

UC Santa Barbara

UC Santa Barbara Electronic Theses and Dissertations

Title

Computational Design of Multifunctional Surfaces using Molecular Dynamics Simulations

Permalink

<https://escholarship.org/uc/item/01s47310>

Author

Jiao, Sally

Publication Date

2023

Peer reviewed|Thesis/dissertation

UNIVERSITY of CALIFORNIA
Santa Barbara

**Computational Design of Multifunctional Surfaces using Molecular
Dynamics Simulations**

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

Doctor of Philosophy

in

Chemical Engineering

by

Sally Jiao

Committee in charge:

Professor M. Scott Shell, Chair

Professor Kevin W. Plaxco

Professor James B. Rawlings

Professor Rachel A. Segalman

December 2023

The dissertation of Sally Jiao is approved:

Professor Kevin W. Plaxco

Professor James B. Rawlings

Professor Rachel A. Segalman

Professor M. Scott Shell, Committee Chair

December 2023

Copyright © 2023
by Sally Jiao

Acknowledgments

This dissertation would not be possible without the outstanding mentorship of Professor Scott Shell. Scott- it has been my great privilege to work with you. You are a uniquely wonderful adviser, in both your scientific insight and your unfailing support, patience, and kindness.

I have so enjoyed collaborating with Prof. Rachel Segalman, Prof. Songi Han, and Dr. Audra DeStefano on polypeptoids. Their experimental measurements underpin Chapter 4, and their feedback has tremendously improved Chapters 4 and 5. I would also like to thank Prof. Lynn Katz for her contributions to Chapter 2 and generally for her insight into (and inspiring passion for) water treatment. Dr. Dennis Robinson Brown's PEO surface measurements were a valuable contribution to Chapter 6. Finally, I am so glad I had the opportunity to mentor and collaborate with Daniela Rivera Mirabal on Chapters 5 and 7.

I am also indebted to my committee members, Rachel, Prof. Kevin Plaxco, and Prof. Jim Rawlings, whose excellent suggestions have shaped this work. Jim's questions inspired much of the method development in Chapter 3.

The members of the Center for Materials for Water and Energy Systems (M-WET) provided invaluable feedback on this work and cheerful companionship at our weekly pizza meetings. Special thanks to Audra, Dennis, Dr. Thomas Webber, and Dr. Rahul Sujanani, as well as Dr. Eileen Dawson, M-WET's fantastic Operations Director.

I would like to thank the members of the Shell Group for helpful suggestions, coding

tips, and general office camaraderie. I would especially like to thank Prof. Jacob Monroe and Dr. Nick Sherck, for their mentorship early on in my PhD; Evan Pretti, for patiently answering my questions about sim; and both Evan and Charles Li for attendance at Shell Group lunches in the early post-COVID era.

Our department is fortunate to have a friendly and capable staff, who helped take care of innumerable logistical issues. Many thanks as well to Dr. Paul Weakliem and Fuzzy Rogers, who enable all this research through their excellent stewardship of the campus computing clusters. Paul was also kind enough to lend his expertise when we had issues with our group's cluster.

I am lucky to have had many mentors prior to UCSB who taught me how to do research and encouraged me to go to graduate school, including Prof. Lynn Loo, Dr. Anna Rubinstein, Prof. Helge Stein, Prof. Thanos Panagiotopoulos, Dr. Andrew Santos, Dr. Nathan Mahynski, and Dr. Wick Hatch.

I have made so many wonderful friends at UCSB and I will miss our hikes, TV binging, board game nights, and boba outings. To Dr. Kellie Heom, especially- thank you for 5 years of roommateship and insightful evening discussions of the day's events.

To Mike Schmithorst, who patiently endured the joys and frustrations of research with me- thank you for all the California adventures. You made Santa Barbara so much more beautiful than it already is.

And to my family, thank you for all your support, sacrifices, and love. I love you all very, very much.

Curriculum Vitæ

Sally Jiao

Education

- 2023 Ph.D., Chemical Engineering, University of California, Santa Barbara
- 2018 B.S.E., Chemical and Biological Engineering, Princeton University, *summa cum laude*, certificate in Applications of Computing

Professional Experience

- 2018-2023 Graduate Researcher, University of California, Santa Barbara
- Summer 2022 Development Engineering Intern, Bristol Myers Squibb
- Summer 2018 Undergraduate Research Fellow, National Institute of Standards and Technology

Publications (* denotes equal contribution)

1. A. J. DeStefano, S. D. Mengel, M. W. Bates, **S. Jiao**, M. S. Shell, S. Han, R. A. Segalman, "Control Over Conformational Landscapes of Short Polypeptoids by Monomer Sequence Patterning," *Macromolecules*, *Under Review*.
2. **S. Jiao**, D. C. Robinson Brown, M. S. Shell, "Relationships between Water's Structure and Solute Affinity at Polypeptoid Brush Surfaces," *Langmuir*, *Accepted*.
3. D. Robinson Brown, T. Webber, **S. Jiao**, D. M. Rivera Mirabal, S. Han, M. S. Shell, "Relationships between Molecular Structural Order Parameters and Equilibrium Water Dynamics in Aqueous Mixtures," *J. Phys. Chem. B*, 127 (**2023**) 4577-4594.
4. J. S. Straub, M. S. Nowotarski, J. Lu, T. Sheth, **S. Jiao**, M. P. A. Fisher, M. S. Shell, M. E. Helgeson, A. Jerschow, S. Han, "Phosphates Form Spectroscopically Dark State Assemblies in Common Aqueous Solutions," *Proc. Natl. Acad. Sci.*, 120 (**2022**) e2206765120.
5. **S. Jiao**, L. E. Katz, M. S. Shell, "Inverse Design of Pore Wall Chemistry to Control Solute Transport and Selectivity," *ACS Cent. Sci.*, 8 (**2022**) 1609-1617.
6. **S. Jiao***, D. M. Rivera Mirabal*, A. J. DeStefano, R. A. Segalman, S. Han, M. S. Shell, "Sequence Modulates Polypeptoid Hydration Water Structure and Dynamics," *ACS Biomacromolecules*, 23 (**2022**) 1745-1756.

7. **S. Jiao**, A. DeStefano, J. I. Monroe, M. Barry, N. Sherck, T. Casey, R. A. Segalman, S. Han, M. S. Shell, "Quantifying Polypeptoid Conformational Landscapes through Integrated Experiment and Simulation," *Macromolecules*, 54 (**2021**) 5011-5021.
8. J. I. Monroe, **S. Jiao**, R. J. Davis, D. Robinson Brown, L. E. Katz, M. S. Shell, "Affinity of Small-Molecule Solutes to Hydrophobic, Hydrophilic, and Chemically Patterned Interfaces in Aqueous Solution," *Proc. Natl. Acad. Sci.*, 118 (**2020**) e2020205118.
9. R. Sujanani, M. R. Landsman, **S. Jiao**, J. D. Moon, M. S. Shell, D. F. Lawler, L. E. Katz, B. D. Freeman, "Designing Solute-Tailored Selectivity in Membranes: Perspectives for Water Reuse and Resource Recovery," *ACS Macro Lett.*, 9 (**2020**) 1709-1717.
10. N. Sherck, T. Webber, D. Robinson Brown, T. Keller, M. Barry, A. DeStefano, **S. Jiao**, R. A. Segalman, G. H. Fredrickson, M. S. Shell, S. Han, "End-to-End Distance Probability Distributions of Dilute Poly (ethylene oxide) in Aqueous Solution," *J. Am. Chem. Soc.*, 142 (**2020**) 19631-19641.
11. J. Monroe, M. Barry, A. DeStefano, P. A. Gokturk, **S. Jiao**, D. Robinson Brown, T. Webber, E. J. Crumlin, S. Han, M. S. Shell, "Water Structure and Properties at Hydrophilic and Hydrophobic Surfaces," *Ann. Rev. Chem. Biomol. Eng.*, 11 (**2020**) 523-557.
12. **S. Jiao**, A. P. Santos, A. Z. Panagiotopoulos, "Differences in Free Surfactant Concentration and Aggregation Properties for Amphiphiles with the Same Critical Micelle Concentration," *Fluid Ph. Equilibria*, 470 (**2018**) 126-133.
13. N. A. Mahynski*, **S. Jiao***, H. W. Hatch, M. A. Blanco, V. K. Shen, "Predicting Structural Properties of Fluids by Thermodynamic Extrapolation," *J. Chem. Phys.*, 148 (**2018**) 194105.
14. H. W. Hatch, **S. Jiao**, N. A. Mahynski, M. A. Blanco, and V. K. Shen, "Communication: Predicting Virial Coefficients and Alchemical Transformations by Extrapolating Mayer-Sampling Monte Carlo Simulations," *J. Chem. Phys.*, 147 (**2017**) 231102.
15. H. S. Stein, **S. Jiao**, A. Ludwig, "Expediting Combinatorial Data Set Analysis by Combining Human and Algorithmic Analysis," *ACS Comb. Sci.*, 19 (**2017**) 1-8.

Oral Presentations

1. "Inverse Design of Pore Wall Chemistry to Control Solute Transport and Selectivity," *AIChE Annual Meeting*, Phoenix, AZ, Nov. **2022**.
2. "Automated Design of Chemically Functionalized Nanopores to Control Solute Transport and Selectivity," *ACS Spring Meeting*, San Diego, CA, Mar. **2022**.
3. "Sequence Modulates Polypeptoid Hydration Water Structure and Dynamics," *AIChE Annual Meeting*, Boston, MA, Nov. **2021**.

4. “Overcoming the Permeability-Selectivity Tradeoff via the Design of Heterogeneous Pore Wall Chemistries,” *AIChE Annual Meeting*, Boston, MA, Nov. **2021**.
5. “Quantifying Polypeptoid Conformational Landscapes through Integrated Experiment and Simulation,” *Online Peptoid Symposium*, virtual, Dec. **2020**.
6. “Design of Pore Wall Chemistry to Control Solute Transport and Selectivity,” *AIChE Annual Meeting*, virtual, Nov. **2020**.

Awards

2021	Schlinger Fellowship, Chemical Engineering, U.C. Santa Barbara
2021	Teacher-Scholar Fellowship, Chemical Engineering, U.C. Santa Barbara
2018	National Science Foundation Graduate Research Fellowship
2018	Chancellor’s Fellowship, Graduate Division, U.C. Santa Barbara
2018	James Hayes-Edgar Palmer Prize in Engineering, Princeton University
2018	Award for Overall Excellence in Chemical Engineering, Princeton University
2018	Richard K. Toner Thermodynamics Prize, Princeton University
2018	Christodoulos A. Floudas Chemical Process Design Award, Chemical and Biological Engineering, Princeton University
2018	Student Teaching Award, Computer Science, Princeton University
2018	Class of 1939 Princeton Scholar Award, Princeton University
2017	Phi Beta Kappa early election, Princeton University
2017	Barry M. Goldwater Scholarship

Leadership, Outreach, and Service Activities

2019-2023	HPC Cluster System Administrator, Shell Group
2022	Publications Committee, Chemical Engineering Graduate Student Symposium

2021-2022	Graduate Scholars Program Mentor, Graduate Division, U.C. Santa Barbara
2021-2022	Basic Energy Sciences Early Career Network Representative, Center for Materials for Water and Energy Systems
2020-2022	President, Graduate Simulation Seminar Series
2021	Organizing Committee, Early Career Researcher Peptoid Symposium
2020-2021	Secretary, Chemical Engineering Graduate Student Association
2020-2021	UCSB Site Lead, Center for Materials for Water and Energy Systems Student Leadership Council
2020	Future Leaders in Advanced Materials Research Mentor, Materials Research Laboratory, U.C. Santa Barbara
2020	Chemical Engineering Prospective Graduate Student Visitation Weekend Co-Leader
2019-2020	Professional Development Co-Chair, Chemical Engineering Graduate Student Association
2019-2020	GAP 1 Lead, Center for Materials for Water and Energy Systems Student Leadership Council
2019-2020	Editorial Board, Frontiers in Energy Research Newsletter

Undergraduate Students Mentored

2020	Daniela M. Rivera Mirabal, University of Puerto Rico at Mayagüez
------	--

Abstract

Computational Design of Multifunctional Surfaces using Molecular Dynamics Simulations

by

Sally Jiao

Surface modification of water filtration membrane materials has the potential to imbue attractive qualities, such as enhanced antifouling or solute rejection/selectivity capabilities. Development of next-generation membrane materials will therefore rely on design rules relating surface structure and functional properties. Heuristics exist for simple surfaces comprising homogeneous chemistries and have largely been discovered through trial-and-error synthesis and characterization. However, we lack a complete understanding of nanoscopic mechanisms governing the water-mediated interactions between contaminants/foulants and realistic surfaces with chemical and topological heterogeneity. Furthermore, trial-and-error methods are not able to efficiently explore the large design space offered by multifunctional surfaces.

This dissertation develops relationships between surface chemistry and local water/solute behavior for model membrane materials, using computational inverse design approaches. By coupling molecular dynamics simulations with optimization algorithms, we identify non-intuitive chemistries showing enhanced functional properties. Genetic algorithm optimization of patterns of nonpolar and polar chemistries in a model nanopore suggests that arrangement of polar chemistry in rings along the nanopore wall will enhance selectivity of water over small, neutral solutes. Active-learning-based optimization of more complex patterns with both chemical and topological heterogeneity suggest that

rings of large nonpolar groups create a rough surface that achieves similar effects. Importantly, these non-intuitive functionalities discovered by the inverse design approach suggest previously unexplored mechanisms governing water-mediated solute transport in model membrane materials.

