Modeling and Experiments in the Self-Assembly of Biomolecules at Interfaces and in Solution. Chemistry of Materials, 32, 19, 8043–8059 (2020). Copyright 2020 American Chemical Society.[6]

4.2 Introduction

Hierarchical self-assembly, the phenomenon by which a system's components organize into higher ordered, functional structures as a result of specific local interactions, is ubiquitous in nature and can be harnessed in the laboratory to produce materials with superior properties and specific functionalities.[168–173] Naturally occurring biomolecules have information encoded in their chemistries that enables them to spontaneously organize into complex structures *in vivo*.[174–176] While some of these structures have been mimicked *in vitro*, by identifying and exploiting key interactions between units that self-assemble in biomolecular systems, we can predictively tune and control the assembly pathway to obtain a final, desirable structure.

Two naturally occurring, self-assembling systems that have been studied extensively are proteins and DNA.[177–180] The assembly of proteins plays a vital role in regulating bodily functions. For instance, misfolded proteins can trigger abnormal aggregation such as the amyloid beta protein into plaques that are prevalent in patients suffering from Alzheiemer's disease.[181–184] Another example is the aggregation of globular hemoglobin protein that distorts red blood cells causing hemophilia.[185–187] Proteins such as capsids that house viral RNA self-assemble within host cell and have a high degree of stability.[188] Relatedly, proteins like osteocalcin and silaffin interact with inorganic interfaces, to template and modulate calcite and silica growth.[189,190] Proteins have inspired countless synthetic self-assembling systems, a prominent example are the biomimetic sequence controlled foldamers known as peptoids, which possess tunable properties that give rise to a wide range of architectures.[191,192]While protein assembly mechanisms are rather complex and varied, DNA assemblies characterized by few primary interactions are highly specific, mediated by hydrogen bonding between two base pairs.[177,193–197] This has made it possible to produce 2D and 3D DNA nanoarchitectures with a high degree of stability, as the interactions are well characterized and the molecule is rigid, enabling a convenient route for predictive design. Unlike DNA, a delicate balance between different non-covalent interactions, such as electrostatic, hydrophobic, hydrogen bonding, and van der Waals gives rise to the final, self-assembled structure in proteins and protein-mimetic materials, which are the primary focus of this review. We believe that presenting these systems together will enable a unified framework for analysis, as findings and observations can be directly translated from one system to another.

Experimentally, biomolecular assembly can be studied using a range of techniques, including fluorescence and optical microscopy,[198] electron microscopy (scanning and

transmission),[199] Nuclear Magnetic Resonance (NMR),[200] X-Ray Diffraction (XRD),[201] and *in situ* Atomic Force Microscopy (AFM).[202] Surface Plasmon Resonance (SPR) can be used to model assembly kinetics on surfaces.[203] However, none of these techniques offer a complete picture of the atomic interactions and molecular driving forces that ultimately give rise to macroscopic structures. Molecular simulations can be utilized as a 'computational microscope' that provides the insight required to augment experiment and advance the field towards predictive synthesis of hierarchical soft materials.[204] In the context of the important applications noted above, this perspective aims to shed light on the methods and applications of molecular simulations that have advanced our understanding of interfacial phenomena of biomolecules. We also discuss opportunities for expanding these studies to more complex systems, with the ultimate goal of making computer simulations of these systems a robust predictive tool in the design of biomolecular self-assembly.

Many current approaches to biomolecular self-assembly are built on extensive past theoretical and experimental studies of rigid colloidal assemblies that produce well-ordered nano structures with huge diversity of self-assembled motifs across a range of length scales.[205,206] This was due to two contributing factors, the first was an unprecedented advance in particle synthesis methods that gave rise to a large variety of building blocks with exotic shapes and functionalities. The second was the ability of statistical mechanical methods to accurately estimate assembly phase space, enabling predictive design of self-assembled colloidal building blocks. Unlike colloids, small biomolecules (approximately 50 residues in size) and their analogs are often highly flexible, and this makes characterizing their propensity to self-assemble challenging. Larger globular proteins (hundreds of residues in length) possess a stable tertiary structure, however, these assemblies are governed by the association of side chains that experience large structural fluctuations. In both extremes, backbone and sidechain flexibility is a key determinant of the final structure. The driving forces which govern self-assembly of biological soft materials can be described using a common physical framework. This framework motivates the current perspective to use advanced computer simulations in tandem with experiments to formulate design rules to control and direct self-assembly of biomolecules, informed by well characterized systems such as micelles and colloids.

In the remainder of this perspective, we first highlight the simulation methods that have been applied to study self-assembly. A number of methods can be applied to tackle these complex

questions, but here we focus on molecular dynamics (MD) simulations and the advanced sampling technique known as metadynamics as an approach to estimating equilibrium structures and thermodynamics. This subset of methods when coupled with experimental observations, has previously been widely used to capture phenomena that spans across broad length scales in the field of biomineralization. Then, building off inspiration from progress made in biomineralizing systems, we present three types of biomolecular systems whose functionalities are dependent on the assembly of individual units – biomineralizing peptides, hierarchical peptoids, and large-scale protein assemblies, which share common themes driving self-assembly in this interesting milieu. We discuss significant advances that have been made in understanding the structure-function relationships in these systems, as well as open questions that remain in these areas, while focusing on the next generation of modeling and simulation studies needed in order to drive the field forward. These simulation methods have the potential to inform and guide experimental design, shifting the paradigm from trial and error based synthesis to predictive, *in silico* predictive design and optimization. We also highlight studies where simulations and experiments deviate considerably from each other and, attempt to provide a direction for these differences. We conclude by discussing how simulations and the expanding role of data science methods will help propel molecular modeling of biomolecular self-assembly in these important research areas.

4.3 Simulations of Biomolecules at Interfaces: Progress in Biomineralization

The span of time and length scales accessible using simulations is wide, ranging from the interactions between few atoms (governed by phenomena like covalent bonding or nonbonded interactions), to the slow conformational rearrangement of glassy systems.[207–211] We would like to note that it is beyond the scope of this review to describe all the computational approaches which could be applied to study self-assembly, therefore we refer to the reader to more general review articles about computational studies of self-assembly in both biomolecular and non-biomolecular contexts.[205,212–215] One major challenge of applying simulations to resolve these assemblies is in the development and use of methods that are widely applicable and predictive of the unique structure-property relationships in flexible self-assembling biomolecular systems. Overcoming this obstacle requires careful consideration of the model systems studied (including system size, and chemistries), ensuring the potentials used in the simulation can capture the underlying physics

at least at a semi-quantitative level, and the proper selection and implementation of enhanced sampling methods that correlate with the spatiotemporal scale of interest.

Biomineralization is one area that has been studied extensively with multi-scale models.[216,217] Starting with the most accurate, and the most computationally expensive, density functional approaches apply a quantum mechanical (QM) description of energy to the system to gain information about electronic structure. Due to the expensive nature of these calculations QM methods are often applied to understand small model systems (on the order of 100-200 atoms). In the context of biomineralization, QM approaches have mainly been employed to study the behavior of bulk minerals in vacuo, the interactions between crystalline surfaces and water or ions, or to validate the behavior of forcefields used to study complex interactions in biomineralization.[218–220] DFT methods can also be used to understand the interplay between part of a mineral surface acting with part of a binding site on a protein and a few surrounding waters.[221]

Classical potentials are a popular choice when studying the behavior of more than a handful of atoms, or when looking to answer more physiologically complex problems like the influence of ions or solvent conditions on biomolecular structure. All-atom molecular dynamics simulations implement classical potentials (forcefields) built from QM calculations or fit to experimental data, integrating out information about electronic structure, to inform dynamics as a function of the bonded and nonbonded interactions at the atomistic level.[222] CHARMM and AMBER are two of the more popular forcefields used for studying biomolecules, which have been developed for use with specific water models (TIP3P[223] and SPC[224], respectively). Consistency between these popular forcefields and the surface models developed for organic and inorganic components is critical. One important factor (aside from compatibility with the biomolecules), is in capturing the unique behavior of water which is known to behave differently at mineral apatite or silicate interfaces. In classical potentials, like the INTERFACE forcefield, water dissociation is treated rigidly where hydroxylated sites are pre-defined before running a simulation.[77] Within the family of INTERFACE models, Emami el. al. introduced a surface model to capture a range of silica surface chemistries and pH conditions (e.g. variable Q2, Q3, and Q4 silanol groups) compatible with CHARMM.[225] Later, Lin. et al. developed a hydroxyapatite model, also compatible with CHARMM, that accounts for changes in pH by modeling different ratios of hydroxylated species at the interface.[58] These are just a few examples of how classical potentials capture complex behavior of water and ionic strength, but there are a number of potentials parameterized for other

mineralizing chemistries (i.e. metal oxides like titania) outside the scope of this work which have also been generalized to be compatible with commonly used forcefields.[69,212,226–228]

In the context of biomineralization and self-assembly, classical simulations are useful in describing at high atomistic resolution the stability of proposed conformations and structures that can be correlated with mechanisms of assembly. These types of simulations probe stability by monitoring properties calculable from thermal fluctuations, or by identifying dominant Coulombic and Lennard-Jones interactions. This can be applied while probing candidate binding poses at interfaces, or the disassembly of pre-assembled structures to name a few examples. While classical methods are popular for probing hypotheses surrounding stability and structure, these methods fall short when the time and length scales to sample these dynamic interactions (e.g., modeling the process of assembly) are larger than computationally accessible. This is typically in the range of running calculations for two to four biomolecules of chain lengths up to fifty residues. However, the calculation of many important thermodynamic quantities, such as binding free energy, requires sampling of biomolecular structure and rare events (such as binding and unbinding) which is often inaccessible on classical timescales. The introduction of a solid or biomolecular interface to model different aspects of assembly further exacerbates these problems due to an increase in a complexity of the thermodynamic landscapes, and system size.

Enhanced sampling methods are one approach that can be used to circumvent these limitations. The possible methods that can be used to circumvent time and length scales are too numerous to name here. For example, these methods include replica exchange, simulated annealing, log mean force dynamics, nudged elastic band, and blue-moon sampling, to name a few. These approaches and many others have also previously been applied to study biomolecules and have been discussed extensively.[9,229,230] One popular suite of enhanced sampling methods is known as metadynamics (MetaD). This enhanced sampling in particular has been used to study mechanisms of self-assembly and interfacial phenomena, making significant strides in the field of biomineralization.

Briefly, in metadynamics a bias in the form of a gaussian potential is applied to a few (1-3) slow degrees of freedom (defined as collective variables, or CV's), and used to explore the thermodynamic landscapes of these systems.[231] In some cases, hidden slow degrees of freedom can limit the scope of using metadynamics. In these cases, the addition of more robust sampling methods can aid in the rapid convergence of system energetics. There are reviews dedicated to

describing MetaD and its variants in detail, herein we highlight how select variants of MetaD have been developed to enable investigations of model systems related to biomineralization.[9,232] Parallel tempering metadynamics (PT-MetaD) found great use in overcoming hidden slow degrees of freedom not captured in the CV's definition.[84] However the challenges of system scaling prevented wide application to more realistic systems at the bio/nano interface. Application of the well-tempered ensemble (WTE) to biomolecular systems extended the use of these methods to many studies of peptide adsorption and folding at interfaces.[70,71,88,233] An additional approach is to improve the number of CVs that can be sampled. Traditional MetaD uses 2-3 CVs for soft-matter systems with severe convergence issues introduced beyond this number. Building on bias exchange MetaD[234], the parallel bias metadynamics (PBMetaD) framework introduces the ability for arbitrary numbers of new CVs.[235] Building on this concept with an eye toward self-assembly, a new variant of this, PBMetaD with partitioned families (PBMetaDPF), demonstrated increased efficiency and convergence when degenerate CVs are partitioned to share the same bias potential.[236]

Several approaches outside of enhanced sampling have also been used to address simulating these large all-atom systems. Coarse-grained (CG) models reduce the degrees of freedom in a system, providing a more efficient route for probing longer time and length scales.[237] Many find the popular MARTINI[238] model a convenient starting place for CG modeling in biomolecular and bio-inspired systems. Parameterization of new molecules can be accomplished with all-atom MD, an approach previously used to study self-assembly. [239–241] A popular statistical method known as Markov state models (MSM) is an example of an unbiased approach which applies a stochastic method used to describe long-scale dynamics and kinetics of a system.[242] In the MSM scheme, a large number of short simulations taken from a large conformational space are used to probe the relevant transitions which can be used to understand the dynamics that give rise to aggregation mechanisms, or kinetics of assembly. More recently, researchers are finding innovative ways to overcome systematic error in physics-based potentials through inclusion of experimental data in simulations. The scope of this area of research is large and has been described by Bonomi et al. in detail.[243] In the spirit of developing ways to improve molecular simulations, MetaD was combined with a Bayesian inference method known as Metadynamics Metainference (M&M), allowing for a natural way to bias a collection of simulations toward agreement with ensemble-averaged experimental data.[244] Combining this approach with additional MetaD biasing

of conformational CVs provides an additional capacity to fully explore conformational of biomolecular assemblies at interfaces.[245]

Biomineralizing Peptides

Bottom up self-assembly enables biomolecules to precisely template mineral precursors to produce hierarchical nanostructures.[189,246–250] This is commonly seen in marine organisms, such as diatoms,[251] and is also observed in the human body, where proteins regulate the formation of tooth enamel and bones.[217] This process, termed as biomineralization, gives rise to inorganic materials with unique characteristics depending on the proteins that template them. For instance, avian egg proteins like osteopontin, lysozyme, and ovalbumin template calcium carbonate to give rise to a highly ordered, semipermeable membrane that protects the egg white against bacterial infection.[252,253] Enamel proteins such as amelogenin and amelotin, along with the salivary protein statherin are responsible for regulating the crystal growth of hydroxyapatite, giving it superior mechanical properties.[254,255] These examples highlight the idea that protein structure is intimately linked to the mechanism by which it recognizes mineral precursors in solution, which determines the morphology of the mineral structure that is formed. Importantly, biomineralizing proteins have stereochemical matching to crystal surfaces that can inhibit or mediate growth in a specific direction. Detailed atomistic level insight into the precise interactions of these proteins with minerals, specifically molecular recognition patterns is needed to control and regulate mineral formation. The understanding of these mechanisms can potentially be applied to treatments for diseases that arise from erroneous biomineralization, such as kidney stones, atherosclerosis, etc.[256] Simulations have helped achieve a comprehensive understanding of molecular interactions that govern the process of biomineralization.[90,212,257–259] Specifically, by using small fragments of protein sequences that have high affinities to mineral interfaces in tandem with complex sampling methods, simulations have successfully probed the thermodynamic and structural properties that drive biomineralization. Below, we highlight simulations that have advanced our understanding of biomineralization, specifically with respect to hydroxyapatite and silica, the two most common biominerals. We conclude this section with open questions pertaining to the role of aggregation of biomolecules on biomineralization, and how future simulations can be designed to address this.

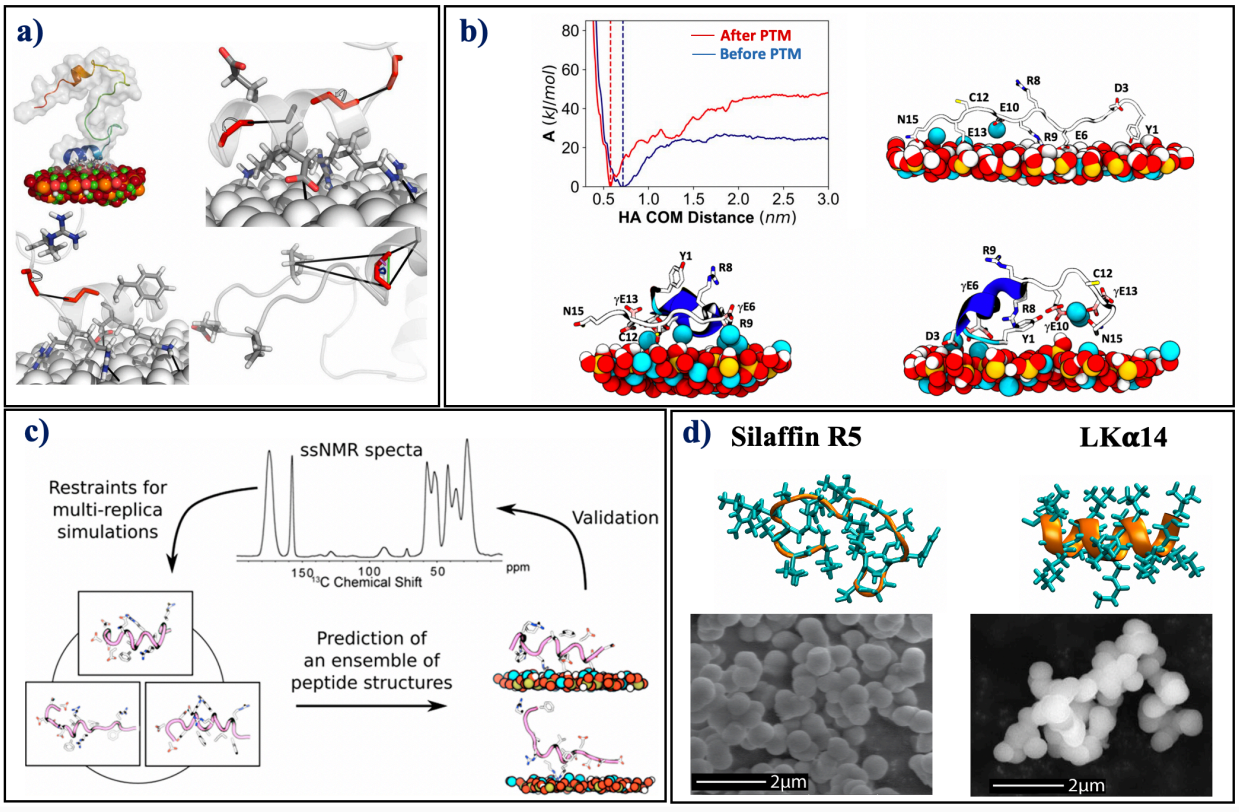


Figure 4.1. Biomineralizing peptide systems (**a**) Molecular structure of Statherin adsorbed to the HAP (001) surface, with the final structure shown on the colored interface; the amino acids that are known to interact with HAP are magnified and shown in licorice representation. Reprinted with permission from ref 260. Copyright Elsevier, 2010. (**b**) Binding free energy of osteocalcin α-1 domain on HAP interface with (red) and without (blue) post translational modification; snapshots pertain to the most probable interfacial peptide structures. Reprinted with permission from ref 259. Copyright Royal Society of Chemistry, 2019. (**c**) Schematic of the M&M workflow used to restrain the structure of N-terminal statherin peptide based on ssNMR chemical shifts. Reprinted with permission from ref 64. Copyright American Chemical Society, 2019 (**d**) Images of silica precipitated from silaffin R5 and LKα14 peptides. Reprinted with permission from refs 261 and 262. Copyright American Chemical Society, 2017 and 2014.

The Gray group extensively characterized the surface structure of the full salivary protein statherin on hydroxyapatite (HAP), shown in **Figure 4.1a**, using an approach that incorporated geometric restraints measured by solid state NMR (ssNMR) into the protein design suite Rosetta.[263] Experimental methods such as ssNMR that measure the structure of proteins at interfaces have high resolution, but often are limited in the total number measurements that can be taken (compared to requirements for structure prediction).[264,265] Therefore, there is an immense benefit from integration with high resolution atomistic MD simulation. NMR chemical shifts can be calculated with relative ease from classical simulations, providing a means of cross validating results between experiments and simulations.[266] Using experimental NMR shifts with

RosettaSurface to restrain the structure of proteins at mineral interfaces, Masica et al.[260] proposed structural models that showed remarkable consistency with prior studies, while greatly reducing the number of ssNMR measurements required to resolve the entire structure of the statherin.[263,267] Specifically, they found that while the protein is extended in solution, it adopts a structured conformation when it is adsorbed on the surface, with a helical binding domain at the C-terminal of the molecule.[260,268] Additionally, they probed the specificity of statherin on three HAP faces – 001, 100, and 010 and found that the 001 surface compliments the helical binding domain of statherin, proving that the biologically relevant HAP face is the 001 surface, in line with prior experimental outcomes. This combined computational-ssNMR approach greatly increased the resolution of residue-specific contributions of proteins to the process of biomineralization, ultimately accelerating the abilities of experiment and computation to inform these processes. Despite the success of this method in predicting the preferential binding residues for statherin to HAP *a priori*, the restraint of an individual structure to an experimental measurement may not fully reflect the ensemble-averaged nature of NMR measurements. Moreover, knowledge-based potentials like Rosetta do not incorporate the effects of explicit water and ions at the interface.

The M&M method was specifically developed to integrate experiments and simulations reflecting the need to guide the ensemble average of some simulation observables toward the experimental measurement.[245] By employing this technique, we were able to elucidate how SNa15, a peptide derived from the HAP recognition domain of the biomineralization protein statherin, interacts with HAP, silica, and titania interfaces. In contrast with traditional ssNMR approaches for predicting biomolecular conformation, we were able to use MD simulations + M&M to predict orientation and conformation of peptides at interfaces. Our work suggested that SNa15 remains helical at the interface of HAP and TiO_2 but unravels upon adsorption to SiO_2.[64] The use of enhanced sampling in these systems provides a powerful combination by giving simultaneous estimates of thermodynamics and interfacial structure in an all atom simulation.

While the above simulations have been very useful in elucidating preferred surface conformations of the full protein sequence, these methods do not allow for explicit characterization of the thermodynamic landscape, specifically in capturing the binding free energies of the protein adsorbing to the surface. Advanced sampling methods allow for complete sampling of phase space for systems that are often trapped in metastable states, resulting in full exploration of the free energy landscape. Using metadynamics, our group has characterized many solid binding proteins

that are known to promote biomineralization, including the effect of pH and post-translational modifications (PTM).[48,70,88,90,230,269,270] **Figure 4.1b** shows the binding free energy profile for the non-collagenous bone matrix protein osteocalcin on HAP before and after PTM of the sidechain glutamic acid group, using parallel tempering metadynamics in the well-tempered ensemble.[259] We find that post translational carboxylation of glutamic acid residues resulted in increased protein structure in solution and at the surface, mediated through strong intrapeptide contacts with ions. This structural stability led to differences in the thermodynamic driving forces on different biologically relevant materials like HAP and TiO_2, where the affinities between the surface and ions was shown to play a critical role in adsorption behavior. By exploring the molecular recognition mechanisms present in surface binding, these findings help shed light on how osteocalcin assembly templates mineralization.

Although the overarching goal of the above computational studies was to correlate the thermodynamics of peptide-surface interactions to the protein mediated formation of inorganic substrates, peptide-surface binding affinities may not be the primary indicator of biomineralization efficiency. For example, Patwardhan et al. suggested that peptides which have a strong affinity to silica perform poorly in early stages of silica precipitation.[271] To build a complete mechanistic understanding of biomineralization, characterizing peptide-peptide and peptide-precursor interactions is vital. To this end, the Drobny group has extensively characterized the intrinsically disordered peptide, silaffin R5, and the helical peptide LKα14, focusing on their aggregation in solution.[261,262,265,272,273] They found that although the two peptides have extremely different structures in solution, their ability to precipitate silica with remarkably similar morphologies suggests that the properties of the peptide aggregate could play a large role in how biomolecules template minerals. In a prior study, we observed that LKα14 molecule binds less strongly to a silica surface than R5, signifying that surface binding alone does not predict biomineralization.[270] Future simulations comparing the aggregation states of R5 and LKα14, using experimental structures for verification are required to quantify interpeptide driving forces that control protein assembly, which could be a mechanism that templates biomineralization. It is also theorized that surface properties of aggregated assemblies are predictive of mineralization. However, it is difficult to fully explore the conformational landscape of multiple proteins binding in atomistic simulations due to convergence issues described in the Simulations section. One way to address this is to extend the simulation-NMR technique that was applied to protein-surface assemblies to

capture protein-protein assemblies in solution. Chemical shifts of the aggregate can be used to bias the collective structure of the proteins in solution, speeding up the simulation process. Understanding how proteins aggregate can help build a more complete picture of protein-precursor interactions, before the onset of biomineralization. These interactions obtained from simulations can be exploited experimentally to give rise to nanostructures with desired morphology.

Cooperative Surface Binding

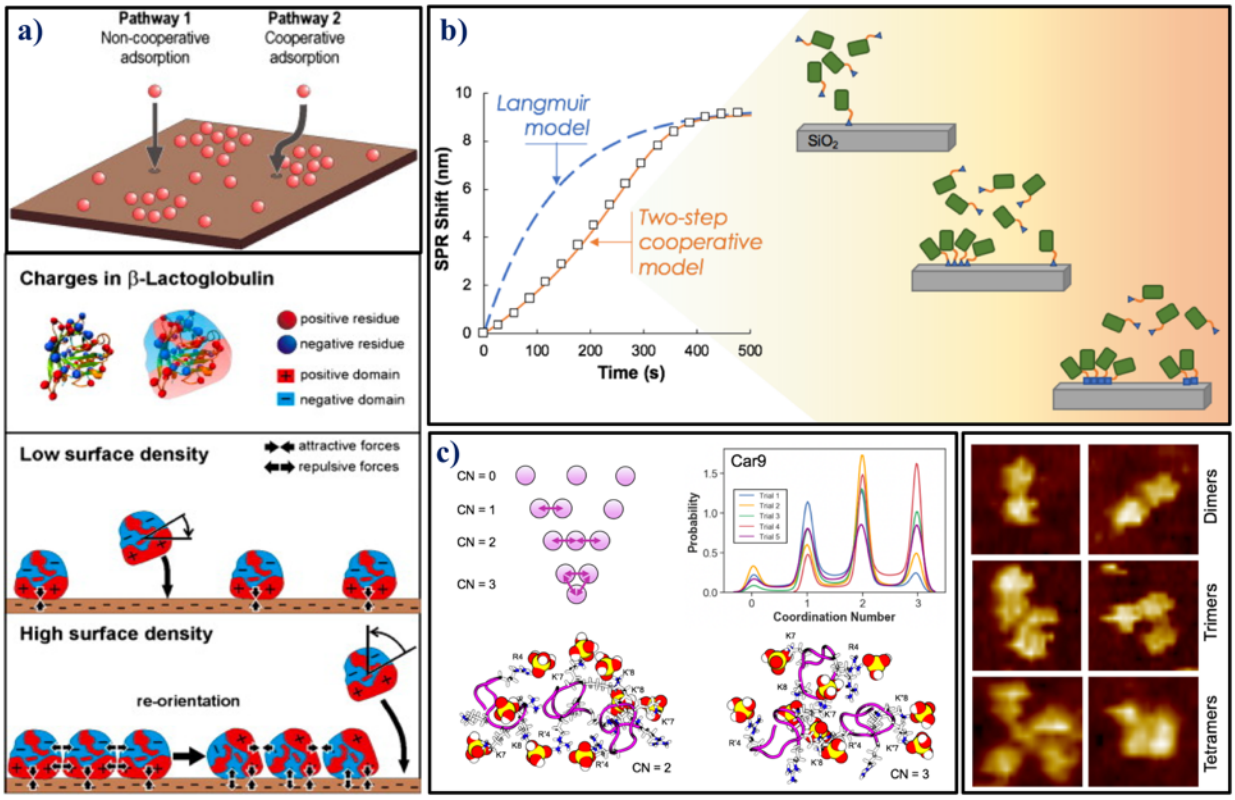


Figure 4.2. Cooperative protein – surface binding (**a**) Schematic representation of non-cooperative vs. cooperative adsorption pathways as a function of surface density; at low surface densities, protein orientation is determined by protein – surface interactions; at high surface densities protein-protein interactions dominate orientation of proteins. Reprinted with permission from ref 274. Copyright Elsevier, 2010 (**b**) Two step binding process for Car9 fusion protein at a silica interface is described by a sigmoidal binding isotherm, in contrast to the monolayer Langmuir isotherm which is used to model single step binding process. Reprinted with permission from ref 203. Copyright American Chemical Society, 2019 (**c**) Cooperative behavior between three Car9 peptides is observed using molecular dynamics under conditions that represent surface bound behavior, and additional AFM images capture similar cooperative behavior when Car9 is fused to sfGFP at a silica interface. Reprinted with permission from ref 275. Copyright American Chemical Society, 2019.

Protein adsorption on solid surfaces is a phenomenon crucial to applications such as biological implants, membranes, analytical sensors, etc.[276–278] Simulations have successfully been

able to elucidate how individual proteins adsorb to surfaces, by providing a picture of the thermodynamic landscape, and elucidate the complex structural rearrangements that proteins undergo at a solid interface, characteristic of non-rigid particles.[48,269,270,279,280] As external conditions such as temperature, pH, surface chemistry etc. can be precisely controlled in simulations compared to experiments, their effects on peptide adsorption has been made clear.[48,280] However, an important aspect that needs to be addressed in both simulations and experiments is the issue of cooperative protein adsorption, as the primary focus so far has been the behavior of single biomolecules on a surface.[274,281] The commonly implemented Langmuir adsorption model, which assumes single monolayer adsorption on the surface, may be too simplistic when dealing with protein adsorption, as proteins assume a variety of conformations at a surface, unlike rigid gas molecules which the theory was originally devised for. A protein adsorption model that incorporates cooperative or nearest neighbor effects can account for deviations from the Langmuir model and provides a mechanistic explanation for the sigmoidal shaped adsorption isotherm. A schematic representation of a cooperative vs. non cooperative adsorption pathway is shown in **Figure 4.2a**.[274] The surface density of proteins can also determine the overall orientation of proteins on a surface. For example, at high surface density, protein-protein interactions can lead to changes in orientation compared to low surface densities, where only protein surface interactions determine orientation.[281] Thus, it is important for future research studies to devise simulations that can probe cooperative binding and still include nanoscale features of the protein-protein-surface interface.

Biomineralization is one field where there has been evidence in using combined simulation/experimental approach to uncover mechanisms of multi-protein adsorption at interfaces. Using Surface Plasmon Resonance (SPR), Hellner et al. developed a two-step cooperative model for the technologically relevant solid binding peptide and superfolder green fluorescent protein scaffold, sfGFP-Car9/silica system (**Figure 4.2b**).[203] The observed sigmoidal adsorption behavior is best captured with a two-step cooperative model. In this model, a first step describes intermolecular cooperative contacts between Car9 linkers, followed by an isomerization step where the Car9 and the protein scaffold are locked in a high affinity interaction with the surface. Building on this work a further combined simulation and experimental study related the correlation between the overall adsorption mechanism and specific chemical interactions at the peptide-surface interface.[275] In this study, MD simulations were used to capture SBP binding

(without the scaffold) to a silica interface. Using restrained MD simulations, they were able to predict mechanisms of cooperative behavior in the Car9 at the silica interface. In this study AFM images validated these predictions, demonstrating similar cooperative behavior when minimal scaffolded to sfGFP, showing strong agreement across both techniques in the mechanisms of cooperativity at a silica interface (**Figure 4.2c**).

An important aspect of multi-peptide adsorption is the interplay between entropic vs. enthalpic forces, and the idea that certain peptides are able to bind tightly due to increased access to configurational space when surface-adsorbed. Estimation of entropy, including breaking entropy down into translational, rotational and conformational parts using MD simulations has been done before, however, reconciling these features with experiments remains elusive. Early work by Tang et al.[282] and Sprenger et al.[88] has proposed different mechanisms based on entropic driving forces, however the quantitative link to experimental observations needs further refinement for these more complex systems. Still there remains a number of questions that have yet to be answered in the area of cooperative adsorption. The application and extension of these combined theory, simulation, and experimental methods that have showed initial success, can be used to study cooperative assembly in-depth. Overall, the combination of these approaches will form a complete mechanistic picture of the processes that drive these complex self-assembling systems.

4.4 Hierarchical Sequence Defined Polymers

Peptoids are a class of synthetic foldamers that have characteristics of both polymers and proteins.[192,201,283–285] The side chain is appended to the backbone nitrogen as opposed to the alpha-carbon atoms in amino acids. The N-substituted glycine backbone prevents the formation of hydrogen bonds, and this absence of hydrogen bond interactions renders the backbone floppy, making peptoid oligomers unstable. However, it is possible to form stable higher order structures like helices, sheets, and tubes through the appropriate choice of sidechain chemistries. As these structures depend on sidechain interactions and are not governed by backbone hydrogen bonds, peptoids remain stable under a host of different solvent and temperature conditions, opening different avenues of application, from therapeutics and diagnosis to catalysis. Zuckerman et al. pioneered the synthesis of peptoids using the solid-phase sub monomer method, enabling a vast array of amine sidechains to be conjugated to the backbone, making these molecules highly tunable.[286] The ability of these molecules to self-assemble into long ranged hierarchical structures

make them perfect candidates to study through simulations, as uncovering key molecular interaction will enable control of the material design space.

Hydrophobicity is often used as a knob to control the self-assembly of peptoids, as quaternary structures such as nanosheets arise from hydrophobic collapse of sidechain groups. In an attempt to tune the crystallization pathway of biomimetic polymers, Ma et al. used two peptoid systems which contained alternate polar and hydrophobic groups, with one system bearing a hydrophobic 'tail' at the end.[202] I*n situ* AFM was used to study the crystallization of the two peptoid systems on mica (**Figure 4.3a**).[202] They found the peptoid without the tail readily form crystalline particles in solution, which deposit on the surface at a rate that decreases as coverage increases. The tail system in contrast forms small micellar clusters in solution that then deposit on the surface and undergo conformational transformation to form crystalline bilayers. MD simulations by our group in this work shed light on the dimerization propensity of these peptoids, where the no-tail peptoid was seen to dimerize in the presence of Ca^{2+} ions, and in contrast, the tail peptoid remains dissociated with the tail wrapped inwards.[202] This is consistent with prior work demonstrating the importance of the parameterization of the interaction between the divalent and the carboxy sidechain.[287]

Stable peptoid nanosheets that self-assemble in solution have been widely characterized. Sanii et al. elucidated the mechanism of nanosheet formation from amphiphilic peptoid chains; they postulate that a monolayer intermediate first forms at the air water interface, and irreversible compression of this monolayer forms a stable bilayer.[288] Using lipid like sequence defined peptoids, Jin et al. successfully self-assembled membrane mimetic 2D materials in solution and on substrates (**Figure 4.3b**).[289] By using all-atom MD simulations, they observed that membrane stability was due to extensive pi stacking of the hydrophobic groups, to minimize contact with water. They were also able to incorporate functional moieties into the peptoid membranes, and these membranes demonstrated self-healing tendencies similar to lipid bilayers.

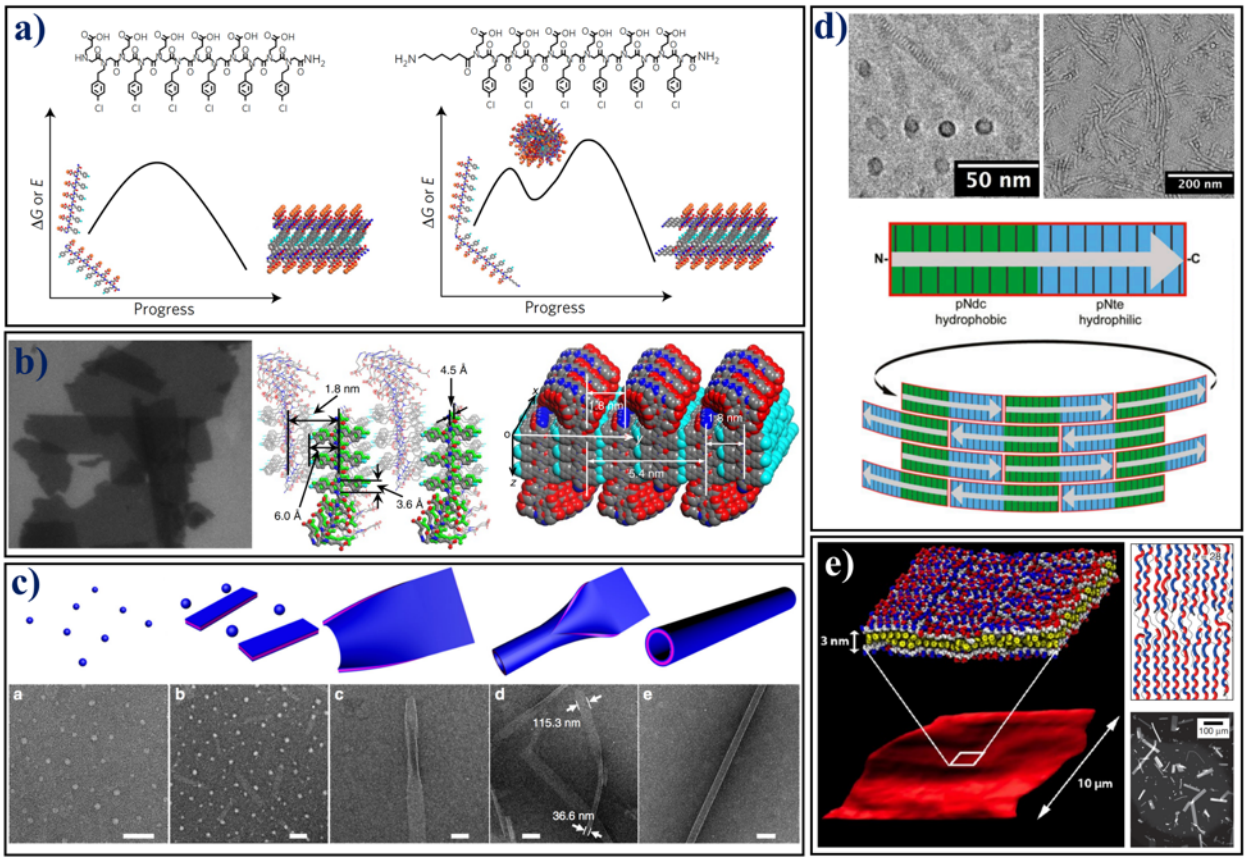


Figure 4.3. Peptoid assemblies at interfaces and in solution (**a**) Proposed peptoid crystallization pathways showing the effect of the hydrophobic tail on the propensity for aggregation on a mica interface; there is an alteration in free energy landscape from a single step to a two-step process. Reprinted with permission from ref 202. Copyright Springer Nature, 2017 (**b**) Self-assembly of lipid like peptoids into highly stable and crystalline membrane-mimetic 2D nanosheets, with a molecular model showing the proposed packing of the peptoid subunits. Reprinted with permission from ref 289. Copyright Springer Nature, 2016 (**c**) Time dependent TEM images showing the assembly pathway of peptoid nanotubes. Reprinted with permission from ref 290. Copyright Springer Nature, 2018 (**d**) Assembly of amphiphilic peptoid tiles into hollow nanotubes. Reprinted with permission from ref 291. Copyright 2016, National Academy of Sciences (**e**) Atomistic simulation of peptoid nanosheets matches well with experimental dimensions. Reprinted with permission from ref 201 and ref 292. Copyright American Chemical Society, 2016 and Springer Nature, 2015, respectively.

Another interesting self-assembled motif that peptoids adopt are tubes, as shown in **Figures 4.3c** and **4.3d**. Using sequence defined peptoids Jin et al. were able to design peptoid nanotubes that demonstrated tunable diameter, wall thickness, and mechanical properties through pH-triggered, reversible contraction-expansion motions.[290] Using time-resolved Transmission Electron Microscopy (TEM), they showed the self-assembly pathway of peptoid nanotubes. Nanospheres are seen immediately after the peptoid is dissolved in a mixture of water-acetonitrile solvent. Soon after (30 mins), a mixture of spheres and ribbons are observed. After 24hrs, nanoribbons with partially rolled up edges are seen, which form nanotubes after 48hrs. The

measured width of the nanoribbon is pi times the diameter of the nanotube, confirming that the nanoribbon folds up and closes to form the nanotube. Sun et al. observed that amphiphilic diblock copolypeptoids spontaneously self-assemble into nanotubes in water, with diameters that are dependent on chain length.[291] These tubes are held together primarily by side chain van der Waals interactions, and are formed without a central hydrophobic core. Although the authors didn't propose a mechanistic pathway for the formation of these structures, they were able to show that crystallization of both hydrophobic and hydrophilic domains play a prominent role in the formation of supramolecular structures.

In order to resolve atomistic details of peptoid assembly, the Whitelam group developed a CHARMM-based forcefield for peptoids to perform MD simulations of the nanosheets, as shown in **Figure 4.3e**.[292–294] Dimensions such as nanosheet thickness and sidechain distances matched up well between experiments and simulations. Results from this work revealed that peptoid backbone adopts novel secondary structure, an alternating binary motif characterized by two different regions of the Ramachandran plot. This enables the nanosheets to adopt a relatively flat configuration. The simulations also revealed that the nanosheets formed by short peptoids were less stable than those formed by long peptoids, in line with experimental observations. While the all-atom model has been useful in describing structural features of peptoid sheets, it can only access length scales in the range of 10 nm, and time scales up to a few microseconds, which is not long enough to observe the formation of higher order structure. Key questions remain about the mechanism of by which tubular peptoids are formed, and their self-assembly pathway from sheets to tubes remains elusive; simulations can provide insights into their formation. To capture the dynamic self-assembly process across a broad spatiotemporal scale, peptoid coarse-grained models with reduced degrees of freedom is necessary.[237] The CG forcefield for N,N-dimethylacetamide developed by Du et al. parameterized against OPLS-AA forcefield captures the appropriate local structure and water solvation behavior as the atomistic model. To remain consistent to the CHARMM based peptoid forcefield by Whitelam et al. described above, development of Martini based peptoids, similar to the peptide model parameterized by Mansbach et al. is necessary.[239,295]

4.5 Protein Assemblies

Naturally occurring proteins have shown to form a wide range of hierarchical, supramolecular structures due to an interplay of chemistry, sequence, and solution conditions.[174,176,178] Due to the high information content encoded in proteins, they are the cornerstone in numerous biological applications and are paving the way for future technologies.[276,296,297] To develop new functional materials it is important to understand self-assembly in natural systems, and proteins are excellent candidates in this regard. There has been a lot of effort in emulating natural protein structure and function using different building blocks, however, this work is often driven by intuition and trial and error approaches vs. a rational design approach. A clear knowledge of the overarching principles that govern in vivo self-assembly can guide the design and synthesis of in vitro systems. In applications such as biological sensors and 2D crystal peptide arrays, the underlying structure of the surface, or the lattice interactions dictates the long-ranged structure of adsorbed peptide assembly.[298] Below, we present studies where computational models have been used in tandem with experiments, to more effectively design large scale protein assemblies.

The Tobacco Mosaic Virus (TMV) capsid made from 2130 molecules of coat protein (CP) is widely considered as the paradigm of self-assembly in biology, as it spontaneously organizes into a helical, cylindrical rod around an RNA strand (**Figure 4.4a**).[175] As it was the first macromolecular structure observed to self-assemble in vitro, this mechanism has been studied in detail. The CP first assembles into disks, after which the viral RNA inserts itself into the core of the disc, that elongates and grows into a helix encapsulating the RNA. The entire structure is formed through competing attractive hydrophobic and repulsive electrostatic interactions. As the structure is well understood, TMV has been used in a number of applications, from vaccines to micromachines.[299–301] The hollow nature of TMV makes it an excellent scaffolding agent that encapsulates and deliver therapeutic drugs to target sites in in the body.[301] The capsid has been shown to reform into other shapes when subjected to different temperatures, and is used to increase the efficiency of electrodes by adhering to the surface of electrodes, thereby increasing the surface area for reactions to take place.[299,300]

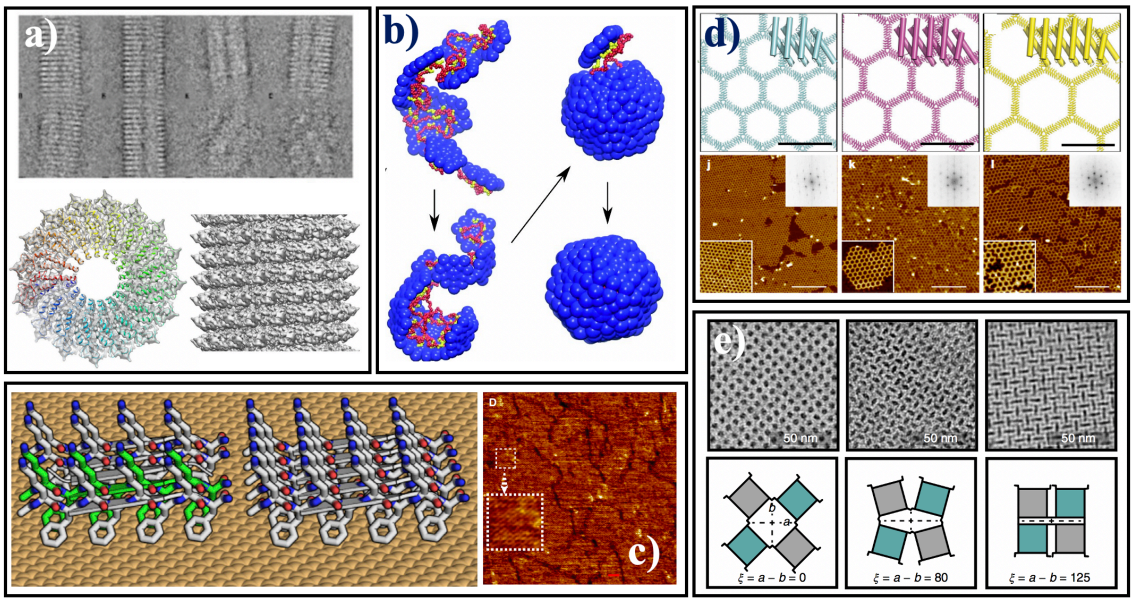


Figure 4.4. Natural and de novo protein assemblies (**a**) helical reconstruction of tobacco mosaic virus using cryo TEM, showing the central axis and cross sections. Reprinted with permission from ref 175. Copyright 2010, Elsevier. (**b**) Mechanism of viral capsid assembly using Brownian dynamics simulations, where a small partial capsid nucleates on the polymer followed by reversible addition of one or two subunits at a time. Reprinted with permission from ref 302. Copyright 2014, Elsevier. (**c**) Schematic representation and AFM image of a designed peptide adsorbing on to the surface of graphite as beta sheets. Reprinted with permission from ref 303. Copyright 2016, Elsevier. (**d**) Protein-protein interactions drive the formation of tunable hexagonal lattices on the surface of Mica. Reprinted with permission from ref 298. Copyright 2019, Nature Springer. (**e**) TEM images and schematic representations of the conformational states accessible by RhuA lattices. Reprinted with permission from ref 304. Copyright 2018, Nature Springer.

To elucidate the role of the encapsulated polymers (RNA, DNA, proteins, etc.) in conferring stability to the viral capsid, Elrad and Hagan used coarse-grained Brownian dynamics to study the self-assembly process in an icosahedral virus (**Figure 4.4b**).[302,305] They find that the nature of assembly can be tuned by controlling polymer length, solution conditions, and protein concentration. Moreover, they also find that the nature polymer has dramatic effects on assembly kinetics, as the polymer significantly lowers the free energy barrier for nucleation. This study sheds light on different aspects of multicomponent assembly processes and provides a robust framework for designing *in vitro* assembly experiments, guided by simulations.

The bacterial caboxysome shell consisting of thousands of protein subunits is another example of *in vivo* protein assembly giving rise to a complex 3D structure. Tanaka et al. constructed atomic 3D models of the shell from two known classes of carboxysomes. Their model showed that an icosahedron unit could be formed from smaller hexameric and pentameric protein subunits.[176] This study highlighted the fact that intricate molecular structures are formed from

common atomic interactions and a few basic building blocks, opening up strategies for designing various molecular machines and scaffolds *in vitro*.

Mustata et al. designed peptides that could self-assemble on graphite and graphene surfaces, forming beta sheets with microdomains corresponding to the underlying surface lattice (**Figure 4.4d**).[303] They were also able to assemble DNA along the peptide that adhered to the surface, creating a 2D platform for studying DNA-peptide interactions. In a similar vein, Pyles et al. showed that large protein assemblies could be precisely assembled on inorganic interfaces through surface lattice matching (**Figure 4.4c**).[298] The interface that they chose was muscovite mica, that has a well-defined sublattice upon which biomolecules have been shown to bind. Using a *de novo* designed protein scaffold with regularly repeating carboxylate units to match the spacing of the underlying mica surface, they were able to assemble 2D protein arrays that showed order over tens of millimeters. Aided by Rosetta, they created a tunable hexagonal protein lattices on the surface, by changing the number of repeat units and K^+ ion concentration in the solution. The primary driving force behind the formation of the honeycomb structure is entropic packing effects, modulated by strong electrostatic interactions. This feature of lattice matching is also observed in ice binding proteins assembling on ice crystals,[306] proving that simple principles of molecular recognition can lead to tunable, long ranged assemblies on surfaces. These studies are prime examples of computationally driven experimental design of protein assemblies at solid/water interfaces.

Using principles of *de novo* design, Suzuki et al. constructed stimuli responsive crystalline assemblies of the C4 symmetric tetrameric protein RhuA through functionalized cystines that direct self-assembly via disulfide bonds.[307] The strong, reversible bonds between the protein subunits allows the lattice to expand and contract without losing crystallinity, and while remaining isotropic. This assembly was shown to possess unique mechanical properties like negative Poisson's ratio (leading to increased toughness and resistance to fracture), and stimuli responsive behavior giving rise to molecular membranes with tunable porosity from 4 nm to 1 nm. By mapping out the free energy landscape of this system using atomistic MD simulations, Alberstein et al. showed that the dynamics of the protein crystalline lattice was driven by solvent entropy, and with this knowledge, elucidated specific interactions that could be tuned to achieve the desired mechanical or chemical response (**Figure 4.4e**).[304]

Using the enhanced sampling toolkit described in the Simulations section, we have successfully used MD simulations with explicit solvent and ions to model peptide dimerization in solution, as well as the underpinnings of molecular recognition of solid binding peptides to various surfaces.[48,90,270] However, in order to achieve a physics-informed prediction of protein self-assembly at interfaces, there is still much work to be done improving both coarse-grained potentials as well as sampling algorithms.

4.6 Statistical Learning and Artificial Intelligence

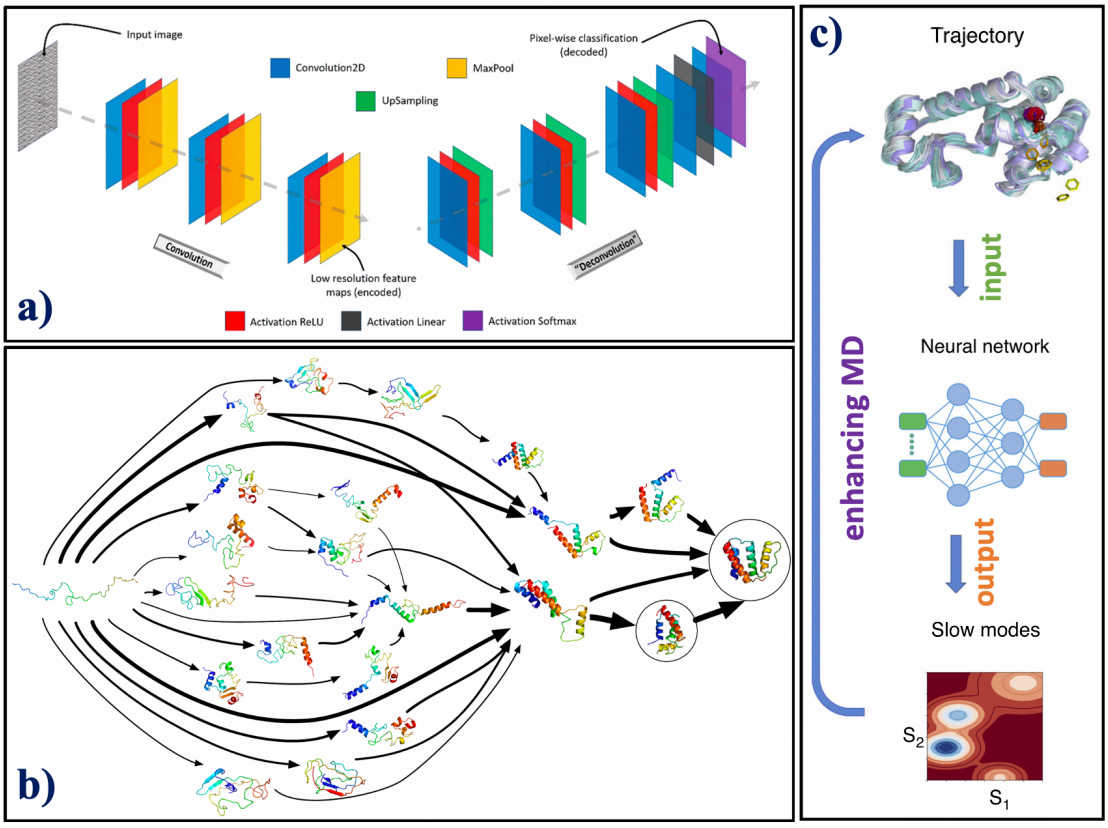


Figure 4.5. Machine learning approaches (**a**) Depiction of a fully convolutional neural network with encoder-decoder type architecture used to characterize defects in STEM and STM images. Reprinted with permission from ref 308. Copyright 2017, American Chemical Society. (**b**) Generalized schematic of a markov state model describing protein folding, thickness of arrows indicate relative probability of transitioning from one state to the next and describe overall likely pathway of folding. Reprinted with permission from ref 309. Copyright 2012, American Chemical Society. (**c**) Schematic illustrating the typical workflow of methods that use machine learning to analyze and enhance MD simulations. Reprinted with permission from ref 310. Copyright 2020, Elsevier.

Machine learning methods (ML) have been finding increasing use across the spectrum of problems described above with examples ranging from experimental characterization of protein self-assembly to improving MD simulation methods. Detailed description of the use of ML in these types of problems is outside the scope of this review, and have been previously characterized elsewhere.[311–313] Broadly, ML methods for use in molecular and chemical studies can be categorized as accomplishing two types of tasks; 1) To infer a quantitative relationship between inputs and outputs (supervised learning), or 2) To predict the relationship between a set of inputs (unsupervised learning). These tools provide a large and scalable toolbox for identifying trends in massive data sets, which can be applied to both simulation and experiments alike.

Experimentally, many approaches that have made use of these methods have largely been directed at high-throughput analysis of microscopic and spectroscopic imaging data. One approach towards analyzing these datasets involves the application of convolutional neural networks (CNN), an example of which is shown in **Figure 4.5a**. In this example, Ziatdinov et al. used a weakly supervised approach (which applied theoretical data as "labeled" predictions of a lattice with defects), to determine the atomic positions and vacancies in unprocessed Scanning Transmission Electron Microscopy (STEM) and Scanning Tunneling Microscopy (STM) imaging data of different silicon dopeants and graphene.[308] Deep learning and ML analysis of imaging data have already been widely successful in the medical field[314,315] and have become critical in materials design and discovery[313,316,317]. Only recently, has self-assembly begun to be explored in this respect. Ziatdinov et al. demonstrated the potential of an approach utilizing Markov random field models and convolutional neural networks to incorporate ab-intio data to analyze spatial, rotational, and structural states of complex molecular assemblies onto a surface.[318] These strides demonstrate an important first step at applying imaging-based methods towards the analysis of more complex biological assemblies on surfaces. Further, the ability to extract physics-based information and theory from these approaches and combine them with first-principles simulations can overall strengthen the use of these techniques to inform mechanistic insight into self-assembly.

With respect to simulation, there are a number of application areas and this review is far from comprehensive. We refer the readers to a number of excellent review papers which have covered these applications in detail.[210,310,319,320] In the context of self-assembly DS methods are aimed at improving three areas; 1) forcefield development, 2) sampling limitations, and 3) property prediction. Specifically, in the case of assembly on complex materials (e.g., protein adsorption on

solid surfaces), the availability validated force field still limits the application of MD to a small number of material categories. The recent development of new methods for building forcefields using neural network trained on ab initio data[321–331] is promising for the rapid development of this area. Bayesian approaches such as metadynamics metainference have successfully been used to overcome systematic errors in forcefield parameterization, but often require input from complex experimental methods or rigorous QM calculations.[245,332–334] Sampling limitations in molecular simulation have been a long-standing issue. Many statistical approaches, as described earlier, have been proposed to overcome these limitations. More recently, many ML-based approaches are an exciting outlook this area. For example, Noé et al. recently demonstrated an approach using deep machine learning and statistical methods to produce unbiased equilibrium samples of rare events, without a priori knowledge of the reaction coordinate, known as Boltzmann generators.[335] A number of techniques rooted in the use of neural networks have been shown powerful in predicting potential energy surfaces, structure, and coarse-grained models.[336–340] Progress in this area will be critical as we seek to use atomistic simulations to study the types of problems described such as understanding cooperativity in solution or atomic scale driving forces that underlie protein self-assembly

Property prediction has been one of the more enticing applications of data science to simulation. This progress suggests that, given a sufficient amount of training data, ML could assist in the prediction of hierarchical assembly of structures based on chemistry, sequence, and structure – an exciting frontier of this field. Some progress has already been made in this respect; Le at al. have predicted the assembly of lipid-based delivery nanoparticle systems[341], Li et al. developed a combinatorial approach to predicting the assembly of hydrogels[342]. Beyond the need for large amounts of training data, an additional challenge is development of chemical descriptors which can effectively correlate structure with assembly. For example, Shamay et al. have shown electrotopological molecular descriptors are highly predictive of nano-assembly and nanoparticle size[343]. Physics-based simulations may be able to inform many of these decisions and are critical to the development of these models. Further, biological systems like proteins peptides, proteins, and peptoids are complex structures which require further advances in structure prediction and assembly, likely from coarse-grained simulations.

4.7 Conclusions and Future Directions

Controlling and tuning biomolecular self-assembly will pave the way for applications in therapeutics, materials synthesis, diagnostics, etc. Beyond the individual success stories we described here, the foundational challenge faced by the field is the development of entirely new heuristics and theories that will enable predictive power over the design and optimization of these systems. The design space offered by the different building blocks that can be used in hierarchical self-assembly is vast, and future advances in the field require reliable models to help focus the efforts of expensive synthesis and characterization experiments. Physics based simulations like MD, in spite of the advances the field has made, will not be a panacea. Like the accuracy of the empirical potentials that describe key fundamental interactions, the speed of calculations and the emergence of new statistical mechanical sampling approaches continues to improve each year. However, it is not realistic to expect that these approaches will either replace experiments or provide a universal tool for materials design as the compute requirements are simply too far beyond today's capabilities. Instead, we expect simulation will emerge as a reliable middleman between even faster computational design and the needed work of careful wet lab experimentation. Integrating experiments and simulations will further propel advancements in the field of biomolecular assembly. As demonstrated by many examples in this review (protein design, biomineralizing peptides, etc.) simulation aided experiments can effectively reduce the vast chemical and process design space of biomolecular assembly, mitigating the need for performing a battery of experiments. In addition, experimental data (NMR chemical shifts, binding free energy, long range structure from X-ray diffraction, etc.) are useful to validate simulations, and to correct artifacts that may arise due to forcefields used. Therefore, a major priority in the near term is the rapid expansion of the use of data driven methods like machine learning. This frontier region in hierarchical materials science will be critical in the coming years across the full spectrum of approaches in the scientific pipeline from the design and execution of MD simulations all the way to the interpretation of sophisticated high throughput experimental results. While applications of these new methods in hybrid and hierarchical materials are still in their infancy, we expect a rapid adoption in the field and an exciting future.

4.8 Acknowledgements

This work was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, as part of the Energy Frontier Research Centers program: CSSAS--The Center for the Science of Synthesis Across Scales under Award Number DE-SC0019288

Chapter 5
Efficient sampling of high-dimensional free energy landscapes: A Review of Parallel Bias Metadynamics[4]

5.1 Abstract

A number of enhanced sampling methods have been developed to overcome the length and time scale barriers of classical simulations. Metadynamics has made considerable strides in the last decade as a technique for constructing free energy landscapes as a function of a few low dimensional descriptors of atomic positions, commonly referred to as collective variables (CVs). In particular, parallel bias metadynamics (PBMetaD) and its variants enable the sampling of many CVs without prohibitively increasing simulation time. This parallelizable scheme allows for its implementation in systems that necessitate the use of more than a few CVs, or in the case that the CVs corresponding to the slowest modes of a process are not easily identifiable. Here we present a review of notable enhanced sampling schemes with a focus on the underlying theory of PBMetaD. We discuss various examples where this method has been applied, and future areas of impact.

[4] This chapter was reproduced with permission from S. Alamdari, J. Sampath, A. Prakash, L. Gibson, and J. Pfaendtner. Efficient sampling of high-dimensional free energy landscapes: A Review of Parallel Bias Metadynamics. Foundations of Molecular Modeling and Simulation. 123-141, 2021. Copyright 2021 Springer, Singapore.[495]

5.2 Introduction

Molecular dynamics (MD) and Monte Carlo simulations are powerful methods used to elucidate the conformational and structural properties of molecules by the dynamic or stochastic evolution of a molecular configuration. Simulations find value across a number of fields, and range from understanding the fundamental driving forces in a simple Lennard Jones system to gaining thermodynamic and kinetic insight of complex biological and chemical systems. MD simulations, while a powerful tool for sampling the dynamic evolution of a system, usually fails to sufficiently sample all the relevant configurations of these complex systems. Further, when the interesting states of the system are separated by free energy barriers larger than kBT, it becomes impossible to comprehensively sample a 3N-dimensional phase space. In order to improve sampling in such simulations, a number of enhanced sampling methods have been proposed. These methods have been successfully applied to study rare events in biological processes (e.g., protein unfolding[231,344,345], peptide adsorption[70,229,346], drug and ligand docking[347–349], phase transitions[350,351], and nucleation[352,353]) and chemical reactions[230,354–358] in combustion, pyrolysis, and enzymology to name a few. This review is not meant to be comprehensive, but an overview of the development of metadynamics and a few of its variants. Notable alternative enhanced sampling methods including, but not limited to, transition path sampling[359], log mean force dynamics[360], multiscale enhanced sampling[361], average force[362], nudged elastic band[363], Onsager-Machlup multiscale enhanced sampling (MSES)[364], boxed molecular dynamics (BXD)[365], blue-moon sampling[366], and string method[367] are outside of the scope of this chapter. These methods and more can be found in excellent detail in various reviews.[368–371]

Broadly, one convenient way to characterize enhanced sampling methods is to categorize them as either tempering or external biasing methods. In tempering methods, temperature is exploited to enhance the sampling of the system. The canonical form of these methods is simulated annealing[372], a generalized probabilistic method used to quickly reach a global optimum. The probability of accepting an exchange relies on a temperature variable which either promotes or restricts exploration of new states. This process rapidly becomes computationally intensive, which introduced a need for more advanced methods with respect to molecular simulations. Replica exchange is the most generalized sampling approach used in MD, where several replicas of a molecular configuration are simulated at the same time. A swap of two replicas can occur under an acceptance probability that preserves detailed balance. Replicas are weakly coupled and only

interact under an exchange attempt. This promotes a highly parallelizable and enhanced method for sampling a system. While it is not a requirement, the most common variant of replica exchange involves simulating replicas at different temperatures; this is commonly known as parallel tempering (PT).[373,374] Other sampling subsets of replica exchange include Hamiltonian replica exchange[375], and solute tempering[376]. These methods and many more such as adiabatic free energy dynamics[377], and temperature accelerated molecular dynamics (TAMD)[378] are described in much greater detail in a number of prominent reviews.[87,229,379,380]

In external biasing methods, sampling of the system is enhanced by applying an external bias on the slow modes of the system. These slow modes of the system can be described by defining a few choice collective variables (CVs). A CV is a differentiable function that provides a low dimensional projection of all atomic coordinates in the system. This function can be as simple as the distance between two atoms, or be extended to more complex ones like Pickett angles and Cremer-Pople coordinates[381], which can quantify the conformation of N-member rings. A good CV (or set of CVs) should capture the slow dynamic evolution of the system (e.g., breaking/forming bonds, adsorption/desorption on a surface, protein unfolding, etc.) and distinguish between states of interest. CVs can provide meaningful insight about the energy differences between two stable states, but should not be confused with a reaction coordinate, which carries a special connotation in the literature. Reaction coordinates are capable of uniquely quantifying the dynamics of a system as it transitions between two stable states. Discussion on the identification and assessment of reaction coordinates has been thoroughly accomplished elsewhere.[382] Choosing a CV is a non-trivial task, as it can be difficult to gain chemical intuition about a system prior to launching a simulation.

External biasing methods greatly benefit from reducing the free energy landscape of a system in terms of a few CVs. One early example of an external biasing method is umbrella sampling (US), developed by Torrie and Valleau.[383] In US, a static restraint (or bias) in the form of a harmonic potential is applied along multiple intervals along a CV. In the simplest terms, a bias is an additional term added to the potential energy of the system. In turn, this allows for greater sampling of high energy regions. Sampling many overlapping “umbrellas” in phase space allows for the reconstruction of the underlying free energy surface. This can be done by using the weighted histogram analysis method (WHAM)[384], popularized by Alan Grossfield[385], or by applying the multistate Bennett acceptance ratio (MBAR) estimator introduced by Michael Shirts

and John Chodera[386]. Often, however, this method is not sufficient to fully sample the phase space and interesting chemistries of the system because integration of the underlying potential of mean force becomes degenerate with higher dimensional CVs. However, the most significant drawback of US is that most CPU time during a simulation is spent sampling high energy, and uninteresting, regions of phase space more so than the relevant ones. A number of external biasing methods have been developed to account for the drawbacks of US, including targeted MD[387], energy landscape paving[366], the Wang–Landau algorithm[388], variationally enhanced sampling[389], and many others.

