

Interpretation of Interfacial Protein Spectra with Enhanced Molecular Simulation Ensembles

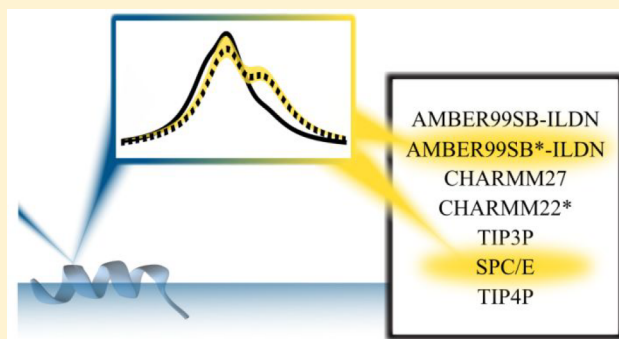
Helmut Lutz,^{*,†,§,#} Vance Jaeger,^{*,‡,§,#} Tobias Weidner,^{†,§} and Bert L. de Groot^{‡,§}

[†]Department of Molecular Spectroscopy, Max Planck Institute for Polymer Research, Mainz 55128, Germany

[‡]Department of Theoretical and Computational Biophysics, Max Planck Institute for Biophysical Chemistry, Göttingen 37077, Germany

Supporting Information

ABSTRACT: An atomistically detailed picture of protein folding at interfaces can effectively be obtained by comparing interface-sensitive spectroscopic techniques to molecular simulations. Here, we present an extensive evaluation of the capability of contemporary force fields to model protein folding at air–water interfaces with a general scheme for sampling and reweighting theoretical conformational ensembles of interfacial peptides. Force field combinations of CHARMM22*/TIP3P and AMBER99SB*-ILDN/SPC/E were found to reproduce experimental observations best.



1. INTRODUCTION

The experimental determination of protein structure at interfaces is a nontrivial problem. The most common methods of structural determination (i.e., X-ray crystallography, nuclear magnetic resonance (NMR) or CryoEM) probe protein conformation only in the solution or crystal phase.¹ However, none of these methods can inform us about the ways in which the protein conformation changes at any given interface. Of the tens of thousands of solved, high-resolution protein structures, none describes a protein at an interface. Because the interfacial structure and dynamics of proteins is of utmost importance for understanding several relevant biological processes, e.g. the handling of protein or peptide drugs on industrial scales, food formulations, biofilm formation, and biomineralization,² the lack of experimental interfacial structures with atomic resolution is unfortunate.

One method to determine the secondary structure and orientation of a protein at an interface is vibrational sum-frequency generation (VSFG) spectroscopy.³ VSFG is a spectroscopic method that probes ordered molecules at an interface with two incident polarized lasers; one is a narrow-band pulse in the visible spectrum, and the second laser's frequency covers regions of the IR spectrum. Due to nonlinear optical frequency mixing at an interface, a photon is emitted at a frequency equal to the sum of the incident beams. The SFG process is physically forbidden in the centrosymmetric bulk phase. Therefore, VSFG is an interface specific method. One shortcoming of VSFG and of other surface-sensitive techniques is that they do not provide molecular-level detail in contrast to common methods such as X-ray crystallography or NMR. To overcome this impediment, researchers rely on molecular

simulations to support VSFG experiments by producing an ensemble of protein conformational states based upon hypotheses about the interfacial coordinates.⁴

Previous studies have not considered two important aspects of the spectral prediction methods. First, while an ensemble of states is generated in molecular dynamics (MD) simulations, often only a single snapshot of the ensemble is used for theoretical spectral prediction. However, one could exploit the large ensemble of conformations provided by the simulation to produce an ensemble spectrum, which in theory should more accurately represent the real-life experimental behavior. Second, the effects of protein and water force fields on proteins' interfacial conformational ensembles have not been thoroughly assessed. The effects of force field selection are known to greatly affect the results of simulations, and therefore care should be taken to select the right force field for a given problem. This study addresses both of these overlooked considerations to develop a more rigorous set of methods and rules for conducting simulations of interfacial proteins. Previous experimental⁵ and theoretical studies⁶ have demonstrated significant changes in protein secondary or tertiary structure at air–solvent interfaces. Therefore, we chose to study and discuss this simplest biologically relevant interface, namely the air–water interface.

Because a protein's adsorbed, interfacial structure can differ greatly from its solution structure, and because there can be multiple stable and metastable protein conformations separated by large energy barriers, advanced sampling techniques

Received: August 12, 2018

Published: November 30, 2018

should be employed to generate a converged structural ensemble. Several enhanced sampling techniques have previously been used to thoroughly sample ensembles of protein conformations at interfaces, including well-tempered metadynamics (WT-MetaD), parallel tempering,⁷ solvent tempering,⁸ and the well-tempered ensemble.⁹ In this study the exploration of the protein conformation is enhanced using WT-MetaD.¹⁰ Because the air–water interface does not restrict the sampling of side chain and backbone conformations in the same way as solid interfaces, we hypothesize that WT-MetaD is sufficient to exhaustively sample the free energy of the system without any further biasing such as parallel tempering which might be used to overcome smaller hidden free-energy barriers associated with protein adsorption.

Since the major branches of both AMBER and CHARMM-based force fields are optimized to reproduce data of aqueous phase proteins and biomolecules,¹¹ there may be considerable effects when introducing an interface into the system, as has been observed for several solid–water interfaces. These effects include the following: (a) differences in solvation free energy of the backbone and side chains among protein force fields,¹² (b) differences in the properties of interfacial water compared to bulk water,¹³ and (c) differences in peptide interactions among other considerations.¹⁴ In this study, several force field combinations were tested for simulations of peptides at the air–water interface. The ability of these simulations to produce peptide ensembles of accurate structure is evaluated by calculating theoretical VSFG spectra from the simulations and comparing these to experimental VSFG data. The protein force fields selected for this study are AMBER99SB-ILDN¹⁵ (A99SB-ILDN), AMBER99*SB-ILDN¹⁶ (A99SB*-ILDN), CHARMM27¹⁷ (C27), and CHARMM22*¹⁸ (C22*). Each of these force fields was paired with one of three water force fields (TIP3P,¹⁹ TIP4P-D,²⁰ and SPC/E²¹). The non-CHARMM version of TIP3P water was used in all cases. A list of force field combinations is provided in Table 1. A99SB-

Table 1. Force Field Combinations Used in Simulations^a

A99SB-ILDN TIP3P	A99SB-ILDN SPC/E	A99SB-ILDN TIP4P-D
A99SB*-ILDN TIP3P	A99SB*-ILDN SPC/E	A99SB*-ILDN TIP4P-D
C27 TIP3P	C27 SPC/E	C27 TIP4P-D
C22* TIP3P	C22* SPC/E	C22* TIP4P-D

^aProtein force fields are abbreviated as follows: AMBER99SB-ILDN (A99SB-ILDN), CHARMM27 (C27), AMBER99*SB-ILDN (A99SB*-ILDN), and CHARMM22* (C22*). TIP3P, SPC/E, and TIP4P-D denote the water force fields, respectively.