Sequence-specific polymers such as polypeptoids provide a synthetically accessible route to chemically precise surface modification. However, exploration of sequence-structure-function relationships for polypeptoids has been lacking, in part because computational workflows to simulate polypeptoids are not well established. In collaboration with detailed experimental characterization, we validate an enhanced sampling simulation workflow and atomistic model for polypeptoids. We then show that polypeptoid sequence modulates local water behavior in dilute solution and at peptoid brush surfaces, and that the changes in water structure can be related to trends in affinity of small-molecule solutes for peptoid brush surfaces. Finally, we apply the previously developed inverse design techniques to build sequence-structure relationships by identifying sequences with extremal conformational behaviors.

Contents

1	Introduction	1
1.1	Background	1
1.2	Overview	8
2	Inverse design of pore wall chemistry to control solute transport and selectivity	10
2.1	Introduction	10
2.2	Methods	15
2.2.1	Simulation models	15
2.2.2	Finite nanopore simulation	16
2.2.3	Infinite nanopore simulation	17
2.2.4	Genetic algorithm structure	18
2.3	Results and Discussion	19
2.3.1	Genetic algorithm optimization	19
2.3.2	Rationally designed patterns suggest mechanism for boric acid rejection	23
2.4	Conclusions	30
3	Inverse design of pore wall chemistry and topology	33
3.1	Introduction	33
3.2	Methods	37
3.2.1	Nanopore simulations	37
3.2.2	Active learning optimization	38
3.3	Results and Discussion	42
3.3.1	Pore wall topology affects water and solute transport	42
3.3.2	Active learning approach identifies diverse and optimal patterns	45
3.4	Conclusions	52
4	Quantifying polypeptoid conformational landscapes through integrated experiment and simulation	55
4.1	Introduction	55
4.2	Methods	61
4.2.1	Polypeptoid forcefield parameterization and simulation	62

4.2.2	Expanded-ensemble sampling	62
4.2.3	Comparison of polypeptoid forcefields	63
4.3	Results and Discussion	66
4.3.1	Expanded-ensemble simulation accurately and efficiently samples isomerization states	66
4.3.2	End-to-end distance distributions from experiments and simulations	68
4.4	Conclusions	73
5	Sequence modulates polypeptoid hydration water structure and dy-	
	namics	77
5.1	Introduction	77
5.2	Methods	80
5.2.1	Peptoid simulation	81
5.2.2	Computational hydration water characterization	84
5.3	Results	86
5.3.1	Peptoids perturb water structure and dynamics	87
5.3.2	Sequence and conformation affect water dynamics	92
5.3.3	Sequence-induced shifts in water dynamics and structure are cor- related	96
5.3.4	The spin label captures changes in water dynamics	98
5.4	Discussion	100
5.4.1	Water shows small-lengthscale responses to disordered polypeptoids	101
5.4.2	In disordered polypeptoids, sequence has a larger effect on water behavior than conformation	102
5.4.3	Water dynamics and tetrahedrality are connected	104
5.5	Conclusions	105
6	Relationships between water’s structure and solute affinity at polypep-	
	toid brush surfaces	109
6.1	Introduction	109
6.2	Methods	113
6.2.1	Peptoid and poly (ethylene oxide) surface construction	113
6.2.2	Surface simulations and analysis	115
6.3	Results and Discussion	118
6.3.1	Solute affinities to surfaces	118
6.3.2	Decomposition of solute free energy of binding	123
6.3.3	Structure of surface hydration water	125
6.4	Conclusions	136
7	Effect of polypeptoid sequence on conformation	139
7.1	Introduction	139
7.2	Methods	143
7.2.1	Atomistic polypeptoid simulations	143

7.2.2	Coarse-grained polypeptoid simulations	144
7.2.3	Genetic algorithm workflow	145
7.3	Results and Discussion	146
7.3.1	Hydrophilic polypeptoids show excluded volume scaling	146
7.3.2	Hydrophobic monomers shift the conformational distribution	149
7.3.3	Coarse-grained model reproduces conformations	150
7.3.4	Genetic algorithm optimization identifies extremal sequences	153
7.4	Conclusions	155
8	Conclusions	157
8.1	Summary	157
8.2	Outlook	161
8.2.1	Improvements to computational workflows	161
8.2.2	Application of workflows to new systems	165
A	Supporting Information for Inverse Design of Pore Wall Chemistry	168
A.1	Simulation models and property calculations	169
A.1.1	Pore geometry	169
A.1.2	Functional group layout	169
A.1.3	Forcefield details	169
A.1.4	Overview of pore simulation steps	172
A.1.5	Finite pore setup	173
A.1.6	Partition coefficient calculation	174
A.1.7	Infinite pore setup	175
A.1.8	Flux calculation	176
A.2	Genetic algorithm workflow	178
A.2.1	Pattern entropy minimization	178
A.2.2	Maximization generations	179
A.2.3	Model-driven generations	180
A.2.4	Generations in each stage	182
A.3	Surrogate function	183
A.3.1	Features	183
A.3.2	Surrogate model performance	184
A.3.3	Alternative surrogate models	185
A.3.4	Feature importance	185
A.4	Rationally designed patterns	187
A.4.1	Pattern clustering	187
A.4.2	Patterns and computed transport properties	187
A.4.3	Diffusivity of water and boric acid	189
A.4.4	Analytical model for boric acid transport	190
A.4.5	Other solutes	191
A.5	Permeability-selectivity trade-off	193

B	Supporting Information for Inverse Design of Chemistry and Topology	195
B.1	Simulation details	196
B.2	Visualizing solute locations in pore	197
B.3	Water properties	198
C	Supporting Information for Quantifying Polypeptoid Conformational Landscapes	201
C.1	Simulation workflow details	202
C.1.1	Parameterization	202
C.1.2	Expanded-ensemble sampling of dihedral rotations	216
C.1.3	Simulation settings	221
C.1.4	Uncertainty calculations	223
C.1.5	Forcefield selection	225
C.1.6	Water model selection	226
C.2	Time-domain simulations	226
C.3	Polypeptoid conformational behaviors	229
C.3.1	Visualization of the conformational distribution	229
C.3.2	Dihedral preferences	230
C.3.3	Comparison of labeled and unlabeled molecules	230
C.3.4	Comparison in DEER-accessible region	231
C.4	Dependence of predicted distribution on cw-EPR optimal region bounds	234
D	Supporting Information for Sequence Modulates Polypeptoid Hydration Water Structure and Dynamics	236
D.1	Simulation details	237
D.1.1	Expanded ensemble simulations	237
D.1.2	NVE simulations	238
D.1.3	Bulk water simulation	239
D.1.4	Silica-tethered peptoid systems	239
D.1.5	Calculation of diffusivity near spin labels	240
D.2	Extracting diffusivity from survival probability	242
D.2.1	Derivation of survival probability model	242
D.2.2	Survival probability around a moving center	243
D.3	Conformational clustering	247
D.3.1	Clustering conformations of NME and alkyl homopolymers	247
D.3.2	Separating out effect of conformation for all sequences	247
D.4	Additional hydration water measurements	250
D.4.1	Structural distributions and dynamical trajectories	250
D.4.2	Trends in rotational correlation time	253
D.4.3	Scaling down NME charges	254
D.4.4	Comparison of water behavior around unlabeled peptoids	255
D.4.5	Varying the cutoff for water behavior calculations	255
D.4.6	Additional structure-dynamics correlations	256

E	Supporting Information for Relationships between Water’s Structure and Solute Affinity	258
E.1	Additional simulation details	259
E.1.1	Expanded ensemble simulations	259
E.1.2	Peptoid brush surface simulations	259
E.1.3	PEO brush surface simulations	262
E.1.4	Bulk water simulations	263
E.2	Water properties	263
E.2.1	Water dynamics	263
E.2.2	Density profiles	264
E.2.3	Hydrogen bonding	264
E.2.4	Three-body angle distribution	265
E.3	Solute-surface interactions	269
E.3.1	Potential of mean force for other solutes	269
E.3.2	Rearrangement of peptoid surface	271
E.3.3	Projections of 3-body distributions	272
E.4	Prediction of $T\Delta S_{\text{rel}}$	273
F	Supporting Information for Effect of Polypeptoid Sequence on Conformation	278
F.1	Simulation details	279
F.1.1	Atomistic simulations	279
F.1.2	Coarse-grained simulations	279
F.2	Comparison to theoretical form of $P(R_{ee})$	280
F.3	Surrogate model	281
	References	282

List of Figures

1.1	Water filtration membrane schematic	3
2.1	Model nanopore setup and genetic algorithm procedure	21
2.2	Genetic algorithm trajectory	22
2.3	Flux and diffusivity in rationally patterned pores	25
2.4	Flux difference for other solutes	30
3.1	Schematic of active learning procedure	40
3.2	Water and boric acid flux in rough pores	44
3.3	Active learning optimization trajectory for 2 headgroups	47
3.4	Active learning optimization trajectory for 3 headgroups	50
3.5	Optimization trajectory for 3 headgroups while varying composition	51
4.1	Hydrophilic polypeptoid chemistries	59
4.2	Comparison of polypeptoid forcefields	64
4.3	Free energy of rotation around the ω dihedral	67
4.4	End-to-end distance distributions from DEER and simulations	70
4.5	Time-domain DEER signal comparison	72
4.6	End-to-end distance distributions from cw-EPR and simulation	73
5.1	Polypeptoids chemistries with varying hydrophilic-hydrophobic sequence	88
5.2	Water structure around hydrophilic peptoid in solution	89
5.3	Water dynamics around hydrophilic peptoid in solution	90
5.4	Dependence of hydration water dynamics on peptoid sequence	93
5.5	Water diffusivity around individual peptoid residues	94
5.6	Water diffusivity around peptoids of varying conformation	96
5.7	Three-body angle distribution around hydrophobic peptoid	97
5.8	Correlated structural and dynamical metrics	98
5.9	Comparison of diffusivity from simulation and ODNP	100
6.1	Surface-tethered polypeptoid system	116
6.2	Potentials of mean force between solutes and peptoid surfaces	119
6.3	Solute affinities for peptoid surfaces	120
6.4	Decomposition of the PMF	126
6.5	Three-body angle distributions near peptoid surfaces	131

6.6	Projections of three-body angle distributions onto principal components .	134
7.1	Peptoids with aromatic groups simulated	144
7.2	Scaling of mean end-to-end distances	148
7.3	End-to-end distance distribution for peptoids with aromatic residues . . .	151
7.4	Scaling of mean end-to-end distances for CG peptoids	152
7.5	Genetic algorithm trajectory for Flory relationship	155
8.1	Relationship between workflow development and fundamental understand- ings	158
A.1	Water diffusivity and density profiles	169
A.2	Nanopore functional group density	170
A.3	Nanopore functional group labels	170
A.4	Trajectory of boric acid dihedral	171
A.5	Independence of flux from boric acid dihedral parameters	172
A.6	Procedure for setting density in nanopore	173
A.7	Procedure for setting density in nanopore	175
A.8	Water and boric acid partition coefficients	175
A.9	Dependence of flux on pressure gradient	176
A.10	Detailed genetic algorithm workflow	178
A.11	Example 3x3 subsection for pattern entropy definition	179
A.12	Minimizing pattern entropy by quenching	179
A.13	Feature diversity with and without pattern entropy minimization	180
A.14	Model optimization with surrogate function	181
A.15	Surrogate model parity plots	184
A.16	Performance of other surrogate models	185
A.17	Spectral clustering of genetic-algorithm-discovered patterns	187
A.18	Water flux versus boric acid flux	189
A.19	Water and boric acid diffusivity	190
A.20	Periodic potential in ringed pores	191
A.21	Pareto front for random and optimized pores	194
B.1	Atom type labels for ethyl and hydroxymethyl groups	196
B.2	Solute in homogeneous nonpolar pores	197
B.3	Solute in pores with dispersed ethyl groups	198
B.4	Water density profiles in homogeneous pores	199
B.5	Three-body angle distributions of water in methyl-ethyl pores	200
C.1	Partial charges for the alcohol monomer	203
C.2	Partial charges for the ether monomer	203
C.3	Polypeptoid atom names	204
C.4	Umbrella sampling distributions	217
C.5	Convergence of expanded ensemble state weights	218