Metadynamics (MetaD), originally developed by Laio and Parrinello[390,391], gained popularity as a robust and flexible technique to address a wide range of problems including protein folding[234], nucleotide complexation[392], chemical reactions[230,393], ligand docking[394,395], and phase transitions[396]. In MetaD, an external bias potential in the form of a Gaussian is applied along a few choice CVs. This enhances the fluctuations of those CVs, discouraging the system from visiting previously visited metastable states, thereby allowing it to sample many configurations. In the first variant of MetaD, constrained MD was used to impose a history dependent bias, allowing the system to diffuse through the free energy surface.[390] In the subsequent versions, the bias is applied directly on the CVs while the system dynamically evolves during an MD simulation.[394,397] A drawback of standard MetaD and its early variants was convergence of the bias potential to the underlying FES, making it unclear when to stop a simulation. This problem was alleviated by the introduction of well-tempered MetaD (WTMetaD), where asymptotic convergence of the FES was proven.[86] Similar to standard MetaD, small, repulsive, history-dependent Gaussian kernels that can adapt to the FES are deposited on the phase space in WTMetaD. The height of the Gaussian bias decreases over time, and in the long-time limit, the bias is shown to converge exactly to the underlying FES.[398] Another variant that achieved convergence better than standard MetaD was flux-tempered MetaD, proposed by Singh et al., an iterative method that maximizes the flux of a random walker in CV space.[399] Still, simulations that apply enhanced sampling methods suffer from being computationally intensive as the number of CVs that are biased exponentially increases the phase space to be sampled.

Numerous variants of MetaD have been proposed over the years,[399–404] and herein we briefly describe some of these methods which have been developed to speed up convergence of the enhanced simulation. In the multiple-walker version of MetaD, “walkers” (or replicas) share a history-dependent bias potential, allowing each walker to concurrently explore the FES. The error

in the energy landscape can be reduced by increasing the number of walkers (note that this number depends on degree of correlation between walkers, and is thus limited).[405] In the adaptive Gaussian flavor of MetaD, the width of the Gaussian biases are adjusted on the fly to better mimic the features of the underlying FES, which is advantageous while dealing with systems with markedly different CV fluctuations.[401] Readers are encouraged to go through prior MetaD literature that detail other variants.[229,231,402,406–409]

In the context of biomolecular applications, variants that couple MetaD with replica exchange are perhaps the most widely employed MetaD methods, and are done so to explore more degrees of freedom.[232] Examples of these are parallel tempering MetaD (PTMetaD), where temperature is used to increase sampling and explore hidden degrees of freedom[84], parallel tempering MetaD in the well-tempered ensemble (PTMetaD-WTE), which addressed the system size limitations of PTMetaD[233], bias exchange MetaD (BEM)[234] which can increase dimensionality by biasing more than 2-3 CVs, and collective variable tempering[410]. The replica exchange methods above have the drawback of being computationally intensive, either requiring a careful, non-trivial choice of temperatures to generate replicas or requiring an additional parameter (bias exchange stride in BEM) to sample the CV space.

Traditionally, MetaD-based sampling has been limited to 1-2 CVs, because the simulation time to converge a multi-CV system increases exponentially with the number of CVs. In the case of BEM, each additional CV increases the number of parallel replicas of the system, as serially biasing many CVs limits collective motion. As an alternative that relies upon the basic premise of replica exchange with concurrent MetaD[410], Pfaendtner and Bonomi developed a scheme known as parallel bias MetaD (PBMetaD). PBMetaD allows a single replica of the system to simultaneously apply 1-dimensional biases to many CVs, allowing for a highly parallelizable simulation that can bias many more CVs when compared to other enhanced sampling methods.[235] Recently, variations of PBMetaD have been proposed including metadynamic metainference (M&M)[400] and PBMetaD with partitioned families (PBMetaD-PF)[236]. This review is a summary of methods and applications related to the PBMetaD framework. Below, we discuss the theory behind PBMetaD and its variants M&M and PBMetaD-PF. To illustrate the use of these methods we describe six examples that deal with: (1) sampling the structure of biomolecules at interfaces, (2) elucidating chemical reaction pathways, or (3) guiding biomolecular simulations toward

experimentally observed structures. We briefly comment on the power of this approach and further emphasize its potential to be incorporated in other research areas.

5.3 Theory

The following section emphasizes how the bias potential calculation is uniquely distributed in the PBMetaD method. In the original metadynamics method,[390] a single multidimensional bias potential is created where the dimensionality of this potential is equal to the number of CVs. In contrast, in the PBMetaD variant, several single-dimensional bias potentials are constructed, and the number of potentials is equal to the number of CVs. As shown empirically by Pfaendtner and Bonomi,[235] this method also converges to the correct free energy profile. This method permits the construction of multiple single-dimensional potentials in lieu of a multi-dimensional potential, which saves computational resources, by utilizing a conditional weight term. Thus, the bias potential for the i^{th} CV(s_i), under the PBMetaD framework, is constructed as following:

$$V_G(s_i,t) = \int_0^t dt'\, W * \exp\left(-\frac{V_G(s_i R(t'),t')}{k_B \Delta T}\right) * \exp\left(\frac{-\left(s_i R(t) - s_i R(t')\right)^2}{2\sigma^2}\right) * W_{PB}(s_i,t') \quad 5.1$$

$$W_{PB}(s_i,t) = \frac{\exp\left(-\frac{V_G(s_i R(t),t)}{k_B T}\right)}{\sum_{j=1}^{n} \exp\left(-\frac{V_G\left(s_j R(t),t\right)}{k_B T}\right)} \quad 5.2$$

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 **Equation 5.1** follow the standard WTMetaD biasing scheme, which allows for the reduction of the height of the Gaussian kernel as bias accumulates. The last term, shown in **Equation 5.2**, is the conditional weight such that CVs that are in less explored regions (or regions with low bias) receive greater bias, while CVs in highly explored regions (or regions of high bias) receive smaller bias. This terms results from a derivation of a cumulative bias that uses an auxiliary variable to stochastically switch among individual 1D MetaD biases. As equilibrium statistics of the auxiliary variable are uninteresting, it is integrated out to obtain Eqs. 1-2.[410] Consequently, these single-dimensional

potentials evolve independently, only interacting through the conditional weight term. This functional form ensures there is no contamination between the individual bias potentials and that the correct free energy is ultimately obtained.

The M&M approach, illustrated in **Figure 5.1**, recently developed by Bonomi et al. uses Bayesian statistics to incorporate experimental observables into the MetaD enhanced sampling scheme.[400] This metainference approach is powerful in that it allows a user to simulate ensembles consistent with experimental data, guiding the simulation to sample conformations obtained from experimental techniques like NMR. Application of PBMetaD to this approach allows for the rapid convergence of these systems to the underlying true free energy profile.

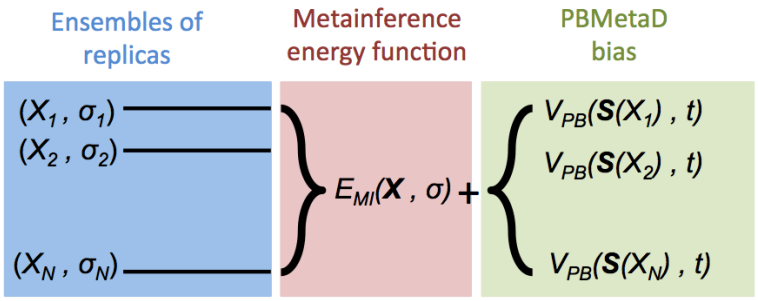


Figure 5.1. Schematic representation of the M&M method. In this method an ensemble of replicas is simulated, where X represents the molecular coordinates of the system, and σ represents the noise in the data. Replicas are coupled to the Metainference function, and sampling of phase space (S) is accelerated by addition of bias from PBMetaD. Adapted from ref [400].

In this combined flavor, each replica (N) is simulated using the metainference energy function (E_{MI}) and the PBMetaD bias factor, giving a total energy function represented by:

$$E_{M\&M}(X, \sigma, t) = E_{MI}(X, \sigma) + \sum_{r=1}^{N} V_{PB}(S(X_r), t) \qquad 5.3$$

This unique approach allows for efficient sampling of the conformational space, which provides higher resolution to results that have a low signal to noise ratio, which is common to experiments of heterogeneous systems. Combined with metadynamics, this is a powerful tool to extract crucial information about biomolecular systems.

Recently, Prakash et al. proposed a variant of PBMetaD called PBMetaD with partitioned families (PBMetaD-PF), to introduce a scalable method for systems that contain degenerate CVs.[236] In this method, CVs that are degenerate are partitioned into the same family such that each member of the family contributes to the same single-dimensional potential, similar to the multiple

walkers framework, but within a single replica.[405] The method is diagrammed in **Figure 5.2**. Consequently, the number of bias potentials is equal to the number of families. And the bias potential for a CV (x) in a partitioned family (say PF1) which has anywhere from 1 to m members, is recovered through:

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

$$\exp\left(\frac{-(s_{PF1-x}(R(t)) - s_{PF1-k}(R(t')))^2}{2\sigma^2}\right) * W_{PB}(s_{PF1-k}, t') \qquad 5.4$$

In this way, the bias potential of the family acts simultaneously and contributes to the enhanced sampling of all the degenerate CVs. Note that the denominator of the conditional weight term (**Equation 5.2**) still sums over all the CVs biased in a system, as is done in regular PBMetaD.

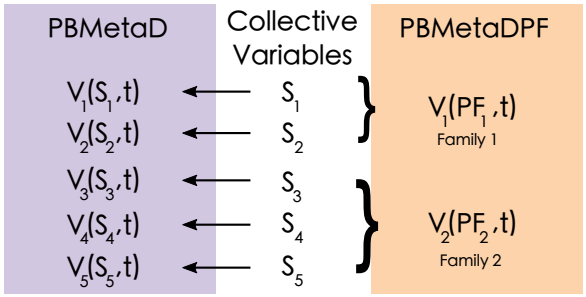


Figure 5.2. Schematic representation of the differences between the PBMetaD and PBMetaD-PF sampling schemes. Under the PBMetaD biasing scheme, an individual bias potential is evolved for each CV and the CV only acts under its own potential. In contrast, the PBMetaD-PF schemes allows for all of the members 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. Reprinted with permission from ref [236]. Copyright 2018 American Chemical Society.

In the following sections we will go on to introduce model systems where these methods have been applied, and comment on areas which might benefit from implementation of newer PB variants.

5.4 Example 1: Alanine and Sarcosine in Solution and Near a Surface

Prakash et al.[411] employed PBMetaD to study the orientation and conformation of alanine and sarcosine in water and near self-assembled monolayers (SAM). The study was aimed towards parameterizing a new model to understand how differences in the conformational flexibility of a peptide and peptoid affect their structure in solution and near a surface. By biasing the following three CVs, they were able to ensure that folding/unfolding of the peptide and binding/unbinding to the surface were simultaneously explored: (1) radius of gyration of the molecule, (2) alpha-beta dihedral parameter, and (3) orthogonal distance between the molecules center of mass and the surface (which was frozen to prevent extensive deformation during simulations). The free energy profile was constructed after performing a three replica multiple walkers PBMetaD simulations (however, using more than two CVs in WTMetaD exponentially reduces the computational efficiency of the simulation). To further understand the effect of surface chemistries, they simulated both hydrophobic and hydrophilic SAMs (**Figure 5.3a**). Due to the rapid convergence of a system biased with three CVs, the authors were also able to employ the scheme developed by Tiwary et al. to reweight two other variables—clusters and number of contact residues.[412] The sampling of the low-energy regions of the three CV space would have been computationally more demanding using a traditional multidimensional MetaD bias.

From **Figure 5.3b**, we see that the binding of both alanine and sarcosine are similar at the hydrophobic and hydrophilic surfaces, although the interaction with the hydrophobic surface for each molecule is stronger than with the hydrophilic surface. The contribution of two types of interaction energies (Lennard-Jones and Coulomb) on the binding of the molecules to the surface was also studied by reweighting the simulation trajectories to calculate the unbiased distribution of these short-range interaction energies. Although the overall interaction energies of both the molecules are comparable (**Figures 5.3c** and **5.3d**), the LJ energy dominates the binding to the hydrophobic surface.

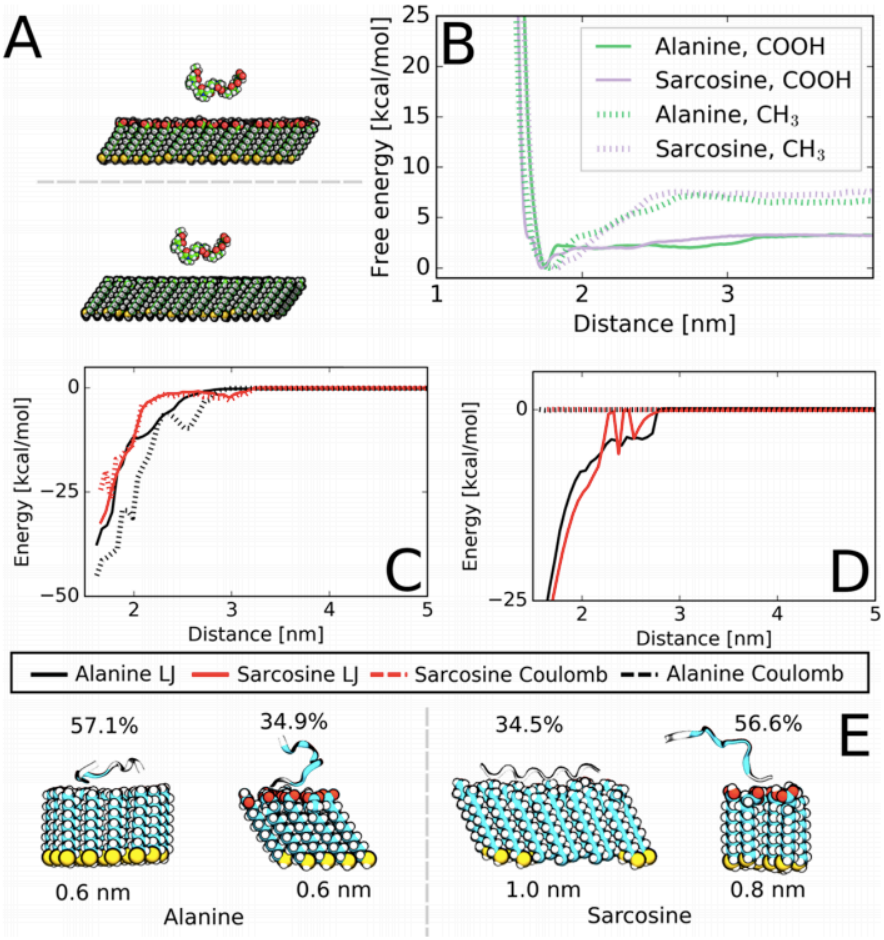


Figure 5.3. Simulation of sarcosine and alanine near SAMs that are either hydrophobic (−CH3 terminated) or hydrophilic (−COOH terminated). (**A**) Simulation setup for the hydrophilic surface (bottom) and hydrophobic surface (top). (**B**) Free energy curve as a function of distance from the surface. Energy contributions from short-range Coulombic and Lennard-Jones interactions with distance from the (**C**) hydrophilic and (**D**) hydrophobic surfaces. Reprinted with permission from ref [411]. Copyright 2018 American Chemical Society.

5.5 Example 2: Elucidating Reaction Networks with SPRINT Coordinates

Fu and Pfaendtner[413] have demonstrated an approach to studying chemical reaction networks by using PBMetaD to explore the connections of atoms in a molecule. The CVs that they utilized to describe these connections are called Social PeRmutation INvarianT (SPRINT) coordinates.[414] A SPRINT coordinate is a type of generic CV that describes the relative connectivity of atoms in a system and is derived from the contact matrix of all relevant atoms. These generalized CVs are unique in that they require no chemical intuition about a system *a priori*, and the only "chemical information required" are the equilibrium bond lengths of the atoms

involved. In practice the number of SPRINT CVs scales massively with the size of the system, due to the nature of creating a contact matrix for every atom in the system. Fu and Pfaendtner have shown that chemical reaction networks can be explored by using the PBMetaD scheme to bias all of the atomic SPRINT coordinates of a system. Particularly, they employed this PBMetaD+SPRINT framework to explore the decomposition of γ-ketohydroperoxide (KHP), also known as the Korcek reaction mechanism.

From their study, they concluded that the recovered reaction networks and pathway selectivities were largely unaffected by the choice of conservative or aggressive bias parameters (i.e., Gaussian hill height and width). It was noted, however, that the use of aggressive bias parameters may increase the number of reactive events that are sampled, but also carries the potential risk of missing reverse reactions.

They also investigated how the choice of starting structure affected the sampled pathways. For KHP decomposition, the lowest energy pathway proceeds via a ring-formation step to form 1,2-dixolan-3-ol (CYCP), shown in **Figure 5.4**, from which the network diverges into four pathways. They found that the choice of starting structure (i.e., KHP or CYCP) did not impact the sampled pathways but affected the recovered selectivity. Since KHP and CYCP exhibit different bonding topologies, the atomic SPRINT coordinates are distributed differently between the two structures, and thus the history-dependent bias potential accumulates at different points in phase space.

Lastly, one important feature of PBMetaD+SPRINT that separates it from other automated reaction prediction methods is the ability to distinguish between pathways that produce enantiomers, whereas other methods lose chirality in their predictions.

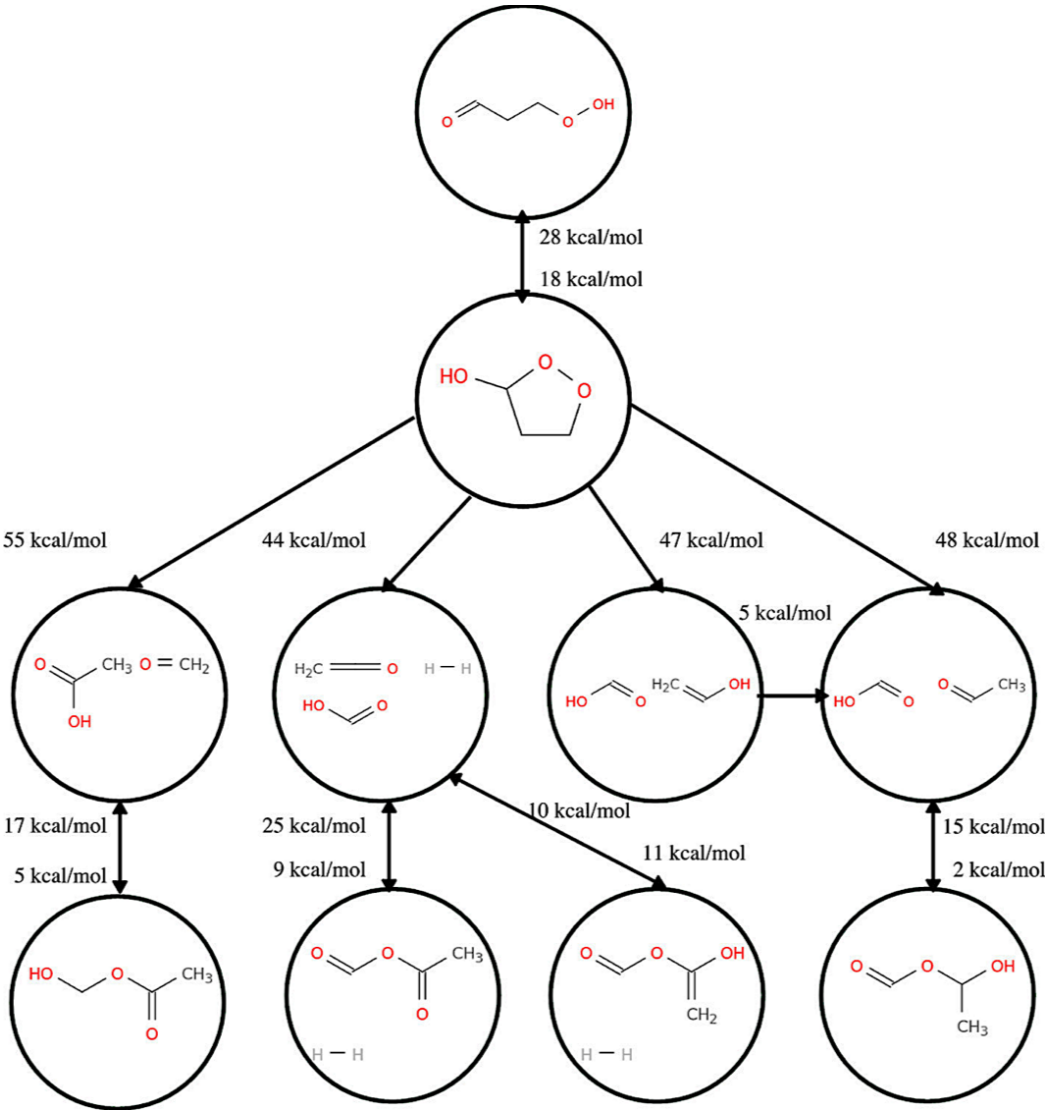


Figure 5.4. Reaction network of KHP after passing through CYCP. This reaction network was produced from simulations with KHP as the starting structure and with a temperature of 800 K. Barrier heights were not determined from the PBMetaD simulation, but added later from additional DFT calculations. Forward reactions barriers are listed as the first value going from top to bottom, the second value listed corresponds to the reverse reaction. Reprinted with permission from ref [413]. Copyright 2018 American Chemical Society.

5.6 Example 3: Effect of Surfactant on the Surface Induced Denaturation of Proteins

It is known that surfactants prevent surface induced denaturation of proteins, and to study this mechanism, Arsiccio et al. employed atomistic MD enhanced by PBMetaD in order to exhaustively sample peptide configurations at three interfaces, namely, air-water, ice-water and water-silica.[415] This study uses GB1, a 16-residue peptide that folds into a stable beta hairpin, and

Tween 80, a polysorbate as the model protein and surfactant, respectively. By employing PBMetaD, the authors were able to include more CVs than is typically possible by utilizing conventional metadynamics, while ensuring that the bias converged to the well-tempered form in the long time limit. The CVs that were used in this study were: (1) distance between protein center of mass and the surface, (2) alpha beta backbone dihedral parameter, (3) peptide radius of gyration, (4) antiparallel beta sheet content, and (5) coordination number of the Tween 80 molecule around the protein surface.

As the effect of surfactant on protein stability is complex, employing a wide spectrum of CVs enabled the authors to study this process over multiple length scales. Different multi-dimensional free energy surfaces were constructed by reweighting the biased simulations using the method of Tiwary et al., as shown in **Figure 5.5a**.[412] Overall, it is seen that GB1 is unstable at air-water and ice-water interfaces; however, it is stabilized in the presence of a silica surface. With the introduction of Tween 80, this trend reverses, with the GB1 being slightly unstable at the silica-water interface. To further investigate the nature of surfactant-protein binding, an orientation parameter was defined as the ratio between the coordination number of the surfactant tails and surfactant heads around the protein, seen in **Figure 5.5b**, for the silica-water interface. It was concluded that the surfactant acts to stabilize/destabilize a protein by an orientation dependent mechanism, where hydrophilic heads oriented toward the protein stabilize it, and hydrophobic tails oriented toward the protein destabilize it.

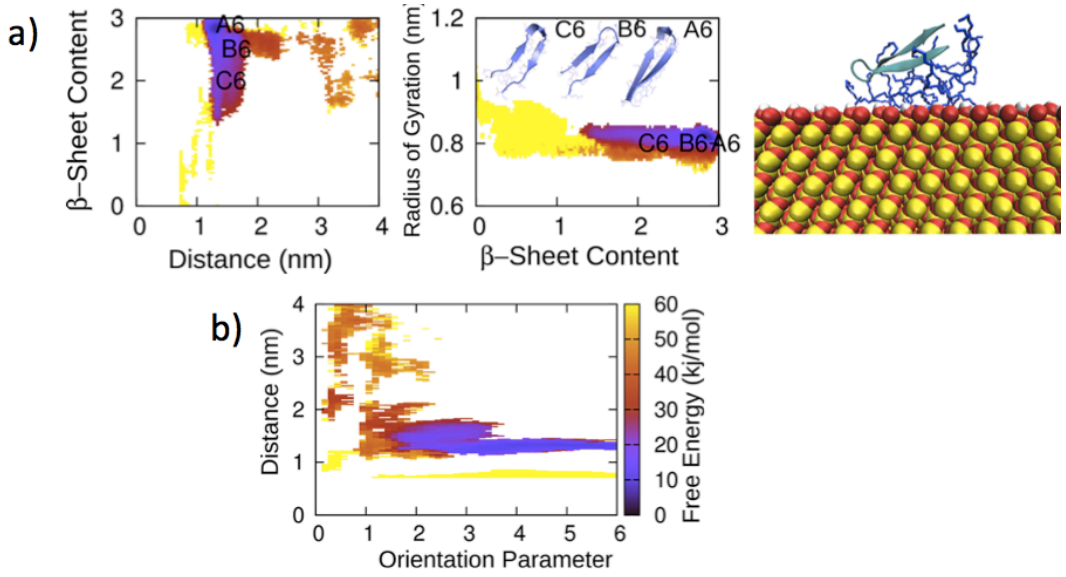


Figure 5.5 (**a**) Free energy surface as a function of beta sheet content vs. distance (left), radius of gyration vs. beta sheet content (center) and snapshot of G1B on silica surface (**b**) Free energy surface of peptide distance vs. orientation parameter for G1B at silica-water interface. Reprinted with permission from ref [415]. Copyright 2018 American Chemical Society.

5.7 Example 4: Protein Simulation Near a Quartz Surface

To compare three variations of MetaD on protein-surface simulations, Prakash et al.[269] simulated a model peptide GGKGG on quartz. Specifically, they used well-tempered (WTMetaD), parallel bias (PBMetaD), and parallel tempering in the well-tempered ensemble (PTMetaD-WTE) to sample the conformational landscape of peptide binding to silica. A second objective of the study was to elucidate the effect that the choice of CV has on the energy landscape and binding free energy of the peptide. All simulations were carried out at a pH of 7.5 and different electrolytes were added to mimic varied experimental conditions.

For the WTMetaD simulations, the CV used was the orthogonal distance between peptide center of mass (COM) and surface. For PTMetaD-WTE, six replicas were simulated between 300-450 K and a bias potential was applied to the distance between the COM of the peptide and the surface. In contrast to the methods above, two CVs were applied for the PBMetaD simulations, namely (1) z-component of the distance of each ion from the surface and (2) distance of the COM of the peptide from the surface.

Based on the type of electrolytic solution used, the authors evaluated the efficiency of the methods based on peptide binding free energy profile and convergence of the simulations. From **Figures 5.6a** and **5.6b**, it is seen that WTMetaD is sufficient for systems containing dilute solutions of electrolytes as well as excess weak biding electrolytes. This is because it yields the same free energy profile as the other two schemes in short simulation times, as many peptide binding/unbinding events occur over the course of the simulation, allowing for correct convergence of the binding free energy profile. In the case of excess competing electrolytes and excess strong binding electrolytes, shown in **Figures 5.6c and 5.6d**, respectively, PBMetaD yielded the most accurate binding free energy profile. This is due to the fact that, along with peptide degrees of freedom, it can also bias ion degrees of freedom which is important as ions have a similar binding affinity to the surface as the peptide, due to which they need to be sampled directly. Overall, in systems like these where ions and peptides have competing slow degrees of freedom,

PBMetaD is shown to be the most accurate sampling scheme as it takes both these effects into account.

This is a choice example of a system where PBMetaD-PF could find application, due to the degeneracy in free energy profiles of each "family" of ions. For example, in the highly concentrated system the PBMetaD simulation was used to bias the distances between the COM of the peptide, 25 sodium, and 26 chloride ions to the surface, for a total of 51 CVs. Alternatively, since the profile of each class of ions is degenerate, a PF simulation would bias only three CVs: (1) the peptide, (2) the family of sodium ions, and (3) the family of chloride ions.

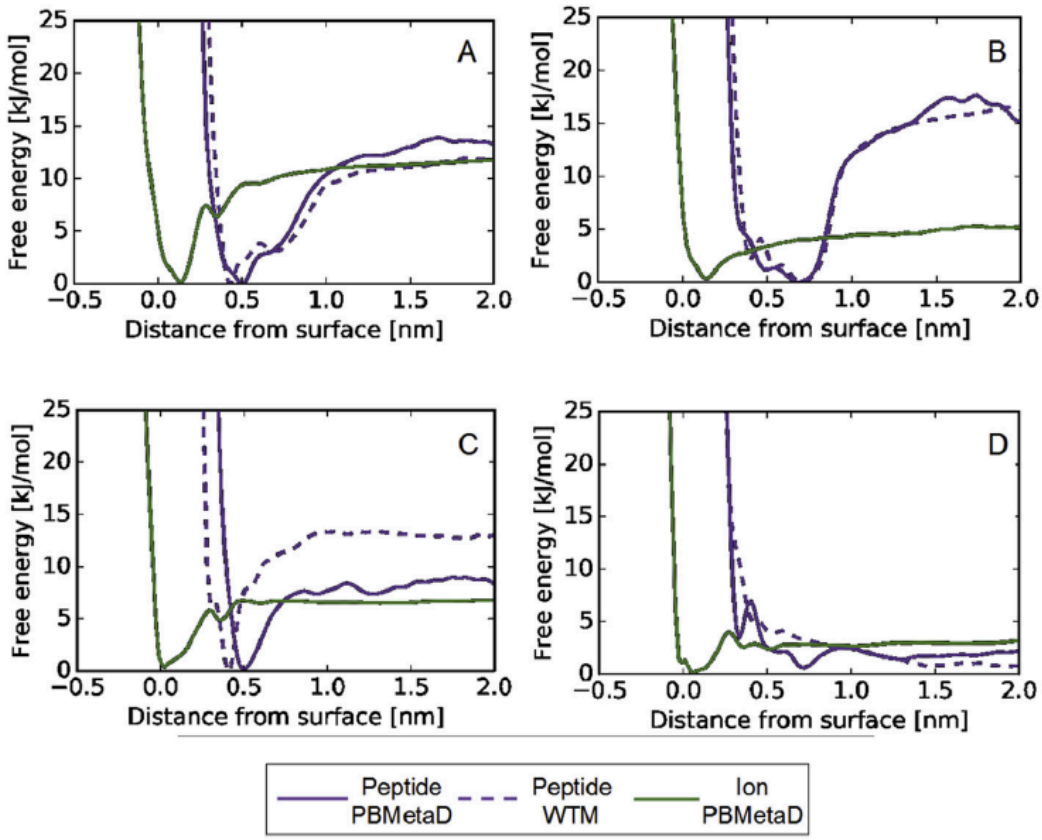


Figure 5.6. Free energy projected onto the center of mass (COM)-surface distance with: (**A**) no electrolyte, (**B**) excess Na0.5+ ions, (**C**) excess Na+ ions, and (**D**) excess Ca2+ ions. Green lines indicate thermally averaged ion binding profiles from PBMetaD simulations. Solid and dotted purple lines indicate peptide-binding profiles from PBMetaD and WTM simulations, respectively. Reprinted from ref [269], with permission from Elsevier.

5.8 Example 5: Sampling Structural Ensembles of Peptides without Forcefields

T. Löhr et al. show that the M&M approach implemented with PBMetaD studies can effectively explore the conformational space of a disordered peptide, with sufficient experimental data, independent of the forcefield.[244] They simulated a disordered peptide with the sequence EGAAWAASS incorporating three sources of experimental data which included: NMR chemical shifts (CS), J-couplings (JC), and Residual Dipolar Couplings (RDCs). They simulated three variations of M&M: unrestrained (no experimental data), using CS and JC data, and using all the data (CS, JC, and RDCs), while also comparing two forcefield/water model combinations (CHARMM22/TIP3P vs AMBER99SB/TIP4P-D). Additionally, they compared the results of an implicit water model (EEF1-SB) using the CHARMM36 forcefield for the unrestrained and full data (CS, JC, and RDCs) cases. This gave a total of eight simulation ensembles to be compared. For the metadynamics calculations, they biased the following CVs: backbone and dihedral angles; E1-S9 Cα–Cα distance; W5 $\chi 1$ and W5 $\chi 2$ angles; and similarities between $\varphi 3$ and $\varphi 6$, as well as $\psi 3$ and $\psi 6$, dihedral angles. Upon reweighting the simulations for the equilibrium distribution, their results indicate that regardless of the forcefield, quality of results from M&M simulations were more similar to the reference data, with the addition of more experimental information into the system. This method allowed for results using two forcefields that were not only in good agreement with each other, but the experimental data as well. To push the limits of this method, they looked at simulations of the protein in implicit solvent and compared them to results from the explicit water models. Compared to the unrestrained models they noted an increase in agreement with experiments, similar to results obtained for the explicit model. A summary of these results is included in **Figure 5.7. Figure 5.7c** demonstrates that in the M&M simulations the results for two different forcefields are indistinguishable from the experimental results, which is not observed without the M&M scheme implemented. The power of this approach is demonstrated by its ability to elucidate an ensemble of data with different priors that share a common solution.

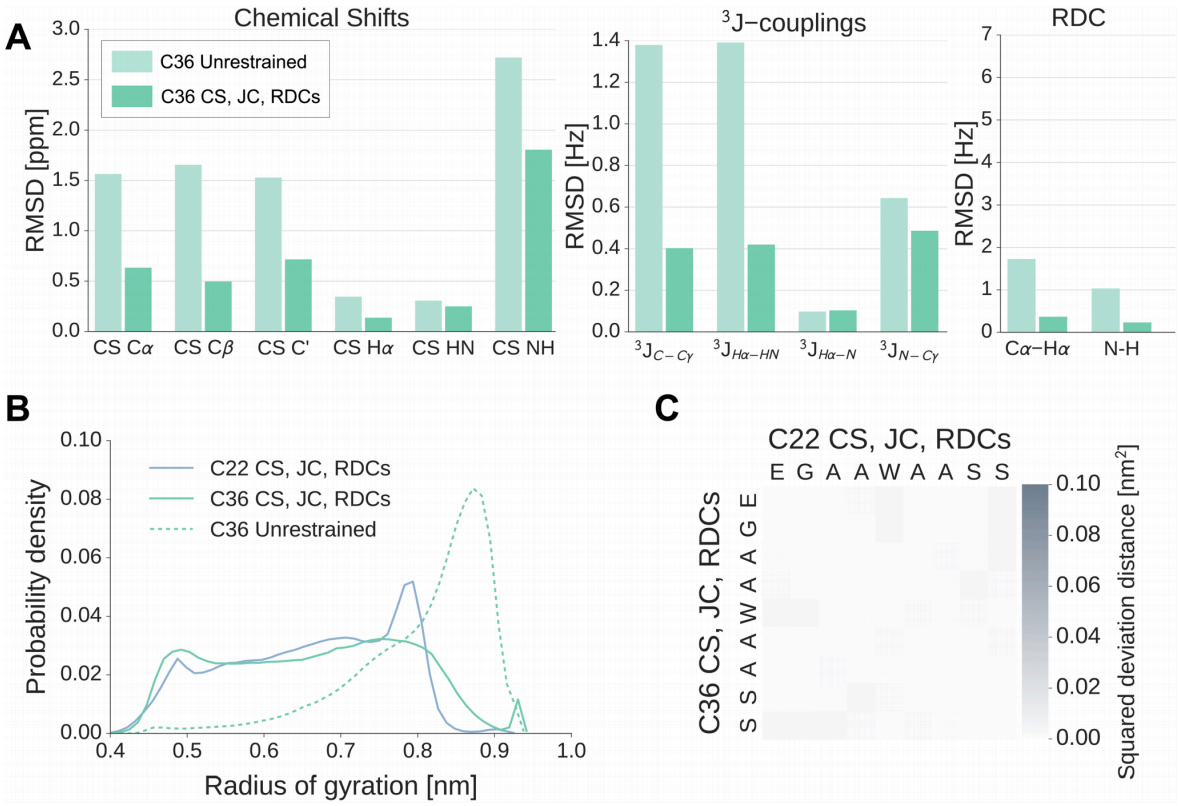


Figure 5.7. Results of M&M simulations comparing results for the explicit (C22) and implicit (C33) solvent models. (**A**) Root-mean-square-deviations (RMSD) of CS, JC, and RDC comparing unrestrained and fully restrained implicit solvent models, using experimental values as reference, (**B**) Probability distributions of the radius of gyration for the fully restrained implicit case (C36), unrestrained implicit case, and restrained explicit case (C22) (**C**) Squared-deviation inter-residue distance matrix between fully restrained explicit, and fully restrained implicit simulations. Reprinted with permission from ref [244].

5.9 Example 6: 7-Particle Lennard-Jones System

Prakash et al. employed PBMetaD-PF to describe the system consisting of 7 LJ particles constrained in two dimensions.[236] All 21 interatomic distances were biased for PBMetaD, and were partitioned into a single family for PBMetaD-PF. For comparison, simulations of the same system were carried out using unbiased MD and WTMetaD, where the CVs biased were the second and third moments of the coordination numbers. The free energy profile for WTMetaD was recovered by reweighting the coordination numbers onto the inter-atomic distances and compared with those obtained from PBMetaD and PBMetaD-PF (**Figure 5.8a**). It is seen that the energy profiles for these schemes were identical, proving that PBMetaD-PF can capture the conformational landscape of a system with high convergence speeds (21 times faster than PBMetaD due to the partitioning

into families). Prior work has shown the 7 LJ particle system exhibits four stable structural minima, and these studies established that the second and third momenta of the coordination number were adequate to differentiate the structures.[416,417] Hence, the inter-atomic distances that were biased in PBMetaD-PF were reweighted for the coordination numbers to remain comparable to prior work. This reweighted FES shown in **Figure 5.8b** correctly captures the four structural minima, consistent with prior results. This example demonstrates that PBMetaD-PF can effectively capture complex free energy landscapes that contain multiple stable minima, such as aggregation of pre-nuclei from solution.

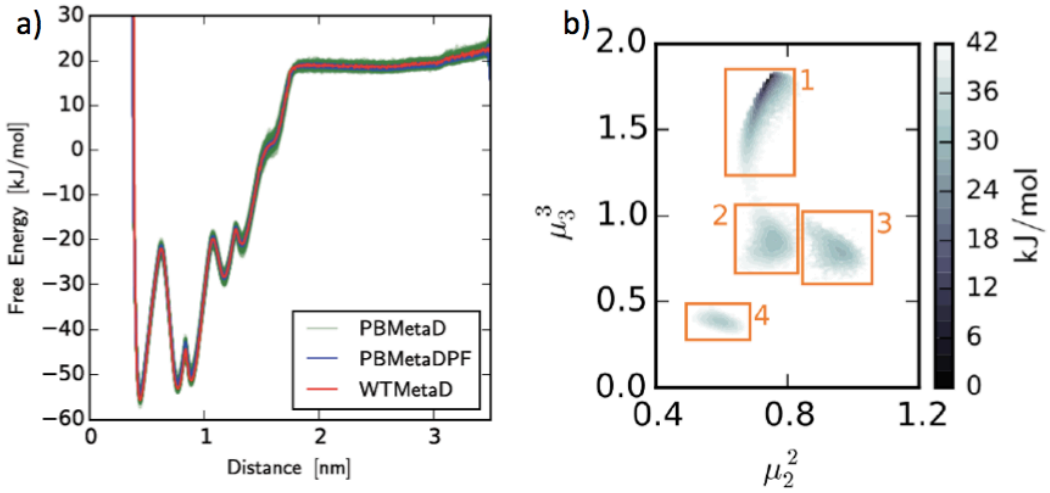


Figure 5.8. (**a**) Mean-aligned free energy profiles of the interatomic distance between LJ particles (**b**) Free energy surface recovered from PBMetaD-PF simulation of the 2D seven-particle LJ system after reweighting for second and third moments of the coordination number. Reprinted with permission from ref [236]. Copyright 2018 American Chemical Society.

5.10 Conclusions

To improve sampling in molecular dynamics simulations, many enhanced sampling methods have been proposed over the years. This review is focused on applications of the PBMetaD method, which has found extensive utility since its introduction in 2015. We review six distinct implementations of PBMetaD, ranging from the effect of ion concentration in peptide binding to protein structure determination in solution. An important, unexplored avenue is the application of PBMetaD-PF to study the driving forces of multiple peptide aggregation in solution. Overall, the main benefit of the method is its versatility and scalability to large numbers of CVs.

Additionally, although the method does not yet have a formal proof of how to reweight the bias potential, various reweighting schemes have been employed using both the rigorous, on-the-fly method of Tiwary and Parrinello[412] as well as the convenient Torrie-Valleu[383] style reweighting introduced by Branduardi et al.[401]. Whereas the former was a convenient ansatz and proved to give physically meaningful results in the study of Prakash et al.[411], the latter method was shown to exactly converge to the underlying free energy (of unbiased variables) for the limited case of LJ clusters.[236] Future work should further explore the application of the Tiwary-Parrinello estimator and compare the convergence efficiency of different types of CVs against the more ad hoc approach. Another frontier area for biasing large numbers of CVs is the case when there are largely different relevant energy scales. It has been long established that the adjustable "bias factor" in the well-tempered variant of MetaD should somewhat correlate to the barriers on the reduced-dimension free energy landscape (c.f., Figure 1 in the original WTM paper[86]), however the case is less clear when using PBMetaD to bias multiple CVs that could benefit from substantially different bias factors. There are ample model systems from which the convergence properties of these instances could be explored. Finally, this high dimensional sampling scheme still suffers from a general problem, which is the proper selection of CVs. Emerging methods for learning CVs such as the SGOOP protocol of Tiwary and Berne may be of use here.[418] Enhanced sampling methods have an important role to play in the molecular simulators toolbox, and the coming years will determine the relative importance of methods that survey many low-dimensional CVs such as PBMetaD.

5.11 Acknowledgements

This work was supported in part by NSF award MCB-1715123 and NIH award 1R21DE026959-01.

Chapter 6
Amino Acid Substitution of Distal and Active Site Residues Reduce Product Inhibition of E1 Endocellulase from *Acidothermus Cellulolyticus*[5]

6.1 Abstract

Cellulases are largely afflicted by inhibition from their reaction products, especially at high substrate loading, which represents a major challenge for biomass processing. This challenge was overcome for endoglucanase 1 (E1) from Acidothermus cellulolyticus by identifying a large conformational change involving distal residues upon binding cellobiose. Having introduced alanine substitutions at each of these residues, we identified several mutations that reduced cellobiose inhibition of E1, including W212A, W213A, Q247A, W249A, and F250A. One of the mutations (W212A) resulted in a 47-fold decrease in binding affinity of cellobiose as well as a five-fold increase in the k_{cat}. The mutation further increased E1 activity on Avicel and dilute-acid treated corn stover and enhanced its productivity at high substrate loadings. These findings were corroborated by funnel metadynamics, which showed that the W212A substitution led to reduced affinity for cellobiose in the +1 and +2 binding sites due to re-arrangement of key cellobiose-binding residues.

[5] This chapter was reproduced with permission from S. Summers, S. Alamdari, C. J. Kraft, R. Brunecky, J. Pfaendtner, J. L. Karr. Substitution of Distal and Active Site Residues Reduces Product Inhibition of E1 from *Acidothermus Cellulolyticus*. Protein Engineering Design & Selection. (In Press) Copyright 2021 Oxford University Press.

6.2 Introduction

Over the last several decades, numerous advances have improved the viability of enzymatic processes for generating renewable biofuels from cellulosic materials. However, the high cost of enzymes required to effectively breakdown cellulose to glucose remains a major bottleneck for the commercialization of such processes.[419] Although many factors contribute to the cost of these enzymes,[420,421] one factor that is prominent in such processes is the inhibition of cellulases by hydrolysate. Specifically, the hydrolysis products cellobiose, and less so glucose,[422] are notorious inhibitors of cellulase enzymes, especially at high substrate loadings.[423,424] As a result of product inhibition by cellobiose, fresh enzyme must be fed intermittently to the reaction, resulting in the need for high enzyme loadings. Furthermore, cellobiose inhibition prevents the recycling of enzyme in the reaction, which is necessary to keep the cost of the enzymes low at commercial scale.[425]

To overcome the challenge of product inhibition and enable enzyme recycling in biorefining processes, several approaches have been investigated. A common approach has involved the removal of hydrolysate, including cellobiose, during biomass conversion via membrane filtration.[426] In one example, Gan and co-workers[427] showed the conversion of cellulose could be increased 1.5 times in a continuous process with simultaneous filtration of hydrolysate compared to in a batch process. Although promising, membrane fouling remains a major challenge in such approaches as is the requirement for large reactors to enable high cellulose conversions.[428] Another conventional approach to alleviate hydrolysate inhibition of cellulases entails the addition of excess β-glucosidase,[429] the enzyme that converts cellobiose to glucose. However, β-glucosidases are also strongly inhibited by glucose,[423] thereby creating an additional challenge that has stymied progress in this field.

Due to the limitations of current approaches, recent efforts to overcome hydrolysate inhibition have shifted to focus on engineering hydrolysate-resistant cellulases. The susceptibility of cellulases to hydrolysate may be reduced through substituting amino acids at residue positions involved in the ligand binding interface.[421,430–432] The most notable example of this involved the substitution of tyrosine to glycine at position 245 in endocellulase 1 (E1) from Acidothermus cellulolyticus.[433] Substitution of tyrosine 245 to glycine in the active site of E1 resulted in dramatic (15-fold) reduction in the affinity of cellobiose for E1, which enhanced cellulose conversion. Crystallographic evidence showed that the reduction in affinity of cellobiose for E1 was due to the

loss of the ability of E1 to hydrogen bond with cellobiose at this position. However, such a decrease in affinity with cellobiose frequently coincides with a reduction in the binding of substrate, resulting in an important trade-off.[430,434,435] For example, Atreya and co-workers reported that while mutations in the active site improved the tolerance of Trichoderma reesei Cel7A to cellobiose inhibition, these same mutations resulted in a dramatic decrease in Cel7A activity.[430] We have observed a similar trade-off for E1 where the Y245G mutation results in an increase in Km and thus a reduction in the catalytic efficiency of the enzyme. To date, the rational engineering of cellulase variants that maintain high substrate affinity while showing reduced cellobiose inhibition has remained elusive.

The overall aim of this work was to identify amino acid substitutions in E1 that reduce cellobiose inhibition and improve enzyme activity relative to the Y245G variant. To identify such amino acid substitutions, 1H-15N HSQC NMR was used to discover residues in E1 that were perturbed as a result of binding by cellobiose. Having observed that binding of cellobiose induced a conformational change in E1, we systematically mutated residues that were perturbed to alanine to understand how each of these residues contributed to cellobiose binding. These findings led to the identification of multiple residues that when mutated to alanine had large effects on cellobiose binding and inhibition of E1. Interestingly, while some of these residues were already known to have a role in product binding, other residues whose role in product binding was less clear were also identified. Discovery of a variant with reduced product inhibition and increased catalytic turnover prompted the use funnel metadynamics (FM) to investigate the effect of amino acid substitutions on cellobiose binding. Finally, we investigated the utility of this variant by testing its activity on Avicel and dilute-acid treated corn stover (DACS) as well as modeling productivity at high substrate loading. Our findings are significant given we have shown it is feasible to engineer cellulases with reduced product inhibition and improved catalytic turnover frequency via mutations in a distal ring

6.3 Methods

Mutagenesis and Preparation of E1

Alanine-substituted E1 variants were generated using the QuikChange (Agilent) multi-site mutagenesis kit and confirmed via sequencing. Plasmid DNA with the desired mutations was transformed into BL21(DE3*) Escherichia coli*. Cells were grown in Lysogeny broth at 37°C

until an OD_{600} of 0.5-0.8 after which the cells were induced with isopropyl β-D-1-thiogalactopyranoside overnight at 16°C. E1 variants were purified by hydrophobic and anion-exchange chromatography, followed by heating to 65°C to remove contaminant proteins.[436]

^{1}H-^{15}N HSQC NMR

We previously analyzed the chemical shift perturbations from the interaction of wild-type and Y245G E1 with cellobiose by 1H-15N HSQC NMR.[437] Briefly, 1H-15N HSQC NMR spectra were collected using ~50-200 μM 15N-labeled wild-type and Y245G E1 in 10 mM citrate buffer, pH 5.0 with 5 v/v% D2O at 50°C. Cellobiose was titrated into the solution containing E1 at concentrations ranging from 0-0.11 M. The dissociation constant for residues that were significantly perturbed by the presence of cellobiose were estimated from chemical shift perturbations using a single-site binding model that accounted for reversible binding with fast exchange.

Screening Cellobiose Inhibition of E1 in Cell Lysate

Cells containing E1 variants were grown and induced in 5 mL cultures as described above. Cells were harvested at 16,000 x g for 10 min and the resulting cell pellets were re-suspended in 1 mL of BugBuster® Master Mix solution to lyse the cells. Cell samples were gently shaken for 10-20 min at room temperature, and then spun at 16,000 x g for 20 min to separate the insoluble fraction. E1 activity was assayed by mixing 40 μL of soluble lysate with 0.5 mM 4-nitrophenyl β-D-cellobioside (pNPC), cellobiose (0, 0.03, 0.1 M), and buffer (50 mM citrate, pH 5.0), resulting in a final reaction volume of 400 μL. Reactions were incubated at 50°C while removing 50 μL aliquots at various time points. The reaction in each of the aliquots was quenched by adding 50 μL of 0.5 M sodium carbonate (pH 11.5). The concentration of the product, 4-nitrophenolate, in the aliquots was measured spectroscopically at 405 nm. Lysate reactions at each condition were measured in triplicate using the same soluble fraction sample.

Determination of Kinetic and Inhibition Constants for E1

Michaelis-Menten kinetic constants for E1 were determined by monitoring the hydrolysis of pNPC (0-9 mM) in 50 mM citrate buffer (pH 5.0) at 50 °C using 0.25 μM of enzyme. Inhibition by cellobiose was assessed using 0, 22, 60, and 100 mM cellobiose, which was added to the solution before initiating the reaction with E1. The resulting Michaelis-Menten curves were fit to

a competitive inhibition model through nonlinear regression analysis using JMP Pro 14 to extract values for kcat, Km, and KI.

Molecular Simulation Details

Molecular simulations were performed on the wild-type as well as Y245G and W212A E1 systems in their holo form using GROMACS2018.3[72]. The structure of E1 was obtained from Protein Data Bank (accession code: 1ECE)[438], whereas the structures for the Y245G and W212A E1 were generated by *in silico* point mutations. All systems were simulated using the Amber 14SB forcefield[439] in an 9.0 x 9.0 x 9.5 nm^3 simulation box containing approximately 22,000 TIP3P water molecules and 11 sodium counter ions, and 2 disulfide bonds between pairs 34/120 and 168/171. A cellobiose ligand was generated from the cellotetraose bound to the +1 and +2 product binding sites in 1ECE, and was modeled using the GLYCAM_06j-1[440] carbohydrate parameter set.

Systems were minimized using a steepest descent method over a maximum of 10,000 steps. Each system was then equilibrated to 300 K in NVT over 1 ns and to 1 bar in NPT over 1 ns. In all simulations, a Parrinello-Rahman barostat[82] (τ = 1.0 ps) and a global stochastic thermostat [81](τ = 0.1 ps) were used. A 1.2 nm cutoff was used for calculating both short range van der Waals interactions and particle-mesh Ewald electrostatics. Periodic boundaries were applied in all directions. Hydrogen bonds were constrained in all simulations by the LINCS[80] algorithm, and a 2 fs timestep was used in all simulations.

Following equilibration, metadynamics simulations were carried out in the NVT ensemble at 300 K using Plumed 2.3[73] and applying the same thermostat and parameters used in equilibration. Using the FM Advanced Protocol[441] a funnel restraint was applied to each system with the following parameters: Z_{cc} was set to 2.5 nm, α was 0.55 radians, and a value of 0.1 nm was used for R_{cyl}, and a harmonic restraint to keep the ligand within a funnel-shaped region was used with a kappa value of 50,000 kJ/mol/nm^2. A 2D well-tempered metadynamics potential was applied to each system biasing two collective variables (CVs): 1) The z-axis of the funnel and 2) the distance between the center of mass of the ligand and residue 213, using a sigma of 0.05 nm for both variables. This sufficiently described binding of cellobiose to the product binding site, which is located between residues 245 and 213 in wild-type E1. For metadynamics an initial gaussian height of 2.5 kJ/mol, a bias factor of 8, and a deposition rate of 1hill/2 ps was used. Upper

and lower walls were placed to constrain sampling within the region of the funnel, and an additional lower wall was placed on the wild-type system at 0.7 nm on collective variable 2 to prevent simulation errors. The use of walls in the FM scheme to aid in sampling these systems has previously been described in detail.[441] RMSD of the backbone α-carbons was monitored during the metadynamics run to confirm that the bias added from the funnel restraint did not lead to any conformational changes (**Figure C-S1**). All corresponding PLUMED input files required to reproduce the results reported in this paper are available on PLUMED-NEST (https://www.plumed-nest.org/) the public repository of the PLUMED consortium (plumID:21.012).[89]

Systems were simulated for 400 ns for WT E1 and 300 ns for both Y245G and W212A E1 to allow for convergence of free energy. Systems were characterized as being converged when ΔG fluctuations were on the order of thermal fluctuations ($\pm K_BT$) while the system was exploring the CV space without additional hills being deposited. This criterion was met for all systems over the at least the last 30% of each simulation and is given in **Figure C-S2**. Free energy was reweighted along the z-axis of the funnel (describing solvation) and the distance perpendicular to the funnel z-axis (describing location in the binding pocket) for each system.[442]

Conversion of Avicel and DACS

Progress curves for the conversion of Avicel and DACS were generated for wild-type, Y245G, and W212A E1 using 1.5% wt Avicel and DACS at 50°C in 20 mM sodium acetate buffer (pH 5.0) with 100 mM NaCl. For these studies, a low (0.5 mg/g substrate) and high (2 mg/g substrate) concentration of E1 was used. To promote complete conversion from cellulose to monomeric glucose, exocellulase Cel7A from *Trichoderma reesei* (15 mg enzyme/g cellulose) and *Aspergillus niger β*-glucosidase (0.5 mg enzyme/g cellulose) from Megazyme were added to the reaction mixture with E1. Prior to the reaction, *β*-glucosidase was de-salted using a Hi-Trap 26/10 de-salting column to remove ammonium sulfate. Hydrolysis of Avicel and DACS was monitored for 4 days. At various time points throughout the 4-day reaction, aliquots were removed from the reaction and boiled for 15 min to inactivate the enzymes. Samples were the filtered using a 0.45 μm Acrodisc syringe filter and the sugar content released during hydrolysis was analyzed via HPLC using an Agilent 1100 HPLC system equipped with a BioRad Aminex HPX-87H 300 mm x 7.8 mm column heated to 55°C. A 20 μL injection volume as well as a constant flow of 0.6

mL/min was used with 0.1 M sulfuric acid in water as the mobile phase. The concentration of glucose, xylose, and cellobiose were determined using independent standard curves and converted to total anhydrous glucan equivalents. All experiments were performed in triplicate and the resulting extents of conversion are shown as percent glucan converted.

Modeling of Theoretical E1 Productivity

Theoretical productivity curves of E1 variants were generated using the integrated Michaelis-Menten equation (**Equation 4.1**) for linear competitive inhibition (assuming no product is initially present):[443]

$$t = \frac{1}{V_{max}}\left\{K_m\left(\frac{S_{t=0}}{K_I}+1\right)\ln\left[\frac{S_t}{S_{t=0}}\right]+\left(1-\frac{K_m}{K_I}\right)(S_t - S_{t=0})\right\} \quad 6.1$$

The Km and KI values used were derived from the initial velocity experiments with pNPC and cellobiose. Vmax was calculated using kcat values derived from Michaelis-Menten analysis as well as an industrially relevant enzyme concentration of 0.13 mg/g substrate.[419] For substrate concentration, a value of 0.23 mM, which corresponds to a high substrate loading from the literature, was used.[424] Error for the productivity curves was calculated using the errors from kinetic and inhibition parameters. Productivity curves were generated by solving the implicit function in Mathematica for the substrate concentration (S_t) over a time period of 20 h.

6.4 Results and Discussion

Structural Analysis of E1-Cellobiose Interaction

We have previously used 1H-15N HSQC NMR to characterize the impact of active site mutations in wild-type (WT) and Y245G E1 on the binding of cellobiose.[437] In these studies, cellobiose was used as substrate analogue to elucidate how mutations altered the affinity of E1 for cellulose relative to 1-butyl-3-methylimidazolium chloride, an ionic liquid that has been exhaustively studied as a possible solvent for biomass processing given its capacity to solubilize cellulose. In this work, we investigated the extent to which the binding of cellobiose induces structural changes that are outside the active site of E1. Interestingly, analysis of chemical shift perturbations from 1H-15N HSQC NMR revealed a significant number of distal residues with large perturbations in the structure of WT E1 (**Figure 6.1A** and **Figure C-S3**, which shows the near complete 1H-15N HSQC spectra for WT E1 in the presence of 0-30 mM cellobiose). In total,

we observed 29 residues that were significantly perturbed, which indicated a large conformational change in the enzyme. These residues formed a ring along the surface of WT E1 that involved the helix interface between alpha helices 7 and 8 as well as helix 6 (**Figure 6.1B** and **C**). The ring further initiates and ends within the active site near where cellobiose binds. Given the connectivity of the ring with the active site, this finding explains how cellobiose binding may induce a conformational change in the ring. To confirm perturbations of the residues in the ring were coordinated, the dissociation constant for cellobiose binding was determined from the chemical shifts of several distal residues, including F283, T286, L287, Q293 and W295 (**Figure C-S4**). Notably, the Kd values for E1-cellobiose binding based on the chemical shifts of this subset of residues were similar (2.61-6.31 mM), suggesting the perturbation of these residues was indeed connected. Additionally, the Kd values based on these residues was nearly the same as the Kd for E1-cellobiose binding based on the chemical shift of residue W213 (3.56 ± 0.002 mM), which directly interacts with cellobiose in the active site. When fit with the average Kd of the subset of distal residues, there was excellent agreement between the chemical shifts for the remaining residues and the fit of the binding isotherms. This further confirmed the structural changes in the ring were indeed coordinated and that these changes were the result of a large conformational change in WT E1 upon cellobiose binding.

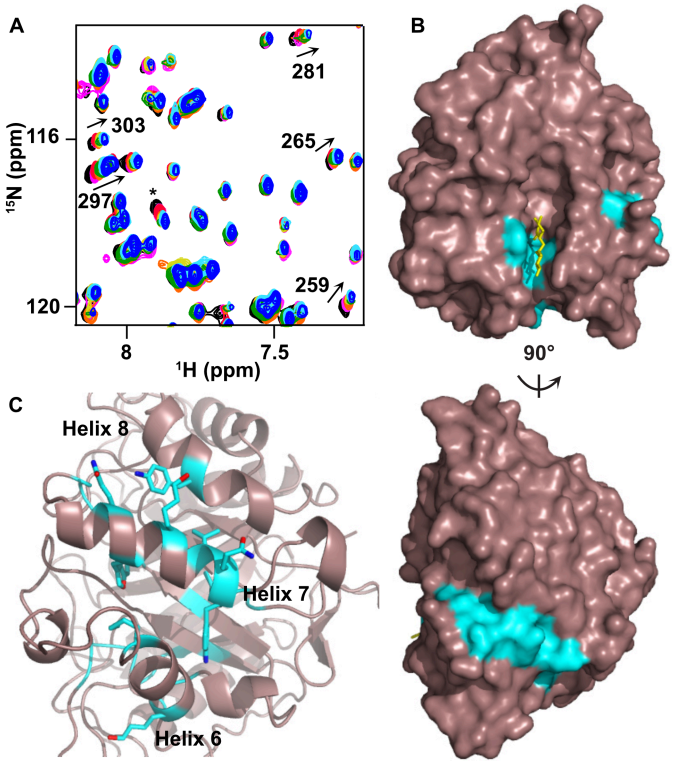


Figure 6.1. ^{1}H–^{15}N HSQC NMR analysis of the interaction of WT E1 and cellobiose. (**A**) Overlay of partial ^{1}H–^{15}N HSQC spectra of WT E1 in the presence of 0 (black), 1 (pink), 2 (red), 6 (orange), 10 (yellow), 15 (green), 21 (cyan), and 30 (blue) mM cellobiose. The partial overlay of the spectra show select distal residues in the ring region that are significantly perturbed as defined by having a $\Delta\delta > 1\sigma$ from average. The labels and arrows indicate the residue number and direction of the chemical shift perturbations, respectively. Additionally, the asterisk indicates a peak that was significantly perturbed, but for which there is no residue assignment. (**B**) Projection of residues with significant chemical shift perturbations on the surface of WT E1 (shown in cyan). The structure of cellobiose bound in the active site of WT E1 is shown in stick representation (yellow). (**C**) A close-up view of helices 6-8 that were part of the ring region.

Although related studies have examined how cellobiose affects the structure of other cellulases,[435,444] this is the first report of a binding-induced conformational change outside the active site for an endoglucanase. Since substrate binding to the -1 to +1 sites in the active site can lead to small structural changes in cellulases,[445–447] we verified that cellobiose was bound to the consensus product binding position (+1 to +2 sites). Consistent with this finding, the chemical shifts of residues E282, W319, and H116 that are involved in substrate binding were negligible. Additionally, the residues adjacent to the ring that is perturbed by cellobiose binding have relatively high B-factors compared to the rest of WT E1. This suggests the ring may have different conformers in the crystal structure of WT E1 when bound to cellobiose and cellotetraose, which is consistent with our NMR findings. In contrast with WT E1, the relative magnitude of the chemical shift perturbations in the critical ring were considerably smaller for Y245G E1 at high

cellobiose concentrations (**Figure C-S5**). This is notable since Y245G E1 is significantly less inhibited by cellobiose compared to WT E1 as mentioned above. Furthermore, the apparent Kd values for the residues that shifted were dissimilar, suggesting the lack of a coordinated structural change. Based on our findings, we hypothesized it may be possible to inhibit cellobiose binding to E1 through amino acid substitutions that disrupt the connectivity of the critical ring. Specifically, if the conformational change in the ring stabilizes the binding of cellobiose, disrupting this change may reduce the affinity of cellobiose for the active site of E1.

Impact of Mutation of Loop Region on Cellobiose Inhibition

Given our structural observations, we sought to determine if the mutation of one or more of the residues in the ring could block cellobiose binding without adversely impacting E1 activity. This question was investigated by mutating each residue in the ring to alanine and measuring the binding of cellobiose as well as the activity of E1 for each mutation. Such an approach is analogous to the use of alanine scanning to determine the contribution of important residues to the stability of proteins as well as protein-protein interactions.[448] Notably, alanine is classically used in scanning mutagenesis experiments because it is small, has minimal effects on protein structure, and does not interact strongly with other residues. To rapidly screen the effect of alanine substitutions on cellobiose binding and inhibition, the relative activity of E1 in the lysate of *E. coli* cells was measured using pNPC as a substrate. Variants were assessed by comparing their relative activities in the presence of 0, 0.03, and 0.1 M cellobiose to that of WT and Y245G E1. In addition to screening alanine variants for the 29 residues that were structurally perturbed, alanine variants for 6 other residues that were unassigned in the $^{1}H-^{15}N$ HSQC NMR spectra and presumed to be part of the ring were also screened.

Of the 35 alanine variants screened, enzymatic activity was detected for all the variants except five, which involved substitution of highly conserved and buried residues. However, five of the variants, including W212A, W213A, Q247A, W249A, and F250A E1, exhibited reduced cellobiose inhibition (**Figure 6.2**). Specifically, E1 with each of these substitutions was significantly more active (based on 95% confidence) in the presence of 0.03 and 0.1 M cellobiose relative to WT E1. It was perhaps not surprising that the replacement of residues 212, 213 and 247 with alanine led to reduced cellobiose inhibition given they are located at the ends of the ring in the active site. Moreover, residues W213 and Q247 have previously been shown to stabilize

cellobiose binding through stacking and hydrogen-bonding interactions.[433,438] Although also in the active site, it is less obvious how replacement of W212 with alanine disrupts cellobiose binding given it is unclear if and how it interacts with cellobiose. However, even more interesting was the impact of the replacement of W249 and F250 with alanine given that these residues are located outside the active site and are largely non-solvent accessible. Unlike the residues in the active site, the replacement of these residues may disrupt cellobiose binding indirectly through altering the conformational change in the critical ring. Despite being outside the active site and thus distal, it is interesting to note that the replacement of residues furthest from the active site with alanine had a negligible impact on product inhibition. As such, proximity to the active site may be critical for disrupting the conformational change of the ring, although this remains to be confirmed through detail structural studies. While replacement of P260 with alanine also resulted in reduced cellobiose inhibition, this was presumed a false positive due to unusually large variability in activity measurements.