ILDN and C27 were selected because of their wide adoption in the literature and their ability to reproduce data from NMR experiments.¹⁸ A99SB*-ILDN and C22* were selected because of some important modifications to backbone dihedral parameters that can greatly affect preferred secondary structures.^{16,18} All four of these force fields are either implemented in GROMACS or readily available as user contributions, which makes them available to a wide range of researchers. The three water models were selected because various water force fields are well-known to affect protein structure differently. Moreover, interfacial properties of each water model differ with respect to structuring and surface tension. These parameters are expected to greatly affect protein secondary structure formation and stability.²²

Four peptides of varying composition and secondary structure have been selected to test these force fields and our ability to reproduce VSFG spectra from molecular dynamics-derived ensembles. The four peptides are aurein 1.2 (A1.2, helical, PDB: 1VMS, GLFDIIKKIAESF),²³ mastoparan X (MPX, helical, PDB: 2CZP, INWKGAAM-AKLL), tryptophan zipper 2 (TZ2, beta hairpin, PDB: 1LE1, SWTWENGKWTWK), and minimal beta-hairpin (MBH, beta hairpin, PDB: 1N09, CTWEGNKLTC).

These structures were selected because each peptide adopts a defined secondary structure in its native state either in solution (TZ2, MBH) or bound to a membrane (A1.2, MPX), each is small enough to quickly simulate for the large force field-parameter space tested in this study, and each has hydrophobic and polar elements to anchor it to the air–water interface without any additional bias. In practice, an air–vacuum interface is used for the simulations, because the density of air is so much smaller than water, and because the addition of a few gaseous molecules is not expected to affect results. A representative air–vacuum simulation box is illustrated in Figure 1A alongside the experimental solution or membrane-bound NMR structures of each of the four peptides in Figure 1B–E.

In the interest of brevity, the results presented in the main text will focus on aurein 1.2 as a demonstration of the methods developed herein with a brief summary of results for all other

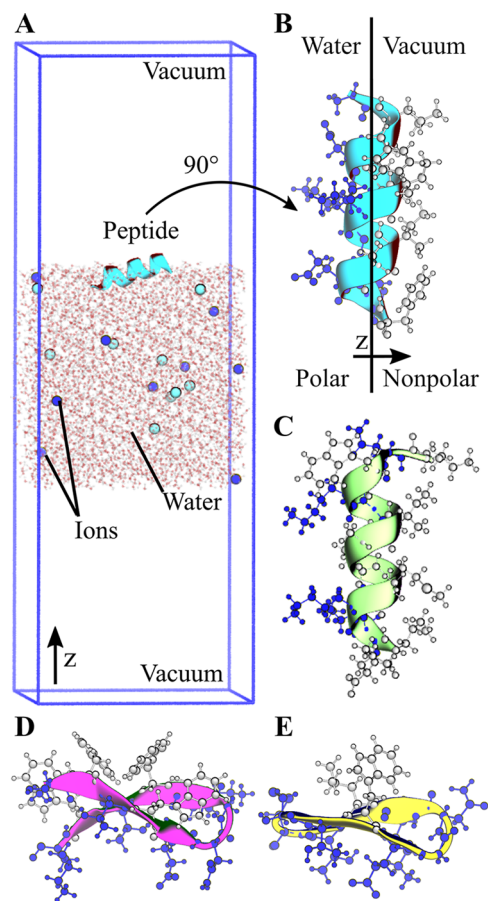


Figure 1. (A) A representative simulation box. (B) Aurein 1.2. (C) Mastoparan X. (D) Tryptophan zipper 2. (E) Minimal beta-hairpin. Hydrophobic side chains in white. Polar side chains in blue.

peptides discussed afterward. Additional figures for other peptides are also included in the [Supporting Information](#).

2. RESULTS AND DISCUSSION

Experimental VSGF Spectra. The four selected model peptides differ greatly in their observed interfacial structures as evident from the various spectral shapes presented in [Figure 2](#). For each peptide, two spectra are displayed—one SSP spectrum and one SPS spectrum. By convention, these letter

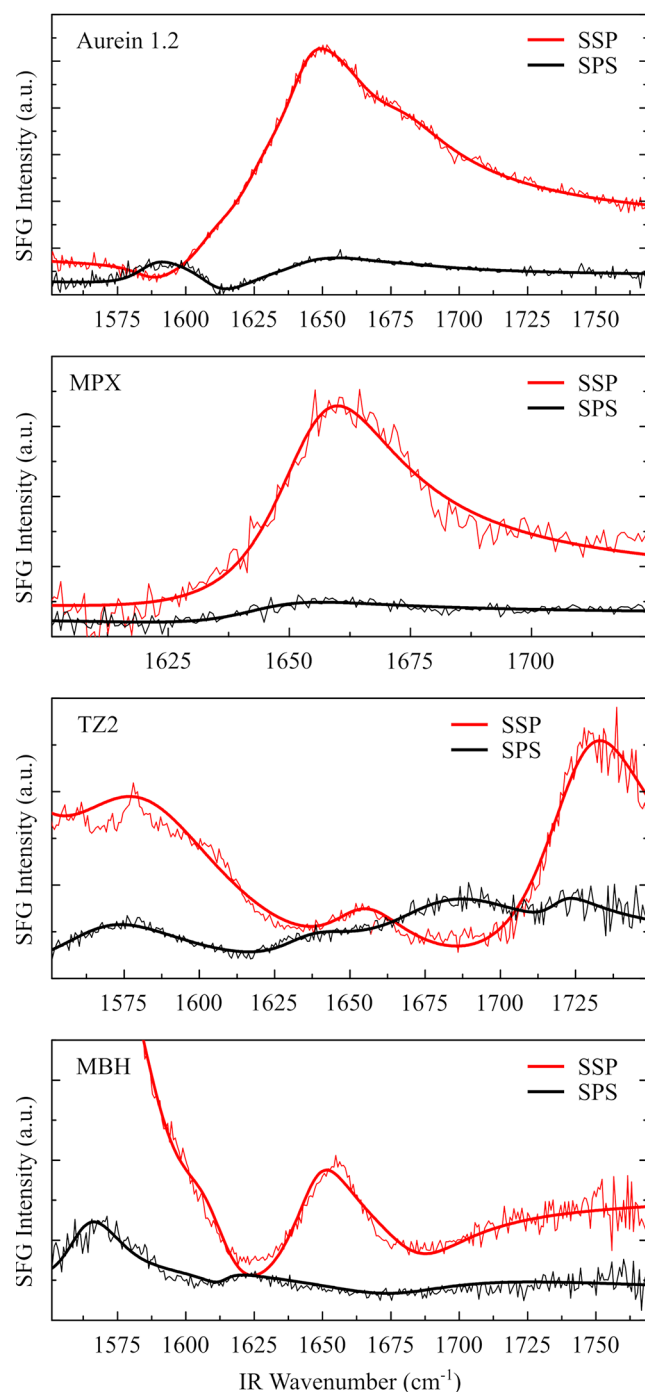


Figure 2. Experimental VSGF spectra of four model peptides: Aurein 1.2 (A1.2), mastoparan X (MPX), tryptophan zipper 2 (TZ2), and minimal beta-hairpin (MBH). SSP and SPS polarization combinations are in red and black, respectively, each with raw data and fitted curves.

