

UC Santa Barbara

UC Santa Barbara Electronic Theses and Dissertations

Title

A molecular-scale investigation of water's response to interfaces: towards rational design of water-mediated interactions

Permalink

<https://escholarship.org/uc/item/92d099n9>

Author

Monroe, Jacob Ian

Publication Date

2019

Peer reviewed|Thesis/dissertation

University of California
Santa Barbara

**A molecular-scale investigation of water's response
to interfaces: towards rational design of
water-mediated interactions**

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

Doctor of Philosophy
in
Chemical Engineering

by

Jacob I. Monroe

Committee in charge:

Professor M. Scott Shell, Chair
Professor Baron Peters
Professor Rachel Segalman
Professor Kevin Plaxco
Professor Joan-Emma Shea

September 2019

The Dissertation of Jacob I. Monroe is approved.

Professor Baron Peters

Professor Rachel Segalman

Professor Kevin Plaxco

Professor Joan-Emma Shea

Professor M. Scott Shell, Committee Chair

July 2019

A molecular-scale investigation of water's response to interfaces: towards rational design
of water-mediated interactions

Copyright © 2019

by

Jacob I. Monroe

Acknowledgements

None of this work would exist if it were not for a tremendous amount of assistance from many friends, family, teachers, coworkers, and collaborators, not only during my time as a graduate student, but throughout my entire life.

First and foremost, I am truly indebted to M. Scott Shell for all that he has done in guiding and advising my research. Throughout my time as a graduate student, he has been my closest collaborator and most valuable resource. Thank you, Scott!

I have also been lucky enough to collaborate closely with a number of faculty, post-docs, and graduate students who have freely shared both their advice and an infectious excitement for their research. This includes my committee, who have provided frequent and pertinent suggestions concerning both the direction of my studies as well as my future career.

I am deeply indebted to Markus Valtiner, Thomas Utzig, and Philipp Stock for sharing with me their knowledge and expertise concerning atomic force microscopy and for many stimulating discussions that profoundly shaped Chapter 2.

Songi Han and Alex Schrader are very much responsible for directing my attention towards the manipulation of water dynamics at interfaces, with many of their insightful thoughts concerning this matter making their way into Chapter 4. Continued collaboration with Songi has been exciting and very fruitful, and I am thankful to her for often stepping in as an informal advisor.

In the final year of my PhD, I have been thankful for the guidance provided by the leadership of the Center for Materials for Water and Energy Systems: Benny Freeman, Rachel Segalman, and Lynn Katz. All of you, especially Lynn, have played a prominent role in directing the work presented in Chapter 5. A direct result of your leadership, and conspiracy with Scott and Songi, has been my continued growth as a researcher despite

my attempts to relax and ride my previous accomplishments in this final year.

Before starting at UCSB, I was lucky enough to work with Michael Shirts as an undergraduate. His continued career advice and mentorship has been greatly appreciated, and his influence from my time working with him many years ago is nonetheless readily apparent in Chapter 5.

My time at UCSB has been both pleasant and intellectually stimulating. There are far too many people to thank for this, both within and outside of the Chemical Engineering Department, but I am particularly indebted to the other members of my cohort, whose comradery was essential to making it through classes, as well as colleagues in the Shell and Peters groups with whom I have shared office space. I have enjoyed countless discussions at white boards with members of both groups and will sorely miss these interactions, which have resulted in many good ideas that I subsequently appropriated into publishable material. Of particular importance to this thesis, the ideas of Dave Smith were instrumental in creating Chapters 2 and 3, and I am indebted to Dennis Robinson Brown and Sally Jiao for contributions to Chapter 5, including Sally's development of the linear regression model that was incorporated into the genetic algorithm described therein.

Finally, I would like to thank all of my friends, many of whom are also my colleagues, as well as my family, mainly for simply putting up with me. My obsession with my work has certainly put strain on relationships at times, but I have always felt supported and am appreciative for all of your sacrifices and love. I love all of you and only hope that you will continue to forgive my compulsive tendencies and always help me to experience aspects of life outside of work.

Curriculum Vitæ

Jacob I. Monroe

Education

Ph.D. in Chemical Engineering, University of California, Santa Barbara 2019
B.S. University of Virginia 2014

Awards, Honors, Grants

UCSB Graduate Opportunity Fellowship - 2018
AIChE COMSEF Graduate Student Award - 2017
NSF Graduate Research Fellowship - 2015
UVA NanoStar Undergraduate Summer Research Grant - 2012 and 2013
Omega Chi Epsilon Member - since 2012
Tau Beta Pi Member - since 2013

Publications

Monroe, J. I.; Jiao, S.; Brown, D. R.; Katz, L.; Shell, M. S. “Chemically-patterned interfaces for the controlled affinity and selectivity of small-molecule solutes.” In preparation.

Monroe, J. I.; Shell, M. S. “Decoding signatures of structure, bulk thermodynamics, and solvation in three-body angle distributions of rigid water models.” *J. Chem. Phys.* In review.

Braun, E.; Gilmer, J.; Mayes, H. B.; Mobley, D. L.; **Monroe, J. I.**; Prasad, S.; Zuckerman, D. M. “Best Practices for foundations in molecular simulations.” *LiveCoMS*. 2019, 1 (1). <https://doi.org/10.33011/livecoms.1.1.5957>.

Monroe, J. I.; Shell, M. S. “Computational Discovery of Chemically Patterned Surfaces That Effect Unique Hydration Water Dynamics.” *Proc. Natl. Acad. Sci.* 2018, 115, 8093–8098.

Schrader, A. M.; **Monroe, J. I.**; Sheil, R.; Dobbs, H. A.; Keller, T. J.; Li, Y.; Jain, S.; Shell, M. S.; Israelachvili, J. N.; Han, S. “Surface Chemical Heterogeneity Modulates Silica Surface Hydration.” *Proc. Natl. Acad. Sci.* 2018, 115, 2890–2895.

Stock, P.*; **Monroe, J. I.***; Utzig, T.; Smith, D. J.; Shell, M. S.; Valtiner, M. “Unraveling Hydrophobic Interactions at the Molecular Scale Using Force Spectroscopy and Molecular Dynamics Simulations.” *ACS Nano* 2017, 11, 2586–2597.

- Robinson, M. K.*; **Monroe, J. I.***; Shell, M. S. “Are AMBER Force Fields and Implicit Solvation Models Additive? A Folding Study with a Balanced Peptide Test Set.” *J. Chem. Theory Comput.* 2016, 12, 5631–5642.
- Monroe, J. I.**; Shirts, M. R. “Converging Free Energies of Binding in Cucurbit[7]Uril and Octa-Acid Host–Guest Systems from SAMPL4 Using Expanded Ensemble Simulations.” *J. Comput. Aided. Mol. Des.* 2014, 28, 401–415.
- Monroe, J. I.**; El-Nahal, W. G.; Shirts, M. R. “Investigating the Mutation Resistance of Nonnucleoside Inhibitors of HIV-RT Using Multiple Microsecond Atomistic Simulations.” *Proteins Struct. Funct. Bioinforma.* 2014, 82, 130–144.

*Authors contributed equally to this work.

Oral Presentations

- Monroe, J. I.**, Shell, M. S. (2018, July). “Chemical patterning of heterogeneous surfaces induces unique dynamics in proximal water.” Gordon Research Symposium for Water and Aqueous Solutions, Holderness, NH.
- Monroe, J. I.**, Schrader, A. M., Han, S., Shell, M. S. (2017, November). “Tuning proximal water diffusion via silanol patterning on quartz surfaces.” AICHE Annual Meeting, Minneapolis, MN.
- Monroe, J. I.**, Schrader, A. M., Han, S., Shell, M. S. (2017, August). “Tuning proximal water diffusion via silanol patterning on quartz surfaces.” ACS National Meeting in Washington, DC.
- Monroe, J. I.**, Stock, P., Utzig, T., Smith, D. J., Valtiner, M., Shell, M. S. (2016, November). “Single-molecule hydrophobic interactions: from experiment to simulation and equilibrium to non-equilibrium.” AICHE Annual Meeting, San Francisco, CA.

Poster Presentations

- Monroe, J. I.**, Shell, M. S. (2018, July). “Chemical patterning of heterogeneous surfaces induces unique dynamics in proximal water.” Gordon Research Symposium for Water and Aqueous Solutions, Holderness, NH.
- Monroe, J. I.**, Stock, P., Utzig, T., Smith, D. J., Valtiner, M., Shell, M. S. (2017, November). “Scaling of peptide sequence-dependent hydrophobic interactions from experiment and simulation.” AICHE Annual Meeting, Minneapolis, MN.
- Monroe, J. I.**, Stock, P., Utzig, T., Smith, D. J., Valtiner, M., Shell, M. S. (2016, October). “Scaling of single-molecule hydrophobic interactions from experiment and simulation.” UCSB Department of Chemical Engineering Graduate Student Symposium, Santa Barbara, CA.

Research Experience

Graduate Researcher, UCSB Department of Chemical Engineering 2014 - Present
Undergraduate Researcher, UVA Department of Chemical Engineering 2012 - 2014
Protein Purifications Intern, MedImmune, Gaithersburg, MD Summer 2013

Teaching Experience

Instructor for School for Scientific Thought Program, UCSB Fall 2017
Graduate Teaching Assistant, UCSB Spring and Fall 2016, Spring 2017
Undergraduate Teaching Assistant, UVA Spring 2013 and 2014

Leadership, Professional, and Outreach Activities

Participant in MOLSSI Workshop for Best Practices in Molecular Simulation August 2017
Organizing Committee for the Graduate Simulation Seminar Series 2015 - 2016
Student Organizer for Prospective Student Visits Spring 2016
Participant in FUSE Program, CSEP at UCSB 2014 - Present
OXE UVA Chapter President 2013 - 2014
AIChE Member 2011 - Present
ACS Member 2017 - Present

Abstract

A molecular-scale investigation of water's response to interfaces: towards rational design
of water-mediated interactions

by

Jacob I. Monroe

Understanding and manipulating the behavior of water is a crucial component of designing materials at interfaces (i.e. for use in membranes, filters, chromatographic columns, etc.) as well as materials based on self-assembling molecules (i.e. micelles, hydrogels, etc.). It has long been known that water plays a significant role in mediating surface adsorption as well as interactions between biomolecules, but effective strategies to modify this behavior based on atomistic physical intuition has been lacking. Many past efforts have focused on unrealistic, toy systems in order to develop an extensive theory of water behavior in idealized inhomogeneous environments. Here I put this theory to the test in realistic yet simple and well-controlled molecular simulations, beginning the process of identifying practical extensions to inform the design of real materials.

To isolate and understand key water physics, I investigate ever-important hydrophobic interactions and explore the thermodynamic information contained within water's structural response to hydrophobes. Specifically, consideration of the behavior of three-body angle distributions, which relate to the tetrahedral structure of water, allows for the construction of simple thermodynamic models for adsorption and solvation. I show that, in general, the success of such thermodynamic models may be attributed to fundamental connections between perturbations to water structure and the relative entropy of a solvation process, which includes indirect thermodynamic contributions associated with solvent restructuring.

In proteins and other biological settings, hydrophobic regions do not occur in isolation, but instead are part of heterogeneous interfaces involving a variety of chemical groups. To study the role of chemical patterning in inducing unique water behavior at such surfaces, I present methods coupling a genetic algorithm to molecular dynamics simulations to optimize arbitrary surface properties through the reorganization of functional groups. Water mobility near an interface is optimized in this way, ultimately revealing novel relationships between water dynamics, structure, and thermodynamics. I also use this algorithm to discover interfacial chemical patterns that repel or attract a diverse set of small-molecule solutes, as might be necessary to prevent fouling during water purification processes or selectively capture desirable products during chromatography. Overall, I extend our current knowledge of water's molecular-scale response to both solutes and macroscopic interfaces, effectively expanding the design space for interfacial materials in terms of these atomistic-level insights.

Contents

Abstract	ix
List of Figures	xiii
List of Tables	xxx
1 Introduction	1
1.1 Overview	1
1.2 Designing interfacial and self-assembling materials	2
1.3 Water-mediated interactions	5
1.4 Outline	6
References	8
2 Scaling of the hydrophobic interaction between short peptides and extended hydrophobic surfaces	12
2.1 Introduction	12
2.2 Methods	14
2.3 Results and Discussion	19
2.4 Conclusions	39
References	41
3 Characterization of three-body angle distributions in bulk and solvation shells of model solutes	48
3.1 Introduction	48
3.2 Methods	56
3.3 Results and Discussion	59
References	83
4 Computational design of surface chemical patterns to optimize proximal water mobility	90
4.1 Introduction	90
4.2 Methods	94

4.3	Results and Discussion	102
4.4	Conclusions	114
	References	116
5	Chemically-patterned interfaces for the controlled affinity and selectivity of small-molecule solutes	122
5.1	Introduction	122
5.2	Methods	128
5.3	Results and Discussion	135
5.4	Conclusions	158
	References	161
6	Conclusions	168
6.1	Summary	168
6.2	Outlook	171
	References	175
A	Characterization of peptide structure in equilibrium and non-equilibrium simulations	176
B	Derivation of solvation-thermodynamic relations in terms of the relative entropy	184
B.1	Free energy of solvating colloids	184
B.2	Free energy to solvate a hard-sphere colloid in a monatomic ideal gas . .	186
C	Additional structural and thermodynamic data associated with water three-body angle distributions	189
	References	205
D	Additivity of dynamics in simple Markov models	206
E	Genetic algorithm validation and additional, spatially-resolved measures of surface hydrophilicity and water mobility	210
	References	222
F	Relating Henry's Law constants to free energies of solvation	223
	References	227
G	Additional data and figures associated with solute affinities for interfaces	228
	References	237

List of Figures

- 2.1 The dimensions of the simulation box are initially set to $60 \times 60 \times 100 \text{ \AA}$, with pressure equilibration resulting in an increase of the box height to around 103 \AA . The x- and y-coordinates of the system are defined to be in the plane of the surface, with the z-coordinate normal to the SAM. In all simulations the peptide is pulled along only the z-coordinate, as shown in the right figure with waters removed. 15
- 2.2 Schematic of (a) experimental and (b) computational setting used for quantifying dynamic hydrophobic interactions between a set of different hydrophobic peptides (shown in Table 2.1) and an extended hydrophobic surface as a function of the experimental distance DX and the simulation distance DS , respectively. (c) Hydrophobic mutations are introduced specifically at the C-terminal tail end of a hydrophilic $(GS)_n$ based scaffold. In experiment (a) a PEG12-diamine-based polymer linker is used to covalently grafted peptides on a molecularly smooth OH/COOH terminated self-assembled monolayer on a gold surface (Roughness, $\sigma_{RMS} \leq 2 - 3 \text{ \AA}$). The contour length of $L_C = 10 \text{ nm}$ is indicated. Gold-coated AFM tips (radius, $R_{AFM} \sim 15 \text{ nm}$) are hydrophobized with a self-assembled monolayer and interactions between different surface grafted peptide mutations (see Table 2.1) and the hydrophobic tip is measured. In the simulation experiment the PEG linker is not included and the contour length of the individual peptides amounts to $L_C \sim 3.4 \text{ nm}$. In (b) the N-terminal peptide end and applied pulling direction is marked with an arrow. 19

2.3	(a) Typical master curve with about 80 individual rupture events measured between $-N-(GS)3LLLL$ functionalized surface and hydrophobic tip, aligned by the best worm like-chain fit with a contour length of 10 nm and a persistence length of 0.37 nm. Inset shows a typical individual SM-AFM force distance profile with both a primary adhesive minimum and a single molecular rupture signature at 60% full extension of the linker. The fit in the inset shows a fit by an extended DLVO theory. (b) Normalized histogram of the measured work distribution. The inset in (b) shows a plot of Jarzynski's free energy ΔG_0 as a function of the number of force trajectories.	22
2.4	Comparison of experimentally and theoretically estimated interaction free energies as a function of the number of leucines. The interaction free energy is referenced to the one leucine case. For the fast-pulling (10 Å/ns) and slow-pulling (2 Å/ns) MD cases, the R^2 values for the linear fits are 0.71 and 0.80, respectively. All other R^2 values are greater than 0.90.	25
2.5	Potentials of mean force from umbrella sampling are shown with reference to three reaction coordinates: the center of mass of all leucine residues, the alpha-carbon pulled in non-equilibrium runs, and the peptide center of mass in the direction perpendicular to the interface, the z-axis. Distances for each coordinate begin at the SAM surface and are normalized to the peptide contour length. Though the abscissa for the isobutane case has been truncated for clarity, the PMF minimum is much sharper in this case. All error bars are those computed by pymbar. ⁴³	30
2.6	(a) Forward, $P(W_{for})$ (red), and negated reverse, $P(-W_{rev})$ (black), work distributions show that slower pulling rates (2 Å/ns, solid bars) significantly reduce the hysteresis and spread observed at faster pulling speeds (10 Å/ns, lightly shaded). At slower pulling rates, many negative work values (positive here, as negated reverse distributions are shown) are observed during reverse pulling runs. (b) Black dots are force versus distance profiles from single, typical trajectories and red lines are force profiles averaged over all trajectories. Inset structures show peptide chains in green (hydrophilic) and orange (hydrophobic) and pulled atoms (black) at the time of pull-off for all trajectories of a particular peptide. These structures are sized and aligned such that they are proportional to the scale on the x-axis.	32

2.7	Contributions to the work are shown from each pulling regime for slow-pulling runs of the simulated peptides. Pulling regimes correspond to changes in the slope of the work versus distance curve, and are physically defined as: (1) from the beginning of the simulation until the pulled atom is off the surface, (2) until the centers of geometry of one less than the number of non-leucine residues are off the surface, (3) until the last atom in the peptide is off the surface, and (4) from pull-off until the end of the simulation. For all events described, the cut-off for being on the surface was set to a height of 21 Å above the fixed carbons of the dodecane molecules. Without any leucines, the largest contribution to the work appears in regime 1, or the perturbation of the peptides configuration on the surface. With addition of leucines, stronger interactions result in larger contributions to the work after the peptide is extended (regime 3).	33
2.8	Change in average peptide properties moving from a surface-restrained state to solution. (a) With increasing leucines, peptides gain fewer peptide-water and more intramolecular hydrogen bonds in moving from the surface to solution, consistent with the slight increase in α -helical tendency (Figure A.1 in Appendix A). (b) Changes in SASA are similar between peptides, while changes in hydrophobic SASA roughly increase with the number of leucines. Contributions to $\Delta SASA_{hyd}$ are broken down by leucine: LEU1 refers to the leucine closest to the C-terminus of the peptide, LEU2 the next inwards along the chain, and so on. Note that for (GS) ₃ L ₄ , backbone constraints prevent all four leucines from simultaneously becoming solvent-shielded in the surface-associated state (Figure A.1).	35
2.9	(Top panel) Probability density distribution of the triplet angle formed by water molecules with two nearest neighbors, in which the angle is measured at the oxygen site. For comparison, a simple Lennard-Jones fluid with icosahedral coordination has a distribution peaked slightly below 60° with a much smaller secondary peak just above 109°. An ideal gas, non-interacting fluid has a broad distribution with a peak near 90°. (Middle and bottom panels) Changes to the triplet angle distribution for the peptide-SAM systems. The dashed curve in the central figure pertains to hydration waters within 5 Å of the SAM surface, while the dashed curve at bottom refers to waters near a fully solvated isobutane molecule. Other curves give the perturbation to the system distribution for each peptide either restrained on the surface (middle) or far away in solution (bottom panel). Gray arrows show changes with increasing leucines, which, with the exception of the 3 and 4 leucine distribution comparison, are found to be statistically significant at the 5% confidence level by the Kolmogorov-Smirnov test. .	37

3.1	The definition of the three-body angle is shown schematically as the angle between two vectors connecting a central water oxygen atom to oxygens on two of its neighbors within 0.34 nm. For computation of the three-body angle distribution, all angles formed by unique pairs of neighbors within the cutoff are considered, regardless of the number of such angles.	50
3.2	Three-body angle distributions of bulk LJ water (top) and TIP4P/2005 water (bottom) as a function of temperature at fixed density (left) and as a function of density at fixed temperature (right).	60
3.3	Principal components of changes in the three-body angle distributions of LJ and TIP4P/2005 water with varying temperature and density, with the average of all such distributions used as the reference for performing PCA. The percentage of total variance explained by each principal component is shown in the legend, with PC1 and PC2 together capturing over 99%. . .	62
3.4	Projections of three-body angle distribution perturbations onto PC1 and PC2 (shown in Fig. 3.3) are shown as a heat-map in temperature-density space. Negative projections are observed in the regime of structural anomalies, while positive projections in the region of expected simple-fluid behavior, as demarcated by the white contour line at zero. White circles indicate the locus of translational order parameter minima (high-density structural anomaly bound) while white triangles indicate isothermal tetrahedral order parameter maxima (low-density anomaly bound). The dotted white curve on the left indicates the mechanical stability boundary.	64
3.5	(a) Isotherms of excess entropy changes (with the lowest temperature and density as a reference) and (b) the entropy of three-body angle distributions. TIP4P/2005 (circles) show anomalous entropy increases over a range of densities while LJ (triangles) show monotonically decreasing behavior. While excess entropy defines a lower bound to the anomalous region via a minimum, no such minimum is observed for the three-body angle entropy.	66
3.6	(a) Colloid-water oxygen radial distribution functions (shifted so the colloid surface, where the RDF goes to zero, is at zero). (b) Perturbations to bulk three-body angle distributions for waters in the first hydration shell of colloids, with (c) principal components of those perturbations. Variations from dark purple to orange indicate small colloids; dashed orange lines indicate colloids of infinite radius (flat 9-3 walls). The dotted lines in the bottom panel of (c) are the resulting PCs if all colloid three-body angle distributions are considered together instead of separately. The percentage of variance for these first, second, and third PCs of the combined set of colloids are 66.3%, 32.0%, and 1.1%.	69

3.7	<p>Perturbations from bulk water distributions for waters in the first hydration shells of WCA colloids are shown with solid lines. Dashed lines give controls that indicate variations due merely to hydration shell geometry (i.e., the volumes over which the distributions are computed), without the explicit presence of a colloid.</p>	71
3.8	<p>(a) Projections of colloid first solvation shell perturbations to $P(\theta)$ onto principal components for colloids of varying size, indicated by the hard-sphere radius, and attractive strength (increasing from the top to bottom panel). The product of projections onto components 1 and 2 is also shown as triangles. (b) The relative entropy of solvation, as defined here and in Appendix B, as a function of the projections onto the first two principal components.</p>	73
3.9	<p>Model predictions of Equation 9 with (left) and without (right) the cross-term for the relative entropies for solvating all colloids. The dotted line indicates a perfect match of the model to the true S_{rel} values while light gray and light blue shaded regions indicate 5% and 20% error. Without cross-terms, the fit near small colloid sizes is poor. The fit for large colloid sizes is much improved when colloids of different levels of attractions are fit separately (Appendix C Fig. C.13).</p>	78
4.1	<p>A schematic of the workflow for the genetic algorithm to optimize (minimize or maximize) hydration water dynamics via re-patterning of surfaces at various fixed coverages of hydrophilic groups. Surfaces studied include the $10\bar{1}$ face of α-cristobalite with varied silanol coverage, self-assembled monolayer (SAM) surfaces with mixed methyl- and hydroxyl-terminated chains, and idealized surfaces of mixed binary Lennard-Jones (LJ) particles with either strong or weak LJ-water interactions.</p>	93
4.2	<p>Evolution of the genetic algorithm for various densities of the cristobalite surface. Blue points are minimum diffusivity surfaces for a given generation during minimization runs of the algorithm while orange points are maximum diffusivity surfaces from a separate maximization run. Blue and orange lines represent convex hulls. Generations 0 through 10 represent surfaces that were randomly generated without any optimization. For these early points, the blue and orange give the minimum and maximum diffusivities in the set of initial random surfaces.</p>	101

4.3	Regions of accessible hydration water diffusivities due to surface repatterning for the model surfaces. The hydrophilic surface fraction is defined for cristobalite as the percentage of the maximum possible silanol (Si-OH) density, for the SAMs as the fraction of hydroxyl-terminated chains, and for the Lennard Jones surfaces as the fraction of strongly attractive van der Waals sites. Filled symbols show the minimum and maximum diffusivity found by the genetic algorithm, while open symbols show the average diffusivity for randomly generated surfaces. Red curves show diffusivities for super-attractive LJ surfaces, using patterns optimized from the original LJ case. The blue dashed lines and blue diamonds show the expanded diffusivity range for SAM surfaces at surface fractions of 0.125 and 0.25 with regularly spaced and circularly-clustered hydroxyl groups (see Fig. 4.5). The black dashed line gives the average lateral diffusivity of simulation bulk water. Error bars, which are typically smaller than the symbol size, give 1% confidence intervals assuming Student's t-distribution from repeated simulation runs of the same surface.	103
4.4	(a) 2D radial distribution functions for SAM hydrophilic groups show that maximum-diffusivity cases strengthen spatial correlations at short distances. Solid lines for RDFs indicate minimum diffusivity surfaces, while dashed lines give maximum diffusivity cases. (b, c) Simulation snapshots of the minimum and maximum diffusivity patterns for SAM and cristobalite surfaces at hydroxyl fractions of 0.25 and 0.50. Clustering of hydroxyl groups (shown in red) on maximum diffusivity surfaces is clear from visualization of surface structures.	107
4.5	Top-views of surfaces with specific patterns to further test the genetic algorithm. Regularly-spaced and circularly-clustered hydroxyl groups represent the extremes of diffusivity suggested by populations of surfaces discovered by the genetic algorithm. The single stripe configuration for 25% hydroxyl-terminated chains represents a more exotic patterned to rudimentarily test for anisotropy in the diffusivity in directions parallel to and perpendicular to the stripe.	108
4.6	(a) Hydration water diffusivity correlates with the excess chemical potential for hard-sphere insertion near the interface across all surface types and patterning. Hatched points represent surfaces at hydrophilic fractions of zero or one, filled points minimum diffusivity surfaces, and open points maximum diffusivity ones. (b) The logarithm of water diffusivity correlates with the orientational contribution to the one-body intensive entropy. . .	110
4.7	Spatial variations as a function of distance from the mean interface (water density at 0.3 of its bulk value) are shown for water density, hard-sphere excess chemical potential, water diffusivity, and water re-orientation times.	111

4.8	Minimum and maximum diffusivity LJ surfaces from the genetic algorithm optimization, as well as LJ surfaces with the optimum patterns from SAM surfaces (values for these surfaces with super-attractive particles are shown in parentheses). Lateral diffusivities are averaged over 8 independent runs. The central column shows the P-value for the hypothesis that the minimum value is greater than the maximum using Welch’s t-test, which in all cases shows that there are no statistical differences between the minimum and maximum surfaces.	113
5.1	Schematic demonstrating the functioning of the genetic algorithm coupled to iterative training of a linear regression model.	133
5.2	While the solvation free energies at all interfaces studied are correlated with bulk solvation free energies for all solutes, they are not correlated with binding affinities. Thus, solute hydrophobicity/hydrophilicity as measured by solvation free energies is not a good metric for solute binding affinity.	136
5.3	Potentials of mean force are shown for all solutes studied at methylated (left) and hydroxylated (right) interfaces. The distance from the interface is calculated from the solute heavy-atom centroid to the fixed sulfur atoms of the SAM chains. All PMF values are relative to the bulk solvation free energies of solutes, which are shown as points at the furthest distances from the interface sampled. Error bars are those reported by pymbar. ⁶⁸	138
5.4	Contributions to binding free energies as described in the text and defined in Equations 5.3 and 5.4 are shown for the methylated (top) and hydroxylated (bottom) SAM surfaces. Summing the $\Delta\Delta G_{rep}$ and $\Delta\Delta G_{attr}$ terms yields the change in LJ interactions $\Delta\Delta G_{LJ}$. The repulsive component, which involves creating a cavity in which the solute may be inserted, is the predominant thermodynamic driving force for a solute’s preference for the interface over bulk solution.	139
5.5	The progress of the genetic algorithm is shown for minimization (decreasing regions) and maximization (increasing regions) of the affinity of methane. As described in Methods, it is advantageous to perform optimization in both directions simultaneously. For each generation, or round of optimization, the minimum and maximum affinity surfaces are shown connected by a vertical line. Green circles and black triangles represent generations where MD and the linear regression model, respectively, were used to determine solvation free energies. The points for the linear regression model are based only its final version – that trained from results of all MD simulations, including the second set of green points at much later generations.	143
5.6	Binding free energies to all surfaces are correlated to each other despite different chemistries and chemical patterns. “Patch” and “spread” surfaces refer to the inferred globally optimal patterns shown in Table 2.	148

5.7	Genetic algorithm optimization successfully identifies patterns that optimize the selectivity of an interface for methanol (2) compared to methane (1). Note that a minimization of methanol selectivity is a maximization of methane selectivity and vice versa. The optimization procedure is the same as that shown in Figure 5.5 but MD simulations are run for both methane and methanol for each surface pattern and the difference in interfacial solvation free energy between the two solutes is optimized.	151
5.8	Free energies of binding are broken down into differences between direct solute-system energetics ($\Delta\langle U_{sw} \rangle_2$) of solvation and relative entropies of solvation ($\Delta S_{rel,1\rightarrow 2}$), which for relatively rigid surfaces and solutes are dominated by water restructuring.	153
5.9	Changes from bulk water structure of three-body angle distributions for water oxygens within first solvation shells of bare surfaces with no solutes (a), solutes in bulk (b), and solutes near methylated (c) and hydroxylated (d) surfaces. Interfaces at bare surfaces are defined by the point at which average water density reaches 0.3 of its bulk value of 33.2 1/nm^3 with the first solvation shell defined as 0.4 nm from this interface, which is approximately the first minimum in the associated density profile for all surfaces. Solvation shells for solutes include all water oxygens within 0.55 nm of solute heavy atoms. Dotted vertical lines show the locations of expected peaks for liquid systems with LJ particles (63.4 \circ), ideal gas particles (90.0 \circ), and water (109.5 \circ). ⁸¹	156
A.1	Highest percentage RMSD clusters are shown for peptides in solution and with the pulling atom, restrained or free, near the simulated SAM surface. Differences do not appear with the release of restraints, though in many cases the GS part of the chain extends more readily into solution. For 3 and 4 leucines, at least one leucine is left dangling into solution, regardless of restraints, due to constraints introduced by backbone connectivity. Especially for structures in solution, low top cluster percentages indicate a lack of well-defined secondary structure.	178

A.2	Two-dimensional free energy surfaces in intrapeptide hydrogen bond (oxygen acceptor to hydrogen donor distance cutoff of 2.1 Å and donor-H-acceptor angle cutoff of 150°) and radius of gyration space are shown for all simulated peptides. Heat maps in the first column show results for the peptide restrained to be near the surface, in the second for the peptide free on the surface, and in the third for the peptide in solution. Each plot has been shifted in reference to its point of lowest free energy, with the scale at right in $k_B T$. In the case of the peptide restrained on the surface, re-weighting to account for the energy of the restraint has not been performed. Such re-weighting results in negligible changes to the free energy surfaces shown in the first column. In general, peptides explore a wide range of configuration space, though this space is reduced for peptides near the surface compared to those in bulk.	179
A.3	Similar to Figure A.2, but with total SASA per atom (top) or hydrophobic SASA per atom (bottom) on the y-axis. Once the restraint is released, the increases in R_g on the surface also occasionally correspond to increases in the SASA per atom, indicating that the peptide extends its hydrophilic region away from the surface into solution.	180
A.4	Radius of gyration for each peptide restrained on the surface, free of restraints but on the surface, and in solution. From the large jumps in R_g , it is clear that the peptide frequently folds and unfolds, indicating a lack of well-defined structure. This matches results from the free energy surfaces above, but further indicates that the wide regions of stability in SASA, R_g , and hydrogen bond space are due to peptide disorder rather than lack of equilibration. Autorrelation times are shown at the top left corner of each figure. Note that simulation times for peptides in solution are 90 ns, while those on the surface are 18 ns.	181

A.5	For the (GS) ₅ peptide, various quantities are shown throughout the slow-pulling process. Black curves are averages over configurations from individual trajectories falling within pulling coordinate (z-axis) bins of 1 Å width. Red curves are averages over all trajectories, or black curves. The minimum z-coordinate refers to the atom in the peptide closest to the SAM surface, with a pull-off event defined as the time at which this value exceeds 21 Å. Peptide z-extension refers to the z-separation between the pulled atom and the first backbone atom in a leucine residue (or the backbone atom preceding the C-terminal cap in (GS) ₅). R_g refers to the peptide radius of gyration. Shell waters are defined as waters within 5 Å from exposed peptide atoms. Total SASA/atom is the per-atom solvent accessible surface area for the entire peptide, while SASA _{hyd} is the solvent accessible surface area of only leucine residues. Numbers of hydrogen bonds (distance cutoff of 2.1 Å between heavy atoms and hydrogen bonded hydrogens, and donor-H-acceptor angle cutoff of 150°) are shown between the peptide and water, as well as between backbone atoms in the peptide. Force and work curves are additionally shown at the bottom.	182
A.6	Metrics shown for the (GS) ₃ L ₄ peptide, slow-pulling simulations. From Figure A.5 to Figure A.6, distinct qualitative differences in many of the curves reveal a starkly different pull-off mechanism with the addition of leucine residues. Mainly, the radius of gyration develops a clear maximum that corresponds to a force peak around full extension of the peptide. . .	183
C.1	Radial distribution functions for the LJ water model at various temperatures and densities (in g/cm ³). Vertical dashed lines show the cut-off of 0.34 nm used to define neighbors for calculating three-body angle distributions in bulk TIP4P/2005 and LJ water.	190
C.2	Radial distribution functions for the TIP4P/2005 water model at various temperatures and densities (in g/cm ³).	191
C.3	Three-body angle distributions for the LJ water model at various temperatures and densities (in g/cm ³).	192
C.4	Three-body angle distributions for the TIP4P/2005 water model at various temperatures and densities (in g/cm ³).	193
C.5	Two-dimensional probability densities in the number of neighbors within the 0.34 nm cutoff and the three-body angles are shown for the lowest-density, lowest-temperature and highest-density, highest-temperature conditions for both LJ and TIP4P/2005 water. For the low-density, low-temperature TIP4P/2005 distribution (top right), the peak at low angles (around 50°), is due to rare configurations involving 5 neighbors, where the tetrahedral peak around 109.5° is still prevalent but a neighbor interstitial to this tetrahedral arrangement also contributes low angles.	194

- C.6 Translational order parameters t and q are computed as described by Errington and Debenedetti.¹ Isotherms in t - q space are shown at far left, with isotherms as a function of density shown for q and t shown in center and right, respectively. The low-density structural anomaly boundary occurs where q exhibits maxima, and the high-density boundary occurs where t shows minima. In the center and right plots, solid lines are fits of the points shown to 4th-order Legendre polynomials, which are used to determine the minima/maxima and the anomaly bounds. In the center figure at a temperature of 280 K (5th from lowest), it appears that a maximum has been found; however, the lowest density point actually represents a decrease in q due to nucleation of voids in the liquid while it is very close to its metastability bound. Thus, the same q behavior is exhibited whether the mechanical stability or structural anomaly boundary is crossed. 195
- C.7 Tetrahedral order parameters isotherms computed via the definition from Errington and Debenedetti,¹ using 4 nearest neighbors (left) and computed from the distributions of three-body angles using a fixed distance cut-off (right). Circles are for TIP4P/2005 water and triangles show LJ water. When using a fixed cut-off, no maxima in q are observed, except in the case of the simulation at 0.87 g/cm³ and 280 K where voids are observed in the metastable liquid. When considering exactly 4 neighbors, some neighbors are occasionally included that reside outside the first solvation shell, leading to decreases in q at low densities where there is non-negligible probability of seeing 3 neighbors within the fixed cut-off instead (Fig. C.5). At high densities, q is much lower due to inclusion of greater numbers of neighbors that have angles different from the tetrahedral angle of 109.5°. 196
- C.8 Translational order parameter distributions are shown using a large cutoff identical to that defined by Errington and Debenedetti¹ (left), and a small cutoff of 0.34 nm equivalent to the neighbor cutoff for computing three-body angles (right). TIP4P/2005 water is represented by circles while LJ water is shown with triangles. Though the order parameter is shifted to different values, the qualitative behavior is almost identical, with minima of isotherms occurring at effectively the same densities. In this way, either cutoff definition produces the same result for the high-density structural anomaly boundary. 197

C.9	<p>Various metrics based on the three-body distribution² are shown for isotherms versus density. f_{tetra} is the fraction of the distribution that is tetrahedral as defined by the weight of the distribution between 90.0° and 130.0°. The top right shows the average of the cosine of the three-body angle within the tetrahedral region defined above. The bottom left shows the variance of the cosine of three-body angles within the tetrahedral region. The bottom right presents the entropy of the three-body angle distribution. Maxima in the three-body angle distribution entropies are suggestive of the high-density structural anomaly bound. However, no metric seems to suggest a lower-density bound.</p>	198
C.10	<p>Perturbations of bulk water three-body angle distributions in the second hydration shell (left) and a shell far enough away to achieve bulk-like behavior, between 2.4 and 2.6 nm from the colloid center (right). The magnitude of second-shell perturbations is almost an order of magnitude smaller than the first shell and the trends are distinct from those observed in the first shell. Far from the colloid, there is only low-magnitude noise compared to the reference distribution.</p>	199
C.11	<p>Perturbations from bulk for three-body angle distributions in the first hydration shell of WCA colloids are shown as solid lines for LJ, TIP3P, TIP4P-Ew, and TIP4P/2005 water models (from top to bottom). Small systems are in darker purple, while larger systems are lighter orange. Dashed lines represent distributions with applied geometric constraints rather than physically present colloids. Specifically, spherical shells corresponding to the same radial range between the hard-sphere radii and RDF first minima of WCA colloids are defined to select water oxygens in bulk-water simulations, with only waters at positions greater than the radius considered when computing the distributions. This effectively imposes geometric constraints without any colloids present. In the case of the LJ fluid, it is notable that the physical colloids perturb the distribution less than the geometric constraints. This is because the physical colloids, though purely repulsive, exhibit a soft interface while the constraints imposed are a perfectly hard interface defined by the effective hard-sphere radius.</p>	200
C.12	<p>Thermodynamic quantities associated with solvation, normalized to the average number of three-body angles formed by waters in the first hydration shell, are shown against projections onto the first two principal components of perturbations by all colloids. Circles represent WCA colloids, triangles low-density colloids, and squares high-density colloids. Smaller colloids are shown with dark purple while large colloids are brighter orange. Only for the relative entropy do we see evidence of a collapse of the projections across all colloid types. This collapse breaks down for larger colloid sizes, possibly indicating a length-scale cross-over or a point at which attractions become increasingly significant.</p>	201

C.13	Fits to the principal-component-based model for the relative entropy (with cross-terms) are shown for independently fitting to each of the colloid types. The gray shaded region represents 5% error while the blue-shaded region represents 20% error. The fits are excellent even at small colloid sizes. Compared to the fit over all colloids (Fig. 3.8), the individual fits are more accurate at large colloid sizes.	202
C.14	Three-body relative entropy in the first solvation shell of colloids is shown against the total solvation relative entropy. For small colloid sizes (smaller relative entropies), all colloid types fall along the same line. For colloids larger than ~ 0.5 nm in hard-sphere radius, the behavior becomes qualitatively different, with slopes for large colloids significantly diverging.	203
D.1	Schematics of 1D Markov models with distinct transition rates that result in average diffusivities that are additive (top) and non-additive (bottom).	208
E.1	For cristobalite at a fraction of 0.25 of the maximum possible number of silanol groups, correlations are shown between the measured diffusivity parallel to the plane of the surface (D_{net} on the x-axis of all panels) and (a) the diffusivity measured below the surface (b) the diffusivity measured above the surface (c) D_{net} itself (d) the measured diffusivity far from the surface (e) the average simulation temperature (f) the total simulation pressure (g) the component of simulation pressure perpendicular to the surface and (h) the equilibrated box dimension perpendicular to the surface. Black lines are linear regressions to the data with R^2 values shown on each plot. While the average simulation box length perpendicular to the interface weakly correlates with diffusivity (larger volumes for faster diffusion), this is due to variation in effective surface hydrophobicity that modulates interfacial width at constant pressure. For cristobalite surfaces at low density (here and in Figure E.2), it appears that the total system pressure (including all Cartesian directions) correlates weakly with the mobility of hydration water. Noting that the surface area is fixed in the plane of the interface and not relaxed during constant-pressure equilibration, it is clear that arrangements of silanol groups can affect the in-plane strain of these surfaces by changing the bonding pattern. Thus, a weak correlation is observed between measured diffusivity and total simulation pressure as the pattern of functional groups in cristobalite is adjusted.	211
E.2	Same as Figure E.1, but for cristobalite surfaces at a fraction of 0.375 of the maximum possible number of silanol groups.	212
E.3	Same as Figure E.1, but for cristobalite surfaces at a fraction of 0.50 of the maximum possible number of silanol groups.	212
E.4	Same as Figure E.1, but for cristobalite surfaces at a fraction of 0.75 of the maximum possible number of silanol groups.	213

E.5	Same as Figure E.1, but for the SAM surface at an OH-terminated chain fraction of 0.25.	213
E.6	Same as Figure E.1, but for the SAM surface at an OH-terminated chain fraction of 0.50.	214
E.7	Same as Figure E.1, but for the LJ surface at an attractive particle fraction of 0.25. For LJ surfaces at all coverages, the pressure normal to the interface correlates with the diffusivity. It should be noted, however, that the range of Z-pressures observed is the same as for other surface types. With surface patterning playing little to no role in determining diffusivity for the LJ surfaces (see main text), system properties that vary with system set-up and equilibration, like the Z-pressure, are more statistically significant, whereas these same fluctuations primarily introduce statistical noise to diffusivities on SAMs or cristobalite.	214
E.8	Same as Figure E.1, but for the LJ surface at an attractive particle fraction of 0.50.	215
E.9	Probability density distributions for the cosine of the tilt angle for SAM surfaces (the angle between the direction perpendicular to the interface and the vector between the restrained sulfur atom and the last united atom before the head group). The top panel shows the distribution for OH-terminated chains while the bottom for CH ₃ -terminated ones. Each color represents an independent simulation, with solid lines corresponding to the minimum-diffusivity surface pattern and dashed lines to the maximum-diffusivity surface. Variations in tilt angle distributions between independent simulation runs are larger than any differences observed between the minimum and maximum diffusivity patterns, implying negligible changes in average surface structure. Additionally, freezing all but the head-group atoms of SAM chains results in nearly identical diffusivity ranges in Figure 4.3 (also see Figure E.12), although the diffusivity is slightly reduced overall (by 0.014 to 0.032 Å ² /ps). The decreases in diffusivity are more pronounced at higher surface densities and stem from the reduction, due to chain rigidification, in hydrogen bonding between surface hydroxyls, increasing the availability of water-surface hydrogen bonds.	216

E.10	Top-views for the normalized fluctuations in the number of waters within probes of radius 3.3 Å sitting with their edge at the mean interface of (a) cristobalite, (b) SAM, and (c) LJ surfaces. The mean interface is defined as the location at which the average water density is 0.3 of the bulk value for TIP4P-Ew of 0.0332 Å ⁻³ . For LJ surfaces, the middle column considers the patterns determined by optimization of the SAM surface at the same density. It is clear that fluctuations are enhanced near larger patches of more hydrophobic groups on both SAM and LJ surfaces. For cristobalite, both very small and large fluctuations are observed near hydroxyl groups. This in turn leads to higher excess chemical potentials for hard-sphere probes near the surface.	217
E.11	Top-views of height fluctuations for instantaneous interfaces for cristobalite (top), SAM (middle), and LJ surfaces (bottom). The instantaneous interface is defined according to the methodology of Willard and Chandler, ¹ but with a surface density cut-off of 0.3 of the bulk value of 0.0332 Å ⁻³ for TIP4P-Ew water.	218
E.12	Correlation between lateral diffusivities of water near minimum and maximum diffusivity surfaces and the average interfacial height fluctuations using the method of Willard and Chandler ¹ to define the instantaneous interface. Hatched points represent surfaces at hydrophilic fractions of zero or one, filled points minimum diffusivity surfaces, and open points maximum diffusivity surfaces. The correlation is enhanced by freezing the united atoms (non-head-group atoms) of the SAM chains, as shown by the black versus blue diamond symbols.	219
E.13	Profiles for water density, hard-sphere excess chemical potential, water diffusivity parallel to the interface, and water dipole re-orientation times as a function of distance from the mean interface (defined as the point where water reaches 0.3 of its bulk value). Dipole re-orientation times were computed by fitting the time decay of the cosine of the angle between the original dipole moment vector with a stretched exponential, as in the work of Debenedetti and co-workers. ² The qualitative trends are the same if an instantaneous interface definition is used instead. In general, hard-sphere excess chemical potentials are correlated with both lateral diffusivities and water re-orientation times. However, these spatial correlations seem to break down very close to the solid surfaces, where geometric constraints due to the presence of surface atoms seem to impact hard-sphere insertion probabilities and water mobility in an unrelated manner.	220

E.14	The natural logarithm of the ratio of lateral diffusivity to its bulk value is shown as a function of the orientational, translational, and total contributions to hydration water entropy 8 Å from the interface (reported relative to bulk waters). Filled symbols are minimum diffusivity surfaces (distributed hydrophilic groups), open symbols are maximum diffusivity surfaces (clumped hydrophilic groups), and hatched symbols are surfaces with hydrophilic fractions of 0.0 or 1.0. When all atoms in SAM chains except head-groups are frozen, all correlations are enhanced, mainly due to more negative translational contributions.	221
G.1	Octanol-water transfer free energies (equal to the natural logarithm of the octanol-water partition coefficient) do not correlated to either aqueous solvation free energies at interfaces (top) or, more importantly, binding free energies (bottom). Transfer free energies were computed as the difference to solvate each solute in water minus solvation in a similarly-sized volume of octanol at the same temperature and pressure. Terms accounting for the difference in average volume have been neglected as they are small ($\sim 0.05 k_B T$) compared to the uncertainty (up to $\sim 0.30 k_B T$).	229
G.2	Contributions to solvation free energies from turning on electrostatics with full LJ interactions on (blue, ΔG_Q), switching solutes from the ideal gas phase to interacting with the rest of the system via only hard-sphere repulsions (gray, ΔG_{rep}), and switching hard-sphere interactions to full LJ interactions (orange, ΔG_{attr}). Due to the chosen alchemical pathway in expanded ensemble simulations it was possible to directly calculate the free energies to turn on only LJ interactions and to turn on electrostatics with LJ on. Repulsive free energies as described here are not along with pathway but were estimated by from Gaussian fits to overlap probability distributions as shown in Figure G.3. The attractive component of the LJ free energy is then estimated by subtracting this hard-sphere solute solvation free energy.	230

G.3	Probability distributions for the number of non-overlapping non-solute atoms are shown for all solutes in bulk (circles) and at methyl- (squares) and hydroxyl-terminated (diamonds) SAMs. Gaussian fits to the distributions used to evaluate the probability at zero non-overlapping atoms are shown as solid, dashed, and dotted lines, respectively. For distributions in bulk and at hydrophilic interfaces, Gaussian distributions have been observed for small solutes for spherical, ¹ cuboidal, ² and irregular volumes, including that defined by a single alanine. ³ Near hydrophobic surfaces, however, the Gaussian fits are anticipated to underestimate the weight in the distribution at low values of N , ^{2,4,5} resulting in underestimates of the hard-sphere solvation free energy at these surfaces. This does not seem to be the case, however, for values of N below the average, where the Gaussian fit actually leads to an overestimate of the weight, possibly due to poor sampling of the tails of the distributions, especially for larger solutes.	231
G.4	Water density profiles as a function of distance from the frozen sulfur atoms of the SAM surface, the same scale as used for the PMFs shown in Figure 5.3 of the main text.	232
G.5	Hard-sphere excess chemical potentials for methane-sized (0.33 nm radius) particles as a function of distance from distance from frozen sulfur atoms in the SAM surfaces. Larger values signify smaller density fluctuations, which is a proxy for measuring hydrophilicity/hydrophobicity, indicating that the surfaces increase in hydrophilicity in the order fully methylated, half and half patchy, half and half spread, and fully hydroxylated.	232
G.6	Genetic algorithm optimization of surface patterns to minimize and maximize the affinity of capped glycine for a 50:50 methyl- and hydroxyl-terminated surface. Procedures are identical to the optimization for methane except that twice as much computational time is utilized for each simulation in both determining biasing weights and for data production. Though the results are far less impressive, the algorithm is beginning to identify patterns that change affinity for the interface, with the trained regression model accurately predicting free energies of solvation (see Table S1). While a longer run might help converge to more optimal patterns, convergence is mainly hindered by the large amount of noise in the MD simulation predictions of affinity.	233
G.7	Radial distribution functions between all solute atoms and water oxygens. Heavy atoms are shown in red while hydrogens are shown in black. Cutoffs defining first and second solvation shells at 5.5 and 8.5 Å are shown as vertical dotted lines.	236

List of Tables

2.1	Peptide mutation sequences and summary of experimentally obtained and simulated interaction free energies with extended hydrophobic surfaces.	21
2.2	Approximate bias in free energy estimates from each methodology at both pulling rates. While increasing trends with the number of leucines suggest slopes are overestimated, the fast-pulling BAR formula yields fairly similar biases, suggesting less error for this estimate. Note that these estimates of the bias still depend on good estimates of the work distributions, possibly explaining the large biases estimated for the forward Jarzynski case at a slower pulling rate.	27
2.3	Free energy estimates from Jarzynski's equality for both forward and reverse distributions, as well as from BAR. In the fast-pulling case, the reverse work distributions are similar across all peptides, yielding similar Jarzynski estimates. However, for the slow-pulling case, relatively few very negative work values result in widely varying reverse Jarzynski estimates.	28
2.4	Fitting parameters for free energy model linear in the number of leucines. Δg_{GS} is the free energy change per residue of removing the scaffold from the surface and Δg_L is the free energy change of removing a single leucine residue.	38
3.1	Projections of principal components of perturbations to $P(\theta)$ in the first hydration shell of colloids onto principal components of perturbations across all temperatures and densities of bulk water.	70
4.1	Performance metrics for the surface design genetic algorithm	105

5.1	Solute structures and solvation free energies at interfaces and in bulk with comparison made in bulk to hydration free energies computed from Henry's constants. Solvation free energies apply to the free energy of solvation either in bulk or when the solute is near to methylated (CH_3) hydroxylated (OH) or 1:1 mixed SAM surfaces with either patchy or dispersed (spread) patterns of hydroxyl groups. Units are reported in $k_B T$ with uncertainties as the standard error in the mean based on 5 independent runs. Henry's Law constants are shown in parentheses in units of mol/m ³ Pa and were obtained from. ⁴² For an accurate force field, we expect that the last two columns be equal. For a derivation relating Henry's Law constants to computed free energies of solvation, see Appendix F.	126
5.2	Surface representations and free energies of solvation for methane, determined by MD simulations and linear regression models. The surfaces shown for the GA are those with minimum and maximum solvation free energies predicted by the linear regression model, not necessarily the overall optimal surfaces reported by the GA. Uncertainties are the standard error in the mean from 5 independent simulations.	145
5.3	Differences in solvation free energies at various interfaces are shown. Errors are based on standard error propagation of the standard error in the mean of the solvation free energy estimates based on 5 independent MD simulation runs. Numbers in parentheses provide the P-value for the two-sided hypothesis test for the equivalence of the mean solvation free energies.	149
C.1	Various thermodynamic components of free energies of solvation for different colloids. All units in $k_B T$ unless otherwise noted. Uncertainties for ΔG_{rep} and ΔG_{attr} derive from estimates of uncertainty using uncorrelated potential energies from the pymbar package. ³ Uncertainties for ΔH and $\langle U_{cs} \rangle$ are standard deviations from averaging all uncorrelated potential energies (and pressure-volume terms in the former case) in the trajectory. Uncertainties for ΔG , ΔS , and S_{rel} are computed via standard error propagation. . . .	204
G.1	Surface representations and free energies of solvation for capped glycine, determined by MD simulations and linear regression models. The surfaces shown for the GA are those with minimum and maximum solvation free energies predicted by the linear regression model, not necessarily the overall optimal surfaces reported by the GA. Uncertainties are the standard error in the mean from 5 independent simulations.	234

G.2	Surface representations and selectivity for methanol compared to methane, determined by MD simulations and linear regression models. Note that minimization of methanol selectivity is maximization of methane selectivity and vice versa. The surfaces shown for the GA are those with minimum and maximum solvation free energies predicted by the linear regression model, not necessarily the overall optimal surfaces reported by the GA. Uncertainties are the standard error in the mean from 5 independent simulations.	235
-----	---	-----

Chapter 1

Introduction

1.1 Overview

This dissertation represents a unique attempt at understanding the role of water in mediating interfacial properties and solute-solute interactions, painting a picture for how nanoscale heterogeneities may be leveraged to modify both proximal water properties and in turn overall properties of an interface. Designing materials for interfacial applications meets with several unique challenges compared with standard materials design. The presence of an interface implies interactions between different materials or phases, with the interfacial structure and properties of each phase usually significantly different than their bulk states. Further, the response of each phase to the other, and hence the overall properties of the interface, may depend subtly on nanoscale heterogeneities in morphology or chemistry.¹⁻⁴ Such considerations are especially important for applications in which solutes are interacting with either an extended interface or each other, as is the case in chromatography, liquid-phase catalysis, or self-assembling materials. For instance, self-assembling peptides show great promise as “smart,” environmentally responsive materials that develop and adapt their structure and interactions on a wide range of

length-scales.⁵⁻⁸ However, it remains very difficult to predict the properties and behavior of such materials based on peptide chemistry alone, mainly due to the poorly understood role of water-mediated interactions.

A rich literature concerning the theory of solvation,^{9,10} adsorption,¹¹ self-assembly,¹² and hydrophobic interactions¹³⁻¹⁵ can provide some guidance in designing interfacial and self-assembling materials. However, these theories either include only implicit descriptions of solvent, which do not fully consider the unique nature of liquid water, or treat only very simple, idealized solutes. To bridge the gap between theory and applications, we study a number of realistic yet simple systems. Our observations and suggested extensions of current theories will greatly assist the future development of computational tools to rationally design interfacial and self-assembling materials for practical applications.

1.2 Designing interfacial and self-assembling materials

Industrial processes that rely on specific material properties or molecular-scale behavior at an aqueous interface are ubiquitous. Water purification, filtration, selective membrane capture, chromatography, aqueous-phase adsorption, and even surface-catalyzed reactions within flow reactors are all processes that will benefit from a deeper knowledge of the interplay between material characteristics and water's interfacial behavior. In water purification, desirable materials facilitate rapid water transport through a membrane while also limiting the passage of other solutes and resisting the adhesion of contaminants to the membrane surface.^{16,17} A key aspect of optimizing adsorption and chromatographic processes involves modifying the selectivity of an interface to bind one solute or another, which is largely dictated by the interfacial thermodynamics and dynamics of water (the

solvent).¹¹ This is well-recognized in biological settings, where water is thought to play an essential role in driving interactions between biomolecules, as well as regulating the timescales involved in molecular-level biologic processes.^{18,19} Similarly, water plays a role in surface-catalyzed reactive processes by not only modifying the reaction-site environment, but also in determining the surface concentration of reaction products and mass transport to and from the interface.

While water's behavior significantly affects all of the above processes, the specific role of nanoscale material features in determining water interfacial properties remains largely a mystery. Investigation of these phenomena is difficult due to the small length-scales and short timescales that are involved – water structural correlations typically persist to around 1.0 nm, with rearrangements of configurations, such as hydrogen bonding, occurring on the order of picoseconds. Measurements of macroscopic properties, such as contact angles, intrinsically average over nanoscale surface features that interact with water on its innate molecular length scale, effectively ignoring the role of such features in modifying water behavior. It is clear from biology, however, that nanoscale variations in both geometry and chemistry of proteins have a tremendous impact on protein function.¹⁸ Mimicking these heterogeneities represents an exciting opportunity in the design of next-generation interfacial materials. While fabrication and control of such nanoscale features will be critically important moving forwards, we focus here on building an understanding of water's response to such material properties. This will extend the material design space and direct advancements in manufacturing and synthesis by identifying features that can be tuned in order to most impactfully manipulate water behavior and hence material performance.

Closely related to the design of nanoscale features on interfacial materials is the design of self-assembling materials. Such systems represent the majority of biological materials, which are built from large networks of interacting proteins, the structure and

properties of which determine the overall material toughness, rigidity, and flexibility.⁵⁻⁸ Examples include materials as diverse as silk²⁰ and the cement that holds barnacles tightly to surfaces in harsh marine environments.^{21,22} Exciting efforts are now pursuing non-naturally occurring, protein-based materials that engineer a protein's unique chemical and structural properties for a variety of biological and non-biological applications – this includes “smart” responses, whereby materials alter their nanostructure and interactions to effect mechanical and thermodynamic property shifts in response to environmental cues like temperature, salt concentration, or pH.^{6,23} Unfortunately, the development of these systems has been predominantly through trial and error, with very few guiding principles and few computational tools to support the rational and efficient design of proteins exhibiting specific material properties.

The concept of self-assembly intrinsically involves the development of an interface, with peptides or other molecules forming phase-separated or solvated structures on various length-scales based on driving forces regulated to a large extent by the nature of the solvent. The central challenge in designing new protein-based materials involves predicting the strength and nature of protein-protein interactions, specifically those that are mediated by hydrating water. While direct atomic-scale interactions between proteins (e.g., electrostatic or van der Waals) are easy to predict, the real challenge is understanding how water surrounding and between them responds to subtle features at the protein surface, and in turn directs more complex inter-protein interactions. The most important yet poorly-understood manifestation of these phenomena is the effective attraction between hydrophobic regions on proteins, which arises not from any fundamental force between these groups but instead from the unique properties of liquid water.¹³⁻¹⁵ As discussed further below, water is in fact an unusual solvent, with many of its bulk anomalies, such as its unique tetrahedral structure, still being researched.

1.3 Water-mediated interactions

Water is instrumental in determining the interfacial properties of a material as well as in driving solute-solute or solute-surface interactions. This involves perturbations to water structure due to solute or material characteristics, which necessitates a detailed understanding of the behavior of this common yet unusual liquid. The majority of anomalous water properties can be associated with its molecular structure.²⁴ Water is capable of donating and accepting two hydrogen bonds, resulting in a strong preference for tetrahedral structure both in the solid and liquid phases. Competition between this tetrahedrality and a spherically-symmetric, simple-liquid structure results in the presence of a temperature of maximum density at constant pressure along with other anomalies.²⁴⁻²⁶ At high temperatures, hydrogen bonding is entropically disfavored, resulting in transition to the simple-liquid structure with increasing temperature along with an initial increase in density near the freezing point as the tetrahedral structure begins to break down.

It is this same structuring of liquid water that has historically been used to develop and interpret theories of solvation, in particular of hydrophobic species.²⁷⁻²⁹ Modern views of aqueous solvation of nonpolar solutes emphasize instead the availability of free space within the liquid.³⁰⁻³² This quantity contains an intertwined dependence on both liquid structure and density fluctuations.^{1-3,33-35} Variations of solvation thermodynamics with temperature are then associated with the shift in water behavior from tetrahedrally-dominated to simple-liquid-like,²⁴ which results in differences in both structure and density fluctuations. With this nuanced view, it has also been established that the nature of hydrophobic solvation, as well as association, depends on both the size and geometry of the solute.^{13,15,36} It has also been demonstrated that the hydrophobicity or hydrophilicity of an interface depends on the surface geometry,^{4,37} such as surface roughness or concavity, as well as the patterning of chemical functional groups at the interface.^{1,38,39}

This work further explores the role of solute geometry, chemistry, and nanoscale interfacial patterning on the behavior of nearby water. Deeper connections between water’s structural response to solutes and interfaces and its thermodynamic and dynamic properties are established while connecting to the theory described above. We also consider systems that are not clearly within the scope of previously established theories, in particular flexible solutes of intermediate size that contain mixtures of both hydrophobic and polar functional groups. These realistic systems allow for rigorous testing of theory on systems that can still be probed in great detail on an atomistic level through molecular dynamics simulations. The insights gleaned from this approach can not only be used to improve theory, but also as preliminary guidelines for designing materials on a nanoscale. This represents an additional feature to tune in interfacial materials, with the approaches presented here also laying the ground-work for designing nanoscale material features through computation.

1.4 Outline

Following the introduction, this thesis is organized as follows:

- Chapter 2: Scaling of the hydrophobic interaction between short peptides and extended hydrophobic surfaces
- Chapter 3: Characterization of three-body angle distributions in bulk and solvation shells of model solutes
- Chapter 4: Computational design of surface chemical patterns to optimize proximal water mobility
- Chapter 5: Chemically-patterned interfaces for the controlled affinity and selectivity of small-molecule solutes

- Chapter 6: Conclusions

In Chapter 2, we present a computational study of hydrophobic interactions in realistic systems that is directly comparable to the experiments of collaborators. Molecular insights from this study relate peptide affinity for hydrophobic surfaces to water structure. These relationships are pursued in Chapter 3 in the context of hydrophobic solvation of a variety of model hydrophobes of varying size and attractive interactions. Chapter 4 examines the effect of surface chemical patterning on water mobility at an interface, establishing fundamental connections between water thermodynamics and dynamics at an interface. Finally, Chapter 5 considers the effect of interfacial chemistry, including chemical patterning, on the surface affinity of a set of chemically diverse solutes. The effect of interfaces and solutes on nearby water structure is investigated in the context of driving adsorption.

References

- [1] Acharya H, Vembanur S, Jamadagni SN, Garde S. Mapping hydrophobicity at the nanoscale: Applications to heterogeneous surfaces and proteins. *Faraday discussions*. 2010; 146: 353–365. ISBN: 1359-6640\l1364-5498.
- [2] Jamadagni SN, Godawat R, Garde S. Hydrophobicity of Proteins and Interfaces: Insights from Density Fluctuations. *Annual Review of Chemical and Biomolecular Engineering*. 2011; 2(1): 147–171. ISBN: 1947-5438\l978-0-8243-5202-8.
- [3] Godawat R, Jamadagni SN, Garde S. Characterizing hydrophobicity of interfaces by using cavity formation, solute binding, and water correlations. *Proceedings of the National Academy of Sciences*. 2009; 106(36): 15119–15124. ISBN: 0027-8424.
- [4] Xi E, Venkateshwaran V, Li L, Rego N, Patel AJ, Garde S. Hydrophobicity of proteins and nanostructured solutes is governed by topographical and chemical context. *Proceedings of the National Academy of Sciences*. 2017; 114(51): 13345–13350. ISBN: 0027-8424, 1091-6490.
- [5] Fairman R, Akerfeldt K. Peptides as novel smart materials. *Current Opinion in Structural Biology*. 2005; 15(4): 453–463. ISBN: 0959-440X.
- [6] Grove TZ, Regan L. New materials from proteins and peptides. *Current Opinion in Structural Biology*. 2012; 22(4): 451–456. Publisher: Elsevier Ltd ISBN: 0959-440X.
- [7] Zhao X, Zhang S. Designer self-assembling peptide materials. *Macromol Biosci*. 2007; 7(1): 13–22.
- [8] Woolfson DN, Bartlett GJ, Bruning M, Thomson AR. New currency for old rope: from coiled-coil assemblies to α -helical barrels. *Current Opinion in Structural Biology*. 2012; 22(4): 432–441.
- [9] Ben-Amotz D, Raineri FO, Stell G. Solvation Thermodynamics: Theory and Applications [†]. *The Journal of Physical Chemistry B*. 2005; 109(14): 6866–6878.
- [10] Ben-Amotz D. Interfacial solvation thermodynamics. *Journal of Physics: Condensed Matter*. 2016; 28(41): 414013.
- [11] Dorsey JG, Dill KA. The molecular mechanism of retention in reversed-phase liquid chromatography. *Chemical Reviews*. 1989; 89(2): 331–346.
- [12] Israelachvili JN. *Intermolecular and Surface Forces*. Burlington, MA: Elsevier 3rd ed. 2011.
- [13] Chandler D. Interfaces and the driving force of hydrophobic assembly. *Nature*. 2005; 437(7059): 640–7.

- [14] Ben-Amotz D. Water-Mediated Hydrophobic Interactions. *Annual Review of Physical Chemistry*. 2016; 67(1): 617–638.
- [15] Hillyer MB, Gibb BC. Molecular Shape and the Hydrophobic Effect. *Annual Review of Physical Chemistry*. 2016; 67(1): 307–329.
- [16] Park HB, Kamcev J, Robeson LM, Elimelech M, Freeman BD. Maximizing the right stuff: The trade-off between membrane permeability and selectivity. *Science*. 2017; 356(6343): eaab0530.
- [17] Werber JR, Osuji CO, Elimelech M. Materials for next-generation desalination and water purification membranes. *Nature Reviews Materials*. 2016; 1(5): 16018.
- [18] Ball P. Water as an Active Constituent in Cell Biology. *Chemical Reviews*. 2008; 108(1): 74–108. arXiv: NIHMS150003 ISBN: 0009-2665.
- [19] Ball P. Water is an active matrix of life for cell and molecular biology. *Proceedings of the National Academy of Sciences*. 2017; 114(51): 13327–13335. ISBN: 0027-8424, 1091-6490.
- [20] Altman GH, Diaz F, Jakuba C, et al. Silk-based biomaterials. *Biomaterials*. 2003; 24(3): 401–416. ISBN: 9783527329168.
- [21] Kamino K. Mini-review: barnacle adhesives and adhesion. *Biofouling*. 2013; 29(6): 735–49.
- [22] Kamino K. Molecular Design of Barnacle Cement in Comparison with Those of Mussel and Tubeworm. *The Journal of Adhesion*. 2010; 86(January 2015): 96–110. ISBN: 0021-8464.
- [23] Kammerer RA, Kostrewa D, Zurdo J, et al. Exploring amyloid formation by a de novo design. *Proceedings of the National Academy of Sciences of the United States of America*. 2004; 101(13): 4435–40. ISBN: 0027-8424 (Print).
- [24] Brini E, Fennell CJ, Fernandez-Serra M, Hribar-Lee B, Lukšič M, Dill KA. How Water’s Properties Are Encoded in Its Molecular Structure and Energies. *Chemical Reviews*. 2017; 117(19): 12385–12414.
- [25] Palmer JC, Poole PH, Sciortino F, Debenedetti PG. Advances in Computational Studies of the Liquid–Liquid Transition in Water and Water-Like Models. *Chemical Reviews*. 2018; 118(18): 9129–9151.
- [26] Cisneros GA, Wikfeldt KT, Ojamäe L, et al. Modeling Molecular Interactions in Water: From Pairwise to Many-Body Potential Energy Functions. *Chemical Reviews*. 2016; 116(13): 7501–7528.

- [27] Bernal JD, Fowler RH. A Theory of Water and Ionic Solution, with Particular Reference to Hydrogen and Hydroxyl Ions. *The Journal of Chemical Physics*. 1933; 1(8): 515–548.
- [28] Frank HS, Evans MW. Free Volume and Entropy in Condensed Systems III. Entropy in Binary Liquid Mixtures; Partial Molal Entropy in Dilute Solutions; Structure and Thermodynamics in Aqueous Electrolytes. *The Journal of Chemical Physics*. 1945; 13(11): 507–532.
- [29] Stillinger FH. Structure in aqueous solutions of nonpolar solutes from the standpoint of scaled-particle theory. *Journal of Solution Chemistry*. 1973; 2: 141–158.
- [30] Pratt LR. Molecular Theory of Hydrophobic Effects: "She is too mean to have her name repeated."*. *Annual Review of Physical Chemistry*. 2002; 53(1): 409–436.
- [31] Ashbaugh HS, Pratt LR. Colloquium: Scaled particle theory and the length scales of hydrophobicity. *Reviews of Modern Physics*. 2006; 78(1): 159–178. ISBN: 0034-6861.
- [32] Hummer G, Garde S, García aE, Pohorille a, Pratt LR. An information theory model of hydrophobic interactions. *Proceedings of the National Academy of Sciences of the United States of America*. 1996; 93(17): 8951–8955. ISBN: 2819960820.
- [33] Lynden-Bell RM, Debenedetti PG. Computational Investigation of Order, Structure, and Dynamics in Modified Water Models [†]. *The Journal of Physical Chemistry B*. 2005; 109(14): 6527–6534.
- [34] Lynden-Bell RM, Giovambattista N, Debenedetti PG, Head-Gordon T, Rossky PJ. Hydrogen bond strength and network structure effects on hydration of non-polar molecules. *Physical chemistry chemical physics : PCCP*. 2011; 13(7): 2748–2757. ISBN: 1463-9076.
- [35] Lynden-bell RM, Head-gordon T. Solvation in modified water models: towards understanding hydrophobic effects. *Molecular Physics*. 2006; 104(22-24): 3593–3605.
- [36] Fennell CJ, Kehoe C, Dill KA. Oil/Water Transfer Is Partly Driven by Molecular Shape, Not Just Size. *Journal of the American Chemical Society*. 2010; 132(1): 234–240.
- [37] Setny P. Water properties and potential of mean force for hydrophobic interactions of methane and nanoscopic pockets studied by computer simulations. *The Journal of Chemical Physics*. 2007; 127(5): 054505. ISBN: 0021-9606 (Print)\n0021-9606 (Linking).
- [38] Giovambattista N, Debenedetti PG, Rossky PJ. Hydration behavior under confinement by nanoscale surfaces with patterned hydrophobicity and hydrophilicity. *Journal of Physical Chemistry C*. 2007; 111(3): 1323–1332. ISBN: 1932-7447.