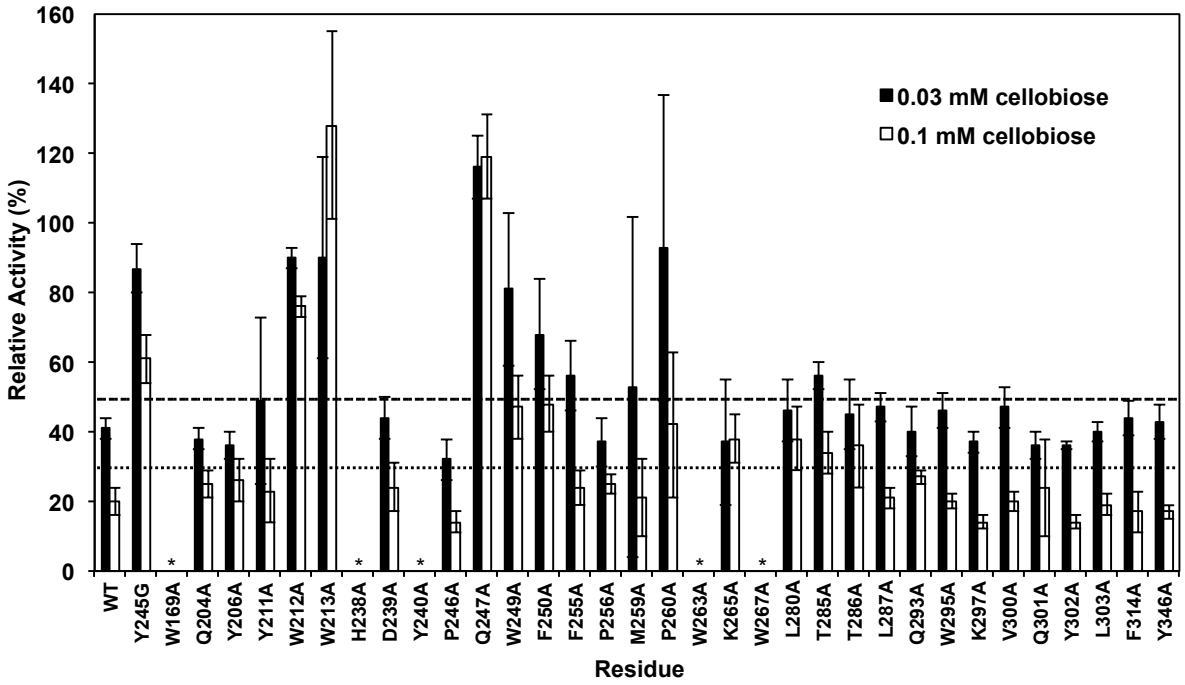


Figure 6.2. E1 lysate screen for cellobiose inhibition of alanine-substituted variants. Relative activity was determined by normalizing the activity of each variant in the presence of 0.03 (black bars) and 0.1 (white bars) M cellobiose to that the activity of each variant in the absence of cellobiose. Error bars represent the standard deviation from the mean for three replicate measurements. Horizontal lines with long and short dashes represent the cutoff for determining significant difference (based on 95% confidence) in the presence of 0.03 and 0.1 M cellobiose, respectively, from the activity of WT E1. Asterisks indicate substitutions that resulted in variants with no appreciable activity above background hydrolysis.

The impact of the replacement of W212A, W213A, Q247A, W249A, and F250A on cellobiose inhibition of E1 was characterized further by kinetic assays using the purified variants. Of specific interest was comparing the kinetic parameters k_{cat} and K_m as well as the inhibitory constant K_I for cellobiose for the variants to that for WT as well as Y245G E1. To determine k_{cat}, K_m, and K_I, the initial rate for E1 with varying substrate (pNPC) concentrations was measured in the presence of 0, 22, 60, and 100 mM cellobiose. Due to unusual kinetic behavior (e.g., non-linear competitive inhibition) and increased aggregation, the k_{cat}, K_m, and K_I could not be determined for W213A, Q247A, W249A, and F250A E1. However, we were able to determine k_{cat}, K_m, and K_I for the W212A variant, which exhibited linear competitive inhibition in the presence of cellobiose. Remarkably, the K_I for the W212A variant (135.3 ± 4.5 mM) was significantly higher than that for WT (2.87 ± 0.09 mM) and Y245G E1 (69.8 ± 3.9 mM), indicating weaker inhibition by cellobiose (**Table 4-1**). Furthermore, the k_{cat} for the W212A variant was approximately five-fold that for WT E1 and two-fold that of Y245G E1, although the K_m for the W212A variant was higher than for WT and Y245G E1. Because of the higher K_m, the overall catalytic efficiency of the W212A variant was slightly lower than for WT and Y245G E1, yet this decrease is compensated for by the reduction in inhibition. The decrease in K_I and increase in k_{cat} for W212A E1 was extremely encouraging for identifying amino acid substitutions that improved the overall productivity of E1 in breaking down cellulose. **Figure C-S6** in the Supporting Information shows the Michaelis-Menten plots for WT, Y245G, and W212A E1 for the different concentrations of cellobiose used in the kinetic analysis. Given the presumption that P260A variant was a false positive, this variant was excluded from this characterization.

Table 6-1. Michaelis-Menten and inhibition parameters for WT, Y245G, and W212A E1 using pNPC as a substrate. The inhibition constant was determined for competitive inhibition using a standard reversible competitive inhibition model. All values represent the mean ± standard error.

Variant	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m ($mM^{-1}\ s^{-1}$)	K_I (mM)
WT	1.11 ± 0.02	0.19 ± 0.02	5.84 ± 0.62	2.87 ± 0.09
Y245G	2.55 ± 0.18	1.57 ± 0.12	1.62 ± 0.17	69.8 ± 3.9
W212A	5.29 ± 0.10	6.01 ± 0.21	0.88 ± 0.03	135.3 ± 4.5

Using the values for k_{cat}, K_m, and K_I from **Table 4-1**, we further modeled the productivity curves for W212A, WT, and Y245G E1 at high substrate loadings. By modeling productivity, we sought to understand if productivity of W212 E1 would be higher than WT and Y245G E1 under industrial conditions despite the reduction in substrate affinity. For this analysis, a concentration of 20% wt of lignocellulose was used to simulate conditions in which high concentrations of hydrolysis products would accumulate over time. This concentration of lignocellulose, which amounted to 8% wt (or 0.23 M) cellulose, was also used since it represents the lower limit for the concentration of solids for economical biomass yields.[424] Additionally, the concentration of E1 used for this analysis was 20 mg enzyme/g cellulose, which was significantly higher than that used for the abovementioned kinetic assays.

As shown in **Figure C-S7**, the theoretical productivity of W212A E1 was greater than that of WT and Y245G E1. This was evident by comparing the time to reach 100% conversion of the lignocellulose, which was the shortest for the W212A variant. Specifically, in the case of W212A, complete conversion of lignocellulose was reached after only 4 h whereas complete conversion for WT and Y245G E1 required 18 and 8 h, respectively. Interestingly, the 4.5-fold difference in the time to reach complete conversion for W212A and WT E1 was nearly the same as the difference in values of k_{cat} for W212A and WT E1 (i.e., the k_{cat} of W212A E1 was 4.7-fold greater than that for WT E1). Similarly, the 2-fold difference in the time to reach complete conversion for W212A and Y245G E1 was nearly the same as the difference in values of k_{cat} for W212A and Y245G E1 (i.e., the k_{cat} of W212A E1 was 2.3-fold greater than that for Y245G E1). This finding suggests that improvements in the k_{cat} of W212A E1 along with the increase in K_I likely had a significant impact on the time to reach complete conversion. It is worth noting that the time to reach complete conversion on lignocellulose substrate would likely be much longer than that predicted here due to greater complexity of the kinetics on an insoluble substrate vs. pNPC, as well as more inhibitors than cellobiose (i.e., lignin) in the hydrolysate. Nonetheless, our results highlight the potential enhancement in productivity of W212A E1 over WT and Y245G E1 under industrially relevant conditions. While such a benefit is clear from this analysis, the economic impact of this benefit on biomass conversion processes at industrial scale remains to be seen.

Funnel Metadynamics of Cellobiose Binding

To better understand how the W212A substitution reduced cellobiose binding, FM was used to characterize the interaction of cellobiose with WT, Y245G, and W212A E1. In FM, a funnel shaped restraint (**Figure C-S8**) is applied to the ligand, confining the ligand in the path of the binding pocket, while slow degrees of freedom are biased with metadynamics to enhance sampling of enzyme-ligand binding/unbinding events[449] By sampling ligand binding events, details of the binding pose of cellobiose in the active site of each variant could be elucidated. The elucidation of such details, in turn, allowed for differences in the interactions of cellobiose with each variant of E1 to be determined, which was used to also explain differences in affinity. To confirm we were able to accurately capture the interactions between cellobiose and E1, the free energy of cellobiose-E1 binding for each variant ($\Delta G^0_{b\,FM}$) was estimated from the results of FM.[449] As shown in **Table C-S1**, the values of $\Delta G^0_{b\,FM}$ were indeed similar to the experimentally determined free energies ($\Delta G^0_{b_{exp}}$), which were computed from the K_I values reported above.

Analysis of the binding poses of cellobiose with the variants from FM showed that each variant had 2-3 energetic minima, or basins, that contributed to cellobiose binding (**Figure 6.3**). These basins correspond to distinct locations in the active site where cellobiose binding was observed as determined based on the measurement of binding energies. Because these basins were determined based on binding energies, they are referred to hereafter as thermodynamic energetic basins. Basins 1 and 3 describe binding poses near the catalytic active site (sites -1 and -2) whereas basin 2 describes binding near the known product binding site (sites +1 and +2). Interestingly, while basin 2 is the known product binding site from the crystal structure of cellotetraose bound to WT E1, basins 1 and 3 represent previously unreported binding sites. Of the three variants, WT E1 was stabilized in the consensus binding site (basin 2) to the largest extent. Upon binding in basin 2, cellobiose was stabilized through stacking interactions with residues W213 and Y245, hydrogen bonding with Q247, and non-CH/π interactions[450] due to the perpendicular arrangement of W212 (**Figure 6.4**). Conversely, the Y245G substitution led to the loss of a critical carbohydrate-stacking interaction, which weakened cellobiose binding in basin 2, although this was compensated for by an increase in binding to basin 3. The enhanced binding to basin 3 explained the increase in binding affinity of Y245G E1 for cellobiose compared to W212A E1. For W212A E1, cellobiose binding to basin 2 was preserved, although the interaction with basin 2

was weaker than for WT E1. The decrease in binding affinity was presumably the result of the rearrangement of W213, which prevented stacking interactions with cellobiose. Additionally, unlike for Y245G E1, a third energetic basin was not observed for W212A, suggesting cellobiose only bound to basins 1 and 2 for this variant. The differences in the apparent binding of cellobiose to the basins explain the differences in affinity of the three variants for cellobiose. Importantly, the overall structure and dynamics of the active site of the variants was similar as observed with unbiased classical MD simulations. In light of this, the differences in the binding free energy of cellobiose was solely due to the loss of favorable energetic interactions upon mutation, and not differences in active site accessibility.

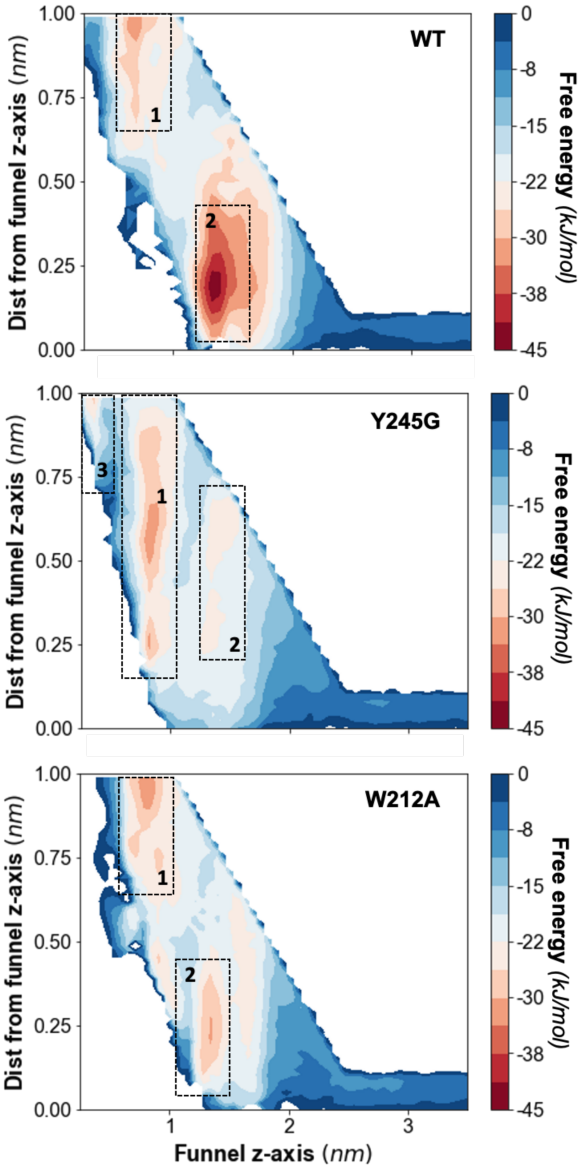


Figure 6.3. Free energy surfaces showing energetic basins in the active site for WT (**top**), Y245G (**middle**), and W212A (**bottom**) E1. The spatial location of the basins is depicted by the dashed squares and labeled according to basin number. Basin 2 for all of the variants represents the consensus product binding site (+1 and +2 sites) in the crystal structure of WT E1 (PDB: 1ECE).

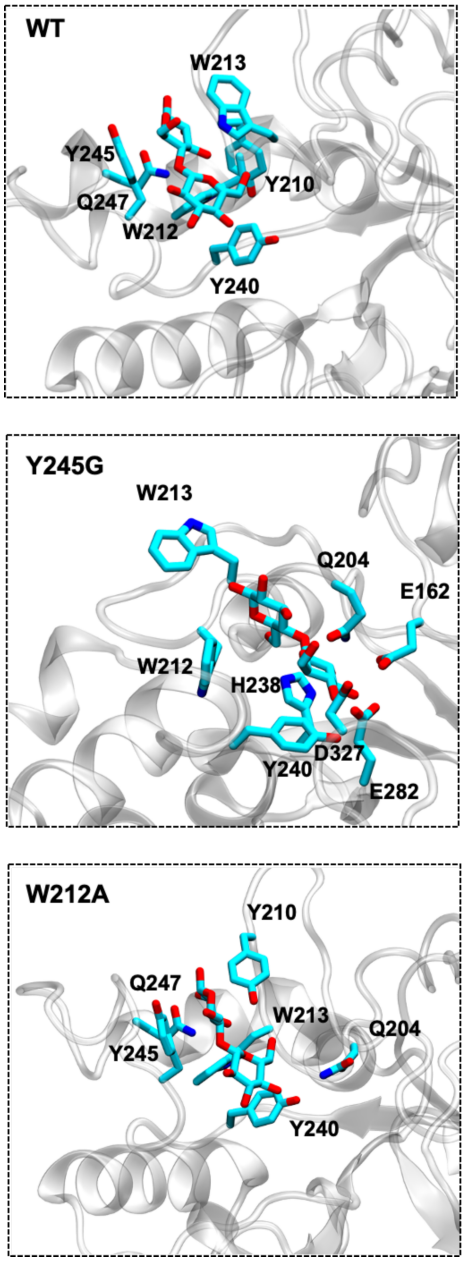


Figure 6.4. Snapshots of cellobiose in basin 2 of WT (**top**), Y245G (**middle**), and W212A (**bottom**) E1. Residues within 4 Å of cellobiose are labeled, and hydrogens and backbone atoms are omitted for clarity. Glycine and valine residues are not included in this visualization.

Impact of Amino Acid Substitutions on Biomass Conversion

Having shown the W212A variant is less susceptible to product inhibition on soluble pNPC, we characterized the activity W212A E1 on the relevant biomass substrates Avicel and DACS. Avicel and DACS are both highly crystalline, insoluble forms of cellulose; however, Avicel is composed of only cellulose, whereas DACS contains lignin in addition to cellulose.[451]

The activity of W212A, Y245G, and WT E1 was tested on both substrates at 1.5% wt substrate and at 0.5 and 2.0 mg enzyme/g substrate enzyme concentrations. To mimic industrial conditions where a cellulase cocktail would be used, the exocellulase Cel7A from *Trichoderma reesei* and *Aspergillus niger* β-glucosidase were also added to the assay. The use of such a cocktail is necessary (and widely used to assay endocellulases) since endocellulases such as E1 primarily serve to create initiation sites for exocellulases and have little activity on insoluble substrates on their own.

As observed with the model substrate pNPC, the W212A variant was significantly more active than Y245G and WT E1 on Avicel (**Figure 6.5**). For comparison, the W212A variant led to ~34% conversion at 2.0 mg enzyme/g substrate after 95 h compared to only ~30 and 27% conversions for Y245G and WT E1, respectively. On DACS, the W212A variant was more active than WT E1, but had similar activity to Y245G E1. Because DACS contains lignin, it is plausible that W212A E1 is inhibited more strongly by lignin than Y245G E1, which would explain the similar activities on DACS. Although not explored here, lignin inhibition of W212A E1 may be overcome via further engineering efforts, including rational modification of enzyme charge as we have shown previously.[452] Ultimately, that W212A E1 was as active or more active than WT and Y245G E1 underscores the potential industrial utility of W212A E1 in commercial biomass conversion processes. Additionally, although the addition of glucosidases can reduce cellobiose inhibition of cellulases, our results show that the W212A mutation still had a beneficial effect.

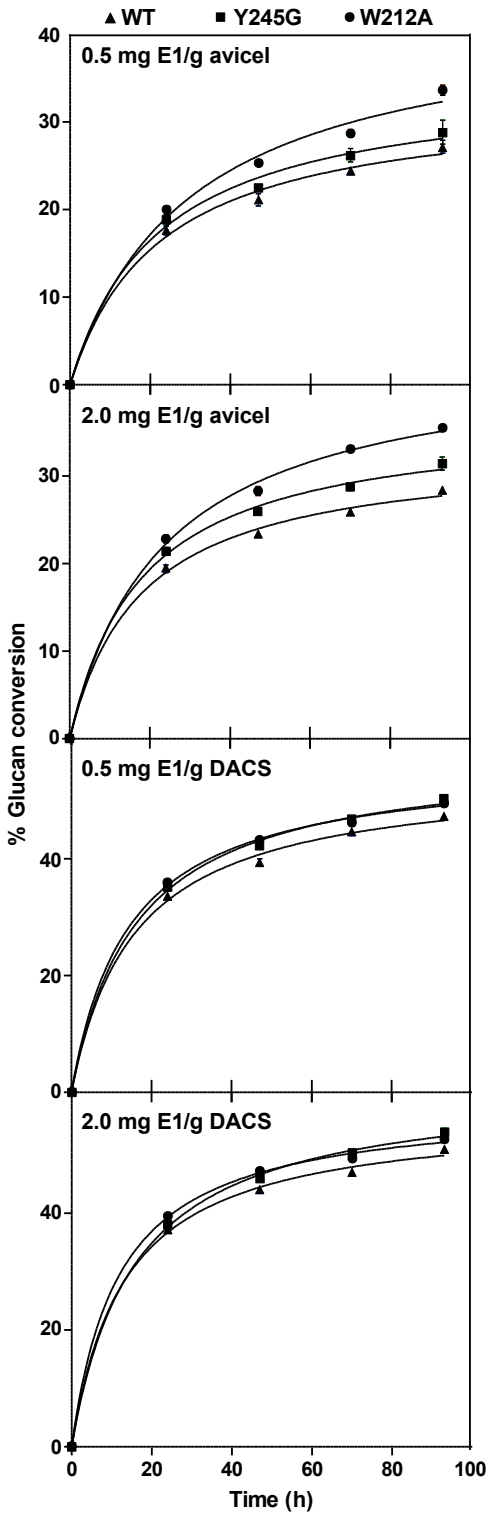


Figure 6.5. Progress curves for the digestion of Avicel and DACS by WT, Y245G, and W212A E1. The progress curves were measured for each variant of E1 using a low (0.5 mg/g substrate) and high (2.0 mg/g substrate) enzyme loading with the addition of Cel7A from T. reesei (15 mg/g) and β-glucosidase from A. niger (0.5 mg/g). Error bars, which are in some cases are smaller than the symbols, represent the mean of three independent reactions. Lines represent the fits to the Michaelis-Menten equation.

6.5 Conclusions

Our findings demonstrate that active site as well as distal mutations in E1 can significantly reduce cellobiose inhibition and thus enhance biomass conversion. Such amino acid substitutions were identified via structural analysis of the binding of cellobiose with E1, which revealed a large conformational change of a ring on the surface of the enzyme. To the best of our knowledge, this is the first evidence of a coordinated conformational change that is induced by cellobiose outside the active site of E1. The W212A substitution was of particular interest since the replacement of tryptophan with alanine led to a dramatic increase in k_{cat} of E1 in addition to weakening the binding of cellobiose. These combined effects translated to increased conversion of Avicel and DACS, as well as greater theoretical productivity at high substrate loadings. Consistent with these findings, FM showed that the W212A substitution led to reduced affinity for cellobiose in the +1 and +2 binding sites due to re-arrangement of key cellobiose-binding residues relative for WT E1. In addition to revealing the structural change in E1 in distal residues, we are not aware of any reports of mutations beyond the cellobiose binding site that enhance activity while reducing product inhibition of cellulases. Ultimately, the understanding from this work provides the framework to develop cellulases with significantly enhanced commercial utility for industrial biomass processing and conversion. Having demonstrated this framework for E1, it is interesting to consider if such understanding can be used to improve the catalytic productivity and thus utility of other enzymes. Of particular interest is exploring how substitutions in distal residues represents a useful approach to engineer other enzymes that suffer from product inhibition.

6.6 Acknowledgements

This work was supported by funding from the National Science Foundation award CBET 1454379. Additionally, S.R.S. was supported by the U.S. Department of Energy (DOE), Office of Science, Office of Workforce Development for Teachers and Scientists, Office of Science Graduate Student Research (SCGSR) program. The SCGSR program is administered by the Oak Ridge Institute for Science and Education (ORISE) for the DOE. ORISE is managed by ORAU under contract number DE-SC0014664. S.A. also acknowledges support from the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1762114. This work was completed in part through the use of advanced computational, storage, and networking

infrastructure provided by the Hyak supercomputer system and funded by the STF at the University of Washington.

Chapter 7
Thermodynamic Basis for Stabilization of Helical Peptoids by Chiral Sidechains[6]

7.1 Abstract

Peptoids are a class of highly customizable biomimetic foldamers that retain properties from both proteins and polymers. It has been shown that peptoids can adopt peptide-like secondary structures through the careful selection of sidechain chemistries, but the underlying conformational landscapes that drive these assemblies at the molecular level remain poorly understood. Given the high flexibility of the peptoid backbone, it is essential that methods to study secondary structure formation possess the requisite sensitivity to discriminate between structurally similar yet energetically distinct microstates. In this work a generalizable simulation scheme is used to robustly sample the complex folding landscape of a polypeptoid backbone and develop a predictive model of the impact of sidechain chemistry on preferential assembly into one of 12 accessible backbone motifs. Using the metadynamics sampling method, three peptoid dodecamers are simulated in water; sarcosine, N-(1-phenylmethyl)glycine (Npm) and S)-N-(1-phenylethyl)glycine (Nspe) – to determine the underlying entropic and energetic impacts of hydrophobic and chiral peptoid sidechains on secondary structure formation. Our results demonstrate that conformational entropic gains from a bulky chiral sidechain (Nspe) promotes overall *cis*- backbone isomerization. However, overall assembly of Nspe into a helix is found to be entropically unfavorable, indicating the overall assembly of Nspe peptoids into polyproline type-I helices in water is enthalpically driven, benefitting from an entropic gain for isomerization and steric strain due to the presence of a chiral center.

[6] This chapter is a draft of an article titled "Thermodynamic Basis for Stabilization of Helical Peptoids by Chiral Sidechains" which has been submitted for review, and was reproduced with permission from S. Alamdari, K. Torkelson, X. Wang, C. L. Chen, A. L. Ferguson, and J. Pfaendtner.

7.2 Introduction

Natural sequence defined polymers such as proteins and DNA have inspired a growing field of synthetic, self-assembling, structured, and functional engineered molecular systems to advance technology [6,453,454]. Well-controlled polymers with high structural and functional complexity hold potential for a wide range of applications ranging from materials science to advanced biomimetics [455]. Peptoids, or N-substituted glycines, are one such class of materials, providing mimicry of the peptide backbone [456]. They have demonstrated extraordinary properties including enhanced thermal, chemical, and enzymatic abilities that are protease resistant, antifouling, and biocompatible [457–460]. In particular, peptoids have been highlighted for their capacity to assemble into complex hierarchical macromolecular architectures like microspheres, nanosheets, and dynamic single-walled nanotube structures [290,456,461]. However, peptoids also have advanced structural and functional complexity, providing a promising avenue for exploring a new class of highly functional molecules. It has previously been shown that peptoids can fold into regular secondary structure motifs that resemble polyproline helices, beta structures, and hierarchically assemble into higher order tertiary-like structures [454,462–465]. Further, the adoption of peptoids into other regular structures found in proteins such as loops and turns has also been explored [466–470].

The chemical and structural design space afforded by synthetic peptoids is vast, providing potentially limitless possibilities. In peptoids the sidechain is attached to the backbone nitrogen (**Figure D-S1**) resulting in two important effects at the molecular level. First, backbone hydrogen bonding, an essential driver in biological secondary structure formation, is eliminated due to the loss of the H-bond donor. Second, in contrast to peptides and proteins which are often sterically confined to a trans- backbone configuration, an isomerizing amide bond in the peptoid backbone results in access to both cis- and trans- backbone configurations. This increased flexibility expands the thermally available configurational landscape and results in conformationally heterogeneous structures, both within a given sequence and with respect to an ensemble of structures. Secondary structure formation in peptoids is instead promoted by the selection of sidechain residues that exploit interchain interactions (either with the backbone, or between residues) to produce folded architectures. In this case, peptoid secondary structure becomes a tunable property through careful selection of the appended sidechain chemistry. Advances in synthesis methods have expanded upon the types of sidechains which can be appended to a peptoid backbone and can accomplish

this with relatively high control. This indicates some promising avenues to independently control peptoid sequence and structure by the careful selection and placement of sidechain chemistries within a peptoid oligomer sequence. This makes these foldamers an exciting route towards applying synthetic analogs to study the structure-function relationships which govern a wide range of natural phenomena [471–474]. However, progress in determining new peptoid sequences and molecular designs is severely limited by a lack of molecular level rules that link the mechanism of folding and stability with its sidechain.

The earliest example of a peptoid that assembles into secondary structure is the N-substituted chiral aromatic oligomer (S)-N-(1-phenylethyl)glycine (Nspe), which has been shown to produce stable polyproline type-I like helices as short as 5 residues in chain length [475–477]. Since the discovery of the helical peptoid, other chiral sidechains such as (S)-N-(1-tert-butylethyl)glycine, a non-aromatic chiral sidechain, and (S)-N-(1-naphthylethyl)glycine, a polycyclic aromatic chiral sidechain, have also been shown to promote backbone cis-isomerization, generating PPI-like helices with increasing homogeneity [478,479]. Stringer et al. has demonstrated that incorporation of N-aryl sidechains stabilize trans- backbone conformations which can resemble PPII helices [480]. Although these achievements have significantly advanced our understanding of the relationship between sequence and secondary structure, corresponding progress in computational design has not kept pace. Despite the tremendous utility of physics-based simulations in aiding in computational design of peptides and intrinsically disordered proteins, there is a severe paucity in the literature of similar applications in biomimetic foldamers.

To date, simulations have provided important knowledge of these systems, complementing the relatively small number of experimental structures available from NMR and x-ray crystallography [481]. Early examples used replica exchange molecular dynamics (REMD) to sample peptoid backbone isomerization [482,483]. However, progress in implementing computational approaches was hindered by a lack of models that adequately described the $N \rightarrow \pi^*$ backbone interaction, and these simulations produced results inconsistent with experimental and density functional theory (DFT) calculations. Recent advancements in quantum calculations and forcefield fitting have led to the development of more robust forcefields and coarse grained models, which more accurately describe the underlying peptoid potential energy landscapes [293,464,477,484–490].

Despite improved potentials, configurational sampling remains a challenge in peptoids as the complex relationship between sequence and secondary structure continues to emerge in the

literature. Early quantum mechanics simulations revealed that the peptoid backbone makes available a highly-dimensional subspace of helical or quasi-helical states which are stable on the order of ~5 kcal/mol [293,464]. Recently, Weiser et al. used DFT calculations to improve the MFTOID forcefield model, uncovering a 12-dimensional subspace (i.e., the existence of 12 unique motifs) available to a solvated peptoid monomer [484]. The potential of biomimetic foldamers to expand the regime of thermally accessible configurations has further been highlighted in a recent discovery by Mannige et al., which demonstrated the existence of a novel secondary structure motif possible through alternating backbone patterns stabilized by intermolecular interactions, leading to the discovery of a new aggregate macromolecular structure [292]. However, previous computational studies have mainly focused on applying simulations to characterize entropic and enthalpic driving forces behind homopolymer assembly into singular known experimental structures such as the PPI like helix, cyclic peptoids, or disordered sequences [411,482,491–494].

Rational design of peptoid secondary structure relies on the ability to select peptoid sidechains which preferentially stabilize a single global minimum over the entire thermodynamic landscape available. To that end, this paper presents a robust, high-dimensional configurational sampling scheme based on the metadynamics family of methods to characterize the underlying thermodynamic preferences within the folding landscape for three aqueous peptoid dodecamers [407]. This approach uses a broadly applicable set of collective variables which can be extended to model other peptoid foldamers systems. The three chemistries simulated in this work are 1) sarcosine ($sarc_{12}$), a polyglycine analog, 2) N-(1-phenylmethyl) glycine ($Npme_{12}$), an achiral aromatic sidechain that is not known to produce helices experimentally, and 3) S)-N-(1-phenylethyl) glycine ($Nspe_{12}$) a chiral aromatic sidechain which has been shown to produce PPI-like helices in solution (**Figure 7.1**). Robust sampling of the 12 relevant states was accomplished by using the parallel bias metadynamics (PBMetaD) scheme, a variant of metadynamics suited to efficiently sample multidimensional free energy landscapesm[235]. In contrast to typical MetaD which uses a high-dimensional bias potential to reconstruct the free energy landscape, in the PBMetaD sampling scheme multiple 1D bias potentials are constructed on the fly to reconstruct the free energy surface (FES) of a highly dimensional CV space. This approach has previously been applied to converge the FES for many high-dimensional systems [235,269,400,413,495]. Mastery of these energy landscapes provides insight into the sidechain interactions that give rise to unique

secondary structures in peptoids and the development of predictive models for the rational design of engineered peptoid structures

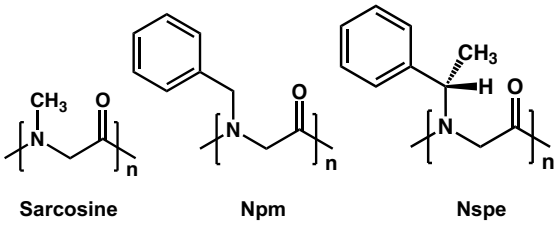


Figure 7.1. Polypeptoid sequences for N-methyl glycine (sarcosine), N-(1-phenylmethyl) glycine (Npm), and (S)-N-(1-phenylethyl) glycine (Nspe).

7.3 Results

I. Sampling the peptoid secondary structure landscape with parallel bias metadynamics

To sample the folding landscape for each 12mer, we used an efficient enhanced sampling method that combines parallel bias metadynamics with parallel tempering in the well-tempered ensemble, described in detail in the **Methods** section. Although the scheme is more complex than the name indicates, for brevity this method is referred to as "PBMetaD" for the remainder of this article. In this work, we applied 1D potentials to each of the backbone and sidechain dihedrals (ϕ, ψ, ω, χ_1, and χ_2), ring-ring coordination (for sequences that contain aromatic sidechains), radius of gyration, and "alphabeta" a collective variable which measures the distances between a set of torsional angles and set of reference values (refer to **SI Methods in Appendix D** for detailed descriptions of these collective variables). This led to a total of 59 (sarcosine) to 138 (nSPE / nPME) CVs used in each simulation. This generalized set of collective variables within the PBMetaD scheme can be extended to model many peptoid folding problems. Complete sampling and convergence was achieved in the limit of 12mer repeat foldamer states requiring a total of 6μs (600 ns/replica) for $sarc_{12}$, 9μs (900 ns/replica) for Npm_{12}, and 10μs (1000ns/replica) for $Nspe_{12}$ (n.b., full demonstration of convergence is discussed in the **SI Methods (Appendix D)** and can be found in **Figs D-S12** to **D-S22.**

Like proteins, the average backbone dihedrals of peptoids can be visualized through Ramachandran plots which project the thermally accessible combinations of the backbone dihedrals ϕ and ψ, illustrated in **Figure 7.2**. To depict the full conformational landscape of

peptoids, inclusive of the backbone *cis*- and *trans*- states, a separate Ramachandran plot is used for each omega value. Specific combinations of ϕ and ψ on the peptoid backbone result in three motifs (α_D, α, and $C_{7\beta}$), which can possess both right-handed (RH, -) or left-handedness (LH, +) and either a *cis*- or *trans*- backbone, resulting in a total of unique 12 relevant states, or motifs, that each residue in a sequence can adopt on the peptoid secondary structure landscape. **Figure 7.2** depicts this secondary structure landscape, with illustrative representative repeat motifs extracted from the enhanced sampling simulations for $sarc_{12}$ in the limit of homogeneous assembly. However, the production of heterogeneous peptoid helices has frequently been observed, and the design of sidechains which can enable robustly homogeneous secondary structure formation has been the focus of many works .[496–498]

As shown, motifs that only differ by a *cis*- or *trans*- backbone configuration can produce dramatically different structures, many bearing strong resemblance to common secondary structures in proteins and peptides. Previously, free-energy plots derived from 46 high resolution peptoid X-ray and NMR structures were used to characterize the peptoid Ramachandran plots, but focused only on four motifs related to the $\alpha_D^{-/+}$-*cis* and -*trans* regions which resemble PPI and PPII helices, respectively[499]. To study oligomeric peptoid structure with enhanced sampling approaches, we propose here the importance of considering all 12 thermodynamically relevant motifs.

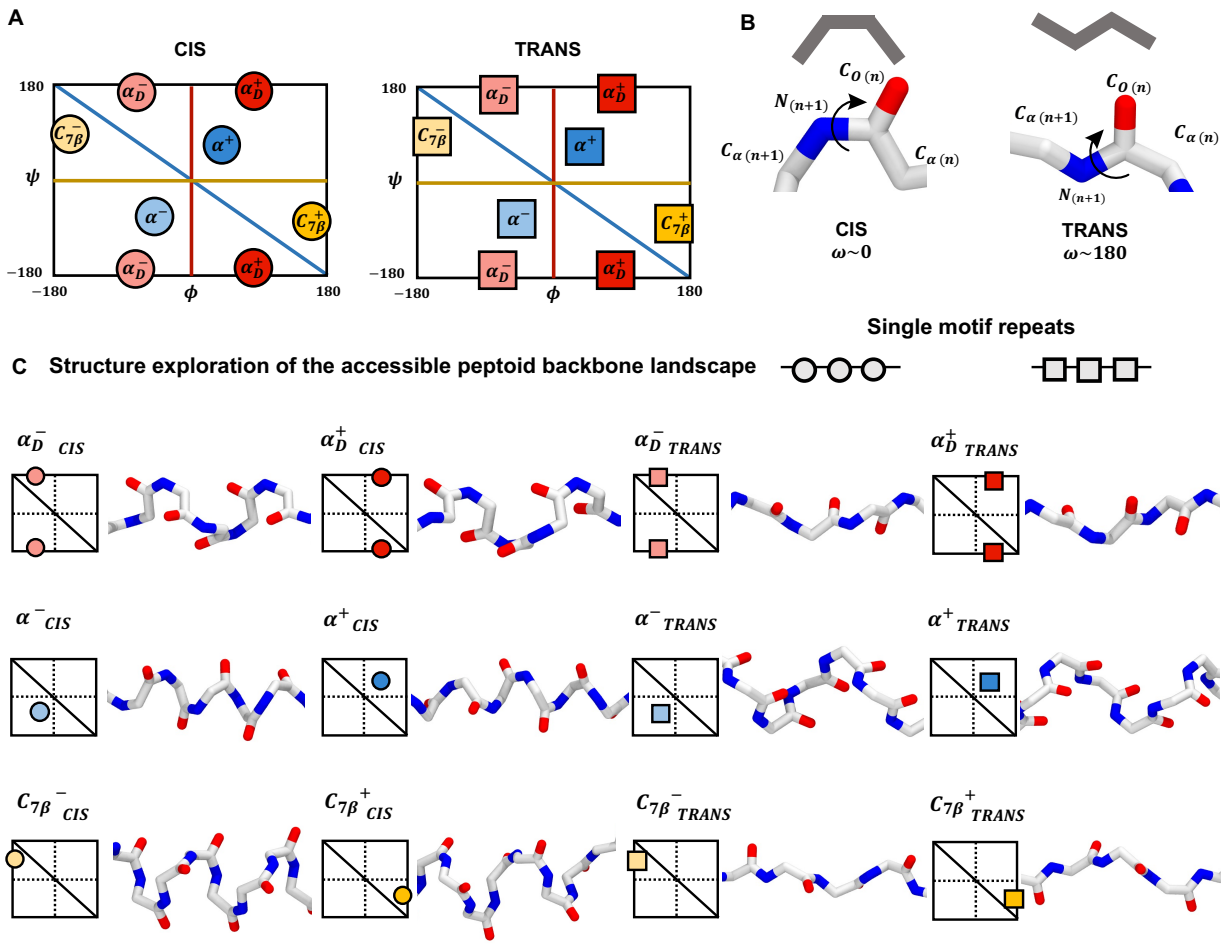


Figure 7.2. Depiction of the accessible thermodynamic folding landscape for the peptoid backbone. (**A**) Ramachandran plots demonstrate the 12 relevant minima α_D^- (light red), α_D^+ (red), α^-(light blue), α^+(blue), C_{7B}^- (light yellow), and C_{7B}^+ (yellow) are given for both *cis*- (circle) and *trans*- (square) conformations. Minus (-) states describe right handed (RH) structures and (+) states describe left handed (LH) structures. Dihedral definitions are included in **Figure D-S2** and this schematic is replotted on a 0 to 360 degree scale in **Figure D-S3**. (**B**) Representative *cis*- and *trans*- backbone conformations. (**C**) Representative snapshots for each homogeneously folded state obtained from sarc$_{12}$ PBMetaD simulations (Ramachandran plots and full structures are given in **Figure D-S4**)

Previous quantum mechanics calculations have supported these 12 states as being thermodynamically relevant (on the order of ~5-10 kcal/mol) for many peptoid sidechain chemistries, with $\alpha^{+/-}$ motifs being the overall the most energetically unfavorable in both the *cis*- and *trans*- conformations for these sidechain chemistries [464,484]. More commonly, sidechains that stabilize heterogeneous or alternating combinations of α_D^-, α_D^+ in the *cis*- and *trans*- conformations have been reported in the literature for a number of cyclical peptoids and coil-like structures [468,470,499,500]. An alternating combination of backbone dihedrals near the $C_{7\beta\,TRANS}^-$ and $\alpha_{D\,TRANS}^-$

local minima as defined here, previously coined the sigma strand, have also been observed in a macromolecular assembly [292]. To date, there are no known sidechains, cyclical structures, or macromolecular assemblies that stabilize any of the other monomeric repeat structures into these notable secondary structures that are depicted in **Figure 7.2**. Further, stable peptoid structures composed of alternating or irregularly patterned backbone motifs (as seen in the novel sigma strand) highlights the possible magnitude of this peptoid backbone structural landscape, which has yet to be explored. Stabilization of the homogeneous $\alpha_{D\,CIS}^{-}$, $\alpha_{D\,CIS}^{+}$ motifs have been of particular biomimetic interest for their resemblance to polyproline type I helices, where dominant handedness has previously been linked to use of S or R enantiomers, for Nspe and similar sidechain chemistries [478]. Other motifs characterized here may be of interest to future sequence design efforts, specifically structures extracted from regions on the *trans*- Ramachandran plot $\alpha_{D}^{-/+}{}_{TRANS}$, $C_{7\mathrm{B}}{}^{-/+}{}_{TRANS}$ bearing closer resemblance to protein beta sheets, and α_{TRANS}^{-}, and α_{TRANS}^{+} to left and right handed protein helices.

To first gain insight into the average backbone structure as a function of the appended sidechain chemistry we analyzed a subset of 20 sample structures for each of the 12 homogeneously folded (alphabeta > 30) lowest-energy peptoid configurations from simulations of sarc_{21} Npm_{12} and Nspe_{12} (240 structures per system). The alphabeta CV undergoes degeneracy quickly, as it does not preserve continuity of the dihedrals within a sequence. A reference value greater than 30 accounts for the most homogeneously folded structures, adjusting for some flexibility. The set of ϕ, ψ, and ω backbone dihedrals for each residue along the sequence (excluding termini) were averaged for each 12mer structure, hereafter referred to as the "sequence-averaged" values and are reported in **Table D-S1**. The sequence-averaged dihedrals reflect the average behavior of each residue's backbone within the 12mer. The values for each motif were found to remain constant independent of the appended sidechain chemistries, given by the low standard deviation. This indicates that the same 12 configurational motifs are accessible to sarcosine, Npm, and Nspe, meaning that the accessible peptoid folding landscape is independent of the appended sidechain chemistry. Therefore, secondary structure formation relies on the sidechain to alter the relative probabilities between these 12 microstates. The following thermodynamic analysis aims to uncover this relationship between sidechain chemistry and stabilization of a single microstate.

II. Energetics of sequence-averaged structural ensembles

Under the assumption of a quasi-static bias potential, these 1D potentials can be reconstructed as a single effective bias potential, which has previously been shown to converge empirically to the correct underlying free energy surface.[235] Because of this, standard metadynamics reweighting approaches can be applied to reconstruct the underlying free energy surfaces. In this work we applied the standard Torrie-Valleau reweighting technique to recover probability distributions.[383] First, we reweighted the energy of every $\phi/\psi/\omega$ dihedral triplet in each sequence, onto a 2D Ramachandran-like FES (**Figure 7.3**). This is what we refer to as the sequence-averaged peptoid structures, or the impact of the appended sidechain on a single residue backbone within the 12mer sequence. We emphasize that the energetics of these basins reflect the average propensity of individual $\phi/\psi/\omega$ dihedral triplets of a single residue to be in one region of the helical phase space and do not necessarily imply anything about possible effects arising from the ordering of structural motifs along a sequence. The overall tendency of a particular helical fold to be adopted homogeneously along a peptoid is further discussed in **section IV**. To accomplish this, a weight was calculated for each $\phi/\psi/\omega$ dihedral triplet in the sequence. A weighted kernel density function (bandwidth=0.05) was then used to project the unbiased energies onto two 2D free energy surfaces (FES) for each system, in this case using the replica at 300K (**Figure 7.3)**.

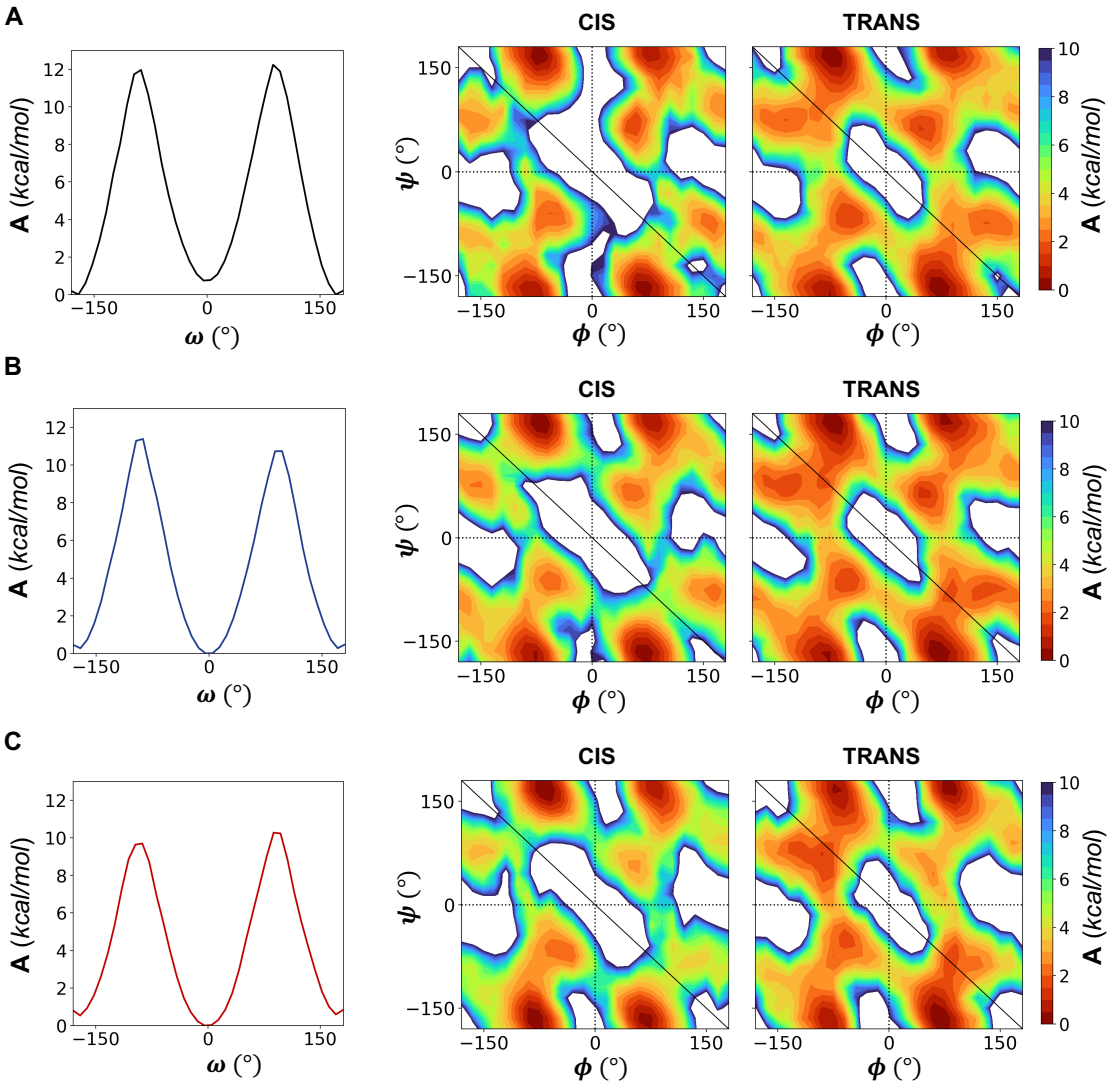


Figure 7.3. Sequence-averaged free energy for backbone dihedrals at 300K. Free energy profiles (FEP) are plotted as a function of the omega dihedral (left) and FES are plotted for *cis*- and *trans*- configurations for **(row A)** sarc$_{12}$ **(row B)** Npm$_{12}$ **(row C)** Nspe$_{12}$.

The data on these information-rich surface representations, are quantitatively represented in **Table 7-1**. To accomplish this, the Helmholtz free energies (A) for each of the 12 folded states were extracted from each FES using basin definitions depicted in **Figure D-S5** and renormalized by basin weights and isomerization weights to extract the relative probability of being in one of the 12 available states, normalized to the entire folding landscape (**Figure D-S6**). In this scheme, P_{folded_i} describes the probability of being in a folded state relative to the entire free energy surface, using the normalized probability of each state ($p_{f_i,norm}$, defined in **Figure D-S6**) as the reference point. Since these values are normalized, the quantity, $1 - p_{f_i,norm}$ describes the probability of

being in outside of this motif. For isomerization, *cis*- regions were defined as $\omega \geq -90$ and $\omega \leq 90$, and all other values were considered to be in a *trans*- backbone configuration. Free energy was then calculated from **Equation 7.3**, using either $P_{isomerization}$ or P_{folded_i}. The probability of a peptoid backbone dihedral falling outside of one of the 12 motif definitions, was found to be insignificant and therefore not included in Table 7.1. This is given by the very small exemplary probabilities provided in **Figure D-S6**, and the very high (> 10kcal/mol) energies needed to describe states outside these regions as presented in **Figure 7.3**.

$$P_{isomerization} = \frac{p_{cis}}{p_{trans}} \quad 7.1$$

$$P_{folded_i} = \frac{p_{f_i,norm}}{1 - p_{f_i,norm}} \quad 7.2$$

$$A = -k_B T ln(P) \quad 7.3$$

The sequence-averaged free energies for the omega dihedral, describes a sequences overall propensity to isomerize into a *cis*- backbone configuration (**Figure 3**), Values reported for isomerization in **Table 7-1** describe the relative energetic propensity for a sidechain to drive *cis*-isomerization for a single residue within the entire 12mer sequence. $Sarc_{12}$ does not favor isomerization (negative energy for isomerization), whereas the presence of the hydrophobic sidechain (Npm) stabilizes the cis-backbone configuration by a very small amount (0.11 kcal/mol), and the addition of the chiral sidechain further stabilizes a *cis*- backbone by 0.45 kcal/mol. Both values are less than kT, but the energies of isomerization can be used to calculate a corresponding equilibrium constant using the equation $K=\exp(-\Delta G/RT)$ and values in **Table 7-1**. From this, we estimated a $K_{cis/trans}$ of 0.23, 1.2, and 2.1 for $sarc_{12}$, Npm_{12}, and $Nspe_{12}$, respectively which are consistent with increasing *cis*- stability. These results suggest that a methyl sidechain stabilizes a *trans*- backbone, whereas the presence of an aromatic sidechain reduces the *trans*- to *cis*-isomerization energy, and a chiral aromatic group further stabilizes *cis*- backbone sequences. In **Section III** we show that the role of these sidechains in stabilizing a specific backbone isomer is correlated with the available configurational entropy in those states. Npm and Nspe sidechains enable a greater diversity of backbone structures in a *cis*-, where as sarcosine supports greater diversity in a *trans*- backbone state.

In **Table 7-1** we report the energy of folding into a specific motif of the possible 12, for a single sidechain chemistry. These values are normalized by the minimum energy value for each chemistry, so that relative stabilization of different motifs by a single chemistry can be compared. In this table we highlight stable motifs within kT of the minimum energy, these are expected to be the most energetically stable states a residue in a sequence can adopt at room temperature. For achiral sidechains, left- (+) and right-handed (-) structures are expected to have the same free energy. In **Table D-S2** we calculated these differences showing that left- and right-handed motifs for sarc and Npm converge to energies within a tenth of a kcal/mol of each other, and significantly less than half a kT at 300K. In contrast, the energy differences in right vs left-handed structures produced by the chiral Nspe sidechain were significantly different. This further highlights the robustness of the sampling method applied in this approach.

In $sarc_{12}$, we find the most stable motifs are $\alpha^{-}_{D\,TRANS}$ and $\alpha^{+}_{D\,TRANS}$ agreeing with previous experimental and computational studies[499,501]. Polysarcosine is exemplary of a random coil peptoid, having been characterized in depth showing significant cis/trans conformational heterogeneity along the backbone and in solution, with an overall preference towards a *trans*-backbone and no overall preference towards a single conformation[502]. This highlights that "randomness" in the context of peptoids may be less related to the propensity of backbone dihedrals to fall outside of these 12 defined states, but instead related to conformational heterogeneity and overall flexibility (or lack of rigidity) along the peptoid backbone.

For Npm_{12} four motifs including both *cis*- and *trans*- backbones ($\alpha^{-}_{D\,TRANS}$, $\alpha^{+}_{D\,TRANS}$, $\alpha^{-}_{D\,CIS}$ and $\alpha^{+}_{D\,CIS}$) are stabilized within kT, indicating that different motifs can be readily adopted by residues within a single oligomer sequence, with nearly equal energetic preference. This would support the overall formation of heterogeneously patterned sequences. In agreement with this, Npm has not previously been reported to form helical structure in solution[475]. Npm however has previously been used as a building block in amphiphilic sequences to form macromolecular peptoid sheets[290], with dominant interactions originating from interchain pi-pi stacking, as opposed to the intrachain interactions identified here, and has been characterized as having a beta-like secondary structure motif. In this case, stability from additional intermolecular driving forces likely drive an overall backbone preference where the sidechain does not sufficiently accomplish this through intermolecular interactions with the backbone or other sidechains. With Npse, $\alpha^{-}_{D\,CIS}$ and $\alpha^{+}_{D\,CIS}$ are the most stable motifs, supporting assembly into the previously reported PPI-like

peptoid helix. Further there is a slight energetic preference (0.5 kcal/mol) for Nspe to form a right-handed motif over a left-handed helix. This matches previous computational predictions for an octamer Nspe sequence, and predicted behavior from the forcefield applied in this study [477,484]. Since we did not initiate the simulations from any pre-populated structures, and our biasing scheme did not favor one structural ensemble over any others, it is very significant that the subtle effects of the sidechain on shifting the helical ensemble seem to be naturally captured by MD, provided there is a suitable enhanced sampling scheme to fully explore helical phase space.

Table 7-1. Sequence-averaged energies extracted from reweighted FES. Values for structures are in kcal/mol relative to the lowest-energy conformation, except isomerization energies. Values within 0.59 kcal/mol (1kT) of the absolute minima for each dodecamer chemistry are bolded.

	Sarc$_{12}$ (kcal/mol)	**Npm$_{12}$** (kcal/mol)	**Nspe$_{12}$** (kcal/mol)
Isomerization (trans to cis)	0.88	-0.11	-0.45
cis- $\boldsymbol{\alpha_D^-}$	1.03	**0.00**	**0.00**
cis- $\boldsymbol{\alpha_D^+}$	1.09	**0.13**	**0.50**
cis- $\boldsymbol{\alpha^-}$	3.01	2.33	2.36
cis- $\boldsymbol{\alpha^+}$	2.83	2.29	3.19
cis- $\boldsymbol{C_{7B}^-}$	3.24	2.54	3.50
cis- $\boldsymbol{C_{7B}^+}$	3.15	2.29	4.15
trans- $\boldsymbol{\alpha_D^-}$	**0.00**	**0.34**	1.13
trans- $\boldsymbol{\alpha_D^+}$	**0.05**	**0.33**	0.74
trans- $\boldsymbol{\alpha^-}$	1.75	2.00	2.60
trans- $\boldsymbol{\alpha^+}$	1.73	1.91	3.34
trans- $\boldsymbol{C_{7B}^-}$	2.03	1.93	2.64
trans- $\boldsymbol{C_{7B}^+}$	2.03	2.04	3.41

To further understand the mechanism of stabilizing interactions, we determined the structure and energetics of the peptoids along the two prominent sidechain dihedrals, χ_1 and χ_2 (**Figure D-S7**). Comparison of the FES describes the impact on freedom of rotation about the sidechain through the addition of the -CH3 (and creation of a chiral center) in Nspe. In Npm$_{12}$, the sidechain is flexible in both a *cis-* and *trans-* backbone configuration, given by the large connected low energy regions along the FES. In Nspe$_{12}$, the FES show higher energy states between rotation of the sidechains, indicating less rotational freedom. The addition of a chiral center also results in

preferential stabilization of two of the minima in the *cis*- form where there are four stable minima present in the *trans*- form, indicating some steric repulsion due to the chiral center in the *cis*- conformation. Exploration of this impact of chirality on homogenous structure stability is further explored in **section IV**.

III. Entropic driving forces in helical peptoid stability

The role of entropy in stabilizing peptoid helices has long been of interest. In the parallel tempered metadynamics approach applied in this work, temperature aids in sampling the conformational landscape. Because of this, we gain access to free energy over a range of temperatures (T). Using the relationship in Equation 4 between A and T, we can uncover the underlying entropic (S) contributions to the observed energetics. In this work, we evaluate the role of entropy and enthalpy as a function of two events necessary for overall helix formation; (1) isomerization into a cis- backbone, and (2) folding into an $\alpha_{D_{CIS}^{-}}$ motif. Free energies were calculated using equations 1-3 over a range of temperatures and plotted in **Figure 7.4.** In these plots, slope indicates the entropic contributions for each system. Under the general assumption that U and S do not significantly vary with temperature, a negative slope indicates an entropic gain, and a positive slope indicates an entropic penalty. The linear fit of these relationships, and respective slopes are provided in **Table D-S3**. The entropic contributions for these processes at temperatures ranging from 300K to 500K are included in Table **D-S4** and **D-S5**.

$$A = U - TS \tag{7.4}$$

Isomerization energy is plotted as a function of temperature in **Figure 7.4,** providing insight into how the role of the sidechain drives cis- backbone stability. For sarc12, there is an overall energetic preference towards the trans- state. The entropic contributions for cis- isomerization at 300K are nearly zero, indicating no entropic driving force for the backbone to isomerize into a cis- form. There is a small energetic driving force for Npm12 and Nspe12 to isomerize into a cis- backbone. However, the entropic component is nearly zero for Npm12 at 300K, while Nspe12 benefits from an entropic gain on the order of 0.30 kcal/mol at 300K. Therefore, while Npm12 appears to have no backbone preference, cis- isomerization for the chiral bulky Nspe12 is entropically driven.

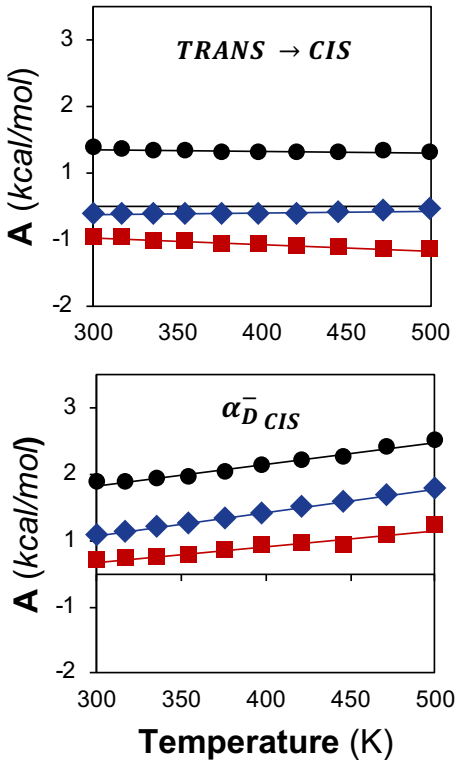


Figure 7.4. Entropic driving forces. Free energy of isomerization into cis- (**top**) and free energy of folding into $\alpha_{\bar{D}_{CIS}}$ state is plotted as a function of temperature (**bottom**). Sarcosine is plotted in black circles (●), Npm in blue diamonds (◆) and Nspe in red squares (■) linear fits for this data is provided in **Table D-S3.**

One possible explanation for this entropic gain is that Nspe provides greater conformational diversity in a *cis*- backbone conformation compared to *trans*-. Meaning, a residue within a $Nspe_{12}$ sequence is more likely to adopt one of any combination of the 6 motifs with a *cis*- backbones, than the 6 *trans*- motifs. To qualitatively assess this, we can apply clustering and reweighting methods to compare the conformational diversity (or the distribution of unique structures) that can exist given a mostly *cis*- or *trans*- backbone. A cumulative distribution function (CDF) of these reweighted cluster probabilities is plotted in **Figure D-S11** (and further described in the **SI, Appendix D**). By comparing the CDF that more rapidly approaches 1 for *cis*- vs *trans*-, we gain a proxy that describes which backbone enables greater conformational diversity of structures. $Sarc_{12}$ demonstrates greater conformational entropy in a *trans*- backbone configuration, compared to *cis*- (given that fewer *cis*- than *trans*- structures are needed to describe the entire distribution of structures). Npm_{12} appears to have slightly greater conformational entropy in a *cis*- than a *trans*- state, although quite similar preferences for both. $Nspe_{12}$ demonstrates greater

conformational entropy in a *cis*- than *trans*- state. This supports that the addition of a hydrophobic and chiral backbone enables greater diversity of *cis*- backbone structures, and that this entropic gain to isomerization in $Nspe_{12}$ can be attributed to conformational entropic gains.

Alternatively, there is a significant entropic penalty associated with all three sidechain chemistries assembling into the $\alpha^{-}_{D\,CIS}$ motif. There is greater conformational diversity in the unfolded states than within the ensemble of $\alpha^{-}_{D_{CIS}}$ structures, resulting in an overall entropic penalty related to assembling an ordered helix in water. This indicates that folding into an $\alpha^{-}_{D\,CIS}$ motif by Nspe is mainly driven through enthalpic gains. This phenomena has previously been observed experimentally in diblock helix containing polymers, where large entropic gains were associated with peptoid disordering [503]. Alternatively, Mukherjee et al., previously applied the Lifson−Roig Theory of the Helix−Coil Transition to evaluate the role of entropy in relation to the steric clash caused by a bulky chiral center on helix formation, proposing that the sidechain supported an overall greater volume of allowed conformations citing entropy as a major driving force.[491] To accomplish this, they use a broader definition to describe a folded state: a cis ω, and a negative ϕ. In the work presented here, this cutoff applies to 3 different cis minus motifs: $\alpha^{-}_{D_{CIS}}$, $C_{7\mathrm{B}^{-}_{CIS}}$, and α^{-}_{CIS}. In general, our metadynamics results support these conclusions with respect to cis- isomerization but not for overall assembly into a helix. These slight discrepancies are likely due to differences in how the folded state is defined between works.

Complexity between competing entropic and enthalpic contributions due to these sidechains begins to elucidate how the generation of homogenous helices is difficult due to competing driving forces, where design of robust sidechains requires optimization of both entropic and enthalpic interactions. This could begin to provide an explanation for why a non-aromatic chiral (S)-N-(1-tert-butylethyl)glycine peptoid sidechain has previously demonstrated more robust assembly into polyproline-like helices than these bulky aromatic and chiral sidechains like Nspe [479]. Application of the computational methods used in this work can be applied to further evaluate the entropic components as they are related to non-aromatic chiral sidechains, and assembly under non-aqueous experimental conditions to provide further insight to the entropic and steric driving forces of helix stabilization in various sidechains, as well as the impact of solvents.

IV. Homogeneous Assembly of Peptoid Secondary Structure

The thermodynamic analysis presented thus far has been reported in reference to the sequence-averaged energies that give rise to ordered peptoid motif stability due to the presence of a sidechain. However, navigating the large structural landscape of these heteropolymers requires careful consideration related to how they assemble into these homogeneous secondary structure sequences. The alphabeta CV is highly degenerate, as it does not preserve continuity of the backbone dihedral reference values. So, to describe energetics of structure along the full peptoid sequence we introduce a new CV used for analysis called peptoidRMSD. This CV has previously been used to explore protein structure[504], and was adapted here to measure peptoid structure, coined here as peptoidRMSD, and given by **equation 7.5**. While the peptoidRMSD CV was not explicitly biased in these simulations, using reweighting techniques previously described, the converged system energetics can be projected onto this other collective variable to gain insight into sequence helicity along the backbone.

A perfectly ordered $\alpha_{D_{CIS}^-}$ structure was used to construct a 3-residue repeat reference structure, which describes a single $\alpha_{D_{CIS}^-}$ helical turn. With this definition of helical structure, the RMSD between the backbone atoms (N, C_{C-H_2}, $C_{C=O}$, O, and C_β) of each 3-residue block along the peptoid and the reference structure is measured. A sigmoidal function ranging from 0-1 is used with values of n=8, m=12, and $r_0 = 0.08$ nm, to determine how closely each segment along a sequence matches the reference structure. The RMSD between all 12 reference motifs was found to be around 0.12 nm. A smaller value of 0.08 nm was then chosen as r_0 as a tight cutoff to distinguish between unique structures. This value is summed over all the 3-residue blocks (i) within the sequence (e.g., a 12mer is composed of 9, 3-residue repeats). Larger peptoidRMSD values correspond to more helical content, and lower values correspond to less or no helical content. This provides a descriptor for a perfectly helical peptoid (for a 12mer this value is 9) and allows for classification of homogeneously folded helices. This CV for defining structure could be extended in the future to model all 12 peptoid structures, and further expanded to quantify heterogeneous (A/B) alternating structures sampled computationally.

$$peptoidRMSD = \sum_i \frac{1 - \left(\frac{RMSD_i}{r_0}\right)^n}{1 - \left(\frac{RMSD_i}{r_0}\right)^m} \qquad 7.5$$

Free energy of the 12mers were projected onto the peptoidRMSD($\alpha_{D_{CIS}^-}$) CV and radius of gyration to describe energy as a function of how folded and flexible each $\alpha_{D_{CIS}^-}$ containing structure was in solution (**Figure 7.5**). On these plots a peptoidRMSD value of 0 is representative of all the sampled structures which do not contain a single 3-residue $\alpha_{D_{CIS}^-}$ helix turn, low peptoidRMSD values indicates the presence of heterogeneous $\alpha_{D_{CIS}^-}$ kinks within a sequence, whereas values that approach 9 denotes a perfectly homogeneous helix. The energy required to assemble into a $\alpha_{D_{CIS}^-}$ helix as a function of chain length was calculated using **equations 7.2-7.3**, using the peptoidRMSD($\alpha_{D_{CIS}^-}$) CV as a proxy for how folded the structure is. A large peptoidRMSD, is consistent with a more homogeneously folded structure, whereas a small peptoidRMSD indicates there is only some adopted structure within the sequence. Free energy of folding into a helix as a function of this CV value is included in **Figure 7.5D**. Assembly of all sidechains, including $Nspe_{12}$, into a fully homogeneous helix (given by a peptoidRMSD($\alpha_{D_{CIS}^-}$) greater than 8) was not observed to be energetically favorable in water. Future work should be aimed at applying the simulation methods described here to probe assembly in non-aqueous solvents, like acetonitrile, which better approximate experimental conditions.