combinations denote laser beam polarizations in the order: SFG signal, visible laser, and infrared laser. SSP, for example, refers to perpendicular polarized SFG signal, perpendicular polarized visible laser and parallel polarized infrared laser (with respect to the plane of incidence). While the incident beams can probe several nanometers into the solution, a VSGF response can only be generated by ordered molecules adsorbed to the interface and not by isotropically oriented bulk peptides. Experimental VSGF spectra displayed a nonresonant sum-frequency response on the high frequency side of the spectrum. This nonresonant response renders a comparison of experimental and calculated spectra difficult. To make experimental and calculated VSGF spectra comparable, the experimental spectra were fit as described in the [Methods](#) section. Subsequently, the amplitude and the phase of the nonresonant contribution to the spectra were set to zero while keeping all other fit parameters constant.

The parameters for each fit are presented in [Tables S1–S4](#). Fitting of experimental peaks and the calculation of theoretical spectra were blind to one another, and no parameters from the fitting were considered in the theoretical spectra calculations. Apart from excluding nonresonant background, the fits served another purpose. Namely, certain peaks at the edges of the spectra were assumed to not originate from amide I backbone coupling. These peaks were ignored because they give no information about the secondary structure of the peptide, and because the theory used to calculate theoretical spectra considers only backbone interactions. Specifically, certain modes can be assigned to the COO[−] asymmetric stretching vibration (of glutamic and aspartic acid) and the phenyl ring in-plane vibration modes of phenylalanine.²⁴ Each ignored mode is assigned a zero amplitude. Parameters for all modes can be found in [Tables S1–S4](#). After processing the data to remove contributions from side chains and the nonresonant background, each of the four peptides exhibits a unique spectral signature in the amide I regime.

Metadynamics Simulations. Enhanced sampling through WT-MetaD was used to assess the ability of different force fields to reproduce a natural ensemble of peptide secondary structures at the air–water interface. The C α radius of gyration and the number of structural hydrogen bonds were selected as collective variables to be biased. The free-energy landscapes of all peptides with respect to these two biased collective variables are shown in [Figure 3](#) and [Figures S1–S3](#). It should be noted that simulations with different combinations of water and protein force fields return distinct free-energy surfaces, and this result is expected given the variety of force fields selected.

Aurein 1.2 Free Energy Surfaces. The native micelle-bound structure of aurein 1.2, as determined from NMR, consists of a single α helix and sits roughly at a coordinate of [0.65, 6.0] in the plots shown in [Figure 3](#). A common feature in all aurein 1.2 free-energy landscapes is the narrow horizontal region near zero α -helix hydrogen bonds. Peptides located in this area of the free-energy landscape form hydrogen bonds neither at the same positions as seen in the NMR structure nor in a perfect α -helical pattern. However, this does not exclude the presence of other secondary structures. This horizontal band contains a range of diverse structures from highly compacted to completely extended states. Most of the 12 force field combinations retain a basin near the solution structure, except in the cases of C22* and some of the TIP4P-D simulations. Because the “*” variations of the force fields adjust the protein backbone parameters to change the population of

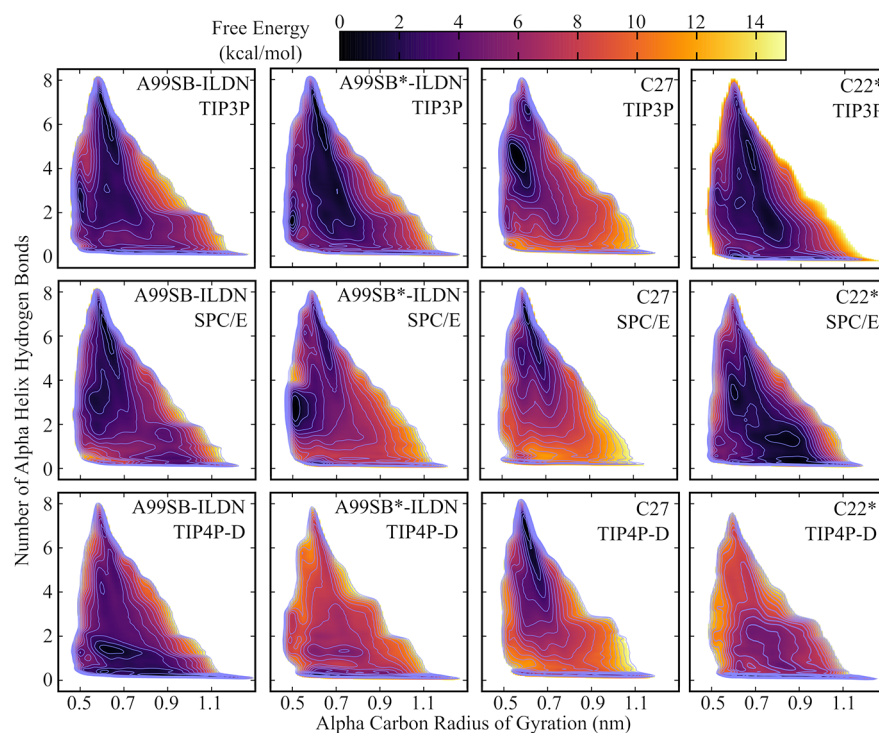


Figure 3. Free energy surfaces from WT-MetaD simulations of aurein 1.2. The protein/water force field combination is indicated in the top right corner of each cell.