C.6	Convergence of mean end-to-end distances	218
C.7	Dependence of isomerization rate on λ_{ee} and temperature	219
C.8	Correlation times for ω dihedrals	220
C.9	Correlation times for ϕ dihedrals	221
C.10	Time between independent and statistically significant samples	224
C.11	Comparison of DEER analysis workflows	226
C.12	Comparison of water models	226
C.13	Effect of time resolution on distance distributions	228
C.14	Effect of distance resolution on time domain signal	228
C.15	Effect of removing short distances on time domain signal	229
C.16	Snapshots at varying distances for $N = 11$ polypeptoid	229
C.17	Dihedral preferences for hydrophilic polypeptoids and disarcosine	230
C.18	Comparison of labeled and unlabeled $N = 11$	231
C.19	Conformational clustering to remove spin-label aggregated states	232
C.20	End-to-end distance distributions without spin-label aggregated states	232
C.21	Mean end-to-end distances without spin-label aggregated states	233
C.22	Mean end-to-end distances without distances below $20\text{\AA}$	234
C.23	Dependence of predicted distribution on cw-EPR bounds	235
D.1	End-to-end distance distribution for singly labeled peptoids	238
D.2	End-to-end distances in NVE simulations	238
D.3	Radial distribution function for bulk water	239
D.4	Computational and experimental silica systems	240
D.5	Atom names for the 4-OH TEMPO molecule	240
D.6	Survival probability and MSD near TEMPO spin probe	241
D.7	Diffusivities from survival probability versus MSD	244
D.8	Fitted survival probabilities around moving versus fixed centers	244
D.9	Fitted diffusivities from moving-center model	246
D.10	Diffusivities from survival probability, corrected, versus MSD	247
D.11	$P(R_{ee})$ for conformational clusters	248
D.12	$P(R_{ee})$ for conformational clusters when clustering all sequences	249
D.13	Hydration water behavior in conformational clusters	250
D.14	RDFs for peptoids of varying sequence	251
D.15	3-body angle distributions for peptoids of varying sequence	251
D.16	Survival probability for peptoids of varying sequence	252
D.17	Dipole correlation for peptoids of varying sequence	252
D.18	Dipole correlation around individual peptoid residues	253
D.19	Dipole correlation around peptoids of varying conformation	253
D.20	Hydration behavior around NME homopolymer with scaled charges	254
D.21	Hydration behavior around unlabeled peptoids	255
D.22	Dependence of dynamics on sequence with a larger cutoff	256
D.23	Correlated structural and dynamical metrics with a larger cutoff	256
D.24	Correlation of hydration metrics with number of hydration shell waters	257

E.1	Cis-trans isomerization free energy for peptoids at SAM	260
E.2	Distribution of the instantaneous interface	262
E.3	PEO surface number density profile	263
E.4	Diffusivities as a function of distance from the surface	264
E.5	O-H bond reorientation correlation function	265
E.6	Water density profile near peptoid surfaces	266
E.7	Density profiles of peptoid brush surface system	267
E.8	Water density profile near PEO surface	267
E.9	Three-body angle distribution around peptoids in dilute solution	268
E.10	Potentials of mean force for other solutes	269
E.11	Decomposition of PMF for other solutes	270
E.12	Electrostatic solute-system interaction	271
E.13	VDW solute-system interaction	271
E.14	Peptoid root mean squared displacement	272
E.15	Projections of 3-body angle distribution for other solutes	273
E.16	Prediction of $T\Delta S_{\text{rel}}$	275
E.17	Solute-water coordination number	276
E.18	Distance-dependence of projections of 3BAD for phenol	276
E.19	Distance-dependence of projections of 3BAD for other solutes	277
F.1	Autocorrelation of the end-to-end distance	280
F.2	Comparison to theoretical excluded volume distribution	280
F.3	Surrogate model for peptoid conformation	281

List of Tables

4.1	Hydrophilic polypeptoids characterized	58
6.1	Hydration properties of peptoid surfaces	127
A.1	Harmonic bond parameters for nanopore functional groups	170
A.2	Harmonic angle parameters for nanopore functional groups	171
A.3	Generations in each genetic algorithm stage	182
A.4	Model-identified predictive pattern features	186
A.5	Rationally designed patterns	189
A.6	Solute properties	192
B.1	Harmonic bond parameters for ethyl and hydroxymethyl groups	196
B.2	Harmonic angle parameters for ethyl and hydroxymethyl groups	196
C.1	Partial charges for the alcohol monomer	205
C.2	Partial charges for the ether monomer	205
C.3	Partial charges for the C-terminal spin-label	206
C.4	Partial charges for the N-terminal spin-label	206
C.5	Partial charges for the C-terminal amino group	207
C.6	Atom types for the alcohol monomer	208
C.7	Atom types for the NME monomer	209
C.8	Atom types for the spin-label monomer	209
C.9	Atom types for the C-terminal amino group	210
C.10	Bond potential parameters for the MFTOID forcefield	212
C.11	Angle potential parameters for the MFTOID forcefield	212
C.12	Dihedral potential parameters for the MFTOID forcefield	213
C.13	Improper dihedral potential parameters for the MFTOID forcefield	213
C.14	Angle potential parameters for the modified AMBER forcefield	214
C.15	Dihedral potential parameters for the modified AMBER forcefield	214
C.16	Bond potential parameters for the modified CGenFF forcefield	214
C.17	Angle potential parameters for the modified CGenFF forcefield	214
C.18	Dihedral potential parameters for the modified CGenFF forcefield	215
C.19	Polypeptoid simulation settings	222
C.20	Distances between experimental and simulated distributions	225

C.21	Time domain signal parameters	227
C.22	Dependence of predicted mean and standard deviation on cw-EPR bounds	235
D.1	Polypeptoid simulation settings for hydration studies	237
D.2	4-OH TEMPO atom types and partial charges	241
D.3	Cluster populations and average end-to-end distances	248
E.1	Location of surface-water interface	261
E.2	Peptoid-water coordination numbers	266
F.1	Polypeptoid simulation settings with increasing hydrophobic sidechains .	279

Chapter 1

Introduction

1.1 Background

Progress in human health and happiness is built on the development of high-performance materials. Materials “inverse design,” or the construction of materials that achieve a desired functional or structural property, is thus not a novel paradigm, but rather has been a constant theme throughout our history, and has encompassed a wide variety of design strategies, some more efficient than others. These activities predate a nanoscopic understanding of materials structure by thousands of years. However, advances in our ability to probe and model the molecular-level behavior of matter is beginning to fundamentally alter the way we design new materials. In many cases, time-consuming and laborious sequential synthesis-characterization cycles (i.e. “trial-and-error” approaches) can be replaced by high-throughput experimentation^{1,2} and/or model-based navigation of the space of candidate materials (i.e. the “design space”).³ These approaches have

already proposed novel materials in a variety of classes, such as catalysts,⁴ nanophotonics,⁵ proteins,⁶ and polymers.^{7–10} This dissertation will demonstrate the application of atomistic modeling to the inverse design of materials for next-generation water filtration membranes.

Severe water scarcity affects 40% of the world’s population,¹¹ and projections suggest the affected population will only grow with time.^{12–14} Future water scarcity will also impair agricultural productivity and exacerbate food insecurity.^{15,16} Water availability is furthermore closely linked to energy generation, as clean water is required to produce energy and vice versa; this interdependence is referred to as the “water-energy” nexus.¹⁷ National and global governmental agencies agree that development of technologies to alleviate water scarcity is critical.^{18,19} Addressing water scarcity requires solutions tailored to local infrastructure¹² and development of a variety of highly tunable approaches is therefore needed. One such approach is treatment and reuse of wastewater,^{11,20} such as produced water,^{21–23} for specific purposes such as crop irrigation,²⁴ termed “fit-for-purpose” water treatment.²⁵ Such wastewater sources are contaminated with high concentrations of solutes with diverse chemical properties and molecular size, requiring complex treatment trains that could benefit from incorporation of polymer-based water filtration membranes.²⁶

Water filtration membranes are typically deployed in a spiral wound module comprising a rolled-up double layer of membrane material (Fig. 1.1a) arranged such that when water flows along the axis, purified water (“permeate”) partitions across the membrane

material to the region in between the two layers, where it is then directed out of the module *via* a dedicated channel. The remaining water that is concentrated in contaminants (“retentate”) flows across the membrane and out the module.²⁷ For reverse osmosis, membrane materials comprise a dense polymeric layer that sits on top of a porous layer (Fig. 1.1b). The dense layer provides selectivity for water over contaminants, while the porous layer provides mechanical support.

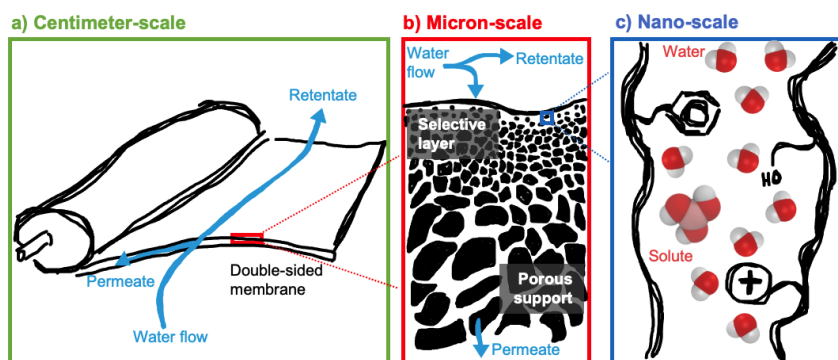


Figure 1.1: (a) Spiral wound module, partly unraveled to show the directions of water flow. Inspired by Fig. 1 of Ref. [27]. (b) In the membrane, the dense layer provides selectivity while the porous layer underneath provides support. Inspired by Fig. 3b of Ref. [28]. (c) Functionalization of the selective layer can enhance rejection of small, neutral solutes such as boric acid or reduce fouling of membrane surfaces.

Membranes separations are typically more energy efficient than those involving phase changes (e.g. distillation).²⁹ However, membranes face significant challenges in purifying highly contaminated water. One key challenge is the separation of small, neutral solutes^{30,31} such as boric acid, phenolic compounds, and 1,4-dioxane, which are present in industrial wastewater at concentrations exceeding standards for human and environmental safety,^{32–35} posing challenges for water reuse. Conventional membranes take advantage of differences in molecular size or charge to perform a separation and thus have

low rejection of uncharged solutes similar in size to water. Current technologies require treatment of contaminated waters in addition to membrane filtration, such as oxidation of phenolic compounds^{36,37} and 1,4-dioxane^{38,39} or manipulation of pH to convert boric acid into charged borate ions,³¹ in order to enhance or bypass low water-solute selectivities, increasing both capital and operating costs of the separation. A related challenge is an inability to separate out specific solutes, i.e. solute-solute selectivity,⁴⁰ which impedes resource recovery efforts. Another issue that plagues membrane technologies is fouling of membrane surfaces, which occurs when solutes adsorb to pore walls or other membrane interfaces, leading to a reduction in permeability, or throughput, over time.^{41,42}

Modification of membrane surfaces has shown promise for addressing these challenges. For instance, chemical functional groups (Fig. 1.1c) at a surface can alter the surface's propensity to foul²⁸ or can tune affinity to specific solutes.⁴³ Such work has largely been based on empirically informed heuristics, for instance that smooth, hydrophilic, non-hydrogen-bonding surfaces generally resist fouling,^{44,45} without an understanding of molecular-level mechanisms. While these heuristics have been useful in developing current membranes, future development of precisely tuned surfaces that can transform highly contaminated water into fit-for-purpose water with high selectivity will likely require optimization of a combination of surface functionalities,⁴⁶ where, for instance, some functionalities contribute antifouling properties while others are tailored for solute selectivity. Design of these multifunctional surfaces requires optimization of chemistries (which groups to use), compositions (surface density of each group), and placement (ar-

rangement of the groups at the surface, if tunable), a large design space that has not been widely searched³¹ and is thus poorly described by current heuristics. A computational inverse design approach, whereby simulations are used to intelligently screen combinations of functionalities, could instead be used to suggest new membrane materials.⁴⁷ These new materials can then be synthesized and tested, and can also (especially if they are non-intuitive) suggest novel mechanisms for surface-modulated behaviors of water and solutes, contributing to an enhancement in the fundamental understanding of aqueous interfaces. This last possibility is a particular advantage of model-driven exploration, which can reach underexplored and poorly understood regions of the design space. In contrast, making iterative improvements on an existing material traverses a relatively smaller, local region.³

Improvements in the molecular-level understanding of interactions at membrane interfaces may also aid the design of solutes and surfaces in other systems involving aqueous interfaces, such as antifouling and antimicrobial coatings, molecular sensors, and protein ligands. While aqueous interfaces are ubiquitous and widely studied, we still lack predictive rules for how water and solutes interact with chemically and topologically heterogeneous surfaces. One factor complicating these interactions is the delicate balance between water’s unique hydrogen bonded structure and the conventional liquid-like structure exhibited by most other solvents.⁴⁸ As a result, the thermodynamics of interfacial water are dependent on the size and chemistry of the structure it is solvating. Around a small hydrophobe (< 1 nm), water can retain, on average, four hydrogen-bonded neigh-

bors by forming a cage-like structure around the hydrophobe, a restriction that results in an entropic penalty of solvation. In contrast, water that solvates a large hydrophobe (> 1 nm) loses, on average, one hydrogen-bonded neighbor, and hence there is an enthalpic penalty of solvation.⁴⁹ These distinct thermodynamic responses determine the driving forces for water-mediated solute-solute and surface-solute interactions,⁵⁰ forming the basis of correlations observed between hydration water thermodynamics and functional properties. Water’s thermodynamic behavior at protein surfaces, for instance, has been shown to be predictive of binding sites.^{51,52}

Unfortunately, hydration water’s thermodynamic behavior remains difficult to predict. The length-scale dependence of solvation is disrupted by the introduction of solute-water interactions,^{53,54} and complicated by solutes with non-trivial topologies,⁵⁵ to which water may exhibit both small and large-lengthscale responses. Analytical models remain suited only to idealized solutes,⁵⁶ with only simple perturbations to their topology⁵⁷ or chemistry,⁵⁸ and simulations to compute hydration free energies rely on expensive enhanced sampling methods.^{59,60} A promising direction is the establishment of correlations between thermodynamic hydration metrics and water’s structural or dynamical behavior, which is less computationally expensive to compute.^{61–65} Additional work to extend such correlations to realistic systems with both chemical and topological heterogeneity is needed.