Of all the sidechains, $Nspe_{12}$ had the overall largest propensity to fold into a partial $\alpha_{D_{CIS}^-}$ helix, preferring a compact helix given by a smaller radius of gyration. It also experienced the smallest energetic penalty associated with assembling into a homogeneous helix (peptoidRMSD($\alpha_{D_{CIS}^-}$) value of 9). The longest stable partial helix was found to be for a peptoidRMSD value of 6, with an overall driving force of -0.94 kcal/mol. The small radius of gyration is correlated with compact structures which are bent or folded onto themselves (**Figure 7.5C**). Npm_{12} also slightly favored assembly into a partial compact helix. The largest partial helix that could be significantly stabilized by Npm was represented by a peptoidRMSD value of 5, with a driving force of -1.39 kcal/mol. Npm_{12} preferred a slightly larger radius of gyration and a less folded structure, indicating more conformational flexibility. Assembly of $sarc_{12}$ sequence into partial $\alpha_{D_{CIS}^-}$ helices was only significantly favorable for a peptoidRMSD value of 4, indicating that a few residues within a sarcosine foldamer sequence could adopt this helical sequence. $Sarc_{12}$ also experienced the largest energetic penalty associated with assembling into more homogeneous $\alpha_{D_{CIS}^-}$ helices. Similarly, $sarc_{12}$ favored the largest radius of gyration values, indicating the greatest

overall conformational flexibility in water. Sequence energetics matched the sequence-averaged trends reported earlier in this work.

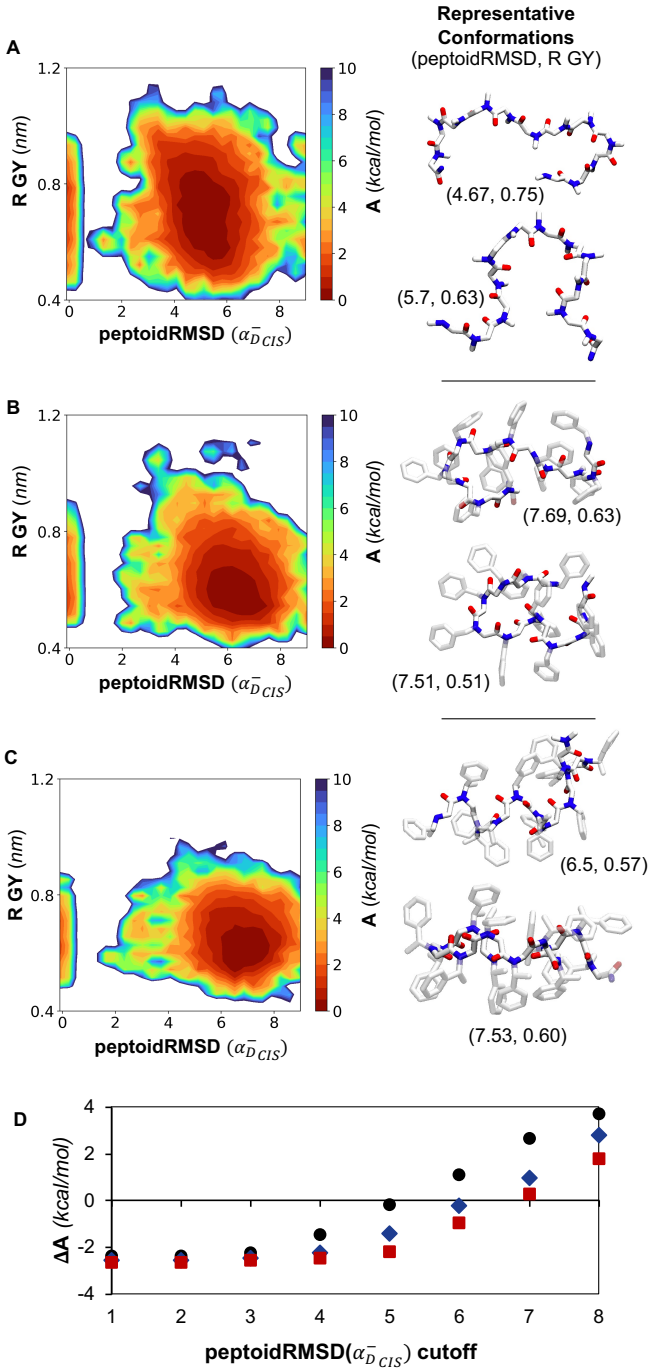


Figure 7.5. FES for homogeneous $\alpha^{-}_{D_{CIS}}$ helix formation. FES given for (**A**) sarc12 (**B**) Npm12 (**C**) Nspe12. Two representative conformations for the minimum energy (>1kcal/mol) regions on FES are provided in the right column for each system. (**D**) free energy associated with folding into a $\alpha^{-}_{D_{CIS}}$ helix as a function of the number of helical segments in a sequence, where peptoidRMSD is used to define the cutoff for total number of helical segments.

To evaluate where the additional stabilization is gained in the $\alpha_{D_{CIS}}^{-}$ partial helix formed by Nspe$_{12}$, the rotational degrees of freedom related to the addition of a chiral center between systems Npm$_{12}$ and Nspe$_{12}$ can be compared. The χ_1 dihedral describes this rotation about the first carbon on the sidechain, directly capturing the effects of a chiral center. The 1D potential of mean force (PMF) integrated along the χ_1 dihedral for Npm$_{12}$ and Nspe$_{12}$ for partial $\alpha_{D_{CIS}}^{-}$ helices (defined as peptoidRMSD > 6) are given in **Figure 7.6**. Rotation about χ_1 in sarcosine describes rotation about a methyl hydrogen, which results in a low energy symmetric rotation, and is not included in this analysis. With Npm$_{12}$ there is a small energetic penalty associated with rotation about χ_1, while the presence of a chiral center in Nspe$_{12}$, resulting in a large energetic penalty to rotate along this dihedral. Additionally, the Nspe$_{12}$ PMF highlights a small energetic preference towards one energetic basin over the other, indicative of a preference towards a right-handed helix for this system. This bias towards a single basin is not present in the achiral Npm$_{12}$ counterpart. Overall, these results support that the role of steric strain due to the bulky chrial center, as previously hypothesized, is a major driving force in helical peptoid formation and stability.[477]

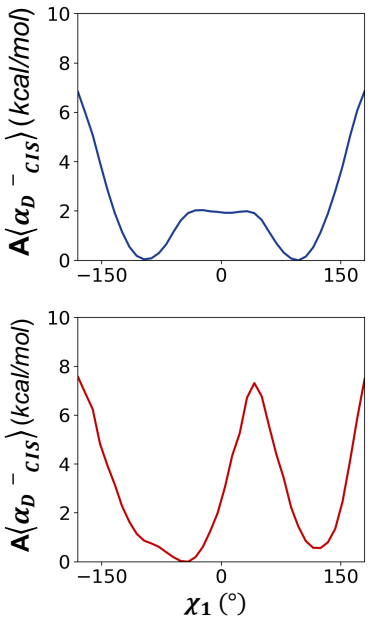


Figure 7.6. 1D PMF of χ_1 rotation integrated out for partially folded $\alpha_{D_{CIS}}^{-}$ helices. (**top**) Npm$_{12}$ and (**bottom**) Nspe$_{12}$.

7.4 Discussion

Peptidomimetics like peptoids are ideal candidates for use in advanced materials applications, but much remains to be learned about the mechanisms that govern their assembly into the large design space of higher order secondary and tertiary structural assemblies. Similarly, the tendency for peptoids to form heterogenous conformers in solution makes them difficult to study experimentally. Without further advances in methods that can uncover the relationships between side chain chemistry and evaluation of structural ensembles, progress in rational design of these systems will be slow. The simulation approach presented in this work holds great potential as an efficient and powerful tool for predicting energetics and structures of complex peptoid chemistries, paving the path towards understanding specific sequence impacts of the sidechain on folding. Beyond demonstrating a general-purpose simulation tool that can be used to study a huge range of biomimetic foldamers, this work provides significant new mechanistic insight into the underling process by which chiral side chains stabilize helical folds in the Nspe system, under aqueous conditions.

Our results indicate isomerization of the omega dihedral into a *cis*- backbone configuration for Nspe is attributed to entropic gains which can be attributed to an increase in conformational entropy for *cis*- vs *trans*- containing backbones. However, entropic penalties associated with assembling into a $\boldsymbol{\alpha_{D_{CIS}^{-}}}$ backbone motif, indicate assembly of peptoid helices in not entirely entropically driven. We found a greater energy required to rotate about the $\boldsymbol{\chi_1}$ dihedral in Nspe compared to Npm, indicating steric strain provides additional stability necessary to promote an $\alpha_{D_{CIS}^{-}}$ motif. Overall, these driving forces relate specific attributes of the sidechain to overall formation of a PPI-like helix (coined here as the $\alpha_{D_{CIS}^{-}}$ helix) in water. This is conclusion is supported by previous insights from experiment and computation, all of which are self-consistent with the new results we present herein. Further, we quantified the overall stability of a right-handed PPI twist formed by Nspe, supporting previous experimental observations for the S-enantiomer. These results are presented in the limit of peptoid sidechain chemistries where neutral steric bulky sidechains drive the main mechanism of assembly for peptoids in water solvents. Future directions for this field could apply this approach to investigate relevant underexplored degrees of freedom and their effects on folding at the molecular level including chain length, solvent effects, sidechain chemistries, and sequence. And further, the assembly of helical peptoids into new hierarchical

nanomaterial macro-assemblies, composed of higher order combinations of the motifs described here. Broad predictions of stability along wider range of side-chain chemistries in specific could possibly allow for the creation of QSPR type relationships for peptoid design.

7.5 Methods

Three different dodecamer polymers were simulated in a water box using a modified CHARMM-based forcefield combining parameters published by Weiser and Santiso [484] and MFTOID forcefield [293], this is described in detail and validated in the **SI (Appendix D)**. The dodecamers studied were sarcosine ($sarc_{12}$), N-(1-phenylmethyl)glycine (Npm_{12}) sequence, and (S)-N-(1-phenylethyl)glycine, ($Nspe_{12}$) with a neutral N-termini and amidated C-termini as shown in **Figure D-S1**. Additional simulation, setup, equilibration, and forcefield details are included in the **SI Methods**. After equilibration, the three peptoid dodecamers were simulated using a variant of parallel tempered metadynamics [505], coined here as parallel bias parallel tempered metadynamics in the well-tempered ensemble (PT-PBMetaD-WTE) [233,235,506].

An ensemble of initial peptoid configurations was generated from a biased simulation of the peptoid exploring different conformational phase space, so that a unique starting configuration was used for each replica. Using 10 equally spaced replicas ranging from 300K to 500K, potential energy was biased with metadynamics for 20ns, to establish the well-tempered ensemble [233]. For this, an initial hill height of 2 kJ/mol, a bias factor of 60, with a sigma of 150 kJ/mol (calculated from the potential energy fluctuations of the lowest temperature replica). A deposition rate of 1hill/0.5ps (250 steps) was used, attempting replica exchange at the same pace. The well-tempered ensemble was achieved by a constant replica exchange probability of 28%, 27%, 25%, for $sarc_{12}$, Npm_{12}, and $Nspe_{12}$, respectively near the recommended value [87]. Following this, additional parallel bias metadynamics bias potentials were introduced, and the WTE deposition rate was reduced by a factor of 20 (1hill/10ps) to maintain constant exchange, and used to smooth out any artificial effects from the second biasing potential [88]. In the parallel bias metadynamics scheme CVs were biased using 1D well-tempered biased potentials. A gaussian potential with an initial hill height of 2.5 kJ/mol, a bias factor of 140, and a hill deposition pace of 1hill/2ps (1000 steps) was used. The bias factor in PBMetaD simulations is considerably higher than standard well-tempered metadynamics (WTM) simulations due to the manner in which the PBMetaD bias is distributed amongst the individual 1D potentials [235] . Experience has shown a reasonable choice is to scale an

expected WTM bias factor for the underlying physics of the system (in this case 12) by the square root of the number of CVs used. In sarcosine simulations the coordination collective variable was omitted due to the absence of ring interactions. This resulted in a total of 138 CVs used to sample the Npm_{12} and $Nspe_{12}$ systems, and 59 CVs used to sample the sarcosine system. These CVs include:

(i) Each ring coordination number (66 CVs, sigma=0.5)

(ii) The total ring coordination number (1 CV, sigma=0.5)

(iii) Radius of gyration (1 CV, sigma=0.5 nm)

(iv) All ω dihedrals (11 CVs, sigma=0.35 rad)

(v) All ϕ dihedrals (11 CVs, sigma=0.35 rad)

(vi) All ψ dihedrals (12 CVs, sigma=0.35 rad)

(vii) All $\chi_1$dihedrals (12 CVs, sigma=0.35 rad)

(viii) All χ_2 dihedrals (12 CVs, sigma=0.35 rad), and

(ix) alphabeta of each identified structural state (12, sigma=0.5)

Each system was simulated for a total of simulation time of 6μs (600 ns/replica) for $sarc_{12}$, 9μs (900 ns/replica) for Npm_{12}, and 10μs (1000 ns/replica) for $Nspe_{12}$. Additional methods, CV descriptions, convergence, and additional supporting classical MD simulations are included and described in the **SI (Appendix D)**. All corresponding PLUMED input files required to reproduce the metadynamics results reported in this paper are available on PLUMED-NEST (www.plumed-nest.org) plumID:21.030, the public repository of the PLUMED consortium.[89]

7.6 Acknowledgements

The research design and simulations in this project were supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, as part of the Energy Frontier Research Centers program: CSSAS--The Center for the Science of Synthesis Across Scales under Award Number DE-SC0019288. SA acknowledges support from the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1762114. KT acknowledges support from National Science Foundation NRT traineeship DGE-1633216. This work was completed in part through the advanced computational, storage, and networking infrastructure provided by the Hyak supercomputer system and funded by the STF at the University of

Washington. The authors are grateful to Dr. Massimiliano Bonomi and Dr. Paul Robustelli for useful discussions.

Chapter 8
Significance and Perspective

The work presented here has advanced the state of computational simulations, by adding fundamental insight to the structure/function mechanisms which regulate biological systems within biomineralization, self-assembly, and enzyme catalysis.

With respect to biomineralization, we have further contributed to the literature on modeling peptide/surface and peptide/peptide interactions, providing fundamental mechanistic insight into experimentally studied systems. This work builds off foundational peptide/surface computational studies done in the Pfaendtner group which involved method development using model systems.[70,88,269] Recently, work from the Pfaendtner group applied these methods to study systems related to silica[48,270,275], titania[90,280], and mineral[64] interfaces. Inspired by this, these methods were applied to study the binding domain of a protein, Osteocalcin, at a physiologically relevant hydroxyapatite and titania surface, as described in Chapter 2. At the time of publication, only a handful of studies had been done to understand the behavior of peptides at these complex interfaces, due to a lack of models that could appropriately describe a hydroxyapatite surface at a physiologically relevant pH. By simulating systems under these conditions, we propose a mechanism for peptide disordering at a titania interface as a function of its post-translational modifications.

Our contributions to enzyme catalysis built off foundational work from the Pfaendtner and Kaar groups related to optimizing enzymes in non-aqueous solvents with the goal of reducing costs related biofuel production. One of many obstacles in the production of biofuels is the costly and time-consuming conversion of biomass into hydrolyzable sugars. Ionic liquids, which can effectively break down cellulose, have been poised as a potential solution to replace multi-step pretreatment steps with a single batch process. Prior work from the Pfaendtner and Karr groups was focused on studying/optimizing the activity enzymes these non-native solvents.[507–509] Combined simulation/experimental approaches had previously revealed ionic liquid cations competitively inhibit the activity of a cellulase known as E1.[436] Following this, it was found that

the activity and stability of this enzyme in ILs could be improved through the rational mutation of active site residues, and the addition of non-native disulfide bonds to reduce structural fluctuations.[437] Inspired by this progress, we looked to further improve activity and lower costs associated with processing biofuels by alternatively reducing the mechanisms of product inhibition for E1 as presented in Chapter 6. Together, these works have contributed different strategies for reducing the inactivity of a model cellulase. This can serve as a template for improving other cellulases and potentially other enzymes.

In the field of self-assembly, we contributed to the literature in two ways; studying protein assembly at the air-water interface, and hierarchical peptoid self-assembly. Through established and ongoing collaborations between the Pfaendtner and Weidner group[149,150,273], we had previously studied the behavior of an antifreeze[138] and ice-nucleating[510] protein at an air-water interface. These works inspired conversations related to resolving fundamental questions in protein surface-science, that SFG methods did not have high enough resolution over. In particular, the SFG signal intensity in the amide-III region which can distinguish characteristic structural features (e.g. α-helixes, β-sheets, β-turns, and random coils) is relatively weak to obtain due to technical challenges. In cases, like Lysozyme, where there is no significant changes to protein structure from bulk to interface, the obtained SFG are difficult to infer due to a weak signal-to-noise ratio. Overcoming this challenge with an integrated simulation/experimental approach led to the publication of Lysozyme at an air-water interface presented in Chapter 3. With respect to studying hierarchical peptoid self-assembly, these challenges were first introduced to us with the goal of using simulations to inform their assembly into macromolecular structures.[201,290,511] In pursuit of this, we contributed to the development of a coarse-grained model.[488] Further interested by the low-level assemblies that can be formed by these molecules, we proposed a thermodynamic basis by which they form secondary structure in solution in Chapter 7. This is the first study to date that has been able to obtain this level of thermodynamic and atomistic resolution for a peptoid 12mer. More importantly, this method can serve as a template for studying and eventually the rational design of new peptoid sequences that enable even greater stability, and functional activity.

Overall, the study of structure-function relationships at interfaces remains to be a great challenge. Taking inspiration from how AI has shifted the field of protein structure discovery, future approaches should aim to leverage big data, high throughput computation, and machine learning methods to advance the field. This dissertation has been a small sample of how

computational methods can be used to study applied systems providing high information content and strong agreement with experiment. Future work should build upon these approaches, applying them to generate large data sets. Information-rich physics-based simulations will be critical for generating data where not experimentally feasible. These *in silico* screening platforms could then be used to uncover the fundamental building blocks of organic biological systems which make them so specific and efficient at their jobs. This molecular template would provide a chemical and physical basis for materials synthesis, and thus inform the design of new highly functional biomaterials. Deep learning models have already been applied to a handful of interfacial systems; RNA and DNA-binding proteins[512–514], binding site predictions[515], and protein-ligand binding[516]. We hope the methods/approaches outlined in these chapters will inspire future development of high-throughput frameworks to study many bioinspired systems.

Appendix A

Supporting Information:

Chapter 2. Impact of Glutamate Carboxylation in the Adsorption of the α-1 Domain of Osteocalcin to Hydroxyapatite and Titania

Sarah Alamdari and Jim Pfaendtner

Dept. of Chemical Engineering, University of Washington, Seattle 98195-1750

Gamma Carboxyl Glutamic Acid Forcefield Parameters

The following list includes the atom names, and the corresponding CHARMM36 atom types that were used to model γ-carboxyglutamic acid (Gla). These were adjusted from existing CHARMM36 atom types. A general CH atom was used to model the $C_\gamma$atom, and partial charges for that atom were adjusted as shown by column 3 so that the entire Gla amino acid had an overall charge of -2, and each carboxylate group retained the same overall charge. This approach has also been used in previous studies.[517,518] Other studies have shown that this Gla-carboxylate/Ca^{2+} interaction has both been over and underestimated by previous iterations of the CHARMM forcefield parameters, but have not yet been investigated yet with the CHARMM36 release, or the $Ca^{1.5+}$ model used by INTERFACE.[53,519] In this study, due to the different ions used within the surface model constraints, a generalized approach parameterized from CHARMM36 was used.

Partial charges for Gla

N	NH1	-0.470	0
HN	H	0.310	1
CA	CT1	0.070	2
HA	HB1	0.090	3
CB	CT2A	-0.180	4
HB1	HA2	0.090	5
HB2	HA2	0.090	6
CG	CT1	-0.290	7
HG1	HA1	0.090	8
JA1	CC	0.620	9
OE1	OC	-0.760	10
OE2	OC	-0.760	11
JA2	CC	0.620	12
OE3	OC	-0.760	13
OE4	OC	-0.760	14
C	C	0.510	15
O	O	-0.510	16

Table A-S1. Specific simulation setup details for PT-WTE-MTD

Simulation	Peptide	Surface	Charge			Counterions		
			Surface	Peptide	Total	$Ca^{1.5+}$	Ca^{2+}	Cl^-
1	dOC	HAP	0	-2	-2	2	-	1
2	OC	HAP	0	-5	-5	4	-	1
3	dOC	Neutral Hydroxylated TiO_2	0	-2	-2	-	1	0
4	OC	Neutral Hydroxylated TiO_2	0	-5	-5	-	3	1

Table A-S2. Specific simulation setup details for wtMTD

Simulation	Peptide	Surface	Charge			Counterions		
			Surface	Peptide	Total	$Ca^{1.5+}$	Na^+	Cl^-
5	GLU	HAP	0	-1	-1	2	-	2
6	GLA	HAP	0	-2	-2	2	-	1
7	GLU	Neutral Hydroxylated TiO_2	0	-1	-1	-	1	0
8	GLA	Neutral Hydroxylated TiO_2	0	-2	-2	-	2	0

Table A-S3. Temperatures used for each replica of PT-WTE

HA		TiO_2	
Replica	**Temperature (K)**	**Replica**	**Temperature (K)**
1	300	**1**	300
2	305	**2**	305
3	310	**3**	311
4	316	**4**	316
5	321	**5**	322
6	326	**6**	328
7	332	**7**	333
8	337	**8**	339
9	343	**9**	345
10	349	**10**	352
11	355	**11**	358
12	361	**12**	364
13	367	**13**	371
14	373	**14**	377
15	380	**15**	384
16	387	**16**	391
17	393	**17**	398
18	399	**18**	405
19	406	**19**	412
20	414	**20**	419
21	421	**21**	426
22	428	**22**	434
23	435	**23**	442
24	442	**24**	450
25	450	**-**	-

Convergence of Enhanced Sampling Simulations and Reweighting

Binding free energies were calculated using a Boltzmann averaged difference of the Helmholtz free energy from the surface adsorbed and solution states, as shown in **Equations A-S1** and **A-S2**. Here r represents the peptide center of mass (COM) distance to the surface. The region from r_0 to r_1 is the region where the peptide COM was considered bound (distance <= 3.0), and r_2 is the distance at which the wall was placed as described in the Methods section. The time series monitoring binding energies are shown in **Figure A-S1**.

$$A_{surface} = \frac{1}{r_1 - r_0} \int_{r_0}^{r_1} e^{-\frac{A(r)}{k_B T}} dr \quad , \qquad A_{solution} = \frac{1}{r_2 - r_1} \int_{r_1}^{r_2} e^{-\frac{A(r)}{k_B T}} dr \tag{A.S1}$$

$$A_{binding} = -k_B T \ln\left(\frac{A_{surface}}{A_{solution}}\right) \tag{A.S2}$$

The binding energy reported in **Figure A-S1** was calculated by taking the mean of the Boltzmann weighted binding energies calculated during the last 30% of simulation time. The binding energies for capped Glu and Gla were calculated to be -17.55 kJ/mol $\pm$ 0.27 kJ/mol and -22.7 kJ/mol $\pm$ 1.05 kJ/mol on HA, and -15.4 kJ/mol $\pm$ 1.13 kJ/mol and -26.0 kJ/mol $\pm$ 0.97 kJ/mol on TiO_2, respectively. The binding energies of dOC are -16.0 $\pm$1.2 kJ/mol and -40.5$\pm$0.9 kJ/mol for OC on HA, and -67.9 $\pm$2.5 kJ/mol for dOC and -6.1 $\pm$1.7 kJ/mol for OC on TiO_2 respectively. Errors reported were calculated by taking the standard deviation of binding energy from the last 30% of the simulation.

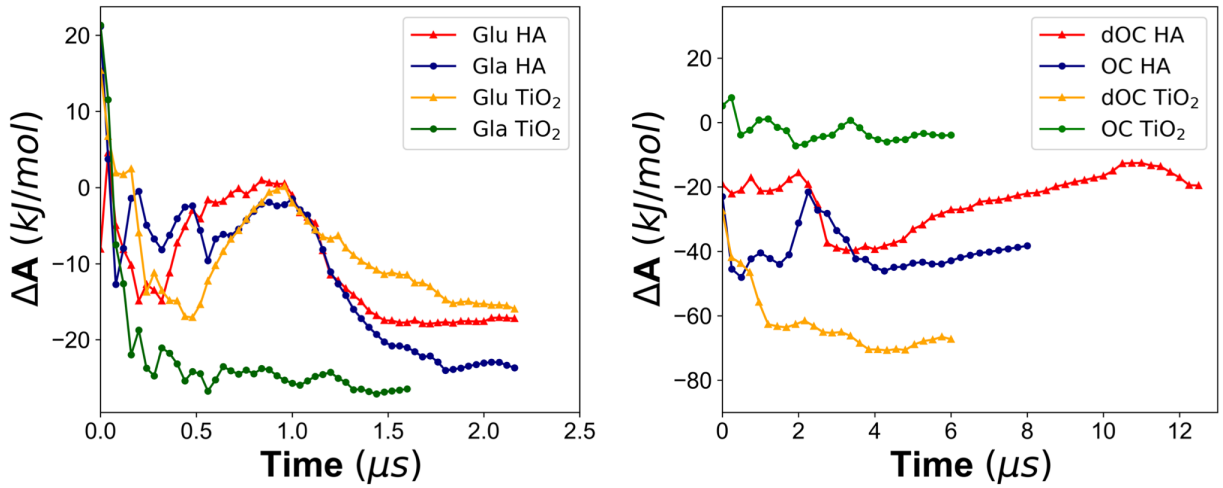


Figure A-S1. Binding free energy calculated from **Equation A-S2** as a function of time for all wtMTD simulations of capped amino acids (left) plotted every 2.5 ns, for clarity. Binding energy calculated for the 300K replica in PTMetaD-WTE simulations (right) plotted every 20 ns, for clarity.

Free-energies and other equilibrium averages were obtained from the biased simulations via reweighting. For all reweighting analyses we used static bias approximation. That is, the trajectory of interest is re-processed using the final MetaD bias potential from each simulation and each frame is assigned a weight based amount of bias given by the use of the final static bias; weights are obtained with the Torrie-Valleau[442] scheme (n.b., this method was inspired by Branduardi et al.[401] and was shown to provide exact agreement with reference potentials in multiple prior studies[235,413])

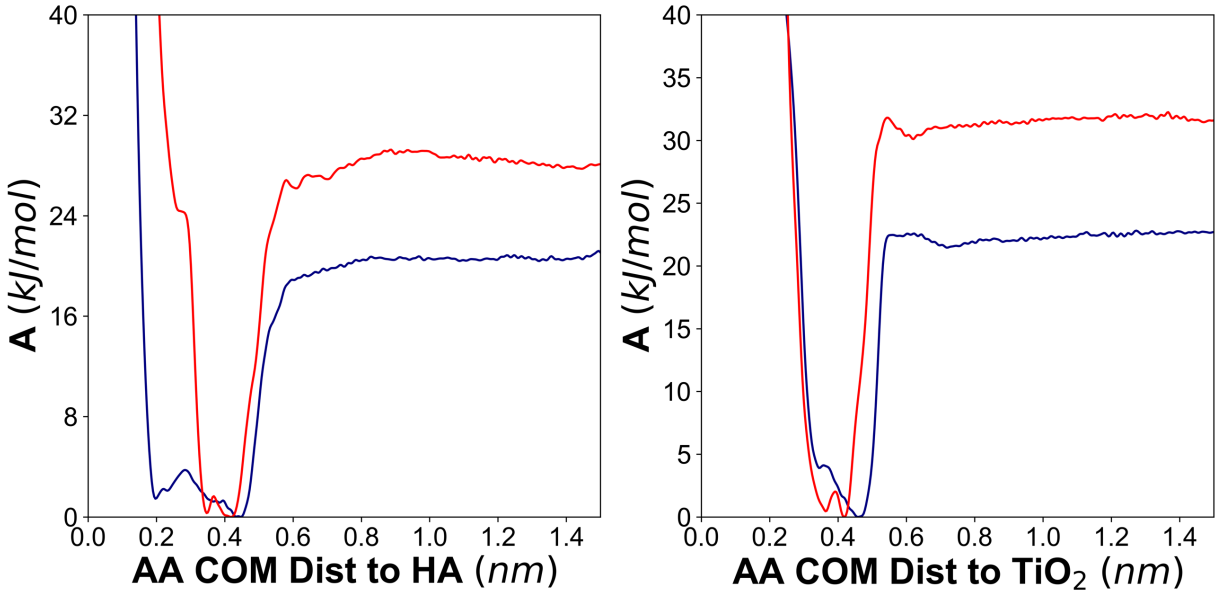


Figure A-S2. Free energy profile for Glu (blue) and Gla (red) binding onto HA (left) and TiO_2 (right). Whereas the binding profiles in Figure 3 and 4 show the reweighted free-energy as a function of the side-chain COM, these profiles represent the free-energy of the COM of the entire amino acid, derived from the sum of the MetaD bias potential

Clustering

Clustering was done using the gromos520 method developed by Daura et al. implemented in GROMACS. Gromos is a geometric clustering method that uses a root mean square deviation (RMSD) distance metric to group structures given a cutoff value. Peptides were clustered based on RMSD of the alpha carbon atoms in the peptide backbone using a cutoff value of 0.2 nm.521 Structures were processed every two to four frames using the ensemble of biased structures at 300K, classified based on their COM distance to the surface (distance <= 3.0 nm away from the surface). The cluster weights from this analysis were then reweighted as described above to correct for the effect of the MetaD bias. The top 3 clusters for dOC on HA and TiO2 are included in Figure A-S3, and OC on HA and TiO2 are included in S4. The top 3 structures in Figure A-S3 and A-S4 represent at least 50% of the peptide conformations observed at the surface. The top 3 solution structures are shown in Figure A-S5 and A-S6, to demonstrate conformational preferences in solution. Convergence of clustering shown in Figure A-S5, was performed by calculating the total number of clusters as a function of trajectory frames clustered.

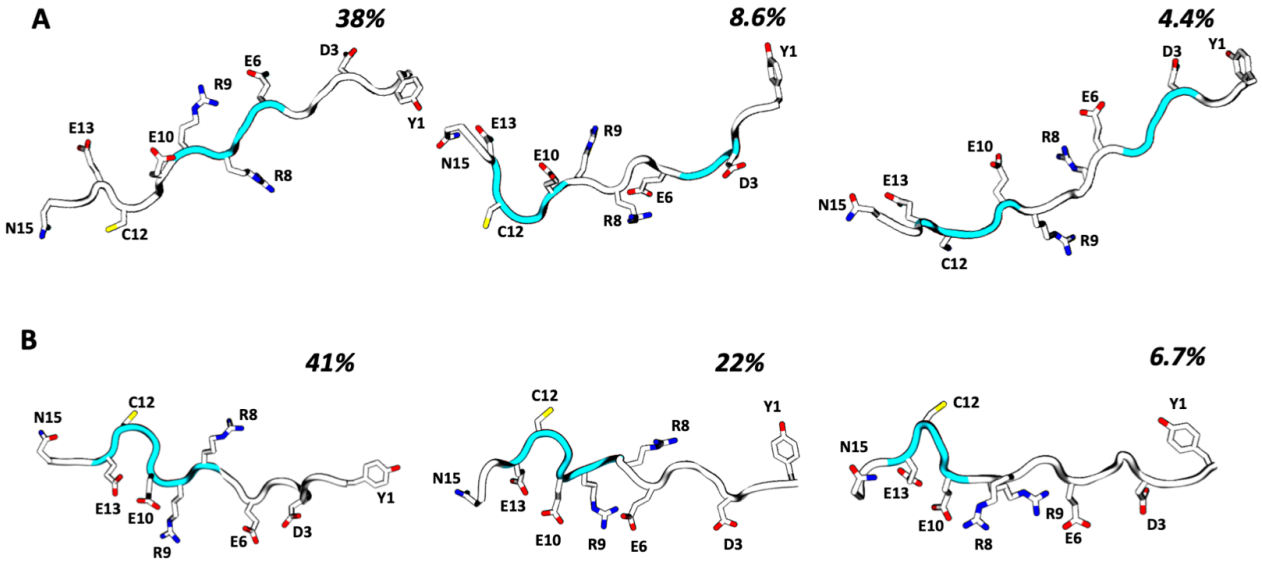


Figure A-S3. Top 3 structures from clustering analysis for adsorbed dOC in row A) HA and row B) TiO_2 surfaces. Percentages listed over the structures correspond to the reweighted % of simulation that structure represents.

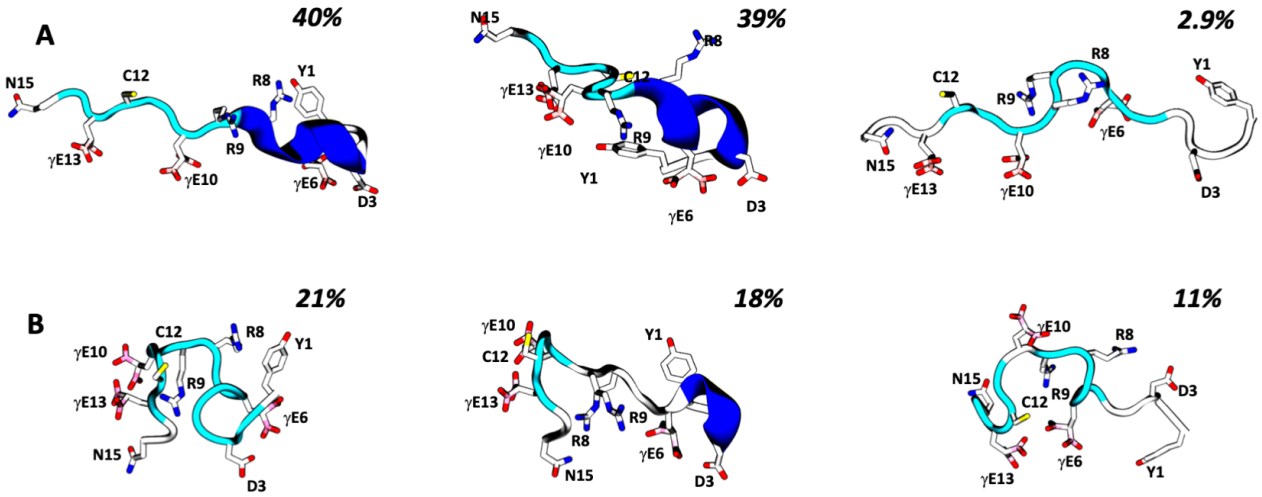


Figure A-S4. Top 3 structures from clustering analysis for adsorbed OC in row A) HA and row B) TiO_2 surfaces. Percentages listed over the structures correspond to the reweighted % of simulation that structure represents.

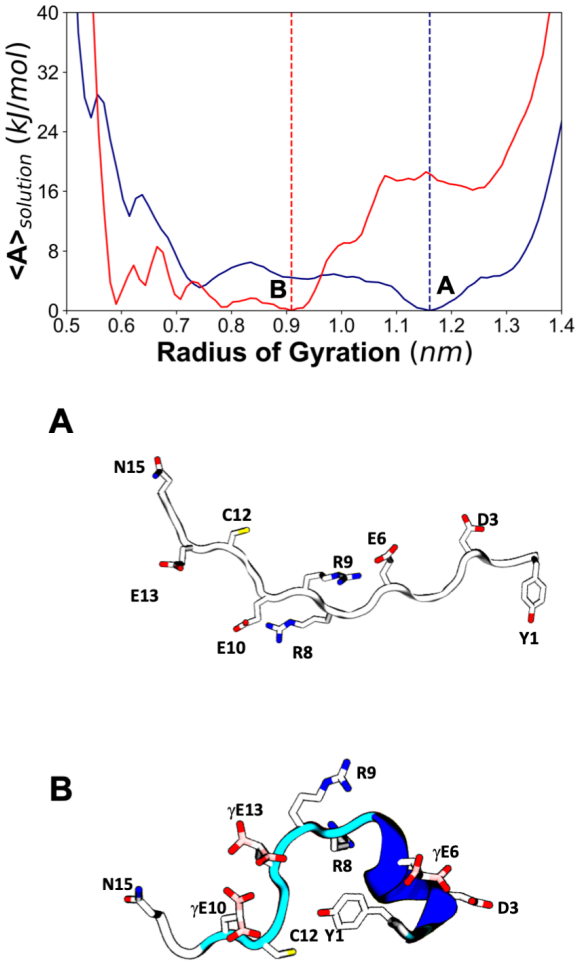


Figure A-S5. Free energy profiles for dOC and OC in HA systems along the radius of gyration for structures in solution (COM distance > 3.0 nm)

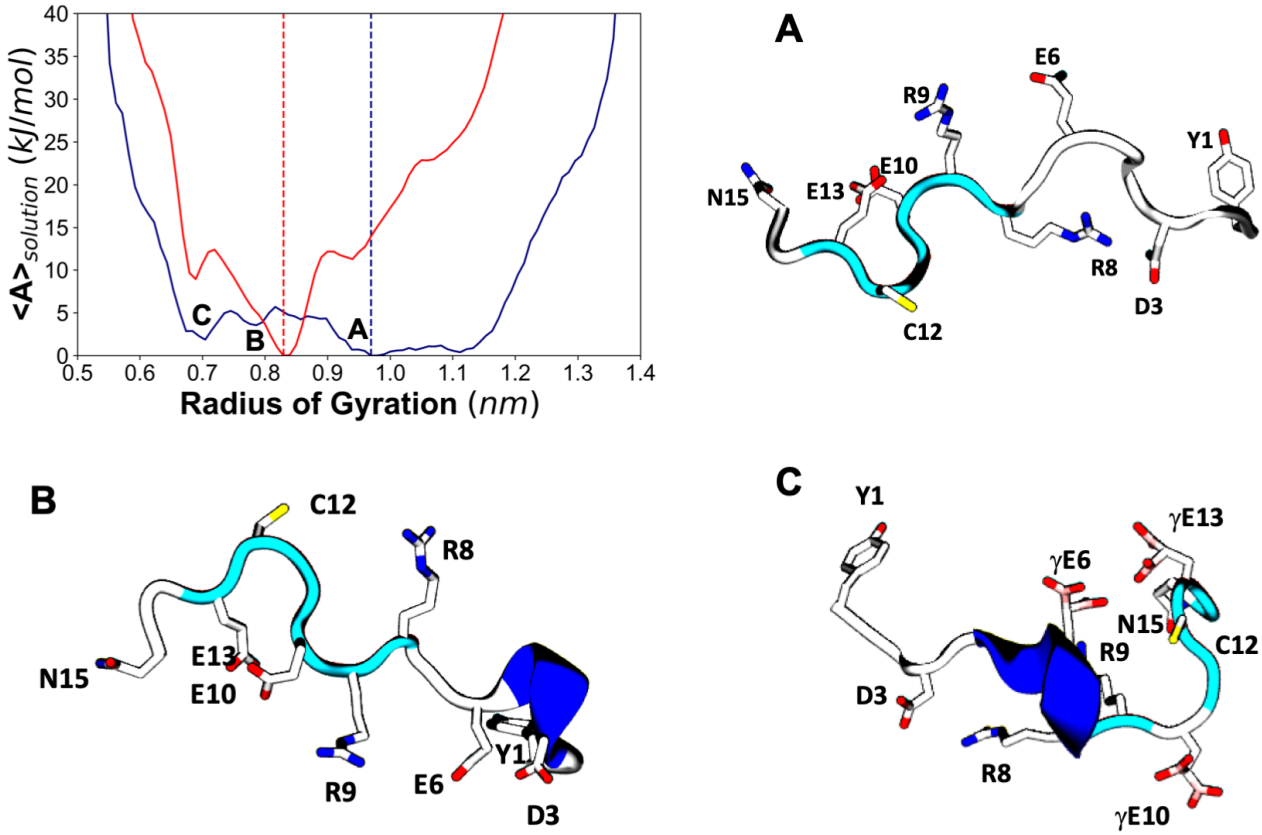


Figure A-S6. Free energy profiles for dOC (blue) and OC (red) in TiO_2 systems along the radius of gyration for structures in solution (COM distance to the surface > 3.0 nm).

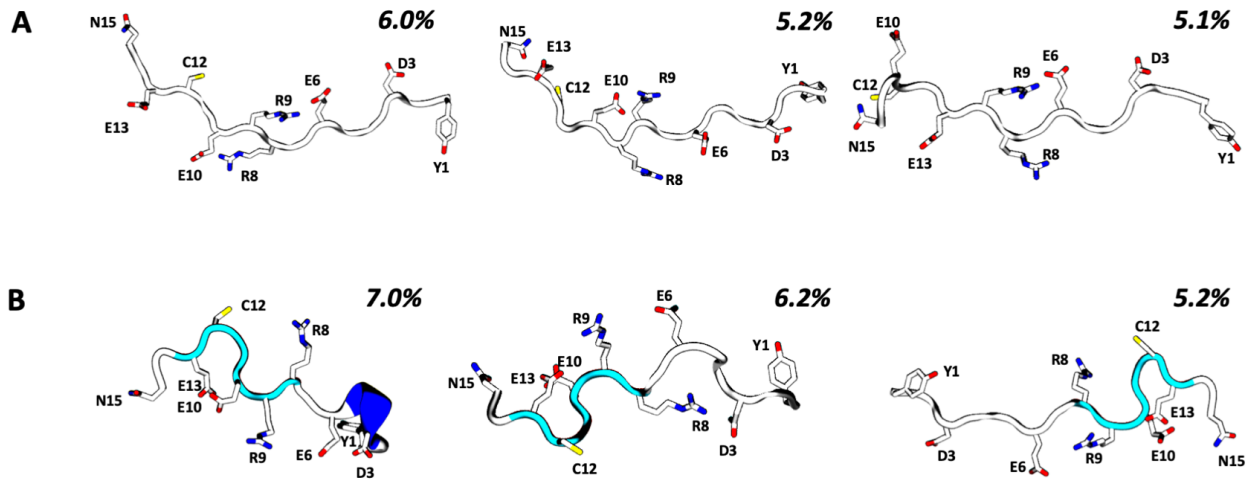


Figure A-S7. Top 3 structures from clustering analysis for solution dOC in row A) HA systems and row B) TiO_2 systems. Percentages listed over the structures correspond to the reweighted % of simulation that structure represents

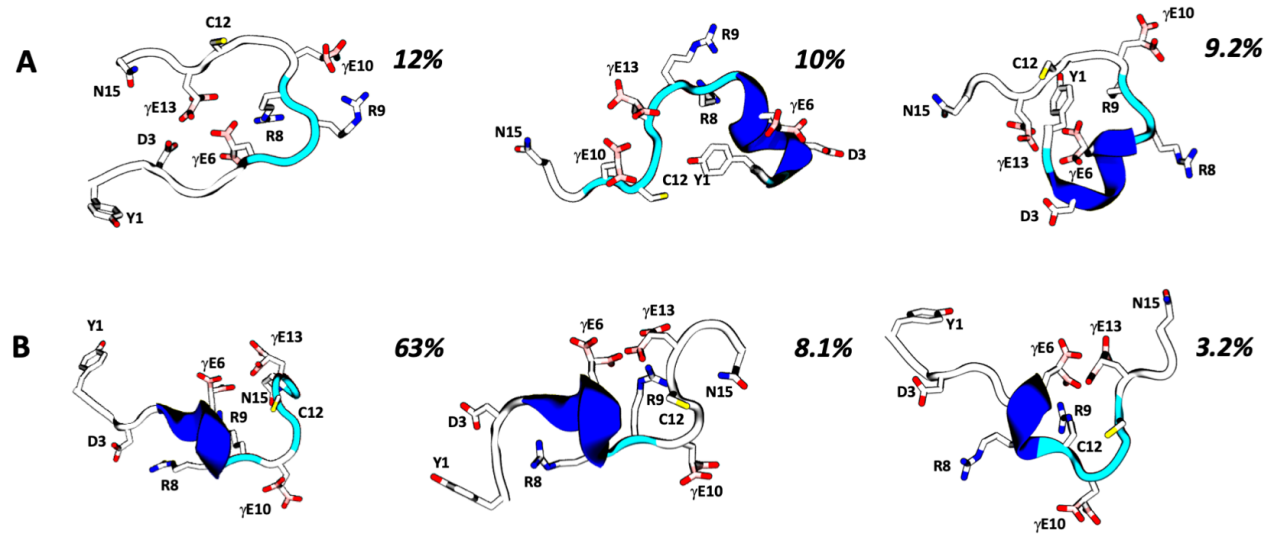


Figure A-S8. Top 3 structures from clustering analysis for solution OC in row A) HA system and row B) TiO_2 systems. Percentages listed over the structures correspond to the reweighted % of simulation that structure represents.

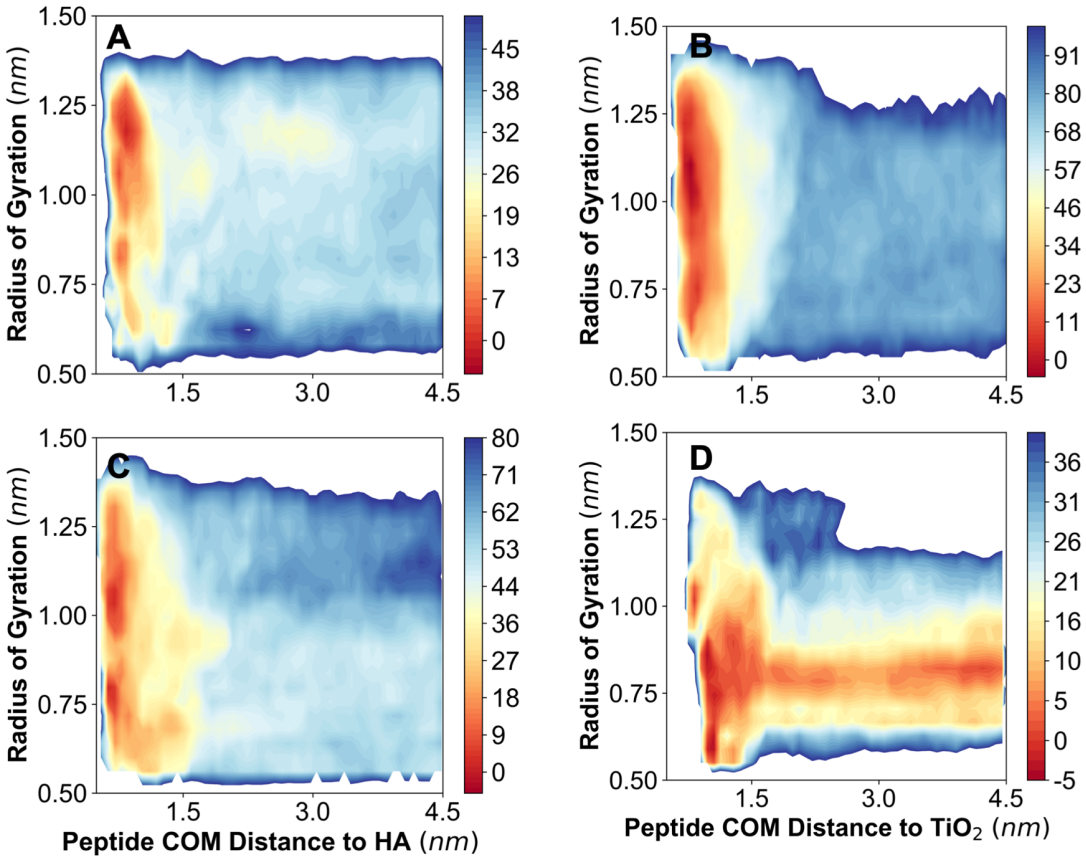


Figure A-S9. FES for A) dOC on HA, B) dOC on TiO_2, C) OC on HA, and d) OC on TiO_2. Note the difference in energy scales, given in kJ/mol for clarity in seeing the minimum energy bound states.

Characterization of surface bound structure

To characterize the minimum energy adsorbed clusters, simulations were reweighted using Torrie-Valleau[442] reweighting scheme using only the surface adsorbed structures (peptide COM distance <= 3.0 nm). The structures were then reweighted along 2 coordination number (CN) CVs, to produce a free energy surface (FES) describing the environment between the peptide, and ions (Ca) or hydroxyls (H) at the surface. The CN is represented by a continuous switching function between atoms i and j following the form given in **Equation A.3** where r_0, n, and m are adjustable parameters to control the shift, and slope of the switching function.

$$CN = \frac{1 - \left(\frac{r_{ij}}{r_0}\right)^n}{1 - \left(\frac{r_{ij}}{r_0}\right)^m} \qquad (A.3)$$

The reweighted CVs shown are: (1) the CN between each oxygen on every carboxylate in Glu or Gla and all calcium ions (r_0 = 0.35 nm), and (2) the CN between each oxygen and hydrogen on surface hydroxyls (r_0 = 0.2 nm). This was done to capture the amount of surface contacts through the unique interaction with each O on the Glu and Gla residues with the surface. Values of 6, and 12 were used for n and m in both CVs.

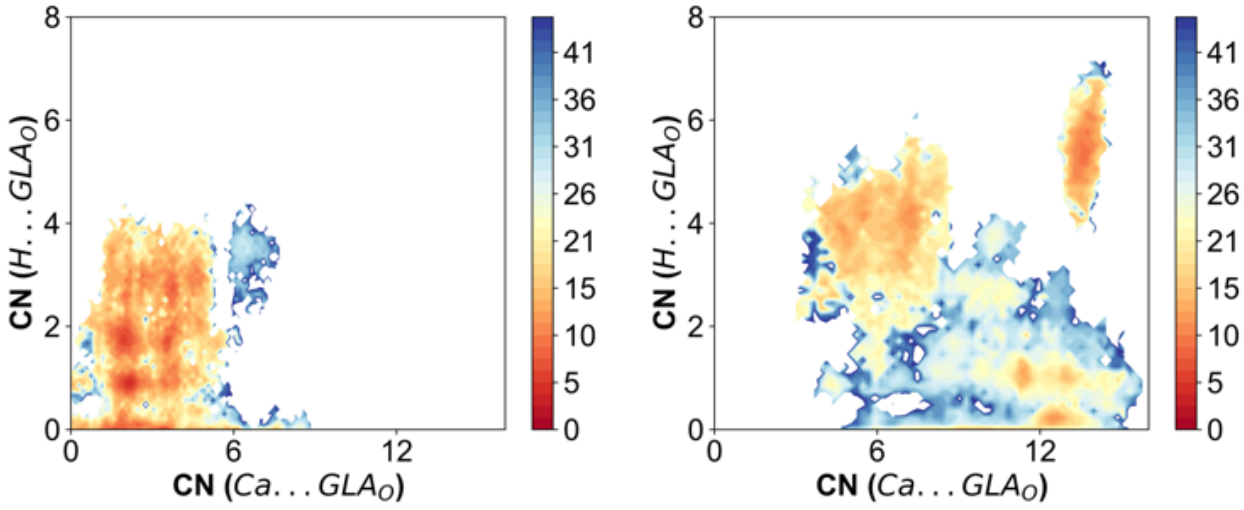


Figure A-S10. Characterization of the surface bound structures of dOC (left) and OC (right) on HA

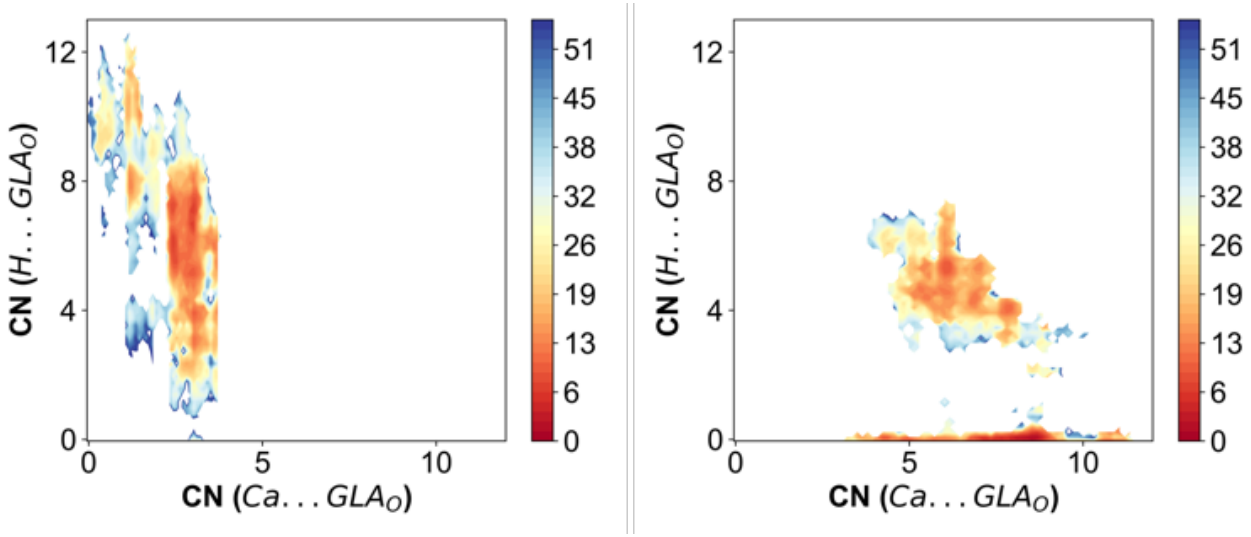


Figure A-S11. Characterization of the surface bound structures of dOC (left) and OC (right) on TiO_2

Characterization of OC helical structure

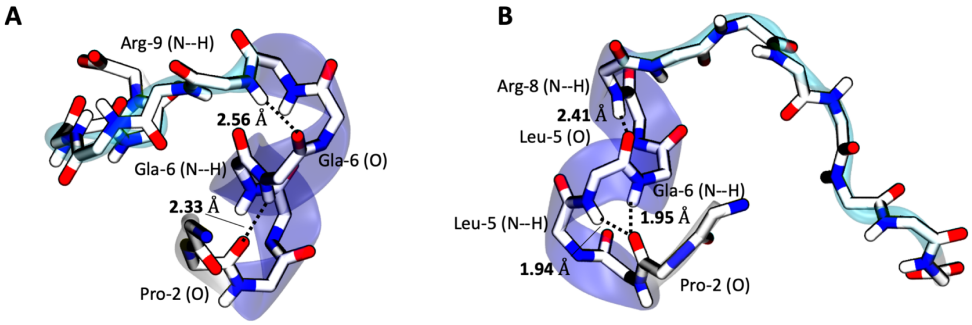


Figure A-S12. Characterization of the kinked helical structure in OC, showing only the backbone atoms in licorice, and transparent overlay of the helical structures obtained from the proteins adsorbed on HA. **A)** A tight helix formed by i+4 h-bonding with a proline kink in the first helix, followed by i+3 h-bonding in the second helix. **B)** Another tight helix formed by combined i+3, i+4 bonding in the first helix with a proline kink with residues Leu-5 and Gla-6, followed by the same i+3 hydrogen bonding mode in the second helix. Protein structure is shown transparently, and backbone residues are shown in licorice following the same color scheme as the **Figure 2.6**.

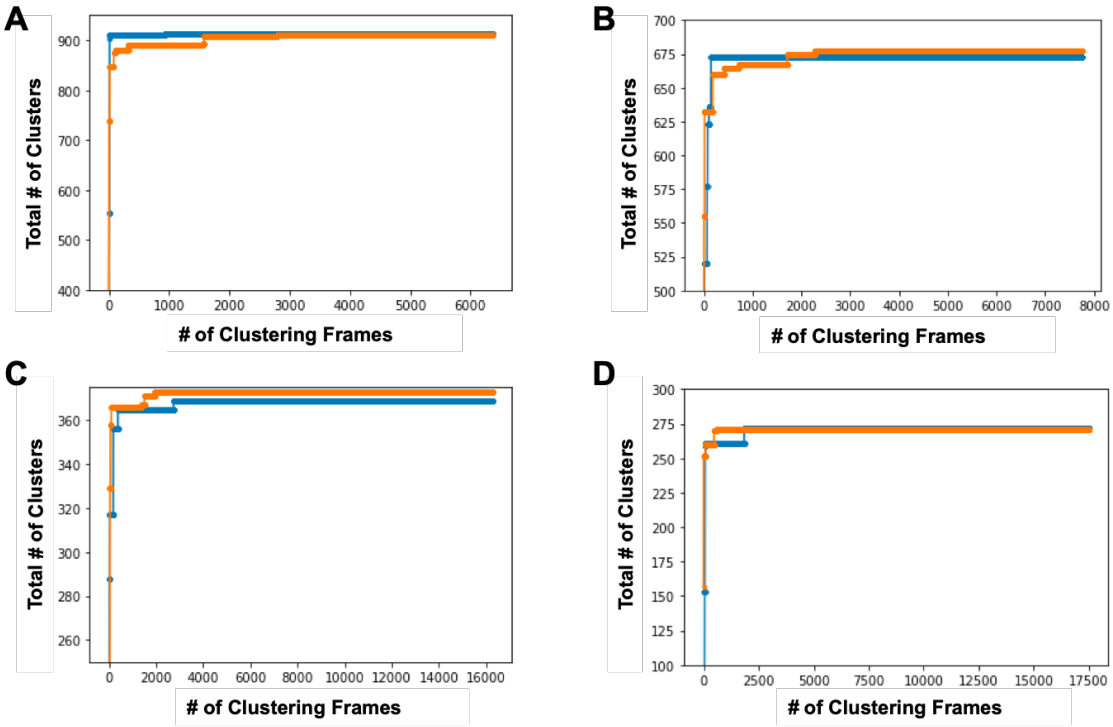


Figure A-S13. Convergence of clustering calculations for (A) dOC and C) OC on HA; and B) dOC and D) OC on TiO2. Convergence of surface clusters (peptide COM distance <= 3.0 nm from the surface) are given by the blue line, and solution clusters (peptide COM distances > 3.0 nm from the surface) are shown by the orange line.

Appendix B

Supporting Information:

Chapter 3. Orientation and Conformation of Proteins at the Air-Water Interface Determined from Integrative Molecular Dynamics Simulations and Sum Frequency Generation Spectroscopy

‡Sarah Alamdari[1], ‡Steven J. Roeters[2], Thaddeus W. Golbek[2], Lars Schmüser[2], *Tobias Weidner[2], *Jim Pfaendtner[1]

[1]*Dept. of Chemical Engineering, University of Washington, Seattle 98195-1750, USA*
[2]*Department of Chemistry, Aarhus University, Langelandsgade 140, 8000 Aarhus C, Denmark*
‡Equal contribution

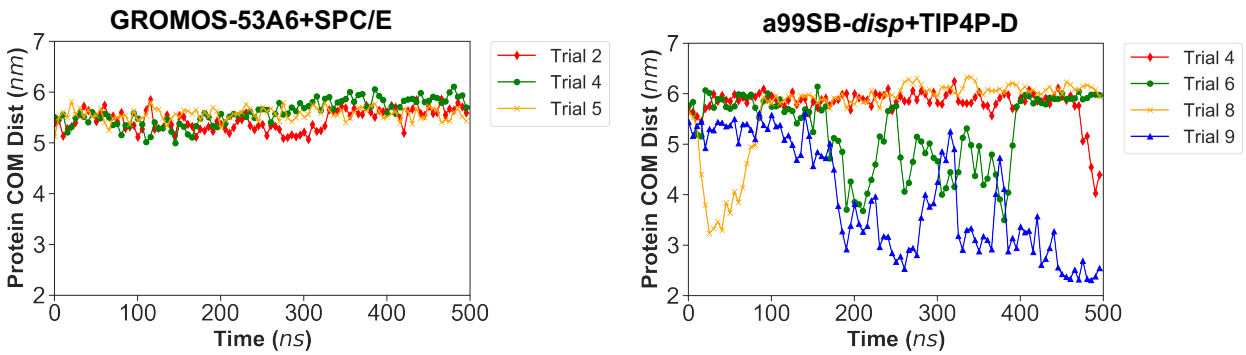


Figure B-S1. Extended time series data for select simulations monitoring the z-distance between the COM of protein to the furthest water atom for (**left**) GROMOS-53A6/SPC/E and (**right**) a99SB-*disp*/TIP4P-D

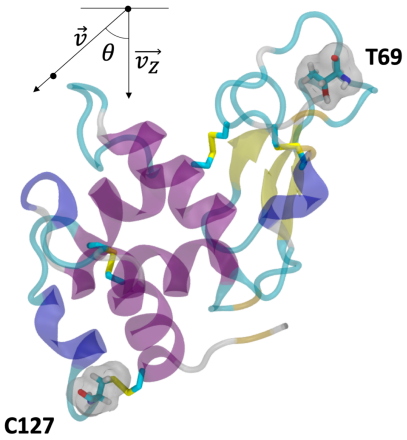


Figure B-S2. Description of the collective variable used to model orientation of lysozyme at the interface, the vector is defined as the distance from the COM of Cystine 127, and Threonine 69. Disulfide bonds are shown in licorice between residues 75 and 93, 63 and 79, 114 and 29, 98 and 126, and the protein is colored according to its secondary structure where alpha helices are given in purple or blue, beta sheets are shown in yellow, and coils are given in cyan.

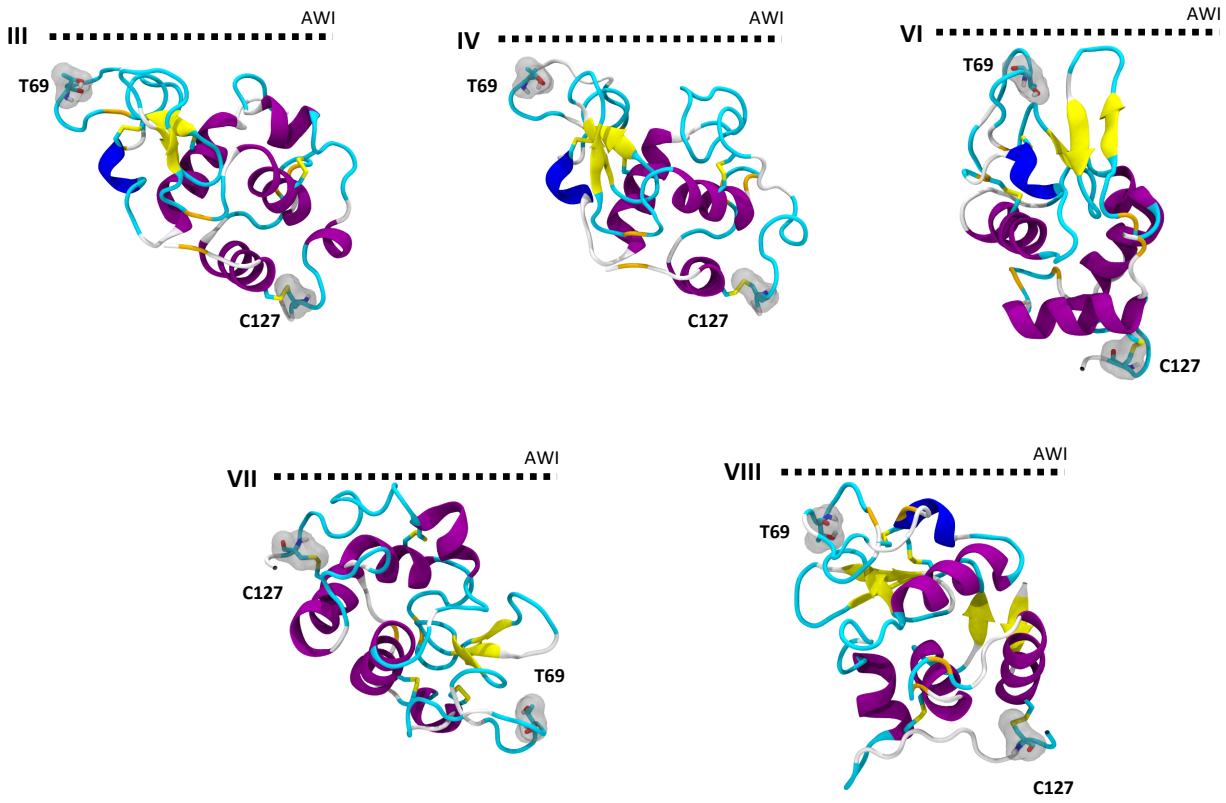


Figure B-S3. Additional interfacial structures obtained from GROMOS-53A6+SPC/E trials, corresponding to labeling scheme in **Figure 3.1**

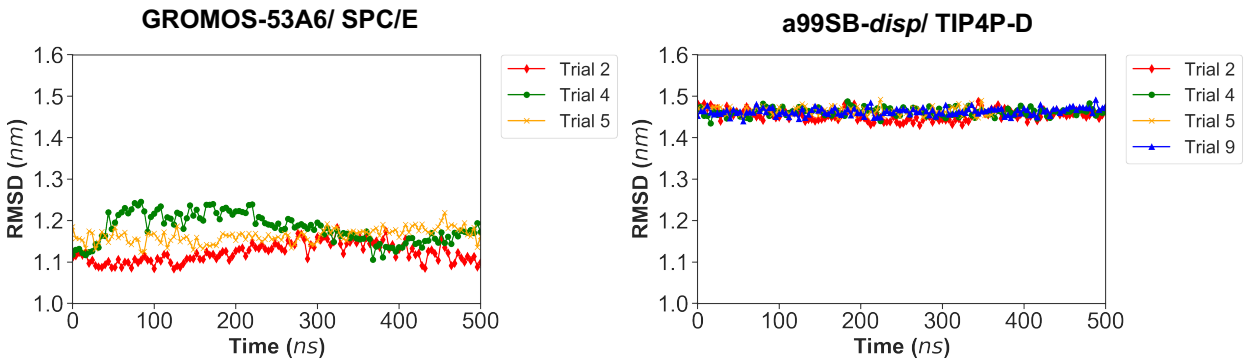


Figure B-S4. Root mean square deviation of α-carbons to monitor stability over time in reference to the PDB structure conformation of (**left**) GROMOS-53A6/SPC/E and (**right**) a99SB-*disp*/TIP4P-D

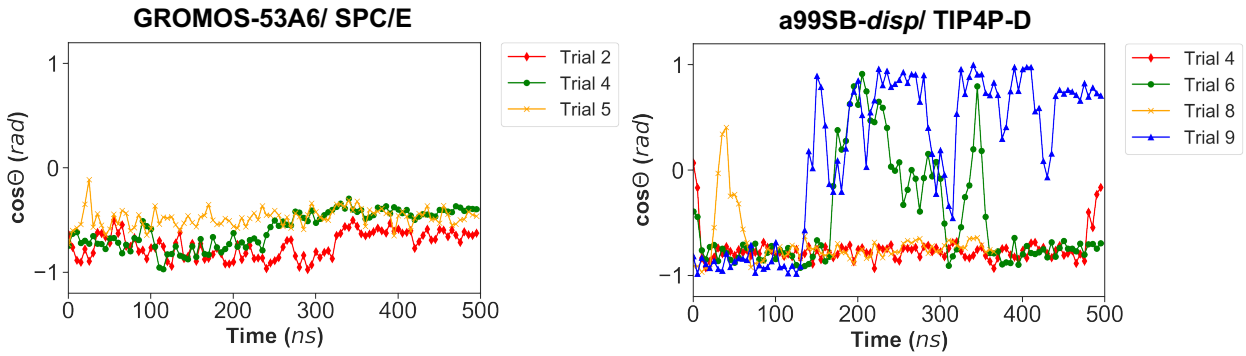


Figure B-S5. Extended time series data of simulations monitoring the orientation of (**left**) GROMOS-53A6/SPC/E and (**right**) a99SB-*disp*/TIP4P-D

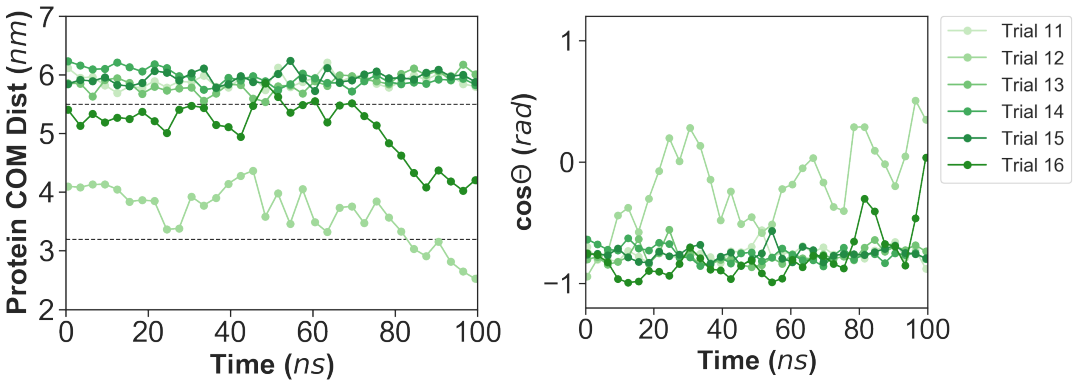


Figure B-S6. Raw data for interfacial distance to monitor adsorption and orientation for trials conducted at pH 11 using a99SB-*disp*/TIP4P-D. Trials 11, 12, 13, 14, 15 and 16 represent the structures started from trial 4, 5, 6, 7, 8, and 9 of the adsorbed poses at pH 7.