- [39] Stirnemann G, Rossky PJ, Hynes JT, Laage D. Water reorientation, hydrogen-bond dynamics and 2d-IR spectroscopy next to an extended hydrophobic surface. *Faraday Discussions*. 2010; 146: 263. ISBN: 1359-6640\|1364-5498.

Chapter 2

Scaling of the hydrophobic interaction between short peptides and extended hydrophobic surfaces

Reproduced in part with permission from:

Stock, P., Monroe, J. I., Utzig, T., Smith, D. J., Shell, M. S., & Valtiner, M. (2017). Unraveling Hydrophobic Interactions at the Molecular Scale Using Force Spectroscopy and Molecular Dynamics Simulations. *ACS Nano*, 11(3). <https://doi.org/10.1021/acsnano.6b06360>

2.1 Introduction

Despite over 30 years of extensive research, discerning the origin, scaling, and even existence of forces between extended hydrophobic surfaces remains a major experimental challenge. Direct measurements of the hydrophobic interaction (HI) not relying on inference from macroscopic properties have been notoriously difficult to execute without

interposing physical forces and processes, such as bridging bubbles and surfactant overturning. In their recent review, Ducker and Mastropietro¹ identify only five experimental results at extended hydrophobic interfaces that show convincing evidence for exponentially decaying attractive hydrophobic interactions (HI) in the range of 3-12 Å,²⁻⁶ where reexamination of the large body of remaining literature reveals principal experimental difficulties.¹ Here, we use ‘hydrophobic interaction’ to refer to the total interaction free energy between solvated hydrophobes and not specifically the water-mediated contribution. In contrast at small length scales, non-linear optical spectroscopy studies^{7,8} and diffraction experiments⁹ have shown how water orients and fluctuates around uncharged small hydrophobes, which may be critical to molecular HI forces, even if such forces themselves have been difficult to measure. Indeed, molecular theory and simulations continue to clarify the relationship between fluctuating water structure and the hydrophobicity of isolated solutes and interfaces.^{10,11}

Understanding how HI forces manifest in a molecular context and can be systematically manipulated, e.g. through variations in amino acid sequence, remains a major unsolved problem. This is particularly relevant to peptides, which depart from the idealized solutes often studied in theory as they comprise non-rigid, heterogeneous molecules with both hydrophobic and hydrophilic groups. For realistic sequences containing a variety of amino acids, a delicate balance of non-covalent driving forces, including the HI as a significant contributor, directs peptide aggregation, self-assembly, folding, and binding behaviors.¹²⁻¹⁴ Several recent single-molecule spectroscopic studies have now probed poly-lysine,¹⁵ spider silk peptides¹⁶ or charged/hydrophobic peptide sequences¹⁷ with extended hydrophobic surfaces,¹⁸⁻²⁰ but systematically understanding how the HI scales with the hydrophobic content in a peptide sequence has yet to be clarified.

Here, we examine the scaling of interactions between a short peptide of mixed hydrophilic/hydrophobic character with an extended hydrophobic surface as the peptide

hydrophobic content is adjusted. Related studies on the single-molecule level have focused on homopolymers²¹ or homopeptides.¹⁹ By mutating a hydrophilic scaffold to increase the number of hydrophobic residues, we can instead probe the additivity and potential cooperation or competition between hydrophilic and hydrophobic moieties related to surface affinity. In this work, fully atomistic nonequilibrium and equilibrium molecular dynamics (MD) simulations explore the same systems experimentally measured by collaborators performing single molecule AFM (SM-AFM) procedures.²² In simulations, as in experiment, we assess the strength of the peptide-surface interaction by rigorously applying Jarzynski’s equality²³ to determine equilibrium free energy differences. Finally, we show that these equilibrium free energy differences may be associated with changes in the equilibrium ensemble of water structure at the level of three-body angular correlations.

2.2 Methods

2.2.1 Simulated system

We constructed surfaces to emulate the properties of the experimental SAM of 11-carbon alkanethiols on an Au111 surface, as shown in Figure 8. For alkanes of this length, a rigid, extended structure of each chain is adopted, with a $\sqrt{3}x\sqrt{3}$ R30° monolayer structure.^{24,25} Simulated SAMs are constructed from dodecane chains aligned to this periodic structure with a nearest neighbor spacing of 4.98 Å. The resulting surfaces were approximately 60 Å x 60 Å, with 12 and 14 chains in the x and y directions, respectively. To maintain atomistic resolution and avoid introducing new parameters, atoms in dodecane chains are assigned standard AMBER force field atom types of “CT” and “H.” For simplicity, charges are assigned based on the work of Siu, et al.²⁶

We minimize initial structures and then equilibrate surfaces for 200 ps with periodic

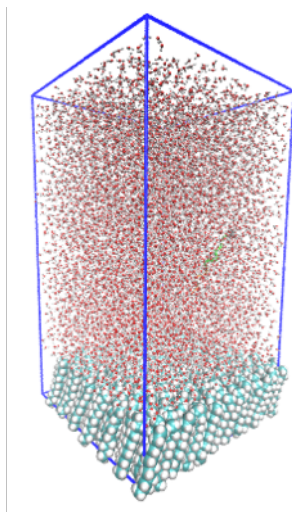


Figure 2.1: The dimensions of the simulation box are initially set to 60 x 60 x 100 Å, with pressure equilibration resulting in an increase of the box height to around 103 Å. The x- and y-coordinates of the system are defined to be in the plane of the surface, with the z-coordinate normal to the SAM. In all simulations the peptide is pulled along only the z-coordinate, as shown in the right figure with waters removed.

boundary conditions at 300 K in the NVT ensemble. A position restraint (spring constant of 10 kcal/mol Å²) fixes one end of each dodecane chain at the correct lattice location. The resulting surface structure exhibits fully extended chains with an average CH₃ density at the non-fixed chain end of 21 Å²/CH₃ and an average chain tilt angle of around 19°. This tilt is smaller than the experimentally-reported value of ~30° for chains of a similar length,^{24,25,27,28} even considering that the tilt angle distribution is expected to be fairly broad.^{27,28} Numerous studies have tried to address simulated tilt angles in alkanethiol SAMs on Au(111) surfaces,²⁷⁻³⁰ using both united atom^{27,29,31} and atomistic models,^{28,30,32} but very often require force-field tuning or significant atomic constraints.^{33,34} We allowed the chains to move freely with the assumption that the discrepancy in tilt angles is not expected to affect the qualitative predictions made in this study, as all peptides will see a similar hydrophobic surface. From the equilibration trajectory, two-dimensional radial distribution functions for the chain termini demonstrates a well-ordered structure as expected.

The simulated system lacks an explicit gold surface and thus also the van der Waals (vdW) attractions that could potentially lead to lower predicted free energies. However, a rough estimate shows that long-range vdW attractions cause the slope in Figure 2.4 to increase by less than $0.1 k_B T$ /leucine.

A periodic box is defined with dimensions of $\sim 60 \times 60 \times 100 \text{ \AA}^3$ (Figure 2.1), which accommodates an extended peptide length of $\sim 34 \text{ \AA}$ and a non-bonded cutoff of 8 \AA . 10,000 waters are added via the crystal solvation modules in AMBER. Solvated systems are energy minimized and then equilibrated with two separate MD simulations: first heating from 100-300 K over 500 ps at constant volume, then 1 ns of NPT dynamics at 1 bar.

2.2.2 Simulation protocol

Molecular dynamics in all steps are carried out using PMEMD in the AMBER 14 package with explicit TIP3P waters³⁵ and the ff14SB force field.³⁶ Though past studies have shown that the choice of water model can influence the precise behavior of hydrophobic phenomena,^{37,38} here we maintain the TIP3P model to ensure consistency with the chosen protein force field. Throughout, a 2 fs timestep is employed and SHAKE and SETTLE are used to constrain all bonds involving hydrogen atoms.^{39,40} We use a Langevin thermostat with a collision frequency of 2 ps^{-1} and an anisotropic Berendsen barostat.⁴¹ Both are weakly coupled and we verify the absence of significant density variations during nonequilibrium pulling.

In nonequilibrium simulations, the pulling coordinate is defined as the height of the pulled peptide end (the α -carbon opposite the hydrophobic leucine residues) above the fixed carbons of the SAM (Figure 2.2). Pulling is performed at rates of 10 \AA/ns and 2 \AA/ns with a spring constant of $0.251 \text{ kcal/mol\AA}^2$ (350 pN/nm), using a custom restraint

potential in AMBER that pulls only in the z-direction, normal to the surface. To generate an ensemble of initial conformations for forward pulling runs, we simulate the entire system for 20 ns at 300 K and 1 bar using a flat-bottom harmonic restraint with spring constant $0.8 \text{ kcal/mol}\text{\AA}^2$ to keep the peptide end within 18 \AA above the bottom of the surface. A total of 30 configurations for fast-pulling runs, or 60 for slow-pulling, evenly spaced in time, are chosen from the last 18 ns of this trajectory as seeds for forward pulling runs, with randomly assigned initial velocities at 300 K.

Initial conformations for the reverse runs (towards the surface) are obtained from bulk simulations. Each peptide is solvated in a water box that extends at least 20 \AA on each side, then simulated for 100 ns at 300 K and 1 bar. 30 (for fast-pulling) or 60 (for slow-pulling) peptide conformations from the last 90 ns are selected at even time intervals to be placed 60 \AA away from the position of the fixed carbon atoms in the surface. Before performing reverse pulling runs, the resulting system is solvated and equilibrated as already described except that the α -carbon positions of the peptide are restrained during heating and density equilibration to preserve the chosen configuration.

In equilibrium umbrella-sampling simulations, the set-up is similar to that described above for equilibrating the peptide on the surface. Instead of restraining the pulled alpha carbon, we instead restrain the center of mass of all leucine residues, or the center of mass of the isobutane molecule, using a harmonic potential with a spring constant of $2.0 \text{ kcal/mol} \cdot \text{\AA}^2$. Depending on the peptide, 23 to 28 windows from 18.0 to 50.0 \AA away from the fixed carbon atoms of the SAM are used (17 windows 16.0 to 29.0 \AA away for isobutane). For each window, 50 ns (10 ns for isobutane) of simulation time is generated and subsequently used to compute a PMF, except for $(\text{GS})_4\text{GL}$, for which 80 ns was used. This was necessary to converge the calculations due to large conformational fluctuations in the unrestrained glycine/serine portion of the peptide.

2.2.3 Free energy estimates

With the exception of free energies reported from umbrella sampling, all estimates of the free energy are computed with JE²³ using only forward realizations of the nonequilibrium pulling process. Variances in these free energy changes are estimated using 10,000 bootstrap re-samples of the original data. To analyze umbrella sampling simulations, the free energy was computed as the difference between the invariant, asymptotic portion at far distances and the minimum of PMFs generated using the Multistate Bennet’s Acceptance Ratio (MBAR) algorithm⁴² implemented in pymbar.⁴³ It should be noted that these free energy differences are subtly different than those computed via JE in two ways. First, they represent the free energy for moving the center of mass of leucine residues to a position far from the surface rather than the alpha carbon opposite these hydrophobic residues. Second, they do not include the free energy of applying a harmonic restraint, as the effect of this has been removed from the PMF through MBAR. To address the first, PMFs computed over both leucine center of mass and pulled alpha carbon indicate that this free energy difference is similar (Figure 2.5), though significantly more complicated for the case of the alpha carbon. In the case of the second difference, this contribution is expected to be small, as mentioned earlier.⁴⁴

With simulations, it is also possible in principle to make use of reverse pulling simulations of the peptide towards the surface. Reverse moves tend to improve convergence of free energy estimates^{45–47} and allow use of Bennet’s Acceptance Ratio (BAR)^{48–50} approach, which often improves upon the forward-only JE for an equivalent number of trajectories.^{47,51} While this proves true for the simulations at faster pulling rates, poor convergence of the reverse distributions in slower-pulling simulations proves detrimental to these estimates. Comparisons of free energy estimates from both JE and BAR, as well as detailed methodologies to account for statistical and sampling errors in these

estimators, are discussed later on.

2.3 Results and Discussion

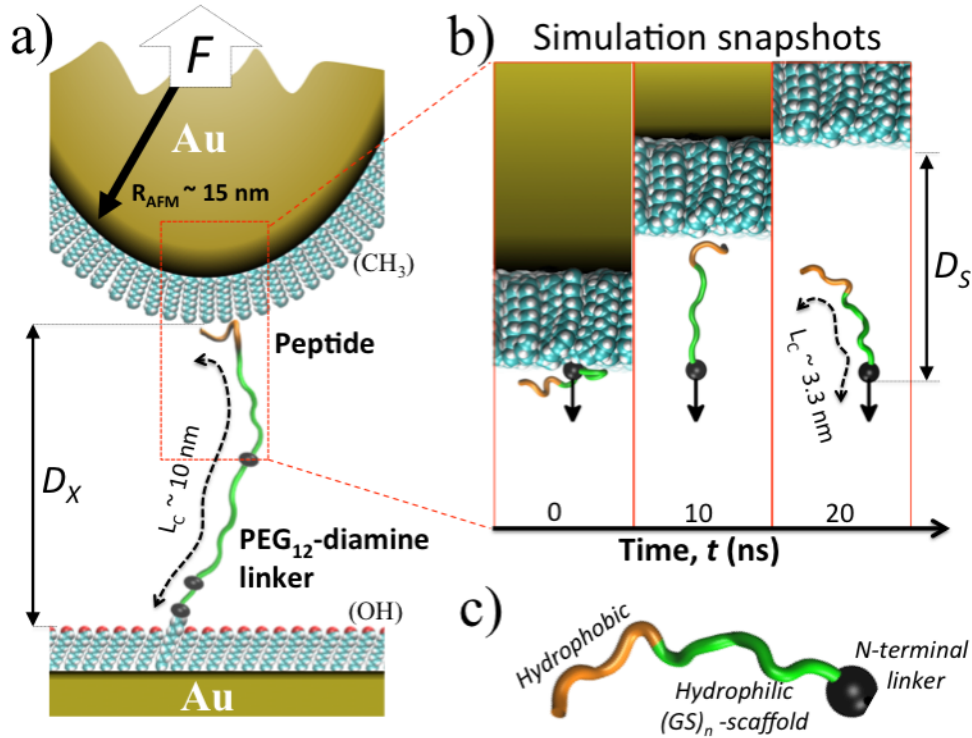


Figure 2.2: Schematic of (a) experimental and (b) computational setting used for quantifying dynamic hydrophobic interactions between a set of different hydrophobic peptides (shown in Table 2.1) and an extended hydrophobic surface as a function of the experimental distance D_X and the simulation distance D_S , respectively. (c) Hydrophobic mutations are introduced specifically at the C-terminal tail end of a hydrophilic $(GS)_n$ based scaffold. In experiment (a) a PEG12-diamine-based polymer linker is used to covalently graft peptides on a molecularly smooth OH/COOH terminated self-assembled monolayer on a gold surface (Roughness, $\sigma_{RMS} \leq 2 - 3 \text{ \AA}$). The contour length of $L_C = 10 \text{ nm}$ is indicated. Gold-coated AFM tips (radius, $R_{AFM} \sim 15 \text{ nm}$) are hydrophobized with a self-assembled monolayer and interactions between different surface grafted peptide mutations (see Table 2.1) and the hydrophobic tip is measured. In the simulation experiment the PEG linker is not included and the contour length of the individual peptides amounts to $L_C \sim 3.4 \text{ nm}$. In (b) the N-terminal peptide end and applied pulling direction is marked with an arrow.

A comparison of our MD systems to the SM-AFM experimental setup is shown in

Figure 2.2. In experiment, fully designed peptide sequences with varying degrees of hydrophobicity are end-grafted with their N-terminus to a polyethylene glycol (PEG) linker, with 12 PEG units and a terminal COOH group (PEG12-COOH), on ultra-smooth gold surfaces modified with self-assembled monolayers ($\sigma \approx 3 \text{ \AA}$).^{5,52} In experiments and our simulations, a hydrophilic (GS)₅ model backbone of alternating glycine (G) and serine (S) amino acids has its C-terminal amino acids selectively mutated with up to 4 consecutively distributed hydrophobic leucine (L) units to adjust the peptide hydrophobic content (Table 2.1). The extended protein with PEG anchor has an estimated contour length, L_C , of 10 nm. The grafted peptide sequences on the PEG12-linker are brought into contact with a 1-undecanethiol terminated hydrophobic gold-coated AFM tip, and single molecule experiment force versus distance characteristics are recorded in so-called fly-fishing mode.⁵³ In this experimental AFM technique, a tip contacts a surface at random points and if – in our context - a single grafted peptide adsorbs at the hydrophobic tip during contact, the formed hydrophobic contact breaks during separation of the tip and the surface, leading to a specific signature in the force profile.⁵³ The experimental data is considered reliable and kept for analysis if the measured contour length matches ≈ 10 nm and the force versus extension curve may be fit well with a worm-like-chain (WLC) model. This ensures that the measurement is only of the peptide’s interactions with the hydrophobic tip surface and does not include surface defects or “dirt.”

The atomically-detailed simulated system (shown in Figure 2.2B) is designed to closely approximate both the physical system and the experimental protocols. The N-terminus of each of the peptides, which in experiment is tethered to a PEG linker and ultimately a hydrophilic surface, is pulled at constant velocity away from a fixed, periodically replicated hydrophobic SAM surface approximating the AFM tip. Numerous models of SAM surfaces,^{29,27,28,30,31} as well as SAM-peptide systems,^{33,34,54–58} have been proposed in the literature; here we build the SAM from dodecane molecules with a single end fixed

Mutation sequence	Experiment ΔG ($k_B T$)	Simulation fast-pull ΔG ($k_B T$)	Simulation slow-pull ΔG ($k_B T$)	Simulation equilibrium umbrella sampling ΔG ($k_B T$)	Three-body angle model ΔG ($k_B T$) *
-N(GS) ₅	Below detection limit	20.9 ± 1.7	7.4 ± 0.5	4.6 ± 0.1	N/A
-N(GS) ₄ GL	7.6 ± 0.8	17.5 ± 3.2	10.3 ± 0.7	8.0 ± 0.1	7.8
-N(GS) ₄ L ₂	8.8 ± 0.9	22.9 ± 4.6	13.4 ± 2.4	7.9 ± 0.1	10.5
-N(GS) ₃ GL ₃	12.2 ± 1.1	31.1 ± 2.4	10.5 ± 3.3	11.6 ± 0.1	13.0
-N(GS) ₃ L ₄	17.9 ± 0.8	33.3 ± 2.1	16.7 ± 1.5	13.1 ± 0.1	14.6

* Errors are not provided for ΔG values from the three-body angle model since this is already an approximate model utilizing reference probability distributions for which the uncertainties are difficult to rigorously assess.

Table 2.1: Peptide mutation sequences and summary of experimentally obtained and simulated interaction free energies with extended hydrophobic surfaces.

at the necessary lattice spacing, with surrounding water explicitly represented. Additional methodological details are described in the Materials and Methods.

For comparison to simulated results shown later, experimental results (provided by collaborators) in Figure 2.3A show a master curve with ~ 80 individual trajectories recorded during unbinding of the (GS)₃L₄-sequence (see Table 2.1) from a hydrophobic AFM tip. The inset in Figure 2.3A shows a typical individual force versus distance curve, with the distance normalized by the contour length, L_C , of the molecular anchor. It should be noted that due to the density of peptides attached to the hydrophilic surface, not all surface to hydrophobic tip interactions yield signatures that involve peptide interactions. In fact, only 3-8% of experimental force profiles indicated successful surface-to-molecule binding and hence allowed direct dynamic measurement of work-trajectories during unbinding of single-molecule hydrophobic interactions. Similar master curves were recorded at different loading rates, as well as for all other peptide sequences. Curves without successful bond formation show only van der Waals, hydration and electric double layer interactions that can be fit well with an extended DLVO theory,^{59,60} while those involving peptide interactions also require a WLC fit.⁴⁴

On the basis of the datasets of individual pulling-trajectories, the interaction free

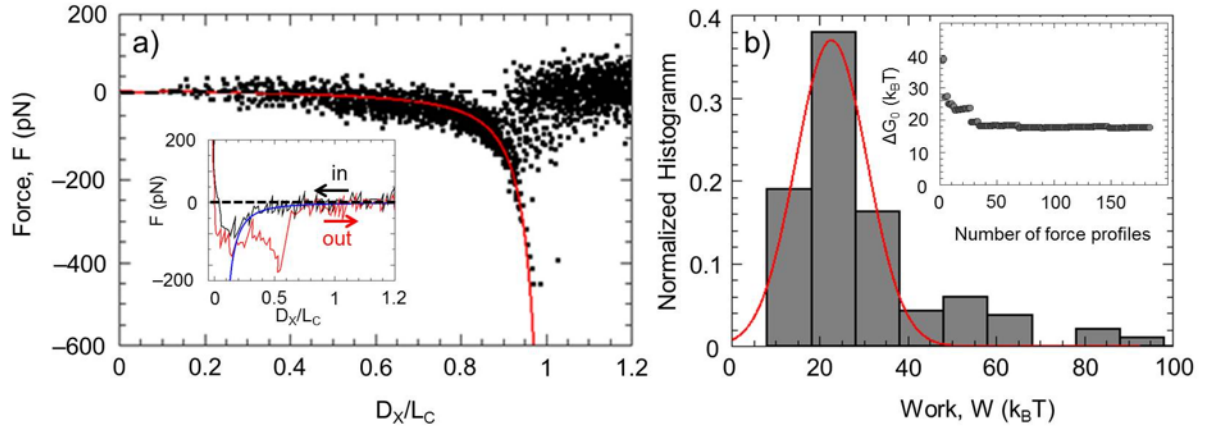


Figure 2.3: (a) Typical master curve with about 80 individual rupture events measured between $-N-(GS)3LLLL$ functionalized surface and hydrophobic tip, aligned by the best worm like-chain fit with a contour length of 10 nm and a persistence length of 0.37 nm. Inset shows a typical individual SM-AFM force distance profile with both a primary adhesive minimum and a single molecular rupture signature at 60% full extension of the linker. The fit in the inset shows a fit by an extended DLVO theory. (b) Normalized histogram of the measured work distribution. The inset in (b) shows a plot of Jarzynski's free energy ΔG_0 as a function of the number of force trajectories.

energy difference between the bound and unbound states of the different pulled peptides with the hydrophobic tip surface can be estimated using Jarzynski's equality (JE):²³

$$e^{-\frac{\Delta G}{k_B T}} = \langle e^{-\frac{W}{k_B T}} \rangle \quad (2.1)$$

The free energy difference, ΔG , is related to an exponential average of the mechanical work W over an infinite number of realizations of the same nonequilibrium protocol that probe the full work distribution. In practice, an estimate for the free energy difference is given by the same equation using a finite number of realizations n and work values W_i ($i = 1, 2, \dots, n$) calculated from individual pulling curves,^{44,61} such that the true difference is approached as $n \rightarrow \infty$.

Through Jarzynski averaging, one can hence exactly obtain the equilibrium free energy difference between the beginning and end states of a non-equilibrium process.

A definition of these states is important for identifying and assessing the dissipated work in an experimental setting. In our experiments, the beginning state consists of a relaxed linker and peptide interacting with an AFM tip in equilibrium, while the end state corresponds to both the linker and peptide similarly relaxed (i.e., under no tension) but with the tip far away from the peptide. Thus, the equilibrium free energy difference between these two equilibrium states consists nearly totally of the contribution from the hydrophobic peptide-AFM tip interaction. We note that the macroscopic AFM spring is in zero position before and after the integration starts. Similarly, the harmonic restraint potential applied to pull the peptide during simulation also starts and ends in an equilibrium position fluctuating around zero. As a consequence, and as shown in previous experimental work, integrating the measurable work from $t = 0$ where the force $F = 0$ until post-bond rupture (where $F = 0$ again), provides an excellent estimate of surface/molecule⁴⁴ and intermolecular binding free energies.⁶²

In practice, the experimentally measured work and calculated free energy include contributions from the PEG tether that are not included in our simulations. As the peptide/linker system is extended, energy will be stored in the PEG prior to bond rupture, after which point the tether relaxes back to equilibrium and this stored energy is dissipated. Thus, the linker will tend to bias work distributions to higher values, further from equilibrium behavior, and in turn also affect how easily the Jarzynski average can be converged with a finite number of runs. Differences in linker configurations in tip contact and withdrawn states may also affect the measured free energy differences. This contribution is expected to be small and increasingly negative with PEG length as the linker gains entropy upon full retraction of the AFM tip (beyond bond rupture). Recent experiments,⁶³ however, have shown that the average measured work and binding free energy increase with tether length. This confirms that contributions to the equilibrium free energy difference are essentially undetectable and that bias attributable to dissipative

work associated with the PEG tether is much more significant. In the experiments that we compare to, a short linker was used such that biases are essentially negligible from an experimental point of view.⁶³ Error estimates and bias are discussed in more detail below, though in the context of simulation results. We also note that Hummer and Szabo discuss in great detail the physics of extracting specific free energy contributions from non-equilibrium measurements and their evolution along a pulling coordinate that involves linker stretching measured by macroscopic spring devices.^{64,65}

Using JE, the interaction free energies for all designed peptide permutations interacting with the hydrophobic tip or interface for both experiments and nonequilibrium simulations. In Figure 2.4 the estimated JE interaction free energies are shown for all peptides. Interestingly, the overall interaction scales approximately linearly with the number of leucine units, with fits showing that each additional hydrophobic unit contributes $3.4 \pm 0.4 k_B T$ for consecutive and $3.0 \pm 0.4 k_B T$ for non-consecutive leucine mutations (see again Table 2.1). This falls within the range of values measured by macroscopic bulk methods for various interacting hydrophobes,⁶⁶ and is consistent with simulation dimerization free energies for methane (~ 1.2 - $3.4 k_B T$),^{47,48} cyclohexane ($4.6 k_B T$),⁴⁶ and benzene ($3.4 k_B T$).⁶⁷

Interestingly, the hydrophobic interaction free energy of 3.0 - $3.4 k_B T$ per leucine unit describes all mutation results within experimental uncertainty. Remarkably, the linear trend of 3.0 - $3.4 k_B T$ per leucine unit holds true even when hydrophobic units are separated by a glycine rather than next to each other,²² though the former case is the only one considered in simulation. This experimental observation suggests that the mutations used are intrinsically unstructured and interact non-cooperatively with the hydrophobic surface. Molecular dynamics simulations of the peptides in bulk and at the surface support this interpretation, showing that none of the consecutive mutations have pronounced formation of secondary structures (Figures A.1-A.4). In solution, all simulated peptides exhibit

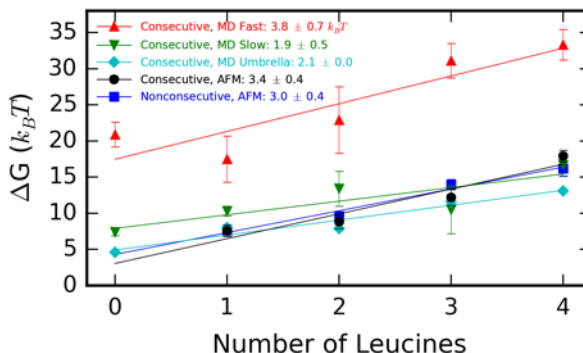


Figure 2.4: Comparison of experimentally and theoretically estimated interaction free energies as a function of the number of leucines. The interaction free energy is referenced to the one leucine case. For the fast-pulling (10 Å/ns) and slow-pulling (2 Å/ns) MD cases, the R^2 values for the linear fits are 0.71 and 0.80, respectively. All other R^2 values are greater than 0.90.

similar and somewhat broad distributions for radius of gyration (R_g), solvent-accessible surface area (SASA), and number of intrapeptide hydrogen bonds that is suggestive of a diverse structural ensemble. The same distributions do reveal some sequence-specific signatures when the peptides are near the hydrophobic surface, but they remain broad and are consistent with unstructured behavior.

These observations strongly suggest a linear model for the peptide-surface interaction free energy with the number of leucines present:

$$\Delta G = n_L(\Delta g_L - \Delta g_{GS}) + n_{tot}\Delta g_{GS} \quad (2.2)$$

Here, n_L is the number of leucines, n_{tot} is the number of residues excluding caps, and Δg_L and Δg_{GS} are the free energies of removing a single leucine or (averaged) glycine/serine residue from the surface, respectively. Thus, in Figure 2.2, the intercept is given by the interaction free energy of the scaffold with the surface, and the slope by the difference in interaction free energy of a leucine residue versus that of the scaffold (Table 2.4). Similar additivity for the solvation of small hydrophobic monomer units in a homopolymer

was previously noted by Li and Walker.⁶⁸ Within experimental uncertainty and in the context of this model the mutation of a peptide residue from hydrophilic to hydrophobic has an additive effect on its interaction with a hydrophobic surface, regardless of the proximity (spaced or adjacent) to other hydrophobic monomers. Thus, experiments did not detect any cooperativity or anticooperativity between the hydrophobic units, although it is possible that such effects may be present at a level smaller than the experimental uncertainty.

Simulation-calculated free energies of peptide removal follow the same approximately linear trend as in experiments, but are slightly higher than those determined experimentally and increase with pulling rate (Table 2.1). This is not surprising for at least two reasons: simulated pulling rates are necessarily orders of magnitude faster than the $\sim 10^{-4}$ Å/ns of experiments, and the simulations are also limited to fewer pulling runs in each case (60 at the slowest rate). It is well-known that both factors magnify the statistical bias introduced by Jarzynski's equality for finite sample sizes,⁶⁹ with this bias increasing with pulling rate and nonequilibrium driving. From Eqn. 2.1, it is evident that this bias originates in the asymmetric weighting by the exponential of low work values in the averaging procedure; thorough sampling of the rare trajectories, with work values well below the mean, are thus necessary for convergence. When the number of runs is limited, as here, our calculations (Table 2.3) show that a single pulling curve with a low work value skews the free energy calculations.

To qualitatively assess the bias our simulations, we use the methods presented by Lu and Kofke⁷⁰ and Lu and Woolf,⁷¹ where the bias in the forward Jarzynski estimate is given by the integral over the distribution of negated reverse works from negative infinity to the most likely minimum value of the forward work:

$$Bias_{JAR,F} = \Delta G_{JAR,F} - \Delta G = -\ln\left(1 - \int_{-\inf}^{min(W_F)} P_R(-W)dW\right) \quad (2.3)$$

Peptide	Fast pulling ($v = 10\text{\AA}/ns$)			Slow pulling ($v = 2\text{\AA}/ns$)		
	Jarzynski, Forward	Jarzynski, Reverse	BAR	Jarzynski, Forward	Jarzynski, Reverse	BAR
-N(GS) ₅	19.7	-10.2	4.8	4.0	-2.6	0.7
-N(GS) ₄ GL	15.4	-8.8	3.3	5.2	-3.2	1.0
-N(GS) ₄ L ₂	17.7	-12.6	2.5	7.3	-6.7	0.32
-N(GS) ₃ GL ₃	26.6	-14.2	6.2	5.1	-5.1	0.0
-N(GS) ₃ L ₄	24.3	-13.6	5.3	16.7	-7.6	4.5

Table 2.2: Approximate bias in free energy estimates from each methodology at both pulling rates. While increasing trends with the number of leucines suggest slopes are overestimated, the fast-pulling BAR formula yields fairly similar biases, suggesting less error for this estimate. Note that these estimates of the bias still depend on good estimates of the work distributions, possibly explaining the large biases estimated for the forward Jarzynski case at a slower pulling rate.

In the above, $\Delta G_{JAR,F}$ represents the free energy as estimated by the application of Jarzynski’s Equality to forward trajectories and ΔG to the true free energy difference. We set the upper limit to the lowest sampled forward work for each peptide and use Gaussian probability density fits, rather than those proposed by Lu and Kofke.⁷⁰ Since the minimum observed forward work for each peptide is much greater than the average of the negated reverse work distribution, the use of a Gaussian is expected to overestimate bias. Table 2.2 gives estimates for the forward, reverse, and BAR biases, the latter using the fact that the Bennet’s Acceptance Ratio (BAR) formula is well-approximated by the average of forward and reverse estimates in the limit of large dissipative work.⁷² It is clear that the magnitude of all biases generally increases with the number of leucines at both fast and slow pulling rates. Thus the net effect of the bias is to increase the effective slope in Figure 2.4, that is, the free energy per leucine. In experiments, the bias is estimated to be low and does not seem to affect the slope in Figure 2.4,²² such that the experimental hydrophobic interaction per leucine appears well approximated in the quoted range from 3-3.4 $k_B T$.

The use of both forward and reverse pulling simulations (towards the surface) with bidirectional free energy estimators, such as BAR, has also been shown to reduce systematic

bias in nonequilibrium studies like this.⁵¹ The reverse protocol is not possible in experiment for the present scenario, but we do compute these with simulations. Still, we find that use of the reverse protocol does not enhance bias or variance, as has been observed previously,^{51,70} for systems in which either the forward or reverse work distribution is very difficult to converge.^{51,70} Here the BAR formula is also sensitive to rare, negative reverse work values from trajectories in which the pulling restraint actively fights against the motion of the peptide as it is attracted towards the surface. Close inspection of Figure 2.6A reveals that these rare trajectories are not well-sampled across all peptides due to the stochastic nature of the simulations.

Peptide	Fast pulling ($v = 10\text{\AA}/ns$)			Slow pulling ($v = 2\text{\AA}/ns$)		
	Jarzynski, Forward	Jarzynski, Reverse	BAR	Jarzynski, Forward	Jarzynski, Reverse	BAR
-N(GS) ₅	20.9 ± 1.7	-5.9 ± 1.3	7.5 ± 1.1	7.4 ± 0.5	2.0 ± 0.9	4.5 ± 0.4
-N(GS) ₄ GL	17.5 ± 3.2	-4.3 ± 2.4	6.6 ± 2.0	10.3 ± 0.7	4.9 ± 1.4	7.1 ± 0.8
-N(GS) ₄ L ₂	22.9 ± 4.6	-6.5 ± 0.6	8.2 ± 2.3	13.4 ± 2.4	2.5 ± 0.6	8.0 ± 1.2
-N(GS) ₃ GL ₃	31.1 ± 2.4	-6.6 ± 1.5	12.2 ± 1.4	10.5 ± 3.3	4.1 ± 1.6	7.5 ± 1.8
-N(GS) ₃ L ₄	33.3 ± 2.14	-6.3 ± 0.8	13.5 ± 1.1	16.7 ± 1.5	1.4 ± 0.9	9.1 ± 0.9

Table 2.3: Free energy estimates from Jarzynski’s equality for both forward and reverse distributions, as well as from BAR. In the fast-pulling case, the reverse work distributions are similar across all peptides, yielding similar Jarzynski estimates. However, for the slow-pulling case, relatively few very negative work values result in widely varying reverse Jarzynski estimates.

For this reason, we find that the combined forward and reverse BAR protocol does not always lead to improved estimates of free energies (Table 2.3). For fast-pulling simulations, where errors from both forward and reverse distributions are large, neither direction is significantly better at predicting the free energy difference, and BAR is beneficial. At slower pulling speeds, convergence of the forward distribution is more rapid than the reverse, and the BAR formula is ultimately not helpful. This is immediately suggested by the idea that faster convergence will occur for the direction for which the conjugate protocol has the smaller dissipative work.⁷³ Table 2.3 shows that there is large variation

between peptides for the reverse Jarzynski estimate at slower pulling because rare, negative work values are not sampled evenly amongst peptides. This translates to BAR estimates that are extremely sensitive to the presence of these rare events, which are characterized by the peptide randomly extending its hydrophobic portion to the surface early in the simulation and in a sense reeling itself in.

To better understand the limitations of the nonequilibrium pulling measurements and Jarzynski’s equality, we perform comparative equilibrium simulations to calculate free energy differences. Specifically, we use the umbrella sampling approach in which the peptide-surface z-distance fluctuations are restrained to a small window using a harmonic potential. Thermodynamic reweighting techniques (i.e. MBAR) then link together multiple such simulations, with systematically varying windows, to produce detailed free energy curves and differences (Figure 2.5). Free energy differences are computed between a potential of mean force (PMF) curve’s minimum and asymptote at far distances. Interaction free energies computed from these equilibrium simulations are generally smaller than those from JE (Table 2.1), as expected, yet yield a slope that is in good agreement with slower-pulling simulation runs at $2.1 k_B T$ per leucine.

Finally, we note that the free energy to separate the hydrophilic (GS)₅ scaffold is not measurable in experiment and so not compared here. Table 2.1 shows that the scaffold has a $\sim 3 k_B T$ smaller interaction free energy than the (GS)₄GL peptide when calculated from simulations. Interestingly, the simulation runs at faster pulling rates predict the scaffold affinity to be large, which is an artifact due to a biasing of the configuration of (GS)₅ near the surface due to the applied restraint (see Appendix A, Figure A.1); the bias is less significant at slower pulling rates that allow the peptide to well-equilibrate with the surface. It should also be re-emphasized that, while the presence of a surface limits the conformational space explored by the peptides compared to fully solvated configurations, all lack well-defined structures in both states (Figures A.2-A.3). Moreover,

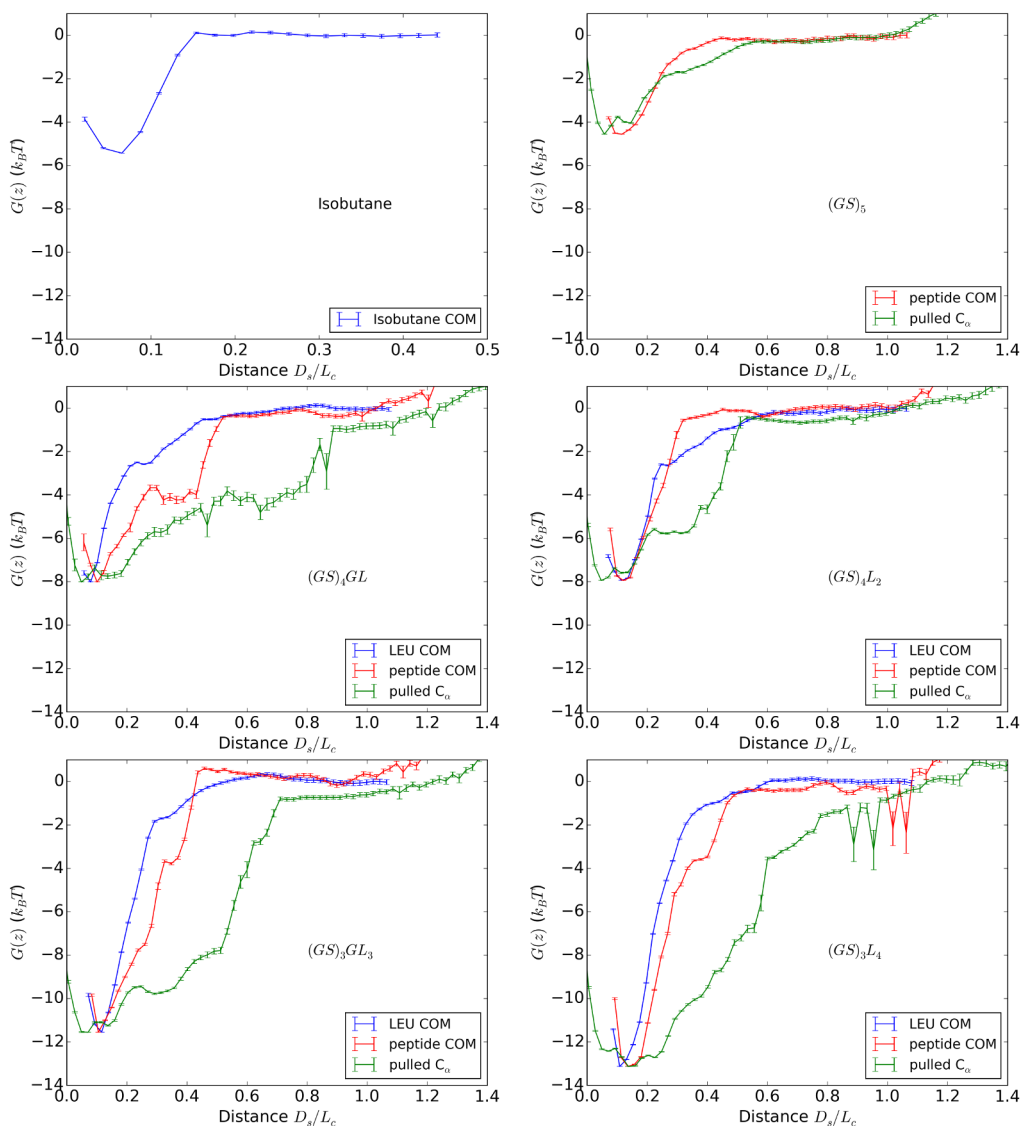


Figure 2.5: Potentials of mean force from umbrella sampling are shown with reference to three reaction coordinates: the center of mass of all leucine residues, the alpha-carbon pulled in non-equilibrium runs, and the peptide center of mass in the direction perpendicular to the interface, the z -axis. Distances for each coordinate begin at the SAM surface and are normalized to the peptide contour length. Though the abscissa for the isobutane case has been truncated for clarity, the PMF minimum is much sharper in this case. All error bars are those computed by pymbar.⁴³

all peptides have relatively fast conformational rearrangement times (assessed by the R_g autocorrelation time) at 0.5-2.0 ns on the surface and slightly longer in solution (Figure A.4).

While experiments are limited to analyzing interaction free energies and measured forces, the nonequilibrium simulations, which nearly quantitatively reproduce these experimental results, suggest a molecular, mechanistic picture of the pull-off process and in turn various contributions to the work distributions. For example, Figure 2.6B shows that detachment for (GS)₅ is quite different from that of the other peptides; once even a single leucine is introduced, the peptide extends substantially prior to detachment in order to maintain leucine residues near the surface. This trend is manifested in peptide pull-off time and distance, radius of gyration, numbers of shell waters, and solvent-accessible surface area (Figures A.5-A.6 in Appendix A).

Simulations also show different regimes of the work-distance curves associated with distinct underlying molecular-mechanistic processes (Figure 2.7): (1) initial perturbation from a stable surface configuration, (2) low-tension peptide extension, (3) high-tension extension in which the peptide exhibits WLC behavior, and (4) dragging through solution. The work for the (GS)₅ peptide primarily lies in the first regime, but as leucines are added, contributions from the high-extension regime grow, reflecting the increased strength of the peptide-surface interaction. For all leucine-containing peptides, work contributions for surface perturbation and dragging through solution are roughly equivalent. Interestingly, this matches the expectation that hydrophobic interactions, and thus surface affinity, should grow with the number of leucine residues present, with the largest energy barrier to detachment being the creation of hydrophobic surface areas.

Importantly, the simulations reveal molecular-structural factors associated with the linear increase in binding free energy with the number of leucines. For example, the bare scaffold shows an increase in peptide-water and a decrease in intrapeptide hydrogen bonds

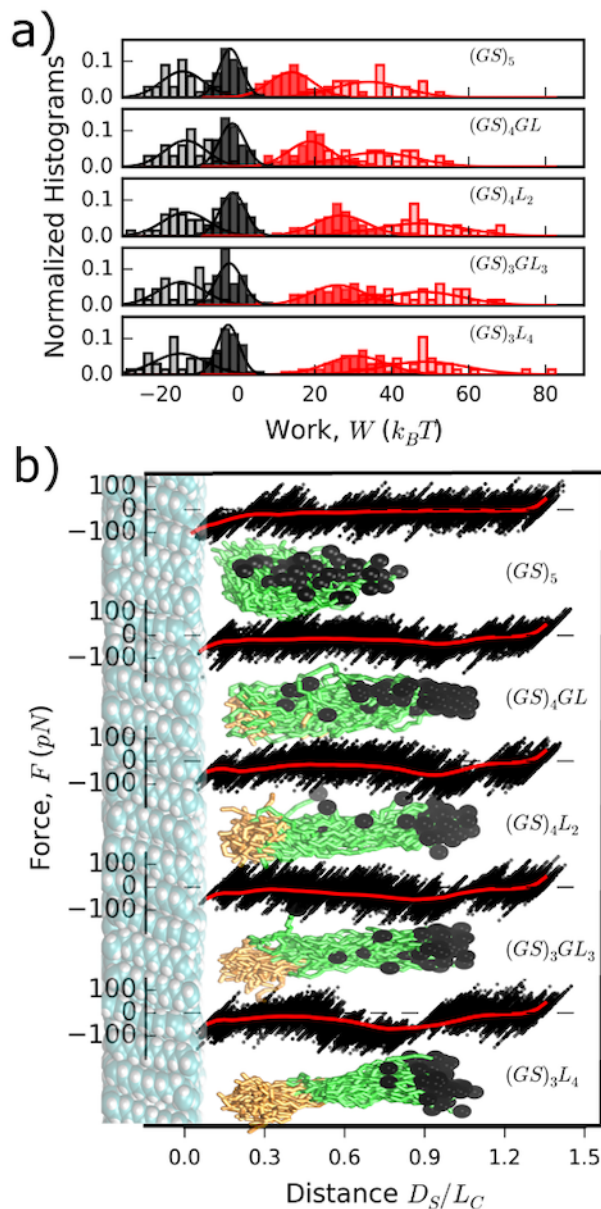


Figure 2.6: (a) Forward, $P(W_{for})$ (red), and negated reverse, $P(-W_{rev})$ (black), work distributions show that slower pulling rates (2 \AA/ns , solid bars) significantly reduce the hysteresis and spread observed at faster pulling speeds (10 \AA/ns , lightly shaded). At slower pulling rates, many negative work values (positive here, as negated reverse distributions are shown) are observed during reverse pulling runs. (b) Black dots are force versus distance profiles from single, typical trajectories and red lines are force profiles averaged over all trajectories. Inset structures show peptide chains in green (hydrophilic) and orange (hydrophobic) and pulled atoms (black) at the time of pull-off for all trajectories of a particular peptide. These structures are sized and aligned such that they are proportional to the scale on the x-axis.

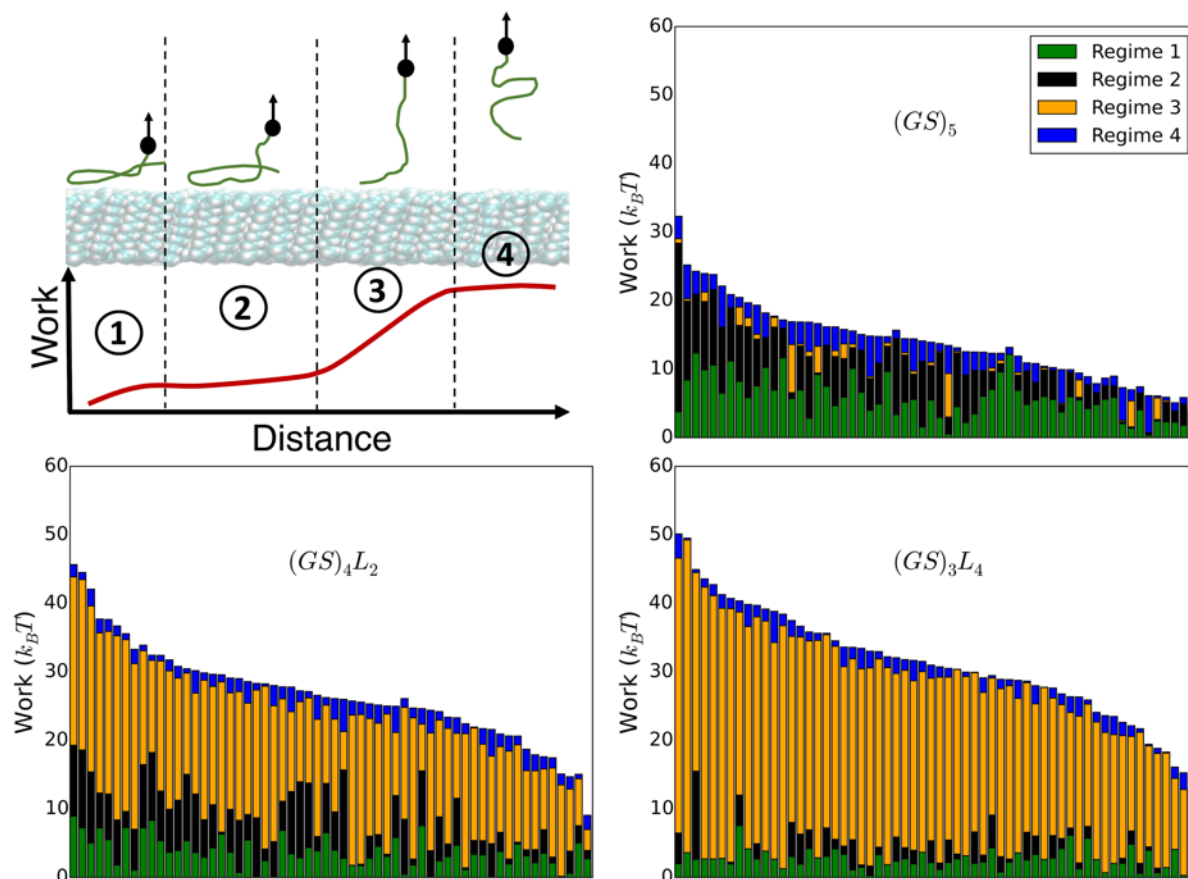


Figure 2.7: Contributions to the work are shown from each pulling regime for slow-pulling runs of the simulated peptides. Pulling regimes correspond to changes in the slope of the work versus distance curve, and are physically defined as: (1) from the beginning of the simulation until the pulled atom is off the surface, (2) until the centers of geometry of one less than the number of non-leucine residues are off the surface, (3) until the last atom in the peptide is off the surface, and (4) from pull-off until the end of the simulation. For all events described, the cut-off for being on the surface was set to a height of 21 Å above the fixed carbons of the dodecane molecules. Without any leucines, the largest contribution to the work appears in regime 1, or the perturbation of the peptides configuration on the surface. With addition of leucines, stronger interactions result in larger contributions to the work after the peptide is extended (regime 3).

upon its removal from the surface, but the trend gradually shifts as leucines are added until it is inverted for the $(GS)_3L_4$ case (Figure 2.8). Moreover, the additional hydrophobic surface area exposed upon detachment from surface into solution generally increases with the number of leucines, and such that each leucine independently contributes a relatively equal amount to the total change in hydrophobic surface area. As more are added, however, the peptide backbone geometry does not permit all leucines to directly contact the surface, forcing at least one leucine in the $(GS)_3GL_3$ and $(GS)_3L_4$ cases to remain partially or fully solvated when the peptide is near to the SAM (see also Figure A.1 for visual renderings).

In such molecularly-sized systems, it is not expected that the changes in interaction free energies should follow a macroscopic trend by scaling with the hydrophobic surface area. Indeed, we find that the effective surface tension, computed as the free energy change per change in hydrophobic SASA, varies with the number of leucines. Such an analysis neglects the contributions of the hydrophilic portion of the peptide, but more importantly ignores well-known differences in the nature of the HI for small and large hydrophobes.⁷⁴ Namely, hydration of small hydrophobes ($< 1nm$) tends to be entropically unfavorable due to water's unique cavity fluctuations and scales with cavity volume, whereas large ones ($> 1nm$) pay an enthalpic penalty for solvation that scales with exposed solute surface area.⁷⁴ Recent efforts have also nuanced these ideas in terms of water-solute interaction energies,¹⁰ solute-local density fluctuations,⁷⁵ and orientation of water dipoles.⁷⁶ However, it is difficult to directly apply these kinds of analyses to the present peptide-surface systems due to the fluctuating, non-rigid nature of the peptide conformations and to the statistical accuracy with which the relevant quantities can be determined in these large, computationally expensive runs.

To seek molecular signatures that may delineate distinct regimes of HI physics, we thus characterize hydration water structure at the initial and final equilibrium states of

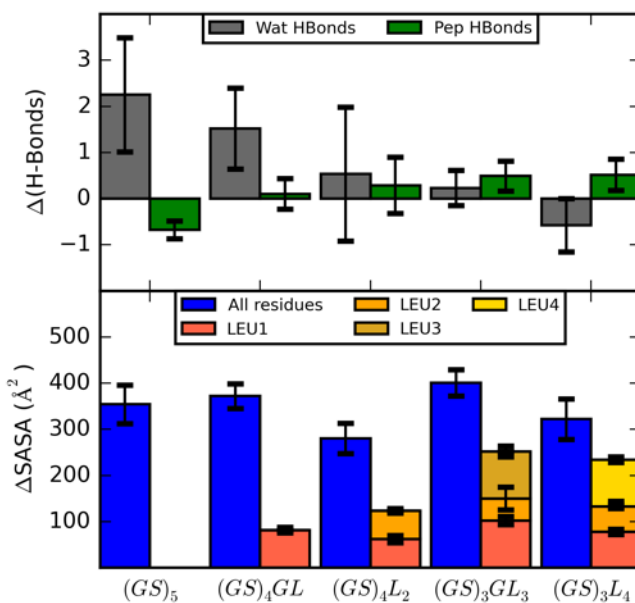


Figure 2.8: Change in average peptide properties moving from a surface-restrained state to solution. (a) With increasing leucines, peptides gain fewer peptide-water and more intramolecular hydrogen bonds in moving from the surface to solution, consistent with the slight increase in α -helical tendency (Figure A.1 in Appendix A). (b) Changes in SASA are similar between peptides, while changes in hydrophobic SASA roughly increase with the number of leucines. Contributions to ΔSASA_{hyd} are broken down by leucine: LEU1 refers to the leucine closest to the C-terminus of the peptide, LEU2 the next inwards along the chain, and so on. Note that for $(GS)_3L_4$, backbone constraints prevent all four leucines from simultaneously becoming solvent-shielded in the surface-associated state (Figure A.1).

the pulling process (Figure 2.9). Longstanding views have sought to understand HIs in terms of water’s unique tetrahedral correlations,⁷⁷⁻⁸¹ and to quantify these, we calculate distributions of the angle formed by triplets of co-neighboring water molecules within a 5 Å hydration shell of both the peptide and SAM surface. In bulk water this distribution peaks near the tetrahedral angle (109.5°), but surfaces and molecules can either induce enrichment or depletion of tetrahedral, random ($\sim 90^\circ$, ideal gas-like), or icosahedral ($\sim 60^\circ$, simple-liquid-like) populations.⁷⁷ As controls for comparison, we characterize how the extended hydrophobic SAM surface alone perturbs water structure (large-scale physics), versus how a single solvated isobutane, a leucine side-chain analogue, does the same (small-scale). We find that the bare SAM enriches the uncorrelated and icosahedral populations of waters, indicating a reduction in tetrahedral configurations, while the opposite is true for isobutane.

These reference cases provide context for the interpretation of similar distributions near interfacial and bulk-solvated peptides. For peptides on the surface, nearby waters exhibit behavior like those near the surface itself, with very little sequence-specific variation. In this context, it appears that water near the SAM surface (with or without peptide) exhibits weaker tetrahedral preference, as expected from large-scale hydrophobicity theory.⁷⁴ Interestingly, however, when the peptides have been pulled off and into solution, the tetrahedral population grows and increasingly so with the number of leucines. While a cluster of one or two leucines should still act as a small hydrophobe in solution, three or four if grouped together should begin to enter the large-scale regime beyond 1 nm. The surprising fact that the tetrahedral region of the angle distribution systematically increases with the number of leucines suggests that each is solvated in an independent fashion from the remaining, possibly due to geometric restrictions imposed by the backbone. Similar trends are also observed in the solvation free energies of moderately short-chain n-alkanes showing that collapse is not favorable, leaving the alkane in an extended conformation.⁸²

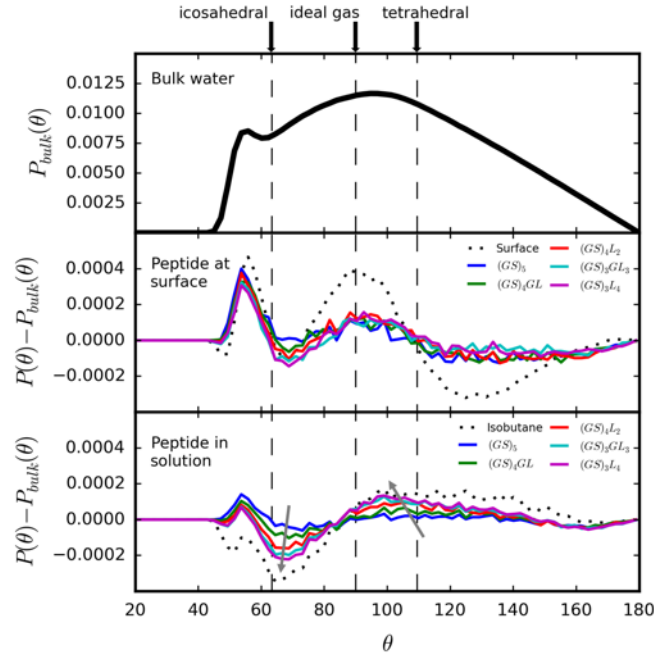


Figure 2.9: (Top panel) Probability density distribution of the triplet angle formed by water molecules with two nearest neighbors, in which the angle is measured at the oxygen site. For comparison, a simple Lennard-Jones fluid with icosahedral coordination has a distribution peaked slightly below 60° with a much smaller secondary peak just above 109° . An ideal gas, non-interacting fluid has a broad distribution with a peak near 90° . (Middle and bottom panels) Changes to the triplet angle distribution for the peptide-SAM systems. The dashed curve in the central figure pertains to hydration waters within 5 \AA of the SAM surface, while the dashed curve at bottom refers to waters near a fully solvated isobutane molecule. Other curves give the perturbation to the system distribution for each peptide either restrained on the surface (middle) or far away in solution (bottom panel). Gray arrows show changes with increasing leucines, which, with the exception of the 3 and 4 leucine distribution comparison, are found to be statistically significant at the 5% confidence level by the Kolmogorov-Smirnov test.

Because the angle distributions of the peptides on the surface all look similar, one might imagine that the sequence variation in the bulk-solvated peptides (far from the surface) may signal the variation of association free energy alone. To examine the relation of shifts in three-body angle distributions to interaction free energies, we conduct equilibrium umbrella-sampling simulations for an isobutane molecule with the SAM surface, which gives a surface-to-bulk free energy change of $5.4 k_B T$ (Figure 2.5). We define a per-water solvation penalty given by $5.4 k_B T$ normalized by the average number of waters in the isobutane hydration shell. We hypothesize that the number of waters around a peptide that exhibit three-body angles similar to the isobutane distribution contribute this penalty to the overall association free energy. Similarly, we hypothesize that the remaining waters exhibit an angle distribution associated with the (GS)₅ scaffold, and contribute based on a reference scaffold per-water solvation free energy.

$(k_B T)$	Experiment				
	Nonconsecutive	Experiment Consecutive	Simulation fast-pull	Simulation slow-pull	Simulation equilibrium
Slope	3.0 ± 0.4	3.4 ± 0.4	3.8 ± 0.7	1.9 ± 0.5	2.1 ± 0.0
Intercept	4.3 ± 1.0	3.1 ± 1.0	17.5 ± 1.9	7.9 ± 0.7	4.9 ± 0.1
Δg_{GS}	0.4 ± 0.1	0.3 ± 0.1	1.7 ± 0.2	0.8 ± 0.1	0.5 ± 0.0
Δg_L	3.5 ± 0.4	3.7 ± 0.4	5.6 ± 0.7	2.7 ± 0.5	2.6 ± 0.0

Table 2.4: Fitting parameters for free energy model linear in the number of leucines. Δg_{GS} is the free energy change per residue of removing the scaffold from the surface and Δg_L is the free energy change of removing a single leucine residue.

To apply these ideas, we fit each peptide’s three-body angle distribution as a weighted sum of the distributions for the (GS)₅ scaffold in solution and isobutane in solution, which bracket all perturbations of the bulk water distribution made by the hydrophobe-containing peptides (Figure 2.9). This gives f_{iso} , the fraction of the distribution that looks like the isobutane one versus that of the scaffold. Given the model assumptions, a

simple model for the association free energy is

$$\Delta G = N_{W,pep} \left(f_{iso} \frac{\Delta G_{iso}}{N_{W,iso}} + (1 - f_{iso}) \frac{\Delta G_{GS}}{N_{W,GS}} \right) \quad (2.4)$$

Here, $N_{W,x}$ is the number of shell waters around either $x =$ the peptide, isobutane, or (GS)₅ scaffold, and ΔG_x is the free energy of removing such a group from the surface. Remarkably, consideration of only water structure near each leucine-containing peptide produces free energies via the above model that closely agree with the actual values (Table 2.1), with a slope of $2.3 k_B T/\text{LEU}$, only slightly higher than $2.1 k_B T/\text{LEU}$ as estimated from detailed free energy calculations. Though this model might be improved through more explicit consideration of how mutations affect peptide chain entropies, it seems clear that three-body structural information of water may be a powerful indicator of solute-solute and solvation thermodynamics.

In summary, these results highlight qualitatively distinct structural behavior for waters in the vicinity of the extended hydrophobic interface versus near the molecularly-sized hydrophobic leucine side chains. Moreover, the three-body angle distribution for hydration waters around a peptide evolves with added leucines in a way that closely tracks the free energies, moving towards that of the purely hydrophobic isobutane side-chain analog. The angle distribution thus offers a convenient signature of hydrophobic interaction regimes suitable to flexible, heterogeneous, soft-matter systems.

2.4 Conclusions

In summary, this close coupling of simulation to experimental approaches provides a molecular view into dynamic single molecule/surface interactions under applied load, in particular the all-important hydrophobic interaction in the context of peptides, which may

easily extend to arbitrary other surface/molecule interactions. The theoretical calculations are in good agreement with experiments given the precision of the measured data, with no parameters fit between the two. For the interaction of hydrophobic model scaffolds, we find that hydrophobic leucine units have an estimated interaction free energy of about $2.1 k_B T$ from simulation, which compares well to the experimental range 3.0 - $3.4 k_B T$, suggesting that hydrophobic amino acids interact independently and additively with the surface. Our results demonstrate how single molecule measurements with AFM can be directly compared to theoretical data at solid/liquid interfaces for realistic systems, providing a road map to increase precision. Modeling of the pulling process reveals how an individual surface-to-molecule interaction dynamically breaks, and how equilibrium interaction free energies can be estimated from such nonequilibrium experiments in both simulation and experiment. This work will prove useful for designing synergistic experimental and simulation approaches for understanding molecular thermodynamics and fluctuations at solid/liquid interfaces under dynamic load in a wide range of fields, including interactions at bio-interfaces and adhesive interactions in technological systems in a broad context ranging from e.g. cell-surface, cell-cell interactions to adhesive bonding in bio-medical applications.

References

- [1] Ducker WA, Mastropietro D. Forces between extended hydrophobic solids: Is there a long-range hydrophobic force? *Current Opinion in Colloid and Interface Science*. 2016; 22: 51–58. Publisher: Elsevier Ltd.
- [2] Meagher L, Craig S.J. V. Effect of dissolved gas and salt on the hydrophobic force between polypropylene surfaces. *Langmuir*. 1994; 10: 2736–2742. ISBN: 0743-7463.
- [3] Donaldson SH, Røyne A, Kristiansen K, et al. Developing a General Interaction Potential for Hydrophobic and Hydrophilic Interactions. *Langmuir*. 2015; 31(7): 2051–2064.
- [4] Mastropietro DJ, Ducker WA. Forces between hydrophobic solids in concentrated aqueous salt solution. *Physical Review Letters*. 2012; 108(10).
- [5] Stock P, Utzig T, Valtiner M. Direct and quantitative AFM measurements of the concentration and temperature dependence of the hydrophobic force law at nanoscopic contacts. *Journal of Colloid and Interface Science*. 2015; 446: 244–251. Publisher: Elsevier Inc.
- [6] Tabor RF, Wu C, Grieser F, Dagastine RR, Chan DYC. Measurement of the hydrophobic force in a soft matter system. *Journal of Physical Chemistry Letters*. 2013; 4(22): 3872–3877. ISBN: 1948-7185.
- [7] Scheu R, Rankin BM, Chen Y, Jena KC, Ben-Amotz D, Roke S. Charge asymmetry at aqueous hydrophobic interfaces and hydration shells. *Angewandte Chemie - International Edition*. 2014; 53(36): 9560–9563.
- [8] Scheu R, Chen Y, Subinya M, Roke S. Stern layer formation induced by hydrophobic interactions: A molecular level study. *Journal of the American Chemical Society*. 2013; 135(51): 19330–19335.
- [9] Mezger M, Sedlmeier F, Horinek D, Reichert H, Pontoni D, Dosch H. On the origin of the hydrophobic water gap: An X-ray reflectivity and MD simulation study. *Journal of the American Chemical Society*. 2010; 132(19): 6735–6741. ISBN: 1520-5126 (Electronic)\r0002-7863 (Linking).
- [10] Ben-Amotz D. Water-Mediated Hydrophobic Interactions. *Annual Review of Physical Chemistry*. 2016; 67(1): 617–638.
- [11] Hillyer MB, Gibb BC. Molecular Shape and the Hydrophobic Effect. *Annual Review of Physical Chemistry*. 2016; 67(1): 307–329.
- [12] Dill KA, Ozkan SB, Shell MS, Weikl TR. The Protein Folding Problem. *Annual Review of Biophysics*. 2008; 37(1): 289–316. ISBN: 1936-122X (Print)\r1936-122X (Linking).

- [13] Fairman R, Akerfeldt K. Peptides as novel smart materials. *Current Opinion in Structural Biology*. 2005; 15(4): 453–463. ISBN: 0959-440X.
- [14] Grove TZ, Regan L. New materials from proteins and peptides. *Current Opinion in Structural Biology*. 2012; 22(4): 451–456. Publisher: Elsevier Ltd ISBN: 0959-440X.
- [15] Krysiak S, Liese S, Netz RR, Hugel T. Peptide desorption kinetics from single molecule force spectroscopy studies. *J Am Chem Soc*. 2014; 136(2): 688–697. ISBN: 1520-5126 (Electronic)\r0002-7863 (Linking).
- [16] Horinek D, Serr A, Geisler M, et al. Peptide adsorption on a hydrophobic surface results from an interplay of solvation, surface, and intrapeptide forces. *Proceedings of the National Academy of Sciences of the United States of America*. 2008; 105(8): 2842–7. ISBN: 0027-8424.
- [17] Ma CD, Wang C, Acevedo-Velez C, Gellman SH, Abbott NL. Modulation of hydrophobic interactions by proximally immobilized ions. *Nature*. 2015; 517(7534): 347–350. ISBN: 1476-4687 (Electronic)\r0028-0836 (Linking).
- [18] Tych KM, Hughes ML, Bourke J, et al. Optimizing the calculation of energy landscape parameters from single-molecule protein unfolding experiments. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*. 2015; 91(1): 1–9. ISBN: 1539-3755.
- [19] Schwierz N, Horinek D, Liese S, et al. On the relationship between peptide adsorption resistance and surface contact angle: A combined experimental and simulation single-molecule study. *Journal of the American Chemical Society*. 2012; 134(48): 19628–19638. ISBN: 1520-5126 (Electronic)\n0002-7863 (Linking).
- [20] Thyparambil AA, Wei Y, Latour RA. Determination of peptide - Surface adsorption free energy for material surfaces not conducive to SPR or QCM using AFM. *Langmuir*. 2012; 28(13): 5687–5694. ISBN: 0743-7463.
- [21] Ray C, Gu C, Brown JR, Kirkpatrick A, Akhremitchev BB. Anisotropy of Pairwise Interactions between Hexadecanes in Water Measured by AFM Force Spectroscopy. *The Journal of Physical Chemistry C*. 2008; 112(46): 18164–18172. ISBN: 1932-7447.
- [22] Stock P, Monroe J, Utzig T, Smith D, Shell M, Valtiner M. Unraveling Hydrophobic Interactions at the Molecular Scale Using Force Spectroscopy and Molecular Dynamics Simulations. *ACS Nano*. 2017; 11(3).
- [23] Jarzynski C. Nonequilibrium Equality for Free Energy Differences. *Physical Review Letters*. 1997; 78(14): 2690–2693.
- [24] Vericat C, Vela ME, Benitez GA, Gago JAM, Torrelles X, Salvarezza RC. Surface characterization of sulfur and alkanethiol self-assembled monolayers on Au(111). *Journal of Physics: Condensed Matter*. 2006; 18(48): R867–R900.