Classical atomistic molecular dynamics (MD) is especially suited to providing insight into surface-modulated water and solute behavior at the nanoscale. Atomistic MD

takes, as input, parameters describing the forces that the atoms in a system exert on each other and propagates these dynamics to provide the trajectory of each atom in the simulation.⁶⁶ Such resolution is necessary because existing continuum models cannot describe the complex interfacial behavior affecting water transport at membrane surfaces and under confinement.^{67,68} Surfaces affect local water and solute structuring, leading to molecular-scale fluctuations in the density profile,^{69,70} and invalidating the typical assumptions used to develop continuum models of flow through a pore, for instance. Classical atomistic MD can easily access (a few days on a single GPU) timescales of up to hundreds of nanoseconds and lengthscales of up to 10 nm.⁷¹ These scales are convenient to access both the perturbations in water structure due to a surface (around 1 nm),⁷⁰ as well as the influence of a surface on water transport (water’s diffusivity is on the order of $0.1 \text{ \AA}^2/\text{ps}$,⁶¹ giving a characteristic timescale of 100 ps for a nanometer-scale surface). Enhanced sampling methods, such as expanded ensemble and umbrella sampling can be applied to study systems or phenomena with longer characteristic timescales (e.g. molecular isomerization or solute-surface binding and unbinding), as will be discussed in this dissertation. In contrast, *ab initio* methods are typically limited to (again, assuming a few days on a single GPU) tens of picoseconds and systems less than 2 nm in size,⁷² making such investigations intractable, especially as part of an iterative inverse design procedure.

While simulations can examine idealized surfaces, validation of these simulations and facile application of the design rules established from simulations requires a synthetically

accessible materials system. Since we are interested in studying the effects of chemical heterogeneity, this system must also be precisely tunable. A class of molecules satisfying these requirements are polypeptoids, which are peptide-like polymers that can be synthesized sequence-specifically with a wide variety of chemistries⁷³ and tethered to surfaces to modulate surface chemistry.⁷⁴ Interest in polypeptoids has grown only recently,⁷⁵ and thus they have significant untapped potential as a platform for examining the relationship between chemical patterning and local water and solute behavior. We will perform such investigations in the middle sections of this dissertation, and then circle back to demonstrate how inverse design methods may be applied to peptoid sequence design.

1.2 Overview

This dissertation first demonstrates a proof-of-concept study of the use of an inverse design workflow to identify chemical patternings of a model nanopore that optimize rejection of a small, neutral solute (Chapter 2). The genetic algorithm automates the discovery of non-intuitive patterns that improve rejection, enabling a targeted assessment of the design space that both confirms the findings of the genetic algorithm and proposes an underlying mechanism for the observed behavior. This first study illustrates how such automated design workflows may lead to insights into nanoscopic solute behaviors at surfaces. In the next chapter, we demonstrate a significant improvement to the inverse design workflow by recasting the design problem in a lower-dimensional design space and utilizing an active learning approach (Chapter 3). We use this approach to investigate

the combined effect of chemical and topological (i.e. surface “roughness”) patterning on solute rejection.

In order to translate these workflows to an experimentally accessible platform, we validate a model and simulation methodology for polypeptoids in collaboration with advanced experimental characterization techniques (Chapter 4). This type of detailed validation of conformational distributions, as opposed to single average quantities, is novel for polypeptoids, and provides confidence in the methods used in subsequent chapters. Next, we demonstrate that changes in polypeptoid sidechain chemistry tune its hydration water structure and dynamics, even in dilute solution (Chapter 5). These trends suggest that water responds differently to these molecular surfaces compared to idealized, extended surfaces. We then show that polypeptoid brush surfaces also give very different hydration behaviors compared to ideal surfaces (Chapter 6). These hydration behaviors are related to changes in small-molecule affinity for these surfaces, providing insight for the design of antifouling coatings.

Characterization of these surfaces points to the underlying complexity of solute-surface interactions and realistic membrane interfaces; both chemical and topological patterning affect hydration and water-mediated surface-solute interactions in correlated and poorly understood ways. We end by exploring one additional complexity, which is that chemical patterning can modulate topology by changing the conformation of the polypeptoid chain (Chapter 7). We again apply the inverse design workflow to show that sequence can be tuned to target specific conformational behaviors.

Chapter 2

Inverse design of pore wall chemistry to control solute transport and selectivity

Adapted with permission from *ACS Cent. Sci.* 2022, 8, 12, 1609–1617. Copyright 2023 American Chemical Society.

2.1 Introduction

The *in silico* inverse design of materials has long been a goal of computational modeling, promising accelerated materials discovery combined with reduced experimental cost. While system size and complexity have challenged earlier efforts, modern hardware and algorithms have given rise to many recent successes in the optimization of

chemical functionalities. These efforts span a wide range of materials, from the design of small molecules,^{76,77} molecular complexes,⁷⁸ and functionalized interfaces^{61,79} to self-assembling colloidal⁸⁰ and polymeric systems.⁸¹ In some cases, existing structure-property relationships or learned empirical models simplify navigation of design space. Often, however, such models are inadequate or nonexistent, and molecular simulations are required to predict material properties, greatly increasing the computational expense of these design efforts. Soft material systems involving solution phases pose particularly difficult design problems, where the difficulty can be due to significant chemical heterogeneity, the diversity of interactions (excluded volume, electrostatic, van der Waals), and the emergence of complex driving forces involving entropies, such as hydrophobic interactions, that necessitate atomistic-level fidelity and long simulation timescales. Such systems often involve very large design spaces that are not easily enumerable or navigable, as they lack obvious order parameters along which optimization can proceed, presenting additional challenges. There is therefore a pressing need for soft material optimization approaches that make efficient use of generated simulation data, such as by incorporating on-the-fly learned models.⁸²

A particularly important complex class of materials that encompasses these challenges, and is a major opportunity for computational soft materials design, is water purification membranes. In this work, we demonstrate an inverse design approach to design the chemical functionality of a nanopore wall to optimize solute rejection in water separation membranes. We establish an optimization workflow coupling molecular

simulations and machine learned surrogate models and demonstrate its capabilities for complex soft material problems. We furthermore demonstrate how algorithm-generated materials can educe novel and unintuitive design rules to guide synthetic efforts and advance the fundamental understanding of the relationship between surface chemistry and the transport properties of water and solutes.

Water scarcity will become an increasingly pressing issue in the coming decades.⁸³ Energy security is intimately linked to the sustainability of current water resources,¹⁷ and thus efficient water treatment processes are crucial for employing alternative lower quality water sources. The maturity of membrane technologies for low-energy, reliable treatment of seawater through reverse osmosis has led to its widespread adoption for desalination^{84,85} and utility for water reuse and other advanced water treatment systems.⁸⁶ However, membranes are not yet capable of economically treating highly impaired waters, such as oilfield produced water, currently an enormous wastewater stream on the order of 2 billion gallons per day.⁸⁷ The development of membrane separation processes providing energy efficient routes to converting such highly contaminated sources into “fit-for-purpose” water²⁶ (e.g. for irrigation, stream flow augmentation, power) as well as recovering valuable resources from these waste streams remains a grand goal,^{31,88–90} but is hindered by key challenges associated with the concentration and broad chemical variety of contaminants. For instance, the separation of small, charge-neutral solutes is particularly difficult for current membranes that leverage charge and size separation mechanisms.³¹ One exemplary solute is boric acid. While the world health organization

guidelines recommend drinking water concentrations of less than 2.4 mg/L of boron,⁹¹ water reuse for agricultural irrigation of most crops requires even lower concentrations (1 mg/L).^{92,93}

Biology provides inspirational design considerations. Transmembrane proteins, such as aquaporin⁹⁴ and potassium ion channels,⁹⁵ achieve impressive permeability and selectivity through nanoscale confinement and precise arrangement of protein side-chain functional groups within the pores, shaping the chemical and topological landscape through which water and solutes transport. A number of efforts have made progress along this direction in synthetic systems.⁴⁶ Nanochannels such as carbon nanotubes can exhibit unique^{96,97} and ultrafast transport of water,^{98,99} where the strength of the wall-water interaction modulates water transport.¹⁰⁰ Functionalizing the pore also affects water adsorption¹⁰¹ and flux¹⁰² and modulates solute rejection,^{103–107} depending on the functionalization chemistry. Previous studies have also demonstrated that the location of the functional groups affects water behavior^{101,102} and solute transport.¹⁰⁸ However, a critical knowledge gap is understanding the potential impact of patterning pore surfaces with more than one type of functional group on engineering solute rejection and selectivity for the large class of neutral small-molecule solutes. The presence of more than one chemistry at a surface exponentially expands the design space. Indeed, previous computational work has shown that both the number and arrangement of chemical functional groups at an extended surface can significantly tune the surface affinity of model solutes,⁷⁹ as well as the local diffusivity of water.⁶¹ Design rules and strategies to iden-

tify optimal, heterogeneous chemistries are largely unknown and would have implications beyond water filtration membranes to other chemically heterogeneous surfaces, such as porous adsorbents, catalytic surfaces, and antifouling interfaces, as well as heterogeneous macromolecules such as proteins and sequence-defined polymers.

In this work, as a case study, we focus on rejection of a neutral solute, boric acid, and demonstrate a novel workflow to precisely tune membrane pore wall spatial chemistry to enhance solute rejection. While Edisonian approaches in the membrane community have identified important heuristic membrane material characteristics affecting e.g. fouling^{28,109} and solute uptake,^{110–113} such approaches only skim the large design space associated with multifunctional chemistries. Instead, here we describe a computational inverse design workflow that optimizes the relative transport of water over boric acid, merely by patterning nonpolar methyl and polar hydroxyl groups tethered to a model pore. This proof of concept demonstrates the potential of computational materials discovery in uncovering the existence of non-intuitive, heterogeneous pore wall chemistries that enhance solute rejection. In particular, the optimization workflow suggests that pores with rings of hydroxyl groups pose selective barriers to boric acid transport over water, suggesting a new class of materials to explore for water separations. Based on these computationally designed pore surfaces, we identify a mechanism behind the enhancement in boric acid rejection captured by a simple, predictive analytical model. We then demonstrate that the optimal pore design gives enhanced selectivity for a range of other solutes that vary with respect to polarity, hydrogen-bonding capability, and surface

affinity, suggesting the generality of the proposed mechanism.

2.2 Methods

2.2.1 Simulation models

To model an ideal nanopore with tunable pore wall chemistry, we simulate a (16,0) zig-zag single-wall carbon nanotube (CNT) with nonpolar methyl and polar hydroxyl functional groups directly tethered to the interior of the pore. At smaller radii, the diffusivity of water inside the pore decreases rapidly due to strong confinement effects (Fig. A.1a), consistent with experimental work on water in porous silica showing an enhancement in water network structures with decreasing pore size.^{114,115} We thus choose a slightly larger radius of 1.25 nm, for which water reaches its bulk density in the middle of the pore (Fig. A.1b), in order to focus on the effects of surface chemistry as opposed to confinement, since confinement can significantly alter water and solute behavior in nanoporous environments.^{116–118} The functional groups are placed every 4 carbons (Fig. A.2) to achieve a surface density of 0.21 nm²/group, comparable to the surface density for realizable functionalized surfaces, such as self-assembled monolayers.¹¹⁹ We obtain nonbonded forcefield parameters for the CNT and functional groups from Ref. [120]. Section A.1.3 describes the bonded parameters in more detail. Parameters for water are from the TIP3P water model¹²¹ and those for boric acid are based on the classical model parameterized in Ref. [122] (section A.1.3). Section A.4.5 describes the parameters for

the other solutes examined.

To compute transport properties of water and boric acid through the pore, we simulate an “infinitely” long nanopore (Fig. 2.1b) with 16 rows of carbon atoms, yielding a pore of length 3.4 nm with 128 functional groups, allowing a diverse set of patterns. To compute water and solute densities inside the pore, we first simulate a “finite” pore in between two water reservoirs bounded by graphene sheets (Fig. 2.1b, details in section A.1). We model all systems using the OpenMM simulation engine.⁷¹ Electrostatic interactions are computed using particle-mesh Ewald. Lennard-Jones (LJ) nonbonded interactions are cut off at $10\text{\AA}$. Hydrogen bonds are constrained with SHAKE¹²³ and water is kept rigid with SETTLE.¹²⁴ We perform an energy minimization before all simulations. We use Langevin dynamics with a timestep of 2 fs, a temperature of 298 K, and a friction coefficient of 1 ps^{-1} . For NPT simulations, we use an anisotropic Monte Carlo (MC) barostat that only adjusts the box size in the axial direction, with 200 fs between MC moves to set the pressure to 1 bar. All carbon atoms in the nanotube and graphene sheets are fixed during the simulation. More details, including the lengths of the equilibration and production stages for each simulation, are given in section A.1.4.