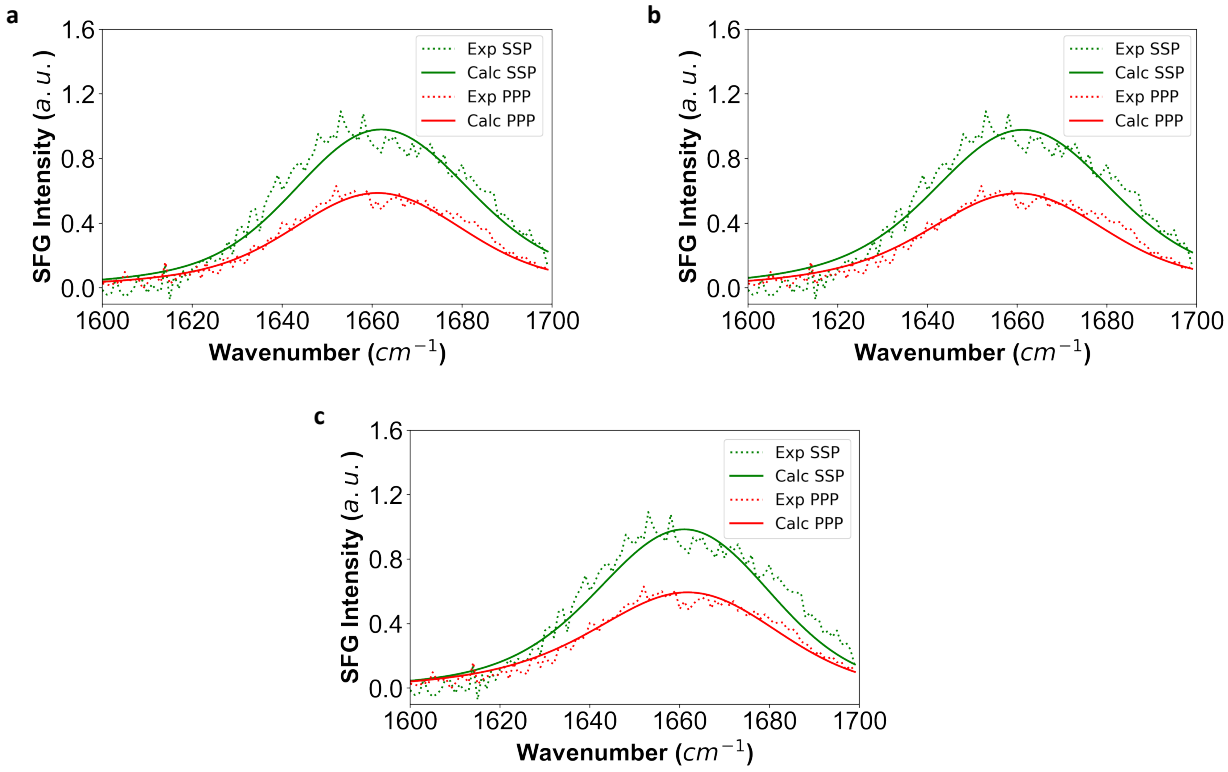


Figure B-S7. Calculated SFG signals given for a99SB-disp+TIP4D (**a**) trial 6 which adsorbs to pose I (axial) (**b**) trial 7 which adsorbs to pose I (axial) and (**c**) trial 9 which adorbs to pose II (head-on). The experimental signal is also shown for comparison.

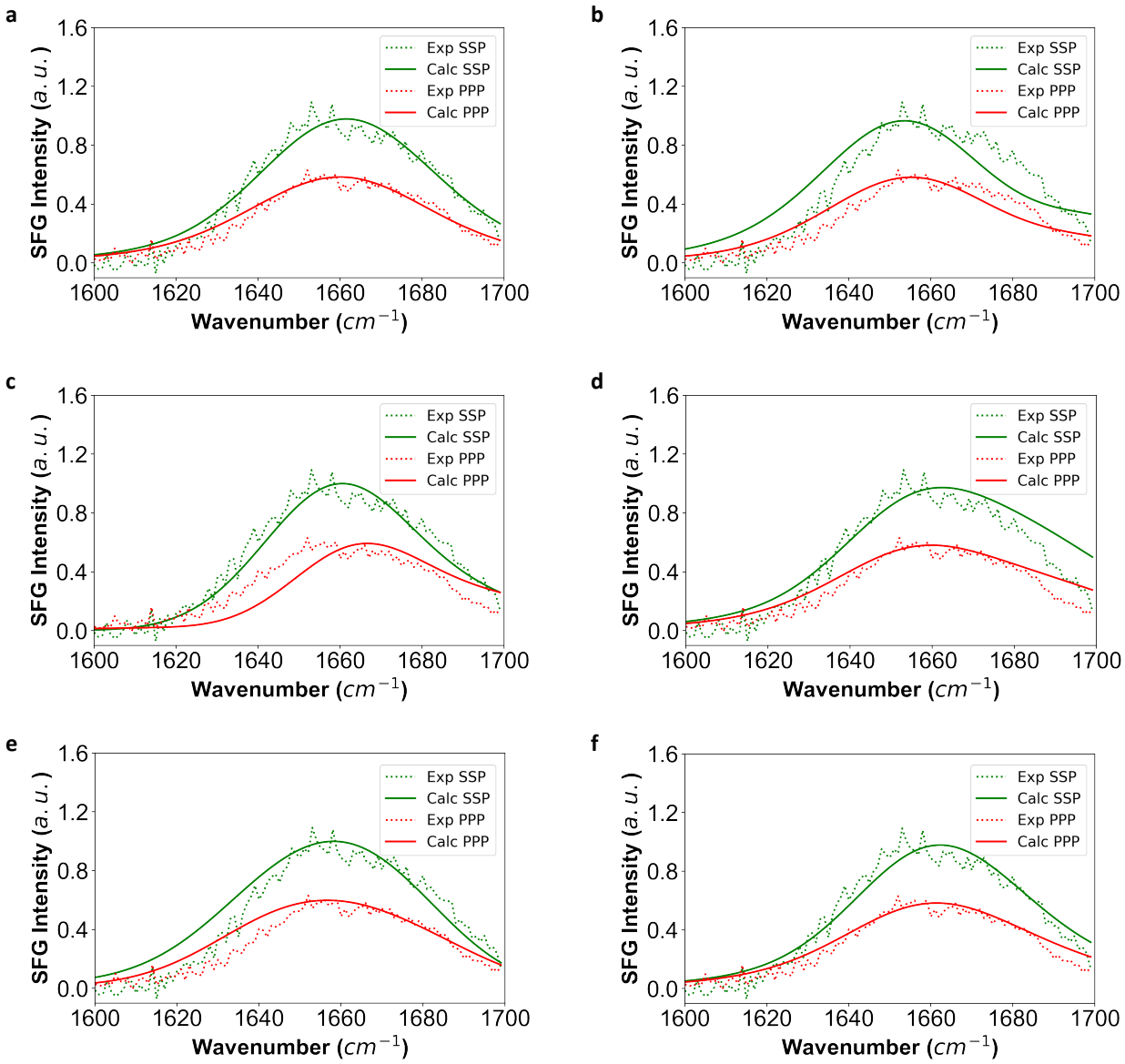


Figure B-S8. Calculated SFG signals given for GROMOS-53A6+SPC/E (**a**) trial 2 which adsorbs to pose III (**b**) trial 4 which adsorbs to pose VIII and (**c**) trial 5 which adorbs to pose V (**d**) trial 6 which adorbs to pose VI (**e**) trial 7 which adorbs to pose VII (**f**) trial 9 which adorbs to pose IV. The experimental signal is also shown for comparison.

Appendix C

Supporting Information:
Chapter 6. Substitution of Distal and Active Site Residues Reduces Product Inhibition of E1 from *Acidothermus Cellulolyticus*

Samantha R. Summers[1†], Sarah Alamdari[2†], Casey J. Kraft[1], Roman Brunecky[3], Jim Pfaendtner[2], Joel L. Kaar[1*]

[1]Department of Chemical and Biological Engineering, University of Colorado, Boulder, CO 80309, USA

[2]Department of Chemical Engineering, University of Washington, Seattle, WA 98195-1750, USA

[3]Biosciences Center, National Renewable Energy Lab, Golden, CO 80401, USA

[‡]Equal contribution

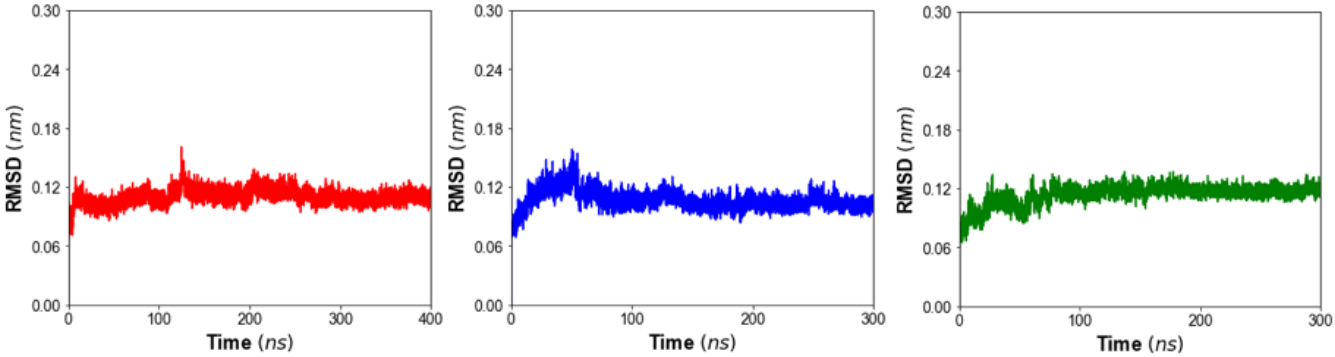


Figure C-S1. RMSD of the alpha-carbon backbones of the course of each metadynamics simulation, (red, left) WT, (blue, middle) Y245G, (green, right) W212A E1.

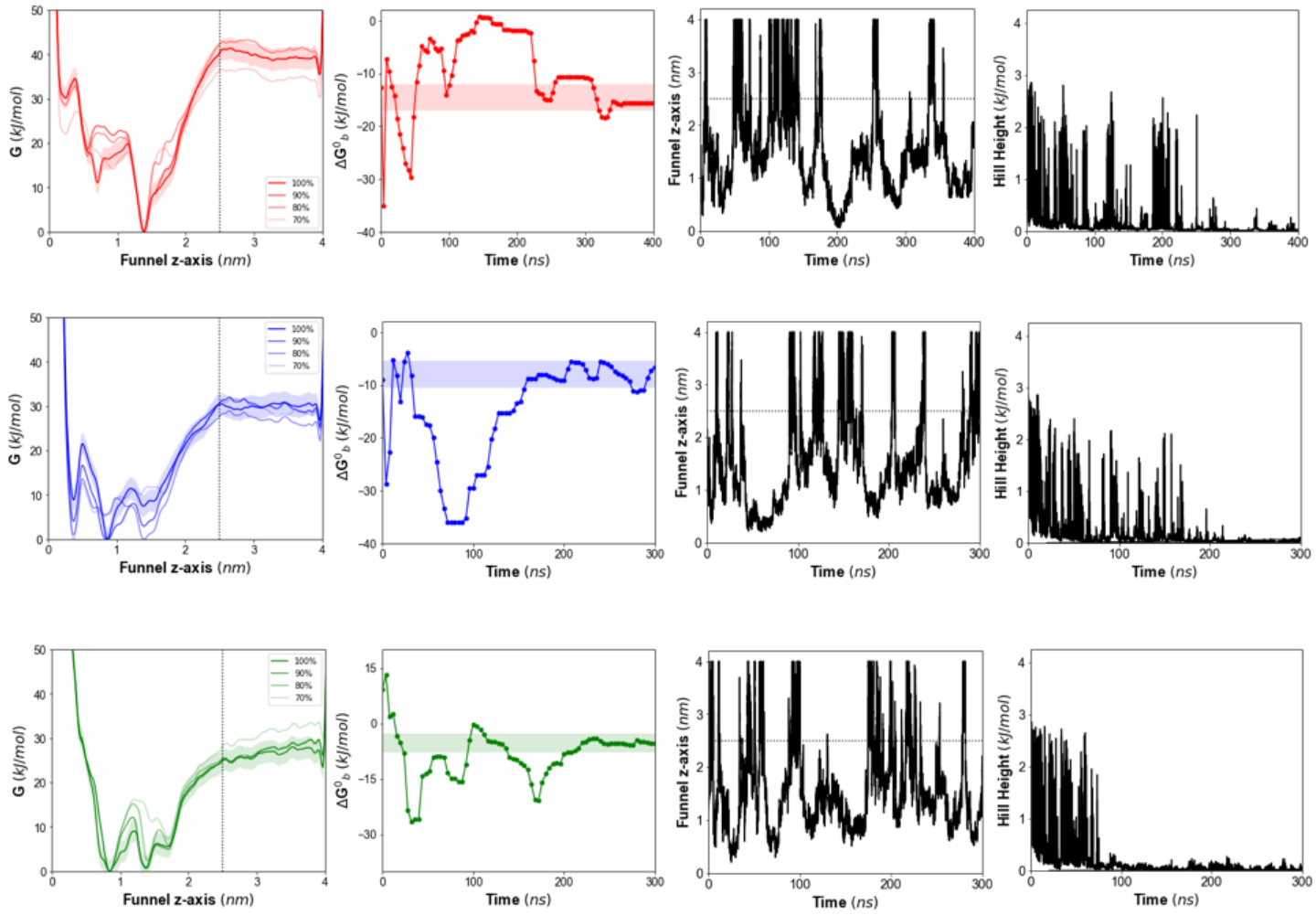


Figure C-S2. Demonstrating convergence for WT (red, top row), Y245G (blue, middle row), and W212A (green, bottom row) E1. For each system (from left to right, column 1) free energy profile (FEP) is plotted as a function of the funnel z-axis, for 70, 80, 90, and 100% of the total simulations time. A dotted line at 2.5 nm is used to indicate the bound state vs. the unbound state. An error bar is plotted over the FEP to show standard deviation over the last 30% of simulation time. (Column 2) The change in free energy as a function of time is plotted with the entropic correction (3.8 kcal/mol) applied. A solid line is used to highlight the average binding energy and its standard deviation over the last 30% of simulation. (Column 3) Protein-ligand binding as a function of simulation time, where 2.5 nm is highlighted with a dotted line to indicate the bound and unbound states. (Column 4) Gaussian hill height as a function of time.

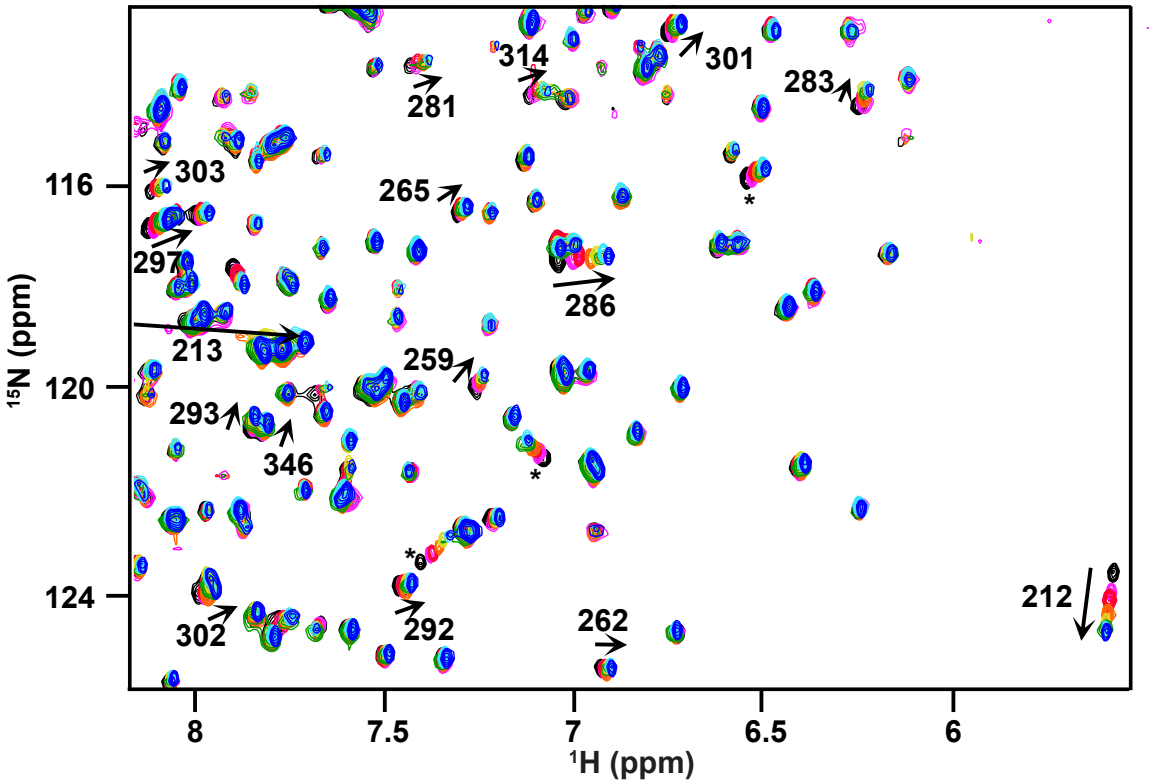


Figure C-S3. Overlay of near complete $^{1}H-^{15}N$ HSQC spectra of WT E1 in the presence of 0 (black), 1 (pink), 2 (red), 6 (orange), 10 (yellow), 15 (green), 21 (cyan), and 30 (blue) mM cellobiose. The overlay of the spectra shows many of the active site and distal residues in the ring region that were significantly perturbed ($\Delta\delta > 1\sigma$ from average). The labels and arrows indicate the residue number and direction of the chemical shift perturbations, respectively. Additionally, the asterisk indicates peaks that were significantly perturbed, but for which there was no residue assignment.

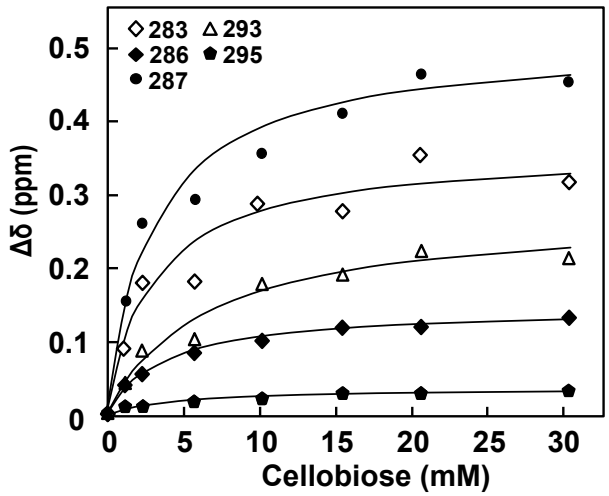


Figure C-S4. Binding isotherms for the interaction of cellobiose with residues in the perturbed ring of WT E1. The lines represent each residue fitted individually to a single-site, reversible binding model for each residue in fast exchange.

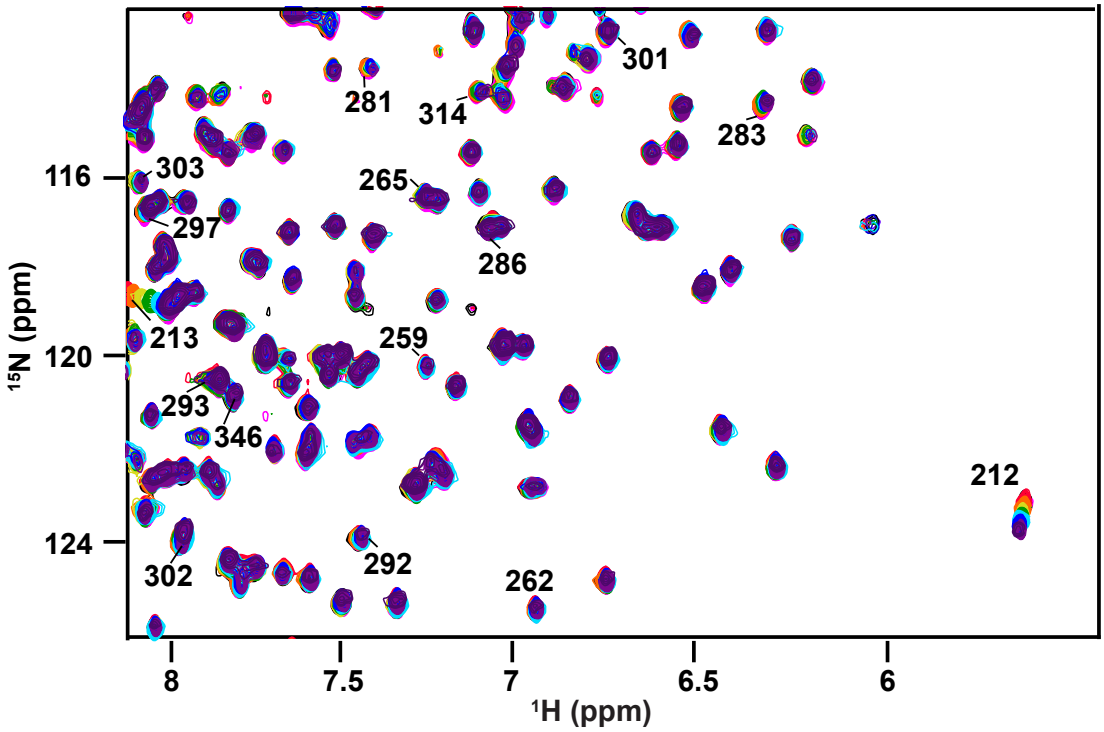


Figure C-S5. Overlay of near complete 1H−15N HSQC spectra of Y245G E1 in the presence of cellobiose. The overlay shows spectra with 0 (black), 0.6 (pink), 3 (red), 9 (orange), 21 (yellow), 31 (green), 46 (cyan), 63 (blue), 90 (light purple), and 110 (dark purple) mM cellobiose. Residues in WT E1 with significant perturbations ($\Delta\delta > 1\sigma$ from average) in the presence of cellobiose are numerically labeled, whereas residues significantly perturbed in the presence of cellobiose for Y245G E1 have arrows indicating the direction of chemical shift perturbations.

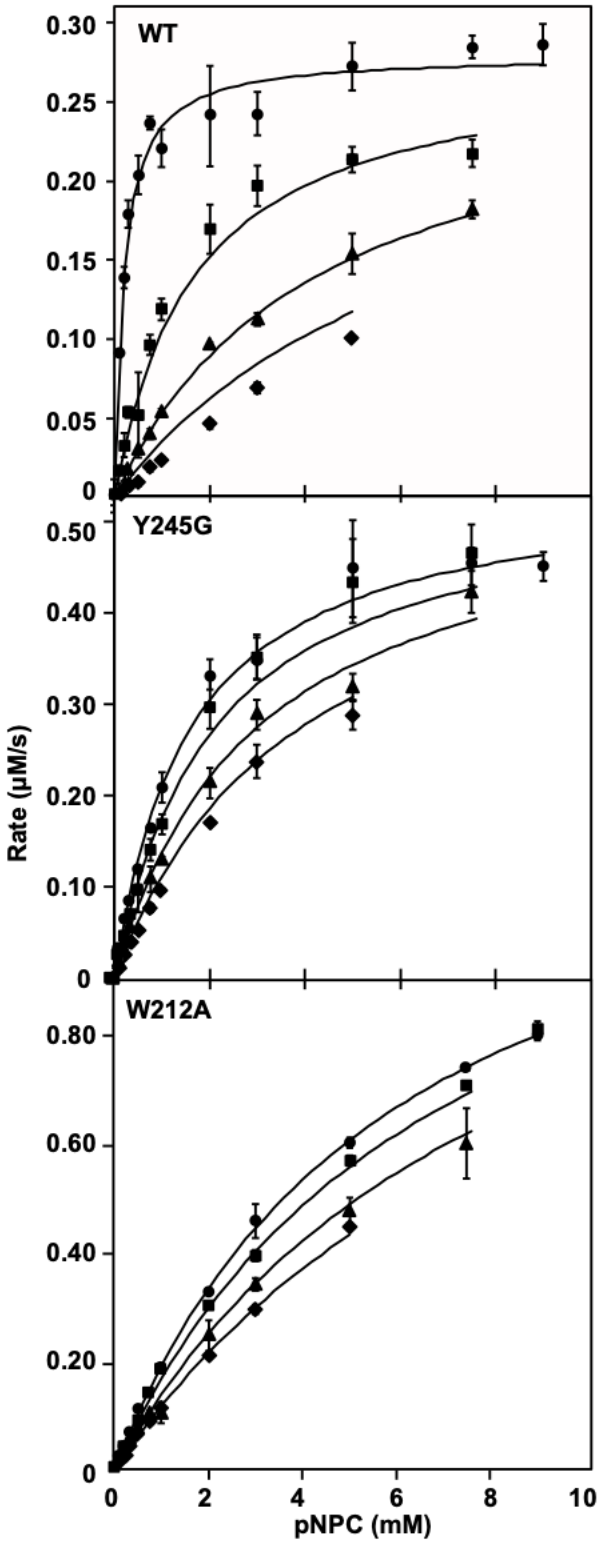


Figure C-S6. Activity of WT, Y245G, and W212A E1 using pNPC as a model substrate in the presence of 0 (circle), 0.022 (square), 0.06 (triangle), and 0.1 M (diamond) cellobiose. Lines represent the fit to a linear competitive inhibition model. Error bars represent the mean of three independent reactions.

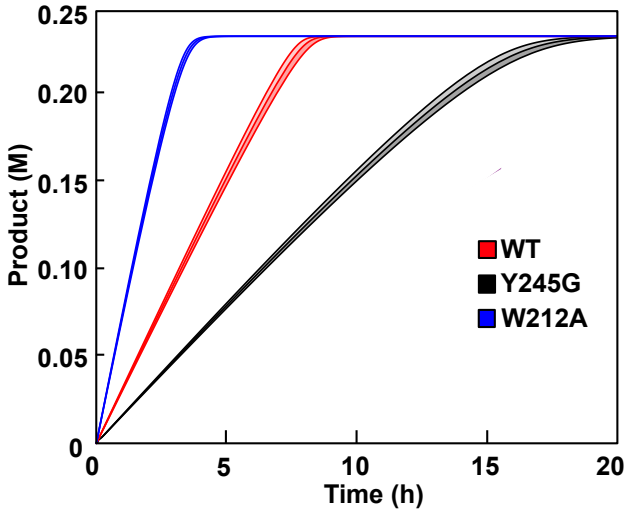


Figure C-S7. Theoretical productivity curves for WT, Y245G, and W212A E1 at 20 wt % (0.23 M) substrate loading. Productivity curves were generated using integrated Michaelis-Menten equations for linear, competitive product inhibition and kinetic parameters from E1 activity on pNPC. Maximum error for each variant was determined from the error of the kinetic constants of each variant on pNPC (shaded area).

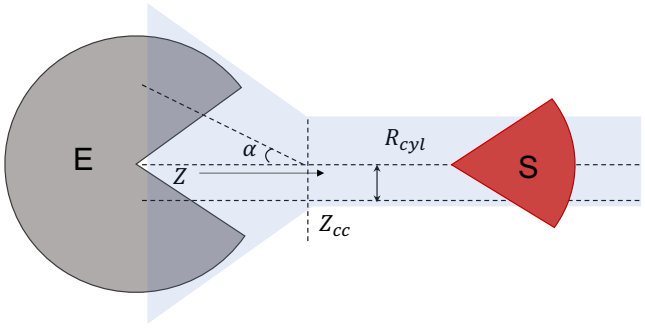


Figure C-S8. Scheme describing tunable parameters in a funnel restraint where E is the enzyme and S is the substrate. $\mathbf{Z_{cc}}$ is the distance at which the restraint switches from a cone into a cylindrical shape, **α** is the angle which describes the amplitude of the cone, $\mathbf{R_{cyl}}$ is the radius of the cylinder, and **Z** describes the axis along which the ligand binds to the protein.

Table C-S1. Free energies of cellobiose binding to WT, Y245G, and W212A E1 determined from experimental kinetic parameters ($\Delta G^0_{b_{exp}}$) and Funnel MetaD ($G^0_{b_{FM}}$).

Variant	$\mathbf{\Delta G^0_{b_{exp}}}$(kJ/mol)	$\mathbf{G^0_{b_{FM}}}$(kJ/mol)
WT	15.7	14.57 ± 2.46
Y245G	7.14	7.97 ± 1.82
W212A	5.37	5.30 ± 0.81

Appendix D

Supporting Information:
Chapter 7. Thermodynamic Basis for Stabilization of Helical Peptoids by Chiral Sidechains

Sarah Alamdari[1], Kaylyn Torkelson[1], Xiaoqian Wang[1], Chun-Long Chen[2,1], Andrew L. Ferguson[3], Jim Pfaendtner*[1,2]

[1]*Dept. of Chemical Engineering, University of Washington, Seattle, Washington 98195*
[2]*Physical Science Division, Pacific Northwest National Laboratory, Richmond, Washington 99354*
[3]*Pritzker School of Molecular Engineering, University of Chicago, Chicago, Illinois 6063*

Department of Chemical and Biological Engineering, University of Colorado, Boulder, CO 80309, USA

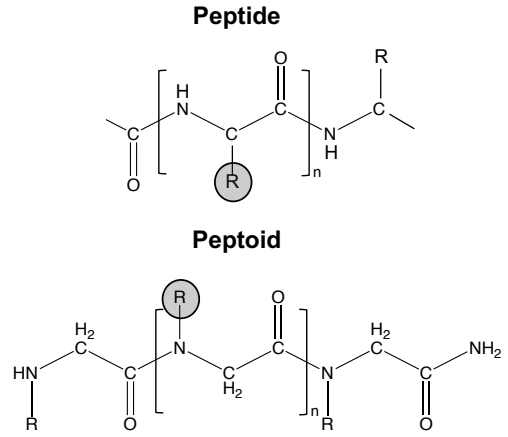


Figure D-S1. Comparison of peptide backbone to peptoid backbone

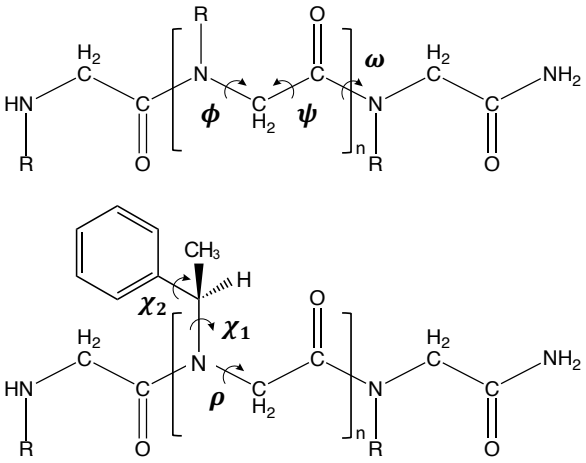


Figure D-S2. Representative polypeptoid backbone and termini used in this study, where R is side chain group. (Top) Three backbone dihedrals are depicted here; ω describes the rotation about the $C_{C-H_2} - C_{C=O} - N - C_{C-H_2}$ dihedral, ϕ describes the rotation about the $C_{C=O} - N - C_{C-H_2} - C_{C=O}$ dihedral and ψ describes rotation about the $N - C_{C-H_2} - C_{C=O} - N$ dihedral. (Bottom) Representative Nspe monomer to depict sidechain dihedrals; ρ describes the rotation about $C_{C=O} - C_{C-H_2} - N - C_\beta$, and $X_1$and X_2 describe the rotation about $C_{C-H_2} - N - C_\beta - C_\gamma$, and $N - C_\beta - C_\gamma - C_\delta$ for longer sidechains, where $C_\beta, C_\gamma,$ and C_δ are the first, second, and third carbon atoms from the sidechain backbone nitrogen.

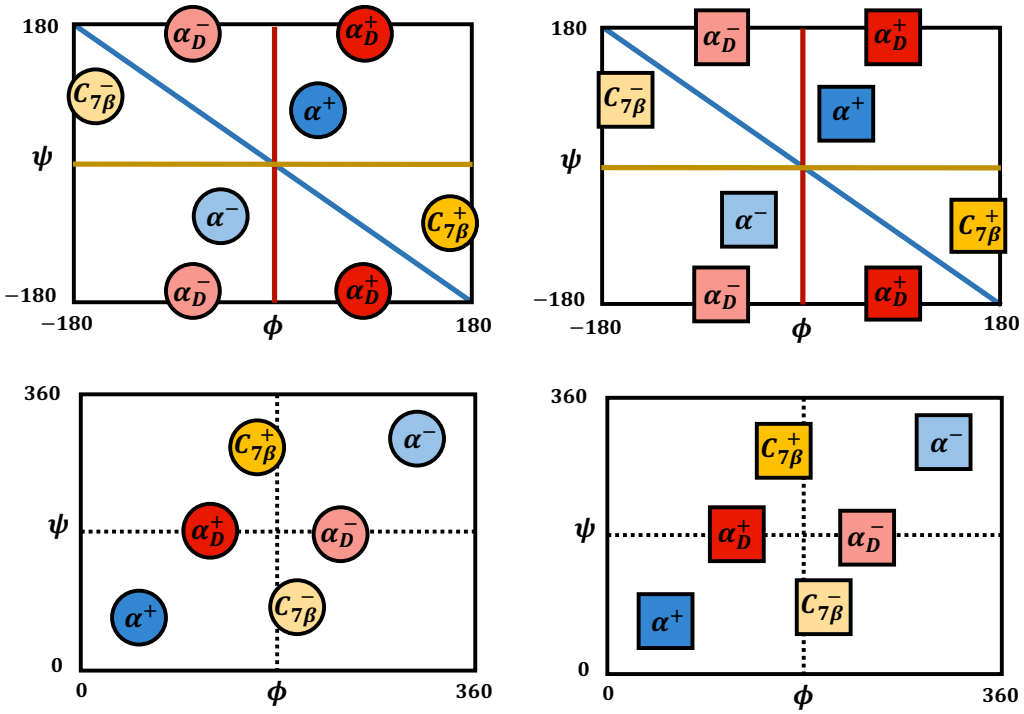


Figure D-S3. Reference Ramachandran plots with helical minimas depicted, plotted on a -180 to 180 degree scale (top row), and again on 0 to 360 scale (bottom row) for both the cis- (left column) and trans- (right column) conformation

Table D-S 1. Sequence-averaged backbone dihedrals for each peptoid chemistry and motif. The average column represents the average dihedral measurement of each of the 12 motifs for each of the three chemistries, where error is the standard deviation of these three values.

	Sarcosine			**Npm**			**Nspe**			**Averaged Motif Backbone**		
	ϕ (°)	**ψ** (°)	**ω**(°)	**ϕ** (°)	**ψ** (°)	**ω**(°)	**ϕ** (°)	**ψ** (°)	**ω**(°)	**ϕ** (°)	**ψ** (°)	**ω**(°)
$\alpha_{D^-\,CIS}$	-82	179	-1	-84	172	-8	-80	173	-1	**-82 ± 2**	**175 ± 4**	**0 ± 2**
$\alpha_{D^+\,CIS}$	81	-176	0	85	-174	-2	80	-180	2	**82 ± 3**	**-176 ± 3**	**0 ± 2**
α^-_{CIS}	-60	-49	-11	-63	-56	-11	-59	-44	-11	**-61 ± 2**	**-50 ± 6**	**-11 ± 0**
α^+_{CIS}	60	51	10	66	56	11	62	41	11	**63 ± 3**	**49 ± 8**	**11± 1**
$C_{7B^-\,CIS}$	-130	74	-1	-146	74	-6	-130	75	-1	**-135 ± 9**	**74 ± 1**	**-3 ± 3**
$C_{7B^+\,CIS}$	128	-72	1	127	-70	6	129	-72	0	**128 ± 1**	**-71 ± 1**	**2 ± 3**
$\alpha_{D^-\,TRANS}$	-85	177	-179	-85	168	179	-86	176	180	**-82 ± 1**	**174 ± 5**	**180 ± 1**
$\alpha_{D^+\,TRANS}$	84	-176	179	84	179	-179	80	-174	180	**83 ± 2**	**-177 ± 4**	**-180 ± 1**
α^-_{TRANS}	-55	-56	175	-60	-67	176	-55	-58	177	**-57 ± 3**	**-60 ± 6**	**176 ± 1**
α^+_{TRANS}	55	60	-176	64	76	-176	58	55	-176	**59 ± 5**	**64 ± 11**	**-176 ± 0**
$C_{7B^-\,TRANS}$	-108	74	-178	-101	91	-176	-108	76	-177	**-82 ± 4**	**80 ± 9**	**-177 ± 1**
$C_{7B^+\,TRANS}$	110	-76	179	97	-91	176	107	-76	178	**105 ± 7**	**-185 ± 9**	**178 ± 2**

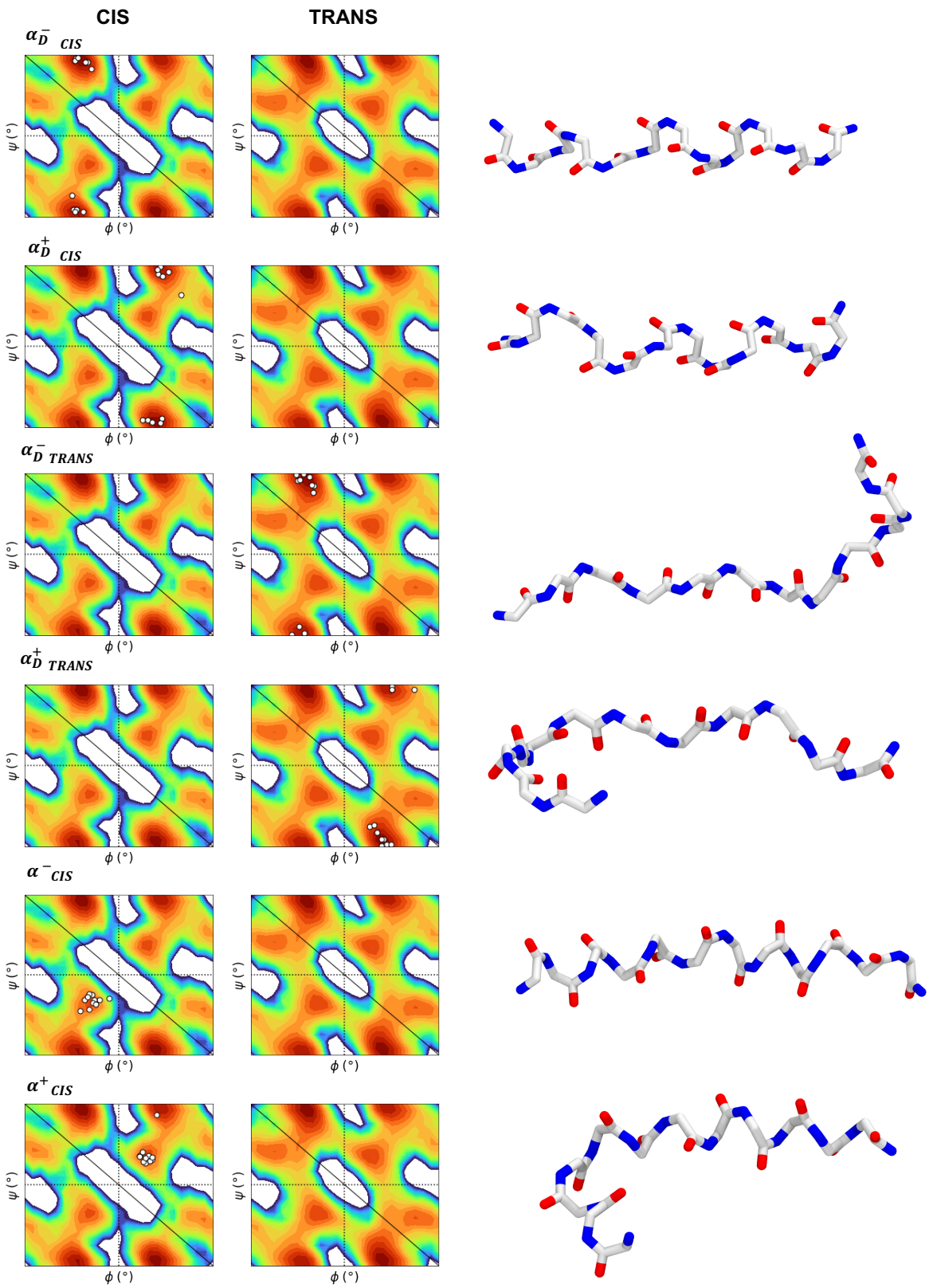


Figure D-S4. Full sequence folded structures obtained from sarcosine PBMetaD simulations, and dihedrals plotted on the converged Ramachandran plots used in **Figure 7.3**.

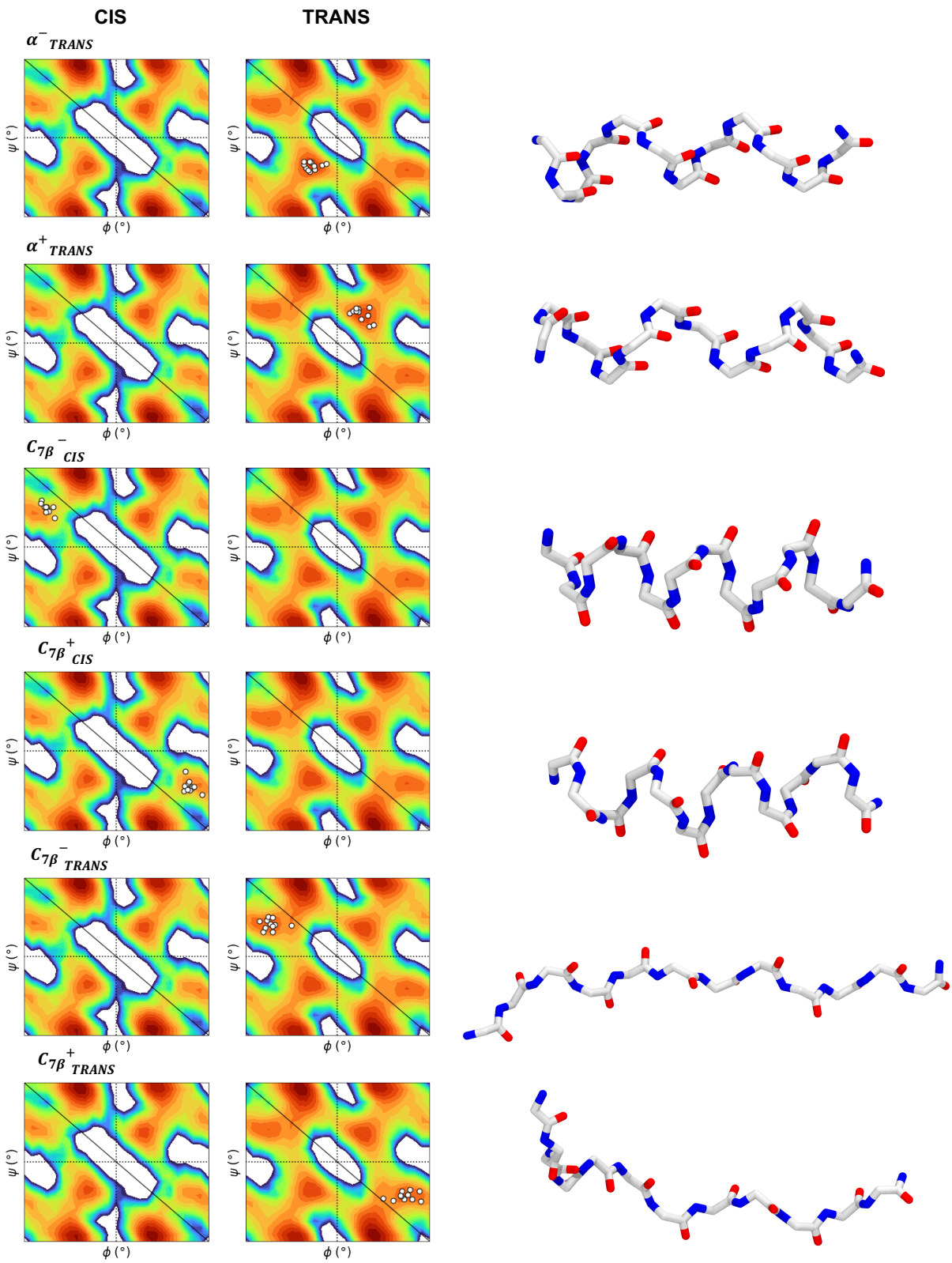


Figure D-S4. Full sequence folded structures obtained from sarcosine PBMetaD simulations, and dihedrals plotted on the converged Ramachandran plots used in **Figure 7.3**.

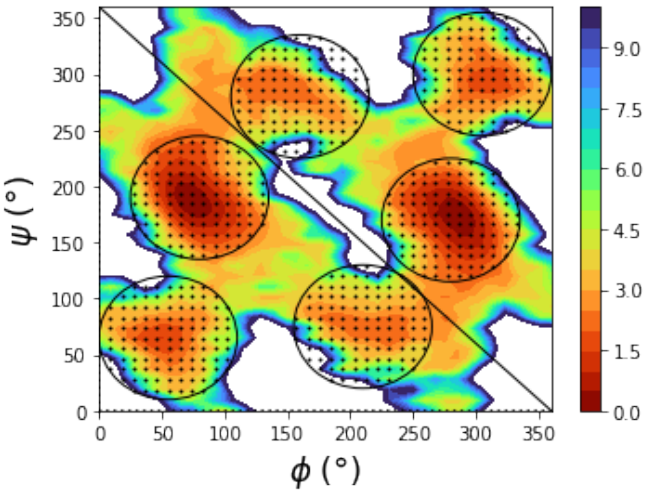
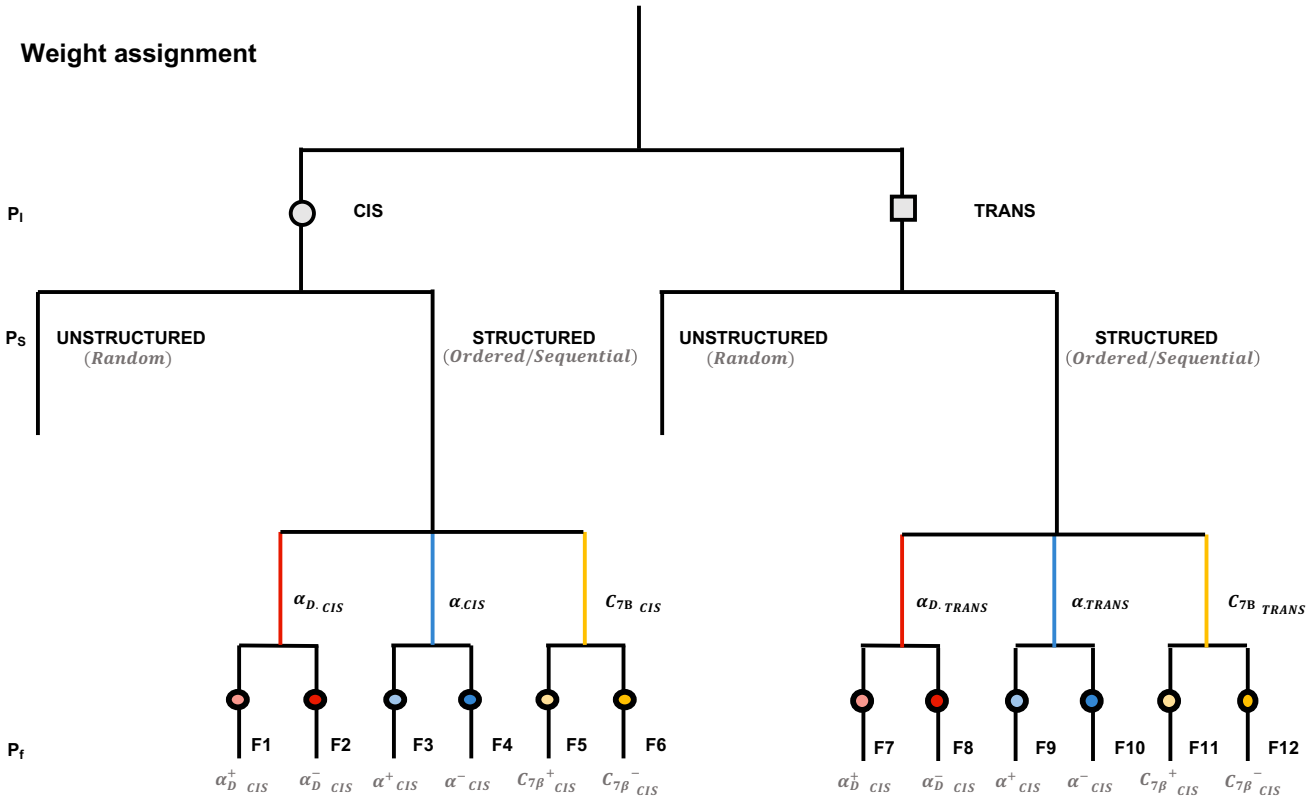


Figure D-S5. Example of basin definitions used for free energy analysis plotted on a 0 to 360 degree scale to remove boundary effects at the regions of interest. Values that lie within any of these basins were considered "structured" and values that lie outside were classified as unstructured. A radius of 55 degrees was used to define each basin. Unstructured regions on the FES accounted for less than 5% of the total weights from each simulation. Energy is plotted in kcal/mol.



Sample weights for Nspe 300K replica

Temperature	300								
CIS ($P_{I=cis}$)					TRANS ($P_{I=TRANS}$)				SUM (P_I)
0.68					0.32				1.00
Structured ($P_{s=structured}$)	0.979	f1	0.351		Structured ($P_{s=structured}$)	0.896	f7	0.588	
		f2	0.619				f8	0.334	
		f3	0.005				f9	0.009	
		f4	0.02				f10	0.031	
		f5	0.001				f11	0.008	
		f6	0.003				f12	0.029	
		SUM (P_f)	1.00				SUM (P_f)	1.00	
Unstructured ($P_{s=unstructured}$)	0.021				Unstructured ($P_{s=unstructured}$)	0.104			
SUM (P_s)	1.00				SUM (P_s)	1.00			

Figure D-S6. Assigning Probability Weights for Free Energy Analysis. Each free energy surface was assigned to 3 different categories 1) Isomerization (cis or trans); *cis*- regions are defined as $\omega \geq -90$ and $\omega \leq 90$ and trans- regions was all omega values outside of that range, 2) Structure (unstructured or structured), structured regions fell into one of the 12 basins described in **Figure S5** whereas unstructured regions fell outside of those regions, 3) Folded (one of 12 unique basins) as described in **Figure S5**. Probability of being in a folded state (**Equation 2**, main text) with respect to the entire folding landscape was calculated by propagating the probability of being in a single state along the tree diagram: $p_{f_i,norm} = p_{f=\{1..12\}} * p_{S=\{structured,unstructured\}} * p_{I=\{cis,trans\}}$

Table D-S2. Absolute energy difference between left-handed and right-handed energies in Table 1. Energy differences are the absolute value calculated by subtracting the RH (+) state from the LH (-) state and are reported in kcal/mol.

	Sarc$_{12}$ (kcal/mol)	**Npm$_{12}$** (kcal/mol)	**Nspe$_{12}$** (kcal/mol)
cis- $\boldsymbol{\alpha_D}$	0.06	0.13	0.5
cis- $\boldsymbol{\alpha}$	0.18	0.04	0.83
cis- $\boldsymbol{C_{7B}}$	0.09	0.25	0.65
trans- $\boldsymbol{\alpha_D}$	0.05	0.01	0.39
trans- $\boldsymbol{\alpha}$	0.02	0.09	0.74
trans- $\boldsymbol{C_{7B}}$	0.00	0.11	0.77

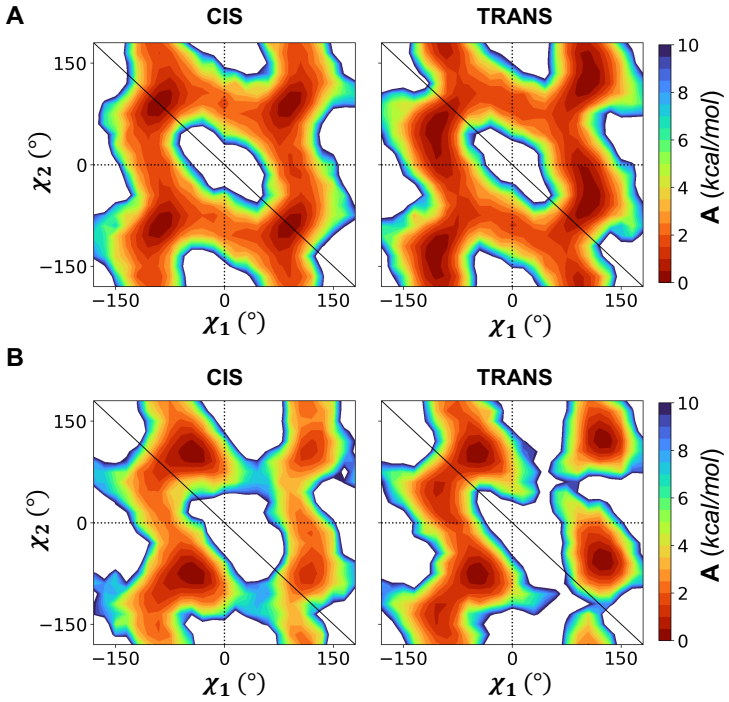


Figure D-S7. Flexibility along residue sidechains. Free energy surfaces projected onto the χ_1 and χ_2 sidechain dihedrals for (**A**) Npm$_{12}$ (**B**) Nspe$_{12}$.

Table D-S3. Entropic fitting parameters for isomerization and folding reported in kcal/mol

	$CIS\ isomerizaiton$			$\alpha_{D\ CIS}^{-}(RH)$		
	$-\Delta S\ (m)$ *(kcal/mol K)*	$\Delta U\ (b)$ *(kcal/mol)*	R^2	$-\Delta S\ (m)$ *(kcal/mol K)*	$\Delta U\ (b)$ *(kcal/mol)*	R^2
Sarcosine	-0.00029	0.93749	0.500	0.00325	0.35502	0.974
Npm	0.00031	-0.21849	0.835	0.00361	-0.51778	0.999
Nspe	-0.00100	-0.15806	0.961	0.00239	-0.54046	0.936

Table D-S4. Entropic contributions to *cis*- isomerization at different temperatures reported in kcal/mol

Temperature (K)	**300**	**317**	**336**	**355**	**376**	**398**	**421**	**446**	**472**	**500**
Sarcosine	-0.086	-0.091	-0.096	-0.102	-0.108	-0.114	-0.121	-0.128	-0.136	-0.144
Npm	0.093	0.099	0.104	0.110	0.117	0.124	0.131	0.139	0.147	0.156
Nspe	-0.301	-0.319	-0.338	-0.357	-0.378	-0.400	-0.423	-0.448	-0.474	-0.502

Table D-S5. Entropic contributions to $\alpha_{D\ CIS}^{-}$ helix formation reported in kcal/mol

Temperature (K)	**300**	**317**	**336**	**355**	**376**	**398**	**421**	**446**	**472**	**500**
Sarcosine	0.976	1.031	1.093	1.155	1.223	1.295	1.370	1.451	1.536	1.627
Npm	1.084	1.145	1.214	1.282	1.358	1.438	1.521	1.611	1.705	1.806
Nspe	0.718	0.758	0.804	0.849	0.900	0.952	1.007	1.067	1.129	1.196

SI Methods

Forcefield Details and Equilibration

All systems were simulated in GROMACS 2018.3[72] and analyzed using the Plumed 2.4.2 plugin[73]. A modified CHARMM-based forcefield was implemented in this work. Backbone partial charges for the amide nitrogen atom, the backbone carbon were taken from the original MFTOID forcefield[293] were used. Lennard-Jones parameters for the carbonyl carbon from the original MFTOID forcefield[293] were also used. Backbone dihedrals (omega, and chi) and tuned-CGENFF parameters for the sidechains as published by Weiser and Santiso[484] were used. A combination of backbone and sidechain parameters from these CHARMM-based forcefields was chosen in order to retain a uniform backbone partial charges for the sequences, similar to how CHARMM-based forcefields are implemented for proteins and peptides. All systems were simulated in a TIP3P water box. GROMACS formatted forcefield rtp files are available on github (https://github.com/UWPRG/peptoid-ffgen/tree/master/docs). To justify combining these parameter sets, we conducted a 500ns, well-tempered metadynamics simulation of a single sarcosine dipeptoid with an acetyl-capped N-terminus, and a dimethylamide-capped C-terminus, following what was used in previous forcefield fitting papers[293,484]. This molecule was simulated in a 3.3 nm x 3.0 nm x 3.1 nm water box, using 1000 water molecules, and equilibrated using the protocol described in the following paragraph. A 3D bias potential was added to the phi, psi, and omega collective variables using a hill height of 2.5 kJ/mol, a pace of 500 steps, and a bias factor of 15. In the original Mirijanian et al., work the α^+ and α^- states were most favorable for a cis backbone, which did not agree with DFT calculations. The FES (**Figure D-S8**) obtained from our metadynamics simulation agrees with the results published by Weiser and Santino, showing overall favorability of the α_D^+ and α_D^- states for both cis and trans backbone configurations. This result validates the choice of forcefield parameters used in this work, in the limit of current forcefields available. Convergence of these calculations are given in **Figure D-S9/D-S10**.

Three different dodecamer polymers were simulated in this work; sarcosine $(\text{sarcosine})_{12}$ which is unsubstituted at N–Cα, an achiral N-(1-phenylmethyl)glycine $(\text{Npme})_{12}$ sequence, and the chiral helix forming (S)-N-(1-phenylethyl)glycine, $(\text{N}s\text{pe})_{12}$ terminated with a free N-termini and an amidated C-termini as shown in **Figure D-S1**. All three systems were simulated in a 6.4 nm x 6.4 nm x 6.1 nm box, with approximately 7,600 water molecules. All systems were equilibrated using the following protocol. First systems were minimized using a steepest descent algorithm to

remove any unfavorable contacts, over 10,000 steps. Then, each system was equilibrated to 300K, over 1 ns, using a stochastic global thermostat[81] to couple temperature (tau=0.5 ps) followed by pressure equilibration to 1 bar, over 100 ps, using the Parrinello-Rahman barostat[82] (tau=5 ps) with compressibility set to 4.5×10^{-5} bar^{-1}. A timestep of 2 femtoseconds was used in all simulations, and hydrogen bonds were constrained using the LINCS[80] algorithm. Van der Waals interactions were calculated below a cutoff of 1.2 nm, and electrostatic interactions were calculated with particle-mesh Ewald[83] summations using a 1.2 nm cutoff.

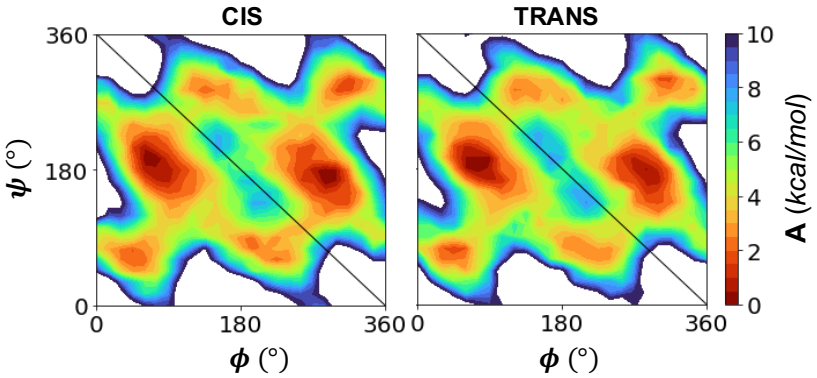


Figure D-S 8. FES obtained from metadynamics of sarcosine dipeptoid in water

Convergence of 3D Metadynamics Simulations of capped diethylamide

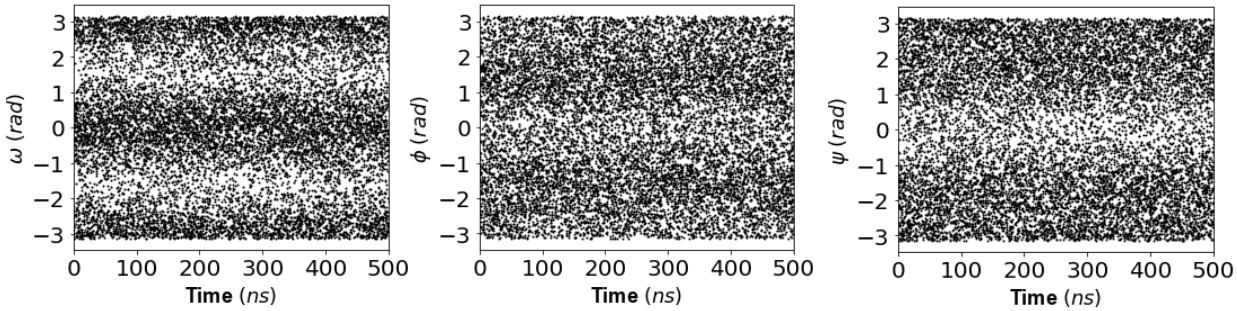


Figure D-S9. CV sampling for metadynamics of sarcosine dipeptoid in water

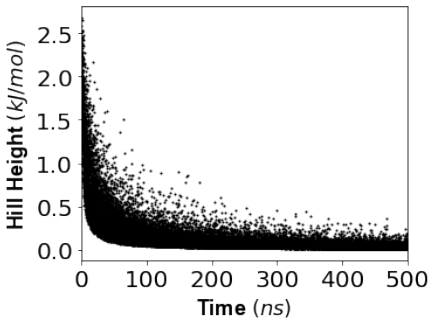


Figure D-S10. Hill height as a function of time for metadynamics of sarcosine dipeptoid in water.