- [25] Vericat C, Vela ME, Benitez G, Carro P, Salvarezza RC. Self-assembled monolayers of thiols and dithiols on gold: new challenges for a well-known system. *Chemical Society Reviews*. 2010; 39(5): 1805.
- [26] Siu SWI, Pluhackova K, Böckmann RA. Optimization of the OPLS-AA Force Field for Long Hydrocarbons. *Journal of Chemical Theory and Computation*. 2012; 8(4): 1459–1470. ISBN: 1549-9618.
- [27] Ramin L, Jabbarzadeh A. Odd-even effects on the structure, stability, and phase transition of alkanethiol self-assembled monolayers. *Langmuir*. 2011; 27(16): 9748–9759. ISBN: 0743-7463.
- [28] Wang Y, Solano Canchaya JG, Dong W, Alcamí M, Busnengo HF, Martín F. Chain-length and temperature dependence of self-assembled monolayers of alkylthiolates on Au(111) and Ag(111) surfaces. *Journal of Physical Chemistry A*. 2014; 118(23): 4138–4146.
- [29] Rai B, Sathish P, Malhotra CP, Pradip , Ayappa KG. Molecular dynamic simulations of self-assembled alkylthiolate monolayers on an Au(III) surface. *Langmuir*. 2004; 20(8): 3138–3144.
- [30] Vemparala S, Karki BB, Kalia RK, Nakano A, Vashishta P. Large-scale molecular dynamics simulations of alkanethiol self-assembled monolayers. *The Journal of Chemical Physics*. 2004; 121(9): 4323.
- [31] Srivastava P, Chapman WG, Laibinis PE. Molecular dynamics simulation of oxygen transport through n-alkanethiolate self-assembled monolayers on gold and copper. *The Journal of Physical Chemistry B*. 2009; 113(2): 456–464.
- [32] Canchaya JGS, Wang Y, Alcamí M, Martín F, Busnengo HF. Study of the interaction between short alkanethiols from ab initio calculations. *Physical Chemistry Chemical Physics*. 2010; 12(27): 7555.
- [33] Deighan M, Pfaendtner J. Exhaustively sampling peptide adsorption with metadynamics. *Langmuir*. 2013; 29(25): 7999–8009.
- [34] Raut VP, Agashe MA, Stuart SJ, Latour RA. Molecular Dynamics Simulations of Peptide - Surface Interactions. *Langmuir*. 2005; 21(4): 1629–1639.
- [35] Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*. 1983; 79(2): 926. arXiv: 1011.1669v3 ISBN: 0131407333.
- [36] Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14sb: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99sb. *Journal of Chemical Theory and Computation*. 2015; 11(8): 3696–3713.

- [37] Paschek D. Temperature dependence of the hydrophobic hydration and interaction of simple solutes: An examination of five popular water models. *Journal of Chemical Physics*. 2004; 120(14): 6674–6690. arXiv: cond-mat/0312252 ISBN: 0021-9606 (Print)\n0021-9606 (Linking).
- [38] Ashbaugh HS, Collett NJ, Hatch HW, Staton JA. Assessing the thermodynamic signatures of hydrophobic hydration for several common water models. *Journal of Chemical Physics*. 2010; 132(12): 124504. ISBN: 1089-7690 (Electronic)\r0021-9606 (Linking).
- [39] Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics*. 1977; 23(3): 327–341. ISBN: 0021-9991.
- [40] Miyamoto S, Kollman PA. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of Computational Chemistry*. 1992; 13(8): 952–962. ISBN: 0192-8651.
- [41] Berendsen HJC, Postma JPM, Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*. 1984; 81(8): 3684. ISBN: doi:10.1063/1.448118.
- [42] Shirts MR, Chodera JD. Statistically optimal analysis of samples from multiple equilibrium states. *The Journal of chemical physics*. 2008; 129(12): 124105.
- [43] Chodera JD, Shirts MR. A python implementation of the multistate Bennet acceptance ratio (MBAR).
- [44] Raman S, Utzig T, Baimpos T, Ratna Shrestha B, Valtiner M. Deciphering the scaling of single-molecule interactions using Jarzynski’s equality. *Nature communications*. 2014; 5: 5539.
- [45] Kosztin I, Barz B, Janosi L. Calculating potentials of mean force and diffusion coefficients from nonequilibrium processes without Jarzynski’s equality. *Journal of Chemical Physics*. 2006; 124(6). arXiv: cond-mat/0510762 ISBN: 0021-9606.
- [46] Nategholeslam M, Gray CG, Tomberli B. Implementation of the Forward-Reverse Method for Calculating the Potential of Mean Force Using a Dynamic Restraining Protocol. *Journal of Physical Chemistry B*. 2014; 118: 14203–14214.
- [47] Pohorille A, Jarzynski C, Chipot C. Good practices in free-energy calculations. *Journal of Physical Chemistry B*. 2010; 114(32): 10235–10253. ISBN: 10.1021/jp102971x.
- [48] Bennett CH. Efficient Estimation of Free Energy Differences from Monte Carlo Data. *Journal of Computational Physics*. 1976; 22: 245–268.

- [49] Crooks G. Path-ensemble averages in systems driven far from equilibrium. *Physical Review E*. 2000; 61(3): 2361–2366. arXiv: cond-mat/9908420 ISBN: 1095-3787.
- [50] Shirts MR, Bair E, Hooker G, Pande VS. Equilibrium free energies from nonequilibrium measurements using maximum-likelihood methods. *Physical review letters*. 2003; 91(14): 140601. ISBN: 0031-9007.
- [51] Shirts MR, Pande VS. Comparison of efficiency and bias of free energies computed by exponential averaging, the Bennett acceptance ratio, and thermodynamic integration. *The Journal of Chemical Physics*. 2005; 122(14): 144107.
- [52] Hegner M, Wagner P, Semenza G. Ultralarge atomically flat template-stripped Au surfaces for scanning probe microscopy. *Surface Science*. 1993; 291(1-2): 39–46. ISBN: 0039-6028.
- [53] Hinterdorfer P, Dufrene YF. Detection and localization of single molecular recognition events using atomic force microscopy. *Nature Methods*. 2006; 3(5): 347–355. ISBN: 1548-7091 (Print)\r1548-7091 (Linking).
- [54] Collier G, Vellore NA, Yancey JA, Stuart SJ, Latour RA. Comparison between Empirical Protein Force Fields for the Simulation of the Adsorption Behavior of Structured LK Peptides on Functionalized Surfaces. *Biointerphases*. 2012; 7(1): 24–43.
- [55] Mijajlovic M, Penna MJ, Biggs MJ. Free energy of adsorption for a peptide at a liquid/solid interface via nonequilibrium molecular dynamics. *Langmuir*. 2013; 29(9): 2919–2926.
- [56] Shen JW, Wu T, Wang Q, Pan HH. Molecular simulation of protein adsorption and desorption on hydroxyapatite surfaces. *Biomaterials*. 2008; 29(5): 513–532.
- [57] Soliman W, Bhattacharjee S, Kaur K. Adsorption of an antimicrobial peptide on self-assembled monolayers by molecular dynamics simulation. *The journal of physical chemistry. B*. 2010; 114(34): 11292–11302. ISBN: 7804921217.
- [58] Abramyan T, Collier G, Kucukkal TG, et al. Understanding protein-surface interactions at the atomistic level through the synergistic development of experimental and molecular simulation methods. *ACS Symposium Series*. 2012; 1120(0): 197–228. ISBN: 9780841227965.
- [59] Pashley R, Israelachvili J. DLVO and hydration forces between mica surfaces in Mg²⁺, Ca²⁺, Sr²⁺, and Ba²⁺ chloride solutions. *Journal of Colloid and Interface Science*. 1984; 97(2): 446–455. ISBN: 0021-9797.
- [60] Valtiner M, Kristiansen K, Greene GW, Israelachvili JN. Effect of surface roughness and electrostatic surface potentials on forces between dissimilar surfaces in aqueous solution. *Advanced Materials*. 2011; 23(20): 2294–2299. ISBN: 0935-9648.

- [61] Utzig T, Raman S, Valtiner M. Scaling from single molecule to macroscopic adhesion at polymer/metal interfaces. *Langmuir*. 2015; 31(9): 2722–2729.
- [62] Stock P, Utzig T, Valtiner M. Soft matter interactions at the molecular scale: Interaction forces and energies between single hydrophobic model peptides. *Phys. Chem. Chem. Phys.*. 2017; 19: 4216–4221. Publisher: Royal Society of Chemistry.
- [63] Moreno Ostertag L, Utzig T, Klinger C, Valtiner M. Tether-Length Dependence of Bias in Equilibrium Free-Energy Estimates for Surface-to-Molecule Unbinding Experiments. *Langmuir*. 2018; 34(3): 766–772.
- [64] Hummer G, Szabo A. Free energy reconstruction from nonequilibrium single-molecule pulling experiments. *Proceedings of the National Academy of Sciences*. 2001; 98(7): 3658–3661. ISBN: 0027-8424.
- [65] Hummer G, Szabo A. Free energy surfaces from single-molecule force spectroscopy. *Accounts of Chemical Research*. 2005; 38(7): 504–513. ISBN: 0001-4842.
- [66] Israelachvili JN. *Intermolecular and Surface Forces*. Burlington, MA: Elsevier3rd ed. 2011.
- [67] Tucker EE, Lane EH, Christian D. S. Vapor pressure studies of hydrophobic interactions. Formation of benzene-benzene and cyclohexane-cyclohexanol dimers in dilute aqueous solution. *J. Sol. Chem.*. 1981; 10(1): 1–20. ISBN: 0100000150.
- [68] Li ITS, Walker GC. Single Polymer Studies of Hydrophobic Hydration. *Accounts of Chemical Research*. 2012; 45(11): 2011–2021. ISBN: 0001-4842.
- [69] Gore J, Ritort F, Bustamante C. Bias and error in estimates of equilibrium free-energy differences from nonequilibrium measurements. *Proceedings of the National Academy of Sciences of the United States of America*. 2003; 100(22): 12564–12569. ISBN: 0027-8424 (Print).
- [70] Lu N, Kofke DA. Accuracy of free-energy perturbation calculations in molecular simulation. II. Heuristics. *Journal of Chemical Physics*. 2001; 115(15): 6866–6875. ISBN: 0021-9606.
- [71] Lu N, Woolf TB. Understanding and Improving Free Energy Calculations in Molecular Simulations: Error Analysis and Reduction Methods. In *Free Energy Calculations* (Pohorille A, Chipot C, eds.) 199–247 Springer 2007.
- [72] Kim S, Kim YW, Talkner P, Yi J. Comparison of free-energy estimators and their dependence on dissipated work. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*. 2012; 86(4): 1–11. arXiv: 1211.4393v1.
- [73] Jarzynski C. Rare events and the convergence of exponentially averaged work values. *Physical Review E*. 2006; 73(4): 046105.

- [74] Chandler D. Interfaces and the driving force of hydrophobic assembly. *Nature*. 2005; 437(7059): 640–7.
- [75] Acharya H, Vembanur S, Jamadagni SN, Garde S. Mapping hydrophobicity at the nanoscale: Applications to heterogeneous surfaces and proteins. *Faraday discussions*. 2010; 146: 353–365. ISBN: 1359-6640\|r1364-5498.
- [76] Remsing RC, Weeks JD. Hydrophobicity Scaling of Aqueous Interfaces by an Electrostatic Mapping. *The journal of physical chemistry. B*. 2014. arXiv: 1502.05587v1.
- [77] Chaimovich A, Shell MS. Tetrahedrality and structural order for hydrophobic interactions in a coarse-grained water model. *Physical Review E*. 2014; 89(2): 022140. ISBN: 1550-2376 (Electronic) 1539-3755 (Linking).
- [78] Bernal JD, Fowler RH. A Theory of Water and Ionic Solution, with Particular Reference to Hydrogen and Hydroxyl Ions. *The Journal of Chemical Physics*. 1933; 1(8): 515–548.
- [79] Chau PL, Hardwick AJ. A new order parameter for tetrahedral configurations. *Molecular Physics*. 1998; 93(3): 511–518. ISBN: 0026-8976.
- [80] Errington JR, Debenedetti PG. Relationship between structural order and the anomalies of liquid water. *Nature*. 2001; 409(6818): 318–321. ISBN: 0028-0836.
- [81] Clark GN, Cappa CD, Smith JD, Saykally RJ, Head-Gordon T. The structure of ambient water. *Molecular Physics*. 2010; 108(11): 1415–1433.
- [82] Ferguson AL, Debenedetti PG, Panagiotopoulos AZ. Solubility and Molecular Conformations of n -Alkane Chains in Water. *The Journal of Physical Chemistry B*. 2009; 113(18): 6405–6414. ISBN: 1520-6106.

Chapter 3

Characterization of three-body angle distributions in bulk and solvation shells of model solutes

3.1 Introduction

Though it is common and essential to life, water is exceptionally unique from a molecular point of view.^{1,2} While simple in its chemical structure, water exhibits a variety of anomalies that may be attributed to its underlying structure and modes of intermolecular interaction.² In particular, hydrogen bonding promotes water associating with four neighbors, resulting in a preference to form tetrahedral structures that are pronounced in both its solid and liquid phases. This tetrahedral structure has historically been the basis for interpreting many of water's unique behaviors 3–6,^{3–6} including anomalous temperature dependence of density, diffusivity, entropy, and gas solubilities.^{2,7} Central to these anomalies is a competition between energetically favorable hydrogen bonds that restrict relative orientations between water molecules and entropic driving forces towards

less tetrahedral geometries that mimic simple-fluid-like behavior in which closest packing is approached.²

Here we examine the range of orientational structure behaviors that rigid water models exhibit in the liquid state and connect these to the thermodynamics of bulk water and aqueous solvation. Specifically, we examine three-body angle distributions of water, which involve the angle θ between two oxygens within approximately the first hydration shell of a central water oxygen:

$$P(\theta) = \frac{V\Omega}{Z} \int \left[\int_{r_{12} < r_c} \int_{r_{13} < r_c} e^{-\beta U} \delta\left(\frac{\mathbf{r}_{2,O}}{|\mathbf{r}_{2,O}|} \cdot \frac{\mathbf{r}_{3,O}}{|\mathbf{r}_{3,O}|} - \cos \theta\right) d\mathbf{r}_2 d\mathbf{r}_3 \right] d\mathbf{r}_4 \dots \mathbf{r}_N \quad (3.1)$$

In the above, translation and rotation of the central water molecule (numbered 1) has been integrated out into the terms V and Ω , and Z represents the proper normalization factor, related to the canonical configurational partition function. Each integral is performed over the translational and rotational degrees of freedom of the corresponding water molecule; the outer integral runs over all waters 4 and higher, while the inner ones correspond to the two nearest neighbors (numbered 2 and 3) with oxygen positions $r_{2,O}$ and $r_{3,O}$. Within this inner integral, the delta function selects angles consistent with the one of interest, and the limits of integration restrict waters 2 and 3 to reside within the neighbor cutoff distance $r_c=0.34$ nm. Though the orientational structure of water, including three-body angles, has been studied in the past, a detailed examination of its behavior as a function of temperature and density in the liquid state is lacking. Further, the specific role that three-body angles play in determining water's unique thermodynamic and solvation behaviors remains unclear, which the present work seeks to clarify.

A number of studies have examined the orientational structure of water, including three-body angle distributions, across a diverse set of computational models. Here, and most commonly in the literature, water is modeled as a rigid molecule with fixed point

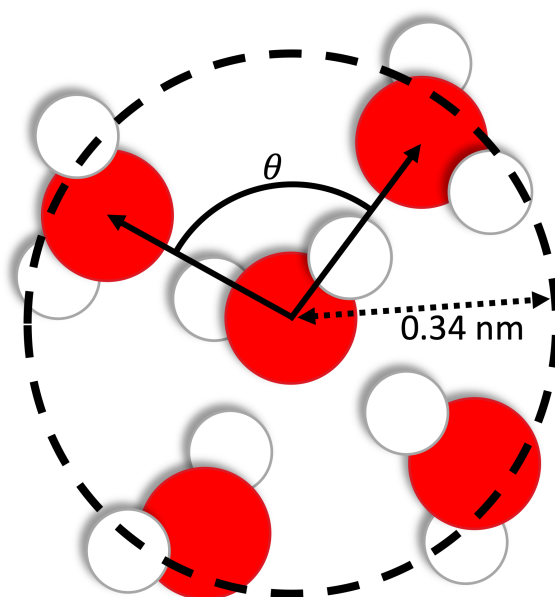


Figure 3.1: The definition of the three-body angle is shown schematically as the angle between two vectors connecting a central water oxygen atom to oxygens on two of its neighbors within 0.34 nm. For computation of the three-body angle distribution, all angles formed by unique pairs of neighbors within the cutoff are considered, regardless of the number of such angles.

charges. Head-Gordon and Stillinger⁸ developed a series of water models based on varying perturbations of an orientational, three-body term enforcing tetrahedrality, studying quenched structures in these models. Subsequent studies by Lynden-Bell and coworkers^{9–12} examined how perturbations to the geometry,^{10–12} hydrogen bond strength,^{9,10} and explicit representation of anisotropy¹⁰ influence water structural, thermodynamic, and solvation properties, including the existence of anomalous behaviors. In general, such perturbations drastically change both water translational (i.e. oxygen-oxygen radial distribution function) and orientational (three-body angle distribution) interparticle correlations, weakening the presence of or even removing anomalous behaviors depending on the magnitude of the perturbation. Interestingly, explicit representation of water’s geometry is not necessary to capture its unique behaviors. Purely repulsive isotropic models with two length-scales of interaction mimic most of water’s anomalous behaviors.¹³ When more realistic attractive interactions are included, isotropic water models may exhibit anomalous behaviors depending on the specific form of the potential and parametrization.^{14,15}

The success of isotropic models might seem to suggest a limited role of purely orientational degrees of freedom in determining water’s behavior other than to create a distinctive second length-scale of interaction. On the other hand, Molinero and Moore¹⁶ developed a single-site water model that includes independent terms to enforce translational and orientational ordering in water, demonstrating that simple Lennard-Jones interactions in combination with an explicit three-body-angle potential function accurately capture many of water’s properties,^{16,17} including structural and thermodynamic anomalies.¹⁸ Other studies involving this water model demonstrate that anomalous properties, including those at supercooled temperatures, may be turned on and off by tuning the strength of the three-body angle term.¹⁹ In other work, the relative role of orientational, three-body interactions in determining water-like behavior was assessed by comparing water to other tetrahedral monatomic liquids (i.e. Si) as well as molten oxides.^{18,20} These and isotropic,

dual-length-scale models both highlight the role of hydrogen bonding and orientational correlations in determining water’s behavior. However, only models including orientational terms are able to simultaneously capture the oxygen-oxygen radial distribution function (RDF) as well as the expected preference for tetrahedral orientations in water.¹⁰ Indeed, even water models developed with hydrogen bonds explicitly constrained to assume unphysical geometries are able to reproduce RDFs for all water atomic pairs,²¹ but exhibit significant deviations in three-body orientational degrees of freedom from what is expected for a tetrahedral geometry.^{20,21}

Relationships between water’s structure and its unique properties have also been assessed through structural contributions to water’s excess entropy. Scala, et al.²² demonstrated that the configurational entropy, as assessed by the number of minima in the potential energy landscape, can predict anomalous behavior of water. Rigid water models and other tetrahedral liquids have subsequently been shown to exhibit regions of phase space where the excess entropy increases with density.^{18,23–25} Two-body contributions to the excess entropy based on pairwise RDFs reflect such anomalous behavior,^{18,25} but contributions from three-body degrees of freedom are largely unstudied. An approximation to such terms has been computed via the entropy of the probability distribution of a popular metric of tetrahedrality based on the full three-body angle distribution.²⁶ In conjunction with a pairwise two-body entropy, this “tetrahedral entropy” can be used to map out regions of phase space that exhibit anomalous water behavior.²⁵ Similar mappings based on orientational and translational contributions to the entropy from two-body degrees of freedom also identify regimes of anomalous water behavior.²⁷ In addition to the excess entropy, many metrics of liquid water structure have been examined in both ambient and supercooled regions,^{7,28,29} generally with a focus on capturing transitions between two structurally distinct populations, which are thought to result in a liquid-liquid phase transition at supercooled temperatures.⁷ Notably, modelling of

configurational entropy, which is closely tied to the excess entropy, gives rise to predictions of a liquid-liquid phase transition²² and critical point.³⁰ Additional (though not exhaustive) examples of metrics that identify two distinct states of water include those based on spatial correlation of tetrahedrality,²⁶ relative structure of first and second neighbor shells,¹⁹ and even Markov models based on states involving clusters of six waters.³¹

The seminal work of Errington and Debenedetti³² suggested two of the most well-studied metrics for assessing water structure and identifying anomalous behaviors, the so-called translational and tetrahedral order parameters. These authors define a translational order parameter that measures the radial ordering of a fluid as

$$t = \frac{\int_0^{r_c} |g(r) - 1| dr}{r_c} \quad (3.2)$$

where $g(r)$ is the oxygen-oxygen RDF and r_c is a chosen cutoff for integration. The tetrahedral order parameter, originally introduced with a distinct normalization by Chau and Hardwick,³³ measures specifically the deviation of a fluid from perfect tetrahedral versus ideal-gas like structuring, and is defined in terms of the three-body angle θ ,

$$q = 1 - \frac{3}{8} \sum_i^4 \sum_{i \neq j}^4 (\cos \theta_{i,j} + \frac{1}{3})^2 \quad (3.3)$$

Here, the sums run over the four nearest neighbors to each water oxygen. When reporting q as a metric, averaging is typically performed over all water oxygens and water configurations of a defined population of water molecules.

The tetrahedral order parameter has been at the heart of a number of studies examining the overall tetrahedrality of water within the hydration shells of solutes. Agarwal, et al.³⁴ found that waters in the hydration shells of peptides show increasing tetrahedrality at more energetically favorable locations. More specifically, the electrostatic energy

is more favorable where tetrahedral structure is higher, implying enhanced hydrogen bonding. A number of other studies examined q within the hydration shells of small hydrophobes.³⁵⁻³⁹ To summarize these briefly, waters within such hydration shells exhibit reduced tetrahedrality very near to the solute (closer than the first maximum in the solute-water oxygen RDF), and enhanced tetrahedrality in the outer edge of the first solvation shell, as first noted by Galamba.³⁵ As hydrophobe size is increased, q for all waters within the first hydration shell rapidly decreases.⁴⁰ This is despite the fact that water at extended hydrophobic and air-water interfaces is expected to assume a unique, not necessarily disordered, structure compared to the bulk.⁴¹⁻⁴⁵ While enhanced structure in the first solvation shell of small hydrophobes is not clearly observed for all rigid water models,³⁸ it has also been observed for the mW model of water,³⁶ as well as for a polarizable model.³⁵ Additionally, experimental measurements of Raman spectra also support the presence of energetically favorable hydrogen bonds (as compared to bulk water) within small-solute hydration shells,^{39,46} which is generally thought to be correlated with more tetrahedral arrangements.

Unfortunately, there is some ambiguity, and associated debate^{38,47,48} surrounding the enhancement of the tetrahedral order parameter near small solutes, as well as its exact behavior as solute size increases. When waters are very close to an interface, q must decrease due to purely geometric constraints – all four nearest-neighbors may no longer reside within the first hydration shell of the central water. Even if all neighbors are within the first hydration shell on average, as has been argued maintains the validity of this metric,²⁹ there is certainly a higher likelihood of having a fourth nearest neighbor outside this first shell than for purely bulk waters. For this reason, it has been suggested that solute heavy atoms should also be considered as neighbors for the purposes of calculating q via Equation 3,^{25,35,40} though the precise effect of this adjustment is unclear.

The studies briefly reviewed above demonstrate that, while useful for mapping the

boundaries of structural anomalies in bulk water, the tetrahedral order parameter is difficult to define at interfaces or near solutes of arbitrary shape and chemical properties. However, understanding the structure of water in the hydration shells of solutes is fundamentally important to understanding solvation processes and solute-solute interactions, especially with regard to biomolecules.⁴⁹ In contrast to q , the distribution of three-body water angles remains well-defined within the hydration shells of solutes, with perturbations from bulk water structure easily assessed.^{50,51} A variant of the tetrahedral order parameter may be calculated from $P(\theta)$, the significant difference being that any number of neighbors may contribute to the angles calculated in Equation 1 as long as they are within the cutoff, while the definition of q requires a restriction to only four nearest neighbors.⁵⁰ In this study, we hypothesize that an analysis of the full distribution, rather than metrics that represent averages, are essential to understanding hydration water structure and properties near arbitrary solutes, and we present results showing this behavior in comparison to conventional order parameters like q .

In a previous study,⁵¹ we noted that perturbations to three-body angle distributions in the solvation shells of peptides can be used to develop thermodynamic models for the free energy of association for peptides with hydrophobic interfaces. This was remarkable in that trends in free energies could be accurately predicted merely by considering the structural response of solvent to the presence of a solute. Here we explore this relationship further for a set of model solutes, specifically idealized spherical colloids of different size and attractive strength. We demonstrate that perturbations in the solvation shell water structures of solutes fall within the variations seen in bulk water as a function of density and temperature. We draw fundamental connections between restructuring of three-body angles within solute hydration shells and thermodynamics of solvation, partially explaining the success of previous thermodynamic models based on angle distributions.⁵¹ More specifically, perturbations to three-body angle distributions track the contribution to the

free energy beyond the direct interactions (potential energy) of the solute with water. We show that these contributions are associated with water restructuring around the solute and in fact are intimately related to the relative entropy associated with the solvation process.

3.2 Methods

3.2.1 Simulation of bulk water

Simulations of bulk water use the TIP4P/2005 water model.⁵² As a control for comparison, we also use a simple fluid model with only Lennard-Jones (LJ) interactions whose energy and length scales were selected using relative entropy coarse-graining to best match SPC/E water at 1.00 g/cm³ and 300 K.⁵³ All simulations are performed with a 0.002 ps timestep using the 2016.1 version of GROMACS⁵⁴ with the “group” cutoff scheme and a neighbor list update every 10 steps. All simulations are run for a total of 5 ns. For both models, a nonbonded cutoff of 1.4 nm is applied, while the TIP4P/2005 case also employs PME treatment⁵⁵ and SETTLE constraints.⁵⁶ A Langevin thermostat ensures sampling at the desired temperature. Simulations are performed at temperatures ranging from 250 K to 350 K. Densities range from 0.80 to 1.40 g/cm³ for TIP4P/2005 water and 0.80 to 1.30 g/cm³ for the LJ water model. From calculations of the pressure (excluding dispersion corrections) at each density for the TIP4P/2005 model and fits to 4th order Legendre polynomials as a function of density, the stability limit is identified as the point at which the pressure derivative with respect to density at constant temperature becomes zero. For nearly all temperatures simulated, the 0.80 and 0.85 g/cm³ densities are below the stability limit and hence the data from these states is excluded in all analyses other than that to establish the stability bound.

3.2.2 Simulation of colloids in water

Colloids are simulated as Hamaker-averaged spheres of LJ particles⁵⁷⁻⁵⁹ with $\sigma=0.3730$ nm and $\epsilon=1.2301$ kJ/mol. We modify the strength of colloid-water attractions by changing the density of LJ particles averaged over the spherical colloidal volume from $19.3/\text{nm}^3$ (low-attractions) to $110.4/\text{nm}^3$ (high-attractions). We also develop purely repulsive colloids by separating the colloidal particle potential according to the Weeks-Chandler-Andersen (WCA) scheme⁶⁰ and only keeping the repulsive part. In the text, the resulting WCA colloids refer to the application of this procedure to only the lowest-density (least-attractive) colloids, though we also develop WCA colloids based on the high-density colloids for the purposes of computing free energies, as described below.

Colloidal simulations are performed as follows. Each colloid is solvated in a box with 5000 waters. Due to the large size of some of the colloids, we use a 2.4 nm nonbonded cutoff. Temperature is kept constant with a Langevin thermostat at 300 K, and pressure is held constant using a Parinello-Rahman barostat.⁶¹ For TIP4P/2005, we fix the pressure at 1 bar, while for the LJ water model, we fix it at 4560 bar, which is the pressure at which this coarse-grained model achieves a bulk water density of approximately 1.00 g/cm^3 at 300 K.⁵³ In calculating the pressure, no dispersion corrections are included as the large size of the colloids renders the environment of the waters effectively heterogeneous. The large cutoffs ensure that long-range attractions important to modeling the behavior of water at interfaces^{62,63} are sufficiently represented. To accurately compute changes in three-body angle distributions for bulk water versus that near colloids, we also perform simulations of water boxes with identical settings but without any colloids present. We perform simulations of WCA, low-density, and high-density colloids for 20 ns each and save simulation frames every 1 ps to ensure high-accuracy three-body angle distributions in the first hydration shell. For the larger colloids, the distributions converge to high

accuracy much faster than this (on the order of 5 ns), but the smaller colloids take longer due to fewer hydration shell waters. To enhance efficiency of calculations, smaller simulation boxes could be used for the smaller colloids, but we maintain the same number of solvated waters for all colloids to facilitate free energy calculations, as described below.

3.2.3 Computation of three-body angle distributions

For simulations of bulk water, three-body angle distributions are computed using a fixed cutoff of 0.34 nm to define neighbors. Within the cutoff, all pairs of neighbors are considered for determination of the angle between vectors connecting the oxygens of the neighbors to the central water oxygen. For simulations with colloids, the water oxygen-oxygen neighbor-cutoff is chosen based on the first minimum in the RDF of bulk water at the same pressure and temperature and under the same simulation conditions, with cutoffs of 0.442 nm, 0.332 nm, 0.332 nm, and 0.358 nm used for LJ water, TIP4P/2005, TIP4P-Ew, and TIP3P, respectively. Hydration shells are defined based on minima in the colloid-water oxygen RDFs, with three-body angle distributions in each shell including all angles determined around central waters in the shell with neighbors anywhere.

3.2.4 Free energy calculations

We calculate solvation free energies for colloids via two steps. In the first step, we simulate purely repulsive WCA colloids with effective radii^{64,65} varying between zero and 1.298 nm for 1 ns each. We then use Bennett’s Acceptance Ratio (BAR) method⁶⁶ to compute the free energy difference for each change in radius. Note that we perform these calculations for WCA colloids based on both weakly and strongly-attractive colloids to broaden the number of reference stages and reduce error. In the second step, the free energy of switching a colloid from a purely repulsive WCA state to a fully-interacting

state is computed via a linear switching path. Simulations of 1 ns each are performed at states along this path in increments of 0.1 from 0.0 (WCA potential) to 1.0 (fully-interacting colloid). We use the Multistate Bennet Acceptance Ratio (MBAR)⁶⁷ estimator, implemented in the pymbar package,⁶⁸ to compute the free energy of this process. By combining the results of the first and second steps, the total free energy of solvation can be computed. Full, tabulated values of the solvation free energies for TIP4P/2005 water are provided in Appendix C.

3.3 Results and Discussion

3.3.1 Structural information in bulk water

We first assess the basic form of the three-body angle distributions in bulk liquid water, its variation with state conditions, and its relationship to other structural properties. Figure 3.2 shows variations in $P(\theta)$ over a range of temperatures and densities for TIP4P/2005 water, using a fixed neighbor cutoff of 0.34 nm that represents the approximate location of the first neighbor shell for the lowest densities considered (Appendix C Figs. C.1-C.2). At all conditions, TIP4P/2005 water clearly exhibits a characteristic preference for tetrahedral three-body arrangements, with $P(\theta)$ peaking near 109.5° . In a tetrahedral crystalline arrangement, one expects only a single, sharp peak at 109.5° , whereas Figure 3.2 shows that the corresponding liquid phase peak is as expected much broader. Moreover, an additional population around 50° emerges, which corresponds to the appearance of a 5th neighbor within the first hydration shell that sits interstitial to the predominantly tetrahedral arrangement (see Fig. C.5 in Appendix C).

For comparison, Figure 3.2 also shows the three-body distribution for a simple liquid, namely a Lennard-Jones (LJ) system using parameters (epsilon and sigma) that were

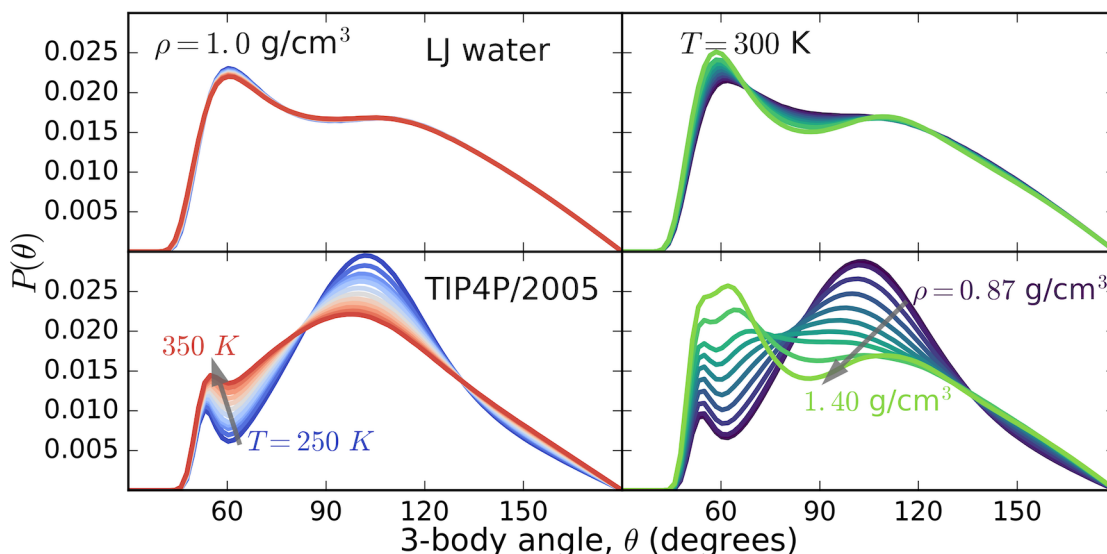


Figure 3.2: Three-body angle distributions of bulk LJ water (top) and TIP4P/2005 water (bottom) as a function of temperature at fixed density (left) and as a function of density at fixed temperature (right).

optimized in an earlier study to match SPC/E water’s length and energy scales at ambient temperature and density 53. In contrast to the tetrahedral preference of TIP4P/2005, the LJ model is peaked around 63.4° , characteristic of a simple fluid with icosahedral closest packing. In an icosahedral crystal, for example, sharp peaks would be observed around 63.4° , 116.6° , and 180.0° , with the first two equal in population and the latter five-times less likely. This explains the broad secondary peak near 116.6° , though it is not of equal probability as the peak near 60° because due to increased broadening at larger angles; the 60° peak is sharper because it includes arrangements of three molecules that are all mutual nearest neighbors, in contrast to larger angles.

As temperature is increased at fixed density, Figure 3.2 shows that TIP4P/2005 water shows a tradeoff in populations: the probability at low angles (near 63.4°) grows at the expense of the tetrahedral configurations near 109.5° . In contrast, the LJ fluid shows very little variation with temperature, and in fact shows a small decrease of the

peak near 63.4° . Changes in density lead to more pronounced shifts in the three body distributions, and the trend is similar for both the TIP4P/2005 and LJ systems, reflecting an increase in population near 63.4° with density. At the highest densities, TIP4P/2005 develops a distinct peak near 63.4° , separate from the one at 50° , indicating that the model has lost substantial tetrahedrality and instead develops a packing-dominated structure more typical of a simple fluid. At both high temperatures and densities, the interstitial waters are squeezed out of the first shell completely (or, alternatively, more waters are forced into the first shell, removing interstitial space) and the peak around 50° merges with the icosahedral, 63.4° peak (Appendix C Figs. C.3-C.4). This represents a significant transition from tetrahedral and water-like at low temperatures and densities to simple-fluid-like at the opposite conditions, at which $P(\theta)$ begins to resemble the LJ distribution.

A natural question is the extent to which the distributions depend on the distance cutoff used to determine neighbors used for computing three-body angles. Because we use a fixed cutoff to define neighbors for calculating angles, the number of angles for each water will track with RDF variations. RDFs at all temperatures and densities for TIP4P/2005 and LJ water are shown in the Appendix C, Figures C.1-C.2. For TIP4P/2005, it is clear that the chosen cutoff of 0.34 nm is close to the first minimum in the RDFs at low densities, but much lower than the first minimum at high densities, around 0.4 nm. By fixing the cutoff at a lower value, we are actually excluding some waters in the first hydration shell at higher densities, but still always capture the high-probability peak most representative of the first hydration shell. We find that using a larger cutoff of 0.42 nm for all temperatures and densities does not change the qualitative behavior of $P(\theta)$ at high densities and temperatures, only acting to broaden all peaks of the distribution as the more disordered outer edge of the first hydration shell becomes included. This is supported by Figure C.5 in Appendix C where the qualitative behavior of $P(\theta)$ varies

little with the number of waters included in the cut-off at high densities of TIP4P/2005 water, or at any density for LJ water. At low densities, however, such a large cutoff includes significant amounts of second-shell neighbors and results in qualitatively distinct behavior.

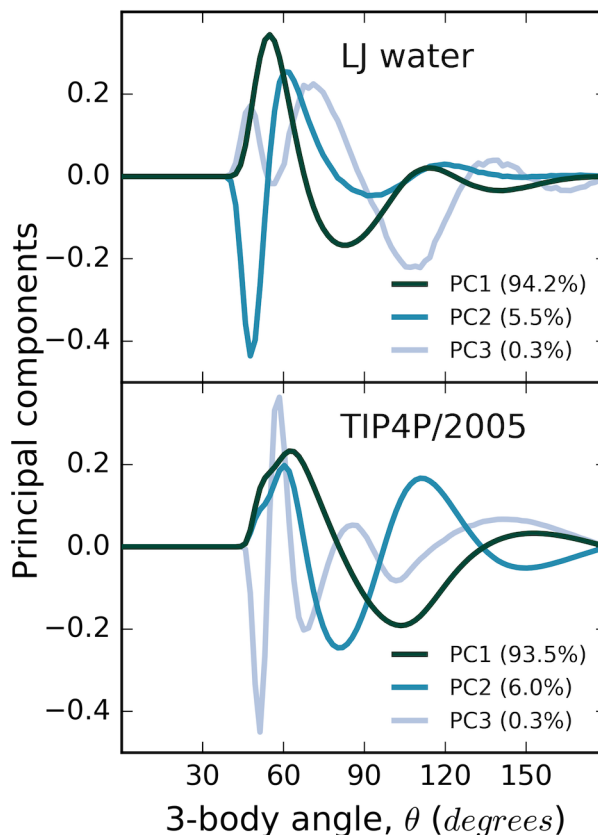


Figure 3.3: Principal components of changes in the three-body angle distributions of LJ and TIP4P/2005 water with varying temperature and density, with the average of all such distributions used as the reference for performing PCA. The percentage of total variance explained by each principal component is shown in the legend, with PC1 and PC2 together capturing over 99%.

A striking feature in Figure 3.2 is the simple manner by which the water distributions shift in response to temperature and density variations, for which intermediate conditions seem to interpolate smoothly between the extremes of highly tetrahedral and simple fluid-like. To quantify the response, we extract the principal components (PCs) of

these variations relative to a reference probability distribution, taken to be the average distribution uniformly-weighted over all conditions. As Figure 3.3 shows, the top two PCs account for greater than 99% of the variance across all temperatures and densities. Moreover, we find that the precise shape of the PCs is largely independent of the reference chosen. Thus, the principal components show rather remarkably that the distributions vary largely according to just two primary degrees of freedom, representing the contributions of principal components 1 and 2 (PC1 and PC2). Their form also reinforces the discussion above, with TIP4P/2005 mainly exhibiting opposite perturbations to $P(\theta)$ around 109.5° and 63.4° , and LJ water exhibiting in general perturbations of smaller magnitude and predominantly at angles less than 109.5° . In turn, any given distribution $P(\theta)$ can be projected onto this small set of PCs (after subtraction of the reference distribution), offering a simple way to interpret contributions from distinct structural propensities. The projection onto PC1 alone accounts for nearly 94% of the variation and is particularly insightful as it indicates, as compared to the average over all temperatures and densities, the relative balance of tetrahedral versus icosahedral-dominant configurations: a negative projection indicates water-like structure while a positive one a simple-fluid-like structure (though note that the sign convention is arbitrary).

Figure 3.4 shows projections onto PC1 and PC2 within the temperature and density space considered in this study. As density is increased at all temperatures, a transition from tetrahedral-dominant to icosahedral-dominant configurations is signaled by a change of sign in projections onto PC1. This crossover happens at lower densities as temperature is increased and the strength of hydrogen bonds diminishes. A similar transition occurs for PC2, except that in this case a minimum is reached where projections onto PC1 pass through zero. This correlation between the first two principal components is required by construction since the other PCs make negligible contributions: the PCs form an orthonormal set based on a Euclidean norm, meaning that the sum of the squares of the

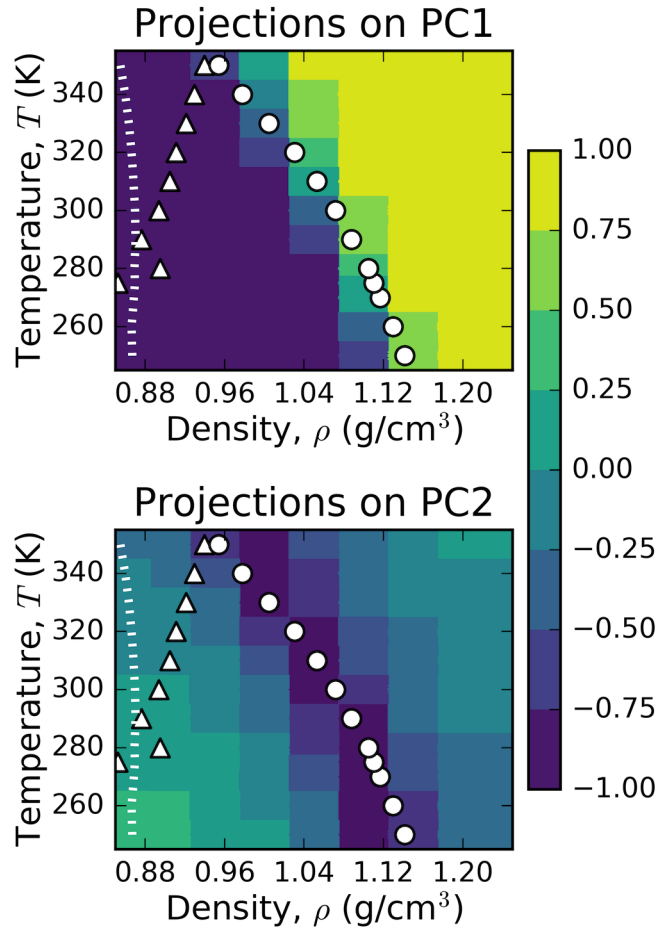


Figure 3.4: Projections of three-body angle distribution perturbations onto PC1 and PC2 (shown in Fig. 3.3) are shown as a heat-map in temperature-density space. Negative projections are observed in the regime of structural anomalies, while positive projections in the region of expected simple-fluid behavior, as demarcated by the white contour line at zero. White circles indicate the locus of translational order parameter minima (high-density structural anomaly bound) while white triangles indicate isothermal tetrahedral order parameter maxima (low-density anomaly bound). The dotted white curve on the left indicates the mechanical stability boundary.

projections must be equal to the magnitude of the vector projected onto this set, which here has been normalized to one.

It is tempting (though incorrect, as shown below) to use the projections mapped out in Figure 3.4, in particular those onto PC1, to define structural anomalies or other water behaviors. Indeed, the high-density structural anomaly boundary, as determined by the pioneering order parameter analysis of Errington and Debenedetti³² (see Appendix C Fig. C.6), seems to coincide with a transition from negative PC1 projections (tetrahedral behavior) to positive projections (simple fluid-like behavior). No evidence, however, is observed for the low-density boundary. This is surprising as the low-density boundary captures a crossover in orientational behavior with density.³² One explanation is that this boundary is intrinsically difficult to define due to its close proximity to the mechanical stability boundary,⁶⁹ where the conventional q order parameter decreases upon development of voids in the liquid in the same manner as is expected for the structural anomaly boundary. However, the anomaly boundary remains identifiable by minima in isotherms of the tetrahedral order parameter (Appendix C Fig. C.6) which occur before the stability boundary.

While analysis of principal components may not identify a low-density structural anomaly boundary, are there other indications in $P(\theta)$ that do? One possibility is the tetrahedral order parameter, which can be computed as an average using the distribution $P(\theta)$. However, this approach fails to identify the lower boundary even though explicit calculations of q do; the reason is that tetrahedral angles contribute strongly to $P(\theta)$ even for configurations with fewer than four neighbors within the first hydration shell, while q is always computed from the first four neighbors regardless of whether or not they fit within a distance cutoff (Appendix C Figs. C.5,C.7-C.8). Other previously studied structural metrics⁵⁰ seem to identify an upper bound, but are also unsuccessful in clearly identifying a lower bound (Appendix C Fig. C.9).

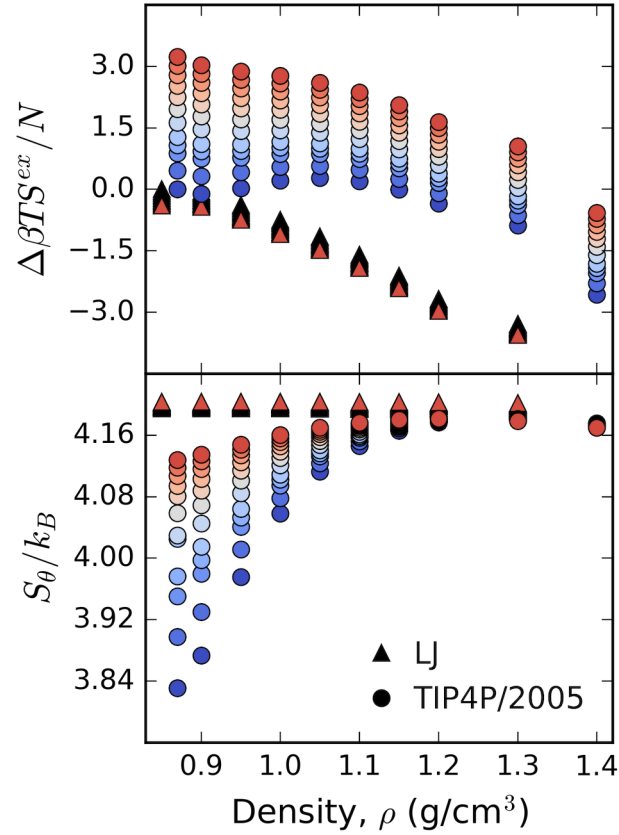


Figure 3.5: (a) Isotherms of excess entropy changes (with the lowest temperature and density as a reference) and (b) the entropy of three-body angle distributions. TIP4P/2005 (circles) show anomalous entropy increases over a range of densities while LJ (triangles) show monotonically decreasing behavior. While excess entropy defines a lower bound to the anomalous region via a minimum, no such minimum is observed for the three-body angle entropy.

Another metric of interest is the entropy associated with the three-body angle distribution, $S_\theta = -k_B \sum P(\theta) \ln P(\theta)$. Excess entropy anomaly boundaries have been shown to be approximated by low-order entropy expansions, in particular 2-body entropies based on atomic radial distributions.^{23,25} Figure 3.5 indicates that three-body angle entropy isotherms behave similarly to the excess entropy, increasing with density anomalously before reaching maxima (at densities slightly higher than for the excess entropy). However, unlike the excess entropy, no minima, and hence no lower density anomaly, emerge in the three-body angle entropy isotherms.

It is clear, therefore, that $P(\theta)$ captures many important features of water's unique behavior, but not all. Moreover, it is also important that the absolute values of the projections shown in Figure 3.4 depend on the choice of the reference distribution, as these projections represent the magnitude of perturbations in a specific direction defined by a principal component. While we find that the principal components themselves do not depend significantly on the choice of reference, the values of the projections certainly do. Therefore, the boundary implied by the projections in state space agrees with the structural anomaly boundary insofar as our choice of reference distribution is appropriate. Here our reference distribution is the average $P(\theta)$ over the range of liquid state points studied, but there may be other useful references, for example, references with equal populations of tetrahedral and simple-fluid-like water structures. Regardless of the reference choice, however, the projections still concretely demonstrate that a boundary exists and provide a reasonable shape for it in state space.

3.3.2 Structural information in solute hydration shells

While water's three body structure is perturbed by changes to bulk temperature and pressure, it is natural to ask whether interfaces and solutes have comparable effects –

and if such inhomogeneous environments perturb $P(\theta)$ in a similarly simple way as state condition, involving just a few principal directions. To test this idea, we consider model spherical solutes (Hamaker colloids^{57–59}) of varying size and attractive strength solvated in TIP4P/2005,⁵² TIP4P-Ew,⁷⁰ TIP3P,⁷¹ and LJ⁵³ water models. Figure 3.6 shows for TIP4P/2005 water variations in $P(\theta)$ compared to bulk for waters in the first colloid hydration shell, as well as colloid-water RDFs; distributions for other water models are similar (with the exception of the LJ model discussed below) and are shown in Figure C.11 of Appendix C. For low attractive strengths, especially for purely repulsive Weeks-Chandler-Andersen (WCA) colloids, water begins to withdraw from the colloid interface as colloid size is increased, while for the attractive colloids, water wets the colloid interface more strongly with increasing colloid radius. This behavior has been well-characterized previously and connected to density fluctuations of hydrating waters.^{59,72,73}

Changes in $P(\theta)$ are distinct across both colloid size and chemistry (i.e., strength of colloid-water interactions). As observed previously,⁵¹ $P(\theta)$ is enhanced relative to bulk around 109.5° for small solutes. For large colloids, including the limit of infinite radius (a flat 9-3 wall), the region around 109.5° is reduced in population relative to bulk water while a population at 90° grows. With increasing attraction, the tetrahedral reduction becomes more pronounced, while only for strongly-attractive colloids is the icosahedral region near 63.4° enhanced relative to bulk. Mittal and Hummer⁷³ noted decreases in density and increases in density fluctuations within probe volumes at solute interfaces. For small hydrophobes, water density and its fluctuations approached bulk values, whereas in contrast here we observe clear perturbations to orientational structure for small solutes. For larger hydrophobes, our results link reduction of the tetrahedral character of three-body angle distributions to previously-observed enhancements of density fluctuations.^{72,73} Perturbations to $P(\theta)$ in the second hydration shell (Appendix C Fig. C.10) are distinct from those in the first but are generally much weaker, the only exception

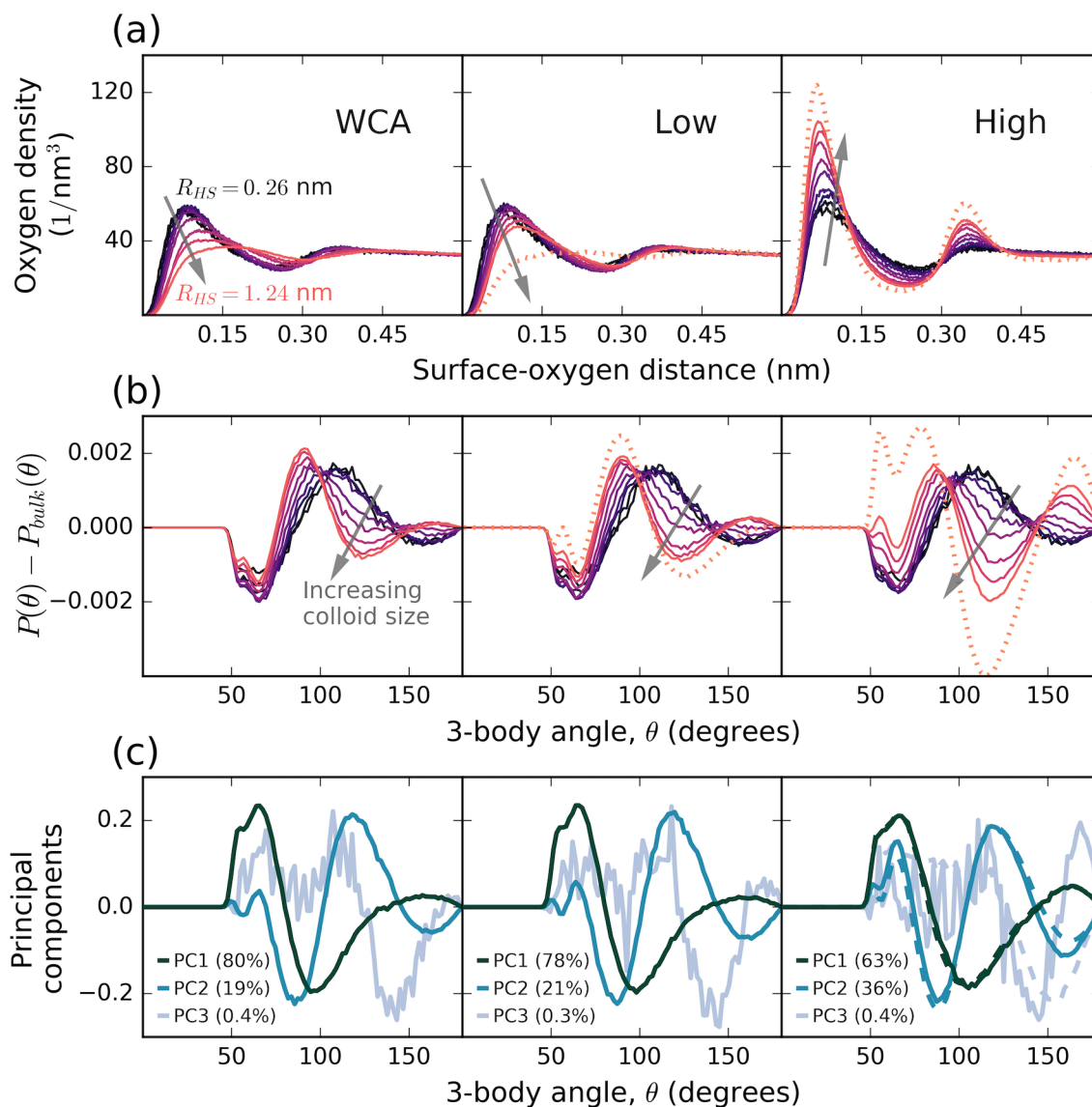


Figure 3.6: (a) Colloid-water oxygen radial distribution functions (shifted so the colloid surface, where the RDF goes to zero, is at zero). (b) Perturbations to bulk three-body angle distributions for waters in the first hydration shell of colloids, with (c) principal components of those perturbations. Variations from dark purple to orange indicate small colloids; dashed orange lines indicate colloids of infinite radius (flat 9-3 walls). The dotted lines in the bottom panel of (c) are the resulting PCs if all colloid three-body angle distributions are considered together instead of separately. The percentage of variance for these first, second, and third PCs of the combined set of colloids are 66.3%, 32.0%, and 1.1%.

being the highly-attractive infinite-radius colloid (i.e., flat wall). For this reason, we only consider the first hydration shell in what follows.

	WCA	Low	High	All colloids
PC1, PC1_{bulk}	0.981	0.984	0.962	0.978
PC1, PC2 _{bulk}	0.082	0.058	-0.187	-0.124
PC1, PC3 _{bulk}	-0.054	-0.055	-0.049	-0.042
PC2, PC1 _{bulk}	-0.131	-0.106	0.195	0.139
PC2, PC2_{bulk}	0.768	0.780	0.728	0.742
PC2, PC3 _{bulk}	-0.024	-0.048	-0.103	-0.058
PC3, PC1 _{bulk}	-0.080	-0.079	-0.086	-0.052
PC3, PC2 _{bulk}	0.252	0.230	0.236	0.135
PC3, PC3_{bulk}	-0.287	-0.245	-0.114	-0.238

Table 3.1: Projections of principal components of perturbations to $P(\theta)$ in the first hydration shell of colloids onto principal components of perturbations across all temperatures and densities of bulk water.

Remarkably, the behavior of the colloids' perturbations to the three-body water structure are highly similar across the different colloid types, and moreover, strongly resemble the kinds of variations seen in bulk water distributions across temperatures and densities (Figure 3.3). While the effect of the colloids' perturbations is an order of magnitude smaller, the striking similarities with the bulk results suggest that the three-body orientational structure of water may only vary in a limited number of ways, even in inhomogeneous scenarios. In this sense, the perturbations induced by the colloids can be viewed as a subset of those in Figure 3.3, which instead correspond to changes in temperature and density. This is tested in Table 3.1, where we calculate projections of all colloid principal components onto bulk water principal components. Projections between first principal components are very close to one in all cases, suggesting high similarity, while second principal components show slightly more differences compared to

bulk. Though we have not pursued it here, it would also be interesting to assess whether perturbations to $P(\theta)$ in the solvation shells of ions also fall within those for bulk water across state space.

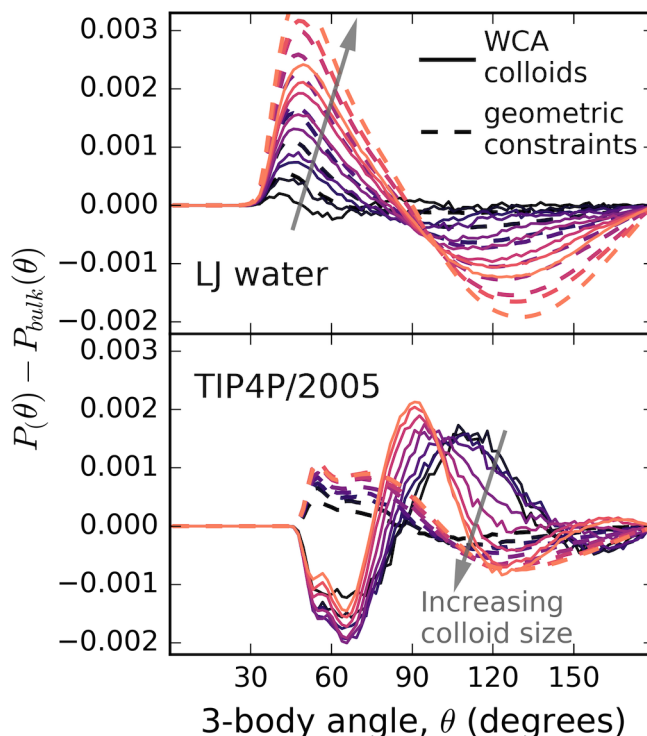


Figure 3.7: Perturbations from bulk water distributions for waters in the first hydration shells of WCA colloids are shown with solid lines. Dashed lines give controls that indicate variations due merely to hydration shell geometry (i.e., the volumes over which the distributions are computed), without the explicit presence of a colloid.

Figure 3.7 shows that perturbations from bulk water for three-body angle distributions of colloids solvated in TIP4P/2005 are distinctly different from those of the LJ fluid, suggesting a fundamentally different response to the presence of a solute. Whereas TIP4P/2005 transitions between enhancement near 109.5° for small solutes and enhancement of the peak near 63.4° for large solutes, the LJ fluid exhibits monotonic behavior as colloid size increases, with the distribution for small angles continuously growing at the expense of larger ones. Moreover, the perturbation in the LJ fluid is exactly what is

expected simply by computing the distribution in a geometry corresponding to a putative hydration shell of some radius. Figure 3.7 shows $P(\theta)$ distributions computed within a radial shell of the same shape as the first hydration layer around WCA colloids, but computed from a simulation of bulk water without an explicit solute present. These distributions with only geometric constraints, no physical solutes, exhibit behavior almost identical to that of the LJ fluid with an actual solute, indicating that the presence of the colloid does not fundamentally perturb the structure of the LJ fluid, but instead only reflects the geometry of the region used to compute $P(\theta)$. In contrast, TIP4P/2005 significantly shifts its structure in response to a solute, exhibiting fundamentally different behavior: the three body distributions show marked differences between the explicit solute case and the geometric control volume.

3.3.3 Thermodynamic information for solvation

The specific variations in $P(\theta)$ for TIP4P/2005 also show signatures of the ~ 1 nm length-scale transition in hydrophobic solvation.⁷⁴ It is well-established that solvation on small and large length-scales exhibit fundamentally different thermodynamic behaviors,⁷⁴ with the structural origin of such behaviors hotly debated.^{75,76} Here we interpret solvation in terms of shifts in the three-body angle distribution, providing a novel structural-theoretical perspective for interpreting distinct solvation contributions solely from changes in solvent configurational degrees of freedom. In Figure 3.8, we project changes in $P(\theta)$ within the first colloid solvation shell onto the overall principal components shown in Figure 3.6C. We observe a crossover in behavior as the colloid size is increased, with projections onto PC1 sharply becoming less negative after ~ 0.5 nm and projections onto PC2 switching from positive to negative around the same point. While the crossover point is unchanged with colloid attractive strength, the rate of increase of the PC1 projection

becomes larger with increasing colloid-water attractions. Importantly, these projections alone seem to be sensitive to both attractive strength and colloid size, exhibiting a fundamental crossover in behavior in the latter that begins to saturate around 1 nm.

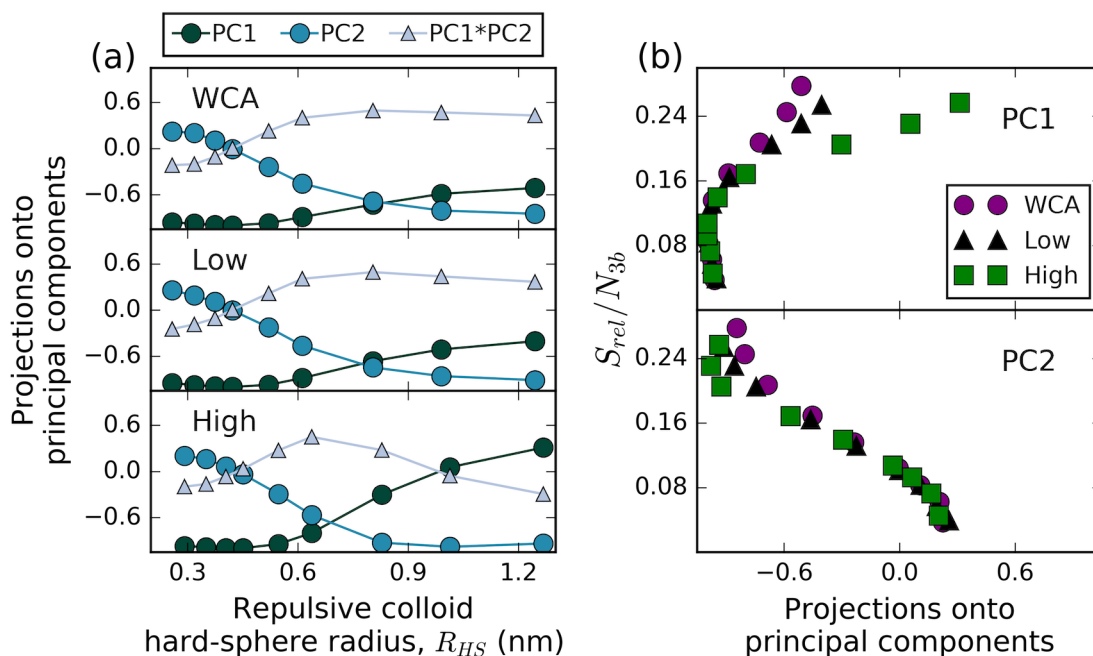


Figure 3.8: (a) Projections of colloid first solvation shell perturbations to $P(\theta)$ onto principal components for colloids of varying size, indicated by the hard-sphere radius, and attractive strength (increasing from the top to bottom panel). The product of projections onto components 1 and 2 is also shown as triangles. (b) The relative entropy of solvation, as defined here and in Appendix B, as a function of the projections onto the first two principal components.

Our past work⁵¹ and inhomogeneous solvation theory⁷⁷ suggest that examination of water degrees of freedom, in particular three-body angles, may also provide thermodynamic information on the solute solvation process. To characterize the relationship between water's structural response and this process, we explore models for solvation thermodynamics based on perturbations to $P(\theta)$. This approach is supported in part by our previous work,⁵¹ in which we showed that $P(\theta)$ appeared to capture association free energies of short peptides of varying hydrophobic content with extended hydrophobic

surfaces. There, the fraction of the three-body angle distribution for peptide hydration waters that corresponded to an idealized reference distribution near hydrophobic leucine sidechain analogues correlated linearly with the free energy for moving the peptide from solution to the hydrophobic surface.

Here, we ask whether a similar parsing of $P(\theta)$ for the colloids correlates with their solvation free energies, i.e., the free energy for transferring them from an ideal gas phase into solution. This quantity corresponds to the difference in free energy ΔG for a solvated colloid (state 2) and that in which the colloid is in the ideal gas phase while the water exists separately in bulk (state 1):

$$\beta\Delta G = \ln \left[\frac{\int e^{-\beta pV} \left(\int e^{-\beta U_{ww}(\mathbf{r}_w^N)} d\mathbf{r}_w^N d\mathbf{r}_c \right) dV}{\int e^{-\beta pV} \left(\int e^{-\beta U_{ww}(\mathbf{r}_w^N) - \beta U_{wc}(\mathbf{r}_w^N, \mathbf{r}_c)} d\mathbf{r}_w^N d\mathbf{r}_c \right) dV} \right] \quad (3.4)$$

Here, p gives the pressure and V the system volume, r_w^N the coordinates of the waters, r_c the position of the colloid, and U_{ww} and U_{wc} the total potential energies for the water-water and water-colloid interactions, respectively. Equation 4 can be further manipulated via the usual Zwanzig-type perturbation:

$$\beta\Delta G = \ln \langle e^{\beta U_{wc}} \rangle_2 \quad (3.5)$$

Equation 5 shows that the solvation free energy is determined by the water-colloid interactions in the colloid-solvated state, which incorporates (through the configurational probabilities) the effect of solvation water structure and its response to the colloid. This result can be made more instructive by separating out the direct interaction contribution

of $\langle U_{wc} \rangle_2$, giving

$$\begin{aligned}\beta\Delta G &= \beta\langle U_{wc} \rangle_2 + \ln\langle e^{\beta(U_{wc}-\langle U_{wc} \rangle_2)} \rangle_2 \\ &= \beta\langle U_{wc} \rangle_2 + S_{rel,1\rightarrow 2}\end{aligned}\tag{3.6}$$

The indirect contribution in the second term of the first line clearly emerges from fluctuations in the solvated state of U_{wc} around the mean $\langle U_{wc} \rangle_2$ 78–80.^{78–80} This term has been previously characterized in the context of the solvation of long-chain alkanes,⁸¹ as well as for the process of turning on attractive interactions for model spherical solutes.⁸² The second line shows that the indirect term is also given by the relative entropy:

$$S_{rel,1\rightarrow 2} = \int \wp_2(V) \left[\int \wp_2(\mathbf{r}_w^N, \mathbf{r}_c; V) \ln \frac{\wp_2(\mathbf{r}_w^N, \mathbf{r}_c, V)}{\wp_1(\mathbf{r}_w^N, \mathbf{r}_c, V)} d\mathbf{r}_w^N d\mathbf{r}_c \right] dV \tag{3.7}$$

where $\wp(\mathbf{r}_w^N, \mathbf{r}_c; V)$ gives the probability for a particular configuration $\mathbf{r}_w^N, \mathbf{r}_c$ at a given volume V in the ensemble of interest, $\wp(\mathbf{r}_w^N, \mathbf{r}_c, V)$ gives the total probability of both a given configuration and volume, and $\wp_2(V)$ gives the volume distribution in the solvated state. The relative entropy integrates over the entire configurational space to measure the overlap of the two probability distributions in states 1 and 2. Relative entropies have been used to develop inverse and coarse-graining algorithms;^{83,84} in the present context, $S_{rel,1\rightarrow 2}$ measures how accurately the pre-solvation state 1 captures the configurational distribution of waters in the solvated state.

Note that the relative entropy still integrates over the colloid degrees of freedom; however, for a rigid, spherically-symmetric colloid in a homogeneous system, these terms do not contribute, such that the relative entropy only captures changes to solvent degrees

of freedom. Equation 7 thus simplifies to

$$S_{rel,1\rightarrow 2} = \int \wp_2(V) \left[\int \wp_2(\mathbf{r}_w^N; V) \ln \frac{\wp_2(\mathbf{r}_w^N, V)}{\wp_1(\mathbf{r}_w^N, V)} d\mathbf{r}_w^N \right] dV \quad (3.8)$$

This result shows that $S_{rel,1\rightarrow 2}$ captures changes to hydration water structure upon solvation, where two distinct effects arise. The first involves the creation of a solvent void volume, at constant pressure, to accommodate the colloid; the corresponding contribution to the relative entropy scales as expected like $\beta p V_c$ where V_c is the colloid volume (see Appendix B). The second effect involves changes to local water structure, relative to the bulk, in regions near to the colloid. In cases where a colloid perturbs the hydration water structure substantially, this effect should be large and positive, while it has a minimum of zero in the case where no perturbation occurs. A more detailed derivation of Equation 3.8 and these results are given in Appendix B.

Because $S_{rel,1\rightarrow 2}$ naturally addresses the manner by which colloids influence hydration water structure, we hypothesize that changes to the three-body angle distribution will correlate strongly with its value and hence the indirect contribution to the solvation free energy in Equation 6. Indeed, Figure 3.8 indicates the near-collapse of curves for all solutes when showing projections onto PC1 and PC2 against the relative entropy of solvation, whereas other thermodynamic quantities do not exhibit such clean behavior (Appendix C Fig. C.12). Of course, water degrees of freedom beyond three-body angles will contribute to $S_{rel,1\rightarrow 2}$, but here we simply examine how well $P(\theta)$ perturbations track local water restructuring and thus the relative entropy. We note that the $\beta p V_c$ contribution to $S_{rel,1\rightarrow 2}$ can be ignored as it is negligible for even the largest colloids studied (see Table C.1 in Appendix C).