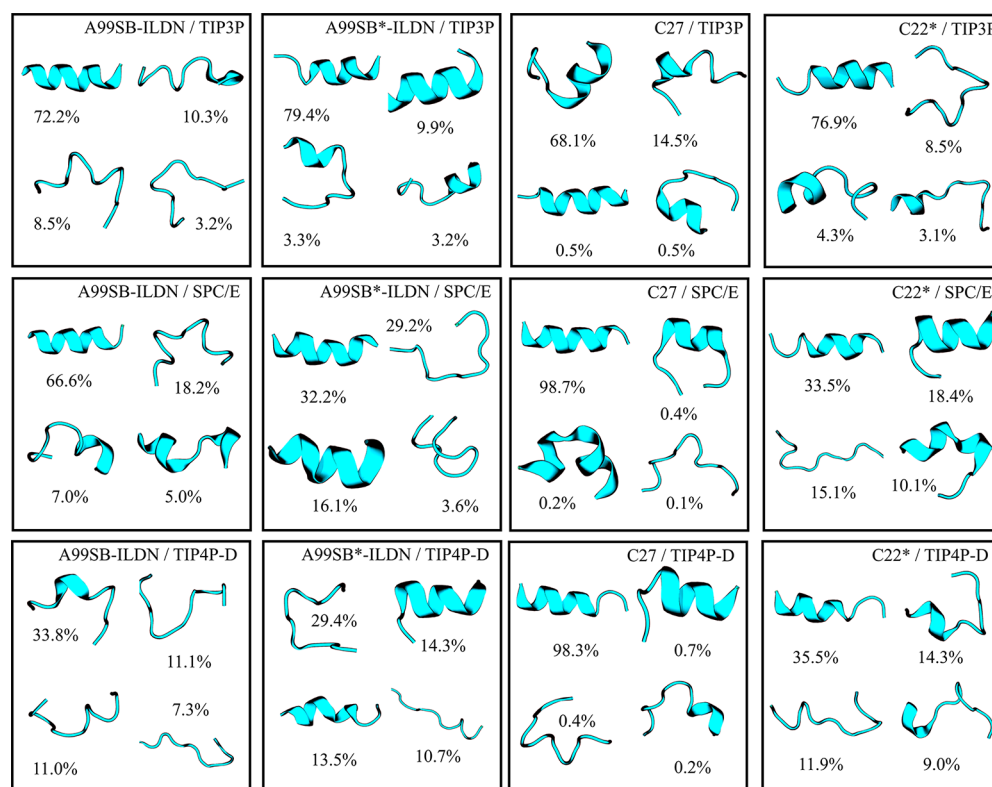


Figure 4. Each cell illustrates the central structure of each of the top four clusters ranked by probability. The force field combinations used in each simulation are depicted in the top right corner of each cell.

α -helices versus coils, a change in the propensity to form a helix at the interface should be expected. TIP4P-D was parametrized in order to better sample intrinsically disordered proteins, and since proteins and peptides at interfaces often

become disordered, it was interesting to test whether this water model might give a population that differs greatly from the solution structure. This was in fact the case for three of the force field combinations containing TIP4P-D water. There are

other states to be explored besides the extended state and the native folded state. For example, in several of the AMBER force field combinations a second well appears for a compacted, less helical structure near the coordinate [0.55, 3.0]. If such a structure maintains some of its helical character, it could be expected to generate a VSFSG signal that has α -helical character with some additional features.

Convergence of the free energy surfaces shown in Figure 3 is illustrated in Figure S4 in which the hill height is shown to decay drastically from its initial height of 2.0 kJ/mol. While this does not definitively demonstrate convergence, it demonstrates that the systems are not exploring new regions of collective variable space.

Most hills added at the end of the trajectory are smaller than 0.02 kJ/mol in CV space, meaning that additional bias is not being added at any appreciable rate. It is common to further assess convergence by comparing the relative free energies of two low energy basins on the free energy surface over time. However, since many of the 48 free energy surfaces do not contain two well-defined energy states, this method was not utilized.

Aurein 1.2 Conformational Clustering. The structures represented in the low-energy basins presented in Figure 3 can be diverse even when their CV coordinates are very similar. Trajectories were clustered as described in the Methods section. Representative structures for the four most populated clusters are shown in Figure 4.

Inspection of the highly populated clusters reveals a large diversity in the structures sampled across the range of protein and water force fields used in this study. This supports the earlier hypothesis that force field selection does matter significantly for problems involving interfaces. All force fields have varying degrees of α -helical character, the least of which are A99SB*-ILDN and A99SB-ILDN with TIP4P-D water. As mentioned before, the TIP4P-D water model is meant to reproduce intrinsically disordered structures, perhaps explaining the observed low populations of helical peptides. Overall, it appears that there is no protein or water force field that consistently produces similar secondary structures no matter which other force field is paired with them.

To verify our assertion that the air–water interface does affect the structuring of these peptides, we simulated aurein 1.2 in the solution state with the A99SB*-ILDN/SPC/E force field combination. This simulation was conducted by simply removing the vacuum layer present in our interfacial simulations. All other parameters remained the same. The free energy surfaces for the interfacial and solution-state peptides are presented in Figure S5. We observe that the peptide, when bound to the interface, prefers a structure similar to the membrane-bound experimental structure along with a partially unfolded compact structure. On the other hand, when simulated in solution, the folded and partially unfolded structures are no longer highly populated, and instead the peptide prefers to assume no helical secondary structure at all. These results indicate that peptide secondary structure can in fact be induced by the introduction of a hydrophobic interface as asserted in the Introduction section.

Theoretical VSFSG Calculation. To test which of the force fields best match experiment, an ensemble of peptide structures obtained from the WT-MetaD simulation was used to calculate a theoretical VSFSG spectrum. Eight samples of 25 conformations were drawn from the trajectory. These 25 conformations were selected by accepting or rejecting frames

with a probability proportional to that obtained from reweighting. The weight of a respective frame was obtained by the Torrie-Valleau method.²⁵ The 25 conformations were aligned in one plane in a 5×5 grid with about 3 nm of space between individual peptides. This file was subsequently used to calculate VSFSG spectra using the method of Roeters et al.^{3b}

Results of the theoretical calculation for aurein 1.2 are presented in Figure 5. Diversity in predicted spectra is quite

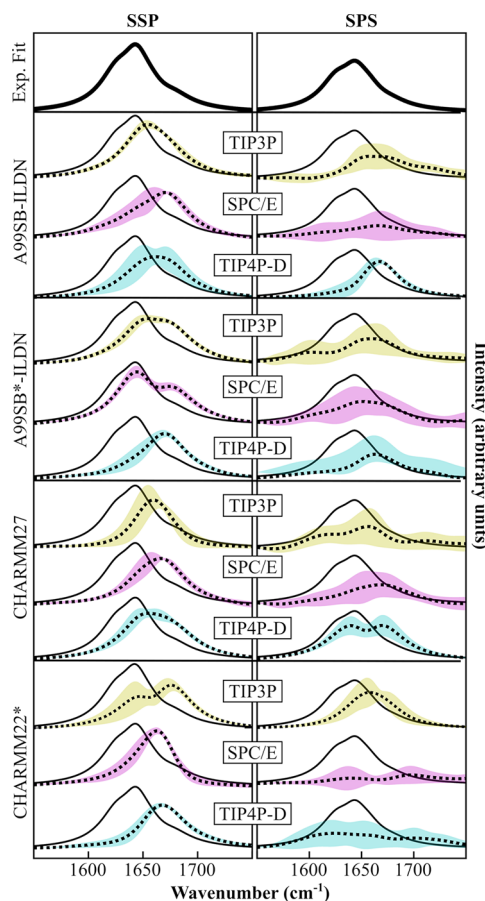


Figure 5. Comparison between experimental and calculated VSFSG spectra for the SSP (left) and SPS polarization (right). Experimental spectra are solid black curves corrected for nonresonant background and side chain related resonances. Theoretical spectra are dotted black curves with colored regions for the standard deviation over eight samples.

large, with predicted peak locations spanning a range of almost 40 cm^{-1} . Most theoretical curves are blue-shifted noticeably compared to experimental data. Some of the blue shift can be explained by the theoretical calculations not considering hydrogen bonds with the solvent. However, Roeters estimates this shift to be about 5 cm^{-1} , which is much smaller than observed in most calculated spectra. Therefore, we hypothesize that most force field combinations are not accurately representing the true ensemble.