Description of Collective Variables (CVs)

Coordination number (CN) was calculated using the functional form in equation 1, following the current implementation in PLUMED. This collective variable calculates the number of contacts between two groups of atoms (i and j). In this definition if a contact between two rings is formed the value is 1, otherwise it is zero, where s_{ij} is a smoothing function provided in equation 2. Coordination number was used in two cases; 1) to bias each individual ring coordination, and 2) bias the overall coordination of aromatic rings in a sequence. For case 1, within this definition each group is represented by the the center of mass of the first carbon and fourth carbon in each aromatic ring (as a proxy for representing the entire ring). This results in a coordination number for each aromatic ring pair possible in a 12mer (CN(ring0, ring1), CN(ring0, ring2)… CN(ring11,ring12), etc.) for a total of 65 unique CVs. Additionally, in PLUMED its possible to capture the overall coordination of a sequence by summing all the coordination between all pairs. In both cases r_0 is 0.5 nm, n is 8, and m is 12. A neighbor list cutoff of 0.75 nm was used and updated every 10 steps to help speedup these calculations. Parameters were chosen based off the average behavior between aromatic rings in classical simulation runs.

$$\sum_i \sum_j s_{ij} \qquad D.1$$

$$s_{ij} = \frac{1 - \left(\frac{r_{ij}}{r_0}\right)^n}{1 - \left(\frac{r_{ij}}{r_0}\right)^m} \qquad D.2$$

Alphabeta is a collective variable defined in PLUMED which can measure the distance between an instantaneous set of torsional values (t_i), and a set of reference values (t_i^{ref}), as

provided in Table D-S6. This collective variable provides a metric to evaluate how many dihedrals of a 12mer are folded into a specific reference. In this work, we applied this CV to describe the sum of all the backbone dihedrals (ϕ_i, ψ_i, ω_i) for an entire 12mer in reference to each of the 12 motifs. Each 12mer consisted of 11 phi, 11 psi and 12 omega dihedrals, thus each alphabeta CV contained i =34 reference dihedrals for each motif. A total of 12 unique alphabeta CVs were used to describe overall homogeneous folding into each of the 12 motifs. The reference dihedrals used in this work are provided in Table S3, and were originally derived by approximating values from the MP2/6-31G* potential energy surface published by Weiser and Santiso[484]. For practical implementation of this CV, we refer the reader to the PLUMED input files published in PLUMED-NEST (plumID:21.030).

$$s = \frac{1}{2}\sum_i \left[1 + \cos\left(t_i - t_i^{ref}\right)\right] \qquad D.3$$

Clustering

Clustering was done using the gromos[520] method developed by Daura et al. implemented in GROMACS. Gromos is a geometric clustering method that uses a root mean square deviation (RMSD) distance metric to group structures given a cutoff value. In this work, each frame of the trajectory was first separated into a cis or trans structure. If more than half the omega dihedrals were in a cis backbone conformation, the structure in that frame was considered a cis structure, otherwise the structure was considered a trans structure. Each set of structures were then separately clustered based on RMSD of the backbone atoms (C_{C-H_2}, $C_{C=O}$, N, O),using cutoff values ranging from 0.12nm to 0.18nm to asses robustness. The probability of each cluster was then reweighted, using Torrie Valleau reweighting as previously described, to correct for the effect of the bias on the cluster's probability. A cumulative distribution function (CDF) of these reweighted probabilities is plotted in **Figure D-S11** as a function of different gromos cutoff values, and different sidechain chemistries.

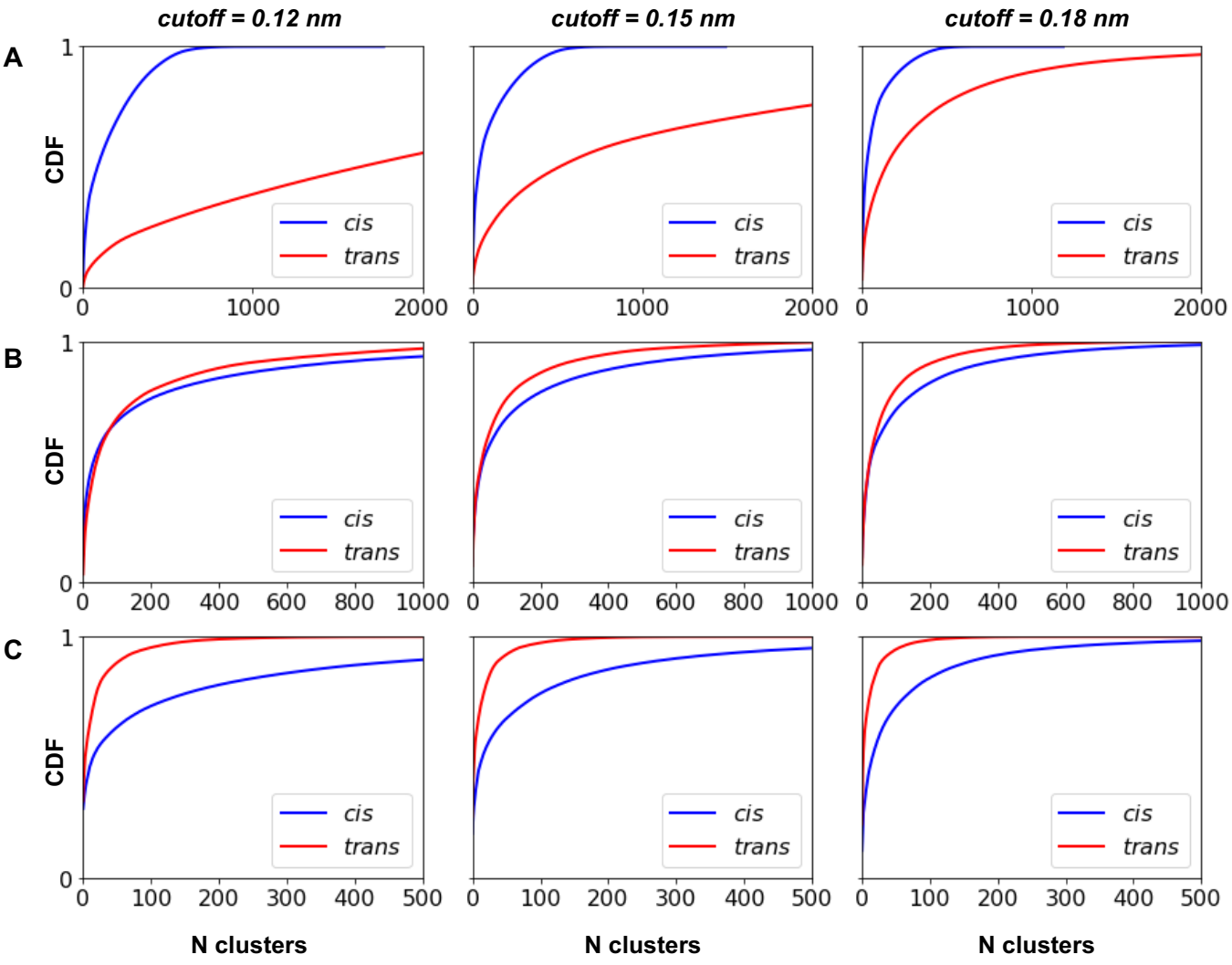


Figure D-S11. Cumulative distribution function (CDF) of reweighted probabilities corresponding to each cluster of structures given a mostly cis or mostly trans backbone for (row A) $sarc_{12}$, (row B) Npm_{12}, and (row C) $Nspe_{12}$ at 300k. The varying cutoffs used in the gromos clustering algorithm are given in each column.

Table D-S6. Reference dihedrals used for the alphabeta collective variable, given in radians

Motif	$\boldsymbol{\phi}$ (rad)	$\boldsymbol{\psi}$ (rad)	$\boldsymbol{\omega}(rad)$
$\alpha_{D^-}{}_{CIS}$	-1.57	3.14	0.00
$\alpha_{D^+}{}_{CIS}$	1.57	3.14	0.00
α^-_{CIS}	-0.87	-1.00	0.00
α^+_{CIS}	0.87	1.00	0.00
$C_{7B^-}{}_{CIS}$	-2.09	1.30	0.00
$C_{7B^+}{}_{CIS}$	2.09	1.30	0.00
$\alpha_{D^-}{}_{TRANS}$	-1.57	3.14	3.14
$\alpha_{D^+}{}_{TRANS}$	1.57	3.14	3.14
α^-_{TRANS}	-0.87	-1.00	3.14
α^+_{TRANS}	0.87	1.00	3.14
$C_{7B^-}{}_{TRANS}$	-2.09	1.30	3.14
$C_{7B^+}{}_{TRANS}$	2.09	1.30	3.14

PT-PBMetaD-WTE Convergence

Convergence was determined by evaluating the 12 Alphabeta CVs over the last 30% of simulation time . Convergence was determined by evaluating 1) fluctuations of this CV, and 2) no significant changes to the FEP minima along alphabeta at long timescales while the overall bias added to the system was low (**Figures D-S12** to **D-S18)**. The minimum and maximum values of alphabeta range from 0 (no dihedral triplets associated with the 12 residues in a sequence) to 34 (all 12 residues are in a local minimum state), providing a metric for evaluating structures traversing the entire free energy landscape. Traversing this CV implies that all the relevant dihedrals are sampled along each sequence. We note that although the reweighted free-energy profiles (cf. Figs S8, S10, S11) experience fluctuations in the small basin at high values of ALPHARMSD, these do not impact our overall conclusions as these states represent unlikely "perfect helices". As discussed in the manuscript, we verified that it is the significantly wider deep energetic minima for all 12 states that have relevance for our key conclusions needed to reproduce sequence averaged free energy surfaces consistent with previous experimental and DFT-level predictions.

To further validate results obtained from metadynamics, a series of classical MD simulations are supplied which supported relative stability of the states extracted from enhanced sampling simulations. Classical MD simulations support those structures obtained from low-energy regions predicted by metadynamics were stable. **Figure D-S19** depicts the time series RMSD for 3-1 μs trials done on sarc$_{12}$, Npm$_{12}$, and Nspe$_{12}$ from a fully folded $\alpha_{D_{CIS}^{-}}$ helix starting configuration. These results show that the Nspe$_{12}$ helix is stable on the order of 1 μs whereas Npm$_{12}$ and sarc$_{12}$ rapidly unfold, given by a large RMSD. This result supports findings reported in **Figure 7.5** in the main text.

Following this, a second validation test was conducted on the Nspe$_{12}$ by running a single 100ns MD simulation starting from structures corresponding to each of the 12 basins. The starting and ending configurations for these simulations are provided in **Figure D-S20**, and the timeseries RMSD and dihedral probability densities related to structure stability are provided in **Figures D-S21** and **D-S22.** As shown in **D-S21 and D-S22** few *trans*- to *cis*- isomerization events, and no *cis*- to *trans*- events were observed. On average *trans*- structures were much less stable *cis*- ones, consistent with isomerization predictions from metadynamics. These results further demonstrate that structures which begin in $\alpha_{D_{CIS}^{-}}$ helices are the most stable, given by no significant changes to the dihedral preferences and structure over 100 ns. The simulation starting from a $\alpha_{D_{CIS}^{+}}$ was slightly less stable than $\alpha_{D_{CIS}^{-}}$, resulting in a final state where 2 of the 12 residues unfolded into a different energetic minimum. Structures started from other higher energy basins demonstrated some additional heterogeneous unfolding into nearby energetic basins, as predicted by the free energy surface. Structures which start in $\alpha_{D_{TRANS}}^{-/+}$ are partially stable, showing initial unfolding into heterogeneous conformers. The $C_{7\beta_{CIS/TRANS}}^{-/+}$ and $\alpha_{CIS/TRANS}^{+/-}$ demonstrated the least stability on MD timescales, showing complete and rapid unfolding into heterogeneous combinations of nearby $\alpha_{D_{CIS/TRANS}}^{+/-}$ states.

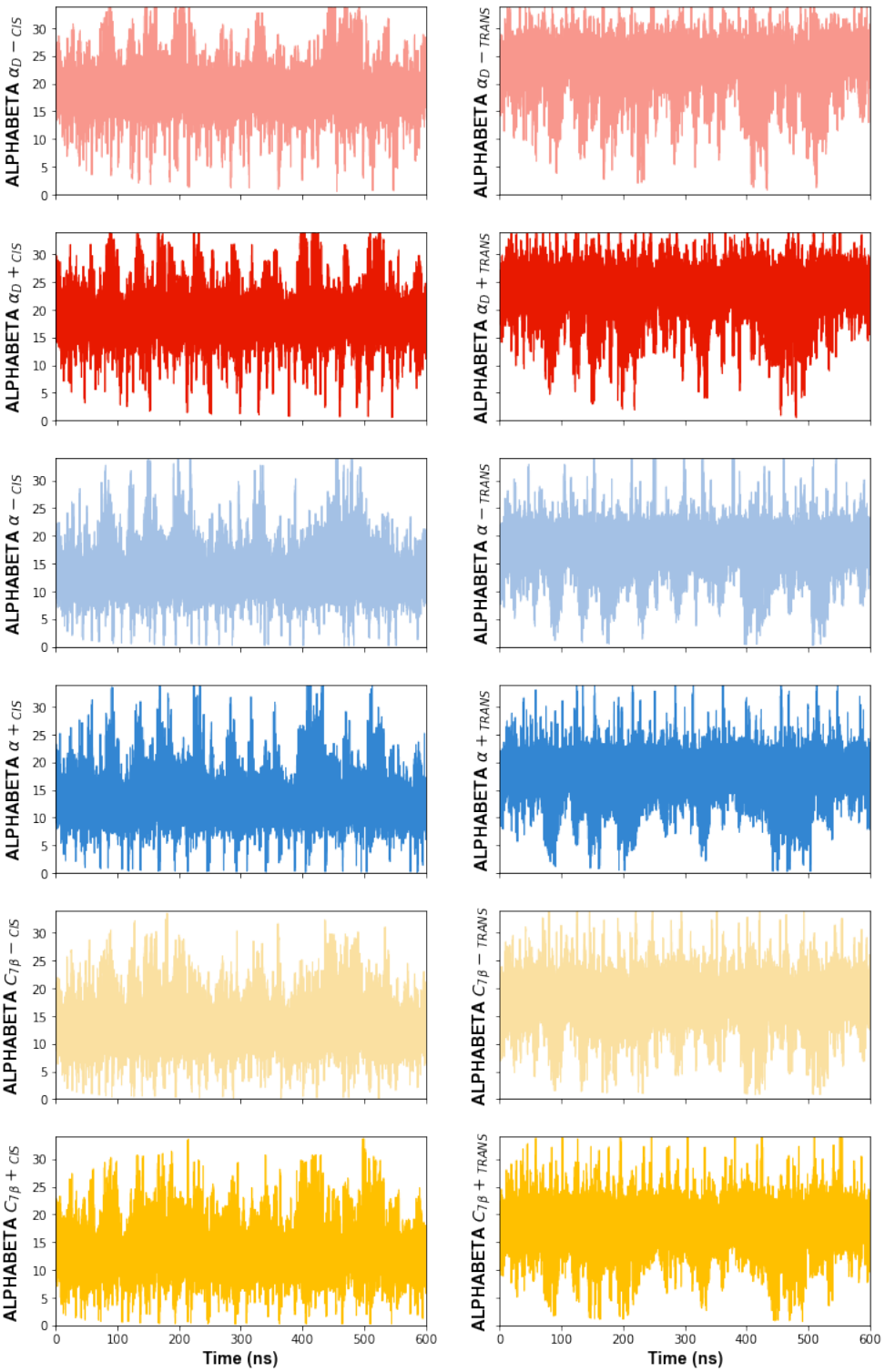


Figure D-S12. $Sarc_{12}$ sampling of the 300K replica monitoring alphabeta of each of the 12 states over the entire simulation time.

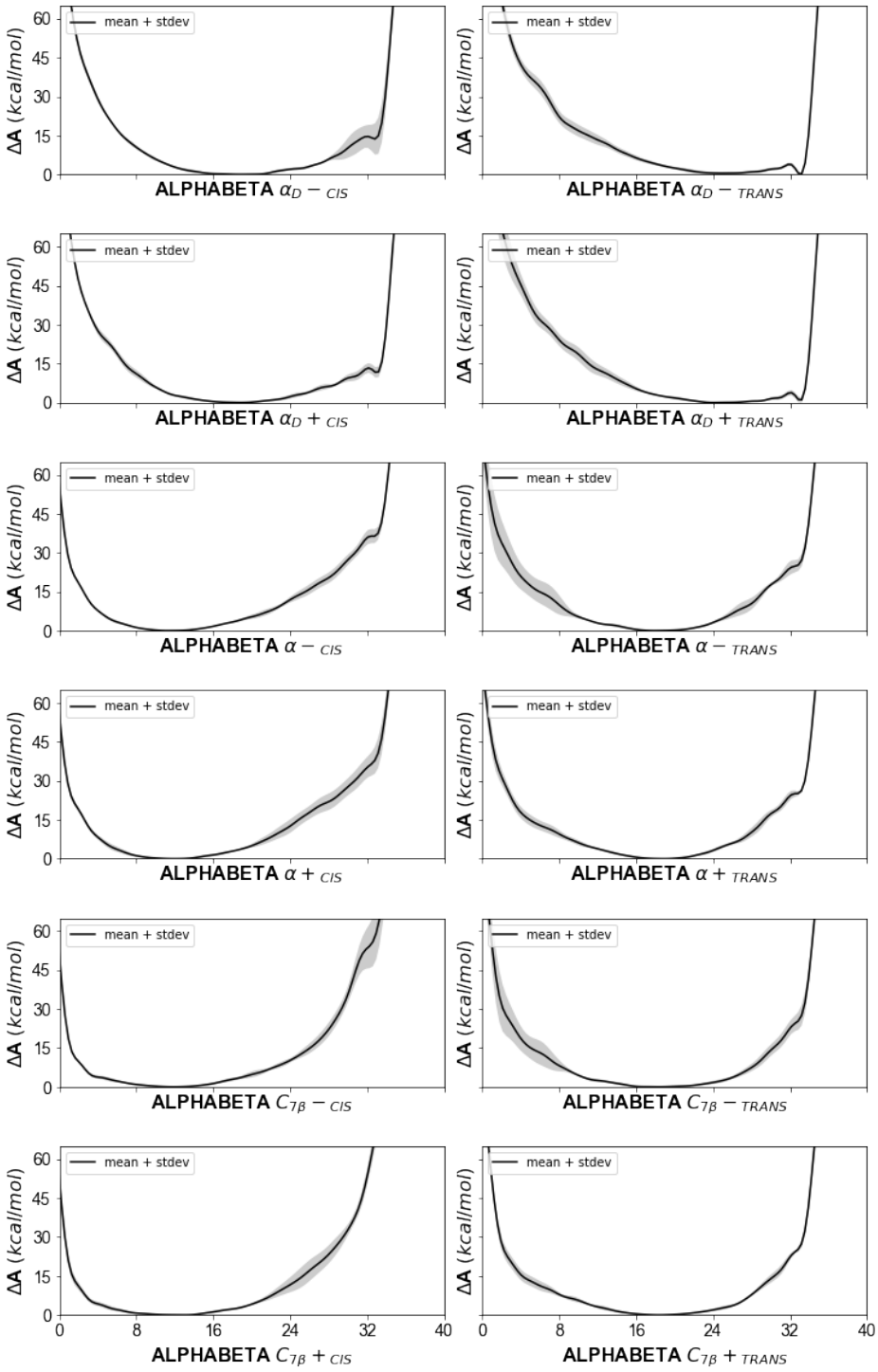


Figure D-S13. Average 1D Free Energy Profile (FEP) for $sarc_{12}$ 300K replica calculated using the last 30 % of simulation time, error bar is the standard deviation of energy fluctuations from the last 30% of simulation time

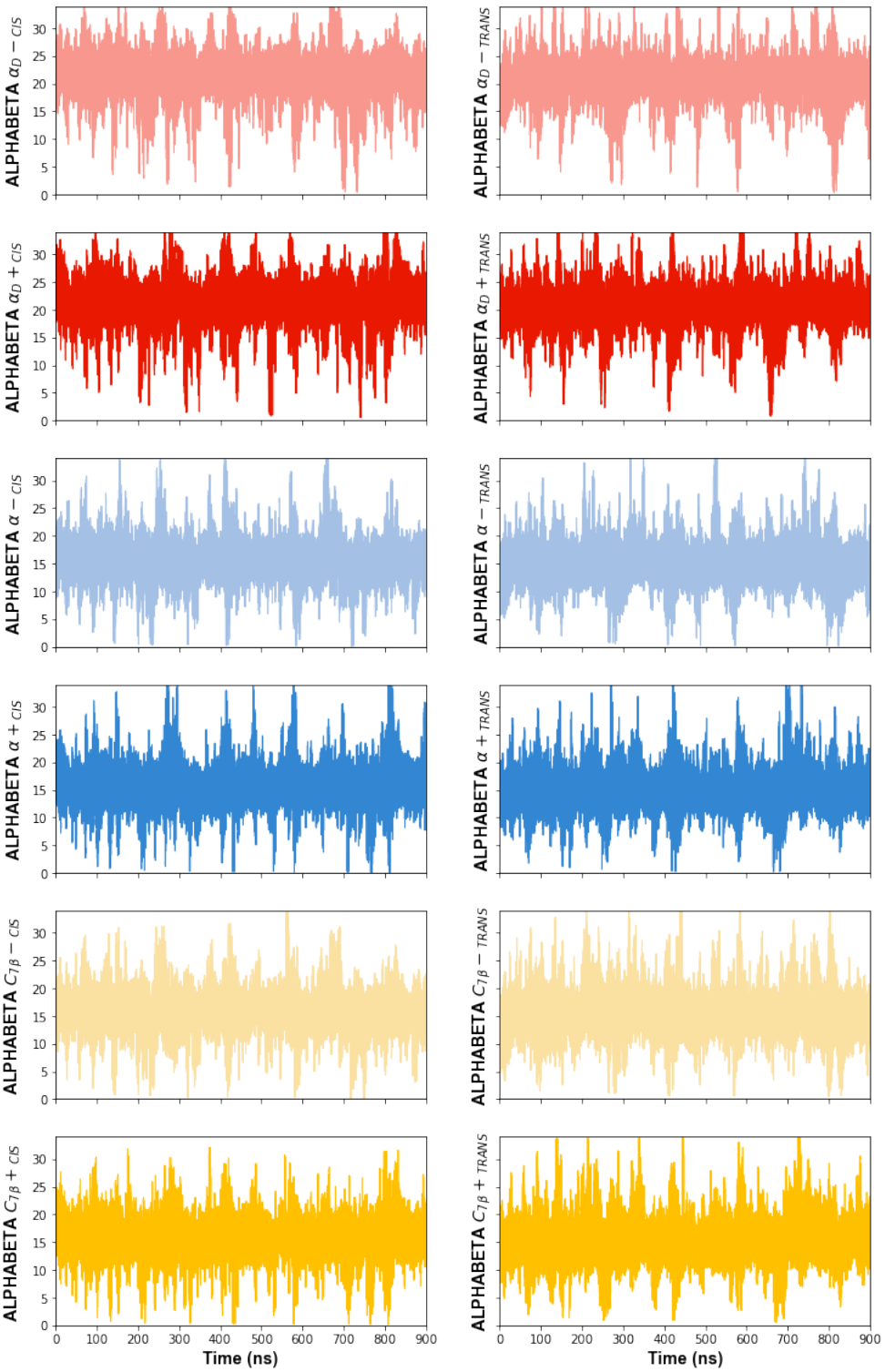


Figure D-S14. Npm_{12} sampling of the 300K replica monitoring alphabeta of each of the 12 states over the entire simulation time.

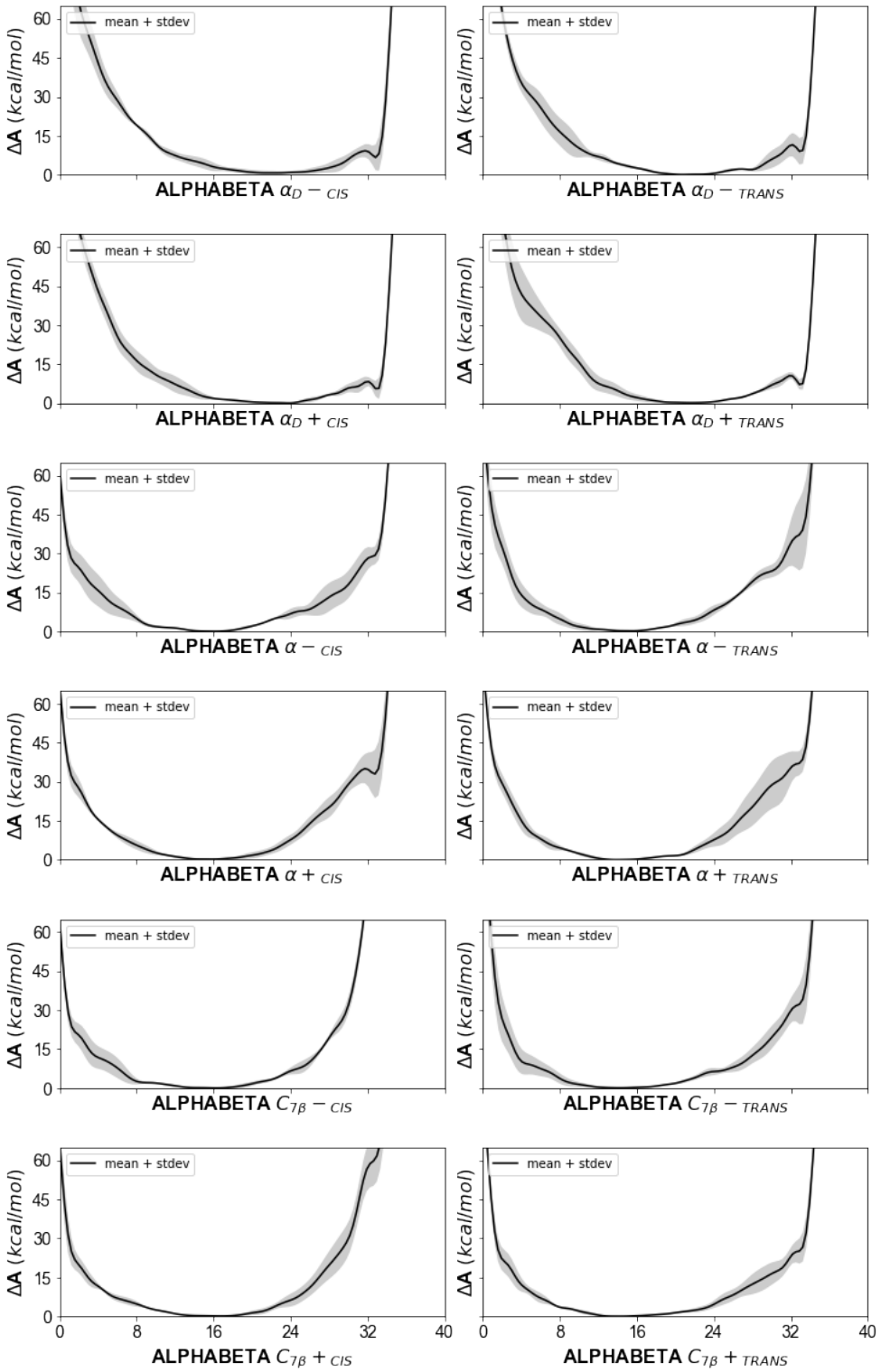


Figure D-S15. Average 1D Free Energy Profile (FEP) for Npm_{12} 300K replica calculated using the last 30 % of simulation time, error bar is the standard deviation of energy fluctuations from the last 30% of simulation time

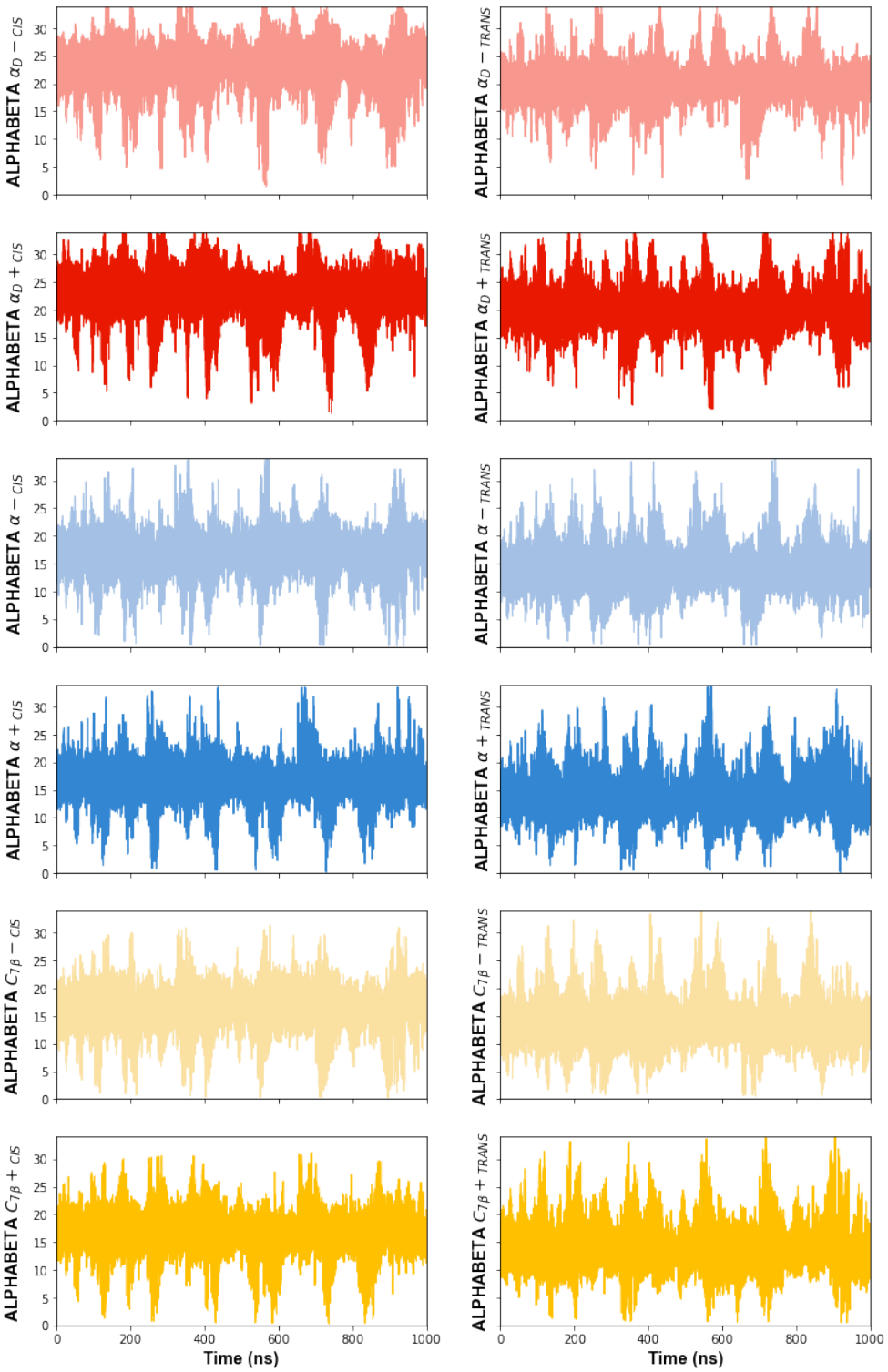


Figure D-S16. $Nspe_{12}$ sampling of the 300K replica monitoring alphabeta of each of the 12 states over the entire simulation time.

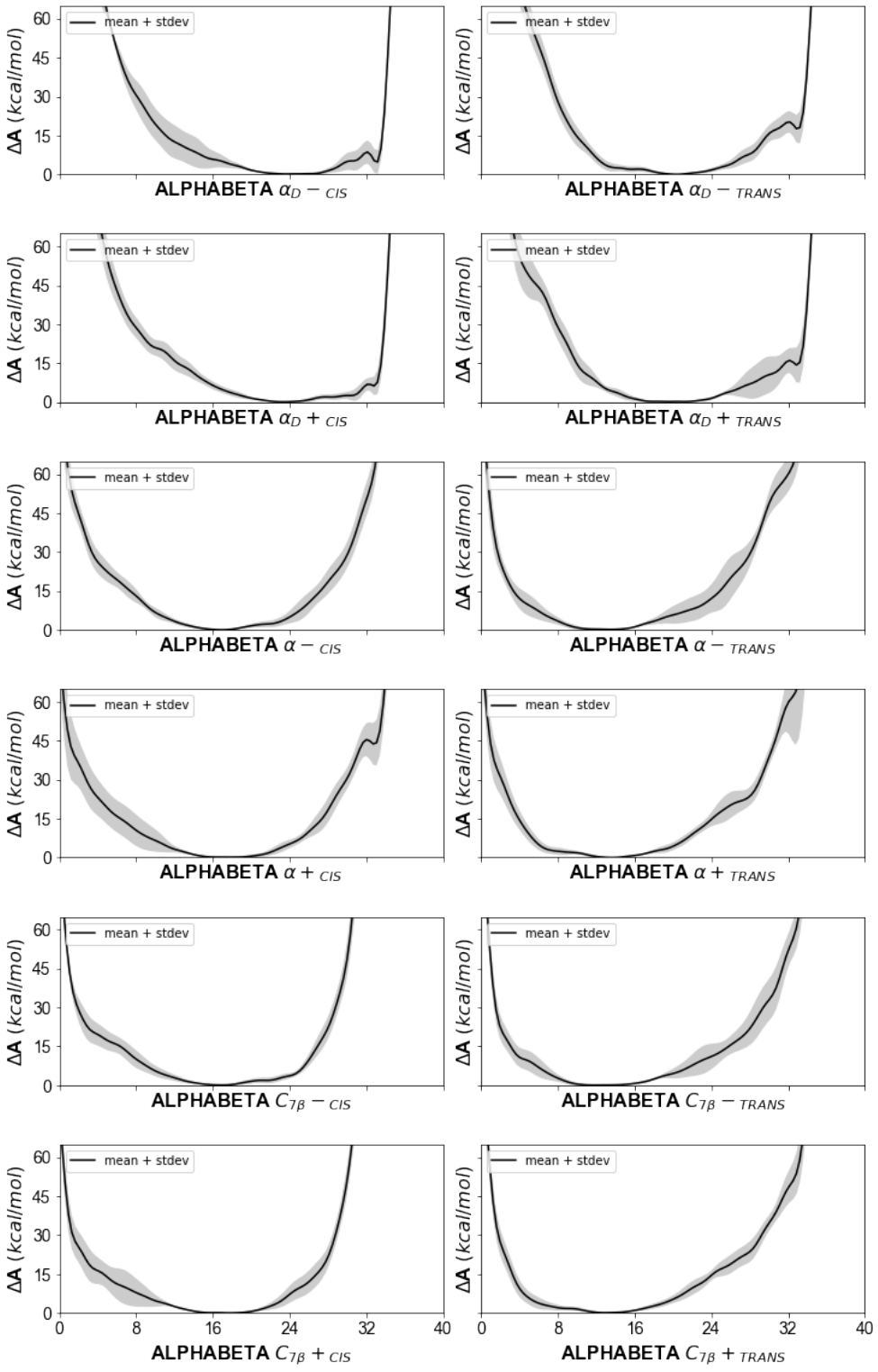


Figure D-S17. Average 1D Free Energy Profile (FEP) for $Nspe_{12}$ 300K replica calculated using the last 30 % of simulation time, error bar is the standard deviation of energy fluctuations from the last 30% of simulation time.

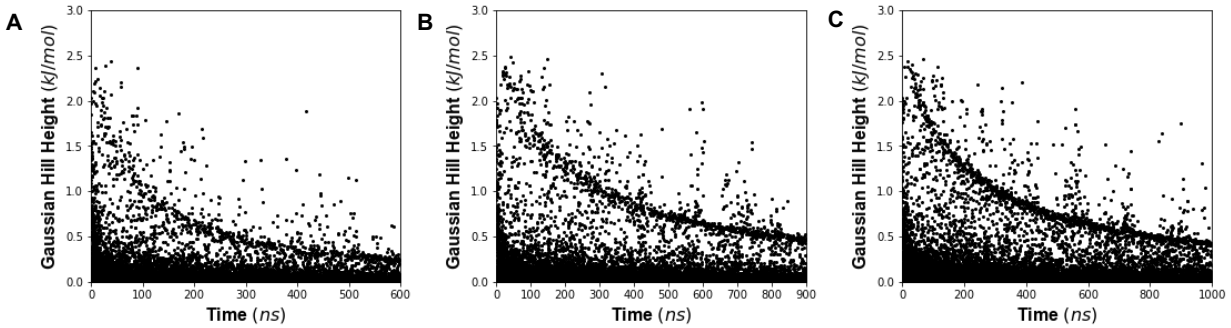


Figure D-S18. Gaussian hill height vs time for all 1D bias potentials added to system at 300K for (A) $sarc_{12}$, (B) Npm_{12}, and (C) $Nspe_{12}$ metadynamics simulations.

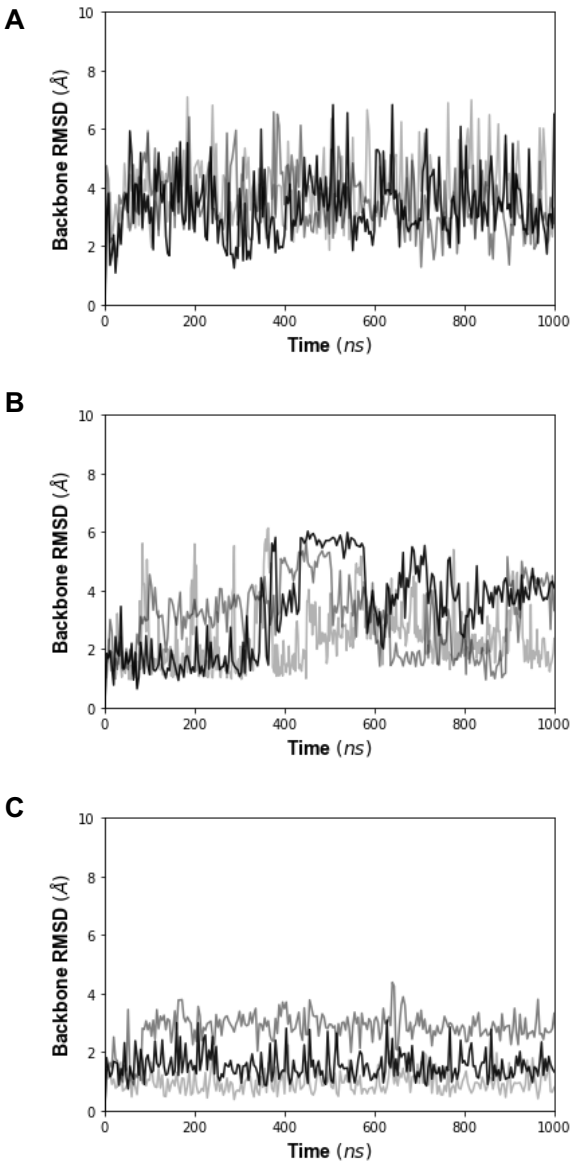


Figure D-S19. Three 1 μs trials of MD were conducted in water from a fully helical starting position. Backbone RMSD vs time was measured for A) $sarc_{12}$ B) Npm_{12} and C) $Nspe_{12}$

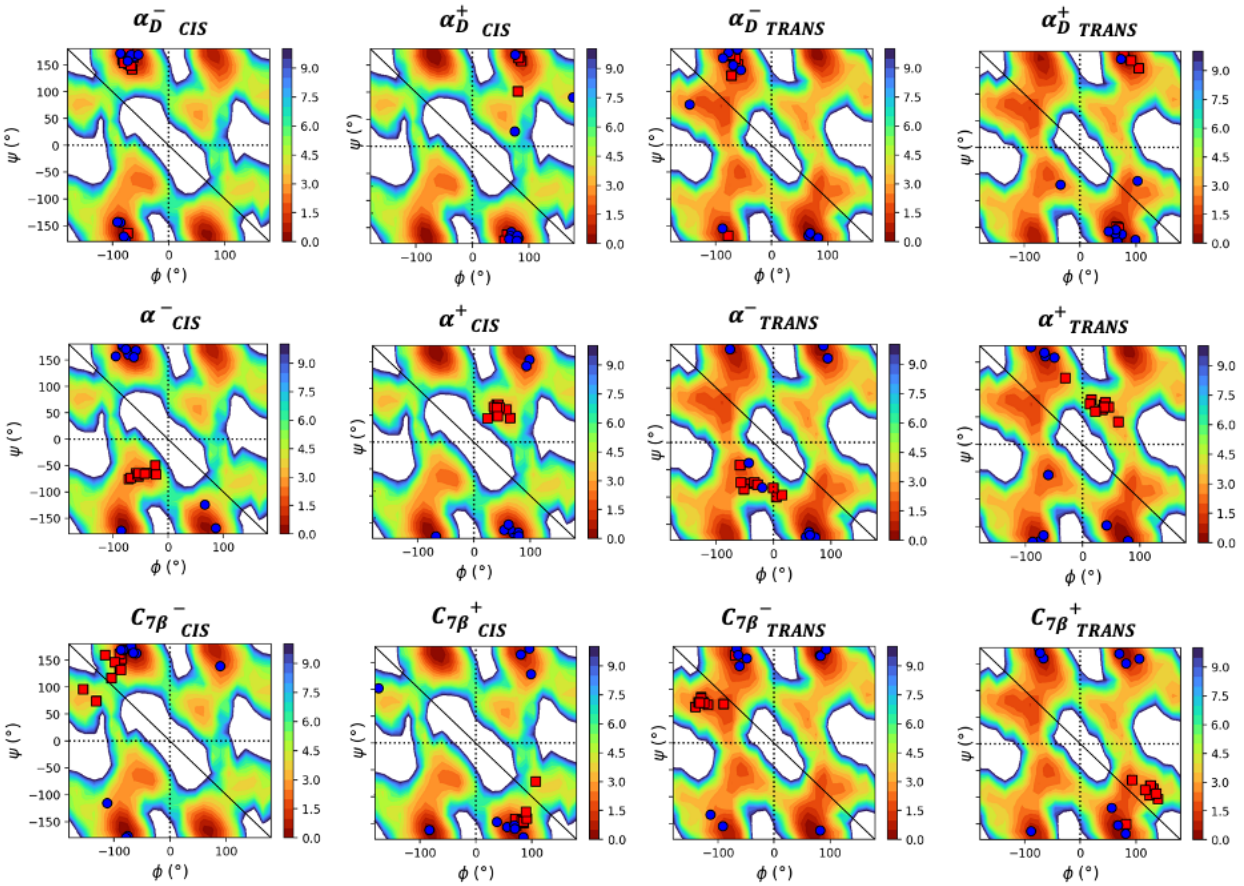


Figure D-S20. Conformational stability measured over 100ns of classical MD. 12 trials of Nspe12 were simulated at 300K with classical MD, initiated from each of the 12 minima. Dihedrals for each residue in the starting (red squares) and ending (blue circles) configurations are plotted on top of the Ramachandran plots obtained using PBMetaD. Each structure contains only either a cis or trans Ramachandran plot for simplicity, although isomerization events occurred over 100ns timescales. Time-series RMSD and probability density related to fluctuations of this system over 100ns timescales are given in **Figures D-S16** and **D-S17.**

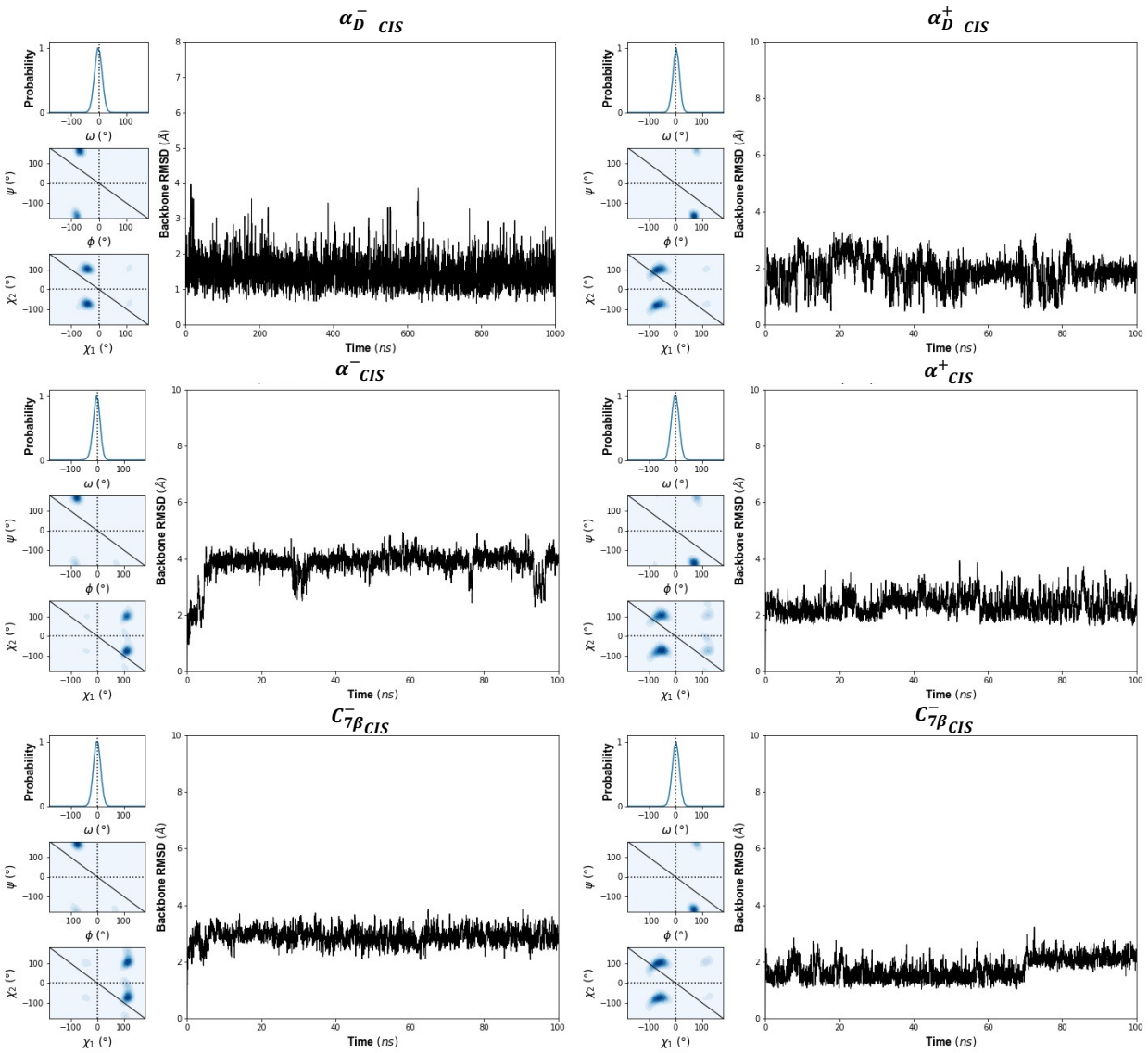


Figure D-S21. Stability of folded Nspe structures on MD timescales starting from *cis*- backbone configurations

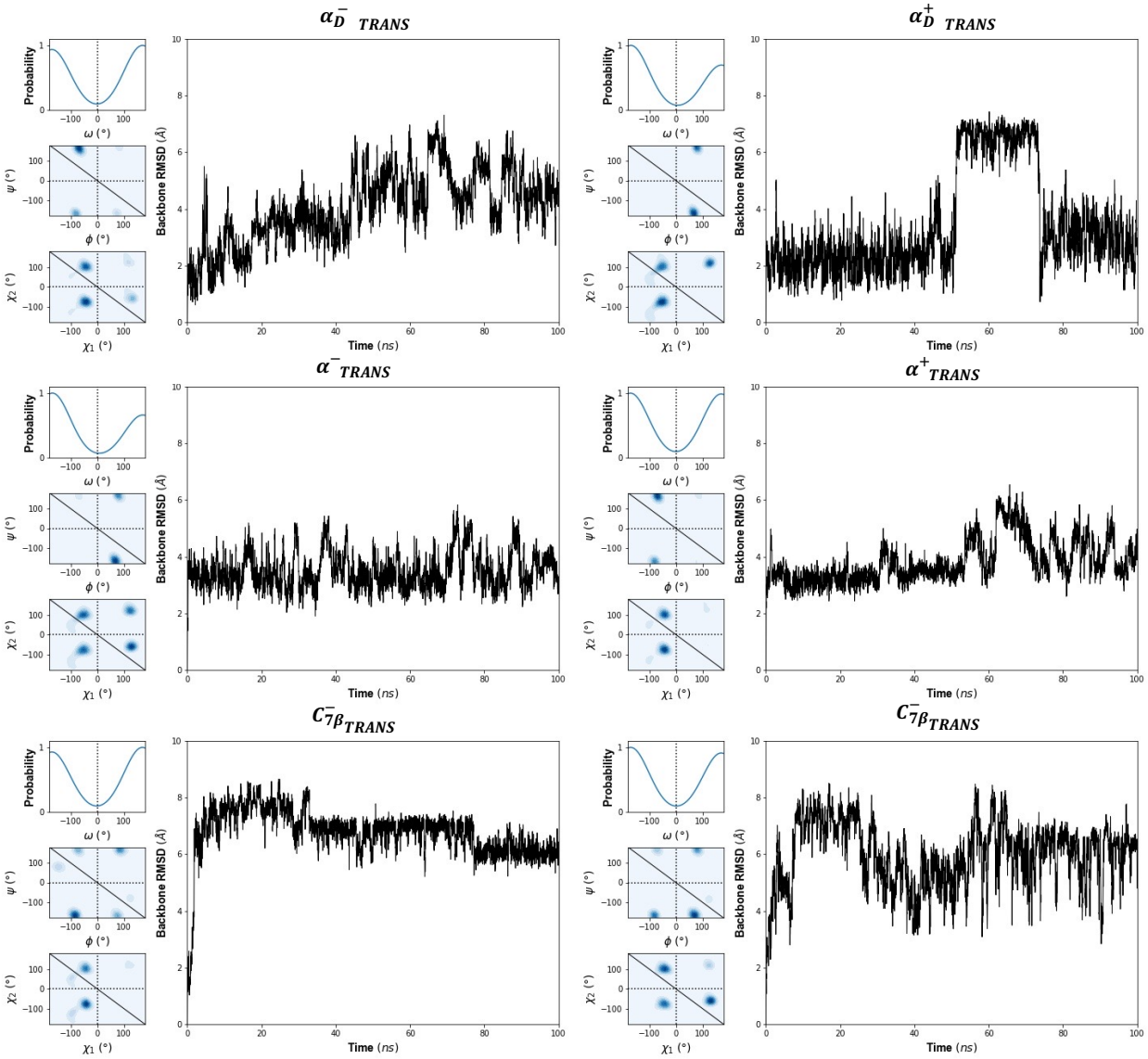


Figure D-S22. Stability of folded Nspe helices on MD timescales starting from *trans*- backbone configurations

References

(1) Arnold, F. H. Innovation by Evolution: Bringing New Chemistry to Life (Nobel Lecture). *Angew. Chemie Int. Ed.* **2019**. https://doi.org/10.1002/anie.201907729.

(2) Ha, N. S.; Lu, G. A Review of Recent Research on Bio-Inspired Structures and Materials for Energy Absorption Applications. *Compos. Part B Eng.* **2020**, *181*, 107496. https://doi.org/10.1016/j.compositesb.2019.107496.

(3) Midura-Nowaczek, K.; Markowska, A. Antimicrobial Peptides and Their Analogs: Searching for New Potential Therapeutics. *Perspect. Medicin. Chem.* **2014**, *6*, PMC.S13215. https://doi.org/10.4137/PMC.S13215.

(4) Gurung, N.; Ray, S.; Bose, S.; Rai, V. A Broader View: Microbial Enzymes and Their Relevance in Industries, Medicine, and Beyond. *Biomed Res. Int.* **2013**, *2013*, 1–18. https://doi.org/10.1155/2013/329121.

(5) Zhu, J.; Avakyan, N.; Kakkis, A.; Hoffnagle, A. M.; Han, K.; Li, Y.; Zhang, Z.; Choi, T. S.; Na, Y.; Yu, C.-J.; et al. Protein Assembly by Design. *Chem. Rev.* **2021**, *121* (22), 13701–13796. https://doi.org/10.1021/acs.chemrev.1c00308.

(6) Sampath, J.; Alamdari, S.; Pfaendtner, J. Closing the Gap Between Modeling and Experiments in the Self-Assembly of Biomolecules at Interfaces and in Solution. *Chem. Mater.* **2020**, *32* (19), 8043–8059. https://doi.org/10.1021/acs.chemmater.0c01891.

(7) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2.

(8) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. How Fast-Folding Proteins Fold. *Science (80-.).* **2011**, *334* (6055), 517–520. https://doi.org/10.1126/science.1208351.

(9) 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. https://doi.org/10.1146/annurev-physchem-040215-112229.

(10) Anselme, K. Osteoblast Adhesion on Biomaterials. *Biomaterials* **2000**, *21* (7), 667–681. https://doi.org/10.1016/S0142-9612(99)00242-2.

(11) Raphel, J.; Holodniy, M.; Goodman, S. B.; Heilshorn, S. C. Multifunctional Coatings to Simultaneously Promote Osseointegration and Prevent Infection of Orthopaedic Implants. *Biomaterials* **2016**, *84*, 301–314. https://doi.org/10.1016/j.biomaterials.2016.01.016.

(12) Sundfeldt, M.; Carlsson, L. V.; Johansson, C. B.; Thomsen, P.; Gretzer, C. Aseptic Loosening, Not Only a Question of Wear: A Review of Different Theories. *Acta Orthop.* **2006**, *77* (2), 177–197. https://doi.org/10.1080/17453670610045902.

(13) Puleo, D. A.; Nanci, A. Understanding and Controlling the Bone-Implant Interface. *Biomaterials* **1999**, *20* (23–24), 2311–2321. https://doi.org/10.1016/S0142-9612(99)00160-X.

(14) Mavrogenis, A. F.; Dimitiou, R.; Babis, G. C. Biology of Implant Osseointegration. *J. Musculoskelet. Neuronal Interact.* **2009**, *9* (2), 61–71.

(15) Anselme, K.; Bigerelle, M.; Noël, B.; Iost, A.; Hardouin, P. Effect of Grooved Titanium Substratum on Human Osteoblastic Cell Growth. *J. Biomed. Mater. Res.* **2002**, *60* (4), 529–540. https://doi.org/10.1002/jbm.10101.

(16) Marco, F.; Milena, F.; Gianluca, G.; Vittoria, O. Peri-Implant Osteogenesis in Health and Osteoporosis. *Micron* **2005**, *36* (7–8), 630–644.

https://doi.org/10.1016/j.micron.2005.07.008.
(17) Bovan, B. D.; Hummert, T. W.; Dean, D. D.; Schwartz, Z. Role of Material Surfaces in Regulating Bone and Cartilage Cell Response. *Biomaterials* **1996**, *17* (2), 137–146.
(18) Orapiriyakul, W.; Young, P. S.; Damiati, L.; Tsimbouri, P. M. Antibacterial Surface Modification of Titanium Implants in Orthopaedics. *J. Tissue Eng.* **2018**, *9*, 204173141878983. https://doi.org/10.1177/2041731418789838.
(19) Nair, M.; Elizabeth, E. Applications of Titania Nanotubes in Bone Biology. *J. Nanosci. Nanotechnol.* **2015**, *15* (2), 939–955. https://doi.org/10.1166/jnn.2015.9771.
(20) Rønold, H. J.; Lyngstadaas, S. P.; Ellingsen, J. E. A Study on the Effect of Dual Blasting with TiO 2 on Titanium Implant Surfaces on Functional Attachment in Bone . *J. Biomed. Mater. Res. Part A* **2003**, *67A* (2), 524–530. https://doi.org/10.1002/jbm.a.10580.
(21) Omar, O.; Karazisis, D.; Ballo, A.; Petronis, S.; Agheli, H.; Emanuelsson, L.; Thomsen, P. The Role of Well-Defined Nanotopography of Titanium Implants on Osseointegration: Cellular And Molecular Events in Vivo. *Int. J. Nanomedicine* **2016**, 1367. https://doi.org/10.2147/ijn.s101294.
(22) Goriainov, V.; Hulsart-Billstrom, G.; Sjostrom, T.; Dunlop, D. G.; Su, B.; Oreffo, R. O. C. Harnessing Nanotopography to Enhance Osseointegration of Clinical Orthopedic Titanium Implants—An in Vitro and in Vivo Analysis. *Front. Bioeng. Biotechnol.* **2018**, *6* (April), 1–14. https://doi.org/10.3389/fbioe.2018.00044.
(23) Salou, L.; Hoornaert, A.; Louarn, G.; Layrolle, P. Enhanced Osseointegration of Titanium Implants with Nanostructured Surfaces: An Experimental Study in Rabbits. *Acta Biomater.* **2015**, *11* (1), 494–502. https://doi.org/10.1016/j.actbio.2014.10.017.
(24) van Hove, R. P.; Sierevelt, I. N.; van Royen, B. J.; Nolte, P. A. Titanium-Nitride Coating of Orthopaedic Implants: A Review of the Literature. *Biomed Res. Int.* **2015**, *2015*, 1–9. https://doi.org/10.1155/2015/485975.
(25) Tao, Z. S.; Zhou, W. S.; He, X. W.; Liu, W.; Bai, B. L.; Zhou, Q.; Huang, Z. L.; Tu, K. K.; Li, H.; Sun, T.; et al. A Comparative Study of Zinc, Magnesium, Strontium-Incorporated Hydroxyapatite-Coated Titanium Implants for Osseointegration of Osteopenic Rats. *Mater. Sci. Eng. C* **2016**, *62*, 226–232. https://doi.org/10.1016/j.msec.2016.01.034.
(26) Baranowski, A.; Klein, A.; Ritz, U.; Ackermann, A.; Anthonissen, J.; Kaufmann, K. B.; Brendel, C.; Götz, H.; Rommens, P. M.; Hofmann, A. Surface Functionalization of Orthopedic Titanium Implants with Bone Sialoprotein. *PLoS One* **2016**, *11* (4), 1–23. https://doi.org/10.1371/journal.pone.0153978.
(27) Damiati, L.; Eales, M. G.; Nobbs, A. H.; Su, B.; Tsimbouri, P. M.; Salmeron-Sanchez, M.; Dalby, M. J. Impact of Surface Topography and Coating on Osteogenesis and Bacterial Attachment on Titanium Implants. *J. Tissue Eng.* **2018**, *9*, 204173141879069. https://doi.org/10.1177/2041731418790694.
(28) Jemat, A.; Ghazali, M. J.; Razali, M.; Otsuka, Y. Surface Modifications and Their Effects on Titanium Dental Implants. *Biomed Res. Int.* **2015**, *2015*, 1–11. https://doi.org/10.1155/2015/791725.
(29) Ducy, P.; Desbois, C.; Boyce, B.; Pinero, G.; Story, B.; Dunstan, C.; Smith, E.; Bonadio, J.; Goldstein, S.; Gundberg, C.; et al. Increased Bone Formation in Osteocalcin-Deficient Mice. *Nature* **1996**, *382* (6590), 448–452. https://doi.org/10.1038/382448a0.
(30) Hauschka, P. V.; Reid, M. L. Timed Appearance of a Calcium-Binding Protein Containing γ-Carboxyglutamic Acid in Developing Chick Bone. *Dev. Biol.* **1978**, *65* (2), 426–434. https://doi.org/10.1016/0012-1606(78)90038-6.

(31) Pollard, B. J.; Kitchen, G. *Bones and Joints*, Sixth Edit.; Elsevier Inc., 2017. https://doi.org/10.1201/9781315164533.
(32) Robey, P. G. Vertebrate Mineralized Matrix Proteins: Structure and Function. *Connect. Tissue Res.* **1996**, *35* (1–4), 131–136. https://doi.org/10.3109/03008209609029183.
(33) Ellingsen, J. E. A Study on the Mechanism of Protein Adsorption to TiO2. *Biomaterials* **1991**, *12* (6), 593–596. https://doi.org/10.1016/0142-9612(91)90057-H.
(34) Palmer, L. C.; Newcomb, C. J.; Kaltz, S. R.; Spoerke, E. D.; Stupp, S. I. Biomimetic Systems for Hydroxyapatite Mineralization Inspired By Bone and Enamel. *Chem. Rev.* **2008**, *108* (11), 4754–4783. https://doi.org/10.1021/cr8004422.
(35) Malashkevich, V. N.; Almo, S. C.; Dowd, T. L. X-Ray Crystal Structure of Bovine 3 Glu-Osteocalcin. *Biochemistry* **2013**, *52* (47), 8387–8392. https://doi.org/10.1021/bi4010254.
(36) Hoang, Q. Q.; Sicheri, F.; Howard, A. J.; Yang, D. S. C. Bone Recognition Mechanism of Porcine Osteocalcin from Crystal Structure. *Nature* **2003**, *425* (6961), 977–980. https://doi.org/10.1038/nature02079.
(37) Atkinson, R. A.; Evans, J. S.; Hauschka, P. V.; Levine, B. A.; Meats, R.; Triffitt, J. T.; Virdi, A. S.; Williams, R. J. P. Conformational Studies of Osteocalcin in Solution. *Eur. J. Biochem.* **1995**, *232* (2), 515–521. https://doi.org/10.1111/j.1432-1033.1995.515zz.x.
(38) Dowd, T. L.; Rosen, J. F.; Mints, L.; Gundberg, C. M. The Effect of Pb2+ on the Structure and Hydroxyapatite Binding Properties of Osteocalcin. *Biochim. Biophys. Acta - Mol. Basis Dis.* **2001**, *1535* (2), 153–163. https://doi.org/10.1016/S0925-4439(00)00094-6.
(39) Andrea, C.; Maset, F.; De Toni, L.; Guidolin, D.; Sabbadin, D.; Strapazzon, G.; Moro, S.; De Filippis, V.; Foresta, C. Carboxylation-Dependent Conformational Changes of Human Osteocalcin. *Front. Biosci.* **2014**, *19* (7), 1105. https://doi.org/10.2741/4270.
(40) Stayton, P. S.; Drobny, G. P.; Shaw, W. J.; Long, J. R.; Gilbert, M. Molecular Recognition at the Protein-Hydroxyapatite Interface. *Crit. Rev. Oral Biol. Med.* **2003**, *14* (5), 370–376. https://doi.org/10.1177/154411130301400507.
(41) Shen, J.-W.; Wu, T.; Wang, Q.; Pan, H.-H. Molecular Simulation of Protein Adsorption and Desorption on Hydroxyapatite Surfaces. *Biomaterials* **2008**, *29* (5), 513–532. https://doi.org/10.1016/j.biomaterials.2007.10.016.
(42) 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. https://doi.org/10.1021/bi010864c.
(43) Hunter, G. K.; O'Young, J.; Grohe, B.; Karttunen, M.; Goldberg, H. A. The Flexible Polyelectrolyte Hypothesis of Protein−Biomineral Interaction. *Langmuir* **2010**, *26* (24), 18639–18646. https://doi.org/10.1021/la100401r.
(44) Scudeller, L. A.; Srinivasan, S.; Rossi, A. M.; Stayton, P. S.; Drobny, G. P.; Castner, D. G. Orientation and Conformation of Osteocalcin Adsorbed onto Calcium Phosphate and Silica Surfaces. *Biointerphases* **2017**, *12* (2), 02D411. https://doi.org/10.1116/1.4983407.
(45) Meretoja, V. V.; Ruijter, A. E. De; Peltola, T. O.; Jansen, J. A.; Närhi, T. O. Osteoblast Differentiation with Titania and Titania–Silica-Coated Titanium Fiber Meshes. *Tissue Eng.* **2005**, *11* (9–10), 1489–1497. https://doi.org/10.1089/ten.2005.11.1489.
(46) Liu, H.; Slamovich, E. B.; Webster, T. J. Increased Osteoblast Functions on Nanophase Titania Dispersed in Poly-Lactic- Co -Glycolic Acid Composites. *Nanotechnology* **2005**, *16* (7), S601–S608. https://doi.org/10.1088/0957-4484/16/7/038.
(47) Chan, K. H.; Zhuo, S.; Ni, M. Priming the Surface of Orthopedic Implants for Osteoblast Attachment in Bone Tissue Engineering. *Int. J. Med. Sci.* **2015**, *12* (9), 701–707.

https://doi.org/10.7150/ijms.12658.
(48) 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. https://doi.org/10.1021/acs.langmuir.7b02868.
(49) Villarreal-Ramirez, E.; Garduño-Juarez, R.; Gericke, A.; Boskey, A. The Role of Phosphorylation in Dentin Phosphoprotein Peptide Absorption to Hydroxyapatite Surfaces: A Molecular Dynamics Study. *Connect. Tissue Res.* **2014**, *55* (sup1), 134–137. https://doi.org/10.3109/03008207.2014.923870.
(50) Boskey, A. L.; Villarreal-Ramirez, E. Intrinsically Disordered Proteins and Biomineralization. *Matrix Biol.* **2016**, *52–54*, 43–59. https://doi.org/10.1016/j.matbio.2016.01.007.
(51) Xu, Z.; Yang, Y.; Zhao, W.; Wang, Z.; Landis, W. J.; Cui, Q.; Sahai, N. Molecular Mechanisms for Intrafibrillar Collagen Mineralization in Skeletal Tissues. *Biomaterials* **2015**, *39*, 59–66. https://doi.org/10.1016/j.biomaterials.2014.10.048.
(52) Huang, B.; Lou, Y.; Li, T.; Lin, Z.; Sun, S.; Yuan, Y.; Liu, C.; Gu, Y. Molecular Dynamics Simulations of Adsorption and Desorption of Bone Morphogenetic Protein-2 on Textured Hydroxyapatite Surfaces. *Acta Biomater.* **2018**, *80*, 121–130. https://doi.org/10.1016/j.actbio.2018.09.019.
(53) Church, A. T.; Hughes, Z. E.; Walsh, T. R. Improving the Description of Interactions between Ca2+ and Protein Carboxylate Groups, Including γ-Carboxyglutamic Acid: Revised CHARMM22∗ Parameters. *RSC Adv.* **2015**, *5* (83), 67820–67828. https://doi.org/10.1039/c5ra11268k.
(54) O'Young, J.; Liao, Y.; Xiao, Y.; Jalkanen, J.; Lajoie, G.; Karttunen, M.; Goldberg, H. A.; Hunter, G. K. Matrix Gla Protein Inhibits Ectopic Calcification by a Direct Interaction with Hydroxyapatite Crystals. *J. Am. Chem. Soc.* **2011**, *133* (45), 18406–18412. https://doi.org/10.1021/ja207628k.
(55) Lai, Z. B.; Wang, M.; Yan, C.; Oloyede, A. Molecular Dynamics Simulation of Mechanical Behavior of Osteopontin-Hydroxyapatite Interfaces. *J. Mech. Behav. Biomed. Mater.* **2014**, *36*, 12–20. https://doi.org/10.1016/j.jmbbm.2014.04.002.
(56) Xu, Z.; Yang, Y.; Wang, Z.; Mkhonto, D.; Shang, C.; Liu, Z. P.; Cui, Q.; Sahai, N. Small Molecule-Mediated Control of Hydroxyapatite Growth: Free Energy Calculations Benchmarked to Density Functional Theory. *J. Comput. Chem.* **2014**, *35* (1), 70–81. https://doi.org/10.1002/jcc.23474.
(57) Zhao, W.; Xu, Z.; Cui, Q.; Sahai, N. Predicting the Structure-Activity Relationship of Hydroxyapatite-Binding Peptides by Enhanced-Sampling Molecular Simulation. *Langmuir* **2016**, *32* (27), 7009–7022. https://doi.org/10.1021/acs.langmuir.6b01582.
(58) Lin, T.-J.; Heinz, H. Accurate Force Field Parameters and PH Resolved Surface Models for Hydroxyapatite to Understand Structure, Mechanics, Hydration, and Biological Interfaces. *J. Phys. Chem. C* **2016**, *120* (9), 4975–4992. https://doi.org/10.1021/acs.jpcc.5b12504.
(59) Walsh, T. R. Molecular Modelling of Peptide-Based Materials for Biomedical Applications. *Adv. Exp. Med. Biol.* **2017**, *1030*, 37–50. https://doi.org/10.1007/978-3-319-66095-0_3.
(60) Xie, Q.; Xue, Z.; Gu, H.; Hu, C.; Yang, M.; Wang, X.; Xu, D. Molecular Dynamics Exploration of Ordered-to-Disordered Surface Structures of Biomimetic Hydroxyapatite Nanoparticles. *J. Phys. Chem. C* **2018**, *122* (12), 6691–6703. https://doi.org/10.1021/acs.jpcc.8b00178.