We thus propose a simple model in which the relative entropy for the solvation of a given colloid is approximated using the projections of its first hydration shell three-body

angle distribution onto the first two principal components of variation. The precise form of our model is

$$\frac{S_{rel,model}}{N_{3b}} = s_1 p_1 + s_2 p_2 + s_{12} p_1 p_2 \quad (3.9)$$

with

$$p_1 = proj\left(\delta P(\theta), PC1(\theta)\right) \quad (3.10)$$

$$p_2 = proj\left(\delta P(\theta), PC2(\theta)\right) \quad (3.11)$$

where N_{3b} is the average number of three-body angles found in the first hydration shell, $proj(\delta P(\theta), PCi)$ is the projection of the perturbation of $P(\theta)$ from bulk onto the i^{th} PC, and s_i are fit parameters representing the free energy contributions associated with changes to projections along the given PCs. Here, we express the relative entropy as linear in the projections of the angle distributions with the addition of a non-linear cross term. The form of Equation 11 might be motivated by the possibility that hydration waters experience two distinct types of environments, corresponding to variations along PC1 and PC2. But generally, we expect more complex behavior that requires a “mixing” term between the two, since a single water’s three body angles (up to six for four-fold coordination) may show among them deviations corresponding to both PCs. Given the model form, we then fit the parameters s_1 , s_2 , and s_{12} by least squares regression to minimize the detailed computed versus approximate values of $S_{rel,1\rightarrow 2}$.

Figure 3.9 shows the performance of the fitted model, both with and without the cross-term. By including the cross-term, the model accurately captures the solvation relative entropy for all colloid sizes. On the other hand, the purely additive model (no cross-term) fails for small colloids, in particular for those that are highly attractive, where the free energy for solvation shifts from positive to negative with colloid size. The

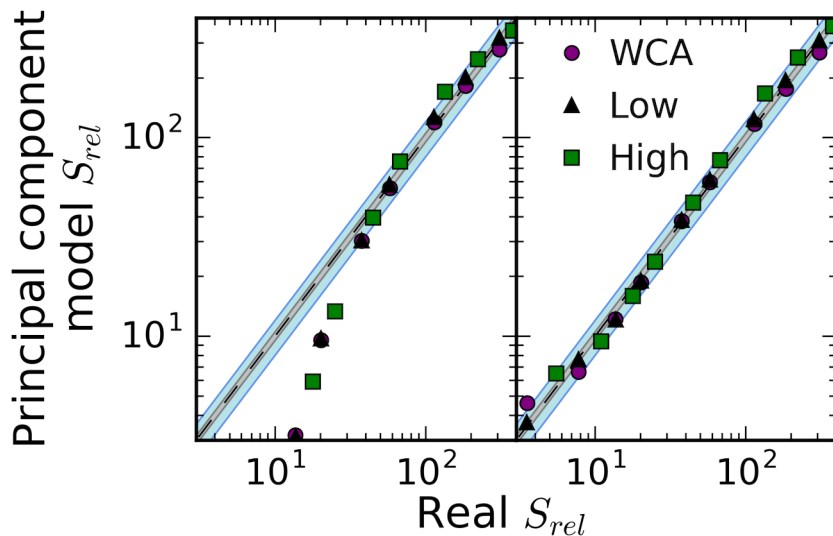


Figure 3.9: Model predictions of Equation 9 with (left) and without (right) the cross-term for the relative entropies for solvating all colloids. The dotted line indicates a perfect match of the model to the true S_{rel} values while light gray and light blue shaded regions indicate 5% and 20% error. Without cross-terms, the fit near small colloid sizes is poor. The fit for large colloid sizes is much improved when colloids of different levels of attractions are fit separately (Appendix C Fig. C.13).

improvement due to the cross term seems to stem primarily from its ability to switch sign depending on the relative signs of projections onto PC1 and PC2. Figure 3.8 indicates that projections onto PC2 change sign with increasing colloid size, while those onto PC1 do not. Thus, the first two terms in Equation 9 represent the amount of the relative entropy, in an empirical sense, that is correlated with independent changes of three-body angles along PC1 (tetrahedral to icosahedral changes) or PC2 (changes in the number of interstitial waters), whereas the cross-term represents joint changes along both of these modes for three-body angle restructuring. Although PC1 and PC2 are orthogonal, we do not expect perturbations induced by the introduction of a colloid to correspond to independent shifts along principal components. By this reasoning, it is not surprising that the cross-term is necessary.

The results in Figure 3.9 also indicate that not all colloids types are fit with the

same accuracy, with the highest errors for the highly attractive species. However, if each type is fit separately (i.e., distinct s parameters for WCA, low, and high versions), both small-colloid and large-colloid behavior is fit much more accurately (Appendix C Fig. C.13); this is likely possible due to the very similar behavior of $S_{rel,1\rightarrow 2}$ for all small colloids, but distinct behavior at larger sizes (Fig. 3.8). This observation indicates intrinsic differences in the way water structure is perturbed in the shells of the largest colloids, beyond what can be explained by changes in three-body angles. As an example, it is clear from Figure 3.6 that water density profiles as a function of radial distance from the colloid are nearly identical for all types of small colloids. For large colloids, however, repulsive and weakly-attractive colloids induce a de-wetting of water from the colloid surface while very attractive colloids induce greater ordering with larger colloid size. Figure C.14 in Appendix C demonstrates that the contribution to $S_{rel,1\rightarrow 2}$ from three-body angles alone is relatively small and does not linearly correlate with the total relative entropy; moreover, colloids with distinct attractions show deviations in the relationship between the two quantities at larger colloid sizes. This highlights the fact that the model in Equation 9 is innately a perturbative model based on a single structural metric – it breaks down once perturbations become too large or the chosen degree of freedom fails to qualitatively track overall shifts in water structure. In this case, perturbations to water structure are not fully captured by three-body angles and information from changes to other degrees of freedom, such as water density profiles or two-body correlations, should be included to generally describe relative entropies between states.

3.3.4 Conclusions

This work has sought to understand the nature of the three-body angle distribution in water, $P(\theta)$, as a measure of its structural and thermodynamic response to different

conditions. This distribution is of great interest because it characterizes the full spectrum of water's orientational structure fluctuations, in contrast to many conventional orientational order parameters that are averages, and because it is readily applied (without modification) to study inhomogeneous environments with solutes or interfaces, since it does not require waters to have a fixed number of nearest neighbors.

We first characterized the behavior $P(\theta)$ as a function of temperature and density in bulk water. At low temperatures and densities, the three-body distribution naturally shows that tetrahedral arrangements of waters dominate, while moving away from these conditions, the distribution transitions to simple-fluid-like. These shifts are captured clearly by the principal components computed from changes to $P(\theta)$. Remarkably, we found that only two principle components explain the shifts across all the distributions, and projections onto these principal components then provide a metric of the relative propensity for tetrahedral or icosahedral structure. Moreover, our results suggest that the principal components can identify some of the boundaries in state space that delineate regions of water's anomalous structural behavior. While here we focused exclusively on $P(\theta)$, similar analysis might be fruitfully applied to other water degrees of freedom, or collections of degrees of freedom including θ , to map water structural shifts more fully. In particular, we found that the significance of structural shifts is more readily obtained through examination of the full distribution of the metrics (and their perturbation), in contrast to averages or moments, such as the conventional tetrahedral order parameter.

We also studied the three-body distribution in hydration water near idealized spherical solutes of varied sizes and strength of solute-water attractions. In the hydration shells of these colloids, we observed enhanced tetrahedrality for small colloids while the corresponding peak in $P(\theta)$ shrinks for larger ones, regardless of the level of attractive strength. For strong attractions, the distribution becomes progressively more like a simple fluid, but to a much lesser extent than high density-induced changes in bulk water. Remarkably, we find

that principal components of perturbations from bulk water of $P(\theta)$ in colloid hydration shells are very similar to those for perturbations of bulk water across temperature and density space. This implies that this degree of freedom may only be perturbed in a limited number of ways and that such responses may be mapped to similar changes induced by state conditions alone. Characterization of three-body angle distributions in the hydration shells of more realistic solutes, in particular ions or flexible polymers, may reveal new responses of $P(\theta)$ beyond what we have presented here. It will be very interesting to see if the principal components derived from bulk water can also track perturbations from a more diverse set of solutes.

Using a very simple model, we find that projections onto principal components of perturbations to $P(\theta)$ in colloid hydration shells also appears to well-predict the indirect, solvent restructuring contribution to solute solvation free energies (specifically, the solvation relative entropy). This represents an exciting explicit connection between solvation thermodynamics and changes in microscopic solvent structure. The proposed solvation model starts to break down for very large colloids when attempting to simultaneously capture the behavior for colloids of all attractive strengths. This represents a limitation of the single degree of freedom θ to comprehensively reflect water restructuring in hydration shells. Future studies exploring shifts in additional water structural metrics and correlations, and their relative contributions to solvation thermodynamics through the relative entropy, will likely prove helpful in enhancing this type of structure-based model of solvation. Similar analyses for bulk water may also be useful in clearly connecting shifts in its structure to its thermodynamic properties. Finally, it is well known that solvation of many solutes in water, particularly hydrophobes, exhibit distinctive temperature dependence. It will be interesting to investigate the role that three-body angles, or other degrees of freedom, play in determining this behavior through their contribution to the relative entropy of solvation. The present type of analysis may identify specific degrees of

freedom that contribute to this temperature dependence, which in turn may suggest key features to include when developing coarse-grained models and solvation theories.

References

- [1] Ball P. Water as an Active Constituent in Cell Biology. *Chemical Reviews*. 2008; 108(1): 74–108. arXiv: NIHMS150003 ISBN: 0009-2665.
- [2] Brini E, Fennell CJ, Fernandez-Serra M, Hribar-Lee B, Lukšič M, Dill KA. How Water’s Properties Are Encoded in Its Molecular Structure and Energies. *Chemical Reviews*. 2017; 117(19): 12385–12414.
- [3] Bernal JD, Fowler RH. A Theory of Water and Ionic Solution, with Particular Reference to Hydrogen and Hydroxyl Ions. *The Journal of Chemical Physics*. 1933; 1(8): 515–548.
- [4] Frank HS, Evans MW. Free Volume and Entropy in Condensed Systems III. Entropy in Binary Liquid Mixtures; Partial Molal Entropy in Dilute Solutions; Structure and Thermodynamics in Aqueous Electrolytes. *The Journal of Chemical Physics*. 1945; 13(11): 507–532.
- [5] Kauzmann W. Some Factors in the Interpretation of Protein Denaturation¹the preparation of this article has been assisted by a grant from the National Science Foundation. In *Advances in Protein Chemistry* (Anfinsen CB, Anson ML, Bailey K, Edsall JT, eds.); Vol. 14 1–63 Academic Press 1959.
- [6] Stillinger FH. Structure in aqueous solutions of nonpolar solutes from the standpoint of scaled-particle theory. *Journal of Solution Chemistry*. 1973; 2: 141–158.
- [7] Palmer JC, Poole PH, Sciortino F, Debenedetti PG. Advances in Computational Studies of the Liquid–Liquid Transition in Water and Water-Like Models. *Chemical Reviews*. 2018; 118(18): 9129–9151.
- [8] Head-Gordon T, Stillinger FH. An orientational perturbation theory for pure liquid water. *The Journal of Chemical Physics*. 1993; 98(4): 3313–3327.
- [9] Lynden-Bell RM, Debenedetti PG. Computational Investigation of Order, Structure, and Dynamics in Modified Water Models[†]. *The Journal of Physical Chemistry B*. 2005; 109(14): 6527–6534.
- [10] Lynden-bell RM, Head-gordon T. Solvation in modified water models: towards understanding hydrophobic effects. *Molecular Physics*. 2006; 104(22-24): 3593–3605.
- [11] Chatterjee S, Debenedetti PG, Stillinger FH, Lynden-Bell RM. A computational investigation of thermodynamics, structure, dynamics and solvation behavior in modified water models. *The Journal of Chemical Physics*. 2008; 128(12): 124511.

- [12] Lynden-Bell RM, Giovambattista N, Debenedetti PG, Head-Gordon T, Rossky PJ. Hydrogen bond strength and network structure effects on hydration of non-polar molecules. *Physical chemistry chemical physics : PCCP*. 2011; 13(7): 2748–2757. ISBN: 1463-9076.
- [13] Yan Z, Buldyrev SV, Giovambattista N, Stanley HE. Structural Order for One-Scale and Two-Scale Potentials. *Physical Review Letters*. 2005; 95(13): 130604.
- [14] Chaimovich A, Shell MS. Anomalous waterlike behavior in spherically-symmetric water models optimized with the relative entropy. *Physical Chemistry Chemical Physics*. 2009; 11(12): 1901.
- [15] Hammer MU, Anderson TH, Chaimovich A, Shell MS, Israelachvili J. The search for the hydrophobic force law. *Faraday Discussions*. 2010; 146: 299–308.
- [16] Molinero V, Moore EB. Water modeled as an intermediate element between carbon and silicon. *Journal of Physical Chemistry B*. 2009; 113(13): 4008–4016. arXiv: 0809.2811 ISBN: 1520-6106.
- [17] Cisneros GA, Wikfeldt KT, Ojamäe L, et al. Modeling Molecular Interactions in Water: From Pairwise to Many-Body Potential Energy Functions. *Chemical Reviews*. 2016; 116(13): 7501–7528.
- [18] Shadrack Jabes B, Nayar D, Dhabal D, Molinero V, Chakravarty C. Water and other tetrahedral liquids: order, anomalies and solvation. *Journal of Physics: Condensed Matter*. 2012; 24(28): 284116.
- [19] Russo J, Akahane K, Tanaka H. Water-like anomalies as a function of tetrahedrality. *Proceedings of the National Academy of Sciences*. 2018; 115(15): E3333–E3341.
- [20] Dhabal D, Singh M, Wikfeldt KT, Chakravarty C. Triplet correlation functions in liquid water. *The Journal of Chemical Physics*. 2014; 141(17): 174504.
- [21] Wikfeldt KT, Leetmaa M, Ljungberg MP, Nilsson A, Pettersson LGM. On the Range of Water Structure Models Compatible with X-ray and Neutron Diffraction Data. *The Journal of Physical Chemistry B*. 2009; 113(18): 6246–6255.
- [22] Scala A, Starr FW, La Nave E, Sciortino F, Stanley HE. Configurational entropy and diffusivity of supercooled water. *Nature*. 2000; 406: 166–169. arXiv: cond-mat/9908301 ISBN: 1476-4687 (Electronic)\n0028-0836 (Linking).
- [23] Sharma R, Agarwal M, Chakravarty C. Estimating the entropy of liquids from atom–atom radial distribution functions: silica, beryllium fluoride and water. *Molecular Physics*. 2008; 106(15): 1925–1938.

- [24] Agarwal M, Singh M, Sharma R, Parvez Alam M, Chakravarty C. Relationship between Structure, Entropy, and Diffusivity in Water and Water-Like Liquids. *The Journal of Physical Chemistry B*. 2010; 114(20): 6995–7001.
- [25] Agarwal M, Alam MP, Chakravarty C. Thermodynamic, diffusional, and structural anomalies in rigid-body water models. *Journal of Physical Chemistry B*. 2011; 115(21): 6935–6945. ISBN: 1520-5207 (Electronic)\r1520-5207 (Linking).
- [26] Kumar P, Buldyrev SV, Stanley HE. A tetrahedral entropy for water. *Proceedings of the National Academy of Sciences*. 2009; 106(52): 22130–22134. ISBN: 0027-8424.
- [27] Esposito R, Saija F, Saitta AM, Giaquinta PV. Entropy-based measure of structural order in water. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*. 2006; 73(4): 1–4. arXiv: cond-mat/0603764 ISBN: 1539-3755\r1550-2376.
- [28] Gallo P, Amann-Winkel K, Angell CA, et al. Water: A Tale of Two Liquids. *Chemical Reviews*. 2016; 116(13): 7463–7500.
- [29] Duboué-Dijon E, Laage D. Characterization of the Local Structure in Liquid Water by Various Order Parameters. *The Journal of Physical Chemistry B*. 2015; 119(26): 8406–8418.
- [30] Handle PH, Sciortino F. Potential energy landscape of TIP4p/2005 water. *The Journal of Chemical Physics*. 2018; 148(13): 134505.
- [31] Hamm P. Markov state model of the two-state behaviour of water. *The Journal of Chemical Physics*. 2016; 145(13): 134501.
- [32] Errington JR, Debenedetti PG. Relationship between structural order and the anomalies of liquid water. *Nature*. 2001; 409(6818): 318–321. ISBN: 0028-0836.
- [33] Chau PL, Hardwick AJ. A new order parameter for tetrahedral configurations. *Molecular Physics*. 1998; 93(3): 511–518. ISBN: 0026-8976.
- [34] Agarwal M, Kushwaha HR, Chakravarty C. Local order, energy, and mobility of water molecules in the hydration shell of small peptides. *Journal of Physical Chemistry B*. 2010; 114(1): 651–659. ISBN: 1520-5207 (Electronic)\r1520-5207 (Linking).
- [35] Galamba N. Water’s Structure around Hydrophobic Solutes and the Iceberg Model. *The Journal of Physical Chemistry B*. 2013; 117(7): 2153–2159.
- [36] Song B, Molinero V. Thermodynamic and structural signatures of water-driven methane-methane attraction in coarse-grained mW water. *The Journal of Chemical Physics*. 2013; 139(5): 054511.

- [37] Ashbaugh HS, Barnett JW, Saltzman A, Langrehr ME, Houser H. Communication: Stiffening of dilute alcohol and alkane mixtures with water. *The Journal of Chemical Physics*. 2016; 145(20): 201102.
- [38] Hajari T, Bandyopadhyay S. Water structure around hydrophobic amino acid side chain analogs using different water models. *The Journal of Chemical Physics*. 2017; 146(22): 225104.
- [39] Wu X, Lu W, Streacker LM, Ashbaugh HS, Ben-Amotz D. Temperature-Dependent Hydrophobic Crossover Length Scale and Water Tetrahedral Order. *The Journal of Physical Chemistry Letters*. 2018; 9(5): 1012–1017.
- [40] Hande VR, Chakrabarty S. Structural Order of Water Molecules around Hydrophobic Solutes: Length-Scale Dependence and Solute-Solvent Coupling. *Journal of Physical Chemistry B*. 2015; 119(34): 11346–11357. ISBN: 1520-6106.
- [41] Lee C, McCammon JA, Rossky PJ. The structure of liquid water at an extended hydrophobic surface. *The Journal of Chemical Physics*. 1984; 80(9): 4448–4455.
- [42] Stirnemann G, Rossky PJ, Hynes JT, Laage D. Water reorientation, hydrogen-bond dynamics and 2d-IR spectroscopy next to an extended hydrophobic surface. *Faraday Discussions*. 2010; 146: 263. ISBN: 1359-6640\|1364-5498.
- [43] Layfield JP, Troya D. Molecular Simulations of the Structure and Dynamics of Water Confined between Alkanethiol Self-Assembled Monolayer Plates. *The Journal of Physical Chemistry B*. 2011; 115(16): 4662–4670.
- [44] Hölzl C, Horinek D. Pressure increases the ice-like order of water at hydrophobic interfaces. *Physical Chemistry Chemical Physics*. 2018; 20(33): 21257–21261.
- [45] Willard AP, Chandler D. Instantaneous liquid interfaces. *The Journal of Physical Chemistry B*. 2010; 114(5): 1954–1958. arXiv: 0909.4761 ISBN: 1520-6106.
- [46] Davis JG, Gierszal KP, Wang P, Ben-Amotz D. Water structural transformation at molecular hydrophobic interfaces. *Nature*. 2012; 491(7425): 582–585. Publisher: Nature Publishing Group ISBN: 1476-4687 (Electronic)\|0028-0836 (Linking).
- [47] Graziano G. Comment on “Water’s Structure around Hydrophobic Solutes and the Iceberg Model”. *The Journal of Physical Chemistry B*. 2014; 118(9): 2598–2599.
- [48] Galamba N. Reply to “Comment on ‘Water’s Structure around Hydrophobic Solutes and the Iceberg Model’”. *The Journal of Physical Chemistry B*. 2014; 118(9): 2600–2603.
- [49] Ball P. Water is an active matrix of life for cell and molecular biology. *Proceedings of the National Academy of Sciences*. 2017; 114(51): 13327–13335. ISBN: 0027-8424, 1091-6490.

- [50] Chaimovich A, Shell MS. Tetrahedrality and structural order for hydrophobic interactions in a coarse-grained water model. *Physical Review E*. 2014; 89(2): 022140. ISBN: 1550-2376 (Electronic) 1539-3755 (Linking).
- [51] Stock P, Monroe J, Utzig T, Smith D, Shell M, Valtiner M. Unraveling Hydrophobic Interactions at the Molecular Scale Using Force Spectroscopy and Molecular Dynamics Simulations. *ACS Nano*. 2017; 11(3).
- [52] Abascal JLF, Vega C. A general purpose model for the condensed phases of water: TIP4p/2005. *The Journal of Chemical Physics*. 2005; 123(23): 234505.
- [53] Shell MS. Systematic coarse-graining of potential energy landscapes and dynamics in liquids. *The Journal of Chemical Physics*. 2012; 137(8): 084503.
- [54] Abraham MJ, Murtola T, Schulz R, et al. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*. 2015; 1-2: 19–25. arXiv: 1503.05249v1 ISBN: 23527110.
- [55] Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. *The Journal of Chemical Physics*. 1995; 103(19): 8577–8593. ISBN: 0021-9606.
- [56] Miyamoto S, Kollman PA. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of Computational Chemistry*. 1992; 13(8): 952–962. ISBN: 0192-8651.
- [57] Chaimovich A, Shell MS. Length-scale crossover of the hydrophobic interaction in a coarse-grained water model. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*. 2013; 88(5): 1–11.
- [58] Hamaker H. The London—van der Waals attraction between spherical particles. *Physica*. 1937; 4(10): 1058–1072.
- [59] Huang DM, Chandler D. The Hydrophobic Effect and the Influence of Solute - Solvent Attractions. *Journal of Physical Chemistry*. 2002; 106(V): 2047–2053.
- [60] Weeks JD, Chandler D, Andersen HC. Role of Repulsive Forces in Determining the Equilibrium Structure of Simple Liquids. *The Journal of Chemical Physics*. 1971; 54(12): 5237–5247.
- [61] Parrinello M, Rahman A. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics*. 1981; 52(12): 7182–7190. arXiv: 1011.1669v3 ISBN: 0021-8979.
- [62] Rensing RC, Weeks JD. Dissecting hydrophobic hydration and association. *Journal of Physical Chemistry B*. 2013; 117(49): 15479–15491. arXiv: 1502.05220v1 ISBN: 1520-6106.

- [63] Remsing RC, Rodgers JM, Weeks JD. Deconstructing Classical Water Models at Interfaces and in Bulk. *Journal of Statistical Physics*. 2011; 145(2): 313–334. arXiv: 1107.5593 ISBN: 0022-4715.
- [64] Barker JA, Henderson D. Perturbation Theory and Equation of State for Fluids. II. A Successful Theory of Liquids. *The Journal of Chemical Physics*. 1967; 47(11): 4714–4721.
- [65] Andersen HC, Weeks JD, Chandler D. Relationship between the Hard-Sphere Fluid and Fluids with Realistic Repulsive Forces. *Physical Review A*. 1971; 4(4): 1597–1607.
- [66] Bennett CH. Efficient Estimation of Free Energy Differences from Monte Carlo Data. *Journal of Computational Physics*. 1976; 22: 245–268.
- [67] Shirts MR, Chodera JD. Statistically optimal analysis of samples from multiple equilibrium states. *The Journal of chemical physics*. 2008; 129(12): 124105.
- [68] Chodera JD, Shirts MR. A python implementation of the multistate Bennet acceptance ratio (MBAR).
- [69] Shell MS, Debenedetti PG, Panagiotopoulos AZ. Molecular structural order and anomalies in liquid silica. *Physical Review E*. 2002; 66(1): 011202.
- [70] Horn HW, Swope WC, Pitera JW, et al. Development of an improved four-site water model for biomolecular simulations: TIP4p-Ew. *Journal of Chemical Physics*. 2004; 120(20): 9665–9678. ISBN: 0021-9606; 0021-9606.
- [71] Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*. 1983; 79(2): 926. arXiv: 1011.1669v3 ISBN: 0131407333.
- [72] Lum K, Chandler D, Weeks JD. Hydrophobicity at Small and Large Length Scales. *The Journal of Physical Chemistry B*. 1999; 103(22): 4570–4577. ISBN: 1089-5647.
- [73] Mittal J, Hummer G. Static and dynamic correlations in water at hydrophobic interfaces. *Proceedings of the National Academy of Sciences*. 2008; 105(51): 20130–20135. ISBN: 0027-8424.
- [74] Chandler D. Interfaces and the driving force of hydrophobic assembly. *Nature*. 2005; 437(7059): 640–7.
- [75] Pratt LR. Molecular Theory of Hydrophobic Effects: "She is too mean to have her name repeated."*. *Annual Review of Physical Chemistry*. 2002; 53(1): 409–436.
- [76] Ashbaugh HS, Pratt LR. Colloquium: Scaled particle theory and the length scales of hydrophobicity. *Reviews of Modern Physics*. 2006; 78(1): 159–178. ISBN: 0034-6861.

- [77] Lazaridis T. Inhomogeneous Fluid Approach to Solvation Thermodynamics. 1. Theory. *The Journal of Physical Chemistry B*. 1998; 102(18): 3531–3541. ISBN: 1089-5647.
- [78] Ben-Amotz D, Raineri FO, Stell G. Solvation Thermodynamics: Theory and Applications [†]. *The Journal of Physical Chemistry B*. 2005; 109(14): 6866–6878.
- [79] Raineri FO, Stell G, Ben-Amotz D. New mean-energy formulae for free energy differences. *Molecular Physics*. 2005; 103(21-23): 3209–3221.
- [80] Ben-Amotz D. Water-Mediated Hydrophobic Interactions. *Annual Review of Physical Chemistry*. 2016; 67(1): 617–638.
- [81] Underwood R, Tomlinson-Phillips J, Ben-Amotz D. Are Long-Chain Alkanes Hydrophilic? *The Journal of Physical Chemistry B*. 2010; 114(26): 8646–8651. ISBN: 1520-6106.
- [82] Underwood R, Ben-Amotz D. Communication: Length scale dependent oil-water energy fluctuations. *The Journal of Chemical Physics*. 2011; 135(20): 201102.
- [83] Shell MS. The relative entropy is fundamental to multiscale and inverse thermodynamic problems. *Journal of Chemical Physics*. 2008; 129(14). ISBN: 0021-9606.
- [84] Shell MS. Coarse-Graining With The Relative Entropy. *Advances in Chemical Physics*. 2016; 161: 395–441. ISBN: 9781119290971.

Chapter 4

Computational design of surface chemical patterns to optimize proximal water mobility

Reproduced in part with permission from:

Monroe J. I., & Shell M. S. (2018). Computational discovery of chemically patterned surfaces that effect unique hydration water dynamics. *Proceedings of the National Academy of Sciences*, 115(32). <https://doi.org/10.1073/pnas.1807208115>

4.1 Introduction

Water dynamics near solid interfaces play a critical role in numerous technologies, including water purification, filtration and adsorption, chromatography, and catalysis. Modifying surface hydrophobicity and chemistry, by altering the average coverage or surface density of functional groups, is well-known to influence the dynamics of hydrating

water.¹⁻⁵ Beyond this macroscopic view, however, there also potentially exists a rich design space for tuning dynamics related to the nanoscale patterning of functional groups at fixed coverage. Indeed, efforts have exploited patterning at nanoscopic to microscopic length-scales to develop biomimetic surfaces with super-hydrophobic⁶ and other anomalous, even time-dependent, wetting behaviors.⁷ Even though a fundamental understanding of the effect of surface chemical heterogeneity on water dynamics is incomplete, there is good reason to expect marked behaviors. Recent theoretical work has shown that surface heterogeneity impacts interfacial thermodynamic properties (including hydrophobicity), in particular highlighting non-additive, pattern-specific effects upon arranging hydrophobic and hydrophilic groups on surfaces.⁸⁻¹¹ Other work has also emphasized non-additivities through the failure of macroscopic theories, like the Cassie-Baxter contact angle equation, applied to heterogeneous surfaces.¹²

Nature is ripe with examples of biological interfaces that capitalize on chemical heterogeneity, using precise functional group patterns to produce unique hydration dynamics.¹³⁻¹⁵ In particular, proteins arrange surface amino acids to cooperatively adjust hydration shell water in ways that impact the mechanism and timescale of protein folding and association.¹⁴⁻¹⁷ For example, large hydrophobic patches at protein surfaces can induce large fluctuations in nearby water density,¹⁸ strengthening and accelerating protein-protein association.¹⁹ More generally, spatial heterogeneity in hydration shell dynamics, influenced by local geometry and chemical patterning, is a hallmark of proteins of all sizes and functions.^{20,21} Recent experimental measurements have also shown that such heterogeneity is unique to the folded, structured protein and does not appear in corresponding peptide fragments or intrinsically disordered proteins, suggesting the importance of spatially-organized heterogeneity in the well-defined folded structure.⁵

Are there design principles for biologically-inspired interfacial patterning that can be translated to arbitrary synthetic systems to control hydration water dynamics? We use

simulations of model surfaces and inverse design algorithms to uncover how functional group patterning alone can influence hydration dynamics in unique and significant ways. While several case-studies on specific systems have established how average surface properties such as functional group density^{2,3,22} and surface-water energetics^{23,24} influence hydration dynamics, to our knowledge this is the first systematic study to provide a general framework for the effect of heterogeneities on dynamics. Here we focus only on the chemical patterning of a surface without the additional effect of confinement, which is also known to induce unusual water dynamics,^{15,25} both alone²⁶ and in conjunction with surface patterning.^{8,27}

We examine three distinct model systems that provide complementary perspectives for the influence of surface heterogeneity: (i) Silica is a model material in many catalytic and separation processes, and the $10\bar{1}$ interface of α -cristobalite allows variations in silanol (Si-OH) densities that span a spectrum of moderately to highly hydrophilic (contact angles of $\theta \approx 42^\circ$ to 0° ²⁸). (ii) Mixed methyl- and hydroxyl-terminated SAMs represent a softer interface with a wider range of hydrophathy, spanning highly hydrophobic ($\theta \approx 121^\circ$) to hydrophilic ($\theta \approx 0^\circ$) surfaces. (iii) As a control, we also examine toy Lennard-Jones (LJ) surfaces containing two uncharged particle types with hydrophobic/weak ($\theta \approx 127^\circ$) and hydrophilic/strong ($\theta \approx 91^\circ$) van der Waals interactions with water molecules. In specific cases noted below, we have also considered LJ surfaces in which the hydrophilic particles are “super-attractive” ($\theta \approx 49^\circ$), with a water-particle interaction energy double the original hydrophilic case. Importantly, all LJ particles interact only isotropically with water, whereas the hydroxyl groups in both the silica and SAM surfaces produce directional hydrogen bonding. All three systems allow distinct coverages and patterning of hydrophilic groups through the spatial arrangement of silanol groups on cristobalite, hydroxyl-terminated SAM chains, or strongly attractive LJ particles.

Rather than exhaustively explore all surface configurations for a fixed number of

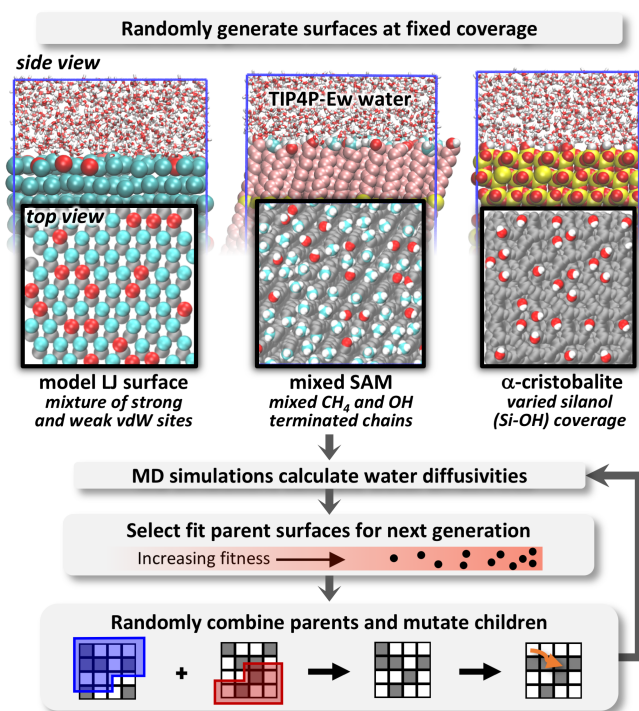


Figure 4.1: A schematic of the workflow for the genetic algorithm to optimize (minimize or maximize) hydration water dynamics via re-patterning of surfaces at various fixed coverages of hydrophilic groups. Surfaces studied include the $10\bar{1}$ face of α -cristobalite with varied silanol coverage, self-assembled monolayer (SAM) surfaces with mixed methyl- and hydroxyl-terminated chains, and idealized surfaces of mixed binary Lennard-Jones (LJ) particles with either strong or weak LJ-water interactions.

hydrophilic groups, we develop a genetic algorithm for discovering surface configurations that either maximize or minimize the dynamics of hydration layer water, as shown in Fig. 4.1. In effect, this approach identifies extremal surfaces that provide bounds for realizable water dynamics due to patterning and, in doing so, magnifies surface characteristics that impact and control mobility. The genetic algorithm rearranges surface chemical groups at fixed coverage, and for each new surface pattern it uses molecular dynamics simulations to assess the hydration water dynamics, quantified by the parallel diffusivity for waters within 8 Å of the surface. The algorithm evolves surface patterns toward extremal water dynamics through creation and evaluation of subsequent “generations” of surfaces. Further details are provided in Methods.

4.2 Methods

4.2.1 Model systems

We model all surfaces using molecular dynamics (MD) simulations of 3D periodic systems with slab geometries (Fig. 4.1). Surfaces are solvated on each side by TIP4P-Ew water,²⁹ although we found no qualitative differences using other water models. Fully hydroxylated crystalline slabs of the $10\bar{1}$ face of α -cristobalite were derived from the fully-silicated library provided by Emami and coworkers.²⁸ To reach lower densities for these models, we perform in-silico “condensations” of two nearest-neighbor silanol groups, using PARMED³⁰ to remove both hydrogens and one oxygen from the topology, then creating all necessary bonds, angles, and dihedrals surrounding the newly bridging oxygen. Force field functional forms and parameters come from the silica portion of INTERFACE-FF.²⁸

We model methyl-and hydroxyl-terminated SAM chains similarly to the efforts of

Garde and coworkers,^{31,32} Levine, et al,³³ and Zerze, et al,³⁴ though we introduce specific modifications as described below. As in previous work, united atoms represent subsurface groups while atomic detail is used for head groups. In the work of Mondello³⁵ referenced by the Garde group, all bonds between united atoms were constrained, whereas here we only constrain bonds involving hydrogens. In previous parametrizations of the SAM surface, it appears based on prior descriptions of methodology that no angle potential energy function was applied to the C-O-H angle of hydroxyl-terminated chains, with the OPLS/AMBER parameters³⁶ for such an angle instead applied to the C-C-O angle. We apply the usual OPLS/AMBER C-C-O and C-O-H angle terms to these atoms, which somewhat limits the flexibility and hydrogen bonding ability of the hydroxyl head group. Chains are arranged with their sulfur atoms in a hexagonal close-packed (HCP) pattern with nearest-neighbor spacing of 4.97 Å. We place a total of 72 chains on this lattice consisting of 9 rows (along the y dimension) and 8 columns (along the x dimension), resulting in surfaces of size 39.7600 x 38.7373 Å².

We generate surfaces of Lennard-Jones particles on a 2D lattice identical to that of the SAM systems, with each layer along the direction perpendicular to the interface translated appropriately to maintain an overall HCP crystal structure. This includes 5 lattice layers, which makes slabs for all systems (including SAM and cristobalite surfaces) at least ~ 20 Å thick. “Hydrophobic” LJ particles use parameters for OPLS united-atom methane ($\epsilon=1.2301$ kJ/mol and $\sigma=3.730$ Å),³⁷ while “hydrophilic” ones have a six times greater ϵ for interactions with water. All inter-surface interactions for LJ particles follow the purely hydrophobic case so that surface-surface energetics are not influenced by patterning. “Super-attractive” LJ particles are also employed in some simulations, with an epsilon that is twelve times greater than that for hydrophobic LJ particles.

The original SAM and LJ surfaces developed in our work entail a defect in the HCP lattice due to the use of an odd number of rows. This is manifested more obviously for

the LJ surfaces, for which the particles are restrained in place whereas the flexible chains of the SAM are able to re-arrange so that the surface-water interface is still approximately HCP. Re-running the genetic algorithm at a hydrophilic fraction of 0.25 for re-constructed SAM and LJ surfaces with 10 rows and no defect (dimensions of $39.7600 \times 43.0415 \text{ \AA}^2$) makes no difference to the diffusivity values or range measured, and thus leads to no change in the conclusions presented.

4.2.2 Molecular dynamics details

We use the 2016.1 release of GROMACS³⁸ for all MD simulations. In all cases, the time step is 0.002 ps and we constrain all bonds involving hydrogens. Short-range, non-bonded interactions are truncated and shifted to zero at a 1.2 \AA cut-off, with default GROMACS treatment for long-range electrostatics. All runs consist of ~ 8000 atoms and utilize GPU-accelerated code on NVIDIA GTX 1080 Pascal GPUs provided by Exxact. For all MD simulations, the velocity Verlet method³⁹ is used to propagate the equations of motion with constraints placed on all bonds involving hydrogens through LINCS.⁴⁰ The Smooth Particle Mesh Ewald algorithm, with default GROMACS parameters, treats long-range electrostatics.⁴¹

Prior to each MD run, we use the following equilibration procedure: (1) The GROMACS solvate command centers and solvates a surface, followed by relaxation with steepest descent energy minimization and a very short NVT simulation with the Berendsen thermostat.⁴² (2) Short NPT equilibration with a Berendsen thermostat and semi-isotropic barostat (volume fluctuations only perpendicular to the surface) equilibrates the box size. (3) A second short pre-production NPT simulation evaluates average box size, which is then fixed. (4) Finally, a short NVT simulation utilizing a weak Nose-Hoover thermostat^{43,44} with 10 chains⁴⁵ precedes a 10 ns production run with the same settings

and frames saved every 0.5 ps.

We separately remove center of mass momentum from the water and surface atoms every 30 time steps. This is sufficient to keep the α -cristobalite surfaces motionless relative to water. For the SAM surface, all united sulfur atoms are harmonically restrained in space, with each sulfur reference position scaled separately during NPT runs. LJ surface particles are also harmonically restrained, with the center of mass reference position of all particles rescaled during NPT simulations.

4.2.3 Calculation of diffusivities

To assess hydration water mobility, we determine the diffusivity via linear fits to the 2D mean-squared displacement (MSD) of water oxygen atoms parallel to the plane of the surface. Similar to earlier simulation studies,^{22,23} we collect MSD data from water oxygen atoms as long as they remain within 8 Å of the surface, as defined by the most extreme surface heavy atom coordinates. For all systems, this includes approximately 2 hydration layers. We focus this calculation to a time window of 2-10 ps, which we find to be optimal for assessing lateral water dynamics near interfaces. At times shorter than this, waters do not exhibit diffusive behavior, while at times beyond 10 ps, too many waters have left the surface and the slope of the MSD curve decreases as only the slowest-moving waters remain. Profiles of water lateral diffusivity in the direction perpendicular to the interface are obtained by the same procedure but using 3 Å slices with their centers spaced 0.5 Å apart along the interface normal.

4.2.4 Calculation of excess hard-sphere chemical potentials

We compute the spatially-varying excess chemical potential for hard-spheres of radius 3.3 Å using the Widom insertion technique.⁴⁶ We discretize the simulation box into cells

and count the number of times a randomly-inserted sphere in any cell is empty of water oxygens or surface heavy atoms. Profiles of μ_{ex}^{HS} result from placing spheres randomly within 0.5 Å slices perpendicular to the interface, with at least 5.04×10^8 insertions attempted for each slice. The negative log probability for a successful insertion (main text Equation 1) then provides the hard-sphere chemical potential and its spatial dependence. In the results presented, we reference hard-sphere excess chemical potentials to the case in bulk TIP4P-Ew water at 300 K and 1 bar, where $\mu_{ex}^{HS} \approx 10.476 k_B T$.

4.2.5 Determination of interfaces and interface fluctuations

We use two definitions for the surface-water interface to spatially map surface hydrophobicity and hydrophilicity. The “mean” interface is defined by the box Z-coordinate where the average density of water becomes 0.3 of its bulk value (0.0332 \AA^{-3} for TIP4P-Ew water). “Instantaneous” interfaces are defined according to the procedure of Willard and Chandler,⁴⁷ utilizing the same parameters except for a density fraction cut-off of 0.3 instead of 0.5. In short, each water oxygen is assigned a truncated and shifted Gaussian function, with the density field represented by the sum of such functions. The interface is the iso-surface within this field where the density is 0.3 of the bulk value. The `marching_cubes_lewiner` method⁴⁸ provided by the `scikit-image` package⁴⁹ identifies the iso-surface after first evaluating the density field on a 1.0 Å grid. For each X-Y bin on the grid (i.e. bins perpendicular to the surface slab), we prune the density field as necessary such that only two Z-coordinate iso-surface points are kept representing an upper and lower interface. Distances of waters to such an interface are defined by projecting the vector between a water oxygen and the nearest surface point onto the surface unit normal of said surface point.

4.2.6 Calculation of water re-orientation times

Water re-orientation times are determined by fitting a stretched exponential function to the time-decay of the cosine of the angle between the water dipole vector at an initial and later time. This was done only for waters remaining in the surface-water distance range of interest, as described by Debenedetti and coworkers.²³ Profiles result from using 3 Å slices in the direction perpendicular to the interface with slice centers spaced 0.5 Å apart.

4.2.7 Calculation of contact angles

We calculate water contact angles by placing an equilibrated water box of 2000 TIP4P-Ew water molecules over SAM or LJ surfaces with dimensions of about 119 x 116 Å². Drops are equilibrated under NVT conditions with molecular dynamics parameters as described in other sections. Once the drop shape stops changing (~1 ns for hydrophobic surfaces, and 5-10 ns for hydrophilic surfaces), a 10 ns production simulation is used to calculate the average drop geometry. The direction perpendicular to the interface is discretized into 1.0 Å bins, with the maximum water oxygen radial distance from the drop center of mass recorded at each timestep for each bin. At each frame, a fixed cut-off distance from the drop center is used to ensure that gas-phase waters were not considered. The average maximum radial distance versus height from the interface is fit to a second order polynomial, with the coefficient on the term linear in radius providing the inverse tangent of the contact angle.

4.2.8 Genetic algorithm implementation

The genetic algorithm workflow of Fig. 4.1 is coded in python 2.7, making extensive use of the PARMED³⁰ and PYTRAJ⁵⁰ packages. “Individuals” refers to specific realizations

of surface patterns, “parents” to a subset of fit individuals selected to produce “children” of the next generation. Boolean arrays store surface configurations with True values indicating that a surface site (i.e. silanol, SAM chain, or LJ particle) is hydrophilic. Initially, a pool of randomly generated surfaces/individuals is created. Tournament selection of parents collects high-fitness candidates (either high or low parallel diffusivity depending on the optimization direction) for generating children. To combine two randomly paired parents, a random fraction of the first’s True values is combined with the conjugate fraction of the second’s, and a second child stems from the flipped version. While this strategy generally preserves surface coverage, in cases where it does not, we randomly mutate child sites until the desired surface density is reached. Once produced, children undergo a fixed number of random mutations proportional to the surface coverage via swapping of True and False Boolean array values; we skip this procedure if the surfaces have already reached the target number of mutations during the process of ensuring fixed surface density. We also explored an advanced algorithm approach involving clustering of surfaces, called “nicheing” in the genetic algorithm literature, but found only marginal gains in efficiency and results that were not distinct from the more basic procedure.

We have tuned the algorithm to efficiently utilize 8 available GPU processors. Each “generation” consists of 8 individuals (surfaces), and the first 11 generations (or 88 surfaces from generation 0 to 10) use random patterns to build a library of potential parents. Subsequently, the genetic optimization produces 20 generations. A second 20 generation optimization starting from the same random 88-surface pool but using a distinct random number seed follows. Finally, we perform another 20 generations of optimization after combining the full surface libraries from the random generations and both independent optimization runs. The progress of genetic algorithm runs to minimize and maximize diffusivity near cristobalite surfaces are shown in Figure X. Each generation requires ~ 2 hours for all surface types, with the majority of time in MD simulations. Since each surface

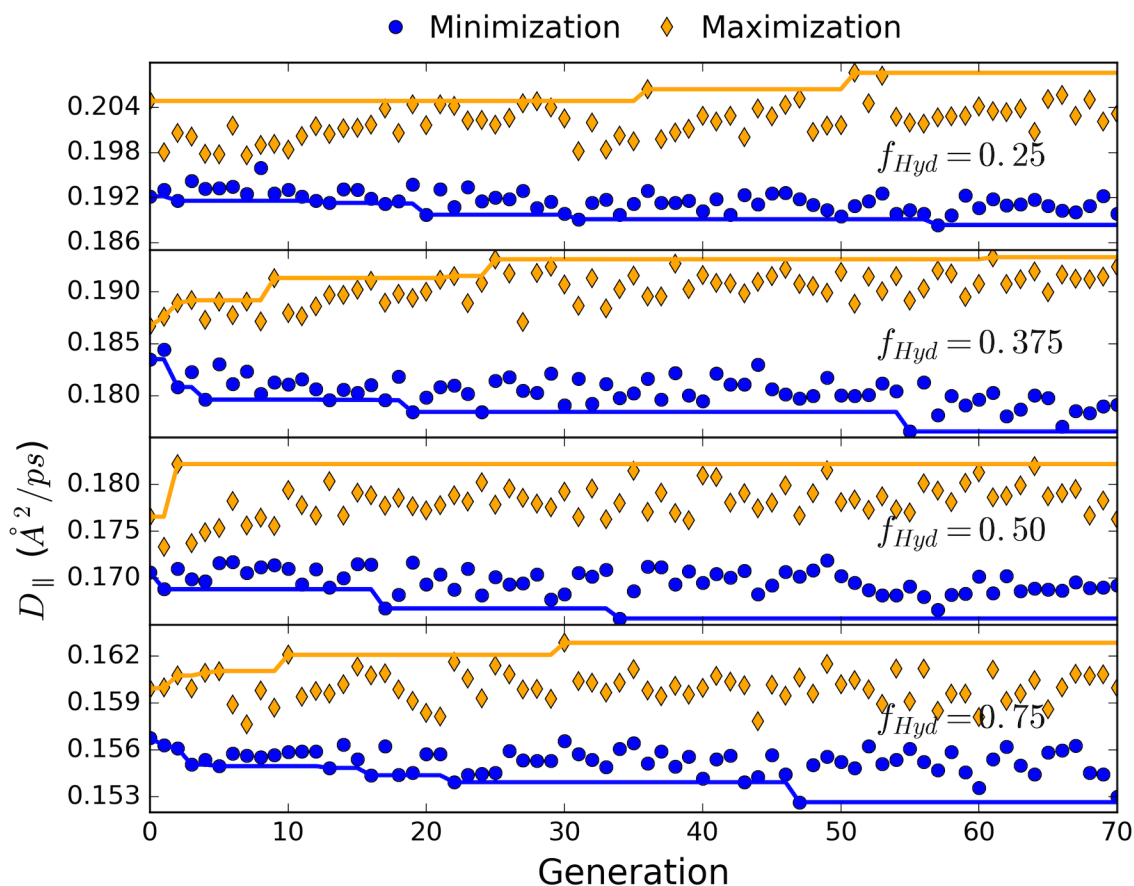


Figure 4.2: Evolution of the genetic algorithm for various densities of the cristobalite surface. Blue points are minimum diffusivity surfaces for a given generation during minimization runs of the algorithm while orange points are maximum diffusivity surfaces from a separate maximization run. Blue and orange lines represent convex hulls. Generations 0 through 10 represent surfaces that were randomly generated without any optimization. For these early points, the blue and orange give the minimum and maximum diffusivities in the set of initial random surfaces.

is identical on its upper and lower faces, 10 ns MD simulations produce effectively 20 ns of data used to evaluate diffusivities. In total for each particular surface and coverage, the genetic algorithm performs 1048 MD simulations requiring 10.5 μ s of simulation time.

4.3 Results and Discussion

4.3.1 Patterning of surfaces produces a wide range of diffusivities at fixed surface coverage

For all surface types studied, the diffusivity monotonically decreases as the number of hydrophilic groups (Si-OH, OH head-groups, or more attractive LJ particles) increases, traversing a range of at least $0.1 \text{ \AA}^2/\text{ps}$ in diffusivity between 0% and 100% hydrophilic surface fraction (Fig. 4.3). Diffusivities generally correlate with surface hydrophobicity as given by the simulation contact angle, increasing in the order of the cristobalite, SAM, and LJ surfaces. Similar to the way that hydration water thermodynamic properties vary with hydrophilic coverage,³¹ we observe an asymmetry in the response of water dynamics to addition of hydrophilic sites: at low hydrophilic fractions, the reduction in mobility is more pronounced than for higher coverages, giving rise to an overall convex relationship in Fig. 4.3. This effect is most apparent on surfaces with a wide range of effective hydrophilicity/phobicity, including the SAM surfaces and LJ surfaces composed of hydrophobic and super-attractive bead types. While the asymmetry is less apparent for the cristobalite and standard LJ cases, the overall relationships still show a discernable convex character.

Importantly, for both the cristobalite and SAM surfaces, the genetic algorithm identifies a wide range of diffusivities resulting from chemical patterning at partial surface coverages, where the number of possible patterns is combinatorically large. At intermediate

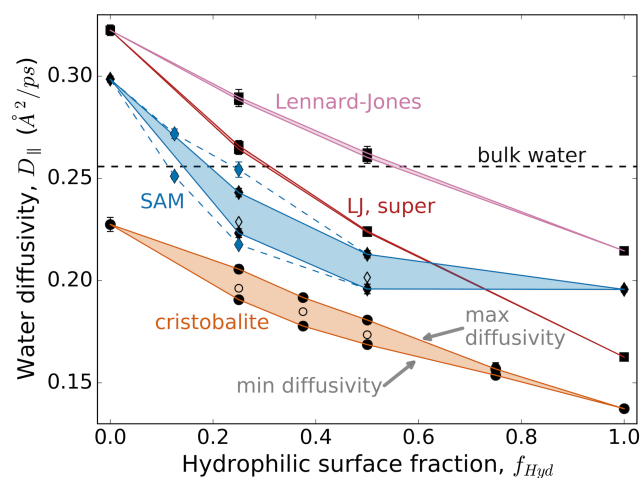


Figure 4.3: Regions of accessible hydration water diffusivities due to surface repatterning for the model surfaces. The hydrophilic surface fraction is defined for cristobalite as the percentage of the maximum possible silanol (Si-OH) density, for the SAMs as the fraction of hydroxyl-terminated chains, and for the Lennard Jones surfaces as the fraction of strongly attractive van der Waals sites. Filled symbols show the minimum and maximum diffusivity found by the genetic algorithm, while open symbols show the average diffusivity for randomly generated surfaces. Red curves show diffusivities for super-attractive LJ surfaces, using patterns optimized from the original LJ case. The blue dashed lines and blue diamonds show the expanded diffusivity range for SAM surfaces at surface fractions of 0.125 and 0.25 with regularly spaced and circularly-clustered hydroxyl groups (see Fig. 4.5). The black dashed line gives the average lateral diffusivity of simulation bulk water. Error bars, which are typically smaller than the symbol size, give 1% confidence intervals assuming Student's t-distribution from repeated simulation runs of the same surface.

hydrophilic fractions, patterning of these surfaces can produce a $\sim 10\%$ change in absolute diffusivity. It is interesting to note, however, that the patterning-induced variation in diffusivity is a substantial fraction of the overall diffusivity range achievable by a given chemistry (i.e., of the range defined between zero and full hydrophilic surface fraction); this relative variation is 16% for cristobalite and 20% for SAM surfaces. Interestingly, for both surface types the largest spread between minimum and maximum diffusivity occurs at coverages skewed towards smaller hydrophilic fractions. As shown in Table 1, the differences in diffusivities due to these surface reconfigurations are statistically significant and reproducible: in nearly all cases, the probability for erroneously predicting that the simulation-calculated minimum surface diffusivity is greater than the maximum is negligible. In contrast, no statistically significant range of diffusivities (for a 5% confidence interval) is obtained from genetic algorithm optimizations of the LJ surface at any coverage, implying that unlike the other two cases, surface patterning for this system has no measurable effect on water dynamics. This is in spite of the fact that both LJ cases in Fig. 4.3 show some non-linear character. While this non-linearity is expected due to the higher probability of finding waters near more attractive sites (see Appendix D), the lack of pattern-dependence is unexpected and discussed shortly.

The ability of patterning to modulate hydration water dynamics is surprising from a macroscopic perspective. Purely continuum arguments would suggest that heterogeneous surfaces consisting of distinct macroscopic regions differing in hydration water diffusivity should see no patterning effect. Instead, the long-time interfacial dynamics on such surfaces should only depend on the fraction of each region type, unless molecular-scale effects are introduced in the form of barrier-crossing dynamics at region boundaries (see Appendix D). Indeed, our results show that molecular-scale (< 1 nm) patterning is essential to produce pronounced changes in water mobility as surface patterns are adjusted, even without confinement or nontrivial variations in surface geometry.

Surface	P-value for hypothesis that min diffusivity is greater than max using Welch's t-test*	Probability to observe min diffusivity surface pattern at least once through random sampling alone**	Probability to observe max diffusivity surface pattern at least once through random sampling alone**
cristobalite-10 $\bar{1}$, $f_{Hyd} = 0.25$	5.20×10^{-12}	9.85×10^{-1}	1.07×10^{-2}
cristobalite-10 $\bar{1}$, $f_{Hyd} = 0.375$	0.00×10^0	5.45×10^{-1}	5.95×10^{-1}
cristobalite-10 $\bar{1}$, $f_{Hyd} = 0.50$	1.5×10^{-11}	9.98×10^{-1}	1.61×10^{-1}
cristobalite-10 $\bar{1}$, $f_{Hyd} = 0.75$	5.90×10^{-3}	9.02×10^{-1}	1.00×10^0
SAM, $f_{Hyd} = 0.25$	5.20×10^{-7}	9.66×10^{-1}	5.21×10^{-6}
SAM, $f_{Hyd} = 0.50$	8.90×10^{-6}	9.83×10^{-1}	1.05×10^{-3}
LJ, $f_{Hyd} = 0.25$	1.80×10^{-1}	1.00×10^0	1.00×10^0
LJ, $f_{Hyd} = 0.50$	6.70×10^{-2}	1.00×10^0	1.00×10^0

*Uses Student's t-distribution, but does not assume equal variances of the two sets of samples

**Assumes a Gaussian distribution of diffusivities for randomly generated surfaces, with the binomial distribution then providing the probability of obtaining at least one equally or more anomalous diffusivity from a draw of the same size as the sample generated by the genetic algorithm optimization

Table 4.1: Performance metrics for the surface design genetic algorithm

The significant diffusivity ranges in Fig. 4.3 hold important implications for experimental measurements of surface hydration water dynamics, direct or inferred. For both the cristobalite and SAM systems, surfaces of intermediate hydrophilicity can have similar water diffusivities as those with nearly half the hydrophilic coverage, if the former are patterned so as to maximize water diffusivity and the latter similar to randomly generated surfaces. Such results call into question hydrophathy scales that paint an additive picture for the effect of surface chemical groups on water. Moreover, these results suggest that surface preparation techniques may critically influence interfacial properties through

surface functional group organization. In particular, significant jumps in diffusivity with surface coverage may signal changes to the spatial clustering and distribution of groups, as recently observed for amorphous silica of varied hydroxylation.⁴ More broadly, measurements to assess the impacts of surface composition on water dynamics may be misleading if surface spatial organization is not also characterized. This further highlights the role of chemical heterogeneity as a significant design parameter in controlling water mobility near interfaces.

4.3.2 The genetic algorithm reveals that dispersion of hydrophilic groups reduces water mobility

Fig. 4.4 demonstrates that water mobility is higher near surfaces with large clusters of hydrophilic groups and low on surfaces that evenly distribute such groups, as is clear from images and two-dimensional surface $-OH$ radial distribution functions. These results for water dynamics now suggest a deep relationship to underlying thermodynamic and structural properties of hydration water. Ongoing efforts have sought to understand how surface chemistry and heterogeneities impact interfacial water density fluctuations that predict surface hydrophobicity,^{8,9,31,51,52} and affect dewetting transitions for water confined between plates or in the cavities of proteins.^{8,10,18,53} Of particular note, Garde and coworkers found that surface groups impact water in a highly non-additive manner for small length scale features; namely, patches of hydrophobic groups on hydrophilic surfaces must become larger than ~ 1 nm before impacting water density fluctuations, while even single hydrophilic groups significantly alter the apparent local hydrophilicity of hydrophobic surfaces.⁸⁻¹¹ Our results are consistent, in that dispersion of hydroxyl groups produces an overall more hydrophilic surface with low water mobility, while hydrophilic clustering permits the exposure of locally more hydrophobic regions that exceed the

critical ~ 1 nm size necessary to impact local water diffusivity.

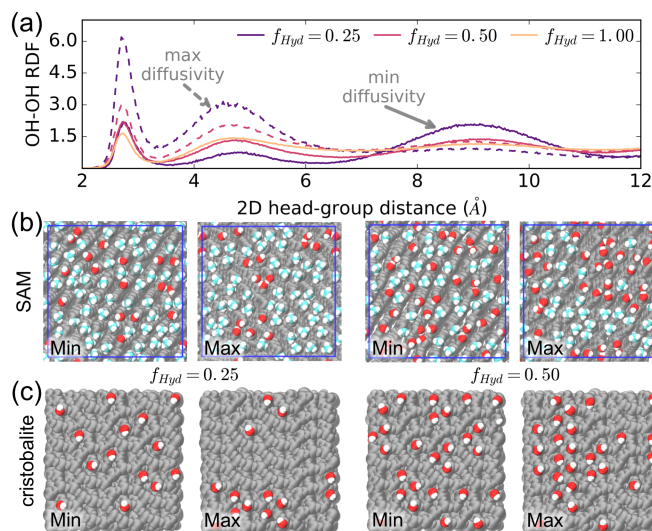


Figure 4.4: (a) 2D radial distribution functions for SAM hydrophilic groups show that maximum-diffusivity cases strengthen spatial correlations at short distances. Solid lines for RDFs indicate minimum diffusivity surfaces, while dashed lines give maximum diffusivity cases. (b, c) Simulation snapshots of the minimum and maximum diffusivity patterns for SAM and cristobalite surfaces at hydroxyl fractions of 0.25 and 0.50. Clustering of hydroxyl groups (shown in red) on maximum diffusivity surfaces is clear from visualization of surface structures.

It is possible that more extremal patterns exist such that the true diffusivity ranges are even broader, since the genetic algorithm finds only minimal bounds due to its finite sampling of the surface landscape. To test the algorithm's convergence, we simulated SAM surfaces that represent the extremes of discovered patterns, namely regularly-dispersed and circularly-clustered hydroxyl-terminated chains for hydrophilic chain fractions of 0.125 and 0.25 (Fig. 4.5). As seen from Fig. 4.3, such patterns do slightly increase the range of diffusivities (dashed blue lines), showing that while the algorithm helps to objectively define the key features of optimal surfaces, it does not find the global optima. As an interesting contrast, we also simulated a more exotic pattern involving a single stripe of hydroxyls along the center of a SAM surface, again at 25% coverage. This pattern exhibits a diffusivity of $0.252 \text{ \AA}^2/\text{ps}$, almost as large as the circularly-clustered configuration,

and notably exhibits an anisotropy of $0.003 \text{ \AA}^2/\text{ps}$ between the directions parallel and perpendicular to the stripe (significant on the 5% confidence interval). Compared to the overall effect of patterning, this anisotropy is small; however, in principle the algorithm could be configured to locate patterns that maximize the anisotropy, resulting in potentially unintuitive arrangements. We leave this for future work.

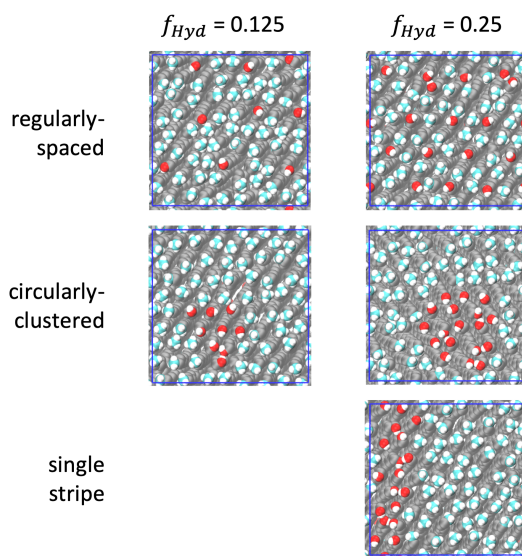


Figure 4.5: Top-views of surfaces with specific patterns to further test the genetic algorithm. Regularly-spaced and circularly-clustered hydroxyl groups represent the extremes of diffusivity suggested by populations of surfaces discovered by the genetic algorithm. The single stripe configuration for 25% hydroxyl-terminated chains represents a more exotic patterned to rudimentarily test for anisotropy in the diffusivity in directions parallel to and perpendicular to the stripe.

In assessing the algorithm, it is also important to ensure that water diffusivity is optimized through functional group patterning and not by other means. Figures E.1-E.9 in Appendix E demonstrate that variations in simulation conditions other than patterning, like temperature, pressure, or even the rigidity and structure of SAM chains, are not significantly modified by the algorithm. Table 1 also characterizes the algorithm's performance by estimating the probability of randomly observing at least one occurrence of an extremal diffusivity surface beyond the minimum and maximum currently found

at each coverage. Overall, minimum diffusivity surfaces are more likely to be observed among random surfaces than maximum diffusivity ones, but the low probabilities for the latter show that the algorithm is helpful, particularly at intermediate coverages. This occurs because higher water mobility surface patterns spatially cluster hydroxyl groups, and such cases are smaller in number and more difficult for the algorithm to locate than randomly dispersed ones. The same behavior is also highlighted by the fact that randomly generated surfaces have diffusivities more similar to minimum-diffusivity ones (Fig. 4.3).

4.3.3 Orientation-specific surface-water interactions drive mobility reductions

We find that molecular measures of hydrophobicity for a given heterogeneous surface also reflect surface patterning in a manner consistent with variations in hydration water diffusivity. Spatially resolved density fluctuations within probes placed near interfaces reveal overall more hydrophilic surfaces (lower density fluctuations) for those with lower water mobilities (Appendix E, Fig. E.10). More hydrophobic surfaces are also expected to exhibit increased fluctuations in the location of the surface-water interface itself.⁴⁷ To define an instantaneous interface, we follow the water density isosurface definition of Willard and Chandler,⁴⁷ and indeed see heightened fluctuations near regions of higher hydrophobicity and lower ones for hydrophilic locations (Appendix E, Fig. E.11). Though not pursued here, we expect that other estimates of local surface hydrophilicity^{54,55} will provide qualitatively similar results.

To quantify and compare the average hydrophobicity of a surface to average water mobility, we consider the excess chemical potential of hard sphere insertion, given by

$$\mu_{ex}^{HS} = -k_B T \ln \wp_V \quad (4.1)$$

where φ_V gives the probability for successful insertion of a hard sphere of volume V randomly placed within a region of interest. Such excess chemical potentials can probe local hydrophobicity as an effective hydrophathy map of heterogeneous, molecular-scale surfaces.^{9,56} They are also intimately tied to the magnitude of molecular scale fluctuations in water density, as a successful hard sphere insertion only occurs when the probe volume density fluctuates to zero.⁵⁷

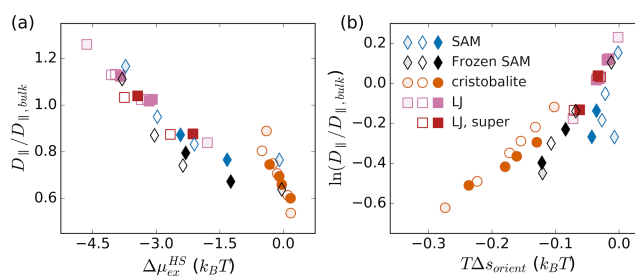


Figure 4.6: (a) Hydration water diffusivity correlates with the excess chemical potential for hard-sphere insertion near the interface across all surface types and patterning. Hatched points represent surfaces at hydrophilic fractions of zero or one, filled points minimum diffusivity surfaces, and open points maximum diffusivity ones. (b) The logarithm of water diffusivity correlates with the orientational contribution to the one-body intensive entropy.

Fig. 4.6a shows that μ_{ex}^{HS} for methane-sized (3.3 \AA) spheres within 8 \AA of the surface correlates remarkably well with hydration water diffusivity, even across a wide range of surface types, coverages, and patterning. Comparable relationships are observed when instantaneous interface location fluctuations are taken into account (although the correlation weakens for the SAM surfaces due to their compressibility; see Appendix E, Fig. E.12). The notion of a relationship between mobility and μ_{ex}^{HS} is not new; Mittal et al.⁵⁸ found a strong correlation in the hard sphere fluid. However, the persistence of this behavior here is surprising given water's very different density fluctuations and liquid structure, which depart from simple liquids like hard spheres and influence many water-unique features of hydrophobic solvation.⁵⁹ To illustrate the distinction, Fig. 4.7 shows water density, excess hard-sphere chemical potential, water parallel diffusivity,

and re-orientation times as a function of distance from the interface. While mobility is spatially correlated with density for confined hard sphere fluids,⁵⁸ here the average water density only weakly predicts diffusivity and it fails to correctly rank surfaces in terms of either hydrophobicity or hydration water mobility. Instead because μ_{ex}^{HS} is predictive, it is clear that water density fluctuations are strongly and locally correlated with both translational and orientational water mobility, a trend that was noted previously for the singular case of a hydrophobic SAM.⁶⁰

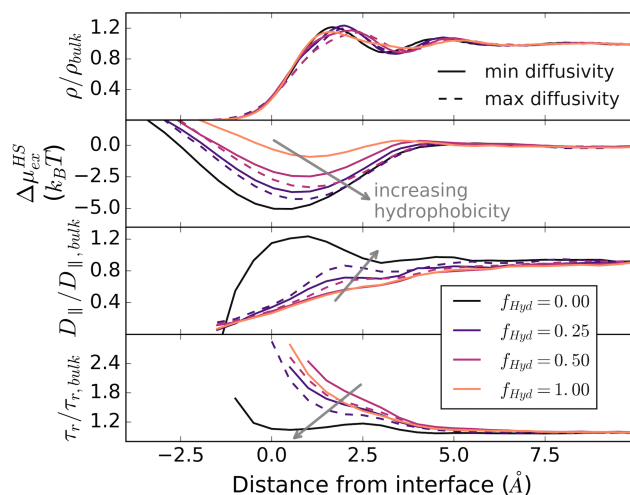


Figure 4.7: Spatial variations as a function of distance from the mean interface (water density at 0.3 of its bulk value) are shown for water density, hard-sphere excess chemical potential, water diffusivity, and water re-orientation times.

While μ_{ex}^{HS} seems to do an excellent overall job of predicting surface dynamics, it does not capture the entire picture. Specifically, binary patterning of the LJ surfaces produces small variations in μ_{ex}^{HS} but fails to impact water diffusivity (Fig. 4.6a). Indeed, earlier efforts for similar LJ surfaces have shown that particle arrangements do affect interfacial thermodynamic properties and alter surface hydrophobicity.¹⁰ It may then seem surprising that the genetic algorithm cannot uncover discernible diffusivity differences due to surface patterning in these systems (Fig. 4.3). Even if we transpose the minimum and maximum patterns from the SAM results to the LJ surface (the two systems share

the same 2D lattice), we still do not find statistically significant variations in diffusivity (Fig. 4.8). This behavior is also independent of the specific LJ interaction parameters: doubling the van der Waals strength of the already six-fold more attractive hydrophilic particles reduces the contact angle of the pure surface from 91° to 49° , but still does not produce a diffusivity difference for dispersed and clustered particle arrangements (Fig. 4.3). Yet, pattern-induced variations in μ_{ex}^{HS} are still apparent and grow larger with the super-attractive particles. This clearly disrupts the overall trend in Fig. 4.6a, suggesting that while water dynamics and local density fluctuations (probed by μ_{ex}^{HS}) are correlated, the nature of the surface-water interaction also plays a role.

The toy LJ surfaces highlight the important role of directional interactions in manipulating water dynamics and nuance the predictive capability of μ_{ex}^{HS} . This observation agrees with models of water mobility based on hydrogen bond dynamics,¹⁶ and with studies showing that interfacial mobility depends on surface affinity, water-surface hydrogen bond strength, and water orientation with respect to the surface.^{61,62} To further this idea, we consider entropic measures that quantify water orientational structure at the interface, since close relationships between fluid entropy and diffusivity are well-established for both simple fluids⁶³ and bulk water.⁶⁴ For small observation volumes, water degrees of freedom may be approximately decoupled,⁶⁵ allowing a separation of entropy into translational and orientational terms. As a first approximation, we use the program GIST⁶⁶ to calculate such contributions to the entropy change from bulk for waters within 8 \AA of the surface. This analysis ignores entropy contributions from intermolecular coupling, focusing only on changes related to single-molecule degrees of freedom.