From these observations, we hypothesize that the true experimental ensemble likely contains high populations of partially unfolded helices and extended structures rather than well-formed helices as would be predicted by most other force fields.

To quantify differences in the experimental and theoretical spectra, the calculated VSFSG spectra were compared to the

experimental curves by first shifting both curves to zero at a wavenumber of 1550 cm^{-1} and then scaling the intensity of the calculated spectrum by a constant to minimize the root-mean-square difference (rmsd) between the calculated and experimental points. Unlike some other spectroscopic techniques, the intensity of the spectrum depends upon various experimental factors. Therefore, the absolute magnitude should not be considered, but instead the relative amplitude and location of SFG peaks with respect to each other are the important spectral features. Thus, a simple rescaling of the amplitude by multiplication of a constant is valid. Calculated points were linearly interpolated to match the exact discrete wavenumbers at which data were collected. To obtain a final scoring, the 12 minimized rmsd values (one for each force field combination) were normalized for all four model peptides. SSP and SPS spectra were normalized independently from one another. Then, the scores for SSP and SPS were summed into a final score. These scores are reported in Table 2 for each force field combination for all four peptides tested. To test the overall performance of each force field combination, the tables for each of the four model peptides were summed into a combined table.

Table 2. Ranking the Force Fields According to Minimum rmsd between Calculated and Experimental Spectra^a

A1.2:	A99SB-ILDN	A99SB*-ILDN	C27	C22*
TIP3P	0.93	0.95	1.39	1.00
SPC/E	1.54	0.48	1.40	1.63
TIP4P-D	1.23	1.51	0.38	1.65
MPX:	A99SB-ILDN	A99SB*-ILDN	C27	C22*
TIP3P	0.36	1.58	0.65	0.60
SPC/E	1.21	0.13	1.19	1.08
TIP4P-D	0.96	0.26	1.11	1.49
TZ2:	A99SB-ILDN	A99SB*-ILDN	C27	C22*
TIP3P	0.51	0.50	0.76	0.72
SPC/E	1.00	1.01	0.86	0.76
TIP4P-D	1.56	1.34	1.61	1.10
MBH:	A99SB-ILDN	A99SB*-ILDN	C27	C22*
TIP3P	1.48	1.18	1.41	0.50
SPC/E	0.49	1.19	1.92	1.37
TIP4P-D	0.63	0.36	0.37	0.24
Overall:	A99SB-ILDN	A99SB*-ILDN	C27	C22*
TIP3P	3.27	4.20	4.21	2.82
SPC/E	4.25	2.82	5.36	4.83
TIP4P-D	4.37	3.47	3.47	4.47

^aThe maximum of each experimental spectrum is set to be 1.0, but this scale is arbitrary. The overall rank is the sum of rmsd of SSP and SPS data. A low rmsd indicates good agreement between theory and experiment.

Table 2 indicates that overall the best force field combinations are A99SB*-ILDN with SPC/E and C22* with TIP3P. The combination of A99SB*-ILDN with SPC/E is exceptional for alpha helices specifically (A1.2 and MPX) but not remarkable for the beta hairpins (MBH and TZ2). Therefore, we recommend this combination for the simulation of systems where helices are expected to dominate the solvated conformational ensemble. Alternatively, C22* with TIP3P displays generally good performance across all tested peptides. If a simulated system contains beta-like structures or a mixture of structures in its solution state ensemble, we recommend this force field combination. The A99SB*-ILDN and the C22* force field are specifically balanced to yield a good representation of the helix-coil transition.¹⁸ Therefore, the good performance of these force fields is not surprising.

On the other hand, the outstanding performance of the TIP4P-D water model for many cases is surprising. An increase in London dispersion interactions of water molecules by using the TIP4P-D water model tends to destabilize the folded states of most peptides. Therefore, the good performance of TIP4P-D may indicate an over-stabilization of folded states at the air-water interface for most protein force fields.

3. CONCLUSION

These findings indicate that force field selection is of utmost importance for the simulation of peptides at air-water interfaces. Moreover, enhanced sampling, in the form of WT-MetaD is necessary to overcome free energy barriers to produce a diverse ensemble of interfacial structures. The method of sampling structural ensembles using molecular simulation and comparing those ensembles to spectral data has been demonstrated to be a feasible and useful technique for generating hypothetical molecular-level structures for interfacial peptides. The combination of molecular simulation and VSFG can thus be symbiotic and helpful for researchers in determining specific, detailed peptide structures provided that the selection of force fields is correctly considered.

Based upon the findings for these four model peptides, we provide some suggestions for researchers interested in performing MD simulations or VSFG experiments of peptides or proteins at an air-water or hydrophobic interface.

- (1) WT-MetaD simulations are a good way to generate structural ensembles for adding atomistic detail to VSFG and other experimental data at the air-water interface. This can be used to generate or test hypothetical structures to be considered together with experimental data. Other enhanced-sampling techniques may need to be explored to overcome hidden free energy barriers in systems that become stuck in certain conformations.
- (2) The C22*/TIP3P and A99SB*-ILDN/SPC/E force field combinations produce structural ensembles whose simulated spectra best match experimental data. Thus, we suggest using C22*/TIP3P for simulations of beta peptides and systems with uncertain or mixed structures. We suggest A99SB*-ILDN/SPC/E for simulations of helical peptides.
- (3) Unbiased molecular dynamics simulations are sometimes insufficient to explore the entire conformational ensemble. Clustering analysis shows that a peptide's interfacial structural ensemble can be diverse. Therefore, we suggest using some enhanced sampling such as

metadynamics or replica exchange in order to generate realistic ensembles.

- (4) The secondary structure of some peptides is greatly affected by the air–water interface. This should be considered in industrial or medical applications in which air–water interfaces are introduced such as in fermentation, filtration, lyophilization, cell culture, protein purification, and protein storage.

These findings are based upon the assumption that the peptides are largely independent of each other within experiments as we know they are in the simulation box. The introduction of high concentrations of peptides at the air–water interface would likely have the effect of causing the peptides to aggregate and move into solution because of strong hydrophobic interactions, as is found in LK α 14 studies.²⁶ In the future, this assumption might be tested by generating structural ensembles from other methods such as parallel bias metadynamics,²⁷ solvent tempering,^{8b} or parallel tempering^{7a} with guidance from the protocols presented herein. We plan future work to reweight these ensembles using a Monte Carlo approach. In this way, we will adjust the weights of each ensemble member by iteratively replacing members to minimize the difference between calculated and experimental spectra. This could help overcome sampling inaccuracies imposed by the force field.