(61) Wang, X.; Zhang, L.; Zeng, Q.; Jiang, G.; Yang, M. First-Principles Study on the Hydroxyl Migration from Inner to Surface in Hydroxyapatite. *Appl. Surf. Sci.* **2018**, *452*, 381–388. https://doi.org/10.1016/j.apsusc.2018.05.050.
(62) Demichelis, R.; Garcia, N. A.; Raiteri, P.; Innocenti Malini, R.; Freeman, C. L.; Harding, J. H.; Gale, J. D. Simulation of Calcium Phosphate Species in Aqueous Solution: Force Field Derivation. *J. Phys. Chem. B* **2018**, *122* (4), 1471–1483. https://doi.org/10.1021/acs.jpcb.7b10697.
(63) Luo, M.; Gao, Y.; Yang, S.; Quan, X.; Sun, D.; Liang, K.; Li, J.; Zhou, J. Computer Simulations of the Adsorption of an N-Terminal Peptide of Statherin, SN15, and Its Mutants on Hydroxyapatite Surfaces. *Phys. Chem. Chem. Phys.* **2019**, *21* (18), 9342–9351. https://doi.org/10.1039/c9cp01638d.
(64) Buckle, E. L.; Prakash, A.; Bonomi, M.; Sampath, J.; Pfaendtner, J.; Drobny, G. P. Solid-State NMR and MD Study of the Structure of the Statherin Mutant SNa15 on Mineral Surfaces. *J. Am. Chem. Soc.* **2019**, *141* (5), 1998–2011. https://doi.org/10.1021/jacs.8b10990.
(65) Skelton, A. A.; Liang, T.; Walsh, T. R. Interplay of Sequence, Conformation, and Binding at the Peptide−Titania Interface as Mediated by Water. *ACS Appl. Mater. Interfaces* **2009**, *1* (7), 1482–1491. https://doi.org/10.1021/am9001666.
(66) Monti, S.; Walsh, T. R. Free Energy Calculations of the Adsorption of Amino Acid Analogues at the Aqueous Titania Interface. *J. Phys. Chem. C* **2010**, *114* (50), 22197–22206. https://doi.org/10.1021/jp107859q.
(67) Sampath, J.; Kullman, A.; Gebhart, R.; Drobny, G.; Pfaendtner, J. No Title. *(under Rev.* **2019**.
(68) Kang, Y.; Li, X.; Tu, Y.; Wang, Q.; Ågren, H. On the Mechanism of Protein Adsorption onto Hydroxylated and Nonhydroxylated TiO 2 Surfaces. *J. Phys. Chem. C* **2010**, *114* (34), 14496–14502. https://doi.org/10.1021/jp1037156.
(69) Předota, M.; Bandura, A. V.; Cummings, P. T.; Kubicki, J. D.; Wesolowski, D. J.; Chialvo, A. A.; Machesky, M. L. Electric Double Layer at the Rutile (110) Surface. 1. Structure of Surfaces and Interfacial Water from Molecular Dynamics by Use of Ab Initio Potentials. *J. Phys. Chem. B* **2004**, *108* (32), 12049–12060. https://doi.org/10.1021/jp037197c.
(70) Deighan, M.; Pfaendtner, J. Exhaustively Sampling Peptide Adsorption with Metadynamics. *Langmuir* **2013**, *29* (25), 7999–8009. https://doi.org/10.1021/la4010664.
(71) 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. https://doi.org/10.1021/ct300297t.
(72) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25. https://doi.org/10.1016/j.softx.2015.06.001.
(73) 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. https://doi.org/10.1016/j.cpc.2009.05.011.
(74) Foloppe, N.; MacKerell, Jr., A. D. All-Atom Empirical Force Field for Nucleic Acids: I. Parameter Optimization Based on Small Molecule and Condensed Phase Macromolecular Target Data. *J. Comput. Chem.* **2000**, *21* (2), 86–104. https://doi.org/10.1002/(SICI)1096-

987X(20000130)21:2<86::AID-JCC2>3.0.CO;2-G.
(75) 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. https://doi.org/10.1021/j100308a038.
(76) Desmond, J. L.; Rodger, P. M.; Walsh, T. R. Testing the Inter-Operability of the CHARMM and SPC/Fw Force Fields for Conformational Sampling. *Mol. Simul.* **2014**, *40* (12), 912–921. https://doi.org/10.1080/08927022.2013.824574.
(77) 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. https://doi.org/10.1021/la3038846.
(78) Sultan, A. M.; Hughes, Z. E.; Walsh, T. R. Effect of Calcium Ions on Peptide Adsorption at the Aqueous Rutile Titania (110) Interface. *Biointerphases* **2018**, *13* (6), 06D403. https://doi.org/10.1116/1.5046531.
(79) Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R. Avogadro: An Advanced Semantic Chemical Editor, Visualization, and Analysis Platform. *J. Cheminform.* **2012**, *4* (1), 17. https://doi.org/10.1186/1758-2946-4-17.
(80) 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. https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H.
(81) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126* (1), 014101. https://doi.org/10.1063/1.2408420.
(82) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52* (12), 7182–7190. https://doi.org/10.1063/1.328693.
(83) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103* (19), 8577–8593. https://doi.org/10.1063/1.470117.
(84) Bussi, G.; Gervasio, F. L.; Laio, A.; Parrinello, M. Free-Energy Landscape for Hairpin Folding from Combined Parallel Tempering and Metadynamics. *J. Am. Chem. Soc.* **2006**, *128*, 13435. https://doi.org/10.1021/ja062463w.
(85) 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. https://doi.org/10.1021/ct200208h.
(86) Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **2008**, *100* (2), 020603. https://doi.org/10.1103/PhysRevLett.100.020603.
(87) Earl, D. J.; Deem, M. W. Parallel Tempering: Theory, Applications, and New Perspectives. *Phys. Chem. Chem. Phys.* **2005**, *7* (23), 3910. https://doi.org/10.1039/b509983h.
(88) Sprenger, K. G.; Pfaendtner, J. Strong Electrostatic Interactions Lead to Entropically Favorable Binding of Peptides to Charged Surfaces. *Langmuir* **2016**, *32* (22), 5690–5701. https://doi.org/10.1021/acs.langmuir.6b01296.
(89) The PLUMED consortium. Promoting Transparency and Reproducibility in Enhanced Molecular Simulations. *Nat. Methods* **2019**, *16* (8), 670–673. https://doi.org/10.1038/s41592-019-0506-8.
(90) Mao, C. M.; Sampath, J.; Sprenger, K. G.; Drobny, G.; Pfaendtner, J. Molecular Driving Forces in Peptide Adsorption to Metal Oxide Surfaces. *Langmuir* **2019**, *35* (17), 5911–5920.

https://doi.org/10.1021/acs.langmuir.8b01392.
(91) Hartvig, R. A.; van de Weert, M.; Østergaard, J.; Jorgensen, L.; Jensen, H. Protein Adsorption at Charged Surfaces: The Role of Electrostatic Interactions and Interfacial Charge Regulation. *Langmuir* **2011**, *27* (6), 2634–2643. https://doi.org/10.1021/la104720n.
(92) Hauschka, P. V.; Carr, S. A. Calcium-Dependent Alpha-Helical Structure in Osteocalcin. *Biochemistry* **1982**, *21* (10), 2538–2547. https://doi.org/10.1021/bi00539a038.
(93) Wierenga, P. A.; Gruppen, H. New Views on Foams from Protein Solutions. *Curr. Opin. Colloid Interface Sci.* **2010**, *15* (5), 365–373. https://doi.org/10.1016/j.cocis.2010.05.017.
(94) Dickinson, E. Proteins at Interfaces and in Emulsions Stability, Rheology and Interactions. *J. Chem. Soc. Faraday Trans.* **1998**, *94* (12), 1657–1669. https://doi.org/10.1039/a801167b.
(95) Wilde, P.; Mackie, A.; Husband, F.; Gunning, P.; Morris, V. Proteins and Emulsifiers at Liquid Interfaces. *Adv. Colloid Interface Sci.* **2004**, *108–109*, 63–71. https://doi.org/10.1016/j.cis.2003.10.011.
(96) McClements, D. J. Protein-Stabilized Emulsions. *Curr. Opin. Colloid Interface Sci.* **2004**, *9* (5), 305–313. https://doi.org/10.1016/j.cocis.2004.09.003.
(97) Dickinson, E. Biopolymer-Based Particles as Stabilizing Agents for Emulsions and Foams. *Food Hydrocoll.* **2017**, *68*, 219–231. https://doi.org/10.1016/j.foodhyd.2016.06.024.
(98) Masschalck, B.; Michiels, C. W. Antimicrobial Properties of Lysozyme in Relation to Foodborne Vegetative Bacteria. *Crit. Rev. Microbiol.* **2003**, *29* (3), 191–214. https://doi.org/10.1080/713610448.
(99) Wu, T.; Jiang, Q.; Wu, D.; Hu, Y.; Chen, S.; Ding, T.; Ye, X.; Liu, D.; Chen, J. What Is New in Lysozyme Research and Its Application in Food Industry? A Review. *Food Chem.* **2019**, *274*, 698–709. https://doi.org/10.1016/j.foodchem.2018.09.017.
(100) Blake, C. C. F.; Koenig, D. F.; Mair, G. A.; North, A. C. T.; Phillips, D. C.; Sarma, V. R. Structure of Hen Egg-White Lysozyme: A Three-Dimensional Fourier Synthesis at 2 Å Resolution. *Nature* **1965**, *206* (4986), 757–761. https://doi.org/10.1038/206757a0.
(101) Lu, J. R.; Su, T. J.; Thomas, R. K.; Penfold, J.; Webster, J. Structural Conformation of Lysozyme Layers at the Air/Water Interface Studied by Neutron Reflection. *J. Chem. Soc. Faraday Trans.* **1998**, *94* (21), 3279–3287. https://doi.org/10.1039/a805731a.
(102) Hunter, J. R.; Kilpatrick, P. K.; Carbonell, R. G. Lysozyme Adsorption at the Air/Water Interface. *J. Colloid Interface Sci.* **1990**, *137* (2), 462–482. https://doi.org/10.1016/0021-9797(90)90421-J.
(103) Perriman, A. W.; Henderson, M. J.; Evenhuis, C. R.; McGillivray, D. J.; White, J. W. Effect of the Air−Water Interface on the Structure of Lysozyme in the Presence of Guanidinium Chloride. *J. Phys. Chem. B* **2008**, *112* (31), 9532–9539. https://doi.org/10.1021/jp800354r.
(104) Delahaije, R. J. B. M.; Gruppen, H.; Giuseppin, M. L. F.; Wierenga, P. A. Quantitative Description of the Parameters Affecting the Adsorption Behaviour of Globular Proteins. *Colloids Surfaces B Biointerfaces* **2014**, *123*, 199–206. https://doi.org/10.1016/j.colsurfb.2014.09.015.
(105) Campbell, R. A.; Tummino, A.; Varga, I.; Milyaeva, O. Y.; Krycki, M. M.; Lin, S.-Y.; Laux, V.; Haertlein, M.; Forsyth, V. T.; Noskov, B. A. Adsorption of Denaturated Lysozyme at the Air–Water Interface: Structure and Morphology. *Langmuir* **2018**, *34* (17), 5020–5029. https://doi.org/10.1021/acs.langmuir.8b00545.
(106) Kim, G.; Gurau, M.; Kim, J.; Cremer, P. S. Investigations of Lysozyme Adsorption at the Air/Water and Quartz/Water Interfaces by Vibrational Sum Frequency Spectroscopy. *Langmuir* **2002**, *18* (7), 2807–2811. https://doi.org/10.1021/la0113365.

(107) Damodaran, S. In Situ Measurement of Conformational Changes in Proteins at Liquid Interfaces by Circular Dichroism Spectroscopy. *Anal. Bioanal. Chem.* **2003**, *376* (2), 182–188. https://doi.org/10.1007/s00216-003-1873-6.
(108) Postel, C.; Abillon, O.; Desbat, B. Structure and Denaturation of Adsorbed Lysozyme at the Air-Water Interface. *J. Colloid Interface Sci.* **2003**, *266* (1), 74–81. https://doi.org/10.1016/S0021-9797(03)00571-X.
(109) Lad, M. D.; Birembaut, F.; Matthew, J. M.; Frazier, R. A.; Green, R. J. The Adsorbed Conformation of Globular Proteins at the Air/Water Interface. *Phys. Chem. Chem. Phys.* **2006**, *8* (18), 2179–2186. https://doi.org/10.1039/b515934b.
(110) Yano, Y. F.; Uruga, T.; Tanida, H.; Toyokawa, H.; Terada, Y.; Takagaki, M.; Yamada, H. Driving Force Behind Adsorption-Induced Protein Unfolding: A Time-Resolved X-Ray Reflectivity Study on Lysozyme Adsorbed at an Air/Water Interface. *Langmuir* **2009**, *25* (1), 32–35. https://doi.org/10.1021/la803235x.
(111) Yano, Y. F.; Arakawa, E.; Voegeli, W.; Kamezawa, C.; Matsushita, T. Initial Conformation of Adsorbed Proteins at an Air–Water Interface. *J. Phys. Chem. B* **2018**, *122* (17), 4662–4666. https://doi.org/10.1021/acs.jpcb.8b01039.
(112) Ibarra-Molero, B.; Sanchez-Ruiz, J. M. Are There Equilibrium Intermediate States in the Urea-Induced Unfolding of Hen Egg-White Lysozyme? *Biochemistry* **1997**, *36* (31), 9616–9624. https://doi.org/10.1021/bi9703305.
(113) Lu, J. R.; Thomas, R. K. Neutron Reflection from Wet Interfaces. *J. Chem. Soc. - Faraday Trans.* **1998**, *94* (8), 995–1018. https://doi.org/10.1039/a707853f.
(114) Lu, J. R.; Su, T. J.; Howlin, B. J. The Effect of Solution PH on the Structural Conformation of Lysozyme Layers Adsorbed on the Surface of Water. *J. Phys. Chem. B* **1999**, *103* (28), 5903–5909. https://doi.org/10.1021/jp990129z.
(115) Kim, G.; Gurau, M.; Kim, J.; Cremer, P. S. Investigations of Lysozyme Adsorption at the Air/Water and Quartz/Water Interfaces by Vibrational Sum Frequency Spectroscopy. *Langmuir* **2002**, *18* (7), 2807–2811. https://doi.org/10.1021/la0113365.
(116) Alahverdjieva, V. S.; Grigoriev, D. O.; Ferri, J. K.; Fainerman, V. B.; Aksenenko, E. V.; Leser, M. E.; Michel, M.; Miller, R. Adsorption Behaviour of Hen Egg-White Lysozyme at the Air/Water Interface. *Colloids Surfaces A Physicochem. Eng. Asp.* **2008**, *323* (1–3), 167–174. https://doi.org/10.1016/j.colsurfa.2007.12.031.
(117) Graham, D. E.; Phillips, M. C. Proteins at Liquid Interfaces: I. Kinetics of Adsorption and Surface Denaturation. *J. Coll* **1978**, *70* (3), 403–414. https://doi.org/090403-12.
(118) Graham, D. .; Phillips, M. . Proteins at Liquid Interfaces: II. Adsorption Isotherms. *J. Colloid Interface Sci.* **1978**, *70* (3), 415–426. https://doi.org/10.1016/0021-9797(79)90048-1.
(119) Graham, D. .; Phillips, M. . Proteins at Liquid Interfaces: III. Molecular Structures of Adsorbed Films. *J. Colloid Interface Sci.* **1978**, *70* (3), 427–439. https://doi.org/10.1016/0021-9797(79)90048-1.
(120) Wertz, C. F.; Santore, M. M. Adsorption and Reorientation Kinetics of Lysozyme on Hydrophobic Surfaces. *Langmuir* **2002**, *18* (4), 1190–1199. https://doi.org/10.1021/la0108813.
(121) Lu, J. R.; Su, T. J.; Thirtle, P. N.; Thomas, R. K.; Rennie, A. R.; Cubitt, R. The Denaturation of Lysozyme Layers Adsorbed at the Hydrophobic Soild/Liquid Surface Studied by Neutron Reflection. *J. Colloid Interface Sci.* **1998**, *206* (1), 212–223. https://doi.org/10.1006/jcis.1998.5680.

(122) Cheung, D. L. Adsorption and Conformations of Lysozyme and α -Lactalbumin at a Water-Octane Interface. *J. Chem. Phys.* **2017**, *147* (19). https://doi.org/10.1063/1.4994561.
(123) Wei, T.; Carignano, M. A.; Szleifer, I. Lysozyme Adsorption on Polyethylene Surfaces: Why Are Long Simulations Needed? *Langmuir* **2011**, *27* (19), 12074–12081. https://doi.org/10.1021/la202622s.
(124) Rzeźnicka, I. I.; Pandey, R.; Schleeger, M.; Bonn, M.; Weidner, T. Formation of Lysozyme Oligomers at Model Cell Membranes Monitored with Sum Frequency Generation Spectroscopy. *Langmuir* **2014**, *30* (26), 7736–7744. https://doi.org/10.1021/la5010227.
(125) Beckner, W.; He, Y.; Pfaendtner, J. Chain Flexibility in Self-Assembled Monolayers Affects Protein Adsorption and Surface Hydration: A Molecular Dynamics Study. *J. Phys. Chem. B* **2016**, *120* (40), 10423–10432. https://doi.org/10.1021/acs.jpcb.6b05882.
(126) Wahlgren, M.; Arnebrant, T.; Lundstrom, I. The Adsorption of Lysozyme to Hydrophillic Silicon Oxide Surfaces: Comparison between Experimental Data and Models for Adsorption Kintetics. 1995, pp 506–514. https://doi.org/0021-9797/95.
(127) Sigal, G. B.; Mrksich, M.; Whitesides, G. M. Effect of Surface Wettability on the Adsorption of Proteins and Detergents. *J. Am. Chem. Soc.* **1998**, *120* (14), 3464–3473. https://doi.org/10.1021/ja970819l.
(128) Su, T. J.; Lu, J. R.; Thomas, R. K.; Cui, Z. F.; Penfold, J. The Adsorption of Lysozyme at the Silica-Water Interface: A Neutron Reflection Study. *J. Colloid Interface Sci.* **1998**, *203* (2), 419–429. https://doi.org/10.1006/jcis.1998.5545.
(129) Hildebrand, N.; Wei, G.; Köppen, S.; Colombi Ciacchi, L. Simulated and Experimental Force Spectroscopy of Lysozyme on Silica. *Phys. Chem. Chem. Phys.* **2018**, *20* (29), 19595–19605. https://doi.org/10.1039/c8cp03747g.
(130) Mulheran, P.; Kubiak, K. Protein Adsorption Mechanisms on Solid Surfaces: Lysozyme-on-Mica. *Mol. Simul.* **2009**, *35* (7), 561–566. https://doi.org/10.1080/08927020802610288.
(131) Carlsson, F.; Hyltner, E.; Arnebrant, T.; Malmsten, M.; Linse, P. Lysozyme Adsorption to Charged Surfaces. A Monte Carlo Study. *J. Phys. Chem. B* **2007**, *108* (28), 9871–9881. https://doi.org/10.1021/jp0495186.
(132) Kubiak, K.; Mulheran, P. A. Molecular Dynamics Simulations of Hen Egg White Lysozyme Adsorption at a Charged Solid. **2009**, No. 113, 12189–12200. https://doi.org/10.1021/jp901521x.
(133) Zhao, Y.; Cieplak, M. Proteins at Air–Water and Oil–Water Interfaces in an All-Atom Model. *Phys. Chem. Chem. Phys.* **2017**, *19* (36), 25197–25206. https://doi.org/10.1039/C7CP03829A.
(134) Brotzakis, Z. F.; Voets, I. K.; Bakker, H. J.; Bolhuis, P. G. Water Structure and Dynamics in the Hydration Layer of a Type III Anti-Freeze Protein. *Phys. Chem. Chem. Phys.* **2018**, *20* (10), 6996–7006. https://doi.org/10.1039/c8cp00170g.
(135) Cheung, D.; Samantray, S. Molecular Dynamics Simulation of Protein Biosurfactants. *Colloids and Interfaces* **2018**, *2* (3), 39. https://doi.org/10.3390/colloids2030039.
(136) Fan, H.; Wang, X.; Zhu, J.; Robillard, G. T.; Mark, A. E. Molecular Dynamics Simulations of the Hydrophobin SC3 at a Hydrophobic/Hydrophilic Interface. *Proteins Struct. Funct. Bioinforma.* **2006**, *64* (4), 863–873. https://doi.org/10.1002/prot.20936.
(137) Ley, K.; Christofferson, A.; Penna, M.; Winkler, D.; Maclaughlin, S.; Yarovsky, I. Surface-Water Interface Induces Conformational Changes Critical for Protein Adsorption: Implications for Monolayer Formation of EAS Hydrophobin. *Front. Mol. Biosci.* **2015**, *2* (November), 1–12. https://doi.org/10.3389/fmolb.2015.00064.

(138) Verreault, D.; Alamdari, S.; Roeters, S. J.; Pandey, R.; Pfaendtner, J.; Weidner, T. Ice-Binding Site of Surface-Bound Type III Antifreeze Protein Partially Decoupled from Water. *Phys. Chem. Chem. Phys.* **2018**, *20* (42), 26926–26933. https://doi.org/10.1039/c8cp03382j.
(139) Cieplak, M.; Allan, D. B.; Leheny, R. L.; Reich, D. H. Proteins at Air–Water Interfaces: A Coarse-Grained Model. *Langmuir* **2014**, *30* (43), 12888–12896. https://doi.org/10.1021/la502465m.
(140) Zhao, Y.; Cieplak, M. Proteins at the Air-Water Interface in a Lattice Model. *J. Chem. Phys.* **2018**, *148* (9), 094704. https://doi.org/10.1063/1.5017789.
(141) Zhao, Y.; Cieplak, M. Proteins at Curved Fluid–Fluid Interfaces in a Coarse-Grained Model. *J. Phys. Condens. Matter* **2020**, *32* (40), 404003. https://doi.org/10.1088/1361-648X/ab94f6.
(142) Schmüser, L.; Roeters, S.; Lutz, H.; Woutersen, S.; Bonn, M.; Weidner, T. Determination of Absolute Orientation of Protein α-Helices at Interfaces Using Phase-Resolved Sum Frequency Generation Spectroscopy. *J. Phys. Chem. Lett.* **2017**, *8* (13), 3101–3105. https://doi.org/10.1021/acs.jpclett.7b01059.
(143) Lu, H.; Lutz, H.; Roeters, S. J.; Hood, M. A.; Schäfer, A.; Muñoz-Espí, R.; Berger, R.; Bonn, M.; Weidner, T. Calcium-Induced Molecular Rearrangement of Peptide Folds Enables Biomineralization of Vaterite Calcium Carbonate. *J. Am. Chem. Soc.* **2018**, *140* (8), 2793–2796. https://doi.org/10.1021/jacs.8b00281.
(144) Hosseinpour, S.; Roeters, S. J.; Bonn, M.; Peukert, W.; Woutersen, S.; Weidner, T. Structure and Dynamics of Interfacial Peptides and Proteins from Vibrational Sum-Frequency Generation Spectroscopy. *Chem. Rev.* **2020**, *120* (7), 3420–3465. https://doi.org/10.1021/acs.chemrev.9b00410.
(145) Perets, E. A.; Yan, E. C. Y. Chiral Water Superstructures around Antiparallel β-Sheets Observed by Chiral Vibrational Sum Frequency Generation Spectroscopy. *J. Phys. Chem. Lett.* **2019**, *10* (12), 3395–3401. https://doi.org/10.1021/acs.jpclett.9b00878.
(146) Tan, J.; Zhang, J.; Luo, Y.; Ye, S. Misfolding of a Human Islet Amyloid Polypeptide at the Lipid Membrane Populates through β-Sheet Conformers without Involving α-Helical Intermediates. *J. Am. Chem. Soc.* **2019**, *141* (5), 1941–1948. https://doi.org/10.1021/jacs.8b08537.
(147) Zou, X.; Wei, S.; Badieyan, S.; Schroeder, M.; Jasensky, J.; Brooks, C. L.; Marsh, E. N. G.; Chen, Z. Investigating the Effect of Two-Point Surface Attachment on Enzyme Stability and Activity. *J. Am. Chem. Soc.* **2018**, *140* (48), 16560–16569. https://doi.org/10.1021/jacs.8b08138.
(148) Boyd, R. W. *Nonlinear Optics*, 2nd ed.; Press, A., Ed.; Elsevier: San Diego, CA, 2003. https://doi.org/10.1016/B978-0-12-121682-5.X5000-7.
(149) 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. https://doi.org/10.1002/anie.201702707.
(150) Donovan, M. A.; Yimer, Y. Y.; Pfaendtner, J.; Backus, E. H. G.; Bonn, M.; Weidner, T. Ultrafast Reorientational Dynamics of Leucine at the Air–Water Interface. *J. Am. Chem. Soc.* **2016**, *138* (16), 5226–5229. https://doi.org/10.1021/jacs.6b01878.
(151) Sauter, C.; Otálora, F.; Gavira, J. A.; Vidal, O.; Giegé, R.; García-Ruiz, J. M. Structure of Tetragonal Hen Egg-White Lysozyme at 0.94 Å from Crystals Grown by the Counter-Diffusion Method. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **2001**, *57* (8), 1119–1126. https://doi.org/10.1107/S0907444901008873.

(152) Oostenbrink, C.; Villa, A.; Mark, A. E.; Van Gunsteren, W. F. A Biomolecular Force Field Based on the Free Enthalpy of Hydration and Solvation: The GROMOS Force-Field Parameter Sets 53A5 and 53A6. *J. Comput. Chem.* **2004**, *25* (13), 1656–1676. https://doi.org/10.1002/jcc.20090.
(153) Robustelli, P.; Piana, S.; Shaw, D. E. Developing a Molecular Dynamics Force Field for Both Folded and Disordered Protein States. *Proc. Natl. Acad. Sci.* **2018**, *115* (21), E4758–E4766. https://doi.org/10.1073/pnas.1800690115.
(154) Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E. Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States. *J. Phys. Chem. B* **2015**, *119* (16), 5113–5123. https://doi.org/10.1021/jp508971m.
(155) 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. https://doi.org/10.1063/1.448118.
(156) Humphrey, W.; Dalke, A.; Schulten, K. VMD - Visual Molecular Dynamics. Journal of Molecular Graphics 1996, pp 33–38.
(157) Golbek, T. W.; Schmüser, L.; Rasmussen, M. H.; Poulsen, T. B.; Weidner, T. Lasalocid Acid Antibiotic at a Membrane Surface Probed by Sum Frequency Generation Spectroscopy. *Langmuir* **2020**, *36* (12), 3184–3192. https://doi.org/10.1021/acs.langmuir.9b03752.
(158) Roeters, S. J.; Van Dijk, C. N.; Torres-Knoop, A.; Backus, E. H. G.; Campen, R. K.; Bonn, M.; Woutersen, S. Determining in Situ Protein Conformation and Orientation from the Amide-I Sum-Frequency Generation Spectrum: Theory and Experiment. *J. Phys. Chem. A* **2013**, *117* (29), 6311–6322. https://doi.org/10.1021/jp401159r.
(159) Hamm, P.; Zanni, M. *Concepts and Methods of 2D Infrared Spectroscopy*; Cambridge University Press: Cambridge, U.K., 2011.
(160) Gorbunov, R. D.; Kosov, D. S.; Stock, G. Ab Initio -Based Exciton Model of Amide I Vibrations in Peptides: Definition, Conformational Dependence, and Transferability. *J. Chem. Phys.* **2005**, *122* (22), 224904. https://doi.org/10.1063/1.1898215.
(161) Krimm, S.; Bandekar, J. Vibrational Spectroscopy and Conformation of Peptides, Polypeptides, and Proteins; 1986; pp 181–364. https://doi.org/10.1016/S0065-3233(08)60528-8.
(162) Lu, H.; Schäfer, A.; Lutz, H.; Roeters, S. J.; Lieberwirth, I.; Muñoz-Espí, R.; Hood, M. A.; Bonn, M.; Weidner, T. Peptide-Controlled Assembly of Macroscopic Calcium Oxalate Nanosheets. *J. Phys. Chem. Lett.* **2019**, *10* (9), 2170–2174. https://doi.org/10.1021/acs.jpclett.9b00684.
(163) Ham, S.; Kim, J.-H.; Lee, H.; Cho, M. Correlation between Electronic and Molecular Structure Distortions and Vibrational Properties. II. Amide I Modes of NMA–ND2O Complexes. *J. Chem. Phys.* **2003**, *118* (8), 3491–3498. https://doi.org/10.1063/1.1536980.
(164) Wang, H.-F.; Gan, W.; Lu, R.; Rao, Y.; Wu, B.-H. Quantitative Spectral and Orientational Analysis in Surface Sum Frequency Generation Vibrational Spectroscopy (SFG-VS). *Int. Rev. Phys. Chem.* **2005**, *24* (2), 191–256. https://doi.org/10.1080/01442350500225894.
(165) Nguyen, K. T.; Le Clair, S. V.; Ye, S.; Chen, Z. Orientation Determination of Protein Helical Secondary Structures Using Linear and Nonlinear Vibrational Spectroscopy. *J. Phys. Chem. B* **2009**, *113* (36), 12169–12180. https://doi.org/10.1021/jp904153z.
(166) Nguyen, K. T.; King, J. T.; Chen, Z. Orientation Determination of Interfacial β-Sheet Structures in Situ. *J. Phys. Chem. B* **2010**, *114* (25), 8291–8300.

https://doi.org/10.1021/jp102343h.
(167) Thakur, G.; Wang, C.; Leblanc, R. M. Surface Chemistry and in Situ Spectroscopy of a Lysozyme Langmuir Monolayer. *Langmuir* **2008**, *24* (9), 4888–4893. https://doi.org/10.1021/la703893m.
(168) McManus, J. J.; Charbonneau, P.; Zaccarelli, E.; Asherie, N. The Physics of Protein Self-Assembly. *Curr. Opin. Colloid Interface Sci.* **2016**, *22*, 73–79. https://doi.org/10.1016/j.cocis.2016.02.011.
(169) Simon, A. J.; Zhou, Y.; Ramasubramani, V.; Glaser, J.; Pothukuchy, A.; Gollihar, J.; Gerberich, J. C.; Leggere, J. C.; Morrow, B. R.; Jung, C.; et al. Supercharging Enables Organized Assembly of Synthetic Biomolecules. *Nat. Chem.* **2019**, *11* (3), 204–212. https://doi.org/10.1038/s41557-018-0196-3.
(170) Eom, K. Computer Simulation of Protein Materials at Multiple Length Scales: From Single Proteins to Protein Assemblies. *Multiscale Sci. Eng.* **2019**, *1* (1), 1–25. https://doi.org/10.1007/s42493-018-00009-7.
(171) Ferrari, S.; Kahl, G.; Bianchi, E. Molecular Dynamics Simulations of Inverse Patchy Colloids. *Eur. Phys. J. E* **2018**, *41* (3), 1–5. https://doi.org/10.1140/epje/i2018-11647-2.
(172) Akcora, P.; Liu, H.; Kumar, S. K.; Moll, J.; Li, Y.; Benicewicz, B. C.; Schadler, L. S.; Acehan, D.; Panagiotopoulos, A. Z.; Pryamitsyn, V.; et al. Anisotropic Self-Assembly of Spherical Polymer-Grafted Nanoparticles. *Nat. Mater.* **2009**, *8* (4), 354–359. https://doi.org/10.1038/nmat2404.
(173) Zhang, Z.; Marson, R. L.; Ge, Z.; Glotzer, S. C.; Ma, P. X. Simultaneous Nano- and Microscale Control of Nanofibrous Microspheres Self-Assembled from Star-Shaped Polymers. *Adv. Mater.* **2015**, *27* (26), 3947–3952. https://doi.org/10.1002/adma.201501329.
(174) Butler, P. J. G. Self-Assembly of Tobacco Mosaic Virus: The Role of an Intermediate Aggregate in Generating Both Specificity and Speed. *Philos. Trans. R. Soc. B Biol. Sci.* **1999**, *354* (1383), 537–550. https://doi.org/10.1098/rstb.1999.0405.
(175) Clare, D. K.; Orlova, E. V. 4.6Å Cryo-EM Reconstruction of Tobacco Mosaic Virus from Images Recorded at 300keV on a 4k × 4k CCD Camera. *J. Struct. Biol.* **2010**, *171* (3), 303–308. https://doi.org/10.1016/j.jsb.2010.06.011.
(176) Tanaka, S.; Kerfeld, C. A.; Sawaya, M. R.; Cai, F.; Heinhorst, S.; Cannon, G. C.; Yeates, T. O. Atomic-Level Models of the Bacterial Carboxysome Shell. *Science (80-.).* **2008**, *319* (5866), 1083–1086. https://doi.org/10.1126/science.1151458.
(177) Chhabra, R.; Sharma, J.; Liu, Y.; Rinker, S.; Yan, H. DNA Self-Assembly for Nanomedicine. *Adv. Drug Deliv. Rev.* **2010**, *62* (6), 617–625. https://doi.org/10.1016/j.addr.2010.03.005.
(178) Bai, Y.; Luo, Q.; Liu, J. Protein Self-Assembly: Via Supramolecular Strategies. *Chem. Soc. Rev.* **2016**, *45* (10), 2756–2767. https://doi.org/10.1039/c6cs00004e.
(179) Welling, R. C.; Knotts, T. A. The Effects of Multiple Probes on the Hybridization of Target DNA on Surfaces. *J. Chem. Phys.* **2015**, *142* (1), 015102. https://doi.org/10.1063/1.4904929.
(180) Fogg, J. M.; Randall, G. L.; Pettitt, B. M.; Sumners, D. W. L.; Harris, S. A.; Zechiedrich, L. Bullied No More: When and How DNA Shoves Proteins Around. *Q. Rev. Biophys.* **2012**, *45* (3), 257–299. https://doi.org/10.1017/S0033583512000054.
(181) De Felice, F. G.; Houzel, J. C.; Garcia-Abreu, J.; Louzada, P. R.; Afonso, R. C.; Meirelles, M. N.; Lent, R.; Neto, V. M.; Ferreira, S. T. Inhibition of Alzheimer's Disease Beta-

Amyloid Aggregation, Neurotoxicity, and in Vivo Deposition by Nitrophenols: Implications for Alzheimer's Therapy. *FASEB J.* **2001**, *5* (7), 1297–1299. https://doi.org/10.1096/fj.00-0676fje.

(182) Finder, V.; Glockshuber, R. Amyloid-Beta Aggregation. *Nerodegenerative Dis.* **2007**, *4* (1), 13–27. https://doi.org/https://doi.org/10.1159/000100355.

(183) Pietrucci, F.; Laio, A. A Collective Variable for the Efficient Exploration of Protein Beta-Sheet Structures: Application to SH3 and GB1. *J. Chem. Theory Comput.* **2009**, *5* (9), 2197–2201. https://doi.org/10.1021/ct900202f.

(184) Chiti, F.; Dobson, C. M. Protein Misfolding, Functional Amyloid, and Human Disease. *Annu. Rev. Biochem.* **2006**, *75* (1), 333–366. https://doi.org/10.1146/annurev.biochem.75.101304.123901.

(185) Li, H.; Lykotrafitis, G. A Coarse-Grain Molecular Dynamics Model for Sickle Hemoglobin Fibers. *J. Mech. Behav. Biomed. Mater.* **2011**, *4* (2), 162–173. https://doi.org/10.1016/j.jmbbm.2010.11.002.

(186) Galamba, N.; Pipolo, S. On the Binding Free Energy and Molecular Origin of Sickle Cell Hemoglobin Aggregation. *J. Phys. Chem. B* **2018**, *122* (30), 7475–7483. https://doi.org/10.1021/acs.jpcb.8b03708.

(187) Christoph, G. W.; Hofrichter, J.; Eaton, W. A. Understanding the Shape of Sickled Red Cells. *Biophys. J.* **2005**, *88* (2), 1371–1376. https://doi.org/10.1529/biophysj.104.051250.

(188) Garmann, R. F.; Goldfain, A. M.; Manoharan, V. N. Measurements of the Self-Assembly Kinetics of Individual Viral Capsids around Their RNA Genome. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116* (45), 22485–22490. https://doi.org/10.1073/pnas.1909223116.

(189) Tsao, Y. T.; Huang, Y. J.; Wu, H. H.; Liu, Y. A.; Liu, Y. S.; Lee, O. K. Osteocalcin Mediates Biomineralization during Osteogenic Maturation in Human Mesenchymal Stromal Cells. *Int. J. Mol. Sci.* **2017**, *18* (1). https://doi.org/10.3390/ijms18010159.

(190) Kröger, N.; Deutzmann, R.; Sumper, M. Polycationic Peptides from Diatom Biosilica That Direct Silica Nanosphere Formation. *Science (80-.).* **1999**, *286* (5442), 1129–1132. https://doi.org/10.1126/science.286.5442.1129.

(191) Sanii, B.; Haxton, T. K.; Olivier, G. K.; Cho, A.; Barton, B.; Proulx, C.; Whitelam, S.; Zuckermann, R. N. Structure-Determining Step in the Hierarchical Assembly of Peptoid Nanosheets. *ACS Nano* **2014**, *8* (11), 11674–11684. https://doi.org/10.1021/nn505007u.

(192) 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. https://doi.org/10.1002/adma.201500275.

(193) Rothemund, P. W. K. Folding DNA to Create Nanoscale Shapes and Patterns. *Nature* **2006**, *440* (7082), 297–302. https://doi.org/10.1038/nature04586.

(194) Wei, B.; Dai, M.; Yin, P. Complex Shapes Self-Assembled from Single-Stranded DNA Tiles. *Nature* **2012**, *485* (7400), 623–626. https://doi.org/10.1038/nature11075.

(195) Zhang, F.; Jiang, S.; Wu, S.; Li, Y.; Mao, C.; Liu, Y.; Yan, H. Complex Wireframe DNA Origami Nanostructures with Multi-Arm Junction Vertices. *Nat. Nanotechnol.* **2015**, *10* (9), 779–784. https://doi.org/10.1038/nnano.2015.162.

(196) Praetorius, F.; Kick, B.; Behler, K. L.; Honemann, M. N.; Weuster-Botz, D.; Dietz, H. Biotechnological Mass Production of DNA Origami. *Nature* **2017**, *552* (7683), 84–87. https://doi.org/10.1038/nature24650.

(197) Gerling, T.; Wagenbauer, K.; NeuNer, A.; Dietz, H. Dynamic DNA Devices and Assemblies Formed by Shape-Complementary, Non–Base Pairing 3D Components. *Science*

(80-.). **2015**, *347*, 1446–1452.
(198) Pujals, S.; Tao, K.; Terradellas, A.; Gazit, E.; Albertazzi, L. Studying Structure and Dynamics of Self-Assembled Peptide Nanostructures Using Fluorescence and Super Resolution Microscopy. *Chem. Commun.* **2017**, *53* (53), 7294–7297. https://doi.org/10.1039/c7cc02176c.
(199) Sung, J. J.; Pardeshi, N. N.; Mulder, A. M.; Mulligan, S. K.; Quispe, J.; On, K.; Carragher, B.; Potter, C. S.; Carpenter, J. F.; Schneemann, A. Transmission Electron Microscopy as an Orthogonal Method to Characterize Protein Aggregates. *J. Pharm. Sci.* **2015**, *104* (2), 750–759. https://doi.org/10.1002/jps.24157.
(200) Jeon, J.; Thurber, K. R.; Ghirlando, R.; Yau, W. M.; Tycko, R. Application of Millisecond Time-Resolved Solid State NMR to the Kinetics and Mechanism of Melittin Self-Assembly. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116* (34), 16717–16722. https://doi.org/10.1073/pnas.1908006116.
(201) 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. https://doi.org/10.1021/acs.accounts.5b00439.
(202) 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. https://doi.org/10.1038/nmat4891.
(203) Hellner, B.; Lee, S. B.; Subramaniam, A.; Subramanian, V. R.; Baneyx, F. Modeling the Cooperative Adsorption of Solid-Binding Proteins on Silica: Molecular Insights from Surface Plasmon Resonance Measurements. *Langmuir* **2019**, *35* (14), 5013–5020. https://doi.org/10.1021/acs.langmuir.9b00283.
(204) Lee, E. H.; Hsin, J.; Sotomayor, M.; Comellas, G.; Schulten, K. Discovery Through the Computational Microscope. *Structure* **2009**, *17* (10), 1295–1306. https://doi.org/10.1016/j.str.2009.09.001.
(205) Glotzer, S. C.; Solomon, M. J. Anisotropy of Building Blocks and Their Assembly into Complex Structures. *Nat. Mater.* **2007**, *6* (8), 557–562. https://doi.org/10.1038/nmat1949.
(206) Cersonsky, R. K.; Van Anders, G.; Dodd, P. M.; Glotzer, S. C. Relevance of Packing to Colloidal Self-Assembly. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115* (7), 1439–1444. https://doi.org/10.1073/pnas.1720139115.
(207) Barrat, J.-L.; Baschnagel, J.; Lyulin, A. Molecular Dynamics Simulations of Glassy Polymers. *Soft Matter* **2010**, *6* (15), 3430. https://doi.org/10.1039/b927044b.
(208) Peter, C.; Kremer, K. Multiscale Simulation of Soft Matter Systems – from the Atomistic to the Coarse-Grained Level and Back. *Soft Matter* **2009**, *5* (22), 4357. https://doi.org/10.1039/b912027k.
(209) Karmakar, S.; Dasgupta, C.; Sastry, S. Length Scales in Glass-Forming Liquids and Related Systems: A Review. *Reports Prog. Phys.* **2016**, *79* (1), 016601. https://doi.org/10.1088/0034-4885/79/1/016601.
(210) Jackson, N. E.; Webb, M. A.; de Pablo, J. J. Recent Advances in Machine Learning towards Multiscale Soft Materials Design. *Curr. Opin. Chem. Eng.* **2019**, *23*, 106–114. https://doi.org/10.1016/j.coche.2019.03.005.
(211) Torres, J. A.; Nealey, P. F.; de Pablo, J. J. Molecular Simulation of Ultrathin Polymeric Films near the Glass Transition. *Phys. Rev. Lett.* **2000**, *85* (15), 3221–3224.

https://doi.org/10.1103/PhysRevLett.85.3221.
(212) 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. https://doi.org/10.1021/acs.accounts.7b00065.
(213) Frederix, P. W. J. M.; Patmanidis, I.; Marrink, S. J. Molecular Simulations of Self-Assembling Bio-Inspired Supramolecular Systems and Their Connection to Experiments. *Chem. Soc. Rev.* **2018**, *47* (10), 3470–3489. https://doi.org/10.1039/C8CS00040A.
(214) Morriss-Andrews, A.; Shea, J.-E. Computational Studies of Protein Aggregation: Methods and Applications. *Annu. Rev. Phys. Chem.* **2015**, *66* (1), 643–666. https://doi.org/10.1146/annurev-physchem-040513-103738.
(215) Taylor, P. A.; Jayaraman, A. Molecular Modeling and Simulations of Peptide–Polymer Conjugates. *Annu. Rev. Chem. Biomol. Eng.* **2020**, *11* (1), 257–276. https://doi.org/10.1146/annurev-chembioeng-092319-083243.
(216) Harding, J. H.; Duffy, D. M.; Sushko, M. L.; Rodger, P. M.; Quigley, D.; Elliott, J. A. Computational Techniques at the Organic−Inorganic Interface in Biomineralization. *Chem. Rev.* **2008**, *108* (11), 4823–4854. https://doi.org/10.1021/cr078278y.
(217) Demichelis, R.; Schuitemaker, A.; Garcia, N. A.; Koziara, K. B.; De La Pierre, M.; Raiteri, P.; Gale, J. D. Simulation of Crystallization of Biominerals. *Annu. Rev. Mater. Res.* **2018**, *48* (1), 327–352. https://doi.org/10.1146/annurev-matsci-070317-124327.
(218) Demichelis, R.; Raiteri, P.; Gale, J. D.; Dovesi, R. Examining the Accuracy of Density Functional Theory for Predicting the Thermodynamics of Water Incorporation into Minerals: The Hydrates of Calcium Carbonate. *J. Phys. Chem. C* **2013**, *117* (34), 17814–17823. https://doi.org/10.1021/jp4048105.
(219) Rimola, A.; Corno, M.; Zicovich-Wilson, C. M.; Ugliengo, P. Ab Initio Modeling of Protein/Biomaterial Interactions: Competitive Adsorption between Glycine and Water onto Hydroxyapatite Surfaces. *Phys. Chem. Chem. Phys.* **2009**, *11* (40), 9005. https://doi.org/10.1039/b913311a.
(220) Buckle, E. L.; Lum, J. S.; Roehrich, A. M.; Stote, R. E.; Vandermoon, B.; Dracinsky, M.; Filocamo, S. F.; Drobny, G. P. Serine–Lysine Peptides as Mediators for the Production of Titanium Dioxide: Investigating the Effects of Primary and Secondary Structures Using Solid-State NMR Spectroscopy and DFT Calculations. *J. Phys. Chem. B* **2018**, *122* (17), 4708–4718. https://doi.org/10.1021/acs.jpcb.8b00745.
(221) Dong, X.; Wang, Q.; Wu, T.; Pan, H. Understanding Adsorption-Desorption Dynamics of BMP-2 on Hydroxyapatite (001) Surface. *Biophys. J.* **2007**, *93* (3), 750–759. https://doi.org/10.1529/biophysj.106.103168.
(222) Bottaro, S.; Lindorff-Larsen, K. Biophysical Experiments and Biomolecular Simulations: A Perfect Match? *Science (80-.).* **2018**, *361* (6400), 355–360. https://doi.org/10.1126/science.aat4010.
(223) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79* (2), 926–935. https://doi.org/10.1063/1.445869.
(224) 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. https://doi.org/10.1021/j100308a038.
(225) 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.

https://doi.org/10.1021/cm500365c.
(226) Freeman, C. L.; Harding, J. H.; Cooke, D. J.; Elliott, J. A.; Lardge, J. S.; Duffy, D. M. New Forcefields for Modeling Biomineralization Processes. *J. Phys. Chem. C* **2007**, *111* (32), 11943–11951. https://doi.org/10.1021/jp071887p.
(227) Liu, J.; Wang, Z.; Zeng, J.; Heinz, H. Molecular Structure and Assembly of Peptide-Derived Nanomaterials. *Curr. Opin. Green Sustain. Chem.* **2018**, *12*, 38–46. https://doi.org/10.1016/j.cogsc.2018.05.007.
(228) Walsh, T. R.; Knecht, M. R. Biointerface Structural Effects on the Properties and Applications of Bioinspired Peptide-Based Nanomaterials. *Chem. Rev.* **2017**, *117* (20), 12641–12704. https://doi.org/10.1021/acs.chemrev.7b00139.
(229) Abrams, C.; Bussi, G. Enhanced Sampling in Molecular Dynamics Using Metadynamics, Replica-Exchange, and Temperature-Acceleration. *Entropy* **2014**, *16* (1), 163–199. https://doi.org/10.3390/e16010163.
(230) Zheng, S.; Pfaendtner, J. Enhanced Sampling of Chemical and Biochemical Reactions with Metadynamics. *Mol. Simul.* **2015**, *41* (1–3). https://doi.org/10.1080/08927022.2014.923574.
(231) Barducci, A.; Bonomi, M.; Parrinello, M. Metadynamics. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1* (5), 826–843. https://doi.org/10.1002/wcms.31.
(232) Pfaendtner, J. Metadynamics to Enhance Sampling in Biomolecular Simulations. In *Biomolecular Simulations:Methods in Molecular Biology*; Bonomi, M., Camilloni, C., Eds.; Springer, 2018; Vol. In Press.
(233) Bonomi, M.; Parrinello, M. Enhanced Sampling in the Well-Tempered Ensemble. *Phys. Rev. Lett.* **2010**, *104* (19), 1–4. https://doi.org/10.1103/PhysRevLett.104.190601.
(234) Piana, S.; Laio, A. A Bias-Exchange Approach to Protein Folding. *J. Phys. Chem. B* **2007**, *111* (4553). https://doi.org/10.1021/jp067873l.
(235) 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. https://doi.org/10.1021/acs.jctc.5b00846.
(236) Prakash, A.; Fu, C. D.; Bonomi, M.; Pfaendtner, J. Biasing Smarter, Not Harder, by Partitioning Collective Variables into Families in Parallel Bias Metadynamics. *J. Chem. Theory Comput.* **2018**, *14* (10), 4985–4990. https://doi.org/10.1021/acs.jctc.8b00448.
(237) Noid, W. G. Perspective: Coarse-Grained Models for Biomolecular Systems. *J. Chem. Phys.* **2013**, *139* (9). https://doi.org/10.1063/1.4818908.
(238) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S.-J. The MARTINI Coarse-Grained Force Field: Extension to Proteins. *J. Chem. Theory Comput.* **2008**, *4* (5), 819–834. https://doi.org/10.1021/ct700324x.
(239) Mansbach, R. A.; Ferguson, A. L. Patchy Particle Model of the Hierarchical Self-Assembly of π-Conjugated Optoelectronic Peptides. *J. Phys. Chem. B* **2018**, *122* (44), 10219–10236. https://doi.org/10.1021/acs.jpcb.8b05781.
(240) Guo, C.; Luo, Y.; Zhou, R.; Wei, G. Probing the Self-Assembly Mechanism of Diphenylalanine-Based Peptide Nanovesicles and Nanotubes. *ACS Nano* **2012**, *6* (5), 3907–3918. https://doi.org/10.1021/nn300015g.
(241) Seo, M.; Rauscher, S.; Pomès, R.; Tieleman, D. P. Improving Internal Peptide Dynamics in the Coarse-Grained MARTINI Model: Toward Large-Scale Simulations of Amyloid- and Elastin-like Peptides. *J. Chem. Theory Comput.* **2012**, *8* (5), 1774–1785. https://doi.org/10.1021/ct200876v.

(242) Husic, B. E.; Pande, V. S. Markov State Models: From an Art to a Science. *J. Am. Chem. Soc.* **2018**, *140* (7), 2386–2396. https://doi.org/10.1021/jacs.7b12191.
(243) Bonomi, M.; Heller, G. T.; Camilloni, C.; Vendruscolo, M. Principles of Protein Structural Ensemble Determination. *Curr. Opin. Struct. Biol.* **2017**, *42*, 106–116. https://doi.org/10.1016/j.sbi.2016.12.004.
(244) Löhr, T.; Jussupow, A.; Camilloni, C. Metadynamic Metainference: Convergence towards Force Field Independent Structural Ensembles of a Disordered Peptide. *J. Chem. Phys.* **2017**, *146* (16), 165102. https://doi.org/10.1063/1.4981211.
(245) Bonomi, M.; Camilloni, C.; Vendruscolo, M. Metadynamic Metainference: Enhanced Sampling of the Metainference Ensemble Using Metadynamics. *Sci. Rep.* **2016**, *6* (1), 31232. https://doi.org/10.1038/srep31232.
(246) Zhou, W.; Moguche, A. O.; Chiu, D.; Murali-krishna, K.; Baneyx, F. Just-in-Time Vaccines : Biomineralized Calcium Phosphate Core-Immunogen Shell Nanoparticles Induce Long-Lasting CD8+ T Cell Responses in Mice. *Nanomedicine Nanotechnology, Biol. Med.* **2014**, *10*, 571–578. https://doi.org/10.1016/j.nano.2013.11.007.
(247) Park, S.; Lee, H.; Lee, S. Y. Effect of Peptide Conformation on TiO2biomineralization. *Dalt. Trans.* **2013**, *42* (38), 13817–13820. https://doi.org/10.1039/c3dt51040a.
(248) 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. https://doi.org/10.1002/psc.2577.
(249) Roehrich, A.; Drobny, G. P. Solid-State NMR Studies of Biomineralization Peptides and Proteins. *Acc. Chem. Res.* **2013**, *46* (9), 2136–2144. https://doi.org/10.1021/bk-2012-1120.ch004.
(250) Sano, K.; Sasaki, H.; Shiba, K. Specificity and Biomineralization Activities of Ti-Binding Peptide-1 (TBP-1). *Langmuir* **2005**, *21*, 3090–3095.
(251) Sumper, M. A Phase Separation Model for the Nanopatterning of Diatom Biosilica. *Science (80-.).* **2002**, *295* (5564), 2430–2433. https://doi.org/10.1126/science.1070026.
(252) Chien, Y. C.; Hincke, M. T.; Vali, H.; McKee, M. D. Ultrastructural Matrix-Mineral Relationships in Avian Eggshell, and Effects of Osteopontin on Calcite Growth in Vitro. *J. Struct. Biol.* **2008**, *163* (1), 84–99. https://doi.org/10.1016/j.jsb.2008.04.008.
(253) Polowczyk, I.; Bastrzyk, A.; Fiedot, M. Protein-Mediated Precipitation of Calcium Carbonate. *Materials (Basel).* **2016**, *9* (11), 1–16. https://doi.org/10.3390/ma9110944.
(254) 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.
(255) Yang, X.; Wang, L.; Qin, Y.; Sun, Z.; Henneman, Z. J.; Moradian-Oldak, J.; Nancollas, G. H. How Amelogenin Orchestrates the Organization of Hierarchical Elongated Microstructures of Apatite. *J. Phys. Chem. B* **2010**, *114* (6), 2293–2300. https://doi.org/10.1021/jp910219s.
(256) Kawasaki, K.; Buchanan, A. V.; Weiss, K. M. Biomineralization in Humans: Making the Hard Choices in Life. *Annu. Rev. Genet.* **2009**, *43* (1), 119–142. https://doi.org/10.1146/annurev-genet-102108-134242.
(257) Skelton, A. A.; Liang, T.; Walsh, T. R. Interplay of Sequence, Conformation, and Binding at the Peptide-Titania Interface as Mediated by Water. *ACS Appl. Mater. Interfaces* **2009**, *1* (7), 1482–1491. https://doi.org/10.1021/am9001666.
(258) Limo, M. J.; Sola-Rabada, A.; Boix, E.; Thota, V.; Westcott, Z. C.; Puddu, V.; Perry, C. C.

Interactions between Metal Oxides and Biomolecules: From Fundamental Understanding to Applications. *Chem. Rev.* **2018**, *118* (22), 11118–11193. https://doi.org/10.1021/acs.chemrev.7b00660.

(259) Alamdari, S.; Pfaendtner, J. Impact of Glutamate Carboxylation in the Adsorption of the α-1 Domain of Osteocalcin to Hydroxyapatite and Titania. *Mol. Syst. Des. Eng.* **2020**, *5* (3), 620–631. https://doi.org/10.1039/C9ME00158A.

(260) 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. https://doi.org/10.1016/j.str.2010.09.013.

(261) Buckle, E. L.; Roehrich, A.; Vandermoon, B.; Drobny, G. P. Comparative Study of Secondary Structure and Interactions of the R5 Peptide in Silicon Oxide and Titanium Oxide Coprecipitates Using Solid-State NMR Spectroscopy. *Langmuir* **2017**, *33*, 10517–10524. https://doi.org/10.1021/acs.langmuir.7b01048.

(262) Zane, A. C.; Michelet, C.; Roehrich, A.; Emani, P. S.; Drobny, G. P. Silica Morphogenesis by Lysine-Leucine Peptides with Hydrophobic Periodicity. *Langmuir* **2014**, *30*, 7152–7161. https://doi.org/10.1021/la501444t.

(263) Gray, J. J. The Interaction of Proteins with Solid Surfaces. *Curr. Opin. Struct. Biol.* **2004**, *14* (1), 110–115. https://doi.org/10.1016/j.sbi.2003.12.001.

(264) 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. *PNAS* **2006**, *103*, 16083–16088.

(265) Breen, N. F.; Weidner, T.; Li, K.; Castner, D. G.; Drobny, G. P. A Solid-State Deuterium NMR and Sum-Frequency Generation Study of the Side-Chain Dynamics of Peptides Adsorbed onto Surfaces. *J. Am. Chem. Soc.* **2009**, *131*, 14148–14149.

(266) 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. https://doi.org/10.1021/ja903772t.

(267) 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. https://doi.org/10.1021/ja074602v.

(268) Pacella, M. S.; Koo, D. C. E.; Thottungal, R. A.; Gray, J. J. *Using the Rosetta Surface Algorithm to Predict Protein Structure at Mineral Surfaces*, 1st ed.; Elsevier Inc., 2013; Vol. 532. https://doi.org/10.1016/B978-0-12-416617-2.00016-3.

(269) Prakash, A.; Sprenger, K. G.; Pfaendtner, J. Essential Slow Degrees of Freedom in Protein-Surface Simulations: A Metadynamics Investigation. *Biochem. Biophys. Res. Commun.* **2018**, *498* (2), 274–281. https://doi.org/10.1016/j.bbrc.2017.07.066.

(270) Sampath, J.; Pfaendtner, J. Amphiphilic Peptide Binding on Crystalline vs. Amorphous Silica from Molecular Dynamics Simulations. *Mol. Phys.* **2019**, *117* (23–24), 3642–3650. https://doi.org/10.1080/00268976.2019.1657192.

(271) 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. https://doi.org/10.1021/ja211307u.

(272) 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. *PNAS* **2010**, *107*, 13288–13293. https://doi.org/10.1073/pnas.1003832107.
(273) Baio, J. E.; Zane, A.; Jaeger, V.; Roehrich, A. M.; Lutz, H.; Pfaendtner, J.; Drobny, G. P.; Weidner, T. Diatom Mimics: Directing the Formation of Biosilica Nanoparticles by Controlled Folding of Lysine-Leucine Peptides. *J. Am. Chem. Soc.* **2014**, *136*, 15134–15137. https://doi.org/10.1021/ja5078238.
(274) Rabe, M.; Verdes, D.; Seeger, S. Understanding Protein Adsorption Phenomena at Solid Surfaces. *Adv. Colloid Interface Sci.* **2011**, *162* (1–2), 87–106. https://doi.org/10.1016/j.cis.2010.12.007.
(275) Hellner, B.; Alamdari, S.; Pyles, H.; Zhang, S.; Prakash, A.; Sprenger, K. G.; De Yoreo, J. J.; Baker, D.; Pfaendtner, J.; Baneyx, F. Sequence-Structure-Binding Relationships Reveal Adhesion Behavior of the Car9 Solid-Binding Peptide: An Integrated Experimental and Simulation Study. *J. Am. Chem. Soc.* **2020**, *142* (5), 2355–2363. https://doi.org/10.1021/jacs.9b11617.
(276) Sarikaya, M.; Tamerler, C.; Jen, A. K. Y.; Schulten, K.; Baneyx, F. Molecular Biomimetics: Nanotechnology through Biology. *Nat. Mater.* **2003**, *2* (9), 577–585. https://doi.org/10.1038/nmat964.
(277) Whaley, S. R.; English, D. S.; Hu, E. L.; Barbara, P. F.; Belcher, A. M. Selection of Peptides with Semiconductor Binding Specificity for Directed Nanocrystal Assembly. *Nature* **2000**, *405*, 665–668.
(278) Care, A.; Bergquist, P. L.; Sunna, A. Solid-Binding Peptides: Smart Tools for Nanobiotechnology. *Trends Biotechnol.* **2015**, *33* (5), 259–268. https://doi.org/10.1016/j.tibtech.2015.02.005.
(279) 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–TiO 2 Interfaces: Isoenergetic Binding via Either Entropically or Enthalpically Driven Mechanisms. *ACS Appl. Mater. Interfaces* **2016**, *8* (28), 18620–18630. https://doi.org/10.1021/acsami.6b05200.
(280) Sampath, J.; Kullman, A.; Gebhart, R.; Drobny, G.; Pfaendtner, J. Molecular Recognition and Specificity of Biomolecules to Titanium Dioxide from Molecular Dynamics Simulations. *npj Comput. Mater.* **2020**, *6* (1), 34. https://doi.org/10.1038/s41524-020-0288-7.
(281) Rabe, M.; Verdes, D.; Seeger, S. Understanding Cooperative Protein Adsorption Events at the Microscopic Scale: A Comparison between Experimental Data and Monte Carlo Simulations. *J. Phys. Chem. B* **2010**, *114* (17), 5862–5869. https://doi.org/10.1021/jp909601m.
(282) Tang, Z.; Palafox-Hernandez, J. P.; Law, W. C.; Hughes, Z. E.; Swihart, M. T.; Prasad, P. N.; Knecht, M. R.; Walsh, T. R. Biomolecular Recognition Principles for Bionanocombinatorics: An Integrated Approach to Elucidate Enthalpic and Entropic Factors. *ACS Nano* **2013**, *7* (11), 9632–9646. https://doi.org/10.1021/nn404427y.
(283) Rosales, A. M.; Segalman, R. A.; Zuckermann, R. N. Polypeptoids: A Model System to Study the Effect of Monomer Sequence on Polymer Properties and Self-Assembly. *Soft Matter* **2013**, *9* (35), 8400–8414. https://doi.org/10.1039/c3sm51421h.
(284) Greer, D. R.; Stolberg, M. A.; Kundu, J.; Spencer, R. K.; Pascal, T.; Prendergast, D.; Balsara, N. P.; Zuckermann, R. N. Universal Relationship between Molecular Structure and

Crystal Structure in Peptoid Polymers and Prevalence of the Cis Backbone Conformation. *J. Am. Chem. Soc.* **2018**, *140* (2), 827–833. https://doi.org/10.1021/jacs.7b11891.

(285) Sun, J.; Teran, A. A.; Liao, X.; Balsara, N. P.; Zuckermann, R. N. Nanoscale Phase Separation in Sequence-Defined Peptoid Diblock Copolymers. *J. Am. Chem. Soc.* **2013**, *135* (38), 14119–14124. https://doi.org/10.1021/ja404233d.

(286) Zuckermann, R. N.; Kerr, J. M.; Moosf, W. H.; Kent, S. B. H. Efficient Method for the Preparation of Peptoids [Oligo(N-Substituted Glycines)] by Submonomer Solid-Phase Synthesis. *J. Am. Chem. Soc.* **1992**, *114* (26), 10646–10647. https://doi.org/10.1021/ja00052a076.

(287) 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. https://doi.org/10.1021/acs.jpcb.5b12277.

(288) 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. https://doi.org/10.1021/ja206199d.

(289) 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* (1), 12252. https://doi.org/10.1038/ncomms12252.

(290) Jin, H.; Ding, Y.-H.; Wang, M.; Song, Y.; Liao, Z.; Newcomb, C. J.; Wu, X.; Tang, X.-Q.; Li, Z.; Lin, Y.; et al. Designable and Dynamic Single-Walled Stiff Nanotubes Assembled from Sequence-Defined Peptoids. *Nat. Commun.* **2018**, *9* (1), 270. https://doi.org/10.1038/s41467-017-02059-1.

(291) Sun, J.; Jiang, X.; Lund, R.; Downing, K. H.; Balsara, N. P.; Zuckermann, R. N. Self-Assembly of Crystalline Nanotubes from Monodisperse Amphiphilic Diblock Copolypeptoid Tiles. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113* (15), 3954–3959. https://doi.org/10.1073/pnas.1517169113.