Fig. 4.6b shows that hydration water orientation entropy strongly correlates with diffusivity. In contrast, the translational contribution to the entropy, which at the single-molecule level is only based on water density, is less predictive (Appendix E, Fig. E.14). Interestingly, the correlation of diffusivity with entropy improves if SAM

	Diffusivity ($\text{\AA}^2/\text{ps}$)	Min Surface	P-value for hypothesis that min diffusivity is greater than max using Welch's t-test	Max Surface	Diffusivity ($\text{\AA}^2/\text{ps}$)
$f_{Hyd} = 0.25$	0.2883 ± 0.0029		1.8×10^{-1}		0.2897 ± 0.0033
$f_{Hyd} = 0.25$ SAM pattern (super- attractive)	0.2874 ± 0.0022 (0.2660 ± 0.0020)		6.1×10^{-2} (9.4×10^{-1})		0.2890 ± 0.0017 (0.2643 ± 0.0021)
$f_{Hyd} = 0.50$	0.2603 ± 0.0026		6.7×10^{-2}		0.2624 ± 0.0030
$f_{Hyd} = 0.50$ SAM pattern (super- attractive)	0.2622 ± 0.0025 (0.2243 ± 0.0017)		6.4×10^{-1} (7.5×10^{-1})		0.2617 ± 0.0023 (0.2237 ± 0.0014)

Figure 4.8: Minimum and maximum diffusivity LJ surfaces from the genetic algorithm optimization, as well as LJ surfaces with the optimum patterns from SAM surfaces (values for these surfaces with super-attractive particles are shown in parentheses). Lateral diffusivities are averaged over 8 independent runs. The central column shows the P-value for the hypothesis that the minimum value is greater than the maximum using Welch's t-test, which in all cases shows that there are no statistical differences between the minimum and maximum surfaces.

chains are frozen, highlighting complexities induced by a flexible, fluctuating interface. Importantly, the LJ surfaces that demonstrate no dynamic effect of surface patterning are also invariant with respect to orientational entropy, showing a consistent and strong diffusivity-entropy correlation over the full range of coverages. Thus, while patterning can effect variations in density fluctuations and μ_{ex}^{HS} , ultimately a perspective that addresses the orientational structure of water is essential to understanding surface-induced dynamics. These findings reveal a close coupling of surface water diffusivity with orientational water entropy and suggest a future opportunity to probe the differential roles of translational and orientational entropies, the impact of surface flexibility, and the relationship between entropic measures and μ_{ex}^{HS} .

4.4 Conclusions

This work illustrates that functional group patterning on heterogeneous surfaces can produce significant variations in hydration water dynamics, even if the surface coverage remains constant. A genetic algorithm treats surface patterning as an adjustable design parameter and can precisely delineate and magnify the effect of surface heterogeneity by locating surfaces that extremize the hydration water diffusion constant. Such an optimization approach seems broadly useful and could be adapted to also discover patterned surface flexibility (stiffness) and local geometry (roughness) that optimize a variety of thermodynamic and kinetic solvent properties across many distinct kinds of surface modalities and solvents. Recent experiments have also shown that careful design of chemical heterogeneity is crucial to controlling both thermodynamic and dynamic interfacial properties in silica materials,⁴ which are ubiquitous in catalytic reaction processes. This suggests exciting opportunities for computational design of novel materials in even more diverse applications, from antifouling membrane surfaces for water purification to the

regulation of interfacial heat transfer.⁶⁷

The patterns that emerge from our optimization procedure exhibit dispersed/clumped hydrophilic groups for high/low mobility surfaces, a result reminiscent of patterning effects known to make surfaces less/more hydrophobic. Indeed, the correlation that we observe between mobility and excess hard-sphere chemical potentials suggests a deep, albeit not fully general, connection between water dynamics and density fluctuations. We also find that surfaces composed of Lennard-Jones particles, without directional or electrostatic interactions, cannot manipulate water diffusivity via their patterning, which is surprising given results here and in previous studies^{10,52,53} that demonstrate distinct, pattern-dependent changes in thermodynamic properties. While more investigation is necessary, this result highlights a key difference between the thermodynamic (water density fluctuations) and dynamic behavior of water at interfaces: interactions that make use of the anisotropy of the water molecule (e.g., multipole electrostatics) may play a more fundamental role in determining dynamics than water density fluctuations. In turn, theoretical efforts seeking to connect water thermodynamics and mobility should explicitly consider effects associated with both translational and orientational degrees of freedom. The strong correlations between orientational hydration water entropies and diffusivities found here lend further support to this idea, and also re-emphasize entropy-dynamic relationships that have been well studied in bulk fluids.^{63,64} Thus, this work represents a first step in not only the discovery of materials with surface patterns that modify water dynamics, but also exposes fundamental connections between local mobility, structure, and entropy of liquid water. In particular, the inhomogeneous yet highly tunable model surfaces presented here, which exhibit pronounced spatial dependence of kinetic, thermodynamic, and structural water properties, provide an excellent testing ground for such theoretical developments.

References

- [1] Fayer MD. Dynamics of Water Interacting with Interfaces, Molecules, and Ions. *Accounts of Chemical Research*. 2012; 45(1): 3–14. arXiv: NIHMS150003 ISBN: 0001-4842.
- [2] Siboulet B, Molina J, Coasne B, Turq P, Dufreche JF. Water self-diffusion at the surface of silica glasses: effect of hydrophilic to hydrophobic transition. *Molecular Physics*. 2013; 111(22-23): 3410–3417. Publisher: Taylor & Francis.
- [3] Siboulet B, Coasne B, Dufreche JF, Turq P. Hydrophobic transition in porous amorphous silica. *Journal of Physical Chemistry B*. 2011; 115(24): 7881–7886. ISBN: 1520-6106.
- [4] Schrader AM, Monroe JI, Sheil R, et al. Surface chemical heterogeneity modulates silica surface hydration. *Proceedings of the National Academy of Sciences*. 2018: 201722263.
- [5] Barnes R, Sun S, Fichou Y, Dahlquist FW, Heyden M, Han S. Spatially Heterogeneous Surface Water Diffusivity around Structured Protein Surfaces at Equilibrium. *Journal of the American Chemical Society*. 2017; 139(49): 17890–17901.
- [6] Wen G, Guo Z, Liu W. Biomimetic polymeric superhydrophobic surfaces and nanostructures: from fabrication to applications. *Nanoscale*. 2017; 9(10): 3338–3366. Publisher: Royal Society of Chemistry ISBN: 2040-3372 (Electronic) 2040-3364 (Linking).
- [7] Mishra H, Schrader AM, Lee DW, et al. Time-Dependent Wetting Behavior of PDMS Surfaces with Bioinspired, Hierarchical Structures. *ACS Applied Materials & Interfaces*. 2016; 8(12): 8168–8174. ISBN: 1944-8244.
- [8] Giovambattista N, Debenedetti PG, Rossky PJ. Hydration behavior under confinement by nanoscale surfaces with patterned hydrophobicity and hydrophilicity. *Journal of Physical Chemistry C*. 2007; 111(3): 1323–1332. ISBN: 1932-7447.
- [9] Acharya H, Vembanur S, Jamadagni SN, Garde S. Mapping hydrophobicity at the nanoscale: Applications to heterogeneous surfaces and proteins. *Faraday discussions*. 2010; 146: 353–365. ISBN: 1359-6640\1364-5498.
- [10] Hua L, Zangi R, Berne BJ. Hydrophobic Interactions and Dewetting between Plates with Hydrophobic and Hydrophilic Domains. *The Journal of Physical Chemistry C*. 2009; 113(13): 5244–5253.
- [11] Xi E, Venkateshwaran V, Li L, Rego N, Patel AJ, Garde S. Hydrophobicity of proteins and nanostructured solutes is governed by topographical and chemical

- context. *Proceedings of the National Academy of Sciences*. 2017; 114(51): 13345–13350. ISBN: 0027-8424, 1091-6490.
- [12] Wang J, Bratko D, Luzar A. Probing surface tension additivity on chemically heterogeneous surfaces by a molecular approach. *Proceedings of the National Academy of Sciences*. 2011; 108(16): 6374–6379.
- [13] Russo D, Murarka RK, Copley JRD, Head-Gordon T. Molecular View of Water Dynamics near Model Peptides. *The Journal of Physical Chemistry B*. 2005; 109(26): 12966–12975. ISBN: 1520-6106 (Print)\r1520-5207 (Linking).
- [14] Ball P. Water as an Active Constituent in Cell Biology. *Chemical Reviews*. 2008; 108(1): 74–108. arXiv: NIHMS150003 ISBN: 0009-2665.
- [15] Biswas R, Bagchi B. Anomalous water dynamics at surfaces and interfaces: synergistic effects of confinement and surface interactions. *Journal of Physics: Condensed Matter*. 2018; 30(1): 013001. Publisher: IOP Publishing.
- [16] Laage D, Elsaesser T, Hynes JT. Water Dynamics in the Hydration Shells of Biomolecules. *Chemical Reviews*. 2017; 117(16): 10694–10725.
- [17] Ball P. Water is an active matrix of life for cell and molecular biology. *Proceedings of the National Academy of Sciences*. 2017; 114(51): 13327–13335. ISBN: 0027-8424, 1091-6490.
- [18] Hua L, Huang X, Liu P, Zhou R, Berne BJ. Nanoscale Dewetting Transition in Protein Complex Folding. *The Journal of Physical Chemistry B*. 2007; 111(30): 9069–9077. ISBN: 1520-6106 (Print)\n1520-5207 (Linking).
- [19] Willard AP, Chandler D. The role of solvent fluctuations in hydrophobic assembly. *Journal of Physical Chemistry B*. 2008; 112(19): 6187–6192. arXiv: 0709.1133 ISBN: 1520-6106.
- [20] Mukherjee S, Mondal S, Bagchi B. Distinguishing dynamical features of water inside protein hydration layer: Distribution reveals what is hidden behind the average. *The Journal of Chemical Physics*. 2017; 147(2): 024901.
- [21] Fogarty AC, Laage D. Water Dynamics in Protein Hydration Shells: The Molecular Origins of the Dynamical Perturbation. *The Journal of Physical Chemistry B*. 2014; 118(28): 7715–7729. ISBN: 1520-6106\r1520-5207.
- [22] Argyris D, Cole DR, Striolo A. Dynamic Behavior of Interfacial Water at the Silica Surface. *The Journal of Physical Chemistry C*. 2009; 113(45): 19591–19600. ISBN: 1932-7447.

- [23] Romero-Vargas Castrillón S, Giovambattista N, Aksay IA, Debenedetti PG. Effect of Surface Polarity on the Structure and Dynamics of Water in Nanoscale Confinement. *The Journal of Physical Chemistry B*. 2009; 113(5): 1438–1446.
- [24] Choudhury N. Effect of surface hydrophobicity on the dynamics of water at the nanoscale confinement: A molecular dynamics simulation study. *Chemical Physics*. 2013; 421: 68–76. Publisher: Elsevier B.V.
- [25] Fayer MD, Levinger NE. Analysis of Water in Confined Geometries and at Interfaces. *Annual Review of Analytical Chemistry*. 2010; 3(1): 89–107. ISBN: 1936-1327.
- [26] Bourg IC, Steefel CI. Molecular dynamics simulations of water structure and diffusion in silica nanopores. *Journal of Physical Chemistry C*. 2012; 116(21): 11556–11564. ISBN: 1932-7447.
- [27] Karzar Jeddi M, Romero-Vargas Castrillón S. Dynamics of Water Monolayers Confined by Chemically Heterogeneous Surfaces: Observation of Surface-Induced Anisotropic Diffusion. *The Journal of Physical Chemistry B*. 2017; 121(41): 9666–9675.
- [28] Emami FS, Puddu V, Berry RJ, et al. Force field and a surface model database for silica to simulate interfacial properties in atomic resolution. *Chemistry of Materials*. 2014; 26(8): 2647–2658. ISBN: 0897-4756.
- [29] Horn HW, Swope WC, Pitera JW, et al. Development of an improved four-site water model for biomolecular simulations: TIP4p-Ew. *Journal of Chemical Physics*. 2004; 120(20): 9665–9678. ISBN: 0021-9606; 0021-9606.
- [30] Swails J, Hernandez C, Mobley DL, Nguyen H, Wang LP, Janowski P. ParmEd: Cross-Program Parameter and Topology File Editor and Molecular Mechanical Simulator Engine.
- [31] Godawat R, Jamadagni SN, Garde S. Characterizing hydrophobicity of interfaces by using cavity formation, solute binding, and water correlations. *Proceedings of the National Academy of Sciences*. 2009; 106(36): 15119–15124. ISBN: 0027-8424.
- [32] Jamadagni SN, Godawat R, Garde S. How surface wettability affects the binding, folding, and dynamics of hydrophobic polymers at interfaces. *Langmuir*. 2009; 25(22): 13092–13099. ISBN: 0743-7463 (Print).
- [33] Levine ZA, Rapp MV, Wei W, et al. Surface force measurements and simulations of mussel-derived peptide adhesives on wet organic surfaces. *Proceedings of the National Academy of Sciences*. 2016: 201603065. ISBN: 1603065113.
- [34] Zerze GH, Mullen RG, Levine ZA, Shea JE, Mittal J. To What Extent Does Surface Hydrophobicity Dictate Peptide Folding and Stability near Surfaces? *Langmuir*. 2015; 31(44): 12223–12230.

- [35] Mondello M, Grest GS, Webb EB, Peczak P. Dynamics of n-alkanes: Comparison to Rouse model. *Journal of Chemical Physics*. 1998; 109(2): 798–805. arXiv: physics/9802022.
- [36] Cornell WD, Cieplak P, Bayly CI, et al. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *Journal of the American Chemical Society*. 1995; 117(19): 5179–5197. arXiv: z0024 ISBN: 0120261464.
- [37] Jorgensen WL, Madura JD, Swenson CJ. Optimized Intermolecular Potential Functions for Liquid Hydrocarbons. *Journal of the American Chemical Society*. 1984; 106(22): 6638–6646. ISBN: 0002-7863.
- [38] Abraham MJ, Murtola T, Schulz R, et al. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*. 2015; 1-2: 19–25. arXiv: 1503.05249v1 ISBN: 23527110.
- [39] Swope WC, Andersen HC, Berens PH, Wilson KR. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *The Journal of Chemical Physics*. 1982; 76(1): 637–649. arXiv: 1011.1669v3 ISBN: 00219606.
- [40] Hess B, Bekker H, Berendsen HJ, Fraaije JG. LINCS: A Linear Constraint Solver for molecular simulations. *Journal of Computational Chemistry*. 1997; 18(12): 1463–1472. ISBN: 0192-8651.
- [41] Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, Pedersen LG. A smooth particle mesh Ewald method. *The Journal of Chemical Physics*. 1995; 103(19): 8577–8593. ISBN: 0021-9606.
- [42] Berendsen HJC, Postma JPM, Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*. 1984; 81(8): 3684. ISBN: doi:10.1063/1.448118.
- [43] Nosé S. A molecular dynamics method for simulations in the canonical ensemble. *Molecular Physics*. 1984; 52(2): 255–268. ISBN: 0026-8976.
- [44] Hoover WG. Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A*. 1985; 31(3): 1695–1697. ISBN: 6318641610.
- [45] Martyna GJ, Klein ML, Tuckerman M. Nosé-Hoover chains: The canonical ensemble via continuous dynamics. *The Journal of Chemical Physics*. 1992; 97(4): 2635–2643. arXiv: 10.1063/1.463940 ISBN: 0021-9606.
- [46] Widom B. Some Topics in the Theory of Fluids. *The Journal of Chemical Physics*. 1963; 39(11): 2808–2812.

- [47] Willard AP, Chandler D. Instantaneous liquid interfaces. *The Journal of Physical Chemistry B*. 2010; 114(5): 1954–1958. arXiv: 0909.4761 ISBN: 1520-6106.
- [48] Lewiner T, Lopes H, Vieira AW, Tavares G. Efficient Implementation of Marching Cubes' Cases with Topological Guarantees. *Journal of Graphics Tools*. 2003; 8(2): 1–15.
- [49] Walt S, Schönberger JL, Nunez-Iglesias J, et al. scikit-image: image processing in Python. *PeerJ*. 2014; 2: e453. arXiv: 1407.6245 ISBN: 2167-9843.
- [50] Nguyen H, Roe DR, Swails J, Case DA. PYTRAJ: Interactive data analysis for molecular dynamics simulations. 2016.
- [51] Jamadagni SN, Godawat R, Garde S. Hydrophobicity of Proteins and Interfaces: Insights from Density Fluctuations. *Annual Review of Chemical and Biomolecular Engineering*. 2011; 2(1): 147–171. ISBN: 1947-5438\n978-0-8243-5202-8.
- [52] Willard AP, Chandler D. Coarse-grained modeling of the interface between water and heterogeneous surfaces. *Faraday Discuss.*. 2009; 141: 209–220. arXiv: 0804.1134 ISBN: 10.1039/B816684F.
- [53] Luzar A, Leung K. Dynamics of capillary evaporation. I. Effect of morphology of hydrophobic surfaces. *The Journal of Chemical Physics*. 2000; 113(14): 5836–5844.
- [54] Remsing RC, Weeks JD. Hydrophobicity Scaling of Aqueous Interfaces by an Electrostatic Mapping. *The journal of physical chemistry. B*. 2014. arXiv: 1502.05587v1.
- [55] Shin S, Willard AP. Characterizing Hydration Properties Based on the Orientational Structure of Interfacial Water Molecules. *Journal of Chemical Theory and Computation*. 2018; 14(2): 461–465.
- [56] Patel AJ, Garde S. Efficient method to characterize the context-dependent hydrophobicity of proteins. *Journal of Physical Chemistry B*. 2014; 118(6): 1564–1573.
- [57] Garde S, Hummer G, García AE, Paulaitis ME, Pratt LR. Origin of Entropy Convergence in Hydrophobic Hydration and Protein Folding. *Physical Review Letters*. 1996; 77(24): 4966–4968. arXiv: physics/9611013 ISBN: 1079-7114.
- [58] Mittal J, Truskett TM, Errington JR, Hummer G. Layering and position-dependent diffusive dynamics of confined fluids. *Physical Review Letters*. 2008; 100(14): 9–12. arXiv: 0801.0757 ISBN: 0031-9007.
- [59] Pratt LR. Molecular Theory of Hydrophobic Effects: "She is too mean to have her name repeated.*". *Annual Review of Physical Chemistry*. 2002; 53(1): 409–436.

- [60] Layfield JP, Troya D. Molecular Simulations of the Structure and Dynamics of Water Confined between Alkanethiol Self-Assembled Monolayer Plates. *The Journal of Physical Chemistry B*. 2011; 115(16): 4662–4670.
- [61] Stirnemann G, Rossky PJ, Hynes JT, Laage D. Water reorientation, hydrogen-bond dynamics and 2d-IR spectroscopy next to an extended hydrophobic surface. *Faraday Discussions*. 2010; 146: 263. ISBN: 1359-6640\|r1364-5498.
- [62] Stirnemann G, Castrillón SRV, Hynes JT, Rossky PJ, Debenedetti PG, Laage D. Non-monotonic dependence of water reorientation dynamics on surface hydrophilicity: competing effects of the hydration structure and hydrogen-bond strength. *Physical Chemistry Chemical Physics*. 2011; 13(44): 19911. ISBN: 1463-9076.
- [63] Rosenfeld Y. Relation between the transport coefficients and the internal entropy of simple systems. *Physical Review A*. 1977; 15(6): 2545–2549. ISBN: 1050-2947.
- [64] Nayar D, Chakravarty C. Water and water-like liquids: relationships between structure, entropy and mobility. *Physical Chemistry Chemical Physics*. 2013; 15(34): 14162.
- [65] Nguyen CN, Kurtzman Young T, Gilson MK. Grid inhomogeneous solvation theory: Hydration structure and thermodynamics of the miniature receptor cucurbit[7]uril. *The Journal of Chemical Physics*. 2012; 137(4): 044101. ISBN: 1089-7690 (Electronic)\\$\\\$n0021-9606 (Linking).
- [66] Ramsey S, Nguyen C, Salomon-Ferrer R, Walker RC, Gilson MK, Kurtzman T. Solvation thermodynamic mapping of molecular surfaces in AmberTools: GIST. *Journal of Computational Chemistry*. 2016; 37(21): 2029–2037.
- [67] Acharya H, Mozdierz NJ, Keblinski P, Garde S. How chemistry, nanoscale roughness, and the direction of heat flow affect thermal conductance of solid-water interfaces. *Industrial and Engineering Chemistry Research*. 2012; 51(4): 1767–1773. ISBN: 0888-5885.

Chapter 5

Chemically-patterned interfaces for the controlled affinity and selectivity of small-molecule solutes

5.1 Introduction

Meeting current and anticipated global demands for clean water is a tremendous immediate challenge.¹⁻³ While water is readily available in the form of seawater and is produced in large quantities as industrial waste-products, these water sources are not currently accessible for domestic or agricultural use without significant energy expenditures.¹² Water filtration membranes, both porous and reverse-osmosis, provide promising avenues towards purifying water at lower energy cost.^{3,4³⁴} Unfortunately, rational design of such membranes is significantly hampered by the lack of molecular-level insight and understanding into the function of these devices. This significantly reduces our ability to rapidly advance this technology, despite an immediate need for its improvement.

Membrane fouling in particular represents a significant problem – adhesion of small

molecule or biological solutes at membrane interfaces leads to reduced water permeation and induces costs associated with removing foulants or replacement of the membrane.¹² A number of strategies to reduce fouling exist, mainly focusing on chemical or geometric modification of the membrane-water interface.² The effects of these types of modifications are often difficult to disentangle as the chemistry and nanoscale structure of an interface are intimately connected. For example, the chemistry of a polymer brush determines surface interactions with foulants and water both through specific chemical interactions as well as through the polymer chain structure (upright, lying flat, etc.), which leads to variations in surface roughness, flexibility, and nanoscale chemical patterning. In this work, we focus on model yet realistic systems in which we assess the impact on solute surface affinity by finely tune the chemistry and chemical patterning of interfaces without modifying the geometry of the surface. Such a fundamental study is of interest not only to protein fouling, but also within the fields of chromatography and surface catalysis.

The pioneering work of Whitesides and coworkers⁵ developed preliminary rules for modifying the chemistry of materials to resist fouling – interfaces should be hydrophilic, net neutral, and avoid hydrogen bond donating groups. These rules, however, only apply generally to surface chemistry, with the effect this chemistry has on surface roughness, flexibility, and nanoscale chemical patterning uncontrolled in the original studies. Later studies show that indeed the density, and hence flexibility and molecular structure, of molecules attached to an interface, such as polymer brushes, also plays a large role in determining fouling resistance.⁶⁷⁸⁹ Here we examine chemistry independently of interface geometry by employing molecular dynamics (MD) simulations of well-studied¹⁰¹¹¹²¹³¹⁴ models of alkanethiol self-assembled monolayer (SAM) systems with either hydroxyl or methyl headgroups. Pure surfaces composed of each headgroup are simulated along with different patterns of the two headgroups at a fixed 1:1 ratio in order to also examine the effect of chemical patterning. In all cases, the surfaces studied are flat surfaces, avoiding

effects due to local curvature, which on its own is known to impact surface hydrophobicity and solute binding.¹⁵¹⁶

Chemical heterogeneity of interfaces may also have a large impact on fouling propensity but has been investigated much less thoroughly. Ostuni, et al.¹⁷ examined how the size of hydrophobic head groups on self-assembled monolayer (SAM) chains affected the binding of a number of proteins. While the head group size and fraction of overall SAM chains were varied, the exact arrangement of the various head-groups on the surface was unknown. In a more recent study, Penna et al.¹⁸ demonstrated that protein adsorption was modified by specific patterns of chemical headgroups, though due to the size and complexity of the system only a very limited set of patterns was investigated. Another study considered the effect on solute affinity of four different hydroxyl group patterns on a graphene interface.¹⁹ In this case, the shape of the solute and its hydrogen bonding ability played a large role in the ability of hydroxyl groups to enhance or reduce affinity. It is well-established theoretically that surface patterning of chemical groups may induce variations in both water thermodynamic²⁰²¹ and dynamic¹³ water properties, but the general effects on affinity are not as clear.


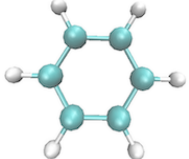
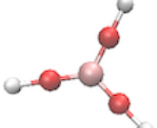
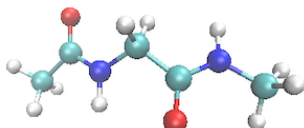
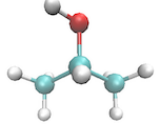


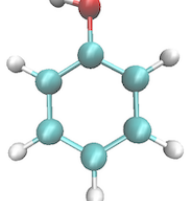
In contrast to many previous studies of fouling, we focus here on a diverse set of simple small-molecule solutes rather than proteins. Molecular simulations exploring the affinities of peptides and proteins for SAM,¹¹¹⁸²²²³²⁴²⁵²⁶²⁷²⁸²⁹ mineral,³⁰³¹ and graphene³²¹² surfaces, to name only a few, have been investigated. In many of these studies, the surface chemistry is well-controlled and varied independently of geometry, yet the fundamental driving forces for adsorption are obscured by the innate complexity of large biomolecules such as proteins. For a set of simpler solutes of varied chemistry, affinities for graphene interfaces have been determined and used to identify thermodynamic trends driving solute binding,³³³⁴ as well as to develop predictive models for affinity based on solute molecular descriptors.³⁵ When considering graphenic interfaces, however, the set of solutes studied is highly biased

towards aromatic compounds, which exhibit strong affinity for these surfaces.³⁶ On the other end of the spectrum are studies exploring surface-solute interactions, primarily hydrophobicity, on a very fundamental level, making use of model solutes with highly simplified geometries and intermolecular interactions.¹⁴³⁷³⁸ These studies provide a rich theory that is unfortunately far-removed from its practical application to protein affinity for interfaces. The solutes shown in Table 1 represent a diverse and realistic chemical set that can begin to bridge this gap. Simplicity is still required, however, to rigorously test and connect with fundamental theories that involve water behavior and structuring.³⁹⁴⁰ This even restricts the set of solutes considered to be neutral, as ions have been shown to exhibit highly complicated, and hotly debated, behavior at interfaces.⁴¹ Small, neutral solutes are also of practical interest to the improvement of reverse osmosis membranes, as such foulants tend to penetrate these membranes along with water.³⁴

To efficiently determine the affinity of the solutes in Table 1 for a variety of interfaces, we implement expanded ensemble simulations and alchemical free energy calculation techniques.⁴³⁴⁴⁴⁵ In brief, we calculate the free energy to transfer a solute from an ideal gas state to a solvated state, both in bulk solution and at an interface. These separate calculations of the solvation free energy, ΔG_{solv} , in different environments relative to the interface can then be directly related to surface affinity, or the free energy of binding

$$\Delta G_{bind} = \Delta G_{solv}^{surf} - \Delta G_{solv}^{bulk} \quad (5.1)$$

This is due to the fact that the free energy for the solute being fully decoupled from the system, in the ideal gas phase, is identical in the states both near and far from the interface. This approach also allows for the facile manipulation of solute surface affinity through patterning of interfacial chemical functional groups: since the surface chemistry does not affect bulk solvation behavior, minimization or maximization of ΔG_{solv}^{surf} will also

Solute	Structure	$\Delta G_{sol}^{CH_3}$	ΔG_{sol}^{OH}	ΔG_{sol}^{patch}	ΔG_{sol}^{spread}	ΔG_{sol}^{bulk}	$\Delta G_{sol}^{Henry's}$
Ammonia		-8.36 ± 0.09	-7.63 ± 0.04	-7.92 ± 0.04	-7.67 ± 0.05	-7.50 ± 0.07	-7.3 (5.9×10^{-1})
Benzene		-5.60 ± 0.07	-2.59 ± 0.06	-4.80 ± 0.09	-3.32 ± 0.09	-1.19 ± 0.11	-1.4 (1.7×10^{-3})
Boric Acid		-6.86 ± 0.06	-5.96 ± 0.10	-6.50 ± 0.06	-6.04 ± 0.11	-5.42 ± 0.04	-23.0 (3.8×10^6)*
Capped Glycine		-27.63 ± 0.08	-25.51 ± 0.13	-26.69 ± 0.09	-25.75 ± 0.18	-23.60 ± 0.15	N/A
Isopropanol		-14.33 ± 0.04	-11.54 ± 0.12	-13.42 ± 0.05	-12.34 ± 0.06	-10.80 ± 0.16	-8.1 (1.3×10^0)
Methane		1.04 ± 0.05	3.32 ± 0.03	1.84 ± 0.03	2.82 ± 0.05	3.99 ± 0.07	3.36 (1.4×10^{-5})
Methanol		-12.44 ± 0.03	-10.92 ± 0.10	-11.74 ± 0.06	-11.46 ± 0.09	-10.43 ± 0.09	-8.5 (2.0×10^0)
Phenol		-16.69 ± 0.08	-13.30 ± 0.11	-15.94 ± 0.10	-14.58 ± 0.10	-12.09 ± 0.12	-11.1 (2.8×10^1)

*This value for boric acid was not measured directly like all of the other constants. Instead it is an approximation based on the vapor pressure of the pure substance divided by the aqueous solubility.

Table 5.1: Solute structures and solvation free energies at interfaces and in bulk with comparison made in bulk to hydration free energies computed from Henry's constants. Solvation free energies apply to the free energy of solvation either in bulk or when the solute is near to methylated (CH_3) hydroxylated (OH) or 1:1 mixed SAM surfaces with either patchy or dispersed (spread) patterns of hydroxyl groups. Units are reported in $k_B T$ with uncertainties as the standard error in the mean based on 5 independent runs. Henry's Law constants are shown in parentheses in units of mol/m³Pa and were obtained from.⁴² For an accurate force field, we expect that the last two columns be equal. For a derivation relating Henry's Law constants to computed free energies of solvation, see Appendix F.

optimize ΔG_{bind} . To perform the optimization, we use a genetic algorithm to iteratively modify surface patterning, using the output of our expanded ensemble simulations and free energy calculations as objective functions (i.e., the genetic algorithm optimizes solute affinity). We previously used a similar computational evolution strategy to repattern surfaces in order to modify water mobility.¹³ Here we improve on this earlier approach by using MD simulations to train on-the-fly a predictive surrogate regression model that can then very quickly estimate a given pattern’s affinity for a particular solute, allowing the genetic algorithm to more expansively explore the design space. This greatly reduces the computational time required for the algorithm to converge to optimal functional group patterns. Finer details of the calculations and optimization procedures are provided in Methods.

Once we have explored pure SAM surfaces with methyl and hydroxyl headgroups, as well as patterns that optimize solute interfacial affinity, we then examine the role of water structure in driving solute-surface interactions. We correlate affinity to solvation shell water structure, finding that chemical patterns induce surprisingly similar behavior in affinities for all small-molecular solutes tested. This is despite notable differences in the solvation-shell structure surrounding these solutes, and the fact that they display a wide range of affinities and bulk solvation free energies. These fundamental connections between surface chemistry and patterning, solute chemistry, and water structure become clear through these extensive simulations of relatively simple systems. Thus, this work represents a preliminary step in connecting detailed theories based on the unique behavior of water to realistic and practical applications involving the design of interfacial materials.

5.2 Methods

5.2.1 Model systems

Self-assembled monolayer (SAM) systems and parameters associated with them have been previously described.^{11,13,46,47} Briefly, systems consist of alkanethiol chains consisting of 11 carbons and either a methyl or hydroxyl head-group. The only difference in the current set-up involves the freezing of SAM sulfur atoms rather than applying position restraints, minimally changing the system. All atoms except for the sulfur move freely in the simulations. We model non-terminal carbon and sulfur atoms within the chains, as well as their associated hydrogens, via united atom representations⁴⁸ while head-groups are modelled with full atomistic resolution, including hydrogens.⁴⁹

Periodic boundary conditions apply in all directions so that SAM interfaces are replicated with approximately 3.5 nm of water between periodic copies, resulting in box dimensions of 2.9820 x 3.4432 x 6.0000 nm³. For solutes in bulk water, simulation boxes are 3.5 x 3.5 x 3.5 nm³. TIP4P/Ew water⁵⁰ models the solvent in all simulations and is added to simulation boxes via the “solvate” command in GROMACS 2016.1.⁵¹ For capped glycine, AMBER ff14SB represents all intramolecular, electrostatic, and Lennard-Jones (LJ) interactions, with no charge reductions applied based on empirical adjustments from MD simulations in explicit water.⁵² Boric acid intramolecular parameters come from the work of Otkidach and Pletnev,⁵³ and include a “bond repulsion” term that simultaneously enforces planarity and angular distances. LJ parameters for boron are provided by the DREIDING force field,⁵⁴ as implemented in simulations of ionic liquids containing boron,⁵⁵ while parameters for the oxygen and hydrogen are taken from GAFF2.⁵⁶ Unintentionally, the LJ sigma parameter for the boron atom was divided by two. Charges for boric acid are determined using the ANTECHAMBER package of AmberTools18⁵⁶ through RESP⁵⁷ fitting of electron densities computed by GAUSSIAN16.⁵⁸ All other solutes are modelled

via GAFF2 parameters and AM1-BCC charges,⁵⁹ with parameter files generated through ANTECHAMBER.

We initially noticed that the solvation free energies of all alcohols studied were systematically larger (less favorable or less negative) than experiment and apply the empirical adjustment to partial charges suggested by Fennel, Wymer, and Mobley.⁶⁰ We do not apply the adjustments to the LJ interactions proposed by these authors, as GAFF2 assigns LJ parameters differently than the GAFF force field used by the authors. However, we do scale the oxygen and hydrogen partial charges by 1.20905 and redistribute charge as described by Fennel, Wymer, and Mobley over other solute atoms to maintain neutrality.⁶⁰ This procedure results in free energies of solvation more in line with experiment, but that are in fact too favorable (more negative than experiment). This is expected for a fairly rigid molecule with fixed partial charges and no polarizability that has a polarity closer to that expected in the liquid phase than the gas phase. For this scenario, there is an unfavorable free energy of polarizing the gas-phase molecule that we do not include here, but which would further bring the alcohol solvation free energies more in line with experiment.⁶⁰

In simulations to determine solvation free energies in octanol, we construct boxes of roughly $3.5 \times 3.5 \times 3.5 \text{ nm}^3$ to hold solutes that are filled with octanol solvent via the ‘insert-molecules’ command of GROMACS. Octanol molecules are modeled with the united-atom TRAPPE force field,⁶¹ with all bonds constrained. In these systems, we use the same cutoff as described above without dispersion corrections, resulting in bulk densities that are $\sim 1.5\%$ lower than literature values at the same temperature of 298.15 K and pressure of 1 bar.⁶¹

5.2.2 Molecular dynamics simulations and free energy calculations

All molecular dynamics simulations utilize GPU-accelerated code in OpenMM,⁶² with all waters constrained with SETTLE,⁶³ non-water bonds involving hydrogens constrained by SHAKE⁶⁴ and equations of motion propagated via a 2 fs timestep with Langevin dynamics. A Monte-Carlo isotropic barostat maintains pressure in bulk systems while in SAM systems an anisotropic barostat holds pressure constant by only changing the z-dimension (normal to the interface). Particle Mesh Ewald (PME) handles long-range electrostatics while a hard real-space cutoff of 1.2 nm is applied to both Coulombic and LJ interactions with no dispersion corrections applied when computing the energy.

We compute potentials of mean force (PMFs) using umbrella sampling⁶⁵ with solute heavy-atom centroids restrained with a harmonic potential at increasing distances from the interface. Starting configurations at each umbrella distance are obtained by pulling the solute at a fixed rate towards the surface in the NVT ensemble. We equilibrate each umbrella in the NVT ensemble for 500 ps and for 1 ns in the NPT ensemble. Production umbrella simulations are run for 10 ns for each umbrella.

To compute solvation free energies, we employ expanded ensemble techniques.^{43–45} This involves Monte-Carlo moves between solute-system interaction states every 250 timesteps (or 0.5 ps). The most extreme interaction states involve the solute fully interacting with the rest of the system (state 0) and the solute maintaining only intramolecular interactions and lacking all interactions with the rest of the system (state 1). Tracking the free energy from state 0 to state 1 then represents the free energy to solvate an ideal gas solute molecule in the same volume occupied by the system. Intermediate electrostatic states are achieved by linearly scaling charges used by the full PME scheme in OpenMM from 1.0 to 0.0 while simultaneously turning on charges at the same rate for atoms within the

solute molecule interacting by a purely real-space coulombic potential. LJ interactions between the solute and the rest of the system are turned off using soft-core interactions⁶⁶ while maintaining intramolecular LJ interactions.

In order to efficiently sample all interaction states, we start all expanded ensemble simulations with 2 ns of Self-Adjusted Mixture Sampling,⁶⁷ which is used to adjust biasing weights until they are approximately equal to the free energies between interaction states so that all states are sampled roughly evenly. At the end of this procedure, the final weights to be used in the production part of the simulation (with fixed biases) are determined via the pymbar⁶⁸ implementation of MBAR⁶⁹ using data from the initial 2 ns. Production runs of 10 ns in bulk solution and 5 ns at interfaces are used to collect potential energies at all interaction states every 1 ps.

At interfaces, the solute must be restrained to prevent it from drifting away from the interface. This is done by harmonically restraining the centroid of all solutes via a flat-bottom harmonic potential that is zero between 1.40 and 1.75 nm from the SAM sulfur atoms and harmonic thereafter. While in state 1 (the fully non-interacting state) the solute can easily escape from free energy minima that exist in state 0 as a function of lateral position on the interface. However, it is still difficult to ensure that the solute samples the entire surface, and hence all surface patterns, which is required for accurate free energies of solvation. To assist this sampling, four simulations of the solute on the surface are conducted, with the first heavy atom of the solute harmonically restrained to a single surface quadrant via two-dimensional flat-bottom restraints. The resulting 20 ns of production simulation data are reweighted through MBAR to remove the effect of the lateral restraints and provide an overall free energy of solvation. In comparing solvation free energies at interfaces and in bulk solution it is implied that a restraint to the same volume in solution is applied as when the solute is near the interface. However, it is not required that restraints are explicitly accounted for in computing the bulk solvation

free energy. This is because the free energy to turn on this restraint is the same in both the coupled and decoupled (ideal gas) states when in bulk, leading to no change in the solvation free energy.

5.2.3 Genetic algorithm optimization

The genetic algorithm optimization procedure has been described previously,¹³ though we have modified it here to incorporate a machine-learned model to more quickly predict the solvation free energy instead of utilizing a full molecular simulation. Briefly, the genetic algorithm treats each surface pattern as an “individual” with a specific level of “fitness.” The goal of the algorithm is to increase the fitness of the population by combining and randomly adjusting (“mutating”) different surface patterns. This can be done to optimize arbitrary interfacial properties by changing the fitness metric associated with each surface that is provided to the algorithm, which here is the solvation free energy of a solute near an interface. The genetic algorithm may then be used to explore surface patterns that minimize or maximize a given property by defining fitness to be either low or high values of the associated metric.

The full process is shown schematically in Figure 5.1 and involves first randomly generating surface patterns and evaluating solvation free energies with MD for 6 generations. To further explore pattern space in order to later train the model more accurately, we next perform 5 generations of minimization and 5 generations of maximization each starting from the randomly generated library. The resulting 128 surface patterns and MD-computed solvation free energies are then used to train a linear-regression model based on two-dimensional radial distribution functions (RDFs) for CH₃-CH₃ and OH-OH chains based on the locations of their frozen SU atoms. This does not reflect the true 2D head-group RDFs when in contact with water but approximates this quantity and may be

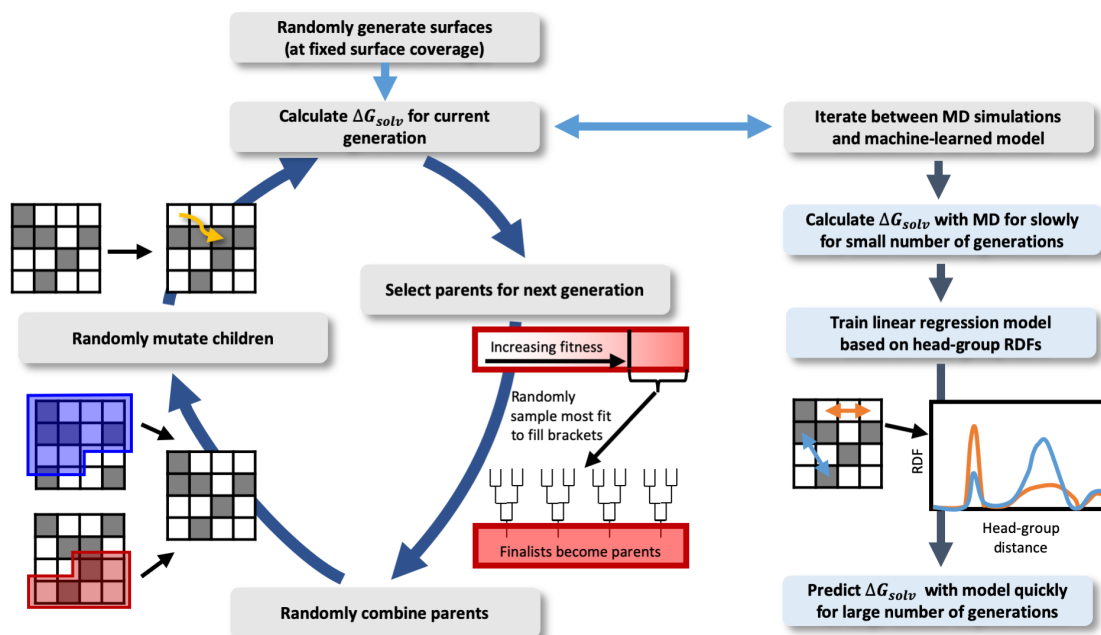


Figure 5.1: Schematic demonstrating the functioning of the genetic algorithm coupled to iterative training of a linear regression model.

computed without solvating the surface or performing any simulation. Once the model is preliminarily trained, it is used instead of MD simulations to estimate solvation free energies to pass as fitness metrics to the genetic algorithm. For both minimization and maximization, 5 independent optimizations of 30 generations are performed each starting from the 128 surfaces used to train the model, with a final 30 generation optimization performed combining all surfaces sampled previously. Another 5 generations of MD-based minimization and maximization are performed including all surfaces generated up to that point in the library of surfaces from which the genetic algorithm draws, regardless of whether the surface fitness is evaluated by MD or the regression model. The regression model is retrained by adding these additional 80 surface patterns to the original MD library and final minimizations and maximizations using the new model are each performed for 60 generations.

5.2.4 Computation of density fluctuations and water structure near interfaces and solutes

Density fluctuations are assessed by examining the probability of finding a given number of water oxygens within a probe volume.²¹ The value of this distribution at zero waters is related directly to the chemical potential of inserting a hard-sphere probe into the fluid.⁷⁰ In this study, we histogram numbers of waters in both spherical probe volumes (see Figure G.4) as well as in volumes based on the effective hard-sphere molecular volume of a solute (see Figure 5.5). Spherical probes are defined as spheres with radii of 0.33 nm such that any heavy-atom center falling within this volume is defined as an overlap. To construct hard-sphere solutes, we compute pairwise effective hard-sphere distances of closest approach for all atom types in the solute to all non-solute atom types, including hydrogens. Pairwise effective hard-sphere radii (or distances of closest approach) are computed as^{71,72}

$$R_{HS,ij} = \int_0^{\text{inf}} \left(1 - e^{-\beta U_{WCA,ij}(r)} \right) dr \quad (5.2)$$

where $U_{WCA,ij}$ is the repulsive WCA⁷³ part of the Lennard-Jones interaction between the two atoms of types i and j . Solute and non-solute atoms are checked for overlaps in a pairwise fashion with atom centers considered to be overlapping if the distance between them is less than $R_{HS,ij}$.

We mainly examine water structure via three-body angle distributions (see Chapter 3) within the first hydration shells of solutes and interfaces. This quantity is defined as the angle between vectors connecting two neighboring water oxygens to a central water oxygen that is within the region of interest. All water oxygens in the system are candidates for being neighbors to water oxygens with the solute or interface hydration shells, with the neighbor cutoff set to 0.332 nm, the first radial distribution function minimum for bulk TIP4P-Ew water at standard temperature and pressure. To assess perturbations from

bulk structure, distributions based on histograms of three-body angles are normalized to sum to 1 and the similarly normalized bulk distribution is subtracted.

5.3 Results and Discussion

5.3.1 Diverse solute chemistries show affinity for both polar and nonpolar surfaces

A primary goal is to understand the influence of surface chemistry on the surface affinity of highly varied solutes. It is well-known that nonpolar, hydrophobic solutes adhere to hydrophobic interfaces, but what level of surface polarity or overall hydrophilicity is required to prevent such adhesion? Conversely, a more polar surface will more strongly attract water and may resist adsorption of nonpolar solutes, but the effect on hydrophilic solutes is less clear. Depending on the solute's polarity or charge, it may adhere more strongly to the polar surface. While we do include solutes of moderate polarity here (Table 1), consideration of highly polar and even charged solutes will be left for future work.

It is first useful to classify the solute molecules by their hydrophobicity/hydrophilicity. One way to define this is in terms of the solvation free energy ΔG_{solv} , or excess chemical potential at infinite dilution in bulk solution.^{40,74} If this quantity is positive, the molecule prefers the ideal gas phase over water and is hydrophobic, while if this is negative, solvation is favorable and the molecule is hydrophilic. All of the solutes studied except for methane exhibit negative solvation free energies (Table 1). For benzene, this is not unexpected once the relatively high spatial density of Lennard-Jones sites in the molecule is considered, which provide attractive interactions with water.

Figure 5.2 demonstrates that solute hydrophilicity, as measured by bulk solvation

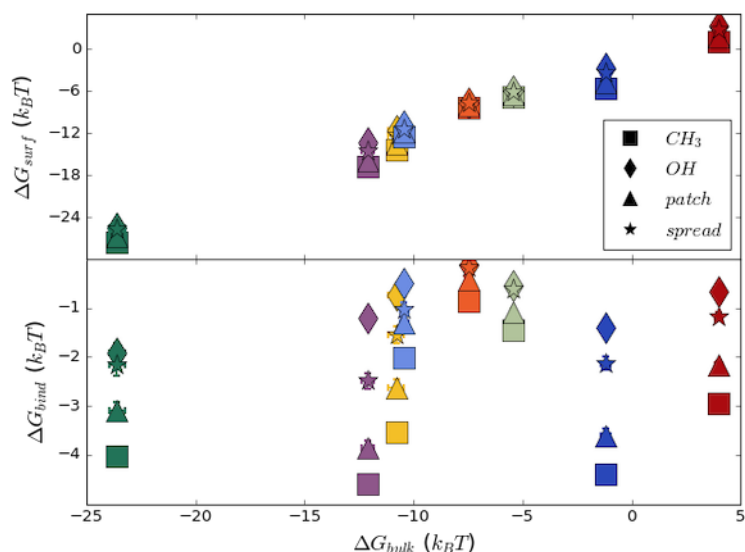


Figure 5.2: While the solvation free energies at all interfaces studied are correlated with bulk solvation free energies for all solutes, they are not correlated with binding affinities. Thus, solute hydrophobicity/hydrophilicity as measured by solvation free energies is not a good metric for solute binding affinity.

free energy, is not correlated with surface affinity. It might seem reasonable that more hydrophobic molecules more tightly adhere to hydrophobic interfaces, but this is not the case here. This is reinforced in Figure 5.3, which shows potentials of mean force (PMFs) as a function of distance of the solute heavy-atom center of geometry from the interface, with the points far from the interface matched to the bulk solvation free energy. Interestingly, Figure 5.3 also demonstrates that all solutes show affinity for both the methylated and hydroxylated interfaces, though the latter affinity is nearly negligible in the cases of ammonia and boric acid. Similar affinity for hydroxylated SAMs has been observed for peptides, both in simulation and experimentally,^{11,12,25} but is still surprising given that the fully hydroxylated surfaces have contact angles of essentially 0° .¹³ While methane is the only solute with an unfavorable (positive) solvation free energy, all solutes, including those that are considered hydrophilic and that form favorable hydrogen bonds with water, have lower free energy compared to bulk when near the fully methylated,

presumably hydrophobic, interface.

An alternative metric for hydrophobicity also fails to correlate with solute affinities for the methylated interface (Figure G.1). Specifically, the logarithm of the octanol-water partition coefficient, which is equal to the difference in chemical potentials between a solute in water and octanol phases, is commonly used to predict affinities in reversed-phase chromatography.⁷⁵ Here, however, we find no correlation between binding affinities and octanol-water transfer free energies, similar to the results observed above for aqueous solvation free energies (Figure G.1). In the case of octanol-water transfer free energies, there is also no correlation to raw interfacial solvation free energies, which appears strongly for bulk aqueous solvation in the top panel of Figure 5.2. Lack of correlations of surface affinities to either common definition of hydrophobicity (aqueous solvation or octanol-water transfer) suggest that binding affinity at interfaces is driven by distinct physics and cannot necessarily be predicted based on common concepts of hydrophobicity/hydrophilicity.

To understand these counterintuitive results, we break the solvation free energy into two separate contributions corresponding to the process of turning on LJ interactions alone ΔG_{LJ} , followed by that of charging of the solute atoms ΔG_Q ,

$$\Delta G_{solv} = \Delta G_{LJ} + \Delta G_Q \quad (5.3)$$

Values for these quantities both in bulk and at methylated and hydroxylated interfaces are shown in Figure G.2. Considering Equation 5.1, we can examine how these contributions change upon binding

$$\Delta G_{bind} = (\Delta G_{LJ}^{surf} - \Delta G_{LJ}^{bulk}) + (\Delta G_Q^{surf} - \Delta G_Q^{bulk}) = \Delta \Delta G_{LJ} + \Delta \Delta G_Q \quad (5.4)$$

This analysis does not make direct connections to conventional physical quantities,

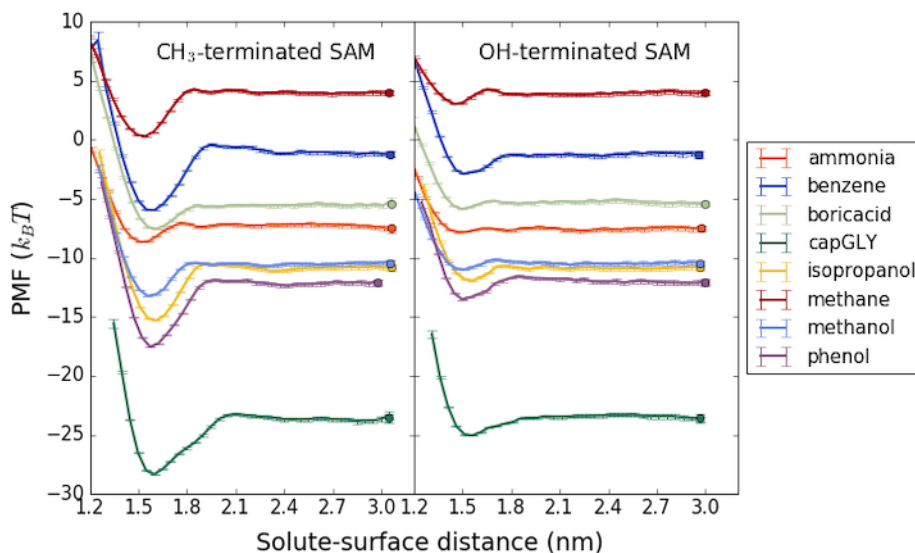


Figure 5.3: Potentials of mean force are shown for all solutes studied at methylated (left) and hydroxylated (right) interfaces. The distance from the interface is calculated from the solute heavy-atom centroid to the fixed sulfur atoms of the SAM chains. All PMF values are relative to the bulk solvation free energies of solutes, which are shown as points at the furthest distances from the interface sampled. Error bars are those reported by pymbar.⁶⁸

such as specific energetic or entropic terms connected to solvent or solute degrees of freedom or interactions, but here it is useful for understanding the broader, more general contributions driving both solvation and adsorption. In Figure 5.4, $\Delta\Delta G_Q$ values are positive for all solutes, indicating that charging a solute at the interface results in less favorable electrostatic interactions than in bulk. On the other hand, all solutes experience a negative $\Delta\Delta G_{LJ}$, signifying a favorable LJ contribution to solvation at the interface as compared to in solution (Figure 5.4). With or without strong solute-water energetic interactions, all of these small solutes are expected to induce unfavorable entropic penalties upon solvation due to excluded volume effects.^{70,76} For all but methane, favorable electrostatic and LJ energies offset these penalties. Once in solution, however, these energetic interactions are not enough to prevent solutes from seeking an interfacial region with enhanced density fluctuations^{38,77,78} and hence more effective volume in which

the solute may reside without disrupting water structure. Intuitively, it might seem reasonable that alcohols have affinity for the hydrophobic interface, potentially orienting their hydroxyl group away from the surface to maintain hydrogen bonds with water while placing their hydrophobic portion close to the interface. However, ammonia and boric acid will both sacrifice hydrogen bonds upon moving to the methylated surface. This could be compensated by more favorable water-water hydrogen bonding upon solute adsorption, but it is clear that this is not the case from the positive values of $\Delta\Delta G_Q$.

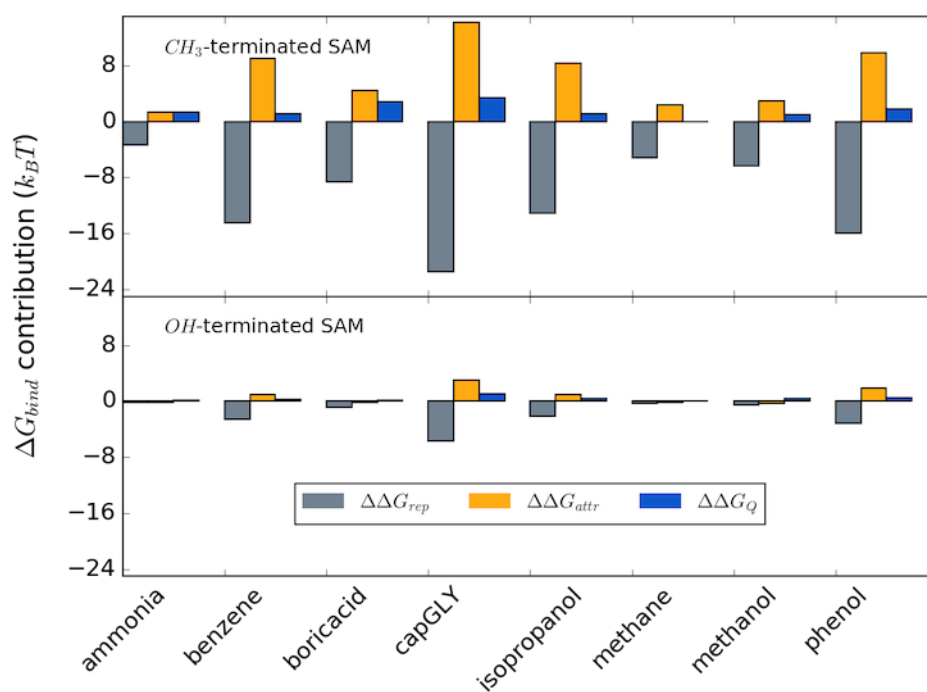


Figure 5.4: Contributions to binding free energies as described in the text and defined in Equations 5.3 and 5.4 are shown for the methylated (top) and hydroxylated (bottom) SAM surfaces. Summing the $\Delta\Delta G_{rep}$ and $\Delta\Delta G_{attr}$ terms yields the change in LJ interactions $\Delta\Delta G_{LJ}$. The repulsive component, which involves creating a cavity in which the solute may be inserted, is the predominant thermodynamic driving force for a solute's preference for the interface over bulk solution.

From the above discussion, it seems that affinity of all solutes for the methylated interface, regardless of solute hydrophilicity, may be driven by excluded volume effects. Such contributions may be estimated by considering a further decomposition of the free

energy to turn on LJ interactions during solvation

$$\Delta G_{LJ} = \Delta G_{rep} + \Delta G_{attr} \quad (5.5)$$

ΔG_{rep} represents the free energy to turn on only repulsive interactions to the free energy and is estimated by the free energy to solvate solutes composed of hard-spheres (HS) through repeated insertions according to Widom's method.⁷⁹ ΔG_{attr} is then the free energy to move from a hard-sphere solute to a LJ solute with soft repulsions and attractions. For a purely HS solute, the free energy of solvation $\beta\Delta G_{rep}$ may be computed based on the probability of finding no solvent (or surface) atoms overlapping with solute atoms, $P(0)$

$$\beta\Delta G_{rep} = -\ln P(0) \quad (5.6)$$

Figure G.3 shows probability distributions of the number of overlapping atoms for HS solutes both in bulk and at interfaces. As the solute size increases, distributions broaden and shift to larger numbers of overlapping atoms. Additionally, the presence of a hydrophobic interface further broadens this distribution, while the hydroxylated interface only induces a slight broadening of the distribution. For larger solutes, such as benzene, phenol, and capped glycine, $P(0)$ cannot be directly evaluated, but Gaussian fits to the probability distributions in Figure G.3 allow for this term to be estimated. For the bulk and hydrophilic cases, such Gaussian fits are justified based on prior observations,^{38,70,80} but result in underestimates of ΔG_{rep} at the methylated interface (Figure G.3). The latter result is unexpected based on previous literature, where super-Gaussian behavior in the tails of the distribution is observed.^{38,77,78} This discrepancy may be due to particularly poor sampling of numbers of overlapping atoms in the tails of the distribution, especially for larger solutes. For both methyl- and hydroxyl-terminated SAMs, $\Delta\Delta G_{rep}$ is highly

favorable when moving solutes from bulk to the methyl-terminated SAM interface (Figure 5.4). By considering the total $\Delta\Delta G_{LJ}$, it is even the case that $\Delta\Delta G_{attr}$ must be positive (unfavorable) for most solutes, indicating a loss of attractive solute-system LJ interactions at the interface compared to in bulk (Figure 5.4).

Extremely favorable $\Delta\Delta G_{rep}$ contributions to binding along with unfavorable $\Delta\Delta G_{attr}$ may be easily rationalized for the methylated interface, but the picture is far less clear at the hydroxylated surface. For a hydrophobic interface, water slightly withdraws from the interface, as demonstrated by water density profiles shown in Figure G.4. Hard-sphere excess chemical potential profiles shown in Figure G.5, which relate to density fluctuations via Equation 5.6, show strongly enhanced fluctuations near the methylated surface. Comparison with PMFs in Figure 5.3 shows that solutes tend to sit in surface locations with reduced water density, leading to a reduction in ΔG_{attr} compared to bulk, while ΔG_{rep} is reduced significantly compared to bulk due to the enhanced density fluctuations. While Figure 5.4 demonstrates similar trends in $\Delta\Delta G_{rep}$ and $\Delta\Delta G_{attr}$ for the hydroxylated interface, similar arguments rationalizing the signs of these terms do not hold. First, the density profile in Figure G.4 indicates that solute centers of geometry are actually separated from the hydroxylated SAM interface by a layer of water. Placing some solute atoms near the first density minimum may result in similar reductions of ΔG_{attr} as compared to the methylated interfaces, but based on a slightly different solute/water interfacial structuring. Figure G.5 demonstrates that solute geometric centers are located where density fluctuations are approximately equivalent to bulk, with atoms expected to experience both reduced and enhanced fluctuations. This is despite slightly enhanced density fluctuations at locations very close to the interface (Figure G.5), which is somewhat at odds with past studies demonstrating repulsion of purely repulsive particles from nearly identically parametrized hydroxylated SAMs in contact with a different water model.¹⁴ This raises doubt as to whether density fluctuations, which create additional free-volume

for the solute to occupy, is the predominant driving force for affinity. It may be that other contributions associated with water restructuring or partitioning back to the bulk may play a more prominent role at the hydroxylated interface.

5.3.2 Genetic algorithm optimization of affinity via surface patterning

Beyond the interplay of solute and surface chemistry on solute affinity, patterning of functional groups represents an additional surface feature that may be tuned. While it is well-established that interface geometry and flexibility also play a role in determining solute adhesion,^{6,7,18} we continue to focus on surface chemistry, including chemical patterning, in a controlled manner for surfaces of essentially identical flexibility and roughness. We start by using our genetic algorithm to optimize the affinity of the simplest solute studied, methane, by repatterning a SAM interface of hydrophobic (methyl) and hydrophilic (hydroxyl) head-groups at a 1:1 ratio of each type of SAM chain (Figure 5.5). For such a simple, small solute the solvation process will be particularly dominated by the magnitude of solvent density fluctuations,⁷⁰ as noted above. It has been well-established that clustering hydrophobic headgroups enhances density fluctuations, while distributing them across an interface reduces these.²¹ Such patterns have recently also been shown to respectively maximize and minimize water dynamics at an interface.¹³ Based on this past work, we expect the global minimum solvation free energy structure to be a patchy surface and the global maximum to be a surface with hydroxyls spread uniformly over the interface, as shown in Table 2.

Figure 5.5 shows the progress of the genetic algorithm optimization (see Methods) for methane, while Table 2 shows representations of the optimal surface patterns. From Table 2 it is clear that the optimization delivers the expected results – a surface with a single

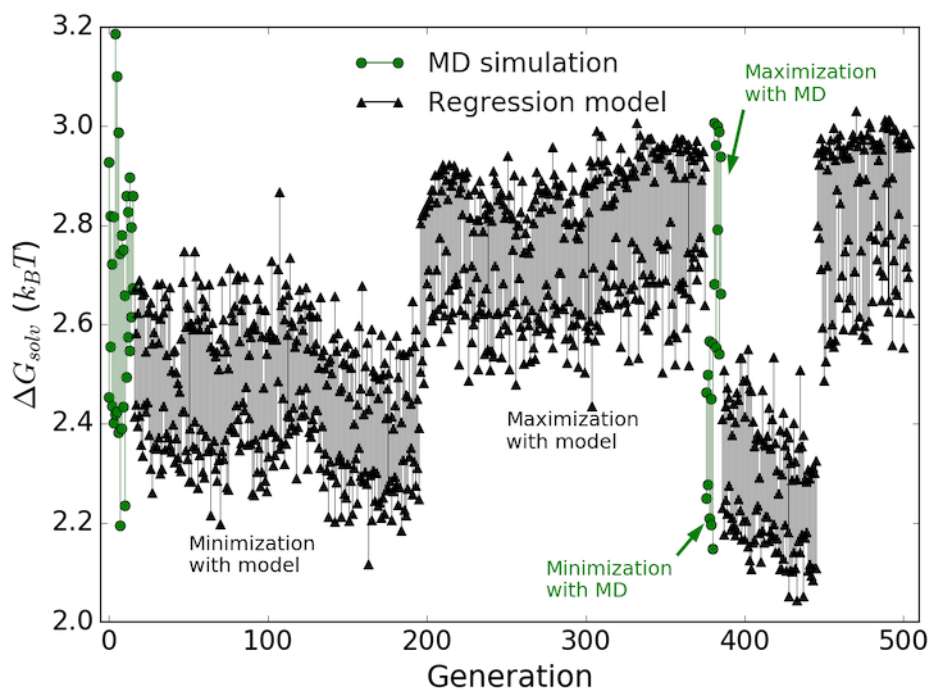


Figure 5.5: The progress of the genetic algorithm is shown for minimization (decreasing regions) and maximization (increasing regions) of the affinity of methane. As described in Methods, it is advantageous to perform optimization in both directions simultaneously. For each generation, or round of optimization, the minimum and maximum affinity surfaces are shown connected by a vertical line. Green circles and black triangles represent generations where MD and the linear regression model, respectively, were used to determine solvation free energies. The points for the linear regression model are based only its final version – that trained from results of all MD simulations, including the second set of green points at much later generations.

large patch of hydrophobic groups minimizes the solvation free energy (increases density fluctuations) and maximizes methane affinity, while a surface with evenly-distributed hydrophobic groups maximizes solvation free energy (decreases density fluctuations) and minimizes affinity. We also test “perfectly” patchy and dispersed surfaces shown in Table 2, revealing that the genetic algorithm comes very close to identifying the surface patterns expected to be optimal. For the maximum solvation free energy surface, the genetic algorithm effectively identifies the optimal pattern within the precision of the free energy calculations. For the minimum solvation free energy surface, however, it only comes within $\sim 0.30 k_B T$ of the anticipated global minimum, though the surface pattern appears visually to be very close to the assumed global optimum. Interestingly, the linear regression model also finds this surface similar to the global minimum, assigning it a nearly identical solvation free energy value. In the case of the global minimum, the model-assigned value is $0.19 k_B T$ too large while for the GA minimum the model prediction is $0.11 k_B T$ too low. This indicates a lack of training, and in turn sensitivity, of the model within the region close to the global minimum involving highly segregated headgroups. In contrast, the regression model uniformly over-predicts the solvation free energy for surfaces with dispersed hydrophilic groups, in the region containing maximal solvation free energy patterns. This is likely due to rare, large overpredictions of solvation free energies by MD simulations during the initial stages of the genetic algorithm, as discussed further below.

From Figure 5.5 it is clear that the probability of randomly identifying a minimum ΔG_{solv} surface, with clusters of hydroxyl or methyl groups, is very low. The first round of optimization using the linear regression model expands the search space, particularly in the direction of minimal solvation free energies to consider surfaces with clustered configurations, but does not reach globally optimal surface patterns. After further refinement of the linear regression model based on short optimizations with MD simulations

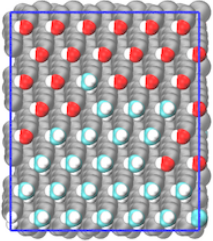
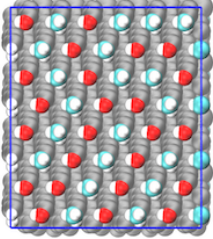
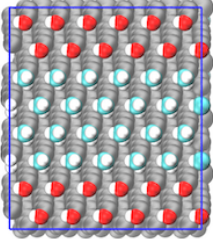
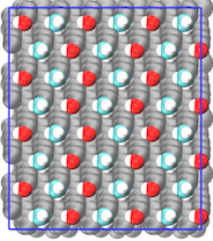
Surface	Surface image	ΔG_{solv} from MD simulation ($k_B T$)	ΔG_{solv} predicted by linear regression model ($k_B T$)
GA Model-predicted Minimum		2.15 ± 0.03	2.04
GA Model-predicted Maximum		2.85 ± 0.04	3.03
Global Minimum		1.84 ± 0.03	2.03
Global Maximum		2.82 ± 0.05	2.95

Table 5.2: Surface representations and free energies of solvation for methane, determined by MD simulations and linear regression models. The surfaces shown for the GA are those with minimum and maximum solvation free energies predicted by the linear regression model, not necessarily the overall optimal surfaces reported by the GA. Uncertainties are the standard error in the mean from 5 independent simulations.

used to evaluate ΔG_{solv} , which should now include surfaces with more clustering of surface groups, we see significantly faster approach of the GA towards the minimum bound. It is then clear that the use of the computationally efficient linear regression model is necessary for enhancing the speed of the calculations, as well as the space searched, but that iterative reassessment and training of the model is crucial.

It is also clear from Figure 5.5 that the effectiveness of the algorithm is limited by the statistical error in the MD simulation predictions of solvation free energies. Use of the linear regression model actually reduces the necessary precision of the MD calculations by stitching together multiple simulation results into a predictive model that is less sensitive to outliers than the underlying genetic algorithm. For instance, the maximum ΔG_{solv} value sampled during genetic algorithm optimization ($3.19 k_B T$) was by MD simulation and is unusually high for the surface pattern with which it is associated. When re-evaluating this surface with multiple MD simulations, we find that its average ΔG_{solv} value is only $2.59 k_B T$, making the high outlier sampled during the genetic algorithm run a statistical anomaly. If running the genetic algorithm with only MD predictions of ΔG_{solv} , such outliers would dominate the algorithm and limit its convergence. However, when training the linear regression model, such statistical outliers have only a small weight compared to more typical predictions, resulting in a reasonably accurate model (Table 2). By using the linear regression model, we actually reduce the influence of these outliers on the genetic algorithm, further enhancing its convergence.

Assuming density fluctuations are the primary driving force for all solutes to adhere to interfaces, we expect that similar patterns as for methane will maximize and minimize the interfacial ΔG_{solv} for the broader set of solutes. This seems to be the case based on affinities of solutes shown in Table 1 for perfectly “patchy” and “spread” patterns (the global minima and maxima in Table 2). As a more specific example, genetic algorithm optimization of the affinity of capped glycine via surface patterning is shown in Figure

G.6. Being the largest, most flexible solute makes this the most challenging case studied here. As such, the results are not as impressive as for methane, largely due to much higher uncertainty in the solvation free energy calculations ($\sim 0.25 k_B T$ compared to $\sim 0.1 k_B T$ for methane). This is despite using twice the simulation time for both convergence of weights and production expanded ensemble simulations in this optimization. Greater uncertainty makes it more difficult to train the regression model and for the genetic algorithm to clearly define patterns that minimize or maximize affinity. However, Table S1 demonstrates that the genetic algorithm does move towards similar surface patterns as discovered for methane. Further, the regression model trained for capped glycine predicts smaller and larger values for the perfectly patchy and spread interfaces than the minima and maxima observed during the genetic algorithm optimization (Table S1). Evaluating the solvation free energy of capped glycine on these perfectly patchy and spread surfaces with repeated MD simulations validates the regression model's accuracy (Table S1). The inability of the genetic algorithm to establish distinct differences in affinity then suggests that it has not been run for long enough. Alternatively, the algorithm could be made to identify these differences more rapidly through even more precise solvation free energy calculations, though this would necessitate significantly more computation time per generation.