2.2.2 Finite nanopore simulation

For all simulations with the finite pore, we restrain the boric acid molecule so that it stays within the nanopore (section A.1.5). To determine the partition coefficient of boric acid, we first compute its free energy of solvation inside the pore *via* an expanded ensemble

simulation where the boric acid LJ ϵ parameters and partial charges are slowly turned on in several distinct states. The procedure is similar to that in Ref. [79] except that an additional restraint keeps boric acid inside the pore (section A.1.5). The free energy of solvating boric acid, ΔG , is the difference between the states in which the boric acid molecule is fully on and off, which we compute using MBAR.^{125,126} We then compute the partition coefficient, K , from $K = K_0 \exp[-(\Delta G - \Delta G_0)]$ where $\Delta G_0 = -3.4k_B T$ and $K_0 = 8.0$ are reference values computed from an unbiased simulation of the fully methylated pore (section A.1.6).

2.2.3 Infinite nanopore simulation

To probe the diffusivity of water and boric acid, we compute the axial mean squared displacement for the water oxygens and boron atom and extract the diffusivities from the slope of the MSDs. To compute the nonequilibrium flux of water and boric acid under an applied pressure gradient, we add a constant acceleration to each water and boric acid atom such that the pressure gradient is 12 MPa/nm. In this regime, water flux remains linear with pressure gradient (Fig. A.9), allowing extrapolation to lower, more realistic gradients. We run a simulation with the applied pressure gradient and compute the total net displacements of the water oxygens and boron atoms in the axial direction. The in-pore flux is the displacement divided by the length of the simulation, the volume of the pore, and the number of molecules. This “flux” calculation measures the rate at which individual water or solute molecules move through the pore and does not have a

direct contribution from their densities in the pore, which is convenient for separating the effects of pore wall chemistry on partitioning into the pore and transport through the pore. To obtain metrics that combine both effects, we compute the permeability by multiplying the partition coefficient by the computed in-pore flux, pore volume, and bulk water density, and dividing by the pressure gradient. We compute the selectivity from the ratio of the product of the partition coefficient and in-pore flux for water and boric acid, and then divide by the approximate mole fraction of boric acid in seawater,²¹ approximately 1.4×10^{-6} .

2.2.4 Genetic algorithm structure

The genetic algorithm, depicted in Fig. 2.1c and in more detail in Fig. A.10, follows the general procedure established in Ref. [79] but is used here to find the arrangement of 128 hydroxyl and methyl groups that optimizes the difference between water and boric acid flux. We perform two genetic algorithm runs, one in which the number of surface hydroxyl groups is fixed at 32 (“fixed fraction”) and one in which it is free to vary (“varying fraction”). Both runs begin with 5 generations of randomly generated patterns. For the fixed fraction run, we generate initial patterns by randomly choosing 32 sites for the hydroxyl groups. For the varying fraction run, the number of hydroxyls is randomly chosen at a given hydroxyl fraction spanning 0 to 1. We then quench these initial randomly generated patterns to minimize the pattern entropy, creating a more feature-diverse set of motifs (Fig. A.13) to initialize the algorithm, improving its convergence. Here, the

pattern entropy is $S_{pat} = -\sum_{i=1}^{512} P_i \log P_i$ where P_i gives the probability (fraction) of 3×3 subsections with a given pattern among all 512 distinct 3x3 motifs (see section A.2.1 for details). In the genetic optimization runs, the initial “random” generations are followed by a subsequent set of maximization generations where the flux difference is evaluated from molecular dynamics (MD) simulations, as described above. Because the simulation-evaluated flux differences are computationally expensive, we fit a surrogate model to the initial simulated generations to predict the flux difference based on a linear combination of 22 pattern features, chosen to elucidate various spatial correlations in the arrangement of the functional groups (section A.3.1). We then rapidly iterate over many generations using the surrogate model, followed by additional generations with explicit MD evaluation of the best candidates, and finally additional surrogate-model-evaluated generations with a retrained surrogate model. Table A.3 gives the number of generations in each stage. Because the varying hydroxyl group fraction case explores a much larger pattern space, we increase the number of generations compared to the fixed fraction run.

2.3 Results and Discussion

2.3.1 Genetic algorithm optimization

Our basic question is: to what extent can surface chemical patterning be leveraged to optimize the relative transport of water versus neutral solutes in a pore? Here, we use boric acid as a model solute, which is difficult to separate due to its small size and charge

neutrality at neutral pH, and we consider nanopores of varied hydrophobic/hydrophilic content, tuned by patterning the interior of a carbon nanotube with hydroxyl and methyl functional groups (Fig. 2.1a). To compute transport properties of water and solutes inside the pore, we first simulate a “finite” pore attached to a water reservoir to equilibrate the density of water inside the pore, and then simulate an “infinite” pore with the same density (Fig. 2.1b), from which we compute water and boric acid transport metrics, such as equilibrium diffusivity and non-equilibrium, pressure-driven flux (details in section 2.2). To assess the potential of chemical patterning to enhance pore transport of water relative to boric acid, we use a bio-inspired evolutionary or genetic algorithm (Fig. 2.1c) that identifies arrangements of methyl and hydroxyl functional groups on the pore wall that optimize the difference between the computed water and boric acid fluxes. To accelerate the optimization, we employ a machine-learned surrogate objective function trained on prior explicitly simulated patterns to predict the flux difference, which is periodically retrained with additional batches of explicit molecular dynamics (MD) calculations. The model allows exploration of orders of magnitude more generations than is possible with MD alone due to computational expense. Because we use a linear model in the feature space, the surrogate model also serves to average out statistical noise from the MD-computed flux, giving the genetic algorithm robustness to uncertainty (Fig. A.14).

This optimization approach detects novel functionalization strategies: Fig. 2.2 shows the trajectory of the flux difference versus number of generations for genetic algorithm runs (a) fixing the hydroxyl group fraction at 0.25 (32 hydroxyl groups) and, (b) allowing

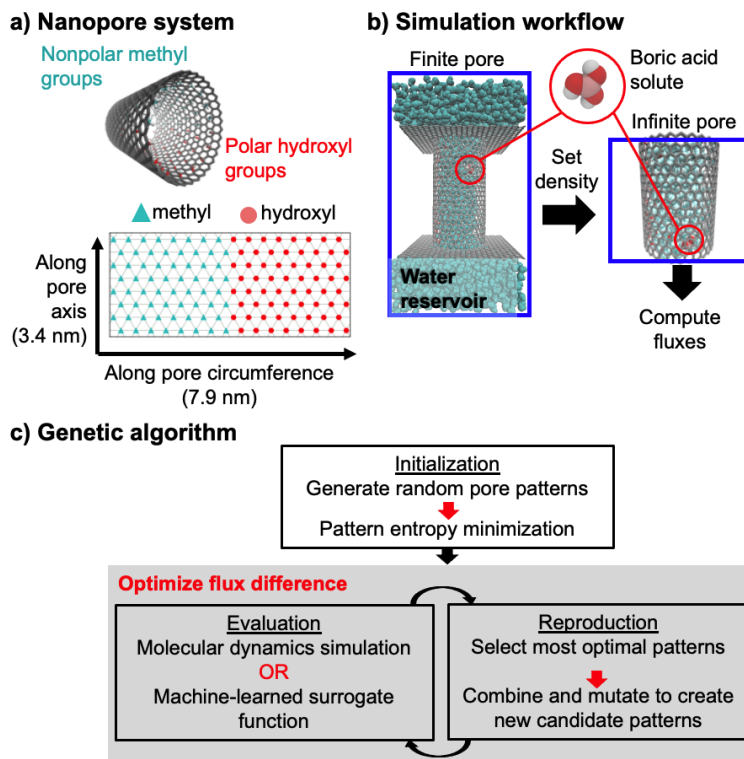


Figure 2.1: (a) The model nanopore system allows for precise patterning of nonpolar methyl groups and polar hydroxyl groups along the CNT backbone. For ease of visualization, we present the flattened pore patterns, with blue triangles representing methyl groups and red circles representing hydroxyl groups. (b) We first simulate the finite pore to compute the density of water, which we then set in the infinite pore to compute water and boric acid transport properties. (c) A genetic algorithm incorporating MD simulations and a learned surrogate function allows for automated identification of pore wall chemical patterns optimizing pore transport properties.

the hydroxyl fraction to vary. Both trajectories span a large range of flux differences and thus demonstrate that chemical patterning can be tuned to enhance solute rejection. The optimal pattern found at fixed hydroxyl fraction contains two nearly perfect rings of hydroxyl groups (shown in the 2D schematic as rows), while the variable hydroxyl case with a much larger search space identifies an optimal pattern that contains a single ring of hydroxyl groups. The optima suggest that rings of hydroxyl groups inside the nanopore particularly inhibit boric acid transport with respect to water transport, a non-obvious

result. Patterns that include these features are not identified in the first set of simulated generations, showing that the surrogate function is key to exploring a sufficiently large number of generations to identify these optimal motifs.

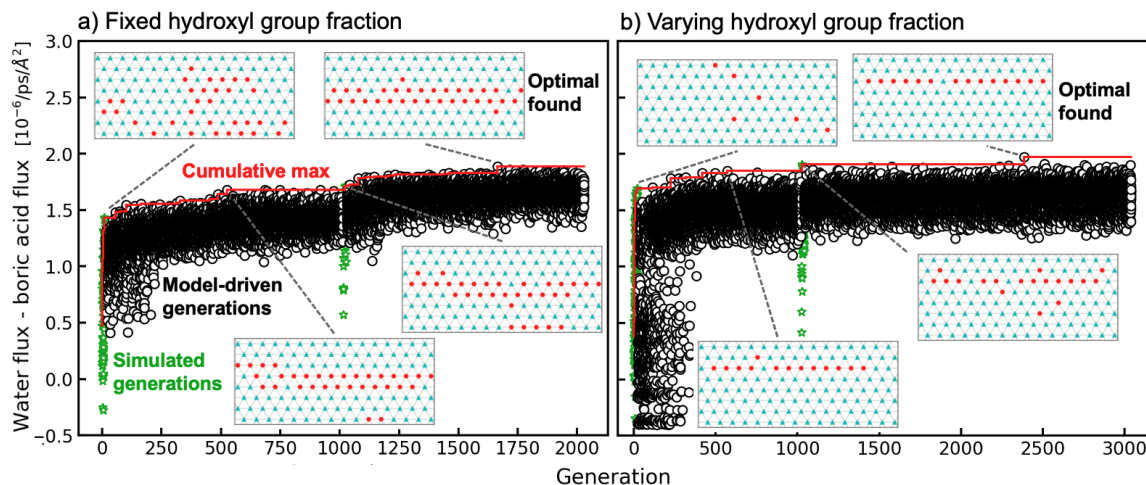


Figure 2.2: Genetic algorithm trajectories maximizing the difference between water flux and boric acid flux for (a) fixed hydroxyl group fraction and (b) varying hydroxyl group fraction show optimization to surfaces with rows of hydroxyl groups. Each marker corresponds to a pattern discovered by the genetic algorithm. The simulated generations are shown as green stars and the surrogate model-driven generations as black circles. The inset patterns show the optimal patterns found after (1) the first set of simulated generations, (2) the first set of model-driven generations, (3) the second set of simulated generations, and (4) the second set of model-driven generations. The red line shows the maximum value of the objective function at each point in the optimization (the cumulative maximum value).

The surrogate model is key to enabling high-throughput optimization. We find that it only needs to maintain accuracy in predicting the flux difference of patterns similar to those currently explored by the genetic algorithm (“local” in pattern space) to enhance performance, with periodic retraining through new MD simulations ensuring its ongoing relevance. Fig. A.15 shows that the linear model is indeed able to fit the “training” data (a single set of explicit MD simulations) well, suggesting good local accuracy. The model also predicts the flux difference of patterns outside of the training set qualita-

tively well, giving accurate directional trends, which is the key requirement to advance the genetic algorithm. The quantitative accuracy is also reasonable given the statistical uncertainty associated with the MD simulations (Fig. A.15). Beyond the linear model, we find that neural-network-based surrogates do not perform significantly better (Fig. A.16) and furthermore require more extensive hyperparameter selection, while sacrificing interpretability. We instead take advantage of the interpretability of simpler models (specifically, LASSO and random forest regression) to compute feature importance metrics demonstrating that spatial correlations of hydroxyl groups along the pore axis and circumference, and especially rows of hydroxyl groups, are the most important features for the regression (Table A.4), reinforcing the idea that these motifs are significant.