4. METHODS

Vibrational Sum-Frequency Generation Spectroscopy. The peptide structure at the air–water interface was probed with VSFG spectroscopy. A regenerative amplifier (*Spitfire Ace PA*, Spectra-Physics, Santa Clara, CA, USA) produced 800 nm pulses with an energy of 9 mJ per pulse at a repetition rate of 1 kHz. One mJ was directed through an etalon to produce a narrow-band laser beam (VIS) with a full width at half-maximum (fwhm) of 20 cm⁻¹. Another 1 mJ was directed into a parametric amplifier (*TOPAS/NDFG*, Light Conversion, Vilnius, Lithuania). The output of the parametric amplifier was an infrared laser beam (IR) with a fwhm of 150 cm⁻¹ at a central wavenumber of 1650 cm⁻¹.

The polarization of both the IR and the VIS beam was controlled by using a polarizer/half-wave plate combination. Overlapping the IR and VIS beam spatially and temporally on the sample surface at angles of incidence of 60 and 55°, respectively, produced a sum-frequency response. This signal was then analyzed with a spectrometer (*Shamrock*, Andor, Belfast, UK) coupled to a CCD camera (*Newton 970*, Andor). Measurements were performed under an atmosphere of dry nitrogen.

The samples were prepared as follows: a Teflon trough was filled with a solution of NaCl in D₂O (40 mL, 150 mM NaCl) and placed on a motorized rotation stage. Surface tension was recorded with a Kibron DeltaPi tensiometer (Helsinki, Finland), calibrated on the NaCl solution. While measuring the surface tension, a peptide solution (1 mL, 1.4 mM aurein 1.2 in D₂O, 150 mM NaCl) was injected through the water surface of the NaCl solution. The peptide layer adsorbed to the air–water interface was assumed to be equilibrated when the surface tension did not change significantly for half an hour. The amount of peptide added to the solution was selected to be just enough to form a monolayer as to exclude the possibility of multiple layers forming at the interface. After the equilibration, VSFG spectra were collected. All data was

acquired for 60 min at 22 °C in the range of 1550–1750 cm⁻¹. The spectra were background corrected by subtracting a spectrum with IR beam blocked and only the VIS beam incident on the sample. Subsequently, spectra were normalized by a background-corrected quartz spectrum. The SFG intensity was fit using the relationship

$$I_{\text{SF}} \propto \left| \sum_i \frac{C_i}{\omega_i - \omega_{\text{IR}} - i\Gamma_i} + |\chi_{\text{NR}}^{(2)}| e^{i\phi} \right|^2 \quad (1)$$

where C is a constant that depends on the Raman- and infrared transition dipole moments, ω_i is the resonance frequency of a particular transition i , Γ is the dipole dephasing rate, and $\chi_{\text{NR}}^{(2)}$ is the nonresonant sum-frequency response of phase ϕ .

Molecular Dynamics Simulations. Simulation boxes were constructed by placing the NMR structure of the peptide with a random spatial orientation at the edge of a water box. These water boxes were cubic with a side length of 2.4 nm longer than the longest axis of the peptide, and 150 mM sodium chloride was added along with any additional ions needed to neutralize the system. Then, the dimension of the box normal to the peptide-containing interface was expanded by a factor of 3 to introduce a vacuum, which is used as a proxy for air. Boxes are simulated with periodic boundary conditions to allow for the application of the particle mesh Ewald sum method for long-range electrostatic calculations.²⁸ Such a simulation setup produces infinite slabs of water separated by several nanometers in space, which is large enough for the electrostatic interactions among the slabs to largely decay. GROMACS 4.6 patched with PLUMED 2.1 was used for all simulations.²⁹ Bonds were constrained using the LINCS algorithm to allow for a numerical integration time step of 2 fs.³⁰ The temperature was held at 300 K using a stochastic velocity-rescaling thermostat.³¹ Lennard-Jones and electrostatic interactions were cutoff at 1.0 nm with a Verlet integration scheme, and long-range electrostatics were handled with the particle mesh Ewald summation method. Steepest descent minimization of 1000 steps was used to relax the system before subsequent production simulations in the canonical ensemble. WT-MetaD simulations were 1 μ s in length.

Metadynamics simulations were facilitated using PLUMED 2.1. In short, the bias potential V , given by

$$V[\mathbf{S}, t] = \sum_{k=1}^{t-1} W e^{-\sum_{i=1}^d \frac{(S_i - S_i[\mathbf{x}(k\tau)])^2}{2\sigma_i^2}} \quad (2)$$

is constructed during the simulation by accumulating small Gaussian hills of potential with a width of σ that are deposited periodically to slow degrees of freedom known as collective variables (CVs) denoted by the letter S . The height of the Gaussian kernel is initially W , but in WT-MetaD subsequent hills decrease in height according to

$$W(k\tau) = W_0 e^{-\frac{V(\mathbf{S}[\mathbf{x}(k\tau)], k\tau)}{k_B \Delta T}} \quad (3)$$

The decrease in hill height is a function of the magnitude of the bias previously added and a temperature parameter ΔT (not to be confused with the temperature of the simulation T) which controls the exponential decay of the hill height. Thus, the system is lifted out of free-energy minima, and eventually the full conformational landscape of these CVs can be sampled. The parameters for the size of the hills are $W = 2.0$ kJ/mol to

be applied every picosecond, and $\Delta T = 2700$ K. The width of the Gaussians was set differently for each model peptide (A1.2 - $\sigma_g = 0.03$, $\sigma_h = 0.1$, MPX - $\sigma_g = 0.03$, $\sigma_h = 0.1$, TZ2 - $\sigma_g = 0.03$, $\sigma_h = 0.05$, MBH - $\sigma_g = 0.01$, $\sigma_h = 0.05$ for radius of gyration and hydrogen bonds respectively) in CV space. The widths of the Gaussians were determined by using unbiased 100 ns simulations of the peptides. The values of the CVs were tracked over the simulation time, the standard deviations of the CVs were calculated, and σ was set to approximately half the standard deviation rounded to a reasonable number of significant digits. For these simulations, we selected the α carbon ($C\alpha$) radius of gyration and the numbers of structural hydrogen bonds as our two CVs. The radius of gyration is

$$g = \left(\frac{\sum_i^n m_i |r_i - r_{COM}|^2}{\sum_i^n m_i} \right)^{1/2}$$

with $r_{COM} = \frac{\sum_i^n m_i r_i}{\sum_i^n m_i}$ (4)

where m_i is the mass, and r_i is the position of atom i . r_{COM} is the position of the center of mass.