5.3.3 Surface patterns modulate density fluctuations to adjust affinity and selectivity

Figure 5.6 demonstrates that affinities for interfaces are highly correlated to each other, further suggesting that affinity for all interfaces is driven by similar physics, namely density fluctuations. The only disruption to the correlations in Figure 5.5 is the affinity of the capped glycine for the hydroxylated interface. Compared to other solutes of similar overall

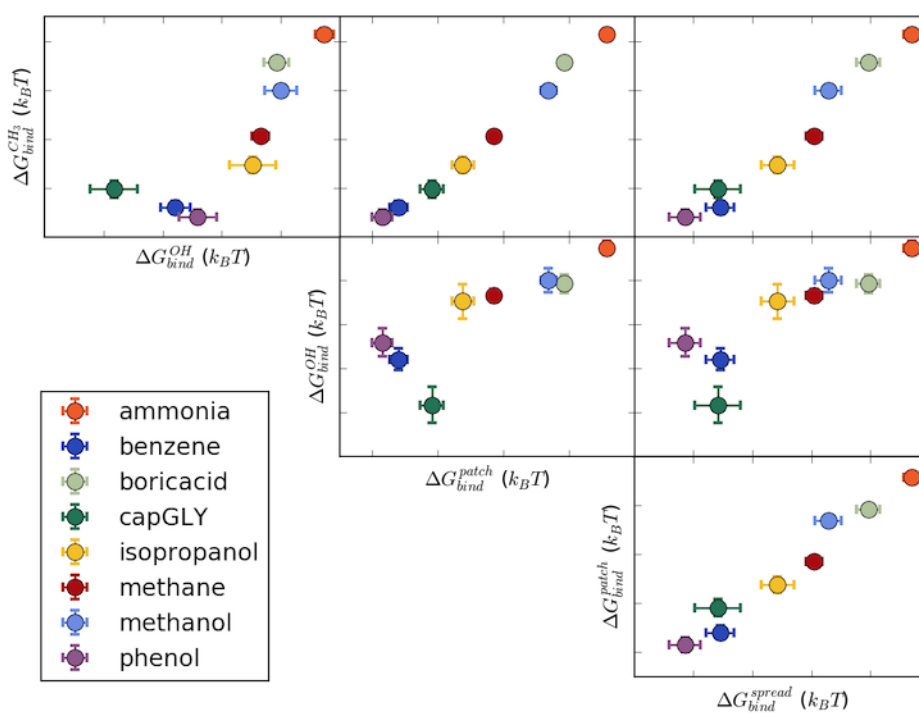


Figure 5.6: Binding free energies to all surfaces are correlated to each other despite different chemistries and chemical patterns. “Patch” and “spread” surfaces refer to the inferred globally optimal patterns shown in Table 2.

affinity, capped glycine exhibits a smaller difference upon moving between hydroxylated or methylated interfaces, or between different surface patterns (Table 3). This is due to a much more favorable bulk solubility compared to other solutes. Additionally, the change in affinity between the hydroxyl and methyl surfaces is lower than for other ligands, though the correlation of this with patterning still holds. The reduced effects of surface chemistry and heterogeneity may be due to the increased flexibility and more varied chemical composition of the capped glycine, which allow it to restructure in different environments.

Solute	$\Delta G_{solv}^{OH} - \Delta G_{solv}^{CH_3}$ ($k_B T$)	$\Delta G_{solv}^{spread} - \Delta G_{solv}^{patch}$ ($k_B T$)	$\frac{\Delta G_{solv}^{spread} - \Delta G_{solv}^{patch}}{\Delta G_{solv}^{OH} - \Delta G_{solv}^{CH_3}}$
Ammonia	0.73 ± 0.09 (3.19×10^{-4})	0.25 ± 0.06 (4.11×10^{-3})	0.34 ± 0.10
Benzene	3.01 ± 0.09 (1.12×10^{-9})	1.47 ± 0.13 (3.00×10^{-6})	0.49 ± 0.05
Boric Acid	0.90 ± 0.11 (1.16×10^{-4})	0.45 ± 0.13 (9.88×10^{-3})	0.50 ± 0.15
Capped Glycine	2.12 ± 0.15 (4.68×10^{-6})	0.94 ± 0.20 (3.17×10^{-3})	0.44 ± 0.10
Isopropanol	2.79 ± 0.12 (3.47×10^{-6})	1.08 ± 0.08 (1.90×10^{-6})	0.39 ± 0.03
Methane	2.28 ± 0.05 (1.39×10^{-7})	0.98 ± 0.06 (3.61×10^{-7})	0.43 ± 0.03
Methanol	1.52 ± 0.10 (2.79×10^{-5})	0.28 ± 0.11 (4.09×10^{-2})	0.19 ± 0.07
Phenol	3.39 ± 0.14 (3.38×10^{-8})	1.36 ± 0.14 (1.26×10^{-5})	0.40 ± 0.05

Table 5.3: Differences in solvation free energies at various interfaces are shown. Errors are based on standard error propagation of the standard error in the mean of the solvation free energy estimates based on 5 independent MD simulation runs. Numbers in parentheses provide the P-value for the two-sided hypothesis test for the equivalence of the mean solvation free energies.

To further assess the relevance of surface patterning, Table 3 compares differences between solvation free energies at the perfectly dispersed and patchy interfaces shown in Table 3 with differences between the fully hydroxylated and methylated interfaces. Differences in solvation free energies between the pure surfaces (contact angles varying

between 0° and 130° 13) are all positive, demonstrating higher solvation free energies at the hydroxylated surface, as expected from the PMFs in Figure 5.3. This includes highly polar, hydrophilic solutes such as ammonia and boric acid, though for these solutes the difference in solvation free energy is much smaller. As the solute size grows, the solvation free energy difference (and hence affinity difference) also generally increases.

Similar trends in Table 3 are observed for surfaces with a 1:1 ratio of hydroxyl and methyl headgroups, with spread surfaces (lower water density fluctuations) always exhibiting higher solvation free energies than patchy surfaces (higher water density fluctuations), even for polar solutes. However, the differences are less statistically significant compared to those for pure surfaces. This is not to say that the differences induced by patterning are insignificant – they are actually 19-50% of the range observed between pure interfaces. From a practical standpoint, however, changes induced by patterning are all less than $2 k_B T$, which would not provide significant improvements to an adsorption process or in terms of preventing fouling. If the trends observed here continue, we expect that patterning may have a significant effect for much larger solutes (i.e. proteins), though this remains out of the scope of the current study.

It is also apparent that it is possible to adjust the interfacial selectivity between two solutes with surface patterning. We define selectivity to be the ratio of the interfacial concentration of solute 2 to that of solute 1 for the same bulk concentration, which relates to the solvation free energies at the interface,

$$\frac{C_2/C_{bulk}}{C_1/C_{bulk}} = \frac{e^{-\beta\Delta G_{bind,2}}}{e^{-\beta\Delta G_{bind,1}}} = \frac{e^{-\beta(\Delta G_{solv,2}-\Delta G_{bulk,2})}}{e^{-\beta(\Delta G_{solv,1}-\Delta G_{bulk,1})}} \quad (5.7)$$

where in the second equality we have made use of Equation 5.1. Because the bulk solvation free energies are unaffected by the patterning at the interface, we may optimize selectivity by considering only the solvation free energies at the interface. Further, we may use

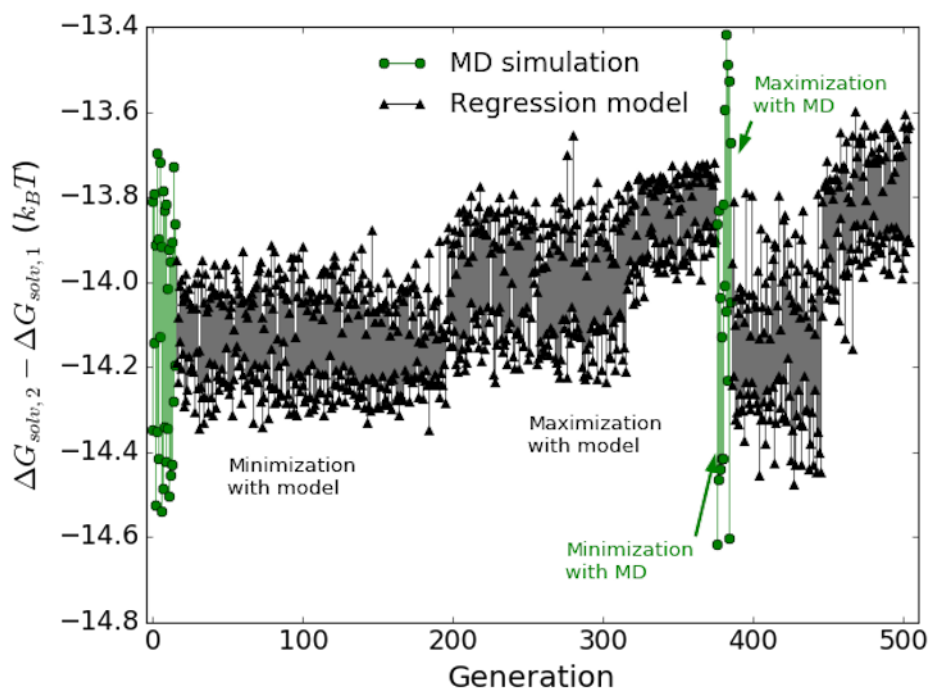


Figure 5.7: Genetic algorithm optimization successfully identifies patterns that optimize the selectivity of an interface for methanol (2) compared to methane (1). Note that a minimization of methanol selectivity is a maximization of methane selectivity and vice versa. The optimization procedure is the same as that shown in Figure 5.5 but MD simulations are run for both methane and methanol for each surface pattern and the difference in interfacial solvation free energy between the two solutes is optimized.

the difference in solute solvation free energies at the interface as a proxy for selectivity; optimizing one will also optimize the other. The solvation free energy increases more drastically for increasing surface patchiness for larger and more hydrophobic solutes, meaning that we expect patchy patterns to generally make such interfaces more selective for these solutes at the expense of slightly enhanced overall binding of the lower affinity solute.

Figure 5.7 demonstrates that the genetic algorithm is successful in optimizing surface selectivity for methanol compared to methane, resulting in surfaces with affinity differences of -13.97 ± 0.07 and $-14.18 \pm 0.08 k_B T$, as evaluated by independent MD simulations of minimum and maximum methanol selectivity surfaces sampled by the regression model during the genetic algorithm (Table S2). This range of $0.21 k_B T$ is somewhat smaller than between the expected minimum (perfectly spread) and maximum (patchy) surfaces of $0.70 k_B T$, and is much smaller than the affinity difference for these surfaces predicted by the regression model of $0.88 k_B T$ (Figure 5.7 and Table S2). We attribute this discrepancy to deficiencies of the linear regression model to distinguish patchy surface patterns, as the genetic algorithm optima shown in Table S2 are visually similar to the expected optima but subtly different. In particular, the algorithm has created a patchy surface through segregation along the longer in-plane box dimension, leading to increased contact between the two domains, reducing the effective size of the patch in terms of its influence on water behavior.

Though the presented genetic algorithm optimizations have been mostly successful, from a practical perspective the increases due to patterning for both selectivity and affinity are quite small. This is potentially a result of the chosen head-group chemistry. Both pure interfaces increase water density fluctuations and hence have affinity for all the solutes studied here. For head-groups that strongly attract water, such as charged species, solutes may actually be repelled from the surface. A similar effect might be observed

with the current head-groups but for solutes, such as ions, that are strongly solvated by water. In either case, we would expect a richer behavior with solute pattern and much larger, more practically useful changes in both affinity and selectivity.

5.3.4 Structural behavior of water associated with solute affinity

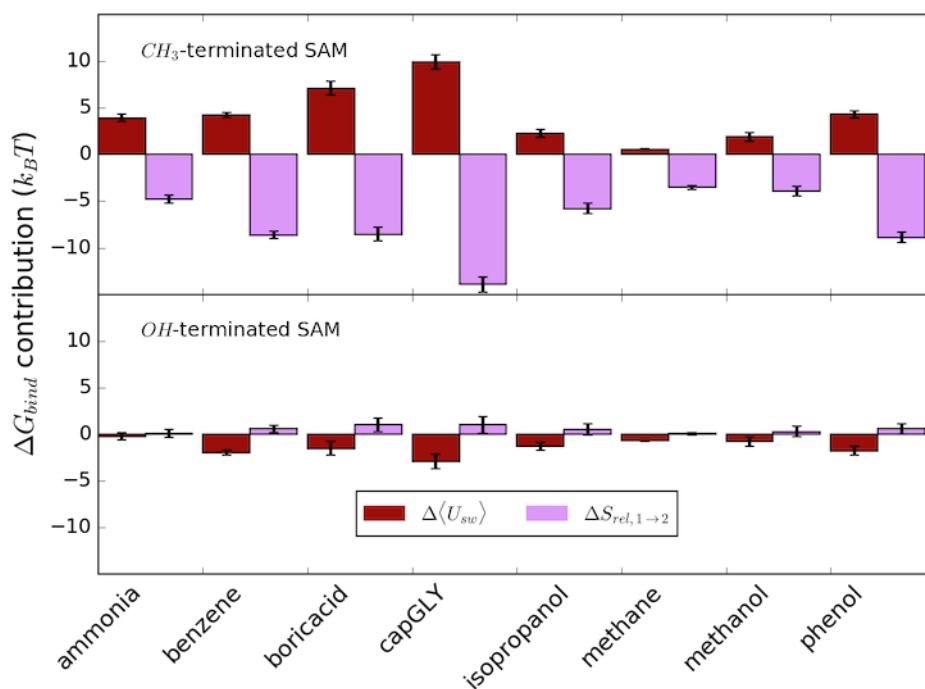


Figure 5.8: Free energies of binding are broken down into differences between direct solute-system energetics ($\Delta\langle U_{sw} \rangle_2$) of solvation and relative entropies of solvation ($\Delta S_{rel,1 \rightarrow 2}$), which for relatively rigid surfaces and solutes are dominated by water restructuring.

So far, we have rationalized differences in solute affinity for interfaces by noting the surface's different abilities to induce water density fluctuations. An alternative perspective, which may be made rigorous through consideration of the relative entropy, involves examination of the way that perturbed water structure changes when the solute

is moved from the bulk to the interface (Chapter 3). Rather than decomposing the free energy of solvation according to Equation 5.3, we can instead write

$$\Delta G_{solv} = \langle U_{sw} \rangle_2 + S_{rel,1 \rightarrow 2} \quad (5.8)$$

The subscripts 1 and 2 represent the ideal gas and fully interacting states, $\langle U_{sw} \rangle_2$ is the average of the solute-water (or overall solute-system for interfacial systems) potential energy evaluated in the coupled ensemble, and $S_{rel,1 \rightarrow 2}$ is the relative entropy associated with moving from the solute being decoupled to coupled. While $\langle U_{sw} \rangle_2$ obviously relates to direct energetic contributions to solvation, $S_{rel,1 \rightarrow 2}$ contributes to the free energy via changes to probability distributions of degrees of freedom within the system (Chapter 3 and Chapter B). The internal degrees of freedom of the solutes and surfaces considered here are not expected to change significantly upon coupling of the solute, and so the relative entropy principally tracks changes to water structuring associated with solute solvation. By then applying Equation 5.1, we can track contributions to the binding free energy from differences between the surface and bulk direct potential energy interactions and solute restructuring

$$\Delta G_{bind} = \Delta \langle U_{sw} \rangle_2 + \Delta S_{rel,1 \rightarrow 2} \quad (5.9)$$

A similar decomposition has been proposed previously, though was not interpreted in terms of the relative entropy.³⁹ Figure 5.8 shows these contributions to the binding free energy for fully methylated and hydroxylated surfaces. In the methyl-terminated surfaces, the solvent restructuring term contributes favorably to affinity while the solute-system energetics resist binding. This is in line with the results shown in Figure 5.4 – the free energy of hard-sphere solvation will not contribute an energetic term and will

be completely described by the relative entropy of water restructuring. This trend is reversed for the hydroxyl-terminated surfaces, but the differences are much smaller and of similar magnitude to the error estimates in the case of the relative entropy contributions. Qualitatively different behavior of the relative entropy between the methylated and hydroxylated surfaces further supports the idea that density fluctuations may drive affinity for the hydrophobic interface, while solvent restructuring may contribute more prominently at the hydroxylated interface. However, further comparison to Figure 5.4 is hampered by the fact that the electrostatic and LJ terms shown there contain contributions not only from solute-system energies, but also from potential energy changes involving water and SAM surfaces, which are not considered through the distinct decomposition shown in Equation 5.8.

In Figure 5.9(a) we observe distinct changes in water structure, as measured by perturbations to three-body angle distributions, for all interfaces studied. In Chapter 3, we have shown that shifts in three-body angle distributions, which contribute to the relative entropy based on water degrees of freedom, are useful for tracking changes to this quantity in the context of solvation. Three-body angles are defined as the angle between the vectors connecting two water oxygens in the first solvation shell of a central water oxygen. Perturbations from bulk for water three-body angle distributions for various bare interfaces without any solutes appear in Figure 5.9(a). All distributions are similar, peaking just below 60° and at 90° . This is in contrast to the behavior for small solutes and has been observed previously for large hydrophobes (Chapter 3) and methyl-terminated self-assembled monolayer surfaces in contact with a different water model.²⁷ It seems that regardless of the hydrophilicity of the interface, geometric effects dominate, resulting in similar perturbations to water three-body angles. There are some differences, notably that the peak near 90° is shifted to higher angles for the fully hydroxylated surface, with this distribution generally smoothed relative to other interfaces. Overall, however,

there are no clear trends that seem to correlate with increased solute affinity or density fluctuations.

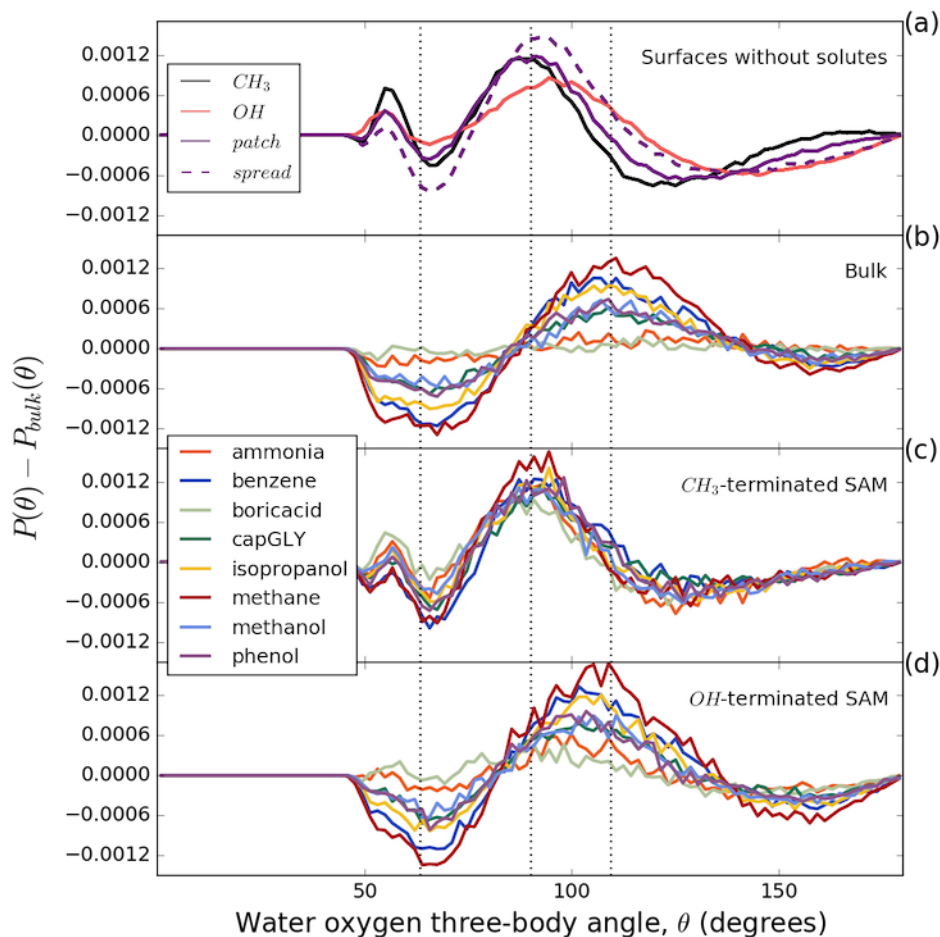


Figure 5.9: Changes from bulk water structure of three-body angle distributions for water oxygens within first solvation shells of bare surfaces with no solutes (a), solutes in bulk (b), and solutes near methylated (c) and hydroxylated (d) surfaces. Interfaces at bare surfaces are defined by the point at which average water density reaches 0.3 of its bulk value of 33.2 1/nm^3 with the first solvation shell defined as 0.4 nm from this interface, which is approximately the first minimum in the associated density profile for all surfaces. Solvation shells for solutes include all water oxygens within 0.55 nm of solute heavy atoms. Dotted vertical lines show the locations of expected peaks for liquid systems with LJ particles (63.4°), ideal gas particles (90.0°), and water (109.5°).⁸¹

In Figure 5.9(b), we focus on the structure of water surrounding solutes in bulk solution. The first solvation shells is defined as all water oxygens within 0.55 nm of any

solute atom except hydrogen. In support of this definition, all RDFs for solute atoms to water oxygens are shown in Figure G.7 of Appendix G. Figure 5.9(b) demonstrates that the smallest, most polar solutes disrupt the structure of water the least, while larger solutes, especially those that are hydrophobic, disrupt structure the most, as measured by three-body angle distributions within the first solvation shell of solutes. In Figure 5.9(b), we see similar behavior for boric acid and ammonia, hardly perturbing the distribution at all. For the other solutes, the distribution near the tetrahedral angle of 109.5° is enhanced while the region near 60° is decreased, as noted previously for small hydrophobes (Chapter 3). Figure 5.9(c,d) also show the effect of an interface on the three-body angle distribution around a solute. When solutes are at a methylated interface, their hydration shell structure is dominated by the presence of the interface. At a hydroxylated interface, however, the three-body angle structure around the solutes is nearly like that of bulk, though the interface itself perturbs water relative to bulk (Figure 5.9(b)). This is likely due to two predominant factors: first, the three-body angle distributions are only weakly perturbed by the hydroxylated interface over the region where the solutes reside, and second, the solutes are only loosely bound (Figure 5.3) and more easily sample regions farther from the interface, allowing for more bulk-like interactions with water.

The magnitude of three-body angle perturbations varies with the size of the solute, as well as the relative amounts of polar and apolar chemical groups present. For instance, methanol exhibits a perturbation approximately half as large as that for methane. If we assume that water structure near the hydroxyl moiety is not modified, similarly to boric acid or ammonia, and that this takes up about half of the hydration shell volume, then this relationship is expected. Overall, methane perturbs water structure the most, followed by benzene, isopropanol, phenol, methanol, and capped glycine. Benzene being larger than methane, this may cause additional strain on the hydrogen-bond structure of surrounding water, resulting in a slightly less tetrahedral structure. Isopropanol and

phenol both have hydrophobic groups larger than methane attached to hydroxyl groups, resulting in similar effects as benzene in combination with regions of the hydration shell that do not affect bulk water structure. While capped glycine is large and has a number of apolar groups, these groups are not clustered together and are directly adjacent to highly polar carbonyl and amine functionalities, resulting in less perturbation of water structure. It is important to note that such shifts are on a per-three-body-angle basis – the net effect on the solute solvation free energy or affinity for an interface will also depend on the number of waters in the solvation shells. This will be an important consideration when constructing a model to connect overall free energies of solvation or binding to shifts in three-body angle distributions, which seems likely to be a fruitful future endeavor.

5.4 Conclusions

We have investigated the effect of surface chemistry and chemical patterning on the affinity of a chemically diverse set of simple solutes. We find that the interfacial affinity of all solutes is similarly modified by surface chemistry and patterning, with surfaces that induce larger density fluctuations universally resulting in higher affinity. This is true even for solutes generally considered to be hydrophilic – surprisingly, we find that surface affinity is poorly correlated to solute hydrophobicity, as evaluated by either aqueous solvation or octanol-water transfer free energies. Surface hydrophobicity, as measured by density fluctuations, seems to be a much better predictor of solute affinity, especially relative to other interfaces. As such, solute affinity increases with the excluded volume of the solute, which is related to the free energy to solvate the same solute with its atoms interacting with the rest of the system as hard-spheres. This term is shown to be the dominant favorable contribution to solute affinity at interfaces, with free energies attributable to the addition of attractive interactions and solute charging resisting solute

association with a surface. However, the importance of density fluctuation physics in driving affinity for the hydroxylated interface is less clear. In this case, affinity may be driven more by reorganization of water. Future work will extend our current investigation of water interfacial structure to more thoroughly elucidate the driving force for binding at hydrophilic interfaces.

We do not here evaluate the role of surface flexibility or curvature in determining solute surface affinity; based on prior literature in fouling^{6,7} and interfacial density fluctuations,^{15,16} however, we expect that increasing flexibility and convexity will further resist binding, while the opposite should hold true for enhancing affinity. In line with these concepts, we find that surfaces with well-segregated domains of hydrophobic and hydrophilic head-groups maximize affinity for all solutes, while those with well-mixed patterns minimize affinity. However, the shifts in binding affinity due to patterning are proportional to the solute volume and overall surface affinity, meaning that such patterns may also be used to modify the selectivity of an interface for one solute or another.

Finally, we have also demonstrated that water structure at interfaces, as well as in the solvation shells of solutes, varies significantly with chemistry. Such shifts in water structure are intimately related to free energies of solvation or binding through the relative entropy. Future work should focus on more concretely connecting shifts in water structure to free energies of binding. This may occur through consideration of an expanded set of water degrees of freedom as well as the development of simple thermodynamic models based on specific structural perturbations, such as changes to three-body angle distributions.

These results hold important implications for the design of interfacial materials for use in water purification membranes, chromatography, and catalysis. In particular, we have identified specific aspects of fundamental theories of solvation and aqueous solutions that seem to apply generally to simple yet realistic solutes. Next steps will involve incrementally increasing solute and surface complexity to eventually connect the theory to

practical applications such as the prevention of biomolecular fouling or the enhancement of selective capture of valuable solvated species. An obvious extension is to charged solutes, as well as charges at an interface. In these cases, it is not clear that density fluctuations will dominate surface affinity, as energetics associated with electrostatic interactions will become increasingly important. Nanoscale patterns that optimize affinity may be very different in these cases. For charged and larger solutes, however, the efficiency and robustness of the methods presented here will need to be improved. Though we have demonstrated an efficient optimization strategy for discovering surface optimal surface patterns, we are far from utilizing this algorithm for full design of interfacial materials. In the meantime, the results presented here provide molecular-level insight into solute affinity for surfaces, helping to establish new rules of thumb for interfacial materials design. Connecting this work with previous studies on the mobility of interfacial water,¹³ it is already interesting to note that the same surface patterns that minimize/maximize water mobility actually minimize/maximize solute affinity. In membrane applications, this immediately suggests that modification of nanoscale patterning to minimize fouling may simultaneously reduce water flux through the membrane. It is then recommended to either consider alternatives to chemical patterning, such as adjustments in nanoscale surface geometry, or localize different patterning modifications to specific mesoscopic membrane regions.

References

- [1] Fritzmann C, Löwenberg J, Wintgens T, Melin T. State-of-the-art of reverse osmosis desalination. *Desalination*. 2007; 216(1-3): 1–76.
- [2] Miller DJ, Dreyer DR, Bielawski CW, Paul DR, Freeman BD. Surface Modification of Water Purification Membranes. *Angewandte Chemie International Edition*. 2017; 56(17): 4662–4711.
- [3] Werber JR, Osuji CO, Elimelech M. Materials for next-generation desalination and water purification membranes. *Nature Reviews Materials*. 2016; 1(5): 16018.
- [4] Park HB, Kamcev J, Robeson LM, Elimelech M, Freeman BD. Maximizing the right stuff: The trade-off between membrane permeability and selectivity. *Science*. 2017; 356(6343): eaab0530.
- [5] Ostuni E, Chapman RG, Holmlin RE, Takayama S, Whitesides GM. A survey of structure-property relationships of surfaces that resist the adsorption of protein. *Langmuir*. 2001; 17(18): 5605–5620.
- [6] Li L, Chen S, Zheng J, Ratner BD, Jiang S. Protein Adsorption on Oligo(ethylene glycol)-Terminated Alkanethiolate Self-Assembled Monolayers: The Molecular Basis for Nonfouling Behavior. *The Journal of Physical Chemistry B*. 2005; 109(7): 2934–2941.
- [7] Ismail AE, Grest GS, Stevens MJ. Structure and Dynamics of Water near the Interface with Oligo(ethylene oxide) Self-Assembled Monolayers. *Langmuir*. 2007; 23(16): 8508–8514.
- [8] Imbrogno J, Williams MD, Belfort G. A New Combinatorial Method for Synthesizing, Screening, and Discovering Antifouling Surface Chemistries. *ACS Applied Materials & Interfaces*. 2015; 7(4): 2385–2392.
- [9] Le TC, Penna M, Winkler DA, Yarovsky I. Quantitative design rules for protein-resistant surface coatings using machine learning. *Scientific Reports*. 2019; 9(1): 265.
- [10] Acharya H, Mozdierz NJ, Koblinski P, Garde S. How chemistry, nanoscale roughness, and the direction of heat flow affect thermal conductance of solid-water interfaces. *Industrial and Engineering Chemistry Research*. 2012; 51(4): 1767–1773. ISBN: 0888-5885.
- [11] Levine ZA, Rapp MV, Wei W, et al. Surface force measurements and simulations of mussel-derived peptide adhesives on wet organic surfaces. *Proceedings of the National Academy of Sciences*. 2016: 201603065. ISBN: 1603065113.

- [12] Zerze GH, Mullen RG, Levine ZA, Shea JE, Mittal J. To What Extent Does Surface Hydrophobicity Dictate Peptide Folding and Stability near Surfaces? *Langmuir*. 2015; 31(44): 12223–12230.
- [13] Monroe JI, Shell MS. Computational discovery of chemically patterned surfaces that effect unique hydration water dynamics. *Proceedings of the National Academy of Sciences*. 2018; 115(32): 8093–8098.
- [14] Godawat R, Jamadagni SN, Garde S. Characterizing hydrophobicity of interfaces by using cavity formation, solute binding, and water correlations. *Proceedings of the National Academy of Sciences*. 2009; 106(36): 15119–15124. ISBN: 0027-8424.
- [15] Xi E, Venkateshwaran V, Li L, Rego N, Patel AJ, Garde S. Hydrophobicity of proteins and nanostructured solutes is governed by topographical and chemical context. *Proceedings of the National Academy of Sciences*. 2017; 114(51): 13345–13350. ISBN: 0027-8424, 1091-6490.
- [16] Setny P. Water properties and potential of mean force for hydrophobic interactions of methane and nanoscopic pockets studied by computer simulations. *The Journal of Chemical Physics*. 2007; 127(5): 054505. ISBN: 0021-9606 (Print)\n0021-9606 (Linking).
- [17] Ostuni E, Grzybowski BA, Mrksich M, Roberts CS, Whitesides GM. Adsorption of Proteins to Hydrophobic Sites on Mixed Self-Assembled Monolayers [†]. *Langmuir*. 2003; 19(5): 1861–1872.
- [18] Penna M, Ley K, Maclaughlin S, Yarovsky I. Surface heterogeneity: a friend or foe of protein adsorption – insights from theoretical simulations. *Faraday Discussions*. 2016; 191: 435–464.
- [19] Comer J, Chen R, Poblete H, Vergara-Jaque A, Riviere JE. Predicting Adsorption Affinities of Small Molecules on Carbon Nanotubes Using Molecular Dynamics Simulation. *ACS Nano*. 2015; 9(12): 11761–11774.
- [20] Jamadagni SN, Godawat R, Garde S. Hydrophobicity of Proteins and Interfaces: Insights from Density Fluctuations. *Annual Review of Chemical and Biomolecular Engineering*. 2011; 2(1): 147–171. ISBN: 1947-5438\n978-0-8243-5202-8.
- [21] Acharya H, Vembanur S, Jamadagni SN, Garde S. Mapping hydrophobicity at the nanoscale: Applications to heterogeneous surfaces and proteins. *Faraday discussions*. 2010; 146: 353–365. ISBN: 1359-6640\r1364-5498.
- [22] Abramyan T, Collier G, Kucukkal TG, et al. Understanding protein-surface interactions at the atomistic level through the synergistic development of experimental and molecular simulation methods. *ACS Symposium Series*. 2012; 1120(0): 197–228. ISBN: 9780841227965.

- [23] Collier G, Vellore NA, Yancey JA, Stuart SJ, Latour RA. Comparison between Empirical Protein Force Fields for the Simulation of the Adsorption Behavior of Structured LK Peptides on Functionalized Surfaces. *Biointerphases*. 2012; 7(1): 24–43.
- [24] Raut VP, Agashe MA, Stuart SJ, Latour RA. Molecular Dynamics Simulations of Peptide - Surface Interactions. *Langmuir*. 2005; 21(4): 1629–1639.
- [25] Vellore NA, Yancey JA, Collier G, Latour RA, Stuart SJ. Assessment of the Transferability of a Protein Force Field for the Simulation of Peptide-Surface Interactions. *Langmuir*. 2010; 26(10): 7396–7404.
- [26] Yancey JA, Vellore NA, Collier G, Stuart SJ, Latour RA. Development of molecular simulation methods to accurately represent protein-surface interactions: The effect of pressure and its determination for a system with constrained atoms. *Biointerphases*. 2010; 5(3): 85.
- [27] Stock P, Monroe J, Utzig T, Smith D, Shell M, Valtiner M. Unraveling Hydrophobic Interactions at the Molecular Scale Using Force Spectroscopy and Molecular Dynamics Simulations. *ACS Nano*. 2017; 11(3).
- [28] Deighan M, Pfaendtner J. Exhaustively sampling peptide adsorption with metadynamics. *Langmuir*. 2013; 29(25): 7999–8009.
- [29] Xie Y, Liu M, Zhou J. Molecular dynamics simulations of peptide adsorption on self-assembled monolayers. *Applied Surface Science*. 2012; 258(20): 8153–8159.
- [30] Prakash A, Sprenger K, Pfaendtner J. Essential slow degrees of freedom in protein-surface simulations: A metadynamics investigation. *Biochemical and Biophysical Research Communications*. 2018; 498(2): 274–281.
- [31] Shen JW, Wu T, Wang Q, Pan HH. Molecular simulation of protein adsorption and desorption on hydroxyapatite surfaces. *Biomaterials*. 2008; 29(5): 513–532.
- [32] Mijajlovic M, Penna MJ, Biggs MJ. Free energy of adsorption for a peptide at a liquid/solid interface via nonequilibrium molecular dynamics. *Langmuir*. 2013; 29(9): 2919–2926.
- [33] Ulissi ZW, Zhang J, Sresht V, Blankschtein D, Strano MS. 2d Equation-of-State Model for Corona Phase Molecular Recognition on Single-Walled Carbon Nanotube and Graphene Surfaces. *Langmuir*. 2015; 31(1): 628–636.
- [34] Azhagiya Singam ER, Zhang Y, Magnin G, et al. Thermodynamics of Adsorption on Graphenic Surfaces from Aqueous Solution. *Journal of Chemical Theory and Computation*. 2019; 15(2): 1302–1316.

- [35] Xia XR, Monteiro-Riviere NA, Riviere JE. An index for characterization of nanomaterials in biological systems. *Nature Nanotechnology*. 2010; 5(9): 671–675.
- [36] Apul OG, Karanfil T. Adsorption of synthetic organic contaminants by carbon nanotubes: A critical review. *Water Research*. 2015; 68: 34–55.
- [37] Wallqvist A, Berne B. Hydrophobic interaction between a methane molecule and a paraffin wall in liquid water. *Chemical Physics Letters*. 1988; 145(1): 26–32.
- [38] Patel AJ, Varilly P, Jamadagni SN, Acharya H, Garde S, Chandler D. Extended surfaces modulate hydrophobic interactions of neighboring solutes. *Proceedings of the National Academy of Sciences*. 2011: 1110703108-. arXiv: 1105.0907 ISBN: 0027-8424.
- [39] Ben-Amotz D. Interfacial solvation thermodynamics. *Journal of Physics: Condensed Matter*. 2016; 28(41): 414013.
- [40] Ben-Amotz D. Water-Mediated Hydrophobic Interactions. *Annual Review of Physical Chemistry*. 2016; 67(1): 617–638.
- [41] Jungwirth P, Tobias DJ. Specific Ion Effects at the Air/Water Interface. *Chemical Reviews*. 2006; 106(4): 1259–1281.
- [42] Sander R. Compilation of Henry’s law constants (version 4.0) for water as solvent. *Atmospheric Chemistry and Physics*. 2015; 15(8): 4399–4981.
- [43] Chodera JD, Shirts MR. Replica exchange and expanded ensemble simulations as Gibbs sampling: simple improvements for enhanced mixing. *The Journal of chemical physics*. 2011; 135(19): 194110.
- [44] Paluch AS, Mobley DL, Maginn EJ. Small Molecule Solvation Free Energy: Enhanced Conformational Sampling Using Expanded Ensemble Molecular Dynamics Simulation. *Journal of Chemical Theory and Computation*. 2011; 7(9): 2910–2918.
- [45] Monroe J, Shirts M. Converging free energies of binding in cucurbit[7]uril and octa-acid host-guest systems from SAMPL4 using expanded ensemble simulations. *Journal of Computer-Aided Molecular Design*. 2014; 28(4).
- [46] Shenogina N, Godawat R, Keblinski P, Garde S. How wetting and adhesion affect thermal conductance of a range of hydrophobic to hydrophilic aqueous interfaces. *Physical Review Letters*. 2009; 102(15). ISBN: 0031-9007.
- [47] Jamadagni SN, Godawat R, Garde S. How surface wettability affects the binding, folding, and dynamics of hydrophobic polymers at interfaces. *Langmuir*. 2009; 25(22): 13092–13099. ISBN: 0743-7463 (Print).

- [48] Jorgensen WL, Madura JD, Swenson CJ. Optimized Intermolecular Potential Functions for Liquid Hydrocarbons. *Journal of the American Chemical Society*. 1984; 106(22): 6638–6646. ISBN: 0002-7863.
- [49] Cornell WD, Cieplak P, Bayly CI, et al. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *Journal of the American Chemical Society*. 1995; 117(19): 5179–5197. arXiv: z0024 ISBN: 0120261464.
- [50] Horn HW, Swope WC, Pitera JW, et al. Development of an improved four-site water model for biomolecular simulations: TIP4p-Ew. *Journal of Chemical Physics*. 2004; 120(20): 9665–9678. ISBN: 0021-9606; 0021-9606.
- [51] Abraham MJ, Murtola T, Schulz R, et al. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*. 2015; 1-2: 19–25. arXiv: 1503.05249v1 ISBN: 23527110.
- [52] Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14sb: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99sb. *Journal of Chemical Theory and Computation*. 2015; 11(8): 3696–3713.
- [53] Otkidach D, Pletnev I. Conformational analysis of boron-containing compounds using Gillespie–Kepert version of molecular mechanics. *Journal of Molecular Structure: THEOCHEM*. 2001; 536(1): 65–72.
- [54] Mayo SL, Olafson BD, Goddard WA. DREIDING: a generic force field for molecular simulations. *The Journal of Physical Chemistry*. 1990; 94(26): 8897–8909.
- [55] Liu Z, Huang S, Wang W. A Refined Force Field for Molecular Simulation of Imidazolium-Based Ionic Liquids. *The Journal of Physical Chemistry B*. 2004; 108(34): 12978–12989.
- [56] Case D, Ben-Shalom I, Brozell S, et al. AMBER 2018. 2018.
- [57] Bayly CI, Cieplak P, Cornell W, Kollman PA. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *The Journal of Physical Chemistry*. 1993; 97(40): 10269–10280.
- [58] Frisch MJ, Trucks GW, Schlegel HB, et al. *Gaussian 16*. 2016.
- [59] Jakalian A, Jack DB, Bayly CI. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *Journal of computational chemistry*. 2002; 23(16): 1623–41.
- [60] Fennell CJ, Wymer KL, Mobley DL. A Fixed-Charge Model for Alcohol Polarization in the Condensed Phase, and Its Role in Small Molecule Hydration. *The Journal of Physical Chemistry B*. 2014; 118(24): 6438–6446.

- [61] Chen B, Potoff JJ, Siepmann JI. Monte Carlo Calculations for Alcohols and Their Mixtures with Alkanes. Transferable Potentials for Phase Equilibria. 5. United-Atom Description of Primary, Secondary, and Tertiary Alcohols. *The Journal of Physical Chemistry B*. 2001; 105(15): 3093–3104.
- [62] Eastman P, Swails J, Chodera JD, et al. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLOS Computational Biology*. 2017; 13(7): e1005659.
- [63] Miyamoto S, Kollman PA. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of Computational Chemistry*. 1992; 13(8): 952–962. ISBN: 0192-8651.
- [64] Andersen C. Rattle : A “ Velocity ” Version of the Shake Algorithm for Molecular Dynamics Calculations. *Journal of Computational ph*. 1983; 52: 24–34.
- [65] Torrie G, Valleau J. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics*. 1977; 23(2): 187–199. ISBN: 0021-9991.
- [66] Beutler TC, Mark AE, Schaik RC, Gerber PR, Gunsteren WF. Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations. *Chemical Physics Letters*. 1994; 222(6): 529–539.
- [67] Tan Z. Optimally Adjusted Mixture Sampling and Locally Weighted Histogram Analysis. *Journal of Computational and Graphical Statistics*. 2017; 26(1): 54–65.
- [68] Chodera JD, Shirts MR. A python implementation of the multistate Bennet acceptance ratio (MBAR).
- [69] Shirts MR, Chodera JD. Statistically optimal analysis of samples from multiple equilibrium states. *The Journal of chemical physics*. 2008; 129(12): 124105.
- [70] Hummer G, Garde S, García aE, Pohorille a, Pratt LR. An information theory model of hydrophobic interactions. *Proceedings of the National Academy of Sciences of the United States of America*. 1996; 93(17): 8951–8955. ISBN: 2819960820.
- [71] Barker JA, Henderson D. Perturbation Theory and Equation of State for Fluids. II. A Successful Theory of Liquids. *The Journal of Chemical Physics*. 1967; 47(11): 4714–4721.
- [72] Andersen HC, Weeks JD, Chandler D. Relationship between the Hard-Sphere Fluid and Fluids with Realistic Repulsive Forces. *Physical Review A*. 1971; 4(4): 1597–1607.
- [73] Weeks JD, Chandler D, Andersen HC. Role of Repulsive Forces in Determining the Equilibrium Structure of Simple Liquids. *The Journal of Chemical Physics*. 1971; 54(12): 5237–5247.

- [74] Ben-Naim A. Standard Thermodynamics of Transfer. Uses and Misuses. *The Journal of Physical Chemistry*. 1978; 82(7): 792–803.
- [75] Dorsey JG, Dill KA. The molecular mechanism of retention in reversed-phase liquid chromatography. *Chemical Reviews*. 1989; 89(2): 331–346.
- [76] Hummer G, Garde S. Cavity Expulsion and Weak Dewetting of Hydrophobic Solutes in Water. *Physical Review Letters*. 1998; 80: 4193–4196. ISBN: 0031-9007.
- [77] Patel AJ, Varilly P, Chandler D. Fluctuations of Water near Extended Hydrophobic and Hydrophilic Surfaces. *The Journal of Physical Chemistry B*. 2010; 114(4): 1632–1637.
- [78] Patel AJ, Varilly P, Jamadagni SN, Hagan MF, Chandler D, Garde S. Sitting at the edge: How biomolecules use hydrophobicity to tune their interactions and function. *Journal of Physical Chemistry B*. 2012; 116(8): 2498–2503. arXiv: 1109.4431 ISBN: 1520-5207 (Electronic) \n1520-5207 (Linking).
- [79] Widom B. Some Topics in the Theory of Fluids. *The Journal of Chemical Physics*. 1963; 39(11): 2808–2812.
- [80] Sosso GC, Caravati S, Rotskoff G, Vaikuntanathan S, Hassanali A. On the Role of Nonspherical Cavities in Short Length-Scale Density Fluctuations in Water. *The Journal of Physical Chemistry A*. 2017; 121(1): 370–380.
- [81] Chaimovich A, Shell MS. Tetrahedrality and structural order for hydrophobic interactions in a coarse-grained water model. *Physical Review E*. 2014; 89(2): 022140. ISBN: 1550-2376 (Electronic) 1539-3755 (Linking).

Chapter 6

Conclusions

6.1 Summary

In summary, we have presented a body of work extending current knowledge of water's response to interfaces and solutes, which drives water-mediated interactions key to aqueous solvation, adsorption, association, and self-assembly phenomena. In particular, we have used realistic yet simple model systems to probe modern theories developed on the basis of artificial, toy problems. This represents an essential and timely bridge necessary to make fundamental insights learned from theory useful in practical scenarios. All of this, of course, relied on the presence and willingness of experimental collaborators devoted to developing fundamental understanding and interested in probing simple systems in a controlled manner in order to make direct comparisons to simulation. All of the research presented here has benefitted significantly from such collaborations, rendering these interactions absolutely essential to the overall goal of developing a fundamental understanding of water's interfacial behavior.

Chapter 2 describes a close-knit collaboration with the Valtiner group to study the affinity of simple yet realistic peptides for a hydrophobic surface. The system at hand was

designed to be amenable to both simulation and atomic force microscopy experiments so that results could be compared as directly as possible with little ambiguity. By changing the hydrophobic content of short peptides and measuring the free energy to separate them from an extended interface, we determined the scaling of the hydrophobic interaction at a molecular level, finding that it grows approximately linearly with the number of hydrophobic units. Molecular simulations of these systems allow for the characterization of molecular scale structure of both the peptides and surrounding water. This analysis highlighted a need to establish structural metrics of water amenable to flexible, realistic solutes heterogeneous in their chemical composition. Perturbations to three-body angle distributions of waters within peptide solvation shells satisfy this criterion and, remarkably, are useful in developing predictive models for peptide-surface affinities.

We further investigate three-body angle distributions and connect this structural signature of water to previously-developed metrics and theories in Chapter 3. Specifically, we consider shifts to this distribution as a function of temperature, density, and the presence of model solutes that have previously been well-characterized. We have shown that three-body angle distributions provide an intuitive way of measuring relative preference of waters for tetrahedral or simple-fluid-like configurations and that three-body angle distributions seem constrained to varying in only a relatively small number of ways. At the same time, consideration of these distributions is still possible in heterogeneous environments, even for flexible solutes. Changes to such water degrees of freedom and their role in thermodynamic processes, as it turns out, are rigorously connected to the relative entropy. With this proof, we have opened up a framework for understanding and assessing the relative role of changes to any water degree of freedom along a thermodynamic pathway. Generalizations to solutes of arbitrary complexity are possible, though the subject of future work. While the practical usefulness of the theory is yet to be explored, it shows great promise for the identification of key solute degrees of freedom during

solvation processes and as a result the development of minimalistic, highly-predictive thermodynamic models. This is deeply connected to the use of relative entropy in coarse-graining, but here we are proposing its merger with dimensionality reduction techniques to simultaneously determine the degrees of freedom present in a coarse-grained ensemble that must be present in order to provide a reliable model, thermodynamic or otherwise.

Rather than examining how specific solutes modulate water properties, Chapters 4 and 5 instead consider the problem of how to design interfaces to bias water to behave in a desired fashion. This is only possible due to the development of a novel genetic algorithm that adjusts surface features in order to modify arbitrary interfacial properties, such as nearby water structure, thermodynamics, or dynamics. While it was previously well-established that surface patterning of more and less hydrophilic groups leads to differences in surface hydrophobicity, as measured by water density fluctuations, Chapter 4 demonstrates that such patterns may also be used to modulate water dynamics. Indeed, the same patterns that maximize water density fluctuations also lead to increased water mobility, resulting in a clear correlation across a diverse set of material interfaces between these two quantities. This trend breaks down, however, for surfaces with only isotropic interaction potentials between surface sites and water. Specifically, without electrostatic interactions that preferentially orient water molecules, surfaces are incapable of modulating water dynamics through patterning. This is despite the ability of such surfaces to modulate water density fluctuations and hence surface hydrophobicity.

Density fluctuations also dominate in determining the affinities of neutral, small-molecule solutes for patterned interfaces in Chapter 5. We demonstrate that surfaces with patterns that lead to higher density fluctuations also maximize surface affinity for all studied solutes, even those that are hydrophilic. In consideration of Chapter 4, this result immediately suggests limitations on materials for use in membranes or other applications, at least for the specific materials considered. In order to prevent fouling (i.e. minimize

affinities of a broad range of solutes), interfaces should be designed at a nanoscale to diminish water density fluctuations, which also reduces water mobility near the interface. The impact of this tradeoff on macroscopic processes, such as water purifications, is unclear, though our work lays the theoretical foundations for understanding such industrial processes from a molecular viewpoint. Further, the presented design algorithms are a significant contribution towards both computational design of materials as well as the introduction of nanoscale features into materials design.

6.2 Outlook

Throughout this work, there is a notable lack of consideration for charged solutes and interfaces. This of course is for good reason – such systems are more difficult to work with from a methodological point of view and are also known to present unique behaviors compared to the neutral systems studied here.¹ However, with the established toolset in hand, it is time to begin exploring the role of water in mediating and modulating charged interactions, as well as the role of charges at interfaces in manipulating water. In the latter case, significantly different behavior is expected for patterning of charged groups compared to that of neutral surface functional groups of differing polarity. With the introduction of charged groups, water will be tightly bound at specific sites on the surface, with the degree of this adhesion highly dependent on the local electrostatic environment, i.e. the presence of other charged species. This is in contrast to even the more polar head-groups, where density fluctuations are at most observed in our simulated systems to be equivalent to bulk water. Ionic surface functional groups, however, will dampen fluctuations relative to bulk, potentially heralding a fundamental shift in observed water behaviors. In terms of solute affinity for interfaces, surface charges will greatly enhance specific energetic interactions and orientational preferences of solutes, potentially leading to dominance of these or

other factors in determining affinity over water density fluctuations. Overall, variations in patterning of charged groups will result in distinctive changes in water structure as well as solute affinity, adding useful, molecular-scale principles to the interfacial materials design toolbox that we have developed.

In terms of solute solvation and association, it would be very interesting to investigate the extent to which charge affects water three-body angle distributions or other degrees of freedom in solute hydration shells. Water orients distinctively relative to solutes of different charge,² as well as relative to neutral solutes, but it will be very interesting to carefully examine the response of water-water degrees of freedom. Our results currently suggest that perturbations to water three-body angle distributions in the solvation shells of ions should be captured by variations in bulk distributions across temperatures and densities. If this is the case, it will be crucial to ascertain whether or not such perturbations are still distinct from those induced by neutral solutes. In particular, if no distinction is made, it clearly highlights the necessity of considering other degrees of freedom when determining solvation or other free energies.

This last point is of particular relevance to extensions of the theory presented in Chapter 3 of this work. In the context of relative entropy associated with thermodynamic processes (such as the case of solvation considered here), there is no reason that multiple water degrees of freedom may not be considered. In analogy to entropy expansions of liquids,³⁻⁶ relative entropy expansions should also be feasible.⁷⁻⁹ As with such expansions in liquids, however, the calculation of higher-order contributions (such as those between three or four molecules) becomes cumbersome. A study examining the convergence properties of various forms of the relative entropy expansion in the context of solvation or other thermodynamic process, such as solute-solute association, is of great interest. It is anticipated that such expansions will converge much faster than their thermodynamic entropy analogs, as the relative entropy only weights changes to distributions of structural

degrees of freedom. If this is the case, this would also provide an efficient route towards calculating solvation free energies.

Beyond simply the efficient calculation of free energies, consideration of the relative entropy should provide insight into which degrees of freedom are most critical to the physics of the system. In coarse-graining applications, minimization of the relative entropy is used to naturally embed contributions from these critical structural features at the atomistic level into a model at the coarse-grained level. A major issue in coarse-graining, however, is selecting an appropriate coarsened model (i.e. the number of atoms to lump into a single bead, the geometry of the coarsened beads, etc.).^{10,11} To solve this, as well as to accelerate free energy calculations, it may prove very fruitful to couple calculations based on the relative entropy to dimensionality reduction techniques. First, it would be essential to establish an orthogonal set of structural degrees of freedom whose probability distributions contained no joint or mutual information. In this scenario, the relative entropy and its expansions should become perfectly additive. Such a coordinate transformation might be determined by techniques such as independent component analysis.¹² By then computing relative entropies associated with thermodynamic processes or coarse-graining in this independent set of coordinates, the collective degree of freedom most informative for constructing a reduced model would be identified. There have been recent efforts to reduce dimensionality in order to identify natural reaction coordinates and subsequently bias sampling along these coordinates for increased simulation efficiency while also discovering key kinetic pathways.^{13,14} However, the methods proposed here may be useful in filling the gap of constructing reduced models preserving thermodynamics. While promising, the method might still result in highly transformed coordinates that lack physical intuition. These coordinates might still prove useful for efficient calculations but may need to be further manipulated in order to provide physical insight useful for theoretical advances.

Finally, the genetic algorithm used for interfacial design stands to benefit from a few

specific improvements. Principally, the current algorithm makes no attempt to reduce the space searched by checking for structures that are equivalent under periodic translation or rotation. This could greatly improve its efficiency by reducing the search space. Additionally, the algorithm is very sensitive to noise in the data, relying on accurate assessments of metrics to be optimized for each proposed surface configuration. Currently this requires that more computer time be devoted to increasing the precision of the metric calculation, which means that longer molecular simulations are required. Improvements to specifically reduce the effect of statistical outliers will improve the accuracy, and likely also efficiency, of the genetic algorithm. Finally, the genetic algorithm itself may need to be reconsidered in deference to more modern machine learning techniques. As the interfaces that we seek to design inevitably become more complicated, more sophisticated optimization strategies will likely prove necessary.

References

- [1] Jungwirth P, Tobias DJ. Specific Ion Effects at the Air/Water Interface. *Chemical Reviews*. 2006; 106(4): 1259–1281.
- [2] Saha D, Mukherjee A. Impact of Ions on Individual Water Entropy. *The Journal of Physical Chemistry B*. 2016; 120(30): 7471–7479.
- [3] Baranyai A, Evans DJ. Direct entropy calculation from computer simulation of liquids. *Physical Review A*. 1989; 40(7): 3817–3822. ISBN: 1050-2947.
- [4] Hernando J. Thermodynamic potentials and distribution functions. *Molecular Physics*. 1990; 69(2): 319–326.
- [5] Laird BB, Haymet ADJ. Calculation of the entropy from multiparticle correlation functions. *Physical Review A*. 1992; 45(8): 5680–5689. ISBN: 1050-2947.
- [6] Wallace DC. On the role of density fluctuations in the entropy of a fluid. *The Journal of Chemical Physics*. 1987; 87(4): 2282–2284.
- [7] Matsuda H. Physical nature of higher-order mutual information: Intrinsic correlations and frustration. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*. 2000; 62(3 A): 3096–3102.
- [8] McClendon CL, Friedland G, Mobley DL, Amirkhani H, Jacobson MP. Quantifying Correlations Between Allosteric Sites in Thermodynamic Ensembles. *Journal of chemical theory and computation*. 2009; 5(9): 2486–2502.
- [9] McClendon CL, Hua L, Barreiro G, Jacobsen MP. Comparing Conformational Ensembles Using the Kullback–Leibler Divergence Expansion. *Journal of Chemical ...*. 2012; 8(6): 2115–2126.
- [10] Shell MS. Coarse-Graining With The Relative Entropy. *Advances in Chemical Physics*. 2016; 161: 395–441. ISBN: 9781119290971.
- [11] Foley TT, Shell MS, Noid WG. The impact of resolution upon entropy and information in coarse-grained models. *The Journal of Chemical Physics*. 2015; 143(24): 243104.
- [12] Hyvärinen A, Oja E. Independent component analysis: algorithms and applications. *Neural Networks*. 2000; 13(4-5): 411–430.
- [13] Tiwary P, Berne BJ. Spectral gap optimization of order parameters for sampling complex molecular systems. *Proceedings of the National Academy of Sciences*. 2016; 113(11): 2839–2844.
- [14] Ribeiro JML, Bravo P, Wang Y, Tiwary P. Reweighted autoencoded variational Bayes for enhanced sampling (RAVE). *The Journal of Chemical Physics*. 2018; 149(7): 072301.

Appendix A

Characterization of peptide structure in equilibrium and non-equilibrium simulations

In Table 2.1 of Chapter 2, the simulation estimate of the free energy difference for the (GS)₅ scaffold is similar to that of the (GS)₄GL peptide, while in experiment no interaction for the former was detected. While this does not affect the slope with respect to the number of leucine residues, it is interesting to explore the source of this deviation. Figure 2.6B shows that the average force profile for the (GS)₅ peptide exhibits a slight peak only at small distances from the surface. In experiments, such an interaction would be hidden by the large attractive forces between the AFM tip and surface at these same distances (Figure 2.3A). By using approach curves as a baseline, experiments thus set the free energy change for the scaffold peptide to zero.

In simulation, however, one still might not expect the scaffold to have a higher affinity for the surface than the (GS)₄GL peptide, as observed at faster pulling rates. Figure A.1 demonstrates that the biasing force to keep the peptide near the surface induces a

β -hairpin-like structure, in which glycine hydrogens are pointed towards the surface while serine hydroxyl groups are directed into solution. This contrasts with observed structures of the unrestrained (GS)₅ peptide on the surface, which show far less time in a β -like conformation and more time protruding into solution (Figure A.1). Thus, the applied biasing force may over-populate the starting ensemble with structures of more favorable affinity for the surface. Pulling at rates faster than the relaxation time of the peptide configuration on the surface then biases the observed work distributions to artificially high values. Unlike the (GS)₅ scaffold, surface conformations of leucine-containing peptides do not seem to be significantly perturbed by a restraint. However, extensions of hydrophilic portions of the peptide into solution seem to become more common (Figure A.1).

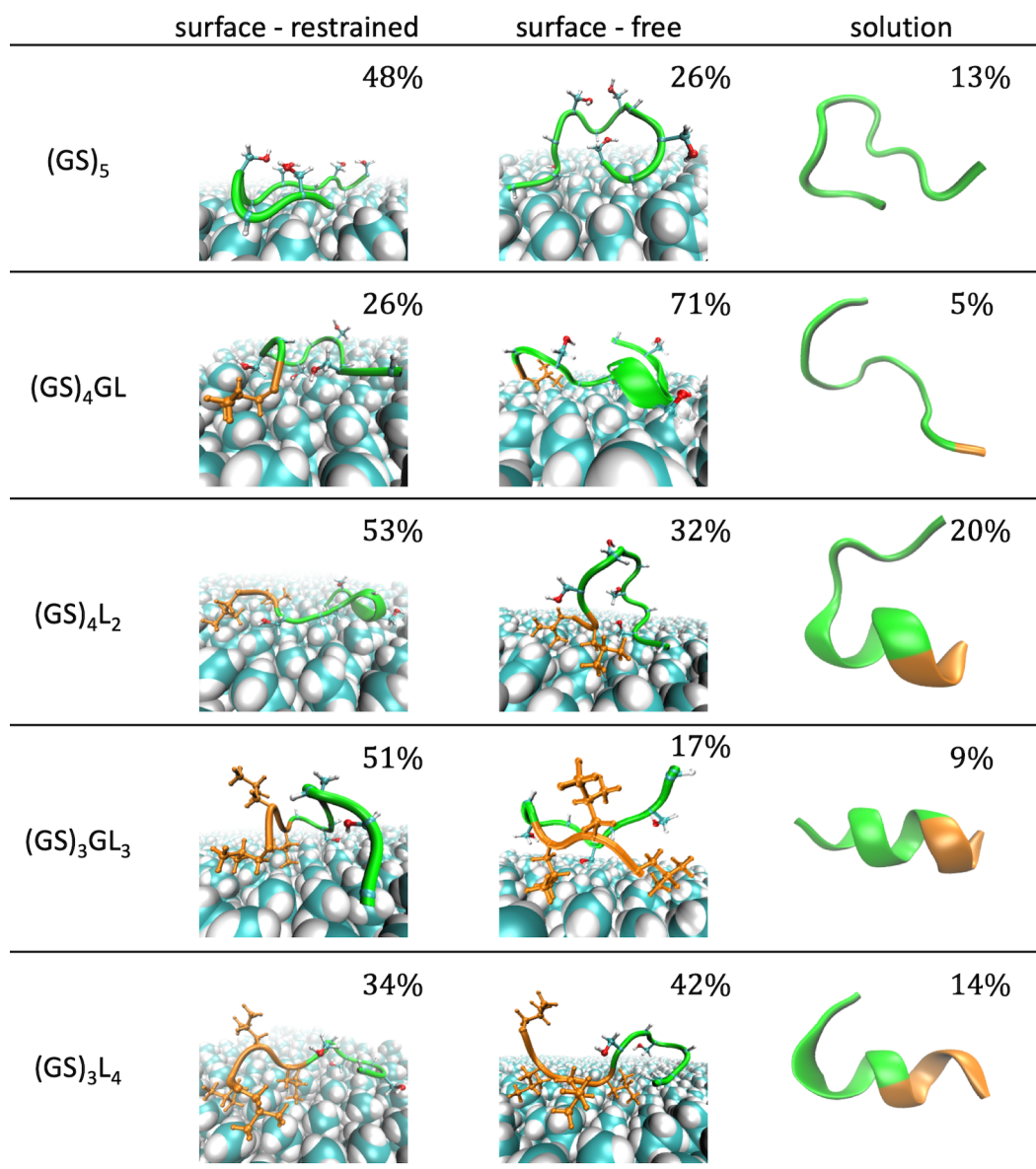


Figure A.1: Highest percentage RMSD clusters are shown for peptides in solution and with the pulling atom, restrained or free, near the simulated SAM surface. Differences do not appear with the release of restraints, though in many cases the GS part of the chain extends more readily into solution. For 3 and 4 leucines, at least one leucine is left dangling into solution, regardless of restraints, due to constraints introduced by backbone connectivity. Especially for structures in solution, low top cluster percentages indicate a lack of well-defined secondary structure.

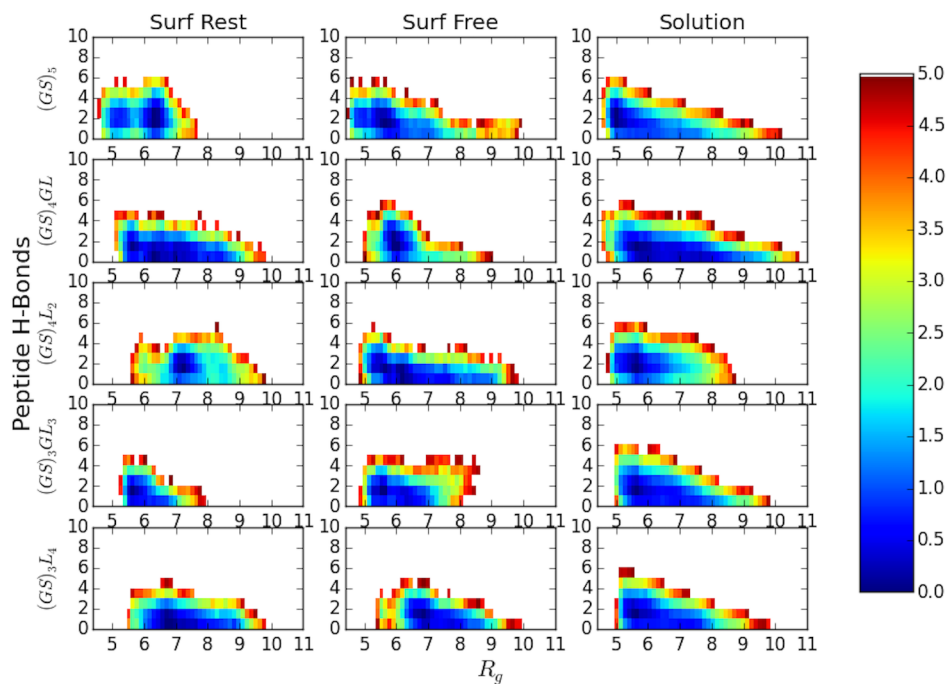


Figure A.2: Two-dimensional free energy surfaces in intrapeptide hydrogen bond (oxygen acceptor to hydrogen donor distance cutoff of 2.1 \AA and donor-H-acceptor angle cutoff of 150°) and radius of gyration space are shown for all simulated peptides. Heat maps in the first column show results for the peptide restrained to be near the surface, in the second for the peptide free on the surface, and in the third for the peptide in solution. Each plot has been shifted in reference to its point of lowest free energy, with the scale at right in $k_B T$. In the case of the peptide restrained on the surface, re-weighting to account for the energy of the restraint has not been performed. Such re-weighting results in negligible changes to the free energy surfaces shown in the first column. In general, peptides explore a wide range of configuration space, though this space is reduced for peptides near the surface compared to those in bulk.

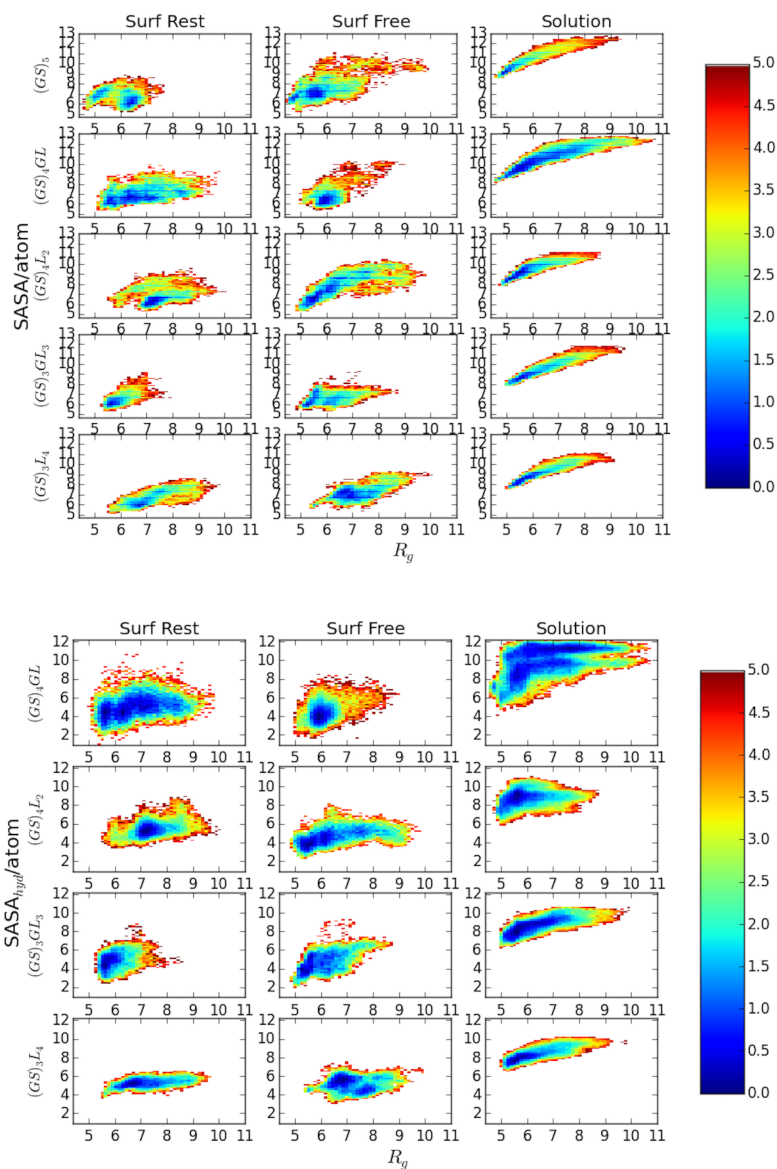


Figure A.3: Similar to Figure A.2, but with total SASA per atom (top) or hydrophobic SASA per atom (bottom) on the y-axis. Once the restraint is released, the increases in R_g on the surface also occasionally correspond to increases in the SASA per atom, indicating that the peptide extends its hydrophilic region away from the surface into solution.