(292) 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. https://doi.org/10.1038/nature15363.

(293) 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. https://doi.org/10.1002/jcc.23478.

(294) Edison, J. R.; Spencer, R. K.; Butterfoss, G. L.; Hudson, B. C.; Hochbaum, A. I.; Paravastu, A. K.; Zuckermann, R. N.; Whitelam, S. Conformations of Peptoids in Nanosheets Result from the Interplay of Backbone Energetics and Intermolecular Interactions. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115* (22), 5647–5651. https://doi.org/10.1073/pnas.1800397115.

(295) Mansbach, R. A.; Ferguson, A. L. Coarse-Grained Molecular Simulation of the Hierarchical Self-Assembly of π-Conjugated Optoelectronic Peptides. *J. Phys. Chem. B* **2017**, *121* (7), 1684–1706. https://doi.org/10.1021/acs.jpcb.6b10165.

(296) Huang, P. S.; Boyken, S. E.; Baker, D. The Coming of Age of de Novo Protein Design. *Nature* **2016**, *537* (7620), 320–327. https://doi.org/10.1038/nature19946.

(297) Leader, B.; Baca, Q. J.; Golan, D. E. Protein Therapeutics: A Summary and Pharmacological Classification. *Nat. Rev. Drug Discov.* **2008**, *7* (1), 21–39. https://doi.org/10.1038/nrd2399.

(298) Pyles, H.; Zhang, S.; De Yoreo, J. J.; Baker, D. Controlling Protein Assembly on Inorganic Crystals through Designed Protein Interfaces. *Nature* **2019**, *571* (7764), 251–256. https://doi.org/10.1038/s41586-019-1361-6.
(299) Gerasopoulos, K.; Pomerantseva, E.; McCarthy, M.; Brown, A.; Wang, C.; Culver, J.; Ghodssi, R. Hierarchical Three-Dimensional Microbattery Electrodes Combining Bottom-up Self-Assembly and Top-down Micromachining. *ACS Nano* **2012**, *6* (7), 6422–6432. https://doi.org/10.1021/nn301981p.
(300) Bruckman, M. A.; Vanmeter, A.; Steinmetz, N. F. Nanomanufacturing of Tobacco Mosaic Virus-Based Spherical Biomaterials Using a Continuous Flow Method. *ACS Biomater. Sci. Eng.* **2015**, *1* (1), 13–18. https://doi.org/10.1021/ab500059s.
(301) Banik, S.; Mansour, A. A.; Suresh, R. V.; Wykoff-Clary, S.; Malik, M.; McCormick, A. A.; Bakshi, C. S. Development of a Multivalent Subunit Vaccine against Tularemia Using Tobacco Mosaic Virus (TMV) Based Delivery System. *PLoS One* **2015**, *10* (6), 1–22. https://doi.org/10.1371/journal.pone.0130858.
(302) Perlmutter, J. D.; Perkett, M. R.; Hagan, M. F. Pathways for Virus Assembly around Nucleic Acids. *J. Mol. Biol.* **2014**, *426* (18), 3148–3165. https://doi.org/10.1016/j.jmb.2014.07.004.
(303) Mustata, G. M.; Kim, Y. H.; Zhang, J.; Degrado, W. F.; Grigoryan, G.; Wanunu, M. Graphene Symmetry Amplified by Designed Peptide Self-Assembly. *Biophys. J.* **2016**, *110* (11), 2507–2516. https://doi.org/10.1016/j.bpj.2016.04.037.
(304) Alberstein, R.; Suzuki, Y.; Paesani, F.; Tezcan, F. A. Engineering the Entropy-Driven Free-Energy Landscape of a Dynamic Nanoporous Protein Assembly. *Nat. Chem.* **2018**, *10* (7), 732–739. https://doi.org/10.1038/s41557-018-0053-4.
(305) Elrad, O.; Hagan, M. F. Encapsulation of a Polymer by an Icosahedral Virus. *Phys. Biol.* **2010**, *7*, 1–17. https://doi.org/10.1038/jid.2014.371.
(306) Bar Dolev, M.; Braslavsky, I.; Davies, P. L. Ice-Binding Proteins and Their Function. *Annu. Rev. Biochem.* **2016**, *85* (1), 515–542. https://doi.org/10.1146/annurev-biochem-060815-014546.
(307) Suzuki, Y.; Cardone, G.; Restrepo, D.; Zavattieri, P. D.; Baker, T. S.; Tezcan, F. A. Self-Assembly of Coherently Dynamic, Auxetic, Two-Dimensional Protein Crystals. *Nature* **2016**, *533* (7603), 369–373. https://doi.org/10.1038/nature17633.
(308) Ziatdinov, M.; Dyck, O.; Maksov, A.; Li, X.; Sang, X.; Xiao, K.; Unocic, R. R.; Vasudevan, R.; Jesse, S.; Kalinin, S. V. Deep Learning of Atomically Resolved Scanning Transmission Electron Microscopy Images: Chemical Identification and Tracking Local Transformations. *ACS Nano* **2017**, *11* (12), 12742–12752. https://doi.org/10.1021/acsnano.7b07504.
(309) Voelz, V. A.; Jäger, M.; Yao, S.; Chen, Y.; Zhu, L.; Waldauer, S. A.; Bowman, G. R.; Friedrichs, M.; Bakajin, O.; Lapidus, L. J.; et al. Slow Unfolded-State Structuring in Acyl-CoA Binding Protein Folding Revealed by Simulation and Experiment. *J. Am. Chem. Soc.* **2012**, *134* (30), 12565–12577. https://doi.org/10.1021/ja302528z.
(310) Wang, Y.; Lamim Ribeiro, J. M.; Tiwary, P. Machine Learning Approaches for Analyzing and Enhancing Molecular Dynamics Simulations. *Curr. Opin. Struct. Biol.* **2020**, *61*, 139–145. https://doi.org/10.1016/j.sbi.2019.12.016.
(311) Ferguson, A. L. Machine Learning and Data Science in Soft Materials Engineering. *J. Phys. Condens. Matter* **2018**, *30* (4). https://doi.org/10.1088/1361-648X/aa98bd.
(312) Beck, D. A. C.; Carothers, J. M.; Subramanian, V. R.; Pfaendtner, J. Data Science: Accelerating Innovation and Discovery in Chemical Engineering. *AIChE J.* **2016**, *62* (5), 1402–1416. https://doi.org/10.1002/aic.15192.

(313) Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A. Machine Learning for Molecular and Materials Science. *Nature* **2018**, *559* (7715), 547–555. https://doi.org/10.1038/s41586-018-0337-2.
(314) Wernick, M.; Yang, Y.; Brankov, J.; Yourganov, G.; Strother, S. Machine Learning in Medical Imaging. *IEEE Signal Process. Mag.* **2010**, *27* (4), 25–38. https://doi.org/10.1109/MSP.2010.936730.
(315) Lemm, S.; Blankertz, B.; Dickhaus, T.; Müller, K.-R. Introduction to Machine Learning for Brain Imaging. *Neuroimage* **2011**, *56* (2), 387–399. https://doi.org/10.1016/j.neuroimage.2010.11.004.
(316) Kalinin, S. V.; Sumpter, B. G.; Archibald, R. K. Big–Deep–Smart Data in Imaging for Guiding Materials Design. *Nat. Mater.* **2015**, *14* (10), 973–980. https://doi.org/10.1038/nmat4395.
(317) Kalinin, S. V; Foster, I.; Kalidindi, S.; Kalinin, S. V; Lookman, T.; van Dam, K. K.; Yager, K. G.; Campbell, S. I.; Farnsworth, R.; van Dam, M. *Handbook on Big Data and Machine Learning in the Physical Sciences*; World Scientific, 2020; Vol. 1. https://doi.org/10.1142/11389-vol1.
(318) Ziatdinov, M.; Maksov, A.; Kalinin, S. V. Learning Surface Molecular Structures via Machine Vision. *npj Comput. Mater.* **2017**, *3* (1), 31. https://doi.org/10.1038/s41524-017-0038-7.
(319) Zhang, L.; Chen, Z.; Su, J.; Li, J. Data Mining New Energy Materials from Structure Databases. *Renew. Sustain. Energy Rev.* **2019**, *107*, 554–567. https://doi.org/10.1016/j.rser.2019.03.036.
(320) Jankowski, E.; Ellyson, N.; Fothergill, J. W.; Henry, M. M.; Leibowitz, M. H.; Miller, E. D.; Alberts, M.; Chesser, S.; Guevara, J. D.; Jones, C. D.; et al. Perspective on Coarse-Graining, Cognitive Load, and Materials Simulation. *Comput. Mater. Sci.* **2020**, *171*, 109129. https://doi.org/10.1016/j.commatsci.2019.109129.
(321) Marques, M. R. G.; Wolff, J.; Steigemann, C.; Marques, M. A. L. Neural Network Force Fields for Simple Metals and Semiconductors: Construction and Application to the Calculation of Phonons and Melting Temperatures. *Phys. Chem. Chem. Phys.* **2019**, *21* (12), 6506–6516. https://doi.org/10.1039/C8CP05771K.
(322) Behler, J. Perspective: Machine Learning Potentials for Atomistic Simulations. *J. Chem. Phys.* **2016**, *145* (17), 170901. https://doi.org/10.1063/1.4966192.
(323) Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1: An Extensible Neural Network Potential with DFT Accuracy at Force Field Computational Cost. *Chem. Sci.* **2017**, *8* (4), 3192–3203. https://doi.org/10.1039/C6SC05720A.
(324) Zeni, C.; Rossi, K.; Glielmo, A.; Baletto, F. On Machine Learning Force Fields for Metallic Nanoparticles. *Adv. Phys. X* **2019**, *4* (1), 1654919. https://doi.org/10.1080/23746149.2019.1654919.
(325) Zhang, L.; Han, J.; Wang, H.; Car, R.; E, W. DeePCG: Constructing Coarse-Grained Models via Deep Neural Networks. *J. Chem. Phys.* **2018**, *149* (3), 034101. https://doi.org/10.1063/1.5027645.
(326) Wang, J.; Cao, H.; Zhang, J. Z. H.; Qi, Y. Computational Protein Design with Deep Learning Neural Networks. *Sci. Rep.* **2018**, *8* (1), 6349. https://doi.org/10.1038/s41598-018-24760-x.
(327) Pfeiffenberger, E.; Bates, P. A. Predicting Improved Protein Conformations with a Temporal Deep Recurrent Neural Network. *PLoS One* **2018**, *13* (9), e0202652.

https://doi.org/10.1371/journal.pone.0202652.
(328) Artrith, N.; Hiller, B.; Behler, J. Neural Network Potentials for Metals and Oxides - First Applications to Copper Clusters at Zinc Oxide. *Phys. status solidi* **2013**, *250* (6), 1191–1203. https://doi.org/10.1002/pssb.201248370.
(329) Behler, J. Neural Network Potential-Energy Surfaces in Chemistry: A Tool for Large-Scale Simulations. *Phys. Chem. Chem. Phys.* **2011**, *13* (40), 17930. https://doi.org/10.1039/c1cp21668f.
(330) Kocer, E.; Mason, J. K.; Erturk, H. A Novel Approach to Describe Chemical Environments in High-Dimensional Neural Network Potentials. *J. Chem. Phys.* **2019**, *150* (15), 154102. https://doi.org/10.1063/1.5086167.
(331) Zhang, L.; Han, J.; Wang, H.; Car, R.; E, W. Deep Potential Molecular Dynamics: A Scalable Model with the Accuracy of Quantum Mechanics. *Phys. Rev. Lett.* **2018**, *120* (14), 143001. https://doi.org/10.1103/PhysRevLett.120.143001.
(332) Boomsma, W.; Ferkinghoff-Borg, J.; Lindorff-Larsen, K. Combining Experiments and Simulations Using the Maximum Entropy Principle. *PLoS Comput. Biol.* **2014**, *10* (2), e1003406. https://doi.org/10.1371/journal.pcbi.1003406.
(333) Wang, L.-P.; Martinez, T. J.; Pande, V. S. Building Force Fields: An Automatic, Systematic, and Reproducible Approach. *J. Phys. Chem. Lett.* **2014**, *5* (11), 1885–1891. https://doi.org/10.1021/jz500737m.
(334) Chen, J.; Chen, J.; Pinamonti, G.; Clementi, C. Learning Effective Molecular Models from Experimental Observables. *J. Chem. Theory Comput.* **2018**, *14* (7), 3849–3858. https://doi.org/10.1021/acs.jctc.8b00187.
(335) Noé, F.; Olsson, S.; Köhler, J.; Wu, H. Boltzmann Generators: Sampling Equilibrium States of Many-Body Systems with Deep Learning. *Science (80-.).* **2019**, *365* (6457), eaaw1147. https://doi.org/10.1126/science.aaw1147.
(336) Bonati, L.; Zhang, Y.-Y.; Parrinello, M. Neural Networks-Based Variationally Enhanced Sampling. *Proc. Natl. Acad. Sci.* **2019**, *116* (36), 17641–17647. https://doi.org/10.1073/pnas.1907975116.
(337) Wang, J.; Ferguson, A. L. Nonlinear Machine Learning in Simulations of Soft and Biological Materials. *Mol. Simul.* **2018**, *44* (13–14), 1090–1107. https://doi.org/10.1080/08927022.2017.1400164.
(338) Noé, F.; De Fabritiis, G.; Clementi, C. Machine Learning for Protein Folding and Dynamics. *Curr. Opin. Struct. Biol.* **2020**, *60*, 77–84. https://doi.org/10.1016/j.sbi.2019.12.005.
(339) Karpatne, A.; Atluri, G.; Faghmous, J. H.; Steinbach, M.; Banerjee, A.; Ganguly, A.; Shekhar, S.; Samatova, N.; Kumar, V. Theory-Guided Data Science: A New Paradigm for Scientific Discovery from Data. *IEEE Trans. Knowl. Data Eng.* **2017**, *29* (10), 2318–2331. https://doi.org/10.1109/TKDE.2017.2720168.
(340) Shin, K.; Tran, D. P.; Takemura, K.; Kitao, A.; Terayama, K.; Tsuda, K. Enhancing Biomolecular Sampling with Reinforcement Learning: A Tree Search Molecular Dynamics Simulation Method. *ACS Omega* **2019**, *4* (9), 13853–13862. https://doi.org/10.1021/acsomega.9b01480.
(341) Le, T. C.; Tran, N. Using Machine Learning To Predict the Self-Assembled Nanostructures of Monoolein and Phytantriol as a Function of Temperature and Fatty Acid Additives for Effective Lipid-Based Delivery Systems. *ACS Appl. Nano Mater.* **2019**, *2* (3), 1637–1647. https://doi.org/10.1021/acsanm.9b00075.
(342) Li, F.; Han, J.; Cao, T.; Lam, W.; Fan, B.; Tang, W.; Chen, S.; Fok, K. L.; Li, L. Design of

Self-Assembly Dipeptide Hydrogels and Machine Learning via Their Chemical Features. *Proc. Natl. Acad. Sci.* **2019**, *116* (23), 11259–11264. https://doi.org/10.1073/pnas.1903376116.

(343) Shamay, Y.; Shah, J.; Işık, M.; Mizrachi, A.; Leibold, J.; Tschaharganeh, D. F.; Roxbury, D.; Budhathoki-Uprety, J.; Nawaly, K.; Sugarman, J. L.; et al. Quantitative Self-Assembly Prediction Yields Targeted Nanomedicines. *Nat. Mater.* **2018**, *17* (4), 361–368. https://doi.org/10.1038/s41563-017-0007-z.

(344) Marinelli, F.; Pietrucci, F.; Laio, A.; Piana, S. A Kinetic Model of Trp-Cage Folding from Multiple Biased Molecular Dynamics Simulations. *PLoS Comput. Biol.* **2009**, *5* (8), e1000452. https://doi.org/10.1371/journal.pcbi.1000452.

(345) van Gunsteren, W. F.; Bakowies, D.; Baron, R.; Chandrasekhar, I.; Christen, M.; Daura, X.; Gee, P.; Geerke, D. P.; Glättli, A.; Hünenberger, P. H.; et al. Biomolecular Modeling: Goals, Problems, Perspectives. *Angew. Chemie Int. Ed.* **2006**, *45* (25), 4064–4092. https://doi.org/10.1002/anie.200502655.

(346) 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. https://doi.org/10.1016/j.bbagen.2014.10.019.

(347) Incerti, M.; Russo, S.; Callegari, D.; Pala, D.; Giorgio, C.; Zanotti, I.; Barocelli, E.; Vicini, P.; Vacondio, F.; Rivara, S.; et al. Metadynamics for Perspective Drug Design: Computationally Driven Synthesis of New Protein–Protein Interaction Inhibitors Targeting the EphA2 Receptor. *J. Med. Chem.* **2017**, *60* (2), 787–796. https://doi.org/10.1021/acs.jmedchem.6b01642.

(348) Clark, A. J.; Tiwary, P.; Borrelli, K.; Feng, S.; Miller, E. B.; Abel, R.; Friesner, R. A.; Berne, B. J. Prediction of Protein–Ligand Binding Poses via a Combination of Induced Fit Docking and Metadynamics Simulations. *J. Chem. Theory Comput.* **2016**, *12* (6), 2990–2998. https://doi.org/10.1021/acs.jctc.6b00201.

(349) Amaro, R. E.; Baudry, J.; Chodera, J.; Demir, Ö.; McCammon, J. A.; Miao, Y.; Smith, J. C. Ensemble Docking in Drug Discovery. *Biophys. J.* **2018**, *114* (10), 2271–2278. https://doi.org/10.1016/j.bpj.2018.02.038.

(350) Invernizzi, M.; Valsson, O.; Parrinello, M. Coarse Graining from Variationally Enhanced Sampling Applied to the Ginzburg–Landau Model. *Proc. Natl. Acad. Sci.* **2017**, *114* (13), 3370–3374. https://doi.org/10.1073/pnas.1618455114.

(351) Fiore, C. E.; da Luz, M. G. E. Comparing Parallel- and Simulated-Tempering-Enhanced Sampling Algorithms at Phase-Transition Regimes. *Phys. Rev. E* **2010**, *82* (3), 031104. https://doi.org/10.1103/PhysRevE.82.031104.

(352) Sosso, G. C.; Chen, J.; Cox, S. J.; Fitzner, M.; Pedevilla, P.; Zen, A.; Michaelides, A. Crystal Nucleation in Liquids: Open Questions and Future Challenges in Molecular Dynamics Simulations. *Chem. Rev.* **2016**, *116* (12), 7078–7116. https://doi.org/10.1021/acs.chemrev.5b00744.

(353) Giberti, F.; Salvalaglio, M.; Parrinello, M. Metadynamics Studies of Crystal Nucleation. *IUCrJ* **2015**, *2* (2), 256–266. https://doi.org/10.1107/S2052252514027626.

(354) Mandal, S.; Debnath, J.; Meyer, B.; Nair, N. N. Enhanced Sampling and Free Energy Calculations with Hybrid Functionals and Plane Waves for Chemical Reactions. *J. Chem. Phys.* **2018**, *149* (14), 144113. https://doi.org/10.1063/1.5049700.

(355) Debnath, J.; Invernizzi, M.; Parrinello, M. Enhanced Sampling of Transition States. *J. Chem. Theory Comput.* **2019**, acs.jctc.8b01283. https://doi.org/10.1021/acs.jctc.8b01283.

(356) Awasthi, S.; Nair, N. N. Exploring High Dimensional Free Energy Landscapes: Temperature Accelerated Sliced Sampling. *J. Chem. Phys.* **2017**, *146* (9). https://doi.org/10.1063/1.4977704.
(357) Miroliaei, M.; Nemat-Gorgani, M. Effect of Organic Solvents on Stability and Activity of Two Related Alcohol Dehydrogenases: A Comparative Study. *Int. J. Biochem. Cell Biol.* **2002**, *34*, 169–175. https://doi.org/10.1016/S1357-2725(01)00109-1.
(358) Wu, D.; Fajer, M. I.; Cao, L.; Cheng, X.; Yang, W. Generalized Ensemble Sampling of Enzyme Reaction Free Energy Pathways. *Methods Enzymol.* **2016**, *577*, 57–74. https://doi.org/10.1016/bs.mie.2016.05.012.
(359) Bolhuis, P. G.; Chandler, D.; Dellago, C.; Geissler, P. L. TRANSITION PATH SAMPLING: Throwing Ropes Over Rough Mountain Passes, in the Dark. *Annu. Rev. Phys. Chem.* **2002**, *53* (1), 291–318. https://doi.org/10.1146/annurev.physchem.53.082301.113146.
(360) Nakamura, M.; Obata, M.; Morishita, T.; Oda, T. An Ab Initio Approach to Free-Energy Reconstruction Using Logarithmic Mean Force Dynamics. *J. Chem. Phys.* **2014**, *140* (18). https://doi.org/10.1063/1.4874654.
(361) Moritsugu, K.; Terada, T.; Kidera, A. Scalable Free Energy Calculation of Proteins via Multiscale Essential Sampling. *J. Chem. Phys.* **2010**, *133* (22), 224105. https://doi.org/10.1063/1.3510519.
(362) Darve, E.; Pohorille, A. Calculating Free Energies Using Average Force. *J. Chem. Phys.* **2001**, *115* (20), 9169–9183. https://doi.org/10.1063/1.1410978.
(363) Henkelman, G.; Jónsson, H. Improved Tangent Estimate in the Nudged Elastic Band Method for Finding Minimum Energy Paths and Saddle Points. *J. Chem. Phys.* **2000**, *113* (22), 9978–9985. https://doi.org/10.1063/1.1323224.
(364) Fujisaki, H.; Shiga, M.; Kidera, A. Onsager–Machlup Action-Based Path Sampling and Its Combination with Replica Exchange for Diffusive and Multiple Pathways. *J. Chem. Phys.* **2010**, *132* (13), 134101. https://doi.org/10.1063/1.3372802.
(365) Glowacki, D. R.; Paci, E.; Shalashilin, D. V. Boxed Molecular Dynamics: A Simple and General Technique for Accelerating Rare Event Kinetics and Mapping Free Energy in Large Molecular Systems. *J. Phys. Chem. B* **2009**, *113* (52), 16603–16611. https://doi.org/10.1021/jp9074898.
(366) Hansmann, U. H. E.; Wille, L. T. Global Optimization by Energy Landscape Paving. *Phys. Rev. Lett.* **2002**, *88* (6), 068105. https://doi.org/10.1103/PhysRevLett.88.068105.
(367) E, W.; Ren, W.; Vanden-Eijnden, E. String Method for the Study of Rare Events. *Phys. Rev. B* **2002**, *66* (5), 052301. https://doi.org/10.1103/PhysRevB.66.052301.
(368) Dellago, C.; Bolhuis, P. G. Transition Path Sampling and Other Advanced Simulation Techniques for Rare Events. In *Advanced Computer Simulation Approaches for Soft Matter Sciences III*; Springer Berlin Heidelberg: Berlin, Heidelberg; pp 167–233. https://doi.org/10.1007/978-3-540-87706-6_3.
(369) Fujisaki, H.; Moritsugu, K.; Matsunaga, Y.; Morishita, T.; Maragliano, L. Extended Phase-Space Methods for Enhanced Sampling in Molecular Simulations: A Review. *Front. Bioeng. Biotechnol.* **2015**, *3* (September), 1–10. https://doi.org/10.3389/fbioe.2015.00125.
(370) Laio, A.; Parrinello, M. Computing Free Energies and Accelerating Rare Events with Metadynamics. In *Computer Simulations in Condensed Matter Systems: From Materials to Chemical Biology Volume 1*; Springer Berlin Heidelberg: Berlin, Heidelberg; pp 315–347. https://doi.org/10.1007/3-540-35273-2_9.

(371) Rodriguez-Gomez, D.; Darve, E.; Pohorille, A. Assessing the Efficiency of Free Energy Calculation Methods. *J. Chem. Phys.* **2004**, *120* (8), 3563–3578. https://doi.org/10.1063/1.1642607.
(372) Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P. Optimization by Simulated Annealing. *Science (80-.).* **1983**, *220* (4598), 671–680. https://doi.org/10.1126/science.220.4598.671.
(373) Hansmann, U. H. E. Parallel Tempering Algorithm for Conformational Studies of Biological Molecules. *Chem. Phys. Lett.* **1997**, *281* (1–3), 140–150. https://doi.org/10.1016/S0009-2614(97)01198-6.
(374) Sugita, Y.; Okamoto, Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.* **1999**. https://doi.org/10.1016/S0009-2614(99)01123-9.
(375) Affentranger, R.; Tavernelli, I.; Di Iorio, E. E. A Novel Hamiltonian Replica Exchange MD Protocol to Enhance Protein Conformational Space Sampling. *J. Chem. Theory Comput.* **2006**, *2* (2), 217–228. https://doi.org/10.1021/ct050250b.
(376) Liu, P.; Kim, B.; Friesner, R. A.; Berne, B. J. Replica Exchange with Solute Tempering: A Method for Sampling Biological Systems in Explicit Water. *Proc. Natl. Acad. Sci.* **2005**, *102* (39), 13749–13754. https://doi.org/10.1073/pnas.0506346102.
(377) Rosso, L.; Mináry, P.; Zhu, Z.; Tuckerman, M. E. On the Use of the Adiabatic Molecular Dynamics Technique in the Calculation of Free Energy Profiles. *J. Chem. Phys.* **2002**, *116* (11), 4389–4402. https://doi.org/10.1063/1.1448491.
(378) Maragliano, L.; Vanden-Eijnden, E. A Temperature Accelerated Method for Sampling Free Energy and Determining Reaction Pathways in Rare Events Simulations. *Chem. Phys. Lett.* **2006**, *426* (1–3), 168–175. https://doi.org/10.1016/j.cplett.2006.05.062.
(379) Meloni, S.; Ciccotti, G. Free Energies for Rare Events: Temperature Accelerated MD and MC. *Eur. Phys. J. Spec. Top.* **2015**, *224* (12), 2389–2407. https://doi.org/10.1140/epjst/e2015-02418-7.
(380) Luitz, M.; Bomblies, R.; Ostermeir, K.; Zacharias, M. Exploring Biomolecular Dynamics and Interactions Using Advanced Sampling Methods. *J. Phys. Condens. Matter* **2015**, *27* (32), 323101. https://doi.org/10.1088/0953-8984/27/32/323101.
(381) Sega, M.; Autieri, E.; Pederiva, F. Pickett Angles and Cremer–Pople Coordinates as Collective Variables for the Enhanced Sampling of Six-Membered Ring Conformations. *Mol. Phys.* **2011**, *109* (1), 141–148. https://doi.org/10.1080/00268976.2010.522208.
(382) Peters, B. Reaction Coordinates and Mechanistic Hypothesis Tests. *Annu. Rev. Phys. Chem.* **2016**, *67* (1), 669–690. https://doi.org/10.1146/annurev-physchem-040215-112215.
(383) 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. https://doi.org/10.1016/0021-9991(77)90121-8.
(384) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. THE Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method. *J. Comput. Chem.* **1992**, *13* (8), 1011–1021. https://doi.org/10.1002/jcc.540130812.
(385) Grossfield, A. WHAM: the weighted histogram analysis method.
(386) Shirts, M. R.; Chodera, J. D. Statistically Optimal Analysis of Samples from Multiple Equilibrium States. *J. Chem. Phys.* **2008**, *129* (12), 124105. https://doi.org/10.1063/1.2978177.
(387) Schlitter, J.; Engels, M.; Krüger, P.; Jacoby, E.; Wollmer, A. Targeted Molecular Dynamics Simulation of Conformational Change-Application to the T ↔ R Transition in Insulin. *Mol. Simul.* **1993**, *10* (2–6), 291–308. https://doi.org/10.1080/08927029308022170.

(388) Wang, F.; Landau, D. P. Efficient, Multiple-Range Random Walk Algorithm to Calculate the Density of States. *Phys. Rev. Lett.* **2001**, *86* (10), 2050–2053. https://doi.org/10.1103/PhysRevLett.86.2050.
(389) Valsson, O.; Parrinello, M. Variational Approach to Enhanced Sampling and Free Energy Calculations. *Phys. Rev. Lett.* **2014**, *113* (9), 1–5. https://doi.org/10.1103/PhysRevLett.113.090601.
(390) Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *PNAS* **2002**, *99*, 12562.
(391) Micheletti, C.; Laio, A.; Parrinello, M. Reconstructing the Density of States by History-Dependent Metadynamics. *Phys. Rev. Lett.* **2004**, *92* (17), 170601. https://doi.org/10.1103/PhysRevLett.92.170601.
(392) Vargiu, A. V.; Ruggerone, P.; Magistrato, A.; Carloni, P. Dissociation of Minor Groove Binders from DNA: Insights from Metadynamics Simulations. *Nucleic Acids Res.* **2008**, *36* (18), 5910–5921. https://doi.org/10.1093/nar/gkn561.
(393) Ensing, B.; De Vivo, M.; Liu, Z.; Moore, P.; Klein, M. L. Metadynamics as a Tool for Exploring Free Energy Landscapes of Chemical Reactions. *Acc. Chem. Res.* **2006**, *39*, 73. https://doi.org/10.1021/ar040198i.
(394) Gervasio, F. L.; Laio, A.; Parrinello, M. Flexible Docking in Solution Using Metadynamics. *J. Am. Chem. Soc.* **2005**, *127*, 2600. https://doi.org/10.1021/ja0445950.
(395) Limongelli, V.; Bonomi, M.; Parrinello, M. Funnel Metadynamics as Accurate Binding Free-Energy Method. *Proc. Natl. Acad. Sci.* **2013**. https://doi.org/10.1073/pnas.1303186110.
(396) Martoňák, R.; Laio, A.; Bernasconi, M.; Ceriani, C.; Raiteri, P.; Parrinello, M. Simulation of Structural Phase Transitions by Metadynamics. *Zeitschrift für Krist. Mater.* **2005**, *220*, 489.
(397) Laio, A.; Rodriguez-Fortea, A.; Gervasio, F. L.; Ceccarelli, M.; Parrinello, M. Assessing the Accuracy of Metadynamics. *J. Phys. Chem. B* **2005**, *109*, 6714. https://doi.org/10.1021/jp045424k.
(398) Dama, J. F.; Parrinello, M.; Voth, G. A. Well-Tempered Metadynamics Converges Asymptotically. *Phys. Rev. Lett.* **2014**, *112* (24). https://doi.org/10.1103/PhysRevLett.112.240602.
(399) Singh, S.; Chiu, C.-C.; De Pablo, J. J. Flux Tempered Metadynamics. *J Stat Phys* **2011**, *145*, 932–945. https://doi.org/10.1007/s10955-011-0301-0.
(400) Bonomi, M.; Camilloni, C.; Vendruscolo, M. Metadynamic Metainference: Enhanced Sampling of the Metainference Ensemble Using Metadynamics. *Sci. Rep.* **2016**, *6* (1), 31232. https://doi.org/10.1038/srep31232.
(401) Branduardi, D.; Bussi, G.; Parrinello, M. Metadynamics with Adaptive Gaussians. *J. Chem. Theory Comput.* **2012**, *8* (7), 2247–2254. https://doi.org/10.1021/ct3002464.
(402) Dama, J. F.; Rotskoff, G.; Parrinello, M.; Voth, G. A. Transition-Tempered Metadynamics: Robust, Convergent Metadynamics via On-the-Fly Transition Barrier Estimation. *J. Chem. Theory Comput.* **2014**, *10* (9), 3626–3633. https://doi.org/10.1021/ct500441q.
(403) Oliver, W. C.; Pharr, G. M. An Improved Technique for Determining Hardness and Elastic Modulus Using Load and Displacement Sensing Indentation Experiments. *Journal of Materials Research*. 1992, pp 1564–1583.
(404) Min, D.; Liu, Y.; Carbone, I.; Yang, W. On the Convergence Improvement in the Metadynamics Simulations: A Wang-Landau Recursion Approach. *J. Chem. Phys.* **2007**, *126* (19), 194104. https://doi.org/10.1063/1.2731769.

(405) 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*, 3533. https://doi.org/10.1021/jp054359r.
(406) Singh, S.; Chopra, M.; de Pablo, J. J. Density of States–Based Molecular Simulations. *Annu. Rev. Chem. Biomol. Eng.* **2012**, *3* (1), 369–394. https://doi.org/10.1146/annurev-chembioeng-062011-081032.
(407) 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. https://doi.org/10.1146/annurev-physchem-040215-112229.
(408) Zhang, Y.; Voth, G. A. Combined Metadynamics and Umbrella Sampling Method for the Calculation of Ion Permeation Free Energy Profiles. *J. Chem. Theory Comput.* **2011**, *7* (7), 2277–2283. https://doi.org/10.1021/ct200100e.
(409) White, A. D.; Dama, J. F.; Voth, G. A. Designing Free Energy Surfaces That Match Experimental Data with Metadynamics. *J. Chem. Theory Comput.* **2015**, *11* (6), 2451–2460. https://doi.org/10.1021/acs.jctc.5b00178.
(410) Gil-Ley, A.; Bussi, G. Enhanced Conformational Sampling Using Replica Exchange with Collective-Variable Tempering. *J. Chem. Theory Comput.* **2015**, *11*, 1077. https://doi.org/10.1021/ct5009087.
(411) 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. https://doi.org/10.1021/acs.biomac.7b01813.
(412) Tiwary, P.; Parrinello, M. A Time-Independent Free Energy Estimator for Metadynamics. *J. Phys. Chem. B* **2015**, *119*, 14. https://doi.org/10.1021/jp504920s.
(413) 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. https://doi.org/10.1021/acs.jctc.7b01289.
(414) 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. https://doi.org/10.1103/PhysRevLett.107.085504.
(415) Arsiccio, A.; McCarty, J. J.; Pisano, R.; Shea, J.-E. The Effect of Surfactants on Surface-Induced Denaturation of Proteins: Evidence of an Orientation-Dependent Mechanism. *J. Phys. Chem. B* **2018**, acs.jpcb.8b07368. https://doi.org/10.1021/acs.jpcb.8b07368.
(416) Nava, M.; Palazzesi, F.; Perego, C.; Parrinello, M. DImer Metadynamics. *J. Chem. Theory Comput.* **2017**, *13* (2), 425–430.
(417) 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.
(418) Tiwary, P.; Berne, B. J. Spectral Gap Optimization of Order Parameters for Sampling Complex Molecular Systems. *Proc. Natl. Acad. Sci.* **2016**, *113* (11), 2839 LP – 2844.
(419) Klein-marcuschamer, D.; Oleskowicz-popiel, P.; Simmons, B. A.; Blanch, H. W. The Challenge of Enzyme Cost in the Production of Lignocellulosic Biofuels. **2012**, *xxx* (xxx), 1–5. https://doi.org/10.1002/bit.24370.
(420) Katz, M.; Reese, E. T. Production of Glucose by Enzymatic Hydrolysis of Cellulose. **1968**, *16* (2), 419–420.
(421) Chen, M.; Bu, L.; Alahuhta, M.; Brunecky, R.; Xu, Q.; Lunin, V. V.; Brady, J. W.; Crowley, M. F.; Himmel, M. E.; Bomble, Y. J. Strategies to Reduce End-Product Inhibition in Family 48 Glycoside Hydrolases. *Proteins Struct. Funct. Bioinforma.* **2016**, *84* (3), 295–304.

https://doi.org/10.1002/prot.24965.
(422) Holtzapple, M.; Cognata, M.; Hendrickson, C. In Hi Bit Io n of Trichoderma Reesei Cel I u Lase by Sugars and Solvents.
(423) Andrić, P.; Meyer, A. S.; Jensen, P. A.; Dam-Johansen, K. Reactor Design for Minimizing Product Inhibition during Enzymatic Lignocellulose Hydrolysis: I. Significance and Mechanism of Cellobiose and Glucose Inhibition on Cellulolytic Enzymes. *Biotechnol. Adv.* **2010**, *28* (3), 308–324. https://doi.org/10.1016/j.biotechadv.2010.01.003.
(424) Modenbach, A. A.; Nokes, S. E. Enzymatic Hydrolysis of Biomass at High-Solids Loadings e A Review. *Biomass and Bioenergy* **2013**, *56*, 526–544. https://doi.org/10.1016/j.biombioe.2013.05.031.
(425) Gomes, D.; Rodrigues, A. C.; Domingues, L.; Gama, M.; Rodrigues, A. C. Cellulase Recycling in Biorefineries — Is It Possible ? **2015**, 4131–4143. https://doi.org/10.1007/s00253-015-6535-z.
(426) Ahmadi, H.; Anne, G.; Mikkelsen, J. D. Enhanced Enzymatic Cellulose Degradation by Cellobiohydrolases via Product Removal. **2013**, 205–212. https://doi.org/10.1007/s10529-012-1067-4.
(427) Gan, Q.; Allen, S. J.; Taylor, G. Design and Operation of an Integrated Membrane Reactor for Enzymatic Cellulose Hydrolysis. **2002**, *12*, 223–229.
(428) Andri, P.; Meyer, A. S.; Jensen, P. A.; Dam-johansen, K. Reactor Design for Minimizing Product Inhibition during Enzymatic Lignocellulose Hydrolysis II . Quanti Fi Cation of Inhibition and Suitability of Membrane Reactors. **2010**, *28*, 407–425. https://doi.org/10.1016/j.biotechadv.2010.02.005.
(429) Payne, C. M.; Knott, B. C.; Mayes, H. B.; Hansson, H.; Himmel, M. E.; Sandgren, M.; Sta, J.; Beckham, G. T. Fungal Cellulases. **2015**. https://doi.org/10.1021/cr500351c.
(430) Atreya, M. E.; Strobel, K. L.; Clark, D. S. Alleviating Product Inhibition in Cellulase Enzyme Cel7A. *Biotechnol. Bioeng.* **2016**, *113* (2), 330–338. https://doi.org/10.1002/bit.25809.
(431) Silveira, R. L.; Skaf, M. S. Molecular Dynamics Simulations of Family 7 Cellobiohydrolase Mutants Aimed at Reducing Product Inhibition. **2015**. https://doi.org/10.1021/jp509911m.
(432) Bu, L.; Beckham, G. T.; Shirts, M. R.; Nimlos, M. R.; Adney, W. S.; Himmel, M. E.; Crowley, M. F. Probing Carbohydrate Product Expulsion from a Processive Cellulase with Multiple Absolute Binding Free Energy. **2011**, *286* (20), 18161–18169. https://doi.org/10.1074/jbc.M110.212076.
(433) Baker, J. O.; McCarley, J. R.; Lovett, R.; Yu, C.-H.; Adney, W. S.; Rignall, T. R.; Vinzant, T. B.; Decker, S. R.; Sakon, J.; Himmel, M. E. Catalytically Enhanced Endocellulase Cel5A from Acidothermus Cellulolyticus. *Appl. Biochem. Biotechnol.* **2005**, *121–124*, 129–148. https://doi.org/10.1385/ABAB:121:1-3:0129.
(434) Teugjas, H.; Väljamäe, P. Selecting β -Glucosidases to Support Cellulases in Cellulose Saccharification. **2013**, 1–13.
(435) Von Ossowski, I.; Ståhlberg, J.; Koivula, A.; Piens, K.; Becker, D.; Boer, H.; Harle, R.; Harris, M.; Divne, C.; Mahdi, S.; et al. Engineering the Exo-Loop of Trichoderma Reesei Cellobiohydrolase, Cel7A. A Comparison with Phanerochaete Chrysosporium Cel7D. *J. Mol. Biol.* **2003**, *333* (4), 817–829. https://doi.org/10.1016/S0022-2836(03)00881-7.
(436) Summers, S. R.; Sprenger, K. G.; Pfaendtner, J.; Marchant, J.; Summers, M. F.; Kaar, J. L. Mechanism of Competitive Inhibition and Destabilization of Acidothermus Cellulolyticus Endoglucanase 1 by Ionic Liquids. *J. Phys. Chem. B* **2017**, *121* (48), 10793–10803.

https://doi.org/10.1021/acs.jpcb.7b08435.
(437) Summers, S.; Kraft, C.; Alamdari, S.; Pfaendtner, J.; Kaar, J. L. Enhanced Activity and Stability of Acidothermus Cellulolyticus Endoglucanase 1 in Ionic Liquids via Engineering Active Site Residues and Non-Native Disul Fi de Bridges. **2020**. https://doi.org/10.1021/acssuschemeng.0c03242.
(438) Sakon, J.; Adney, W. S.; Himmel, M. E.; Thomas, S. R.; Andrew Karplus, P. Crystal Structure of Thermostable Family 5 Endocellulase E1 from Acidothermus Cellulolyticus in Complex with Cellotetraose. *Biochemistry* **1996**, *35* (33), 10648–10660. https://doi.org/10.1021/bi9604439.
(439) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB. *J. Chem. Theory Comput.* **2015**, *11* (8), 3696–3713. https://doi.org/10.1021/acs.jctc.5b00255.
(440) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. GLYCAM06: A Generalizable Biomolecular Force Field. Carbohydrates. *J. Comput. Chem.* **2008**, *29* (4), 622–655. https://doi.org/10.1002/jcc.20820.
(441) Raniolo, S.; Limongelli, V. Ligand Binding Free-Energy Calculations with Funnel Metadynamics. *Nat. Protoc.* **2020**, *15* (9), 2837–2866. https://doi.org/10.1038/s41596-020-0342-4.
(442) 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. https://doi.org/10.1016/0021-9991(77)90121-8.
(443) Manuel, R.; Bezerra, F.; Dias, A. A. Cellulose Hydrolysis by Cellobiohydrolase Cel7A Shows Mixed Hyperbolic Product Inhibition. **2011**, 178–189. https://doi.org/10.1007/s12010-011-9242-y.
(444) Rouvinen, J.; Bergfors, T.; Teeri, T.; Knowles, J. K. C. Three-Dimensional Structure Of. *249* (16), 380–387.
(445) Varrot, A.; Schu, M.; Davies, G. J. Insights into Ligand-Induced Conformational Change in Cel5A from Bacillus Agaradhaerens Revealed by a Catalytically Active Crystal Form. **2000**, *10*. https://doi.org/10.1006/jmbi.2000.3567.
(446) Domı, R.; Lascombe, M.; Alzari, P. M. The Crystal Structure of a Family 5 Endoglucanase Mutant in Complexed and Uncomplexed Forms Reveals an Induced Fit Activation Mechanism. **1996**, 1042–1051.
(447) Davies, G. J.; Mackenzie, L.; Varrot, A.; Dauter, M.; Brzozowski, A. M.; Schülein, M.; Withers, S. G. Snapshots along an Enzymatic Reaction Coordinate: Analysis of a Retaining β-Glycoside Hydrolase. *Biochemistry* **1998**, *37* (34), 11707–11713. https://doi.org/10.1021/bi981315i.
(448) Fersht, A. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding.*; 1999; Vol. 13409. https://doi.org/10.1016/S0307-4412(99)00114-4.
(449) Limongelli, V.; Bonomi, M.; Parrinello, M. Funnel Metadynamics as Accurate Binding Free-Energy Method. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (16), 6358–6363. https://doi.org/10.1073/pnas.1303186110.
(450) Stankovi, I. M.; Blagojevi, J. P.; Zari, D. International Journal of Biological Macromolecules Carbohydrate – Protein Aromatic Ring Interactions beyond CH / π Interactions : A Protein Data Bank Survey and Quantum Chemical Calculations. **2020**, *157*, 1–9. https://doi.org/10.1016/j.ijbiomac.2020.03.251.

(451) Lynd, L. R.; Weimer, P. J.; Zyl, W. H. Van; Pretorius, I. S. Microbial Cellulose Utilization : Fundamentals and Biotechnology. **2002**, *66* (3), 506–577. https://doi.org/10.1128/MMBR.66.3.506.
(452) Nordwald, E. M.; Brunecky, R.; Himmel, M. E.; Beckham, G. T.; Kaar, J. L. Charge Engineering of Cellulases Improves Ionic Liquid Tolerance and Reduces Lignin Inhibition. *Biotechnol. Bioeng.* **2014**, *111* (8), 1541–1549. https://doi.org/10.1002/bit.25216.
(453) Meier, M. A. R.; Barner-Kowollik, C. A New Class of Materials: Sequence-Defined Macromolecules and Their Emerging Applications. *Adv. Mater.* **2019**, *31* (26), 1806027. https://doi.org/10.1002/adma.201806027.
(454) Gangloff, N.; Ulbricht, J.; Lorson, T.; Schlaad, H.; Luxenhofer, R. Peptoids and Polypeptoids at the Frontier of Supra- and Macromolecular Engineering. *Chem. Rev.* **2016**, *116* (4), 1753–1802. https://doi.org/10.1021/acs.chemrev.5b00201.
(455) Lutz, J.-F.; Ouchi, M.; Liu, D. R.; Sawamoto, M. Sequence-Controlled Polymers. *Science (80-.).* **2013**, *341* (6146), 1238149. https://doi.org/10.1126/science.1238149.
(456) Sun, J.; Zuckermann, R. N. Peptoid Polymers: A Highly Designable Bioinspired Material. *ACS Nano* **2013**, *7* (6), 4715–4732. https://doi.org/10.1021/nn4015714.
(457) Xuan, S.; Zuckermann, R. N. Engineering the Atomic Structure of Sequence-Defined Peptoid Polymers and Their Assemblies. *Polymer (Guildf).* **2020**, *202*, 122691. https://doi.org/10.1016/j.polymer.2020.122691.
(458) Maayan, G.; Ward, M. D.; Kirshenbaum, K. Folded Biomimetic Oligomers for Enantioselective Catalysis. *Proc. Natl. Acad. Sci.* **2009**, *106* (33), 13679–13684. https://doi.org/10.1073/pnas.0903187106.
(459) Statz, A. R.; Meagher, R. J.; Barron, A. E.; Messersmith, P. B. New Peptidomimetic Polymers for Antifouling Surfaces. *J. Am. Chem. Soc.* **2005**, *127* (22), 7972–7973. https://doi.org/10.1021/ja0522534.
(460) Nguyen, A. I.; Spencer, R. K.; Anderson, C. L.; Zuckermann, R. N. A Bio-Inspired Approach to Ligand Design: Folding Single-Chain Peptoids to Chelate a Multimetallic Cluster. *Chem. Sci.* **2018**, *9* (47), 8806–8813. https://doi.org/10.1039/C8SC04240C.
(461) 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. https://doi.org/10.1021/acs.accounts.5b00439.
(462) Lee, B.-C.; Zuckermann, R. N.; Dill, K. A. Folding a Nonbiological Polymer into a Compact Multihelical Structure. *J. Am. Chem. Soc.* **2005**, *127* (31), 10999–11009. https://doi.org/10.1021/ja0514904.
(463) Kirshenbaum, K.; Barron, A. E.; Goldsmith, R. A.; Armand, P.; Bradley, E. K.; Truong, K. T. V.; Dill, K. A.; Cohen, F. E.; Zuckermann, R. N. Sequence-Specific Polypeptoids: A Diverse Family of Heteropolymers with Stable Secondary Structure. *Proc. Natl. Acad. Sci.* **1998**, *95* (8), 4303–4308. https://doi.org/10.1073/pnas.95.8.4303.
(464) 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. https://doi.org/10.1021/ja905267k.
(465) Baskin, M.; Zhu, H.; Qu, Z.-W.; Chill, J. H.; Grimme, S.; Maayan, G. Folding of Unstructured Peptoids and Formation of Hetero-Bimetallic Peptoid Complexes upon Side-Chain-to-Metal Coordination. *Chem. Sci.* **2019**, *10* (2), 620–632. https://doi.org/10.1039/C8SC03616K.

(466) Shin, S. B. Y.; Yoo, B.; Todaro, L. J.; Kirshenbaum, K. Cyclic Peptoids. *J. Am. Chem. Soc.* **2007**, *129* (11), 3218–3225. https://doi.org/10.1021/ja066960o.
(467) Pokorski, J. K.; Miller Jenkins, L. M.; Feng, H.; Durell, S. R.; Bai, Y.; Appella, D. H. Introduction of a Triazole Amino Acid into a Peptoid Oligomer Induces Turn Formation in Aqueous Solution. *Org. Lett.* **2007**, *9* (12), 2381–2383. https://doi.org/10.1021/ol070817y.
(468) Gorske, B. C.; Mumford, E. M.; Gerrity, C. G.; Ko, I. A Peptoid Square Helix via Synergistic Control of Backbone Dihedral Angles. *J. Am. Chem. Soc.* **2017**, *139* (24), 8070–8073. https://doi.org/10.1021/jacs.7b02319.
(469) D'Amato, A.; Pierri, G.; Tedesco, C.; Della Sala, G.; Izzo, I.; Costabile, C.; De Riccardis, F. Reverse Turn and Loop Secondary Structures in Stereodefined Cyclic Peptoid Scaffolds. *J. Org. Chem.* **2019**, *84* (17), 10911–10928. https://doi.org/10.1021/acs.joc.9b01509.
(470) Crapster, J. A.; Guzei, I. A.; Blackwell, H. E. A Peptoid Ribbon Secondary Structure. *Angew. Chemie Int. Ed.* **2013**, *52* (19), 5079–5084. https://doi.org/10.1002/anie.201208630.
(471) Yang, W.; Yin, Q.; Chen, C.-L. Designing Sequence-Defined Peptoids for Biomimetic Control over Inorganic Crystallization. *Chem. Mater.* **2021**, *33* (9), 3047–3065. https://doi.org/10.1021/acs.chemmater.1c00243.
(472) Chien, Y.-C.; Tao, J.; Saeki, K.; Chin, A. F.; Lau, J. L.; Chen, C.-L.; Zuckermann, R. N.; Marshall, S. J.; Marshall, G. W.; De Yoreo, J. J. Using Biomimetic Polymers in Place of Noncollagenous Proteins to Achieve Functional Remineralization of Dentin Tissues. *ACS Biomater. Sci. Eng.* **2017**, *3* (12), 3469–3479. https://doi.org/10.1021/acsbiomaterials.7b00378.
(473) Gopalakrishnan, R.; Frolov, A. I.; Knerr, L.; Drury, W. J.; Valeur, E. Therapeutic Potential of Foldamers: From Chemical Biology Tools To Drug Candidates? *J. Med. Chem.* **2016**, *59* (21), 9599–9621. https://doi.org/10.1021/acs.jmedchem.6b00376.
(474) Yashima, E.; Ousaka, N.; Taura, D.; Shimomura, K.; Ikai, T.; Maeda, K. Supramolecular Helical Systems: Helical Assemblies of Small Molecules, Foldamers, and Polymers with Chiral Amplification and Their Functions. *Chem. Rev.* **2016**, *116* (22), 13752–13990. https://doi.org/10.1021/acs.chemrev.6b00354.
(475) Wu, C. W.; Sanborn, T. J.; Huang, K.; Zuckermann, R. N.; Barron, A. E. Peptoid Oligomers with α-Chiral, Aromatic Side Chains: Sequence Requirements for the Formation of Stable Peptoid Helices. *J. Am. Chem. Soc.* **2001**, *123* (28), 6778–6784. https://doi.org/10.1021/ja003154n.
(476) 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 α-Chiral Aliphatic Side Chains. *J. Am. Chem. Soc.* **2003**, *125* (44), 13525–13530. https://doi.org/10.1021/ja037540r.
(477) Armand, P.; Kirshenbaum, K.; Falicov, A.; Dunbrack, R. L.; Dill, K. A.; Zuckermann, R. N.; Cohen, F. E. Chiral N-Substituted Glycines Can Form Stable Helical Conformations. *Fold. Des.* **1997**, *2* (6), 369–375. https://doi.org/10.1016/S1359-0278(97)00051-5.
(478) Stringer, J. R.; Crapster, J. A.; Guzei, I. A.; Blackwell, H. E. Extraordinarily Robust Polyproline Type I Peptoid Helices Generated via the Incorporation of α-Chiral Aromatic N -1-Naphthylethyl Side Chains. *J. Am. Chem. Soc.* **2011**, *133* (39), 15559–15567. https://doi.org/10.1021/ja204755p.
(479) Roy, O.; Dumonteil, G.; Faure, S.; Jouffret, L.; Kriznik, A.; Taillefumier, C. Homogeneous and Robust Polyproline Type I Helices from Peptoids with Nonaromatic α-Chiral Side Chains. *J. Am. Chem. Soc.* **2017**, *139* (38), 13533–13540.

https://doi.org/10.1021/jacs.7b07475.
(480) Stringer, J. R.; Crapster, J. A.; Guzei, I. A.; Blackwell, H. E. Construction of Peptoids with All Trans -Amide Backbones and Peptoid Reverse Turns via the Tactical Incorporation of N -Aryl Side Chains Capable of Hydrogen Bonding. *J. Org. Chem.* **2010**, *75* (18), 6068–6078. https://doi.org/10.1021/jo101075a.
(481) Reese, H. R.; Shanahan, C. C.; Proulx, C.; Menegatti, S. Peptide Science: A "Rule Model" for New Generations of Peptidomimetics. *Acta Biomater.* **2020**, *102*, 35–74. https://doi.org/10.1016/j.actbio.2019.10.045.
(482) Butterfoss, G. L.; Yoo, B.; Jaworski, J. N.; Chorny, I.; Dill, K. A.; Zuckermann, R. N.; Bonneau, R.; Kirshenbaum, K.; Voelz, V. A. De Novo Structure Prediction and Experimental Characterization of Folded Peptoid Oligomers. *Proc. Natl. Acad. Sci.* **2012**, *109* (36), 14320–14325. https://doi.org/10.1073/pnas.1209945109.
(483) 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. https://doi.org/10.1002/bip.21575.
(484) Weiser, L. J.; Santiso, E. E. A CGenFF-based Force Field for Simulations of Peptoids with Both Cis and Trans Peptide Bonds. *J. Comput. Chem.* **2019**, *40* (22), 1946–1956. https://doi.org/10.1002/jcc.25850.
(485) Shah, N. H.; Butterfoss, G. L.; Nguyen, K.; Yoo, B.; Bonneau, R.; Rabenstein, D. L.; Kirshenbaum, K. Oligo(N- Aryl Glycines): A New Twist on Structured Peptoids. *J. Am. Chem. Soc.* **2008**, *130* (49), 16622–16632. https://doi.org/10.1021/ja804580n.
(486) Gorske, B. C.; Stringer, J. R.; Bastian, B. L.; Fowler, S. A.; Blackwell, H. E. New Strategies for the Design of Folded Peptoids Revealed by a Survey of Noncovalent Interactions in Model Systems. *J. Am. Chem. Soc.* **2009**, *131* (45), 16555–16567. https://doi.org/10.1021/ja907184g.
(487) Haxton, T. K.; Mannige, R. V.; Zuckermann, R. N.; Whitelam, S. Modeling Sequence-Specific Polymers Using Anisotropic Coarse-Grained Sites Allows Quantitative Comparison with Experiment. *J. Chem. Theory Comput.* **2015**, *11* (1), 303–315. https://doi.org/10.1021/ct5010559.
(488) Zhao, M.; Sampath, J.; Alamdari, S.; Shen, G.; Chen, C.-L.; Mundy, C. J.; Pfaendtner, J.; Ferguson, A. L. MARTINI-Compatible Coarse-Grained Model for the Mesoscale Simulation of Peptoids. *J. Phys. Chem. B* **2020**, *124* (36), 7745–7764. https://doi.org/10.1021/acs.jpcb.0c04567.
(489) Renfrew, P. D.; Craven, T. W.; Butterfoss, G. L.; Kirshenbaum, K.; Bonneau, R. A Rotamer Library to Enable Modeling and Design of Peptoid Foldamers. *J. Am. Chem. Soc.* **2014**, *136* (24), 8772–8782. https://doi.org/10.1021/ja503776z.
(490) Hoyas, S.; Lemaur, V.; Duez, Q.; Saintmont, F.; Halin, E.; De Winter, J.; Gerbaux, P.; Cornil, J. PEPDROID: Development of a Generic DREIDING-Based Force Field for the Assessment of Peptoid Secondary Structures. *Adv. Theory Simulations* **2018**, *1* (12), 1800089. https://doi.org/10.1002/adts.201800089.
(491) 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. https://doi.org/10.1021/acs.jpcb.5b09625.
(492) Jiao, S.; DeStefano, A.; Monroe, J. I.; Barry, M.; Sherck, N.; Casey, T.; Segalman, R. A.; Han, S.; Shell, M. S. Quantifying Polypeptoid Conformational Landscapes through Integrated Experiment and Simulation. *Macromolecules* **2021**.

https://doi.org/10.1021/acs.macromol.1c00550.
(493) Hoyas, S.; Halin, E.; Lemaur, V.; De Winter, J.; Gerbaux, P.; Cornil, J. Helicity of Peptoid Ions in the Gas Phase. *Biomacromolecules* **2020**, *21* (2), 903–909. https://doi.org/10.1021/acs.biomac.9b01567.
(494) Hoyas, S.; Weber, P.; Halin, E.; Coulembier, O.; De Winter, J.; Cornil, J.; Gerbaux, P. Helical Peptoid Ions in the Gas Phase: Thwarting the Charge Solvation Effect by H-Bond Compensation. *Biomacromolecules* **2021**, *22* (8), 3543–3551. https://doi.org/10.1021/acs.biomac.1c00623.
(495) Alamdari, S.; Sampath, J.; Prakash, A.; Gibson, L. D.; Pfaendtner, J. Efficient Sampling of High-Dimensional Free Energy Landscapes: A Review of Parallel Bias Metadynamics; 2021; pp 123–141. https://doi.org/10.1007/978-981-33-6639-8_6.
(496) Wijaya, A. W.; Nguyen, A. I.; Roe, L. T.; Butterfoss, G. L.; Spencer, R. K.; Li, N. K.; Zuckermann, R. N. Cooperative Intramolecular Hydrogen Bonding Strongly Enforces Cis -Peptoid Folding. *J. Am. Chem. Soc.* **2019**, *141* (49), 19436–19447. https://doi.org/10.1021/jacs.9b10497.
(497) Hoyas, S.; Roscioni, O. M.; Tonneaux, C.; Gerbaux, P.; Cornil, J.; Muccioli, L. Peptoids as a Chiral Stationary Phase for Liquid Chromatography: Insights from Molecular Dynamics Simulations. *Biomacromolecules* **2021**, *22* (6), 2573–2581. https://doi.org/10.1021/acs.biomac.1c00302.
(498) Davern, C. M.; Lowe, B. D.; Rosfi, A.; Ison, E. A.; Proulx, C. Submonomer Synthesis of Peptoids Containing Trans -Inducing N -Imino- and N -Alkylamino-Glycines. *Chem. Sci.* **2021**, *12* (24), 8401–8410. https://doi.org/10.1039/D1SC00717C.
(499) Spencer, R. K.; Butterfoss, G. L.; Edison, J. R.; Eastwood, J. R.; Whitelam, S.; Kirshenbaum, K.; Zuckermann, R. N. Stereochemistry of Polypeptoid Chain Configurations. *Biopolymers* **2019**, *110* (6). https://doi.org/10.1002/bip.23266.
(500) 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. https://doi.org/10.1021/ja0574318.
(501) Simon, R. J.; Kania, R. S.; Zuckermann, R. N.; Huebner, V. D.; Jewell, D. A.; Banville, S.; Ng, S.; Wang, L.; Rosenberg, S.; Marlowe, C. K. Peptoids: A Modular Approach to Drug Discovery. *Proc. Natl. Acad. Sci.* **1992**, *89* (20), 9367–9371. https://doi.org/10.1073/pnas.89.20.9367.
(502) Bovey, F. A.; Ryan, J. J.; Hood, F. P. Polymer Nuclear Magnetic Resonance Spectroscopy. XV. The Conformation of Polysarcosine. *Macromolecules* **1968**, *1* (4), 305–307. https://doi.org/10.1021/ma60004a004.
(503) Davidson, E. C.; Rosales, A. M.; Patterson, A. L.; Russ, B.; Yu, B.; Zuckermann, R. N.; Segalman, R. A. Impact of Helical Chain Shape in Sequence-Defined Polymers on Polypeptoid Block Copolymer Self-Assembly. *Macromolecules* **2018**, *51* (5), 2089–2098. https://doi.org/10.1021/acs.macromol.8b00055.
(504) Pietrucci, F.; Laio, A. A Collective Variable for the Efficient Exploration of Protein Beta-Sheet Structures: Application to SH3 and GB1. *J. Chem. Theory Comput.* **2009**, *5* (9), 2197–2201. https://doi.org/10.1021/ct900202f.
(505) 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. https://doi.org/10.1021/ja062463w.

(506) 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. https://doi.org/10.1021/ct300297t.
(507) Jaeger, V.; Burney, P.; Pfaendtner, J. Comparison of Three Ionic Liquid-Tolerant Cellulases by Molecular Dynamics. *Biophys. J.* **2015**, *108* (4), 880–892. https://doi.org/10.1016/j.bpj.2014.12.043.
(508) Nordwald, E. M.; Armstrong, G. S.; Kaar, J. L. NMR-Guided Rational Engineering of an Ionic-Liquid-Tolerant Lipase. *ACS Catal.* **2014**, *4* (11), 4057–4064. https://doi.org/10.1021/cs500978x.
(509) Burney, P. R.; Nordwald, E. M.; Hickman, K.; Kaar, J. L.; Pfaendtner, J. Molecular Dynamics Investigation of the Ionic Liquid/Enzyme Interface: Application to Engineering Enzyme Surface Charge. *Proteins Struct. Funct. Bioinforma.* **2015**, *83* (4), 670–680. https://doi.org/10.1002/prot.24757.
(510) Roeters, S. J.; Golbek, T. W.; Bregnhøj, M.; Drace, T.; Alamdari, S.; Roseboom, W.; Kramer, G.; Šantl-Temkiv, T.; Finster, K.; Pfaendtner, J.; et al. Ice-Nucleating Proteins Are Activated by Low Temperatures to Control the Structure of Interfacial Water. *Nat. Commun.* **2021**, *12* (1), 1183. https://doi.org/10.1038/s41467-021-21349-3.
(511) Chen, C.-L. Peptoid-Based Membrane-Mimetic Two Dimensional Nanomaterials. In *Micro- and Nanotechnology Sensors, Systems, and Applications X*; Islam, M. S., Dutta, A. K., George, T., Eds.; SPIE, 2018; p 33. https://doi.org/10.1117/12.2303685.
(512) Zhang, S.; Zhou, J.; Hu, H.; Gong, H.; Chen, L.; Cheng, C.; Zeng, J. A Deep Learning Framework for Modeling Structural Features of RNA-Binding Protein Targets. *Nucleic Acids Res.* **2016**, *44* (4), e32–e32. https://doi.org/10.1093/nar/gkv1025.
(513) Alipanahi, B.; Delong, A.; Weirauch, M. T.; Frey, B. J. Predicting the Sequence Specificities of DNA- and RNA-Binding Proteins by Deep Learning. *Nat. Biotechnol.* **2015**, *33* (8), 831–838. https://doi.org/10.1038/nbt.3300.
(514) Zeng, H.; Edwards, M. D.; Liu, G.; Gifford, D. K. Convolutional Neural Network Architectures for Predicting DNA–Protein Binding. *Bioinformatics* **2016**, *32* (12), i121–i127. https://doi.org/10.1093/bioinformatics/btw255.
(515) Jiménez, J.; Doerr, S.; Martínez-Rosell, G.; Rose, A. S.; De Fabritiis, G. DeepSite: Protein-Binding Site Predictor Using 3D-Convolutional Neural Networks. *Bioinformatics* **2017**, *33* (19), 3036–3042. https://doi.org/10.1093/bioinformatics/btx350.
(516) Gomes, J.; Ramsundar, B.; Feinberg, E. N.; Pande, V. S. Atomic Convolutional Networks for Predicting Protein-Ligand Binding Affinity. **2017**.
(517) Banerjee, S.; Prasad, R.; Sen, P. In Silico Mutation and Binding Studies of Human FVIIa γ-Carboxyglutamic Acid-Domain to Endothelial Protein C Receptor: A Molecular Dynamics Simulation Approach. *ACS Omega* **2019**, *4* (1), 496–508. https://doi.org/10.1021/acsomega.8b02647.
(518) Sorensen, A. B.; Madsen, J. J.; Svensson, L. A.; Pedersen, A. A.; Østergaard, H.; Overgaard, M. T.; Olsen, O. H.; Gandhi, P. S. Molecular Basis of Enhanced Activity in Factor VIIa-Trypsin Variants Conveys Insights into Tissue Factor-Mediated Allosteric Regulation of Factor VIIa Activity. *J. Biol. Chem.* **2016**, *291* (9), 4671–4683. https://doi.org/10.1074/jbc.M115.698613.
(519) Project, E.; Nachliel, E.; Gutman, M. Parameterization of Ca+2–Protein Interactions for Molecular Dynamics Simulations. *J. Comput. Chem.* **2008**, *29* (7), 1163–1169. https://doi.org/10.1002/jcc.20876.

(520) 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. https://doi.org/10.1002/(SICI)1521-3773(19990115)38:1/2<236::AID-ANIE236>3.0.CO;2-M.
(521) Yan, Y.; Zhang, D.; Huang, S.-Y. Efficient Conformational Ensemble Generation of Protein-Bound Peptides. *J. Cheminform.* **2017**, *9* (1), 59. https://doi.org/10.1186/s13321-017-0246-7.
(522) Alamdari, S.; Roeters, S. J.; Golbek, T. W.; Schmüser, L.; Weidner, T.; Pfaendtner, J. Orientation and Conformation of Proteins at the Air–Water Interface Determined from Integrative Molecular Dynamics Simulations and Sum Frequency Generation Spectroscopy. *Langmuir* **2020**, *36* (40), 11855–11865. https://doi.org/10.1021/acs.langmuir.0c01881.