2.3.2 Rationally designed patterns suggest mechanism for boric acid rejection

Inspired by the genetic algorithm results, a rational exploration of the effect of pore patterning on water and solute transport reveals a rich design landscape. Fig. 2.3 visualizes this design space, by showing the (a) water flux and (b) boric acid flux as a function of hydroxyl fraction in regularly patterned pores (Table A.5), chosen as idealized versions of the motifs identified from the genetic algorithm-discovered patterns (Fig. A.17). Fig. 2.3a clearly illustrates that pore wall chemistry significantly modulates water transport: increasing the number of hydroxyl groups generally reduces water flux, with the fully methylated and fully hydroxylated pores capturing the extreme cases with

an approximately 4-fold difference in flux. However, even at a fixed fraction of hydroxyl groups, the specific pattern of the groups also significantly affects transport: for instance, patterns with a hydroxyl group fraction of 0.25 span a 2-fold change in flux difference. In contrast, previous work that “patterned” a CNT with chemically-unspecific LJ sites of varying wall-water interaction strengths revealed that functional group arrangement did not affect water flux.¹²⁷ This present study thus suggests that functionalization with chemically-specific moieties involving e.g. hydrogen bonding interactions is required to observe pattern-dependent trends, consistent with previous work showing the dependence of water diffusivity on surface patterning.⁶¹

In particular, the patterns with the hydroxyl groups in a single patch spanning the pore axis (orange triangles) lead to maximal water flux out of all patterns examined, while those that disperse the hydroxyl groups (blue circles) produce minimal flux; intermediate behaviors are seen from patterns with hydroxyl groups arranged in multiple, smaller patches (green diamonds). The diffusivity of water in these pores from equilibrium simulations recapitulates the same nonequilibrium flux trends (Fig. A.19). These behaviors are consistent with earlier observations at extended self-assembled monolayer surfaces in which both increasing the number of hydroxyl groups and dispersing them more evenly across the surface lead to a reduction in surface water diffusivity.⁶¹ More broadly, the pattern-dependent trends follow similar findings for water near surfaces where increased dispersion of polar sites leads to reduced surface hydrophobicity, which manifests as reduced water density fluctuations,^{128,129} reduced tendency to dewet,^{130–132} and reduced

solute affinity.⁷⁹

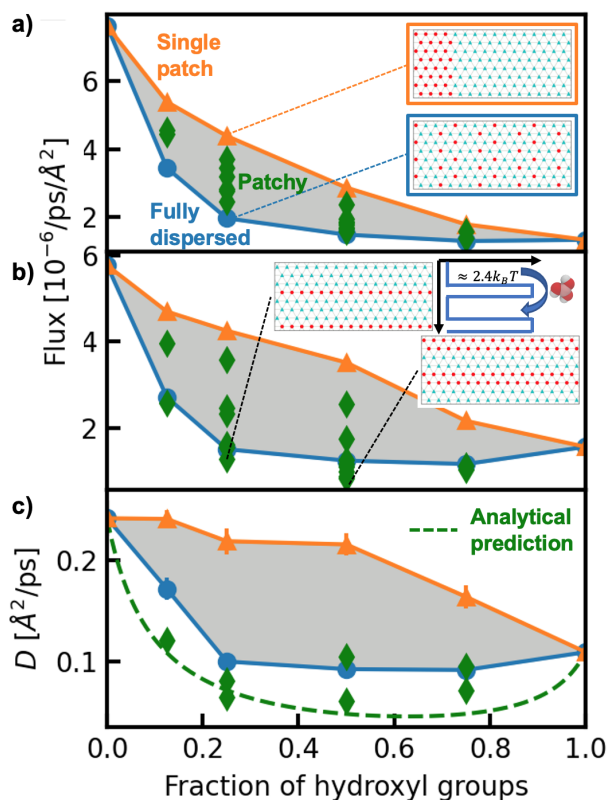


Figure 2.3: The pattern of methyl and hydroxyl groups has a significant effect on the flux of (a) water and (b) boric acid in nonequilibrium simulations and on (c) the diffusivity of boric acid in equilibrium simulations. For each panel, the orange triangles represent patterns with a “single patch” of hydroxyl groups, spanning the pore axis, while the blue circles represent “fully dispersed” patterns (examples in the inset of (a)). The green diamonds represent other patterns where the hydroxyl groups form smaller patches (“patchy”). Two examples of such patterns are shown in the inset of (b). These patterns also show lower boric acid flux compared to the fully dispersed patterns. The inset also includes a schematic of the periodically varying free energy landscape used in the analytical model to predict the reduction in boric acid diffusivity in the ringed pores. In panel (c), the green diamonds are only shown for patterns with rows of hydroxyl groups. The dashed line shows the analytical prediction described in the text based on a model of the periodic free energy landscape depicted in the inset of (b).

Fig. 2.3b’s design space for boric acid shows that pore wall chemistry generally affects its transport similarly to water; however, the patterns that minimize boric acid transport for a given hydroxyl fraction are not fully dispersed but instead contain rows

of hydroxyl groups (Fig. 2.3b inset). These motifs use rings of hydroxyl groups to reduce boric acid flux relative to water (Fig. A.18), rationalizing the optimal surfaces located by the genetic algorithm. For the ringed pores (Fig. 2.3b inset), boric acid flux is less than half that of water, a reduction that is sufficient to reduce boric acid concentration in seawater (4.5 mg/L)⁹² to the target concentration for drinking water (2.4 mg/L).⁹¹ We note, though, that treating groundwater, which can contain higher concentrations of boric acid, or achieving a lower target concentration for irrigation water would require recycling.

To understand the origin of this unique behavior, we hypothesize that two mechanisms are important: (1) boric acid strongly partitions to the hydrophobic, methylated regions of the pore wall, and (2) it is then forced to “hop” over the hydroxyl rings, with an energy barrier determined by rearrangements to the hydrogen bonding network. Boric acid, while highly nonpolar, shows affinity for methylated wall sections of the pore due to its small size; earlier simulation results demonstrate that water drives boric acid to methylated surfaces as a way to minimize the penalty associated with water restructuring around the solute in bulk solution distant from the interface.⁷⁹ To assess this hypothesis, we compare measurements of diffusivity from equilibrium simulations to predictions of an analytical model from Ref. [133] that describes the 1D diffusive transport of a particle (here, boric acid) in a periodic energy landscape (Fig. 2.3b, inset). We estimate the height of the free energy barrier from moving between methylated sections (across hydroxylated sections) to be $2.4k_B T$, based on the partition coefficients of boric acid into

fully methylated versus fully hydroxylated pores. The periodicity of the landscape is then set by the geometry of the pattern, i.e. the number of methylated and hydroxylated rows (details in section A.4.4). We combine this prediction with a simple exponential decay that describes the background reduction in diffusivity due to surface hydroxyl group fraction (section A.4.4). The resulting model then predicts boric acid diffusivity (Fig. 2.3c, dashed line) in agreement with equilibrium simulations for pores with various arrangements of rows of hydroxyl groups. This mechanism is consistent with earlier work on the reduction in water dynamics in rough versus smooth pores, concluding that the primary mechanism is the creation of a rough energy landscape.¹³⁴

At higher fractions of hydroxyl groups (0.50 and 0.75), not all patterns with rows of hydroxyl groups agree with the analytical prediction. We hypothesize that at least two consecutive rows of methyl groups are required to produce an effective hydrophobic surface environment, as previous work has shown that polar groups affect local water behavior near adjacent methylated regions¹³⁰ such that isolated methyl groups have little effect on local water behavior.¹²⁹ Consistent with this hypothesis, the patterns showing the largest deviations from the analytical model each contain isolated single rows of methyl groups. At lower hydroxyl group fraction, the agreement between the analytical model and MD diffusivities is consistent with the proposed hopping mechanism, reinforcing this potential route to inhibiting solute transport while maintaining water transport. These observations underscore the relevance of non-additive, context-dependent interactions to the design of hydrated systems. Here, the arrangement of hydroxyl groups affects

solute transport, independent of the average density of the hydroxyl groups and even beyond basic characterizations of pattern (i.e., degree of clustering), thus demonstrating behaviors that are highly non-linear on a nanoscopic scale.

Transport through the pore is not the sole factor affecting membrane performance. Water and solutes must first partition into it: both the number and arrangement of hydroxyl groups affects water and boric acid partition coefficients (Fig. A.8). While patterns with more hydroxyl groups or dispersed hydroxyl groups lead to enhanced water partitioning and reduced boric acid partitioning and flux, they also produce reduced water flux, suggesting a tradeoff between permeability and selectivity, a commonly observed relationship for other membrane materials that captures how improvements to solute rejection tend to be at the expense of reducing throughput.¹¹⁰ This tradeoff is a key challenge for membrane design efforts. To assess this behavior, we compute the permeability from water’s partition coefficient (tendency to go into the pore) multiplied by its in-pore flux (transport once inside the pore). We compute the selectivity from the ratio of the permeability of water to that of boric acid. The calculated permeability and selectivity values generally reinforce the tradeoff: for instance, water permeability increases from the fully hydroxylated pore (0.703_4 mol-nm/cm²-s-MPa) to the pore with an equal number of hydroxyl and methyl groups arranged in single, axial patches (Fig. 2.1a) (1.300_4 mol-nm/cm²-s-MPa), and again to the fully methylated pore (2.866_8 mol-nm/cm²-s-MPa). However, selectivity for water over boric acid decreases along that same series ($57_1 \times 10^4$, $9.3_2 \times 10^4$, $5.8_3 \times 10^4$). Beyond these specific patterns, randomly gen-

erated pores also demonstrate this tradeoff (Fig. A.21). However, the optimal genetic optimization patterns display improved permeability-selectivity behavior compared to randomly patterned pores (Fig. A.21), suggesting that optimization of the flux difference may produce pattern motifs that address the permeability-selectivity tradeoff.

To further demonstrate the generality of these pattern motifs, we choose a diverse set of other neutral solutes (phenol, benzene, isopropanol, ammonia, arsenous acid) to simulate with several exemplar patterns. These solutes possess a wide range of affinities to methylated and hydroxylated self-assembled monolayers (measured in Ref. [79]), as well as a diverse set of surface hydrogen-bonding behaviors, as quantified by the average number of hydrogen bonds formed with the fully hydroxylated pore (Table A.6). Out of these solutes, phenol, benzene, and isopropanol similarly demonstrate enhanced water flux relative to boric acid flux in pores with rings of hydroxyl groups (Fig. 2.4), suggesting that the proposed mechanism may be broadly generalized to enhance the separation of other solutes from water. Consistent with the proposed diffusive hopping mechanism, there is a correlation between the enhancement in solute rejection in ringed pores and the difference in affinity for methylated versus hydroxylated SAMs (Table A.6), as larger differences in affinity raise the free energy barrier to move between ring-separated methylated sections. For instance, ammonia, which has the lowest difference in affinities for the methylated and hydroxylated surfaces, has the lowest relative water flux in the fully methylated pore. Measures of affinity to homogeneous functionalized surfaces are thus predictive of this effect of patterned functionalizations on solute rejection.

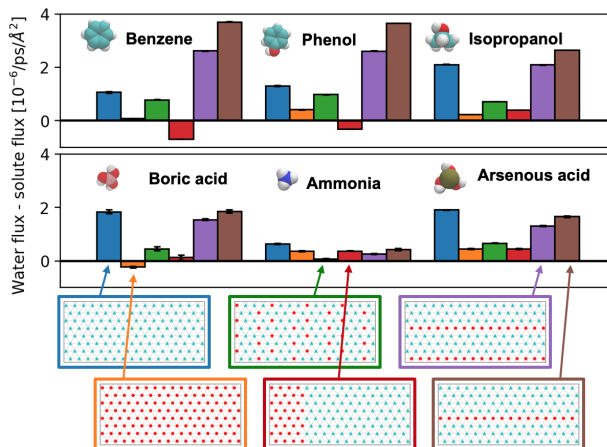


Figure 2.4: Other solutes exhibit similar trends as boric acid in a selection of rationally patterned pores. Benzene, phenol and isopropanol show maximal water flux relative to boric acid flux in pores with rows of hydroxyl groups, while for ammonia and arsenous acid, the optimal pattern is the fully methylated one. For the simulations with solutes besides boric acid, we use the same workflow as described for boric acid except that the production simulations in the infinite pore are run for $1 \mu\text{s}$. The reported uncertainties are the standard error of the mean from three independent simulations.

2.4 Conclusions

In this chapter, we demonstrate that pore wall patterning of nonpolar and polar functional groups significantly affects the partitioning and transport of water and solutes and offers new strategies to design pore chemistries. We show that inverse design algorithms automatically identify unique, non-intuitive chemical patterning strategies and associated mechanisms that decouple and thus optimize the flux of water relative to solutes of interest. Here, for example, we find that pores with higher rates of boric acid rejection contain rings of hydroxyl groups. The underlying mechanism identified by the evolutionary optimization is unexpected: boric acid transport in optimal pores is well-described by a periodic barrier-hopping process, whereby boric acid adsorbs at the methylated regions of the pore wall and then must hop over hydrogen-bonding hydroxyl rings. These results

suggest that leveraging functional groups to create a fluctuating free energy landscape for a solute in a pore may be an important mechanism for enhancing solute rejection at little cost to water permeability. More generally, this approach illustrates the use of computational inverse design algorithms to design membranes and other interfacial systems to identify unique, non-obvious solutions that can then suggest theoretical models for guiding design strategies.

Our specific results imply that as synthetic advances make multifunctional membrane surfaces more accessible, inverse design procedures will be key to exploring the larger design space. While direct functionalization of membrane surfaces with precisely-patterned methyl and hydroxyl groups may not be currently feasible, lithographic techniques with block copolymer assembly, as well as covalent or noncovalent functionalization of nanochannels with biomolecules or polymers,^{106,135,136} may be promising routes to introducing multiple chemistries at a pore surface. Surface-tethered sequence-specific polypeptoids, which we study extensively in subsequent chapters, may also provide a route to chemical functionalization of membrane interfaces. The actual design space therefore potentially includes a large set of possible functionalities, much larger than the two functionalities examined here. Beyond multi-functional pore walls, other possible extensions of this pore design problem include (1) varying the pore radius, as the difference in solute and water affinity for the pore wall may offer a means to enhance water flux in the bulk region of the pore while still relying on the pore wall to enhance solute rejection and (2) varying the topology or “roughness” of the pore surface, which has been shown

to modulate local water behavior. These expanded design spaces make the development of automated design algorithms even more important, an area we will explore further in the next chapter. Beyond membranes, workflows for the precise and efficient *in silico* design of chemical patterns would be useful for identifying optimal antifouling surfaces and directing biomolecular folding and assembly, among other applications.