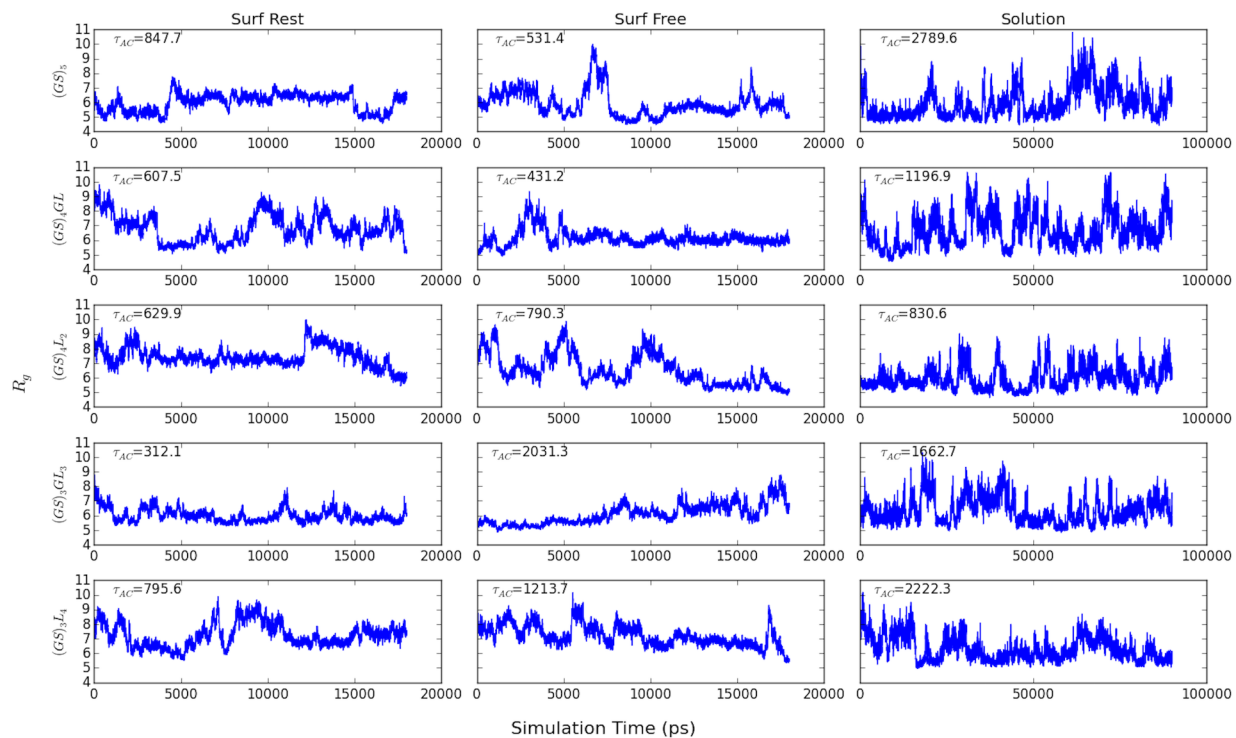


Figure A.4: Radius of gyration for each peptide restrained on the surface, free of restraints but on the surface, and in solution. From the large jumps in R_g , it is clear that the peptide frequently folds and unfolds, indicating a lack of well-defined structure. This matches results from the free energy surfaces above, but further indicates that the wide regions of stability in SASA, R_g , and hydrogen bond space are due to peptide disorder rather than lack of equilibration. Autocorrelation times are shown at the top left corner of each figure. Note that simulation times for peptides in solution are 90 ns, while those on the surface are 18 ns.

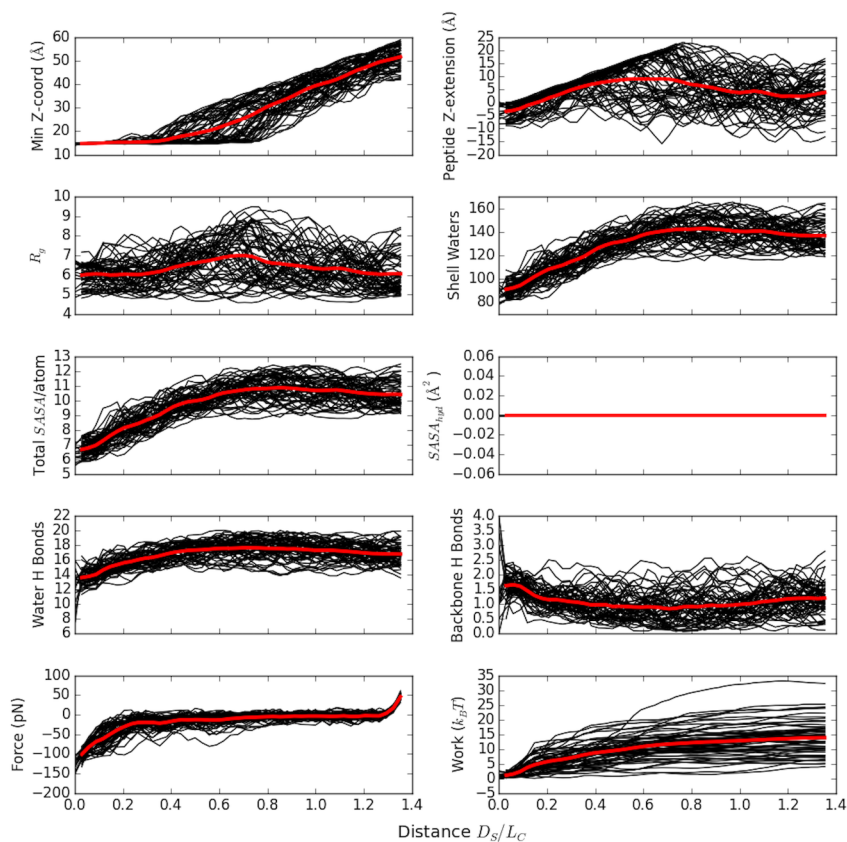


Figure A.5: For the $(GS)_5$ peptide, various quantities are shown throughout the slow-pulling process. Black curves are averages over configurations from individual trajectories falling within pulling coordinate (z -axis) bins of 1 \AA width. Red curves are averages over all trajectories, or black curves. The minimum z -coordinate refers to the atom in the peptide closest to the SAM surface, with a pull-off event defined as the time at which this value exceeds 21 \AA . Peptide z -extension refers to the z -separation between the pulled atom and the first backbone atom in a leucine residue (or the backbone atom preceding the C-terminal cap in $(GS)_5$). R_g refers to the peptide radius of gyration. Shell waters are defined as waters within 5 \AA from exposed peptide atoms. Total SASA/atom is the per-atom solvent accessible surface area for the entire peptide, while SASA_{hyd} is the solvent accessible surface area of only leucine residues. Numbers of hydrogen bonds (distance cutoff of 2.1 \AA between heavy atoms and hydrogen bonded hydrogens, and donor-H-acceptor angle cutoff of 150°) are shown between the peptide and water, as well as between backbone atoms in the peptide. Force and work curves are additionally shown at the bottom.

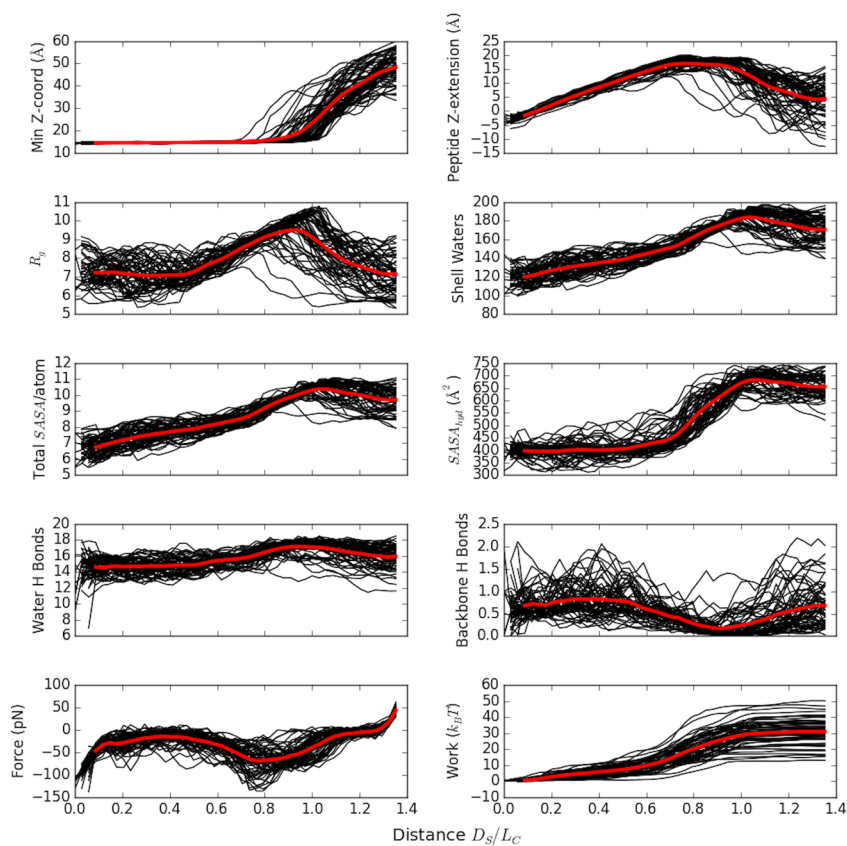


Figure A.6: Metrics shown for the $(GS)_3L_4$ peptide, slow-pulling simulations. From Figure A.5 to Figure A.6, distinct qualitative differences in many of the curves reveal a starkly different pull-off mechanism with the addition of leucine residues. Mainly, the radius of gyration develops a clear maximum that corresponds to a force peak around full extension of the peptide.

Appendix B

Derivation of solvation-thermodynamic relations in terms of the relative entropy

B.1 Free energy of solvating colloids

We begin with the definition of the relative entropy between the system with the colloid solvated (state 2) and that with the colloid in the ideal gas phase and a separate bulk water system (state 1):

$$S_{rel,1\rightarrow 2} = \int \wp_2(V) \left[\int \wp_2(\mathbf{p}_w^N, \mathbf{r}_w^N, \mathbf{p}_c, \mathbf{r}_c; V) \ln \left[\frac{\wp_2(\mathbf{p}_w^N, \mathbf{r}_w^N, \mathbf{p}_c, \mathbf{r}_c, V)}{\wp_1(\mathbf{p}_w^N, \mathbf{r}_w^N, \mathbf{p}_c, \mathbf{r}_c, V)} \right] d\mathbf{p}_w^N d\mathbf{r}_w^N d\mathbf{p}_c d\mathbf{r}_c \right] dV \quad (\text{B.1})$$

In the above, there are N solvent molecules and a single solute and \mathbf{r} and \mathbf{p} represent vectors of positions and momenta for all degrees of freedom of a single molecule and the subscripts w and c signify whether the molecule is a solvent (water) or colloid. We consider the above in the isothermal-isobaric ensemble and rewrite the expression in terms

of partition functions.

$$\begin{aligned}
S_{rel,1\rightarrow 2} &= \int \wp_2(V) \left[\int \wp_2(\mathbf{p}_w^N, \mathbf{r}_w^N, \mathbf{p}_c, \mathbf{r}_c; V) \right. \\
&\quad \left. \ln \left[\frac{e^{-\beta H_2(\mathbf{p}_w^N, \mathbf{r}_w^N, \mathbf{p}_c, \mathbf{r}_c)} e^{-\beta pV}}{\Delta_2} \frac{\Delta_1}{e^{-\beta H_1(\mathbf{p}_w^N, \mathbf{r}_w^N, \mathbf{p}_c, \mathbf{r}_c)} e^{-\beta pV}} \right] d\mathbf{p}_w^N d\mathbf{r}_w^N d\mathbf{p}_c d\mathbf{r}_c \right] dV \\
&= \int \wp_2(V) \left[\int \wp_2(\mathbf{p}_w^N, \mathbf{r}_w^N, \mathbf{p}_c, \mathbf{r}_c; V) \right. \\
&\quad \left. \ln \left[\frac{e^{-\beta U_2(\mathbf{r}_w^N, \mathbf{r}_c)} e^{-\beta K_{2,w}(\mathbf{p}_w^N)} e^{-\beta K_{2,c}(\mathbf{p}_c)} e^{-\beta pV}}{\Delta_2} \right. \right. \\
&\quad \left. \left. \frac{\Delta_1}{e^{-\beta U_1(\mathbf{r}_w^N, \mathbf{r}_c)} e^{-\beta K_{1,w}(\mathbf{p}_w^N)} e^{-\beta K_{1,c}(\mathbf{p}_c)} e^{-\beta pV}} \right] d\mathbf{p}_w^N d\mathbf{r}_w^N d\mathbf{p}_c d\mathbf{r}_c \right] dV
\end{aligned} \tag{B.2}$$

The momenta are not affected by an isothermal solvation process, and hence the kinetic energies are equal ($K_{2,x} = K_{1,x}$) for both solvent and solute. This means that all terms containing momenta or integrals over momenta will cancel within the logarithm, allowing integration over these degrees of freedom and cancelation with the momenta component of Δ_2 . Additionally, the exponentials involving the pressure term cancel. The relative entropy is then more simply written as

$$S_{rel,1\rightarrow 2} = \int \wp_2(V) \left[\int \wp_2(\mathbf{r}_w^N, \mathbf{r}_c; V) \ln \left[\frac{e^{-\beta U_2(\mathbf{r}_w^N, \mathbf{r}_c)}}{\Delta_2} \frac{\Delta_1}{e^{-\beta U_1(\mathbf{r}_w^N, \mathbf{r}_c)}} \right] d\mathbf{r}_w^N d\mathbf{r}_c \right] dV \tag{B.3}$$

where we have now defined Δ as the configurational partition function and the probabilities now only depend on the positions and not the momenta. The full partition functions have been left for simplicity of notation, as will become clear later on. For a pairwise potential function, we may split the potential energy in the solvated ensemble (U_2) into water-water, colloid-water, and colloid only terms. After cancellation and re-arrangement, we arrive at an expression for the free energy of solvation in terms of the relative entropy:

$$\begin{aligned}
S_{rel,1\rightarrow 2} &= \int \wp_2(V) \left[\int \wp_2(\mathbf{r}_w^N, \mathbf{r}_c; V) \right. \\
&\quad \left. \ln \left[\frac{e^{-\beta U_{ww}(\mathbf{r}_w^N)} e^{-\beta U_{wc}(\mathbf{r}_w^N, \mathbf{r}_c)} e^{-\beta U_c(\mathbf{r}_c)} \Delta_1}{\Delta_2 e^{-\beta U_{ww}(\mathbf{r}_w^N)} e^{-\beta U_c(\mathbf{r}_c)}} \right] d\mathbf{r}_w^N d\mathbf{r}_c \right] dV \\
S_{rel,1\rightarrow 2} &= \int \wp_2(V) \left[\int \wp_2(\mathbf{r}_w^N, \mathbf{r}_c; V) \ln \left[\frac{e^{-\beta U_{wc}(\mathbf{r}_w^N, \mathbf{r}_c)} \Delta_1}{\Delta_2} \right] d\mathbf{r}_w^N d\mathbf{r}_c \right] dV \\
S_{rel,1\rightarrow 2} &= \int \wp_2(V) \left[\int \wp_2(\mathbf{r}_w^N, \mathbf{r}_c; V) \left(-\beta U_{wc}(\mathbf{r}_w^N, \mathbf{r}_c) \right) d\mathbf{r}_w^N d\mathbf{r}_c \right] dV \\
&\quad + \int \wp_2(V) \left[\int \wp_2(\mathbf{r}_w^N, \mathbf{r}_c; V) \ln \left[\frac{\Delta_1}{\Delta_2} \right] d\mathbf{r}_w^N d\mathbf{r}_c \right] dV \\
S_{rel,1\rightarrow 2} &= -\beta \langle U_{wc} \rangle_2 + \beta \Delta G \\
\beta \Delta G &= \beta \langle U_{wc} \rangle_2 + S_{rel,1\rightarrow 2}
\end{aligned} \tag{B.4}$$

B.2 Free energy to solvate a hard-sphere colloid in a monatomic ideal gas

We present an example solvation free energy calculation using the relative entropy approach that highlights how the pressure-volume work is included in the relative entropy of solvation. For solvating a hard-sphere in an ideal gas in the isothermal-isobaric ensemble, the colloid-solvent potential energy is zero if no particles are in the hard sphere volume and infinity otherwise. If the volume of the system is smaller than the volume of the hard-sphere, the weight for that configuration is zero. This means that the average potential energy between the colloid and ideal gas is zero and the free energy difference is equal to the relative entropy. In this case, we can write the relative entropy as

$$\beta \Delta G = S_{rel,1\rightarrow 2} = \ln \left[\frac{\Delta_1}{\Delta_2} \right] \tag{B.5}$$

$$S_{rel,1 \rightarrow 2} = \ln \left[\frac{\int_0^{\text{inf}} e^{-\beta p V} V^{N+1} dV}{\int_0^{\text{inf}} e^{-\beta p V} V \left[\int e^{-\beta U_{wc}} d\mathbf{r}^N \right] dV} \right] \quad (\text{B.6})$$

In the numerator, each of the N ideal gas particles contributes a volume term, as well as the colloid in the ideal gas state. Since only positions of the ideal gas particles relative to the colloid matter, integration of the colloid position over the volume yields the V term in the denominator.

As noted previously, the effect of allowing the hard-sphere colloid to interact with ideal gas particles is to restrict the volume over which the ideal gas may be placed. The effect of the integral over positions of ideal gas particles is to restrict the volume integral from the size of the colloid, $V_c = \frac{4}{3}\pi R_c^3$, to infinity, giving an overall available volume of $V - V_c$ per particle.

$$S_{rel,1 \rightarrow 2} = \ln \left[\frac{\int_0^{\text{inf}} e^{-\beta p V} V^{N+1} dV}{\int_{V_c}^{\text{inf}} e^{-\beta p V} (V - V_c)^N dV} \right] \quad (\text{B.7})$$

The numerator may be readily evaluated by substituting $x = \beta p V$ and re-writing the resulting integral as a Gamma function, $\Gamma(N) = \int_0^{\text{inf}} x^{(N-1)} e^{-x} dx = (N-1)!$

$$S_{rel,1 \rightarrow 2} = \ln \left[\frac{(\beta p)^{-(N+2)} \Gamma(N+2)}{\int_{V_c}^{\text{inf}} e^{-\beta p V} (V - V_c)^N dV} \right] \quad (\text{B.8})$$

In the denominator, we first make the substitution $V' = V - V_c$

$$S_{rel,1 \rightarrow 2} = \ln \left[\frac{(\beta p)^{-(N+2)} \Gamma(N+2)}{\int_0^{\text{inf}} e^{-\beta p (V'+V_c)} (V' + V_c) V'^N dV'} \right] \quad (\text{B.9})$$

And next we rearrange and let $x = \beta p V'$ so that

$$\begin{aligned}
S_{rel,1 \rightarrow 2} &= \ln \left[\frac{(\beta p)^{-(N+2)} \Gamma(N+2)}{e^{-\beta p V_c} \int_0^{\text{inf}} e^{-\beta p V'} V'^{N+1} dV' + V_c e^{-\beta p V_c} \int_0^{\text{inf}} e^{-\beta p V'} V'^N dV'} \right] \\
&= \ln \left[\frac{(\beta p)^{-(N+2)} \Gamma(N+2)}{e^{-\beta p V_c} (\beta p)^{-(N+2)} \int_0^{\text{inf}} e^{-x} x^{N+1} dx + V_c e^{-\beta p V_c} (\beta p)^{-(N+1)} \int_0^{\text{inf}} e^{-x} x^N dx} \right] \\
&= \ln \left[\frac{(\beta p)^{-(N+2)} \Gamma(N+2)}{e^{-\beta p V_c} \left((\beta p)^{-(N+2)} (N+1)! + V_c (\beta p)^{-(N+1)} N! \right)} \right] \\
&= \ln \left[e^{\beta p V_c} \frac{(N+1)!}{(N+1)! + \beta p V_c N!} \right] \\
&= \ln \left[e^{\beta p V_c} \frac{N+1}{N+1 + \beta p V_c} \right] \\
&= \beta p V_c + \ln \left[\frac{N+1}{N+1 + \beta p V_c} \right] \\
&\approx \beta p V_c
\end{aligned} \tag{B.10}$$

This shows that the relative entropy includes the pressure-volume work involved in the solvation process because the restriction of volume accessible to the solvent is explicitly taken into account when integrating over solvent configurational degrees of freedom. This results in a contribution from the work to grow the cavity at constant pressure (the first term in the final expression) and a second term that is approximately zero for large N (or even when $N \gg \beta p V_c$, which only requires that $\beta p V_c$ be small and is exact without invoking the macroscopic limit). This work to grow the cavity is not captured by looking at local degrees of freedom between water molecules, such as the three-body angle, but is negligibly small for even the largest of colloids (see Table C.1).

Appendix C

Additional structural and
thermodynamic data associated with
water three-body angle distributions

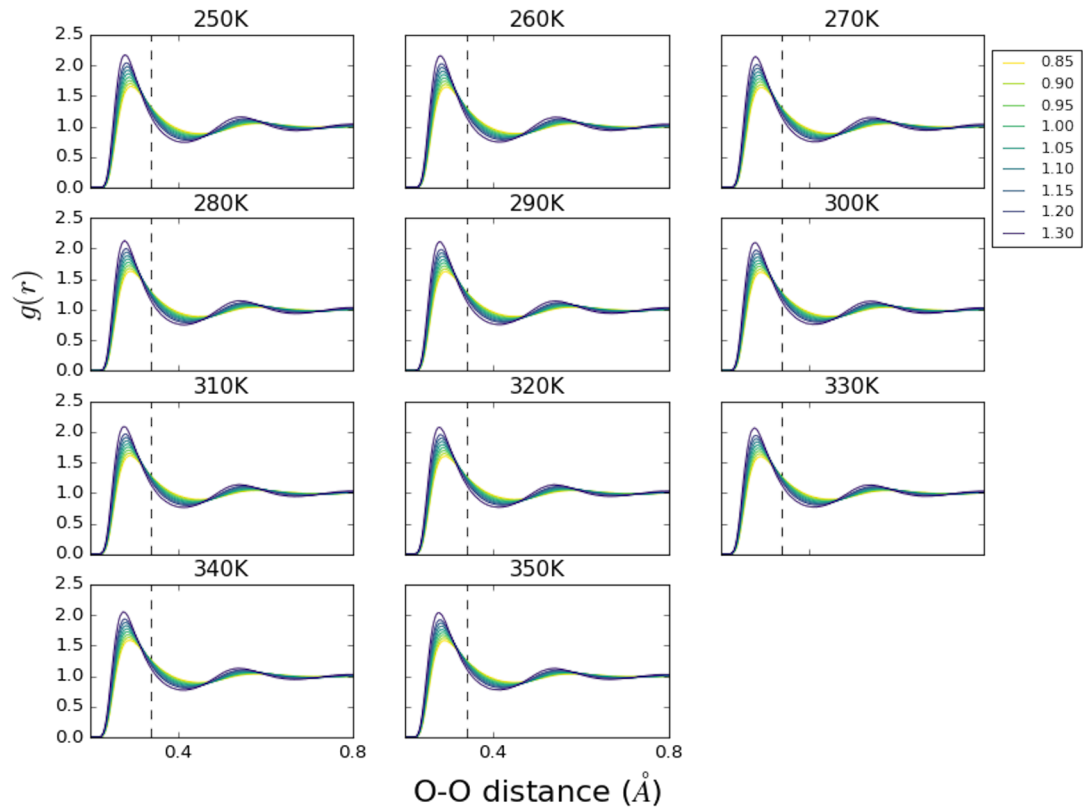


Figure C.1: Radial distribution functions for the LJ water model at various temperatures and densities (in g/cm^3). Vertical dashed lines show the cut-off of 0.34 nm used to define neighbors for calculating three-body angle distributions in bulk TIP4P/2005 and LJ water.

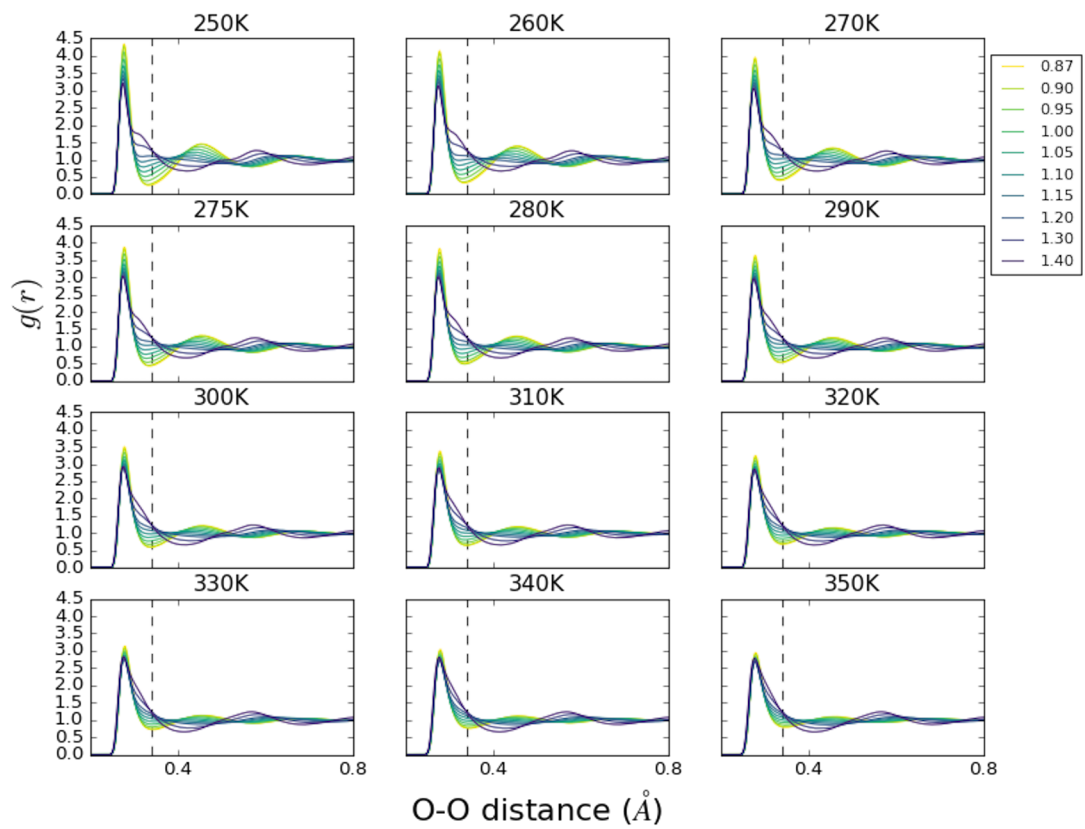


Figure C.2: Radial distribution functions for the TIP4P/2005 water model at various temperatures and densities (in g/cm^3).

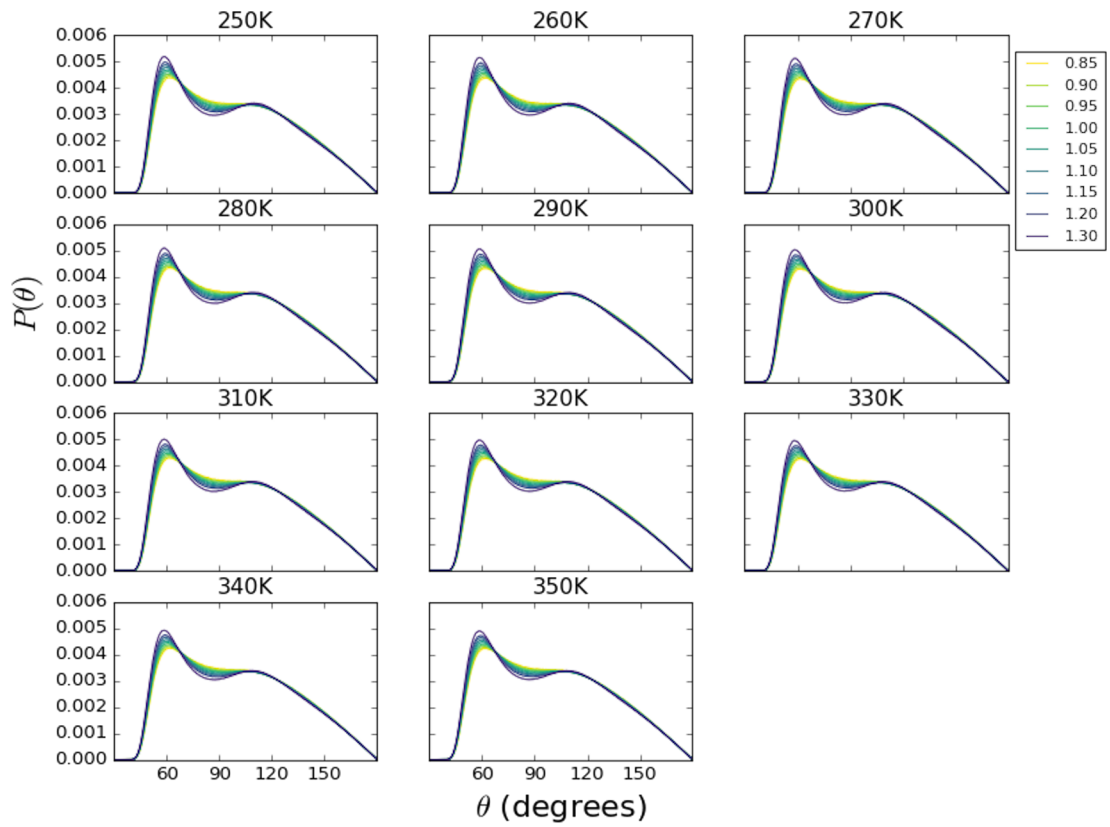


Figure C.3: Three-body angle distributions for the LJ water model at various temperatures and densities (in g/cm^3).

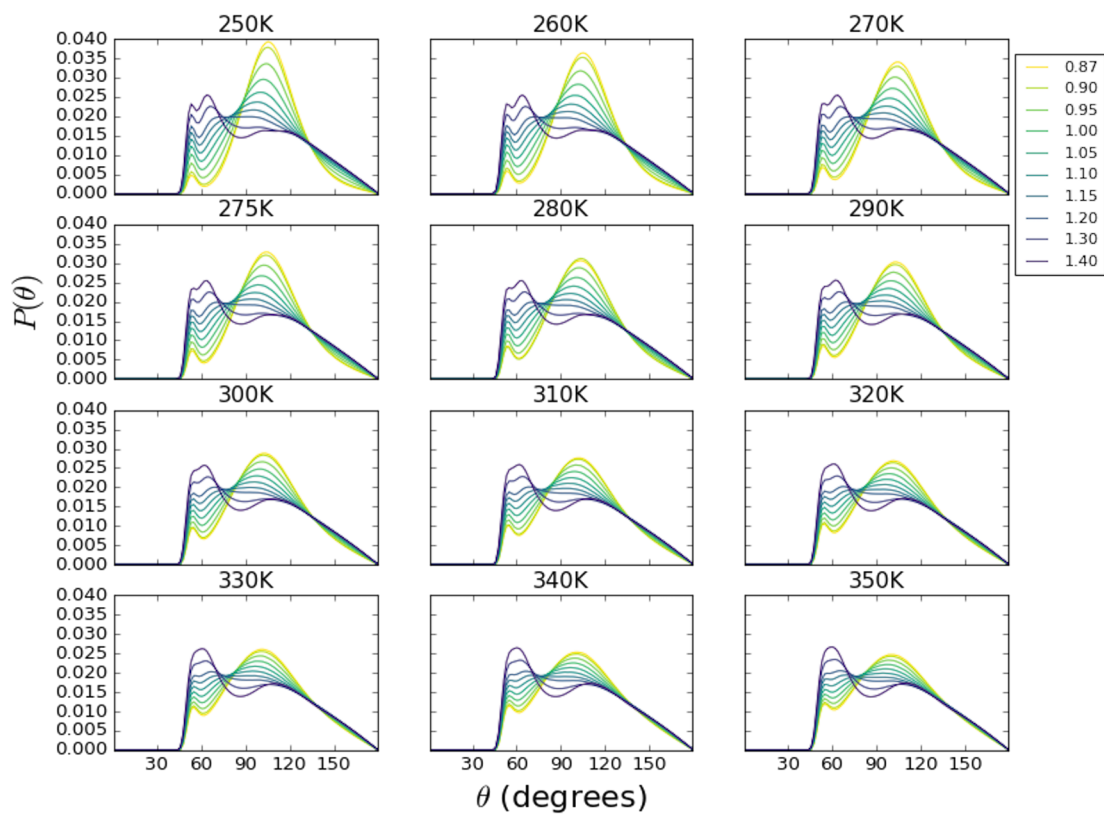


Figure C.4: Three-body angle distributions for the TIP4P/2005 water model at various temperatures and densities (in g/cm^3).

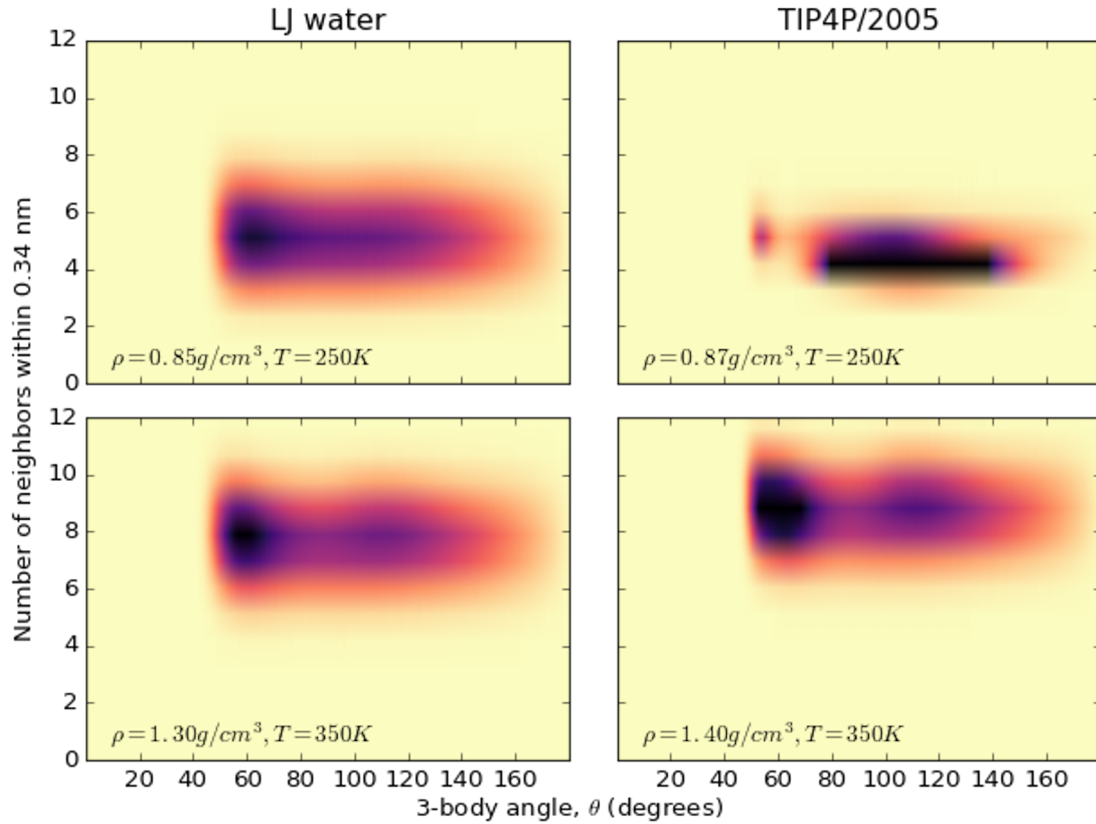


Figure C.5: Two-dimensional probability densities in the number of neighbors within the 0.34 nm cutoff and the three-body angles are shown for the lowest-density, lowest-temperature and highest-density, highest-temperature conditions for both LJ and TIP4P/2005 water. For the low-density, low-temperature TIP4P/2005 distribution (top right), the peak at low angles (around 50°), is due to rare configurations involving 5 neighbors, where the tetrahedral peak around 109.5° is still prevalent but a neighbor interstitial to this tetrahedral arrangement also contributes low angles.

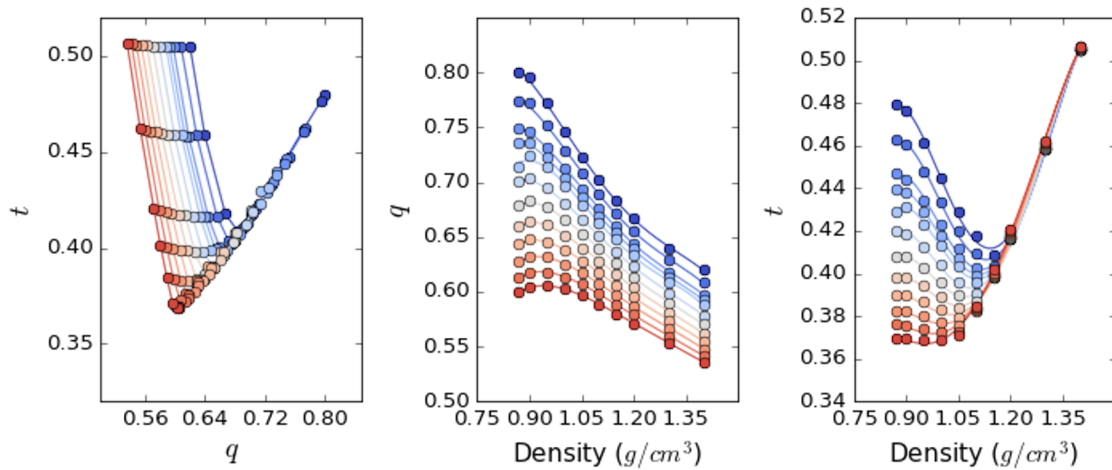


Figure C.6: Translational order parameters t and q are computed as described by Errington and Debenedetti.¹ Isotherms in t - q space are shown at far left, with isotherms as a function of density shown for q and t shown in center and right, respectively. The low-density structural anomaly boundary occurs where q exhibits maxima, and the high-density boundary occurs where t shows minima. In the center and right plots, solid lines are fits of the points shown to 4th-order Legendre polynomials, which are used to determine the minima/maxima and the anomaly bounds. In the center figure at a temperature of 280 K (5th from lowest), it appears that a maximum has been found; however, the lowest density point actually represents a decrease in q due to nucleation of voids in the liquid while it is very close to its metastability bound. Thus, the same q behavior is exhibited whether the mechanical stability or structural anomaly boundary is crossed.

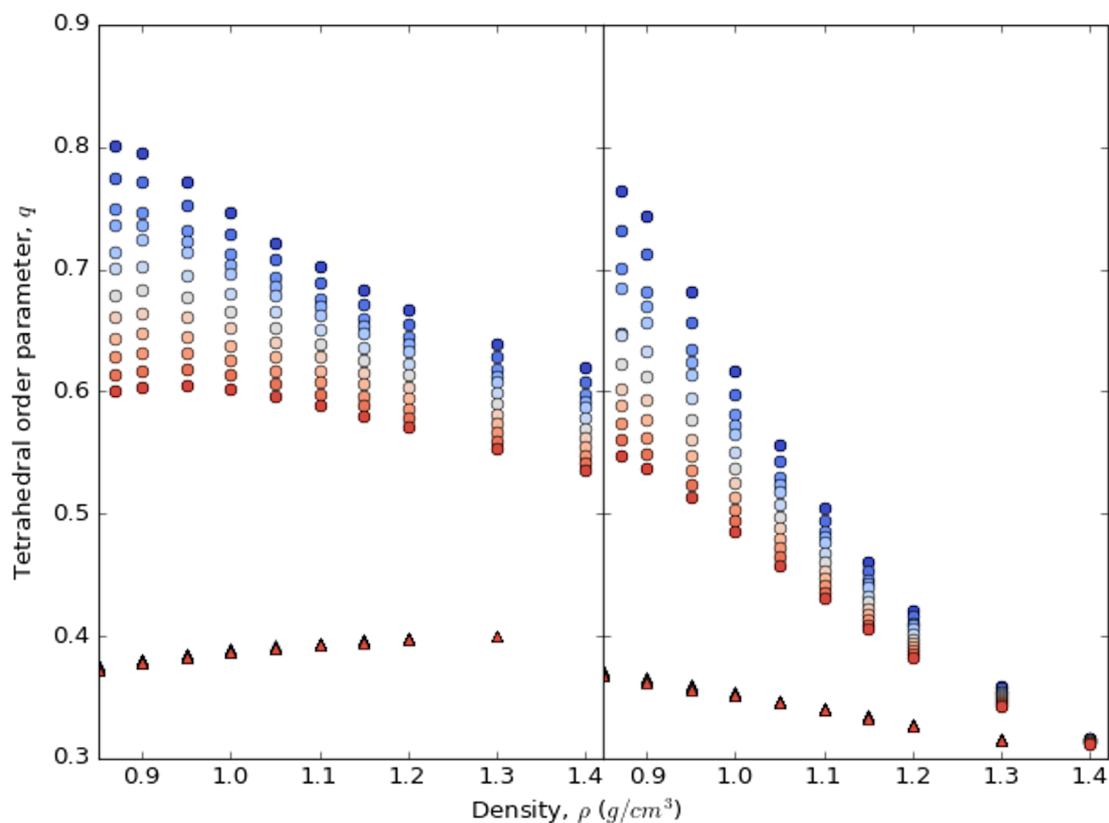


Figure C.7: Tetrahedral order parameters isotherms computed via the definition from Errington and Debenedetti,¹ using 4 nearest neighbors (left) and computed from the distributions of three-body angles using a fixed distance cut-off (right). Circles are for TIP4P/2005 water and triangles show LJ water. When using a fixed cut-off, no maxima in q are observed, except in the case of the simulation at 0.87 g/cm^3 and 280 K where voids are observed in the metastable liquid. When considering exactly 4 neighbors, some neighbors are occasionally included that reside outside the first solvation shell, leading to decreases in q at low densities where there is non-negligible probability of seeing 3 neighbors within the fixed cut-off instead (Fig. C.5). At high densities, q is much lower due to inclusion of greater numbers of neighbors that have angles different from the tetrahedral angle of 109.5° .

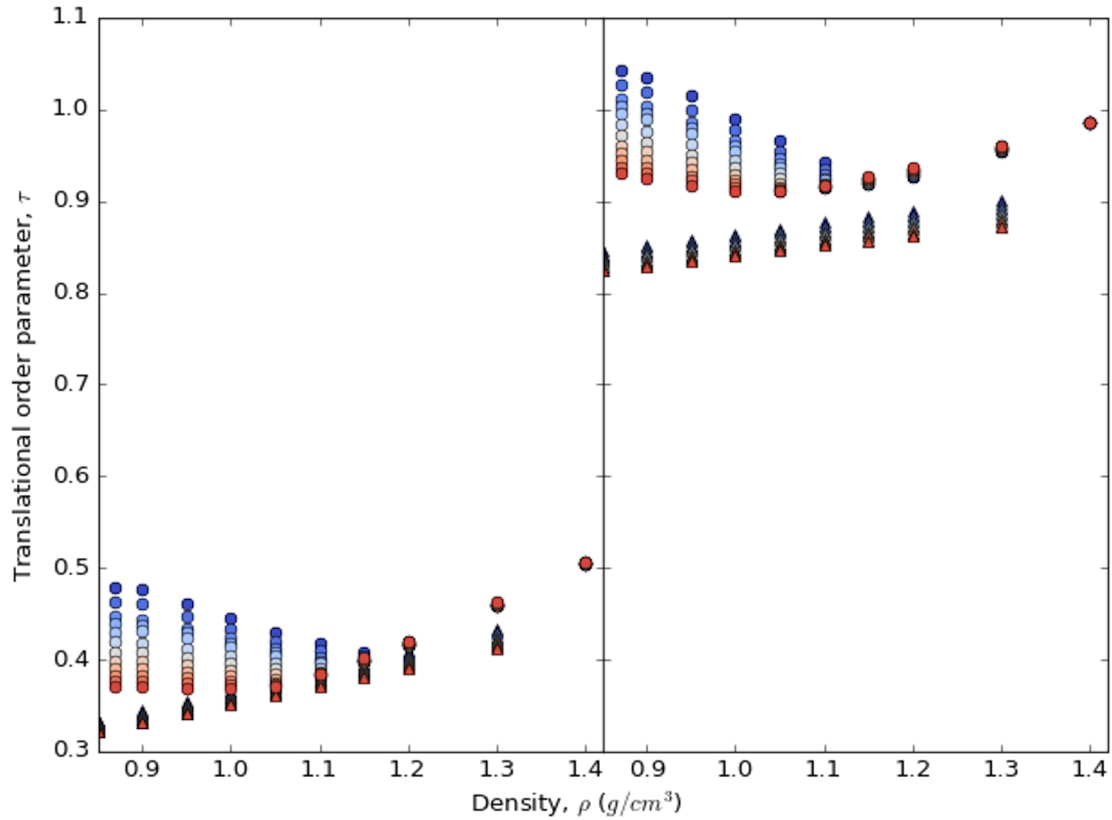


Figure C.8: Translational order parameter distributions are shown using a large cutoff identical to that defined by Errington and Debenedetti¹ (left), and a small cutoff of 0.34 nm equivalent to the neighbor cutoff for computing three-body angles (right). TIP4P/2005 water is represented by circles while LJ water is shown with triangles. Though the order parameter is shifted to different values, the qualitative behavior is almost identical, with minima of isotherms occurring at effectively the same densities. In this way, either cutoff definition produces the same result for the high-density structural anomaly boundary.

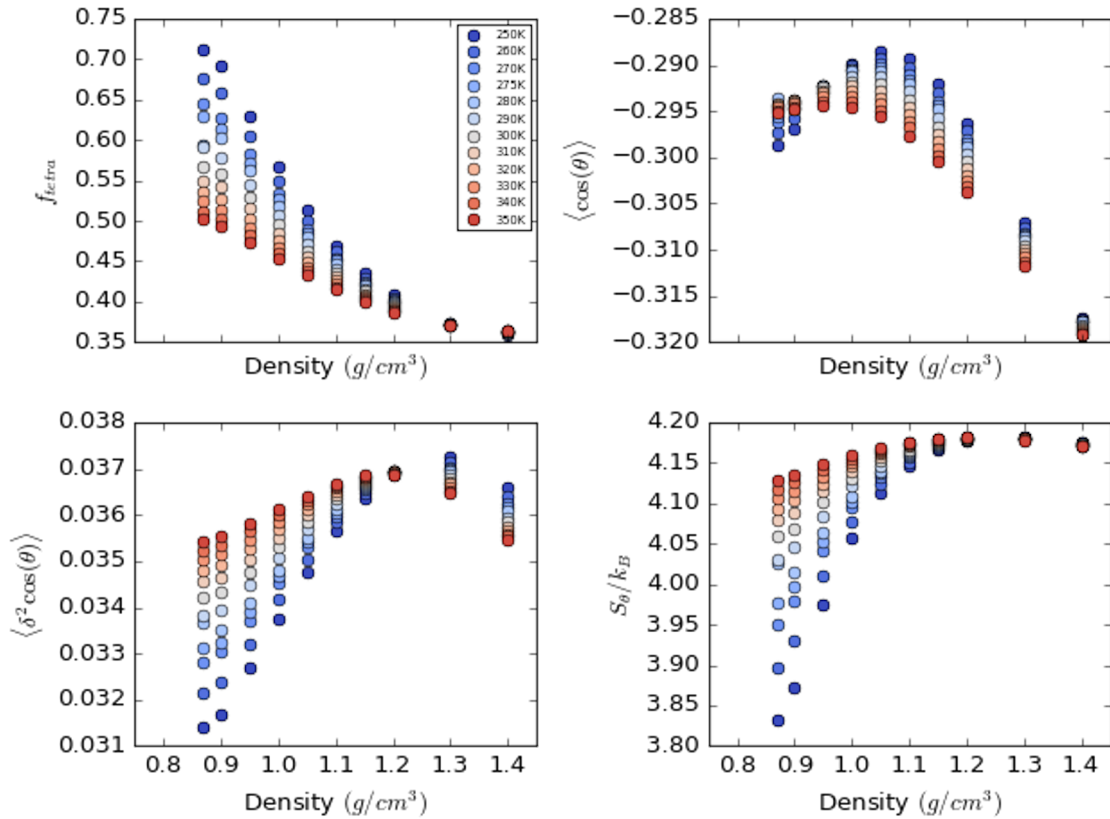


Figure C.9: Various metrics based on the three-body distribution² are shown for isotherms versus density. f_{tetra} is the fraction of the distribution that is tetrahedral as defined by the weight of the distribution between 90.0° and 130.0° . The top right shows the average of the cosine of the three-body angle within the tetrahedral region defined above. The bottom left shows the variance of the cosine of three-body angles within the tetrahedral region. The bottom right presents the entropy of the three-body angle distribution. Maxima in the three-body angle distribution entropies are suggestive of the high-density structural anomaly bound. However, no metric seems to suggest a lower-density bound.

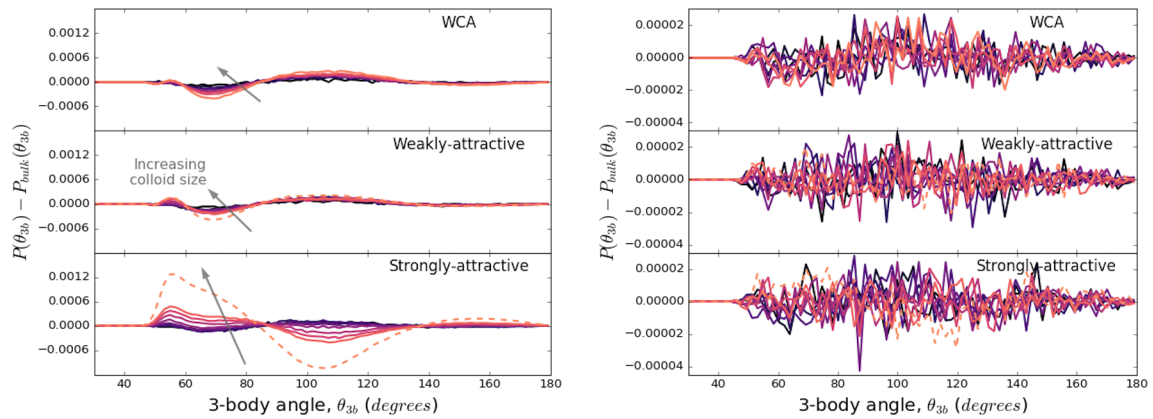


Figure C.10: Perturbations of bulk water three-body angle distributions in the second hydration shell (left) and a shell far enough away to achieve bulk-like behavior, between 2.4 and 2.6 nm from the colloid center (right). The magnitude of second-shell perturbations is almost an order of magnitude smaller than the first shell and the trends are distinct from those observed in the first shell. Far from the colloid, there is only low-magnitude noise compared to the reference distribution.

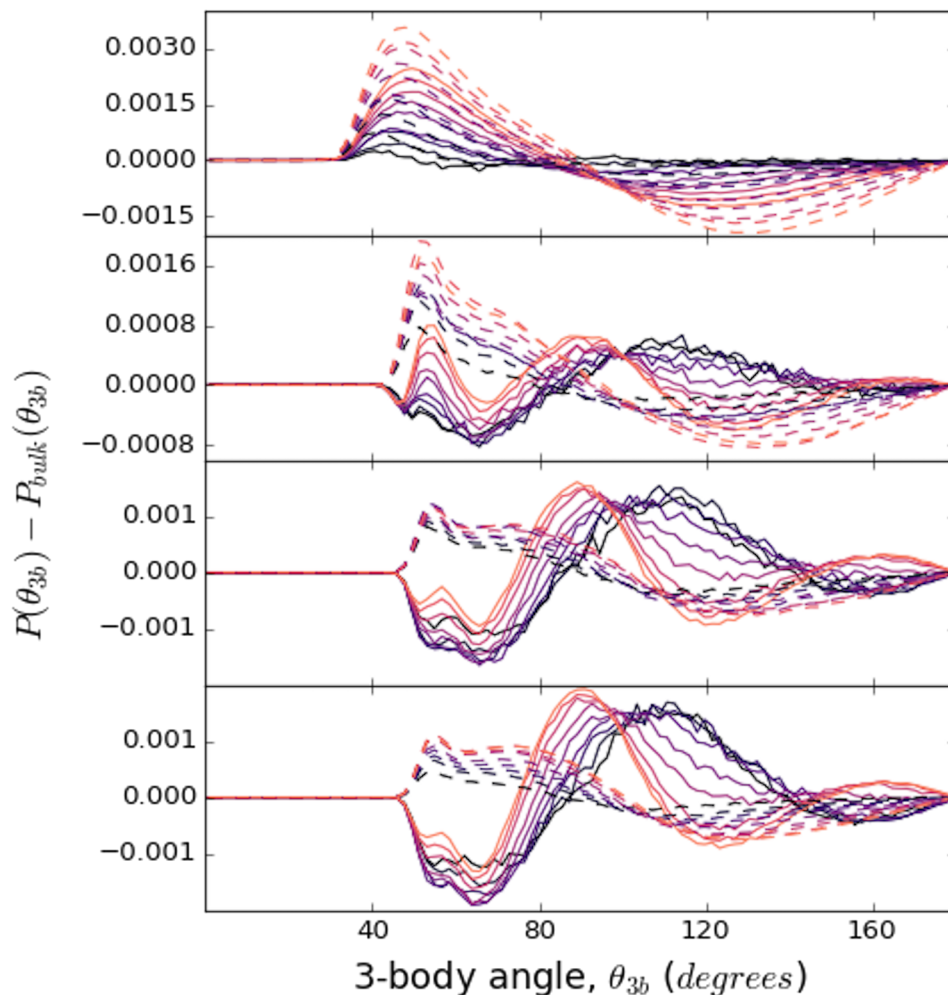


Figure C.11: Perturbations from bulk for three-body angle distributions in the first hydration shell of WCA colloids are shown as solid lines for LJ, TIP3P, TIP4P-Ew, and TIP4P/2005 water models (from top to bottom). Small systems are in darker purple, while larger systems are lighter orange. Dashed lines represent distributions with applied geometric constraints rather than physically present colloids. Specifically, spherical shells corresponding to the same radial range between the hard-sphere radii and RDF first minima of WCA colloids are defined to select water oxygens in bulk-water simulations, with only waters at positions greater than the radius considered when computing the distributions. This effectively imposes geometric constraints without any colloids present. In the case of the LJ fluid, it is notable that the physical colloids perturb the distribution less than the geometric constraints. This is because the physical colloids, though purely repulsive, exhibit a soft interface while the constraints imposed are a perfectly hard interface defined by the effective hard-sphere radius.

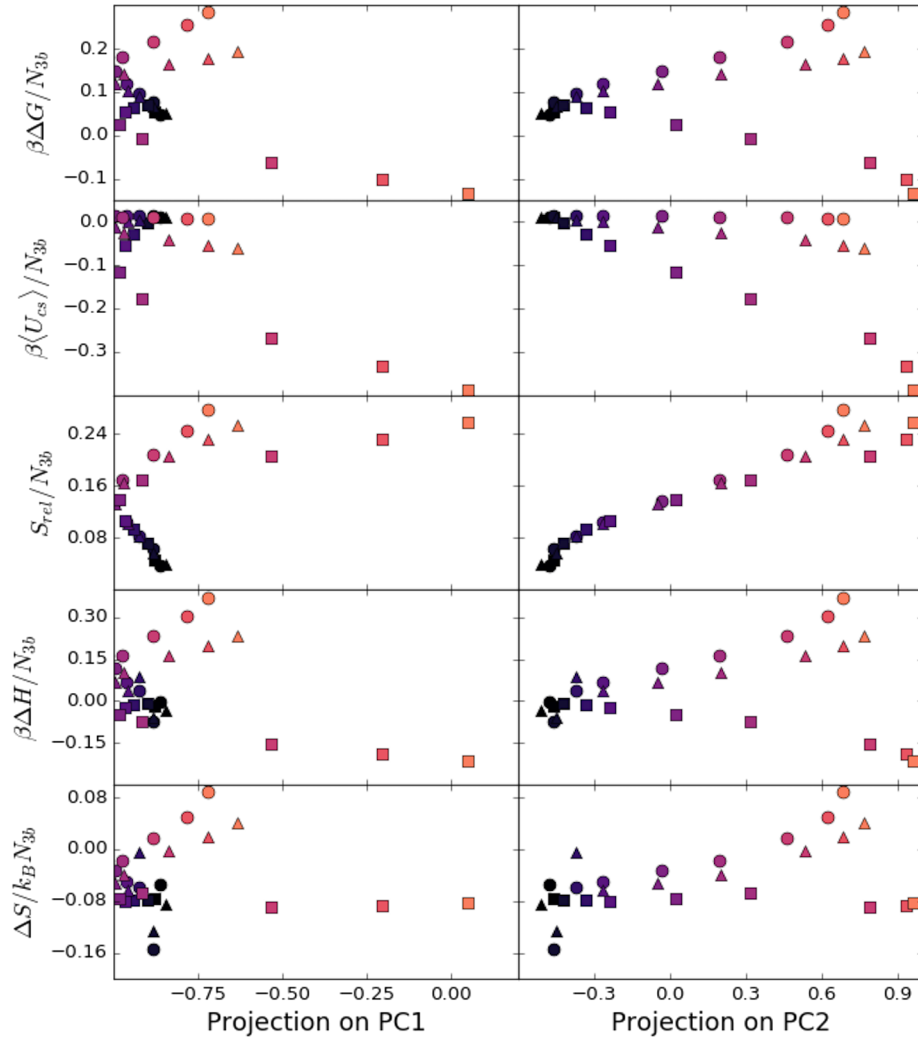


Figure C.12: Thermodynamic quantities associated with solvation, normalized to the average number of three-body angles formed by waters in the first hydration shell, are shown against projections onto the first two principal components of perturbations by all colloids. Circles represent WCA colloids, triangles low-density colloids, and squares high-density colloids. Smaller colloids are shown with dark purple while large colloids are brighter orange. Only for the relative entropy do we see evidence of a collapse of the projections across all colloid types. This collapse breaks down for larger colloid sizes, possibly indicating a length-scale cross-over or a point at which attractions become increasingly significant.

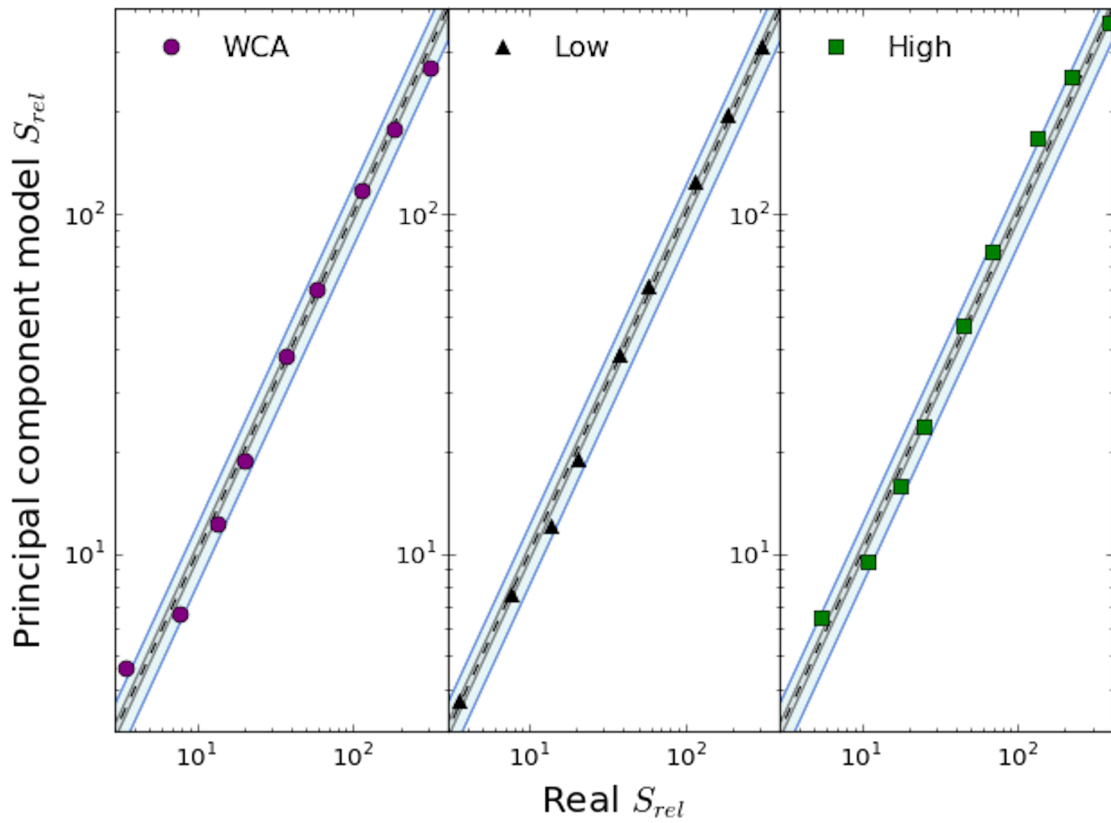


Figure C.13: Fits to the principal-component-based model for the relative entropy (with cross-terms) are shown for independently fitting to each of the colloid types. The gray shaded region represents 5% error while the blue-shaded region represents 20% error. The fits are excellent even at small colloid sizes. Compared to the fit over all colloids (Fig. 3.8), the individual fits are more accurate at large colloid sizes.

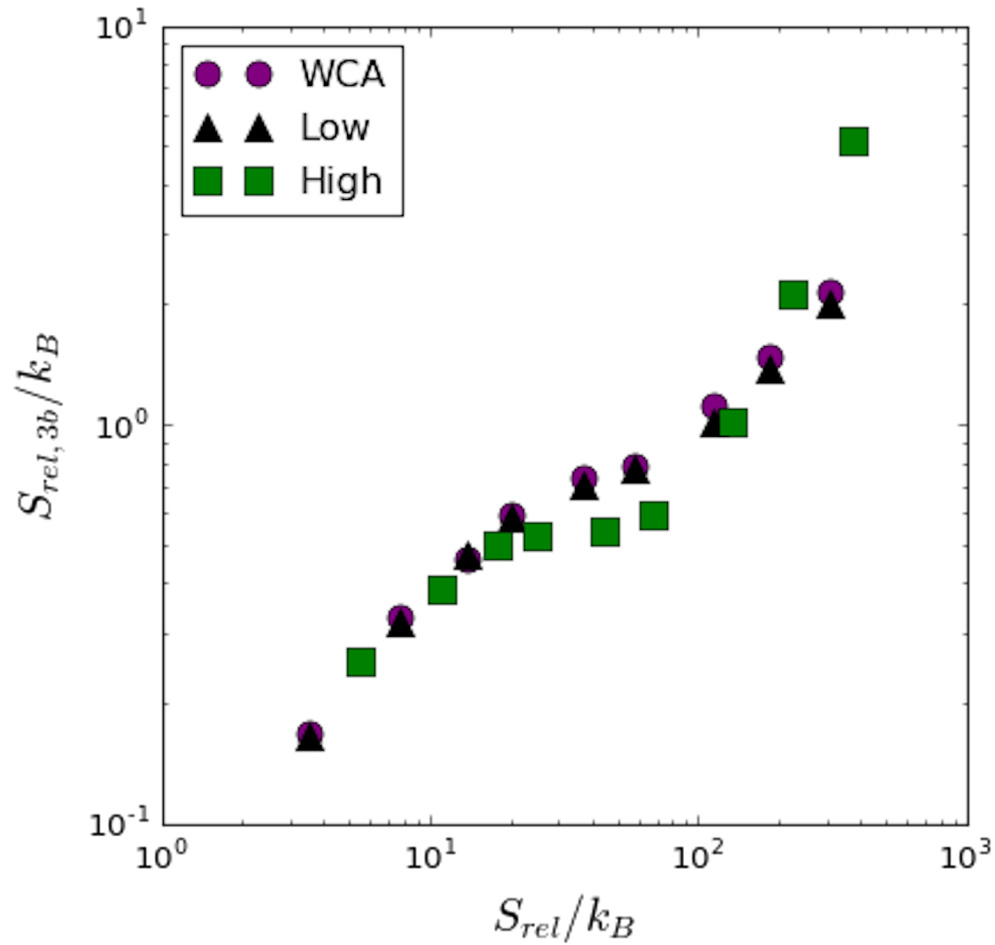


Figure C.14: Three-body relative entropy in the first solvation shell of colloids is shown against the total solvation relative entropy. For small colloid sizes (smaller relative entropies), all colloid types fall along the same line. For colloids larger than ~ 0.5 nm in hard-sphere radius, the behavior becomes qualitatively different, with slopes for large colloids significantly diverging.

R_{hs} (nm)	ΔG_{rep}	ΔG_{attr}	$p\Delta V$	ΔH	ΔS	$\langle U_{cs} \rangle$	S_{rel}	ΔG
WCA colloids								
0.258	4.71 ± 0.08	N/A	0.00 ± 0.00	-0.52 ± 5.18	-5.23 ± 5.18	1.17 ± 0.04	3.54 ± 0.08	4.71 ± 0.08
0.318	9.60 ± 0.10	N/A	0.00 ± 0.00	-9.30 ± 5.03	-18.90 ± 5.03	1.85 ± 0.05	7.74 ± 0.12	9.60 ± 0.10
0.374	16.10 ± 0.13	N/A	0.00 ± 0.00	6.41 ± 5.06	-9.64 ± 5.07	2.43 ± 0.06	13.63 ± 0.14	16.10 ± 0.13
0.422	22.99 ± 0.15	N/A	0.00 ± 0.00	13.30 ± 4.90	-9.68 ± 4.91	2.78 ± 0.06	20.21 ± 0.16	22.99 ± 0.15
0.520	41.19 ± 0.30	N/A	0.01 ± 0.00	32.24 ± 4.95	-8.95 ± 4.96	3.69 ± 0.10	37.50 ± 0.31	41.19 ± 0.30
0.612	62.06 ± 0.44	N/A	0.02 ± 0.00	55.79 ± 4.91	-6.27 ± 4.93	4.14 ± 0.08	57.92 ± 0.45	62.06 ± 0.44
0.804	119.49 ± 0.71	N/A	0.05 ± 0.00	129.48 ± 4.93	9.99 ± 4.98	5.33 ± 0.10	114.16 ± 0.72	119.49 ± 0.71
0.990	191.37 ± 1.13	N/A	0.09 ± 0.00	229.20 ± 4.92	37.82 ± 5.05	6.79 ± 0.11	184.58 ± 1.14	191.37 ± 1.13
1.244	316.99 ± 1.35	N/A	0.20 ± 0.00	414.69 ± 4.86	97.68 ± 5.05	8.02 ± 0.12	308.97 ± 1.36	316.99 ± 1.35
Low-density colloids								
0.258	4.71 ± 0.08	-0.06 ± 0.00	0.00 ± 0.00	-2.93 ± 4.95	-7.57 ± 4.95	1.12 ± 0.01	3.52 ± 0.08	4.64 ± 0.08
0.318	9.60 ± 0.10	-0.44 ± 0.00	0.00 ± 0.00	-7.87 ± 4.96	-17.03 ± 4.96	1.41 ± 0.01	7.75 ± 0.10	9.16 ± 0.10
0.374	16.10 ± 0.13	-1.38 ± 0.00	0.00 ± 0.00	14.10 ± 5.16	-0.582 ± 5.16	0.99 ± 0.01	13.69 ± 0.13	14.68 ± 0.13
0.422	22.99 ± 0.15	-2.71 ± 0.00	0.00 ± 0.00	7.78 ± 4.94	-12.49 ± 4.94	0.10 ± 0.01	20.17 ± 0.15	20.28 ± 0.15
0.520	41.19 ± 0.30	-6.97 ± 0.00	0.01 ± 0.00	19.30 ± 4.94	-14.92 ± 4.94	-3.33 ± 0.02	37.54 ± 0.30	34.22 ± 0.30
0.612	62.06 ± 0.44	-12.56 ± 0.01	0.02 ± 0.00	35.67 ± 4.85	-13.83 ± 4.87	-8.12 ± 0.02	57.62 ± 0.44	49.50 ± 0.44
0.804	119.49 ± 0.71	-28.62 ± 0.02	0.04 ± 0.00	89.07 ± 5.06	-1.80 ± 5.11	-22.68 ± 0.02	113.55 ± 0.71	90.87 ± 0.71
0.990	191.37 ± 1.13	-48.97 ± 0.04	0.09 ± 0.00	158.35 ± 4.96	15.95 ± 5.10	-41.87 ± 0.03	184.28 ± 1.13	142.40 ± 1.13
1.244	316.99 ± 1.35	-82.58 ± 0.06	0.18 ± 0.00	284.92 ± 4.86	50.51 ± 5.04	-74.81 ± 0.04	309.22 ± 1.35	234.41 ± 1.35
High-density colloids								
0.291	7.13 ± 0.09	-0.367 ± 0.00	0.00 ± 0.00	-2.48 ± 5.05	-9.22 ± 5.05	1.22 ± 0.01	5.51 ± 0.09	6.74 ± 0.09
0.350	13.08 ± 0.11	-2.45 ± 0.00	0.00 ± 0.00	-1.10 ± 4.88	-11.73 ± 4.88	-0.36 ± 0.01	11.00 ± 0.12	10.63 ± 0.11
0.404	20.20 ± 0.14	-7.66 ± 0.01	0.00 ± 0.00	-2.38 ± 5.10	-14.92 ± 5.09	-5.25 ± 0.01	17.79 ± 0.14	12.54 ± 0.14
0.451	27.61 ± 0.16	-15.10 ± 0.01	0.01 ± 0.00	-5.98 ± 4.82	-18.49 ± 4.82	-12.50 ± 0.01	25.01 ± 0.16	12.51 ± 0.16
0.547	47.13 ± 0.33	-39.15 ± 0.03	0.01 ± 0.00	-16.37 ± 5.34	-24.34 ± 5.35	-36.75 ± 0.02	44.72 ± 0.33	7.98 ± 0.33
0.637	68.74 ± 0.46	-71.90 ± 0.05	0.02 ± 0.00	-30.37 ± 4.92	-27.22 ± 4.94	-70.73 ± 0.03	67.57 ± 0.47	-3.15 ± 0.47
0.828	127.84 ± 0.74	-169.59 ± 0.11	0.04 ± 0.00	-100.19 ± 4.97	-58.43 ± 5.03	-175.88 ± 0.06	134.13 ± 0.75	-41.75 ± 0.75
1.013	201.77 ± 1.17	-298.78 ± 0.18	0.07 ± 0.00	-180.91 ± 5.03	-83.90 ± 5.17	-319.59 ± 0.08	222.58 ± 1.18	-97.01 ± 1.18
1.267	329.07 ± 1.39	-524.99 ± 0.24	0.15 ± 0.00	-318.99 ± 5.01	-123.06 ± 5.21	-576.72 ± 0.11	380.80 ± 1.42	-195.92 ± 1.41

Table C.1: Various thermodynamic components of free energies of solvation for different colloids. All units in $k_B T$ unless otherwise noted. Uncertainties for ΔG_{rep} and ΔG_{attr} derive from estimates of uncertainty using uncorrelated potential energies from the pymbar package.³ Uncertainties for ΔH and $\langle U_{cs} \rangle$ are standard deviations from averaging all uncorrelated potential energies (and pressure-volume terms in the former case) in the trajectory. Uncertainties for ΔG , ΔS , and S_{rel} are computed via standard error propagation.