Structural hydrogen bonds are those found in the experimental solution NMR plus any other possible hydrogen bonds that would arise in a perfect α helix. The presence of a hydrogen bond was determined by applying a sigmoidal function s_{ij} to the distance between the hydrogen and oxygen in question

$$s_{ij} = \frac{1 - \left(\frac{r_{ij}}{r_0}\right)^n}{1 - \left(\frac{r_{ij}}{r_0}\right)^m}$$
 (5)

where s_{ij} decays from one to zero as the distance between the two atoms i and j grows. Constants $m = 8$, $n = 6$, and $r_0 = 0.25$ nm were used for all peptides. The CV biased during the simulations was the sum of s_{ij} over all defined hydrogen bonds. From the bias applied to these CVs during a WT-MetaD simulation, the free-energy landscape can be computed according to

$$V(\mathbf{S}, t \rightarrow \infty) = -\frac{\Delta T}{T + \Delta T} F(\mathbf{S}) + C$$
 (6)

Clustering. The trajectory was clustered using a sample of 20,000 of the 1,000,000 frames from the simulation with the `g_cluster` tool within GROMACS 4.6 and the GROMOS method.³² This smaller sample was used, because the calculation time and memory required to construct the necessary root-mean-square displacement matrix among the frames scale roughly with the number of frames squared. Thus, clustering of the whole trajectory would be around 2500 times costlier than the reduced set. The cutoff for cluster members in $C\alpha$ -rmsd space was set to different values for each of the four model peptides (A1.2 - 0.35 nm, MPX - 0.35 nm, TZ2 - 0.33 nm, MBH - 0.2 nm) with a goal of keeping the number of clusters near ten. The remaining 980,000 frames were then compared to the central member of the top ten existing clusters and assigned to the first cluster within the cutoff. If no cluster was within the cutoff, the frame was assigned to a separate “junk” cluster that was not considered in the subsequent clustering analysis. Using the Torrie-Valleau method,²⁵ weights were assigned to frames in the trajectory

based upon the metadynamics bias applied during a given frame. Frames were taken after the transient period where the majority of the metadynamics bias was applied. Each cluster was assigned a weight by the sum of the weights of these frames.

Simulating VSG Spectra. Simulation frames were selected from the whole ensemble of conformations with a probability proportional to the frame weight, assigned by the Torrie-Valleau method. These random frames were included in eight samples of 25 frames in total. Through some trial and error, we determined that 25 was the optimum number of frames to consider. Spectral calculations were much slower on larger sample sizes, and these larger samples did not significantly change the results. The sample of 25 peptide structures was arranged in an array in a single pdb file. The pdb file was used to compute a VSG spectrum using the method of Roeters et al.^{3b} Roeters’ model shows that for larger, less structurally disordered proteins, theoretical and experimental spectra match well. Since his model accurately predicts spectral features of large, complex proteins, we expect that for our simple peptides, which contain a small fraction of the complexity of Roeters’ larger protein, the selected theoretical model will also be accurate. The method calculates the second-order nonlinear susceptibility $\chi^{(2)}$. The calculation includes nearest- and non-nearest neighbor coupling and intrapeptide hydrogen bond effects. Hydrogen bonds between the peptide and solvent are not considered due to limits on data storage, because of the high frequency with which the trajectory is written. Nearest-neighbor coupling is included in the calculation by a map of the dihedral angle dependent coupling. The coupling was obtained from literature values calculated by *ab initio* methods at the 6-31G+(d) B3LYP level of theory.³³ In these calculations, glycine dipeptide was used as a model for amide I coupling. Results are tabulated and available on the Web site associated with ref 33b. Non-nearest neighbor couplings were estimated using transition-dipole interactions rather than the *ab initio* map used for nearest neighbor coupling.³⁴

■ ASSOCIATED CONTENT

📄 Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.8b00840.

Free energy surfaces, fitting parameters, convergence (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: helmut.lutz@tum.de (H.L.).

*E-mail: vance.jaeger@louisville.edu (V.J.).

ORCID

Helmut Lutz: 0000-0002-7915-8542

Tobias Weidner: 0000-0002-7083-7004

Bert L. de Groot: 0000-0003-3570-3534

Present Addresses

[§]Theoretical Chemical Biology and Protein Modeling Group, Technical University of Munich, Freising, Germany.

[¶]Department of Chemical Engineering, University of Louisville, Louisville, Kentucky, USA.

Author Contributions

#H.L. and V.J. contributed equally. The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Funding

This work was supported by the Max Planck Society and the Alexander von Humboldt Foundation (V.J.). H.L. and T.W. acknowledge the European Commission (CIG grant #322124) and the Deutsche Forschungsgemeinschaft (WE4478/2-1) for financial support.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Yuki Nagata whose questions inspired this work and for his helpful discussions along the way.

ABBREVIATIONS

VSFG, vibrational sum frequency generation; SFG, sum frequency generation; MD, molecular dynamics; WT-MetaD, well-tempered metadynamics; CV, collective variable; A1.2, aurein 1.2; MBH, minimal beta hairpin; TZ2, tryptophan zipper 2; MPX, mastoparan X.

REFERENCES

- (1) PDB statistics; 2017. <http://pdbeta.rcsb.org/pdb/statistics/holdings.do> (accessed Dec 9, 2018).
- (2) (a) Mauri, S.; Weidner, T.; Arnolds, H. The structure of insulin at the air/water interface: monomers or dimers? *Phys. Chem. Chem. Phys.* **2014**, *16* (48), 26722–26724. (b) Rodríguez Patino, J. M.; Carrera Sánchez, C.; Rodríguez Niño, M. R. Implications of interfacial characteristics of food foaming agents in foam formulations. *Adv. Colloid Interface Sci.* **2008**, *140* (2), 95–113. (c) Meister, K.; Bäumer, A.; Szilvay, G. R.; Paananen, A.; Bakker, H. J. Self-Assembly and Conformational Changes of Hydrophobin Classes at the Air–Water Interface. *J. Phys. Chem. Lett.* **2016**, *7* (20), 4067–4071. (d) Lutz, H.; Jaeger, V.; Berger, R.; Bonn, M.; Pfaendtner, J.; Weidner, T. Biomimetic Growth of Ultrathin Silica Sheets Using Artificial Amphiphilic Peptides. *Adv. Mater. Interfaces* **2015**, *2* (17), 1500282.
- (3) (a) Nguyen, K. T.; King, J. T.; Chen, Z. Orientation Determination of Interfacial beta-Sheet Structures in Situ. *J. Phys. Chem. B* **2010**, *114* (25), 8291–8300. (b) 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. (c) Hennig, R.; Heidrich, J.; Saur, M.; Schmäser, L.; Roeters, S. J.; Hellmann, N.; Woutersen, S.; Bonn, M.; Weidner, T.; Markl, J.; Schneider, D. IM30 triggers membrane fusion in cyanobacteria and chloroplasts. *Nat. Commun.* **2015**, *6*, 7018.
- (4) (a) Schach, D.; Globisch, C.; Roeters, S. J.; Woutersen, S.; Fuchs, A.; Weiss, C. K.; Backus, E. H. G.; Landfester, K.; Bonn, M.; Peter, C.; Weidner, T. Sticky water surfaces: Helix-coil transitions suppressed in a cell-penetrating peptide at the air-water interface. *J. Chem. Phys.* **2014**, *141* (22), 22D517. (b) Apte, J. S.; Collier, G.; Latour, R. A.; Gamble, L. J.; Castner, D. G. XPS and ToF-SIMS Investigation of α -Helical and β -Strand Peptide Adsorption onto SAMs. *Langmuir* **2010**, *26* (5), 3423–3432.
- (5) (a) Wang, J.; Buck, S. M.; Chen, Z. The effect of surface coverage on conformational changes of bovine serum albumin molecules at the air–solution interface detected by sum frequency generation vibrational spectroscopy. *Analyst* **2003**, *128* (6), 773–778. (b) Yano, Y. F. Kinetics of protein unfolding at interfaces. *J. Phys.: Condens. Matter* **2012**, *24* (50), 503101.
- (6) (a) Dalgicdir, C.; Sayar, M. Conformation and Aggregation of LK α 14 Peptide in Bulk Water and at the Air/Water Interface. *J. Phys. Chem. B* **2015**, *119* (49), 15164–15175. (b) Anderson, R. E.; Pande, V. S.; Radke, C. J. Dynamic lattice Monte Carlo simulation of a model protein at an oil/water interface. *J. Chem. Phys.* **2000**, *112* (20), 9167–9185. (c) 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.