Acknowledgments

This work was supported by the Center for Materials for Water and Energy Systems (M-WET), an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Basic Energy Sciences under Award DE-SC0019272. Use was made of computational facilities purchased with funds from the National Science Foundation (OAC-1925717) and administered by the Center for Scientific Computing (CSC). The CSC is supported by the California NanoSystems Institute and the Materials Research Science and Engineering Center (MRSEC; NSF DMR 1720256) at UC Santa Barbara. S.J. acknowledges support by the National Science Foundation Graduate Research Fellowship (DGE 1650114). We thank M. P. Howard for helpful discussions concerning nanopore transport. We also thank M. R. Landsman for insight regarding boric acid solutes. Finally, we thank J. B. Rawlings for useful suggestions related to developing the optimization workflow.

Chapter 3

Inverse design of pore wall chemistry and topology

3.1 Introduction

In the previous chapter, we used an inverse design approach to demonstrate that changes in the pattern of surface-tethered methyl and hydroxyl groups had an effect on the rejection of boric acid through a nanopore. In this chapter, we explore whether changes in the topology of the pore wall (i.e. its “roughness”), in addition to its chemistry, can also modulate the transport of water and solutes.

The topology of a surface has been shown to influence hydrophobicity and solute affinity, though the nature of such effects is dependent on the characteristic length-scale of the surface roughness. “Superhydrophobicity,” for instance, is created by lay-

ered micron-scale and nanometer-scale oscillations in surface topology, termed the lotus-leaf effect.^{137,138} Reduction in tens-of-nanometer-scale surface roughness reduces colloidal fouling.⁴⁵ Experimental studies at smaller lengthscales are difficult to interpret due to the difficulty of controlling and characterizing sub-nanometer-scale roughness. For instance, while contact-angle measurements of self-assembled monolayers (SAMs) comprising mixtures of short and long alkane chains suggested different behaviors compared to single-component SAMs, X-ray photoelectron spectroscopy measurements showed that the composition of mixed SAMs at the surface was not equivalent to the solution composition and also likely not at equilibrium.¹³⁹

Simulations, on the other hand, can characterize surfaces with tunable roughness. One common finding that has emerged is the enhanced hydrophobicity of concave pockets. This is reflected in the behavior of water, which appears to be closer to vapor-liquid equilibrium in such settings. For instance, water in a hemicylindrical carbon nanotube is more compressible than near a flat, graphene surface.¹²⁸ Water near crystalline polymeric surfaces with sub-nm-scale indentations has higher interfacial free energy compared to the planar surface.¹⁴⁰ Cavitation between melittin dimers is attributed to the concave geometry of the dimer surface.^{141,142} Solute affinity is also enhanced near concave pockets.⁵⁵ Even with the addition of strong surface-water interactions, water still dewets near rough surfaces.¹⁴³ An interesting counterpoint is the work by Van Lehn, Abbott and coworkers on disordered versus ordered SAM surfaces, where the degree of order is primarily varied by modulating the length of the alkane chain comprising the SAM, and shorter chains are

shown to result in more disordered SAMs. Atomic force microscopy measurements and hydration free energy calculations suggest that surface hydrophobicity *decreases* with disorder in these SAMs.^{144–146} The authors suggest that water embeds itself in the pockets created by the disordered SAMs, locally pinning the SAM-water interface and reducing the surface’s hydrophobicity,¹⁴⁵ similar to how a single polar group in an otherwise nonpolar surface significantly reduces the hydrophobicity of the surface.¹²⁸

The distinct trends in hydrophobicity between disordered SAMs and other rough surfaces may stem, again, from differences in the lengthscale of roughness and whether water can preserve its hydrogen-bonding structure near the surface topology. Although the agreed upon crossover between water’s response to “small-lengthscale” hydrophobes, around which its hydrogen-bonding network is preserved, and “large-lengthscale” hydrophobes, which disrupt the hydrogen bonding network is 1 nm,⁵⁶ the precise geometry of the surface features may shift this crossover length. Van Lehn and coworkers, for instance, observed enhanced tetrahedrality, a signature of small-lengthscale behavior, near disordered SAMs compared to ordered SAMs.¹⁴⁴ Other computational studies have also observed enhanced tetrahedrality near silica surfaces that are made more “rough” *via* adsorption of benzene molecules.¹⁴⁷ In contrast, the geometry of the surfaces in studies that observe enhanced compressibility of water may be such that it disrupts the hydrogen-bonded network, leading to large-lengthscale behavior. Thus, roughening a surface may lead to either enhanced large-lengthscale structural response of water and enhanced surface hydrophobicity, or enhanced small-lengthscale structural response and diminished

surface hydrophobicity. Additionally, realistic surfaces likely exhibit a complex interplay between small and large lengthscale water physics, which is poorly understood compared to the picture now established for idealized solutes and surfaces.

To address this knowledge gap, it is important to understand the effect of surface roughness and particularly the lengthscale of rough features on local water behavior. A highly tunable platform for this study is the model nanopore system from the last chapter, where we examined patterns of nonpolar methyl and polar hydroxyl groups directly tethered to the pore wall and determined their effect on water and solute dynamics in the pore. In order to add roughness to the nanopore wall, we will examine patterns that include longer nonpolar ethyl and polar hydroxymethyl groups. These groups extend further from the nanopore wall, creating a rougher pore-water interface. We will first examine “rough” pores comprising only nonpolar groups (methyl and ethyl) or only polar groups (hydroxyl and hydroxymethyl) in order to determine how the topology influences water and solute transport independent of chemistry. However, chemical heterogeneity can have a non-intuitive effect on local water behavior and surface hydrophobicity, especially when paired with topological heterogeneity.¹⁴⁸ Thus, we next seek to determine whether chemistry and topology can work synergistically in modulating transport through the pore, using an inverse design approach.

The previous chapter demonstrated the utility of a genetic algorithm in suggesting non-intuitive pore patterns that exhibited extremal functional properties (in particular, hindered boric acid transport). Importantly, the addition of a surrogate model that

rapidly estimates the performance of a novel pattern based on the pattern alone significantly accelerated the genetic algorithm, allowing exploration of orders of magnitude more patterns than would be possible without the model. However, exploration of patterns with more than two headgroups exponentially expands the design space, such that more efficient approaches are needed. In this chapter, we use an active learning approach in order to efficiently search a lower-dimensional space of pattern descriptors. Such active learning approaches have previously been similarly used to efficiently explore large and complex design spaces.^{78,149–151} Here, we demonstrate the algorithm’s relative efficiency compared to the genetic algorithm by performing the optimization of methyl-hydroxyl patterns. We then use the algorithm to suggest non-intuitive patterns comprising methyl, hydroxyl, ethyl, and hydroxymethyl groups.

3.2 Methods

3.2.1 Nanopore simulations

As in Chapter 2, we model an ideal, functionalized nanopore as a carbon nanotube backbone to which we directly tether functional groups. The forcefield parameters used to model the backbone along with the methyl and hydroxyl functional groups are given in section 2.2 and Ref. [152]. In this chapter, in order to study the effect of pore wall roughness, we expand the set of functional groups to include ethyl and hydroxymethyl functional groups. We obtain parameters for these functional groups from the OPLS

forcefield¹⁵³ (details in section B.1). As in Chapter 2, we use a TIP3P water model¹²¹ and the boric acid model from Ref. [122].

Details of the simulation procedure for the nanopore are also given in section 2.2 and Ref. [152]. Briefly, in order to compute transport properties associated with a given pattern of functional groups, we simulate a “finite” nanopore attached to two water reservoirs and containing a single boric acid molecule to compute the equilibrium density of water within the nanopore. We then simulate an “infinite” nanopore (connected to itself by periodic boundary conditions) with the number of waters corresponding to the equilibrium density in the finite pore, as well as a single boric acid molecule, and compute transport properties such as a nonequilibrium flux under an applied pressure gradient. Here, the finite pore simulation is run for 20 ns and the infinite pore simulation is run for 200 ns (after initial equilibration periods, as described in Ref. [152]).

3.2.2 Active learning optimization

The genetic algorithm in Chapter 2 directly optimizes in the space of functional group patterns. As a result, the design space is high dimensional and the design parameters are discrete (functional group identity at each location). Both factors make the optimization difficult. Here, we instead optimize in a low-dimensional, continuous space of functional group “interaction” parameters. These parameters define an energy function that can be minimized to obtain patterns. For instance, for patterns containing methyl and hydroxyl groups, one such energy function could be:

$$U(N_r, N_a; J_r, J_a) = -J_r N_r - J_a N_a \quad (3.1)$$

Where N_r and N_a are the number of nearest-neighbor pairs of hydroxyl groups in the “radial” (along rings of the pattern) and “axial” (along the pore axis) directions, and J_r and J_a are the parameters defining the strength of these interactions. For instance, positive and large J_r values imply that patterns with rings of hydroxyl groups are “energetically” favorable. This energy function is in the spirit of an Ising model with distinct directional interactions and zero external potential.

In a real system, these interactions between functional groups could dictate the self-assembled structure of the surface and the resulting heterogeneous pattern displayed. This approach can therefore directly suggest synthetic strategies, as opposed to patterns that may be difficult to generate.

To optimize in this space, we utilize an active learning approach; a schematic for the example 2-parameter (J_r, J_a) case is presented in Fig. 3.1. We initially choose random values of (J_r, J_a) , drawn using Latin Hypercube Sampling¹⁵⁴ within chosen bounds for each parameter. Here, we bound each J parameter by $[-1, 1]$.

Using these values of (J_r, J_a) , we then generate patterns that minimize equation 3.1 using a simulated annealing approach. Specifically, we first initialize the pattern at the desired composition (i.e. number of hydroxyl and methyl groups) with a random arrangement of hydroxyl groups. Then, randomly chosen hydroxyl and methyl groups are swapped with each other. We accept swaps with probability $P = \exp[-\beta(U_{\text{new}} - U_{\text{old}})]$, where U_{old} and U_{new} are the values of the energy associated with the patterns before

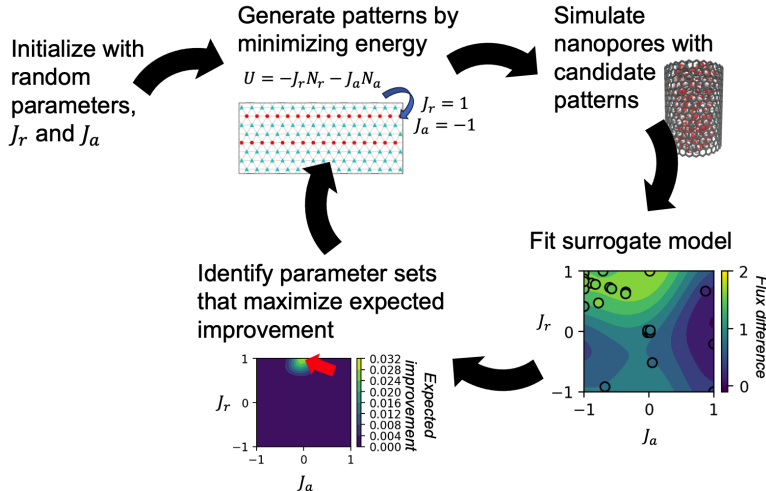


Figure 3.1: This schematic describes the active learning optimization procedure when the design space comprises the parameters J_r and J_a in equation 3.1. Each iteration through the cycle is considered a single “step.” The nanopore snapshot was created in VMD.¹⁵⁵

and after the swap, and $\beta = \beta_\infty[1 - C(-i/N)]$ where $\beta_\infty = 10$, $C = 0.95$, $N = 1000$, and i is the number of swaps attempted so far. This gives a continuously decreasing “temperature” schedule. For each set of parameters, we generate two patterns. The energy surface for these patterns has multiple local minima, so the simulated annealing procedure can produce a variety of patterns. Generating multiple patterns for each set of parameters provides information on the possible spread of objective functions.

We then create nanopores with the generated patterns and run simulations following the methods outlined in section 3.2.1 to compute the flux difference between water and boric acid. We fit a Gaussian Process Regressor (GPR) to the parameter-flux difference data. The GPR is a non-linear regressor that predicts both an expectation and an uncertainty, which can inform the choice of next parameters to simulate. Here, the GPR predicts the flux difference at unsampled $\mathbf{J} = (J_r, J_a)$, providing both a “mean” estimate,

$\mu(\mathbf{J})$, and a quantification of the uncertainty associated with the estimate, a standard deviation, $\sigma(\mathbf{J})$. For situations in which we can only run one simulation at a time, we would then identify the parameter set, $\mathbf{J}$, that maximizes the expected improvement:

$$EI(\mathbf{J}) = (\mu(\mathbf{J}) - C^*)F(z(\mathbf{J})) + \sigma(\mathbf{J})f(z(\mathbf{J})) \quad (3.2)$$

where $z(\mathbf{J}) = (\mu(\mathbf{J}) - C^*)/\sigma(\mathbf{J})$, C^* is the largest flux difference for a pattern found so far, F is the cumulative distribution function for the standard, normal distribution and f is the density function for the standard, normal distribution. The expected improvement is maximal at points with high estimated flux differences, $\mu(\mathbf{J})$, as well as high uncertainties, $\sigma(\mathbf{J})$. The algorithm thus chooses points that are expected to provide the largest increase in flux difference compared to the current optimal.