References

- [1] Errington JR, Debenedetti PG. Relationship between structural order and the anomalies of liquid water. *Nature*. 2001; 409(6818): 318–321. ISBN: 0028-0836.
- [2] Chaimovich A, Shell MS. Tetrahedrality and structural order for hydrophobic interactions in a coarse-grained water model. *Physical Review E*. 2014; 89(2): 022140. ISBN: 1550-2376 (Electronic) 1539-3755 (Linking).
- [3] Chodera JD, Shirts MR. A python implementation of the multistate Bennet acceptance ratio (MBAR).

Appendix D

Additivity of dynamics in simple Markov models

From a continuum point of view, the dynamics of heterogeneous systems can be explored with models as simple as continuous time Markov chains. Imagine a one-dimensional system with two types of sites, either 1 or 2, that can be arbitrarily arranged along a line that wraps periodically in space (Figure D.1). Note that the dimension and boundary conditions of the problem are chosen for convenience of illustration, though in principle the results are more general. Let us assume that if a random walker is at a type 1 site, the rate with which it leaves to either of the neighboring states is a , while this rate is b for

type 2 sites. This results in a transition rate matrix corresponding to Figure D.1 (top) as

$$\mathbf{R} = \begin{bmatrix} -2a & a & 0 & 0 & 0 & a \\ a & -2a & a & 0 & 0 & 0 \\ 0 & a & -2a & a & 0 & 0 \\ 0 & 0 & b & -2b & b & 0 \\ 0 & 0 & 0 & b & -2b & b \\ b & 0 & 0 & 0 & b & -2b \end{bmatrix} \quad (\text{D.1})$$

If we require that detailed balance be satisfied, the equilibrium occupation probability of any site 1 is related to the probability of any site 2 via

$$\wp_2 = \frac{a}{b} \wp_1 \quad (\text{D.2})$$

which fully specifies the problem if we assume that we start from an ensemble in which the equilibrium probabilities of each state are satisfied. For any time t , the time evolution of the occupation probabilities follows

$$\frac{d\vec{P}(t)}{dt} = \vec{P} \cdot \mathbf{R} \quad (\text{D.3})$$

which is formally solved by

$$\mathbf{P}(t) = e^{\mathbf{R}t} \quad (\text{D.4})$$

If we then measure the slope of the average mean-squared displacement at long times, the diffusivity is given by

$$\begin{aligned} D &= \frac{\langle d^2(t + \delta t) \rangle - \langle d^2(t) \rangle}{2\delta t} \\ &= \frac{\sum_i \wp_i \sum_j \mathbf{P}(t + \delta t)_{i,j} (j - i)^2 - \sum_i \wp_i \sum_j \mathbf{P}(t)_{i,j} (j - i)^2}{2\delta t} \end{aligned} \quad (\text{D.5})$$

Note that the time must be long enough that a walker visits domains of both type 1 and 2, but short enough that the periodicity scale of the system remains comparatively large (this difficulty vanishes for infinitely sized systems). Though the above expressions are analytical, numerical solutions for arbitrary arrangements of sites 1 and 2 that address finite-size effects are easiest. Such results show that the diffusivity does not depend on the arrangement of sites of type 1 and 2, illustrating that a macroscopic model holds in which surface types have additive effects (i.e., the influence of a surface type is independent of neighboring site identities).

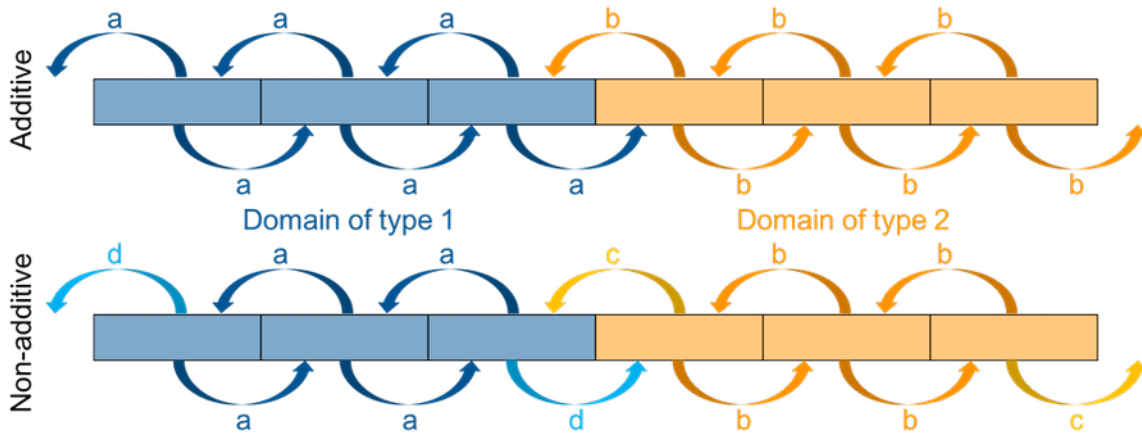


Figure D.1: Schematics of 1D Markov models with distinct transition rates that result in average diffusivities that are additive (top) and non-additive (bottom).

Additivity breaks down, however, as soon as we introduce edge effects between the domains, as shown in Figure D.1 (bottom). This is the equivalent of creating new types of sites with asymmetric rates of transition to adjacent sites. The asymmetry allows specific patterning of domains to influence the average diffusivity since the number of such edges can vary from one arrangement to another. The effect is similar to what is observed for interfaces with spatial chemical heterogeneity as described in the main text. At each surface site, one can consider analogous rates for moving to any neighboring surface site. If these rates are all the same, depending only on the nature of the site itself and not

where the walker is moving, then additivity will hold. However, it is obvious that the arrangement of other nearby chemical groups can produce rate asymmetry, with subtle effects appearing as the free energy landscape changes across the surface. In this way, molecular scale heterogeneities likely make additive models for dynamical behavior break down, instead requiring brute force atomistic molecular dynamics simulations to evaluate the detailed dynamics.

Appendix E

Genetic algorithm validation and
additional, spatially-resolved
measures of surface hydrophilicity
and water mobility

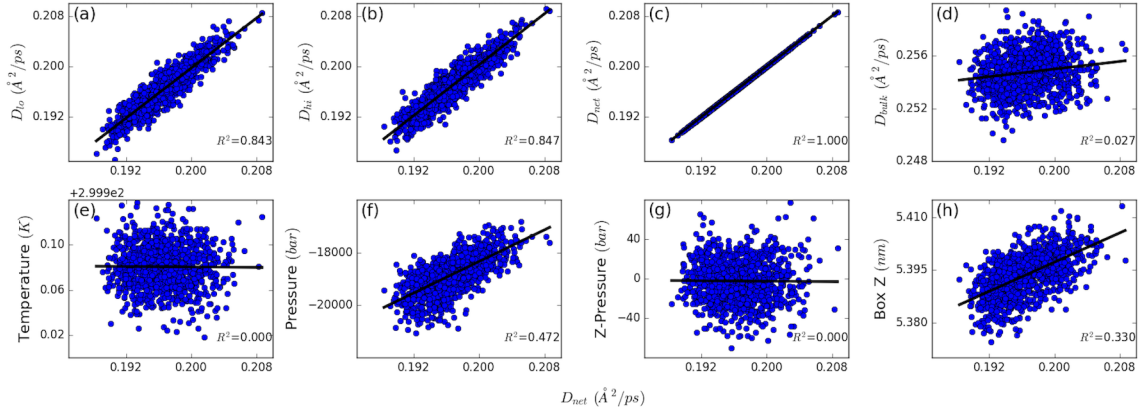


Figure E.1: For cristobalite at a fraction of 0.25 of the maximum possible number of silanol groups, correlations are shown between the measured diffusivity parallel to the plane of the surface (D_{net} on the x-axis of all panels) and (a) the diffusivity measured below the surface (b) the diffusivity measured above the surface (c) D_{net} itself (d) the measured diffusivity far from the surface (e) the average simulation temperature (f) the total simulation pressure (g) the component of simulation pressure perpendicular to the surface and (h) the equilibrated box dimension perpendicular to the surface. Black lines are linear regressions to the data with R^2 values shown on each plot. While the average simulation box length perpendicular to the interface weakly correlates with diffusivity (larger volumes for faster diffusion), this is due to variation in effective surface hydrophobicity that modulates interfacial width at constant pressure. For cristobalite surfaces at low density (here and in Figure E.2), it appears that the total system pressure (including all Cartesian directions) correlates weakly with the mobility of hydration water. Noting that the surface area is fixed in the plane of the interface and not relaxed during constant-pressure equilibration, it is clear that arrangements of silanol groups can affect the in-plane strain of these surfaces by changing the bonding pattern. Thus, a weak correlation is observed between measured diffusivity and total simulation pressure as the pattern of functional groups in cristobalite is adjusted.

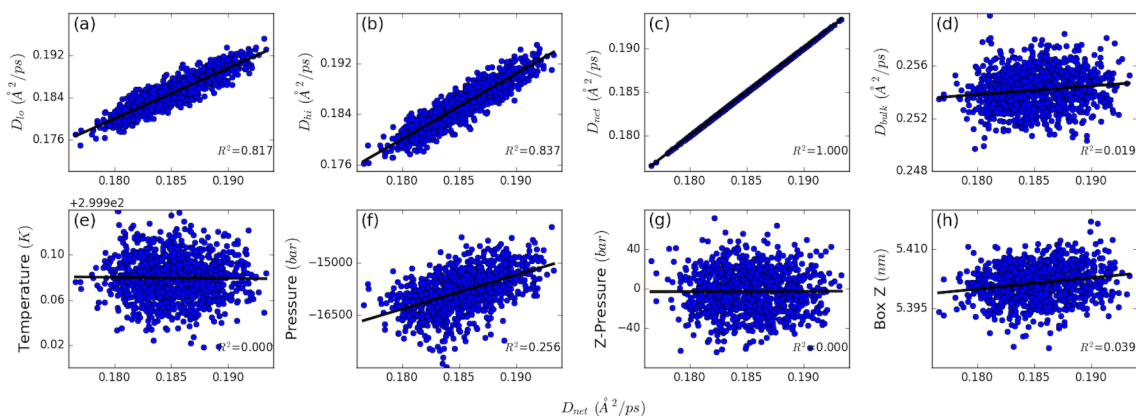


Figure E.2: Same as Figure E.1, but for cristobalite surfaces at a fraction of 0.375 of the maximum possible number of silanol groups.

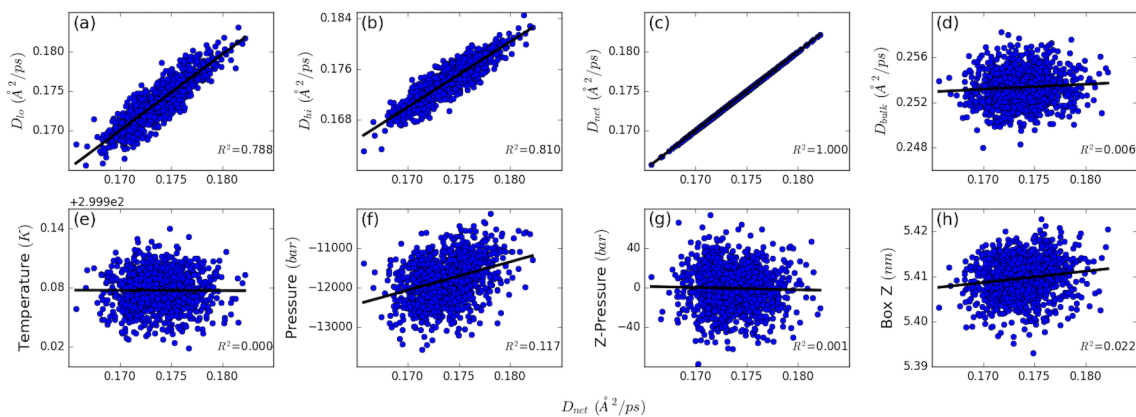


Figure E.3: Same as Figure E.1, but for cristobalite surfaces at a fraction of 0.50 of the maximum possible number of silanol groups.

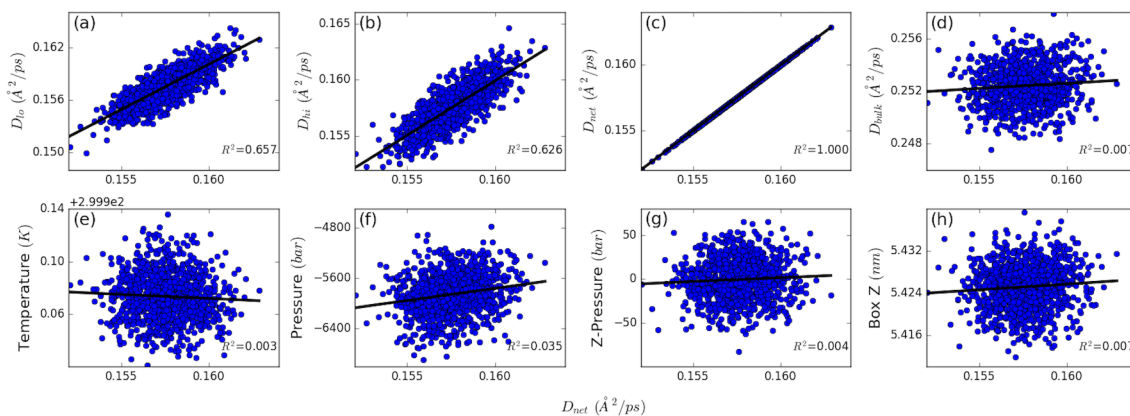


Figure E.4: Same as Figure E.1, but for cristobalite surfaces at a fraction of 0.75 of the maximum possible number of silanol groups.

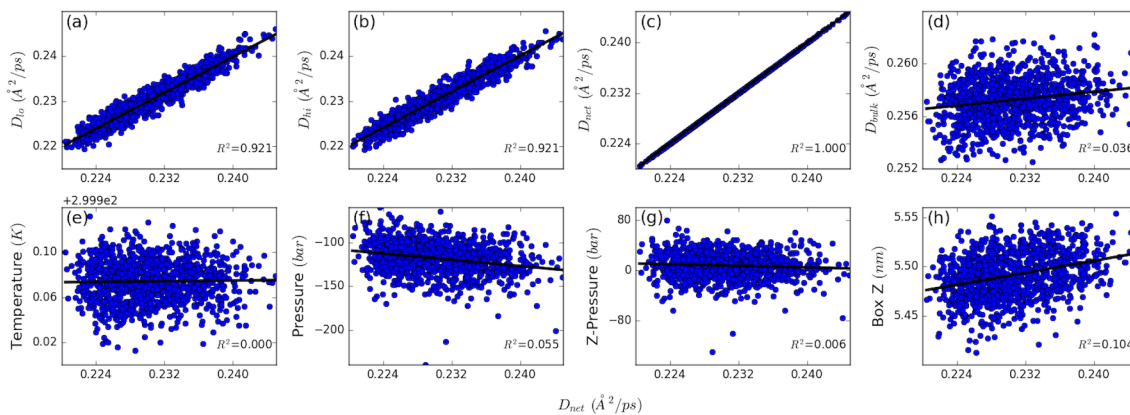


Figure E.5: Same as Figure E.1, but for the SAM surface at an OH-terminated chain fraction of 0.25.

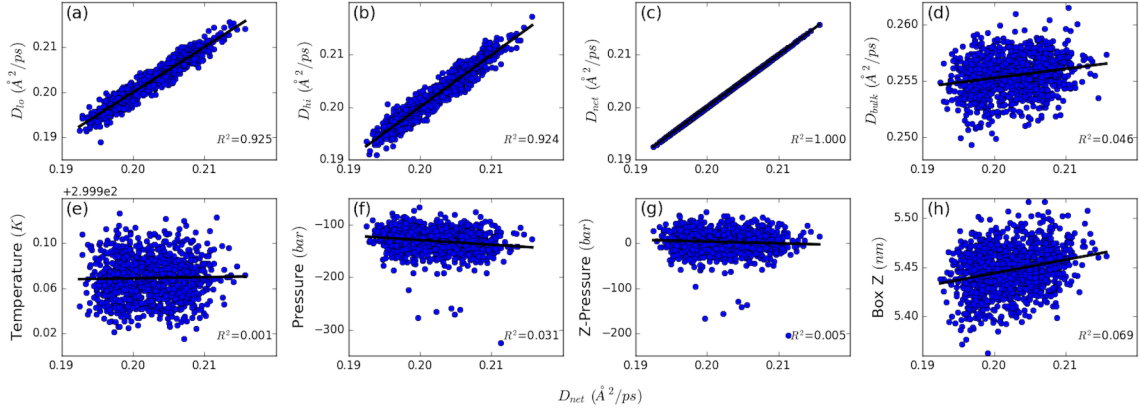


Figure E.6: Same as Figure E.1, but for the SAM surface at an OH-terminated chain fraction of 0.50.

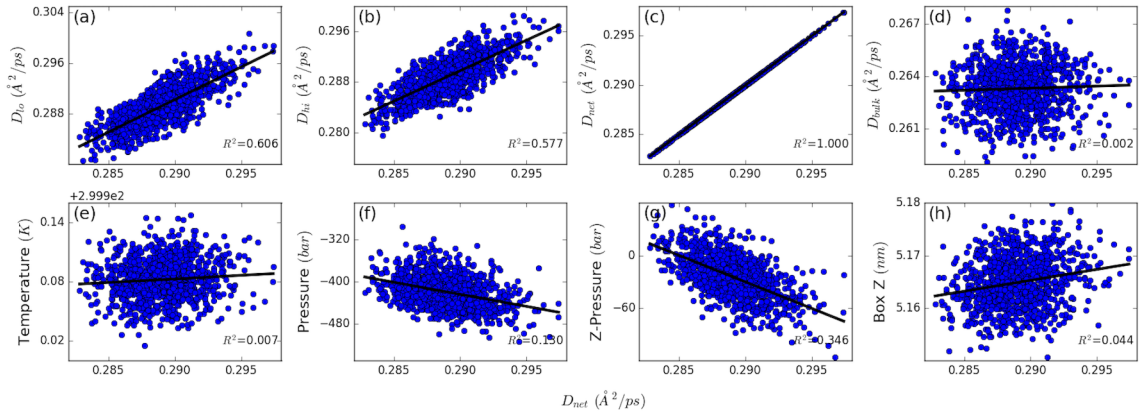


Figure E.7: Same as Figure E.1, but for the LJ surface at an attractive particle fraction of 0.25. For LJ surfaces at all coverages, the pressure normal to the interface correlates with the diffusivity. It should be noted, however, that the range of Z-pressures observed is the same as for other surface types. With surface patterning playing little to no role in determining diffusivity for the LJ surfaces (see main text), system properties that vary with system set-up and equilibration, like the Z-pressure, are more statistically significant, whereas these same fluctuations primarily introduce statistical noise to diffusivities on SAMs or cristobalite.

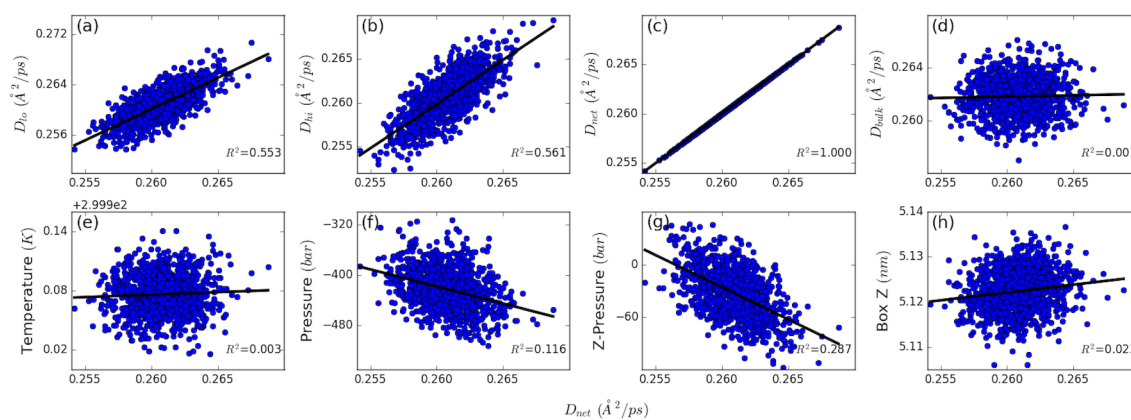


Figure E.8: Same as Figure E.1, but for the LJ surface at an attractive particle fraction of 0.50.

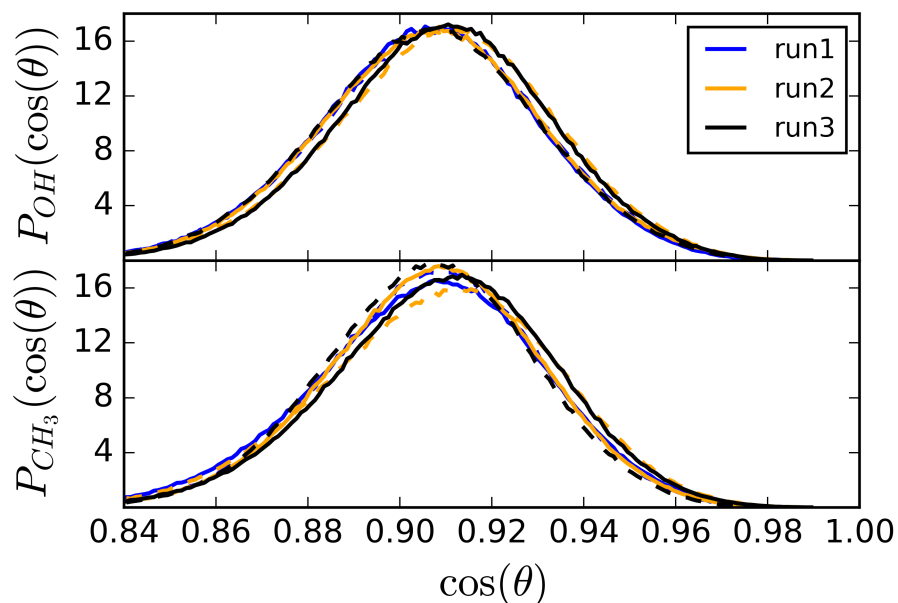


Figure E.9: Probability density distributions for the cosine of the tilt angle for SAM surfaces (the angle between the direction perpendicular to the interface and the vector between the restrained sulfur atom and the last united atom before the head group). The top panel shows the distribution for OH-terminated chains while the bottom for CH₃-terminated ones. Each color represents an independent simulation, with solid lines corresponding to the minimum-diffusivity surface pattern and dashed lines to the maximum-diffusivity surface. Variations in tilt angle distributions between independent simulation runs are larger than any differences observed between the minimum and maximum diffusivity patterns, implying negligible changes in average surface structure. Additionally, freezing all but the head-group atoms of SAM chains results in nearly identical diffusivity ranges in Figure 4.3 (also see Figure E.12), although the diffusivity is slightly reduced overall (by 0.014 to 0.032 Å²/ps). The decreases in diffusivity are more pronounced at higher surface densities and stem from the reduction, due to chain rigidification, in hydrogen bonding between surface hydroxyls, increasing the availability of water-surface hydrogen bonds.

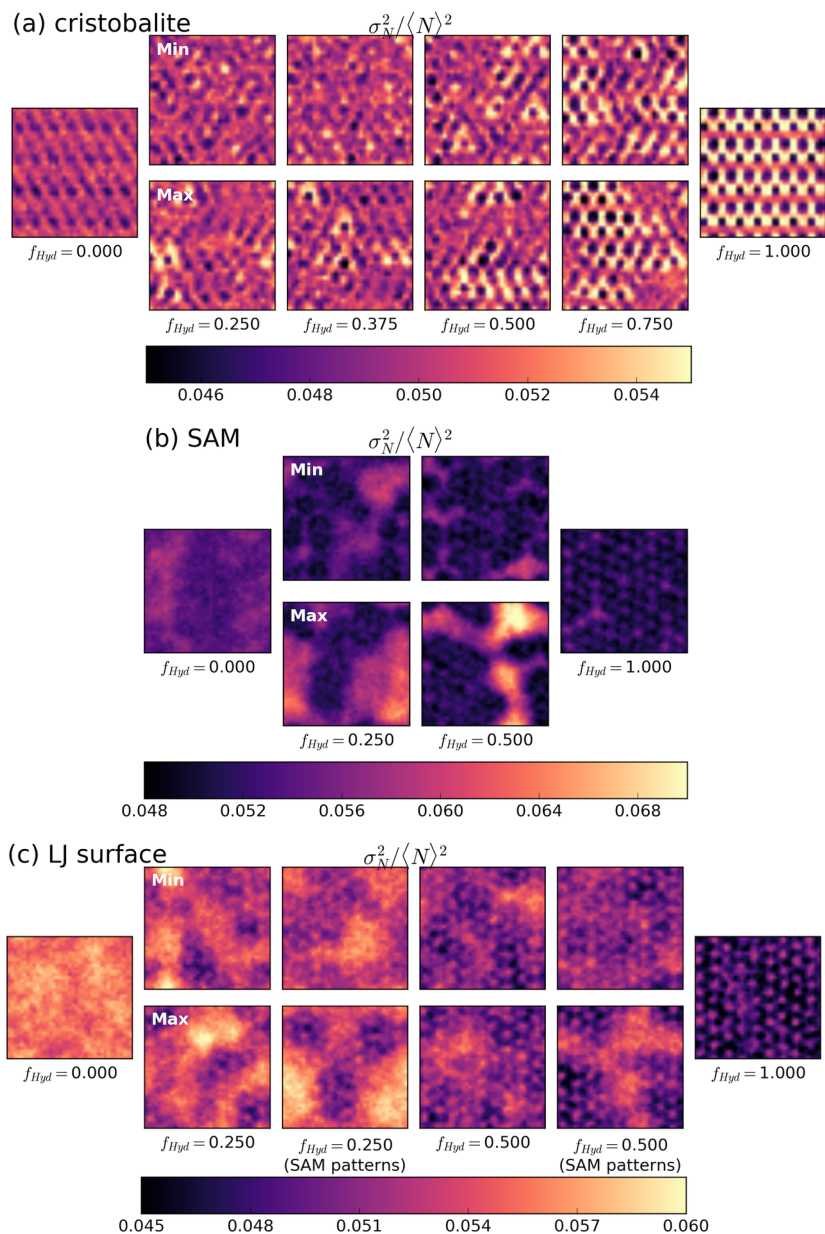


Figure E.10: Top-views for the normalized fluctuations in the number of waters within probes of radius 3.3 \AA sitting with their edge at the mean interface of (a) cristobalite, (b) SAM, and (c) LJ surfaces. The mean interface is defined as the location at which the average water density is 0.3 of the bulk value for TIP4P-Ew of 0.0332 \AA^{-3} . For LJ surfaces, the middle column considers the patterns determined by optimization of the SAM surface at the same density. It is clear that fluctuations are enhanced near larger patches of more hydrophobic groups on both SAM and LJ surfaces. For cristobalite, both very small and large fluctuations are observed near hydroxyl groups. This in turn leads to higher excess chemical potentials for hard-sphere probes near the surface.

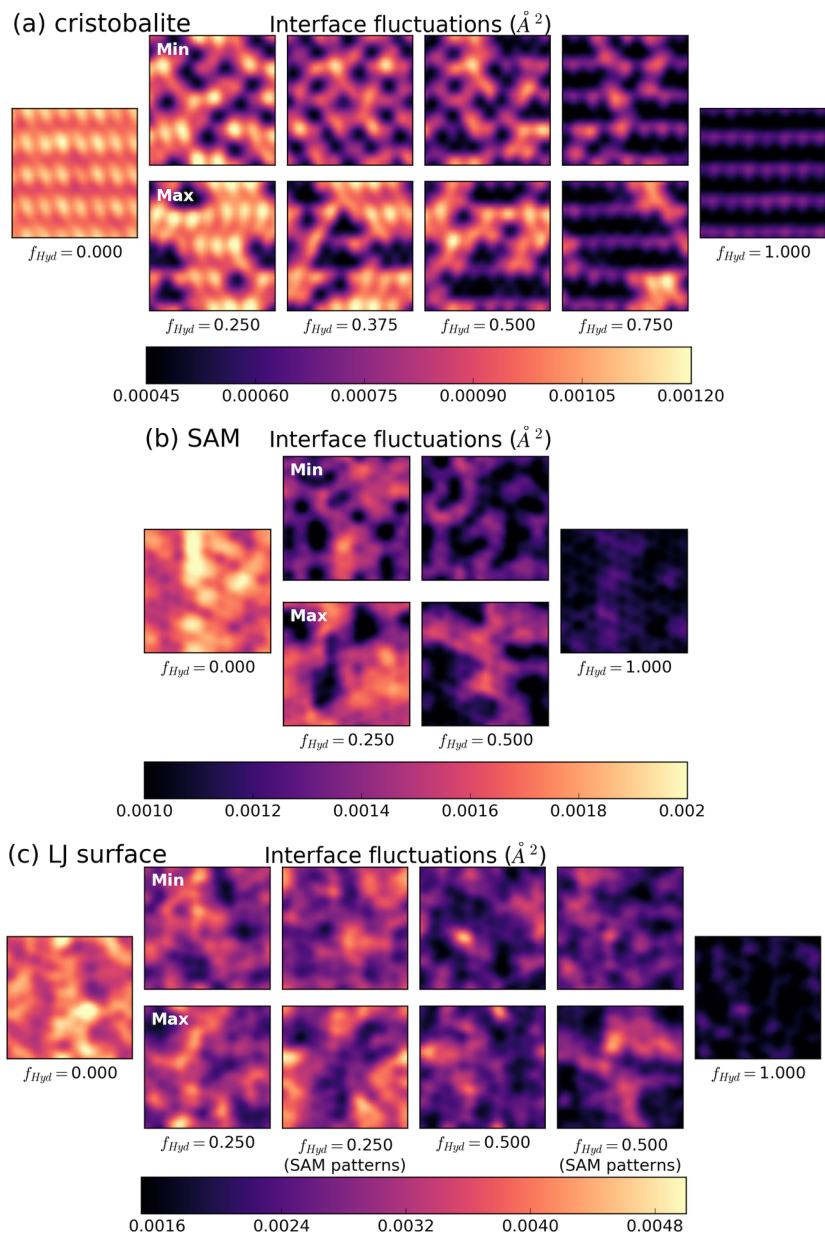


Figure E.11: Top-views of height fluctuations for instantaneous interfaces for cristobalite (top), SAM (middle), and LJ surface (bottom). The instantaneous interface is defined according to the methodology of Willard and Chandler,¹ but with a surface density cut-off of 0.3 of the bulk value of 0.0332\AA^{-3} for TIP4P-Ew water.

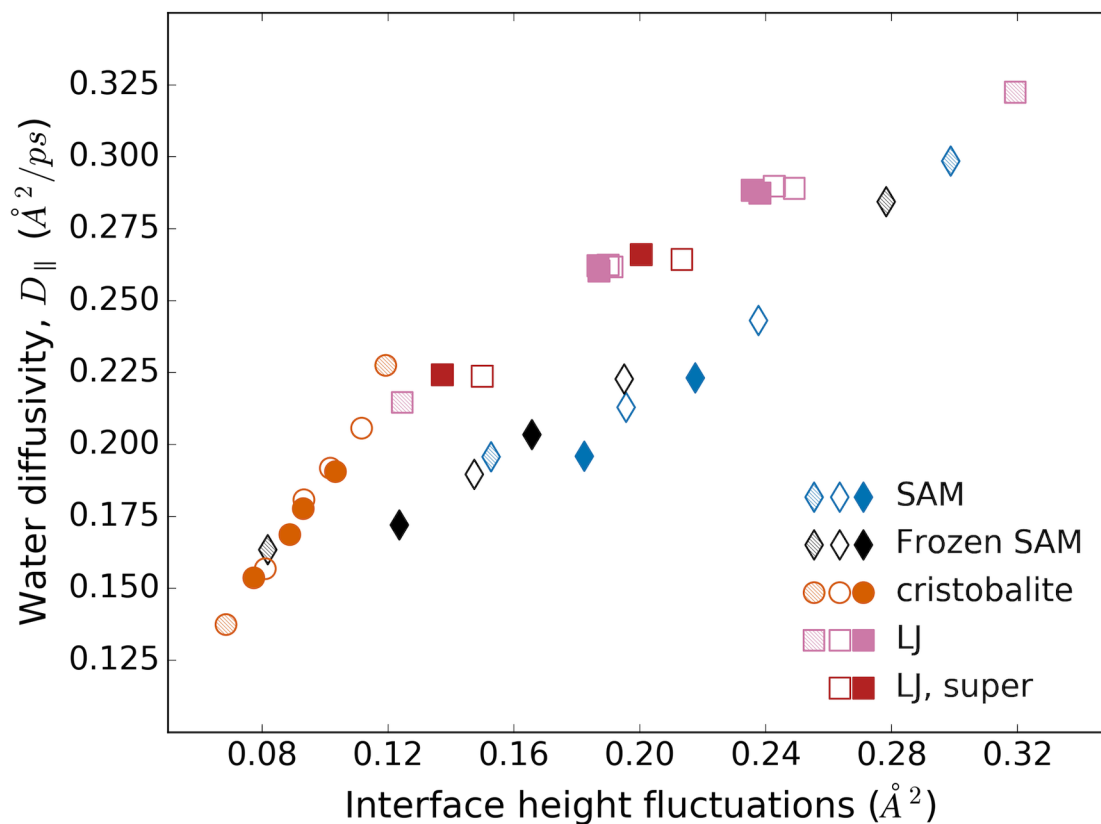


Figure E.12: Correlation between lateral diffusivities of water near minimum and maximum diffusivity surfaces and the average interfacial height fluctuations using the method of Willard and Chandler¹ to define the instantaneous interface. Hatched points represent surfaces at hydrophilic fractions of zero or one, filled points minimum diffusivity surfaces, and open points maximum diffusivity surfaces. The correlation is enhanced by freezing the united atoms (non-head-group atoms) of the SAM chains, as shown by the black versus blue diamond symbols.

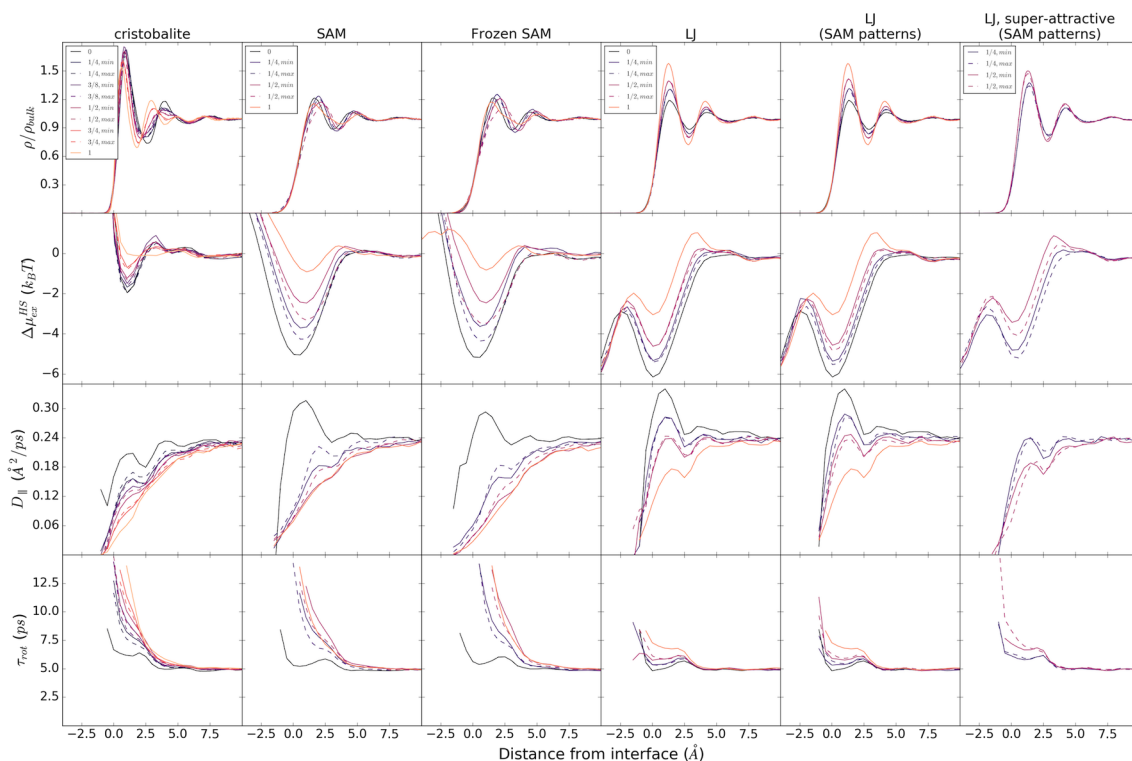


Figure E.13: Profiles for water density, hard-sphere excess chemical potential, water diffusivity parallel to the interface, and water dipole re-orientation times as a function of distance from the mean interface (defined as the point where water reaches 0.3 of its bulk value). Dipole re-orientation times were computed by fitting the time decay of the cosine of the angle between the original dipole moment vector with a stretched exponential, as in the work of Debenedetti and co-workers.² The qualitative trends are the same if an instantaneous interface definition is used instead. In general, hard-sphere excess chemical potentials are correlated with both lateral diffusivities and water re-orientation times. However, these spatial correlations seem to break down very close to the solid surfaces, where geometric constraints due to the presence of surface atoms seem to impact hard-sphere insertion probabilities and water mobility in an unrelated manner.

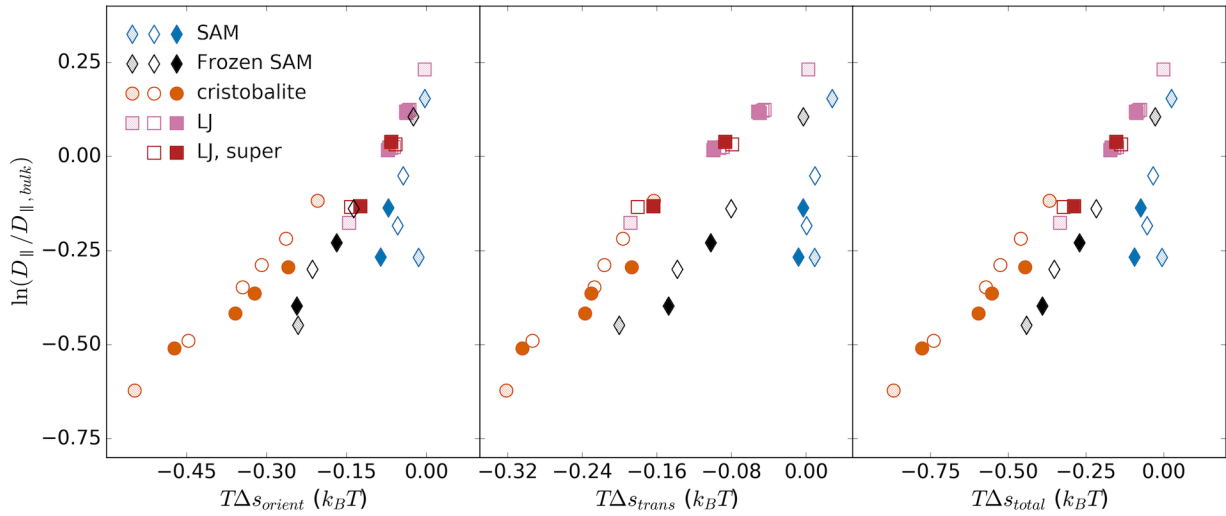


Figure E.14: The natural logarithm of the ratio of lateral diffusivity to its bulk value is shown as a function of the orientational, translational, and total contributions to hydration water entropy 8 \AA from the interface (reported relative to bulk waters). Filled symbols are minimum diffusivity surfaces (distributed hydrophilic groups), open symbols are maximum diffusivity surfaces (clumped hydrophilic groups), and hatched symbols are surfaces with hydrophilic fractions of 0.0 or 1.0. When all atoms in SAM chains except head-groups are frozen, all correlations are enhanced, mainly due to more negative translational contributions.

References

- [1] Willard AP, Chandler D. Instantaneous liquid interfaces. *The Journal of Physical Chemistry B*. 2010; 114(5): 1954–1958. arXiv: 0909.4761 ISBN: 1520-6106.
- [2] Romero-Vargas Castrillón S, Giovambattista N, Aksay IA, Debenedetti PG. Effect of Surface Polarity on the Structure and Dynamics of Water in Nanoscale Confinement. *The Journal of Physical Chemistry B*. 2009; 113(5): 1438–1446.

Appendix F

Relating Henry's Law constants to free energies of solvation

In order to convert Henry's Law constants to free energies of solvation, we first make it clear that the latter quantity is related directly to the excess chemical potentials at infinite dilution. We start by writing the total chemical potential in terms of Gibb's ensemble partition functions at constant T , P , and N solvent molecules while adding a single solute molecule

$$\begin{aligned}\beta\mu &= -\ln \frac{\Delta(T, p, N, N_s = 1)}{\Delta(T, p, N, N_s = 0)} \\ \beta\mu &= -\ln \frac{\Lambda^N \int_0^{\text{inf}} e^{-\beta pV} \int e^{-\beta U(\mathbf{r}^N, \mathbf{r}_s)} d\mathbf{r}^N d\mathbf{r}_s dV}{\Lambda^N \Lambda_s \int_0^{\text{inf}} e^{-\beta pV} \int e^{-\beta U(\mathbf{r}^N)} d\mathbf{r}^N dV}\end{aligned}\tag{F.1}$$

In the above, \mathbf{r}^N refers to the coordinates of the N solvent molecules, \mathbf{r}_s to the solute coordinates, $1/\Lambda^N$ to the kinetic contribution to the partition function for the solvent, and $1/\Lambda_s$ the kinetic contribution from the solute. The factorial terms cancel in this case because there are the same number of identical solvent molecules and only one solute. From here we may follow Widom's derivation of the excess chemical potential resulting in

his Potential Distribution Theorem.¹ Assuming a pairwise potential energy function, we may write $U(\mathbf{r}^N, \mathbf{r}_s) = U(\mathbf{r}^N) + \Delta U(\mathbf{r}^N, \mathbf{r}_s)$ so that

$$\begin{aligned}\beta\mu &= -\ln \frac{\int_0^{\text{inf}} e^{-\beta pV} \int e^{-\beta U(\mathbf{r}^N)} e^{-\beta \Delta U(\mathbf{r}^N, \mathbf{r}_s)} d\mathbf{r}^N d\mathbf{r}_s dV}{\Lambda_s \int_0^{\text{inf}} e^{-\beta pV} \int e^{-\beta U(\mathbf{r}^N)} d\mathbf{r}^N dV} \\ \beta\mu &= -\ln \frac{1}{\Lambda_s} \left\langle \int e^{-\beta \Delta U(\mathbf{r}^N, \mathbf{r}_s)} d\mathbf{r}_s \right\rangle\end{aligned}\tag{F.2}$$

The average runs over all configurations, including realizable volumes, of the NPT solvent system without the solute present. If we introduce the solute canonical configurational partition function z_s to the numerator and denominator inside the average, we can also write the average over all configurations of the solute in the ideal gas phase

$$\begin{aligned}\beta\mu &= -\ln \frac{1}{\Lambda_s} \left\langle \frac{z_s}{z_s} \int e^{-\beta \Delta U(\mathbf{r}^N, \mathbf{r}_s)} d\mathbf{r}_s \right\rangle \\ \beta\mu &= -\ln \frac{1}{\Lambda_s} \left\langle z_s(V) e^{-\beta \Delta U(\mathbf{r}^N, \mathbf{r}_s)} \right\rangle\end{aligned}\tag{F.3}$$

We have noted the dependence of z_s on the volume to remind ourselves of this fact and that the averaging procedure described above will affect this term via the integral over volume. Multiplying and dividing by the factor $\frac{k_B T}{p}$ we find

$$\begin{aligned}\beta\mu &= -\ln \frac{k_B T}{p \Lambda_s} \left\langle \frac{z_s(V) p}{k_B T} e^{-\beta \Delta U(\mathbf{r}^N, \mathbf{r}_s)} \right\rangle \\ \beta\mu &= -\ln \frac{k_B T}{p \Lambda_s} - \ln \left\langle \frac{z_s(V) p}{k_B T} e^{-\beta \Delta U(\mathbf{r}^N, \mathbf{r}_s)} \right\rangle\end{aligned}\tag{F.4}$$

$$\beta\mu = \beta\mu_{id} + \beta\mu_{ex}$$

The calculated free energy of solvation involves the free energy to move a single solute from the ideal gas phase into solution, but this is not identical to the excess chemical potential due to differences in pressure. In the simulations, the ideal gas particle is to a good approximation in a fixed volume equal to the average box size \bar{V} which results in

a pressure of $p' = \frac{k_B T}{\bar{V}}$. This pressure is different from the pressure of the bulk phase p and so the excess chemical potential is related to the solvation free energy computed in simulation by

$$\beta \Delta G_{solv} = \beta \mu_{ex} + \ln \frac{p}{p'} = \beta \mu_{ex} + \ln \frac{\bar{V}}{V} \quad (\text{F.5})$$

In the last equality, V is the volume occupied by a single ideal gas particle at pressure p .

We now assume that the solute is in equilibrium with an ideal gas mixture at the same overall pressure with vapor mole fraction of the solute, y

$$\begin{aligned} -\ln \frac{k_B T}{p \Lambda_s} - \ln \left\langle \frac{z_s(V)p}{k_B T} e^{-\beta \Delta U(\mathbf{r}^N, \mathbf{r}_s)} \right\rangle &= -\ln \frac{k_B T}{y p \Lambda_s} \\ \left\langle \frac{z_s(V)p}{k_B T} e^{-\beta \Delta U(\mathbf{r}^N, \mathbf{r}_s)} \right\rangle &= \frac{1}{y} \end{aligned} \quad (\text{F.6})$$

Henry's Law is usually defined as having the form

$$y = \frac{c}{H^{cp} p} \quad (\text{F.7})$$

where H^{cp} is the Henry's Law constant and c is the concentration, in moles/m³ in the liquid phase. Substituting this into our expression and remembering the above definition of the excess chemical potential

$$\begin{aligned} e^{-\beta \mu_{ex}} &= \frac{H^{cp} p}{c} \\ \beta \mu_{ex} &= -\ln \frac{H^{cp} p}{c} \end{aligned} \quad (\text{F.8})$$

To compute μ_{ex} for our simulated systems via Henry's Law constants, we only need to know the concentration of the solute in the system. For a large enough system, the volume fluctuates little and is well approximated by the average system volume, or $\bar{V} = 1/c$. Substituting the expression relating the free energy of solvation to the excess chemical

potential, we find that

$$\beta\Delta G_{solv} = -\ln \frac{H^{cp}p}{c} + \ln \frac{\bar{V}}{V} = -\ln H^{cp}pV \quad (\text{F.9})$$

$$\beta\Delta G_{solv} = -\ln RT H^{cp} \quad (\text{F.10})$$

The above expression is equivalent to the expression commonly relating the excess chemical potential in the canonical ensemble to the Henry's Law constant

$$\beta\mu_{ex} = -\ln \frac{RT\rho_{solvent}}{k_H} \quad (\text{F.11})$$

where at infinite dilution

$$k_H = \frac{yp}{x} \approx \frac{yp\rho_{solvent}}{c} = \frac{\rho_{solvent}}{H^{cp}} \quad (\text{F.12})$$

Indeed, if our simulations were performed in the canonical ensemble, the solvation free energy would be identical to the excess chemical potential of the canonical ensemble. In the macroscopic limit, both expressions yield the same result in equilibrium with an ideal gas at a fixed pressure.

References

- [1] Widom B. Some Topics in the Theory of Fluids. *The Journal of Chemical Physics*. 1963; 39(11): 2808–2812.

Appendix G

Additional data and figures
associated with solute affinities for
interfaces

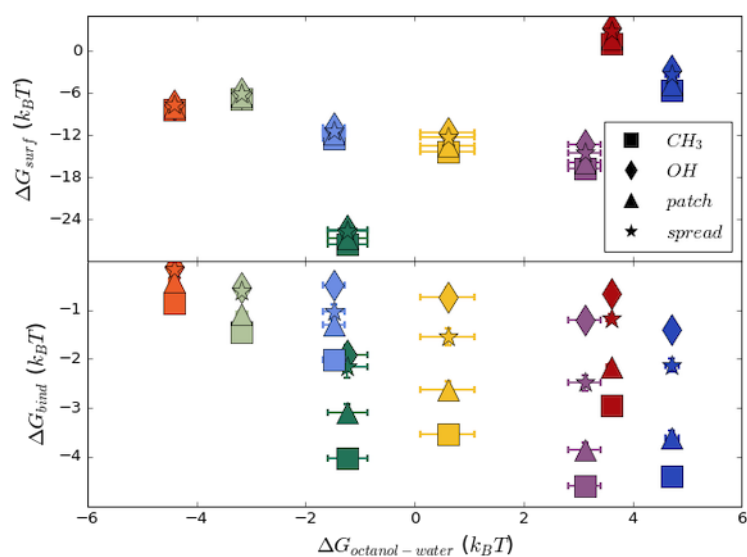


Figure G.1: Octanol-water transfer free energies (equal to the natural logarithm of the octanol-water partition coefficient) do not correlated to either aqueous solvation free energies at interfaces (top) or, more importantly, binding free energies (bottom). Transfer free energies were computed as the difference to solvate each solute in water minus solvation in a similarly-sized volume of octanol at the same temperature and pressure. Terms accounting for the difference in average volume have been neglected as they are small ($\sim 0.05 k_B T$) compared to the uncertainty (up to $\sim 0.30 k_B T$).

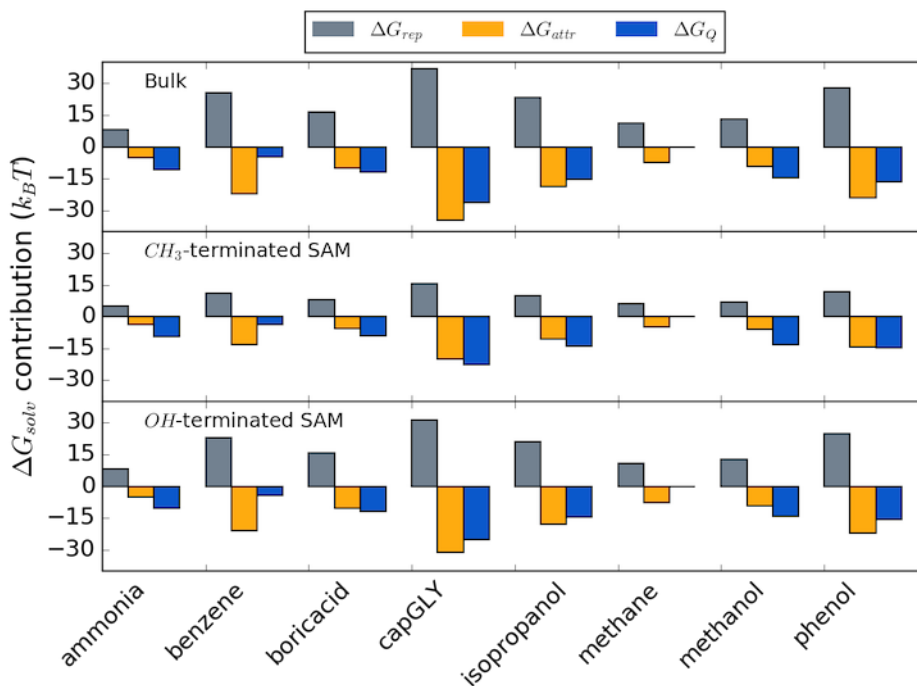


Figure G.2: Contributions to solvation free energies from turning on electrostatics with full LJ interactions on (blue, ΔG_Q), switching solutes from the ideal gas phase to interacting with the rest of the system via only hard-sphere repulsions (gray, ΔG_{rep}), and switching hard-sphere interactions to full LJ interactions (orange, ΔG_{attr}). Due to the chosen alchemical pathway in expanded ensemble simulations it was possible to directly calculate the free energies to turn on only LJ interactions and to turn on electrostatics with LJ on. Repulsive free energies as described here are not along with pathway but were estimated by from Gaussian fits to overlap probability distributions as shown in Figure G.3. The attractive component of the LJ free energy is then estimated by subtracting this hard-sphere solute solvation free energy.

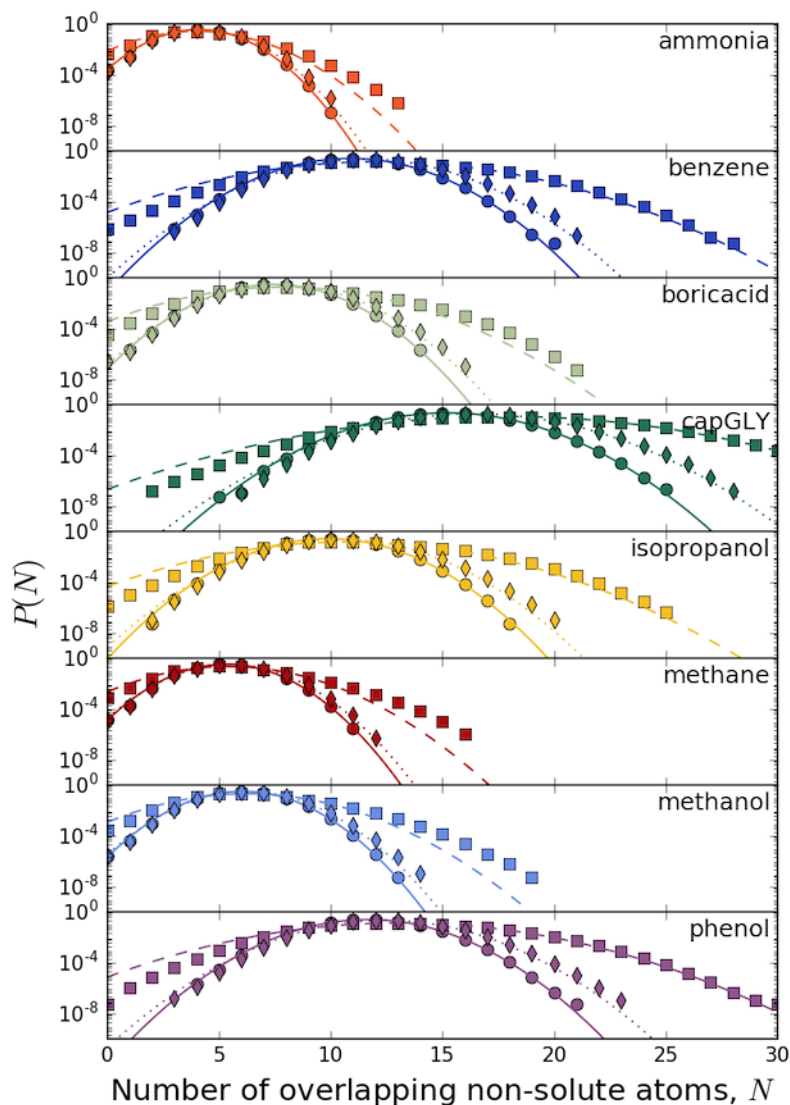


Figure G.3: Probability distributions for the number of non-overlapping non-solute atoms are shown for all solutes in bulk (circles) and at methyl- (squares) and hydroxyl-terminated (diamonds) SAMs. Gaussian fits to the distributions used to evaluate the probability at zero non-overlapping atoms are shown as solid, dashed, and dotted lines, respectively. For distributions in bulk and at hydrophilic interfaces, Gaussian distributions have been observed for small solutes for spherical,¹ cuboidal,² and irregular volumes, including that defined by a single alanine.³ Near hydrophobic surfaces, however, the Gaussian fits are anticipated to underestimate the weight in the distribution at low values of N ,^{2,4,5} resulting in underestimates of the hard-sphere solvation free energy at these surfaces. This does not seem to be the case, however, for values of N below the average, where the Gaussian fit actually leads to an overestimate of the weight, possibly due to poor sampling of the tails of the distributions, especially for larger solutes.

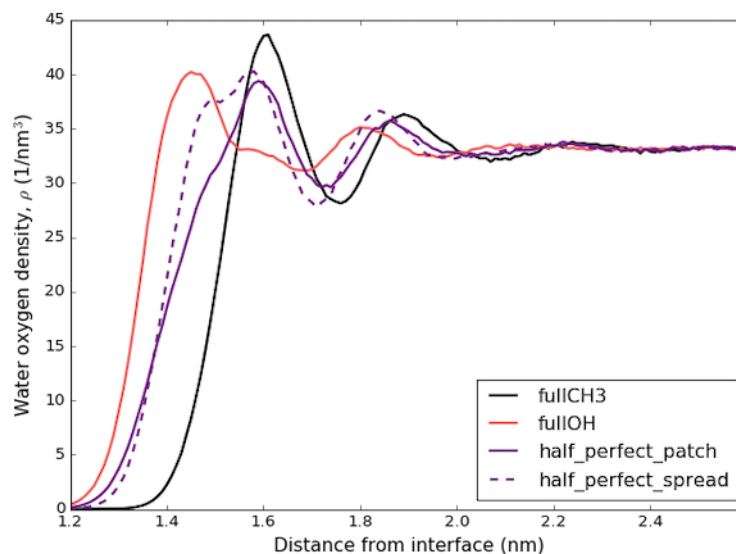


Figure G.4: Water density profiles as a function of distance from the frozen sulfur atoms of the SAM surface, the same scale as used for the PMFs shown in Figure 5.3 of the main text.

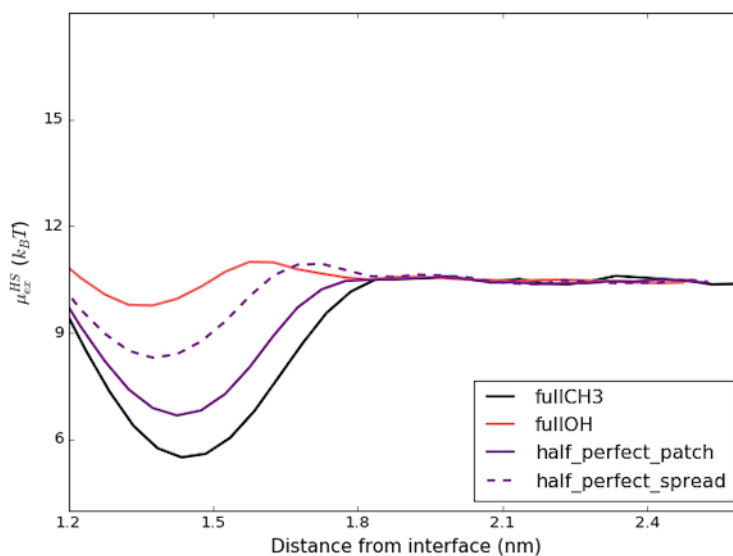


Figure G.5: Hard-sphere excess chemical potentials for methane-sized (0.33 nm radius) particles as a function of distance from distance from frozen sulfur atoms in the SAM surfaces. Larger values signify smaller density fluctuations, which is a proxy for measuring hydrophilicity/hydrophobicity, indicating that the surfaces increase in hydrophilicity in the order fully methylated, half and half patchy, half and half spread, and fully hydroxylated.

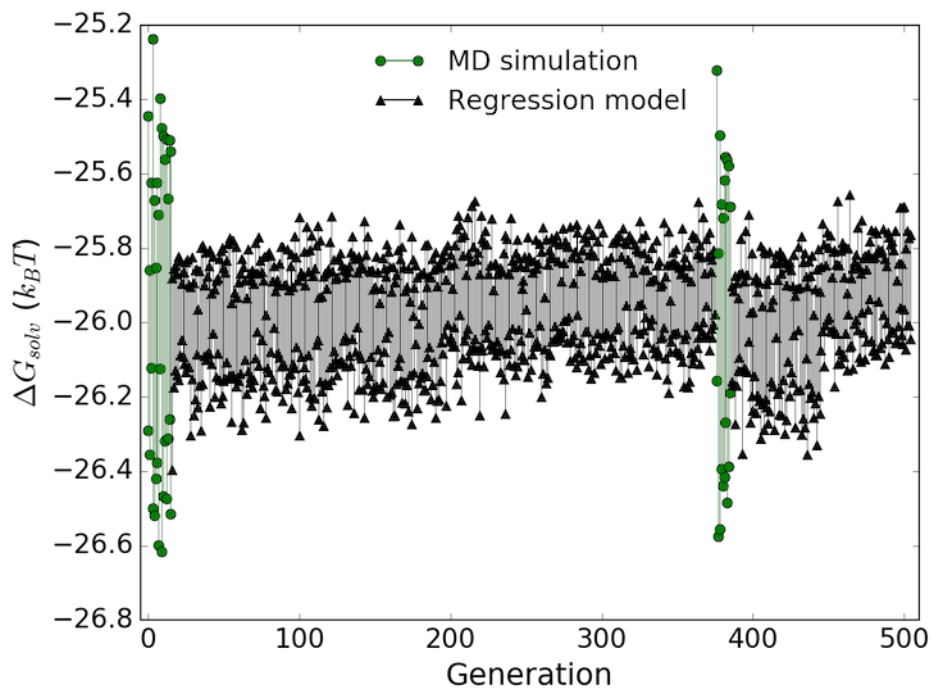


Figure G.6: Genetic algorithm optimization of surface patterns to minimize and maximize the affinity of capped glycine for a 50:50 methyl- and hydroxyl-terminated surface. Procedures are identical to the optimization for methane except that twice as much computational time is utilized for each simulation in both determining biasing weights and for data production. Though the results are far less impressive, the algorithm is beginning to identify patterns that change affinity for the interface, with the trained regression model accurately predicting free energies of solvation (see Table S1). While a longer run might help converge to more optimal patterns, convergence is mainly hindered by the large amount of noise in the MD simulation predictions of affinity.

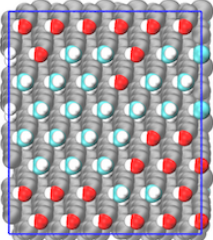
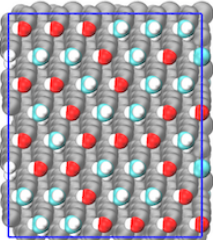
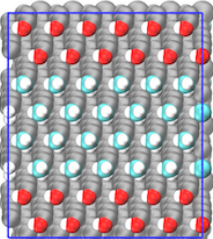
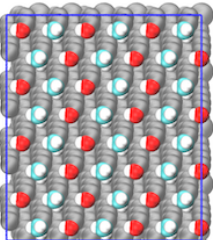
Surface	Surface image	ΔG_{solv} from MD simulation ($k_B T$)	ΔG_{solv} predicted by linear regression model ($k_B T$)
GA Model-predicted Minimum		-26.34 ± 0.18	-26.40
GA Model-predicted Maximum		-25.71 ± 0.15	-25.67
Global Minimum		-26.69 ± 0.09	-26.79
Global Maximum		-25.75 ± 0.17	-25.87

Table G.1: Surface representations and free energies of solvation for capped glycine, determined by MD simulations and linear regression models. The surfaces shown for the GA are those with minimum and maximum solvation free energies predicted by the linear regression model, not necessarily the overall optimal surfaces reported by the GA. Uncertainties are the standard error in the mean from 5 independent simulations.

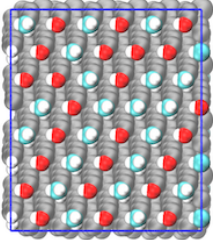
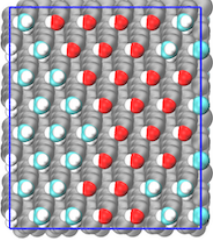
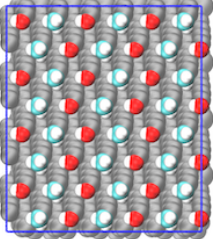
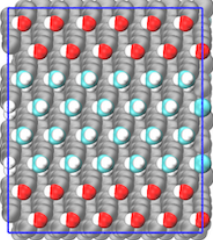
Surface	Surface image	$\Delta G_{solv,2} - \Delta G_{solv,1}$ from MD simulation ($k_B T$)	$\Delta G_{solv,2} - \Delta G_{solv,1}$ predicted by linear regression model ($k_B T$)
GA Model-predicted Minimum		-14.18 ± 0.08	-14.48
GA Model-predicted Maximum		-13.97 ± 0.07	-13.60
Global Minimum		-14.28 ± 0.10	-14.59
Global Maximum		-13.58 ± 0.07	-13.81

Table G.2: Surface representations and selectivity for methanol compared to methane, determined by MD simulations and linear regression models. Note that minimization of methanol selectivity is maximization of methane selectivity and vice versa. The surfaces shown for the GA are those with minimum and maximum solvation free energies predicted by the linear regression model, not necessarily the overall optimal surfaces reported by the GA. Uncertainties are the standard error in the mean from 5 independent simulations.

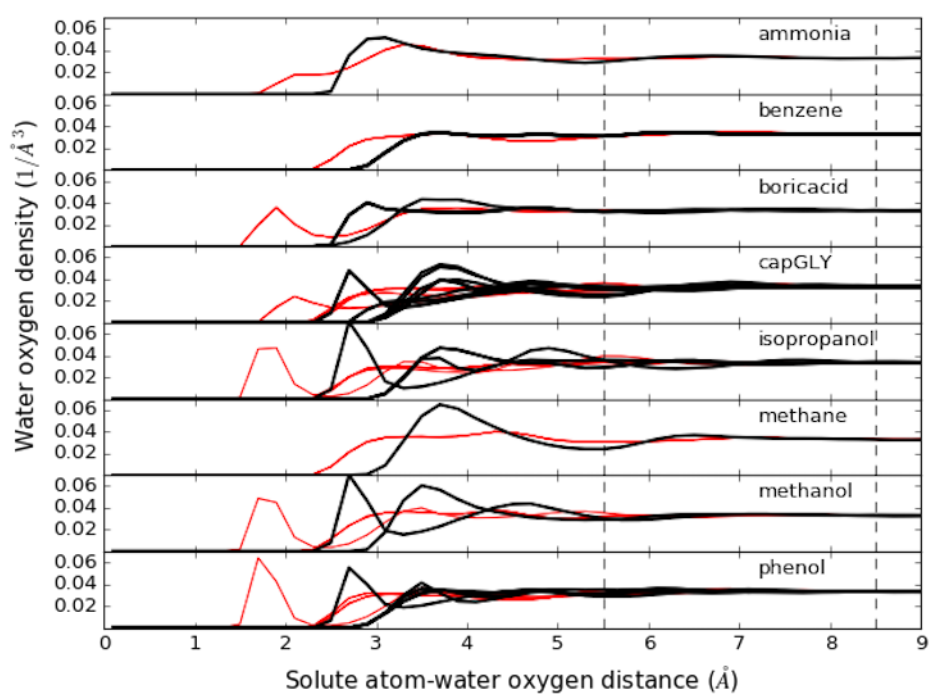


Figure G.7: Radial distribution functions between all solute atoms and water oxygens. Heavy atoms are shown in red while hydrogens are shown in black. Cutoffs defining first and second solvation shells at 5.5 and 8.5 Å are shown as vertical dotted lines.

References

- [1] Hummer G, Garde S, García aE, Pohorille a, Pratt LR. An information theory model of hydrophobic interactions. *Proceedings of the National Academy of Sciences of the United States of America*. 1996; 93(17): 8951–8955. ISBN: 2819960820.
- [2] Patel AJ, Varilly P, Chandler D. Fluctuations of Water near Extended Hydrophobic and Hydrophilic Surfaces. *The Journal of Physical Chemistry B*. 2010; 114(4): 1632–1637.
- [3] Sosso GC, Caravati S, Rotskoff G, Vaikuntanathan S, Hassanali A. On the Role of Nonspherical Cavities in Short Length-Scale Density Fluctuations in Water. *The Journal of Physical Chemistry A*. 2017; 121(1): 370–380.
- [4] Patel AJ, Varilly P, Jamadagni SN, Acharya H, Garde S, Chandler D. Extended surfaces modulate hydrophobic interactions of neighboring solutes. *Proceedings of the National Academy of Sciences*. 2011: 1110703108–. arXiv: 1105.0907 ISBN: 0027-8424.
- [5] Patel AJ, Varilly P, Jamadagni SN, Hagan MF, Chandler D, Garde S. Sitting at the edge: How biomolecules use hydrophobicity to tune their interactions and function. *Journal of Physical Chemistry B*. 2012; 116(8): 2498–2503. arXiv: 1109.4431 ISBN: 1520-5207 (Electronic)\n1520-5207 (Linking).