- (7) (a) Neal, R. M. Sampling from multimodal distributions using tempered transitions. *Stat. Comput.* **1996**, *6* (4), 353–366. (b) Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314* (1–2), 141–151.
- (8) (a) Wright, L. B.; Walsh, T. R. Efficient conformational sampling of peptides adsorbed onto inorganic surfaces: insights from a quartz binding peptide. *Phys. Chem. Chem. Phys.* **2013**, *15* (13), 4715–4726. (b) 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. U. S. A.* **2005**, *102* (39), 13749–13754.
- (9) (a) Sprenger, K. G.; He, Y.; Pfaendtner, J. Probing How Defects in Self-assembled Monolayers Affect Peptide Adsorption with Molecular Simulation. *Mol. Model. Simul.* **2016**, 21–35. (b) Sprenger, K. G.; Pfaendtner, J. Strong Electrostatic Interactions Lead to Entropically Favorable Binding of Peptides to Charged Surfaces. *Langmuir* **2016**, *32* (22), 5690–5701. (c) Bonomi, M.; Parrinello, M. Enhanced Sampling in the Well-Tempered Ensemble. *Phys. Rev. Lett.* **2010**, *104* (19), 190601.
- (10) Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **2008**, *100* (2), 020603.
- (11) (a) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules (vol 117, pg 5179, 1995). *J. Am. Chem. Soc.* **1996**, *118* (9), 2309–2309. (b) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102* (18), 3586–3616.
- (12) Shirts, M. R.; Pande, V. S. Solvation free energies of amino acid side chain analogs for common molecular mechanics water models. *J. Chem. Phys.* **2005**, *122* (13), 134508.
- (13) Nagata, Y.; Ohto, T.; Backus, E. H. G.; Bonn, M. Molecular Modeling of Water Interfaces: From Molecular Spectroscopy to Thermodynamics. *J. Phys. Chem. B* **2016**, *120* (16), 3785–3796.
- (14) Bjelkmar, P.; Larsson, P.; Cuendet, M. A.; Hess, B.; Lindahl, E. Implementation of the CHARMM Force Field in GROMACS: Analysis of Protein Stability Effects from Correction Maps, Virtual Interaction Sites, and Water Models. *J. Chem. Theory Comput.* **2010**, *6* (2), 459–466.
- (15) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Struct., Funct., Genet.* **2010**, *78* (8), 1950–1958.
- (16) Best, R. B.; Hummer, G. Optimized Molecular Dynamics Force Fields Applied to the Helix-Coil Transition of Polypeptides. *J. Phys. Chem. B* **2009**, *113* (26), 9004–9015.
- (17) Mackerell, A. D.; Feig, M.; Brooks, C. L. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **2004**, *25* (11), 1400–1415.
- (18) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. How Robust Are Protein Folding Simulations with Respect to Force Field Parameterization? *Biophys. J.* **2011**, *100* (9), L47–L49.
- (19) 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.

(20) 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.

(21) 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.

(22) Mackerell, A. D. Empirical force fields for biological macromolecules: Overview and issues. *J. Comput. Chem.* **2004**, *25* (13), 1584–1604.

(23) Wang, G. S.; Li, Y. F.; Li, X. Correlation of three-dimensional structures with the antibacterial activity of a group of peptides designed based on a nontoxic bacterial membrane anchor. *J. Biol. Chem.* **2005**, *280* (7), 5803–5811.

(24) (a) Barth, A. Infrared spectroscopy of proteins. *Biochim. Biophys. Acta, Bioenerg.* **2007**, *1767* (9), 1073–1101. (b) Venyaminov, S. Y.; Kalnin, N. N. Quantitative IR spectrophotometry of peptide compounds in water (H₂O) solutions. I. Spectral parameters of amino acid residue absorption bands. *Biopolymers* **1990**, *30* (13–14), 1243–1257. (c) Sjöberg, B.; Foley, S.; Cardey, B.; Enescu, M. An experimental and theoretical study of the amino acid side chain Raman bands in proteins. *Spectrochim. Acta, Part A* **2014**, *128*, 300–311.

(25) Torrie, G. M.; Valleau, J. P. Non-Physical Sampling Distributions in Monte-Carlo Free-Energy Estimation - Umbrella Sampling. *J. Comput. Phys.* **1977**, *23* (2), 187–199.

(26) 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* (43), 15134–15137.

(27) 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.

(28) 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.

(29) (a) Pronk, S.; Pall, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, *29* (7), 845–854. (b) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **2014**, *185* (2), 604–613.

(30) Hess, B. P-LINCS: A parallel linear constraint solver for molecular simulation. *J. Chem. Theory Comput.* **2008**, *4* (1), 116–122.

(31) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126* (1), 014101.

(32) Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E. Peptide folding: When simulation meets experiment. *Angew. Chem., Int. Ed.* **1999**, *38* (1–2), 236–240.

(33) (a) 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. (b) Hamm, P.; Zanni, M. *Concepts and Methods of 2D Infrared Spectroscopy*; Cambridge University Press: Cambridge, 2011; DOI: [10.1017/CBO9780511675935](https://doi.org/10.1017/CBO9780511675935).

(34) (a) Krimm, S.; Abe, Y. Intermolecular Interaction Effects in the Amide I Vibrations of β Polypeptides. *Proc. Natl. Acad. Sci. U. S. A.* **1972**, *69* (10), 2788–2792. (b) Torii, H.; Tasumi, M. Ab initio molecular orbital study of the amide I vibrational interactions between the peptide groups in di- and tripeptides and considerations on the conformation of the extended helix. *J. Raman Spectrosc.* **1998**, *29* (1), 81–86.