In our case, we can run eight simulations in parallel on a single GPU node, so we need to choose four candidate parameter sets (since we generate two patterns per parameter set). To do this, we choose the first parameter set by maximizing equation 3.2 and we choose the subsequent q th parameter sets by maximizing a modified “pseudo” expected improvement¹⁵⁶:

$$EI_{\text{pseudo}}(\mathbf{J}) = EI(\mathbf{J}) \prod_{i=1}^{q-1} [1 - \exp(-|\mathbf{J} - \mathbf{J}_i|^2)] \quad (3.3)$$

where $\mathbf{J}_i$ are the previously chosen parameter sets in this step. This penalizes parameter sets that are close to those already chosen, such that the proposed new sets are not repeats. For each $\mathbf{J}$, two new patterns are generated using the simulated annealing approach described earlier. We consider each iteration of the cycle shown in Fig. 3.1 to be a single “step.”

3.3 Results and Discussion

3.3.1 Pore wall topology affects water and solute transport

We first determine to what extent the addition of “roughness” to the nanopore wall can perturb the transport of water and solutes. Here, we again use boric acid as a model neutral, small-molecule solute of interest to water treatment efforts. To assess the effect of roughness independent of chemistry, we simulate “fully hydrophobic” pores with only methyl and ethyl functional groups, as well as “fully hydrophilic” pores with only hydroxyl and hydroxymethyl functional groups. We arrange the longer ethyl and hydroxymethyl groups in pattern types that were identified as extremal for the methyl-hydroxyl patterns in Chapter 2, namely: single vertical patches, fully dispersed, and rings.

Fig. 3.2 shows the flux of (a) water and (b) boric acid as a function of the fraction of either ethyl or hydroxymethyl groups. In the nonpolar pore, both the composition and arrangement of the pattern affect flux, as with the methyl-hydroxyl patterns in Chapter 2. Specifically, water flux is highest in the homogeneous pores (only methyl groups or only ethyl groups) and is approximately the same in both. There is a slight, approximately linear decline in boric acid flux from the fully methylated to fully ethylated pore, likely due to increased flexibility of the ethyl groups. The longer ethyl groups adopt conformations that produce ring-like roughness down the pore that slightly inhibit boric acid transport (Fig. B.2). When the ethyl groups are arranged in a single vertical

patch (orange triangles), their effect is minimal and both water and boric acid flux are similar to the flux in the homogeneous pores. However, when the ethyl groups are fully dispersed (blue circles), there is a minima in the flux of water at around 25% ethyl groups, suggesting that the creation of a “rough” nanopore wall *via* protruding ethyl groups does reduce water’s dynamics. This trend is roughly correlated with an enhancement in water’s tetrahedrality near the pore wall (Fig. B.5), which peaks at around 50% ethyl groups. Ref. [147] similarly found that roughening a silica surface by adsorbed benzene molecules reduced local water dynamics due to increased water tetrahedrality. In Chapter 5, we will also find a correlation between water’s dynamics and tetrahedral structure near molecular surfaces. For fully dispersed patterns, the dependence of boric acid flux on the fraction of ethyl groups is more complex and has multiple minima below and above 50% ethyl groups; we hypothesize that the pattern at 50% ethyl groups uniquely creates “channels” that enhance boric acid dynamics along the wall (Fig. B.3).

Rings of ethyl groups (green diamonds) also appear to uniquely inhibit boric acid transport over water transport. Water flux in pores with rings is comparable to that in the fully dispersed pore, suggesting that protruding ethyl groups have similar effects on water dynamics whether they are dispersed or arranged in rings. In contrast, boric acid flux in ringed pores is significantly lower than the flux in dispersed pores, especially for pores with two rings of ethyl groups (dashed green line). Boric acid may thus follow a similar “hopping” mechanism described in Chapter 2, where the functional group rings create an energetic barrier that boric acid has to “hop” over in order to transport down

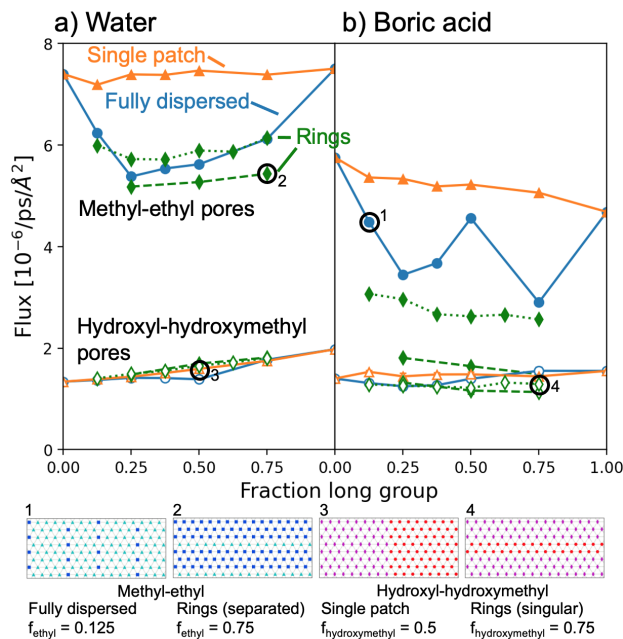


Figure 3.2: (a) Water and (b) boric acid flux in pores with either only nonpolar (methyl and ethyl, filled markers) or only polar (hydroxyl and hydroxymethyl, unfilled markers) functional groups, as a function of fraction of either ethyl or hydroxymethyl (“long”) groups for patterns where the long groups are arranged in a single vertical patch (orange triangles), a fully dispersed pattern (blue circles), or rings (green diamonds). The ringed patterns contain either a single ring (dotted green line) or two rings (dashed green line). Example patterns are shown for a few selected compositions and arrangements.

the pore. Here the mechanism of the energetic barrier created by the ethyl groups is likely distinct from that created by the hydroxyl groups (a hydrogen-bonded layer of water at the rings).

In contrast, in the polar pores, the arrangement of the hydroxymethyl groups has a much smaller effect on water and boric acid transport. There is a slight increase in water flux with the fraction of hydroxymethyl groups, which may be related to the increased flexibility of the hydroxymethyl group that leads to less structured hydration water (Fig. B.4). There is furthermore a slight reduction in water flux when the hydroxymethyl groups are fully dispersed compared to when they are arranged in a single, vertical patch,

for the pores with 50% hydroxymethyl groups. Thus, water’s dynamics may be slightly slowed over a “rough” hydrophilic surface with protruding polar groups. However, the magnitude of this reduction is much smaller than that achieved in the nonpolar pores. There is no significant effect of the hydroxymethyl group arrangement on boric acid flux. We hypothesize that these polar pores strongly hydrogen-bond with water and therefore boric acid preferentially partitions to the center region of the pore, such that the effect of wall topology on boric acid transport is negligible.

3.3.2 Active learning approach identifies diverse and optimal patterns

We first test the active learning procedure on the same optimization problem addressed in Chapter 2, namely identification of methyl-hydroxyl patterns with enhanced transport of water versus boric acid, quantified by the flux of water through the nanopore minus the flux of boric acid. We use the energy function in equation 3.1 to generate patterns, such that we optimize over the parameters (J_r, J_a) . The composition of the pattern is fixed at 25% hydroxyl groups and 75% methyl groups, the same case explored by the genetic algorithm whose trajectory is shown in Fig. 2.2a.

Fig. 3.3a shows the trajectory of this active learning optimization; each marker represents a single pattern explored and the flux difference computed from simulation. To test that the algorithm is able to reach the optimal even when starting from far away, we initialize the algorithm with the feature sets $(J_r, J_a) = (\pm 0.1, \pm 0.1)$, as opposed to

using Latin Hypercube Sampling in the bounded region. These feature sets produce relatively disordered patterns (such as the blue-bordered pattern in Fig. 3.3). Despite these initial patterns, the algorithm is able to rapidly identify within 6 steps the optimal that took almost 2000 generations for the genetic algorithm (Fig. 2.2) to find, namely a pattern with a single ring of hydroxyl groups (the red-bordered pattern in Fig. 3.3). The parameters that achieve this pattern are $(J_r, J_a) = (1.0, 0.014)$, i.e. strongly attractive radial interactions and weakly attractive axial interactions. Compared to the genetic algorithm, the active learning approach provides at least a $5\times$ speedup (6 stages of simulations versus 30 stages of simulations). Furthermore, the active learning approach suggests patterns that are more regular, and thus likely easier to synthesize.

The GPR fit using the entire set of data points is shown in Fig. 3.3b, where the top panel shows predicted flux differences for other combinations of parameters. Fig. 3.3b suggests that higher J_r and negative to small J_a values produce patterns with the highest flux difference. More attractive radial interactions, while axial interactions remain weak, result in patterns with more ring-like structures. In contrast, the GPR predicts the lowest flux difference for higher J_a values, or stronger axial interactions, which would produce patterns with vertical patches. This is consistent with conclusions from Chapter 2, which proposed that rings of hydroxyl groups form energetic barriers that hinder the transport of boric acid, lowering boric acid flux while having a smaller effect on water flux. In contrast, patterns with vertical patches of hydroxyl groups showed the highest fluxes of boric acid (Fig. 2.3). An advantage of the present approach, in addition to its significant

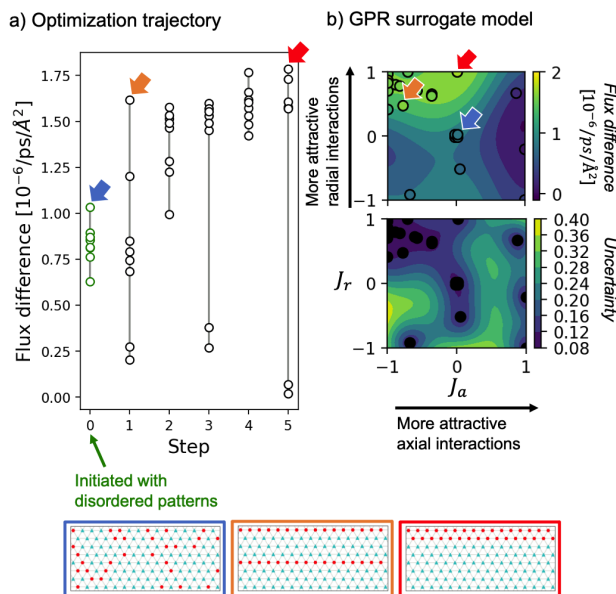


Figure 3.3: (a) The flux difference for patterns explored in each step (represented by individual markers). (b) The GPR surrogate model-predicted (top) flux difference, μ and (bottom) uncertainty, σ , used to compute the expected improvement in equation 3.2. Each marker represents a pattern that is explored, while the colored contours show the fitted GPR. For the top panel, the markers are colored by the simulated flux difference, using the same scale as the contours. The patterns shown at the bottom are optimal patterns discovered along the trajectory; arrows colored by the pattern borders show when each pattern was discovered in the optimization and the (J_a, J_r) parameters that produced the pattern.

speedup, is that the surrogate model directly gives predicted ranges of parameters that can be easily translated to pattern motifs that show the most and least optimal behavior, in contrast to having to infer these design rules from clustering of the discovered patterns or analysis of the coefficients of the linear surrogate model, as was done in Chapter 2.

At step 1 the algorithm discovers a pattern with high flux difference (see Fig. 3.3a), but continues to explore the design space and eventually finds a higher performing pattern. This intermediate optimal is the pattern with two separate rings of hydroxyl groups (orange-bordered pattern) and is created by strongly attractive radial interactions (large, positive J_r) and strongly repulsive axial interactions (large, negative J_a), i.e. parameter

sets on the top-left of the contour plots in Fig. 3.3b. The algorithm continues to explore this region, as the number of markers in that corner show, but also proposes parameters in other regions in the (J_r, J_a) parameter space, due to the higher uncertainty given by the GPR for regions where fewer patterns have been simulated (Fig. 3.3b, bottom). This behavior illustrates the trade-off between exploitation and exploration that is balanced by the expected improvement function that determines the proposal of new points.

Having confirmed that the active learning approach is able to optimize patterns efficiently, we now apply it to the optimization over patterns with three headgroups. With the increased number of functional groups, we necessarily use a different energy function to generate patterns:

$$U(\mathbf{N}; \mathbf{J}, \lambda) = \sum_{i \in (hh, ee, he)} -J_i [(1 - \lambda)N_{i,r} + \lambda N_{i,a}] \quad (3.4)$$

Where $\mathbf{N} = (N_{hh,r}, N_{ee,r}, N_{he,r}, N_{hh,a}, N_{ee,a}, N_{he,a})$ are the counts of various types of nearest-neighbor pairs: hh = hydroxyl-hydroxyl, ee = ethyl-ethyl, and he = hydroxyl-ethyl, and the subscripts r and a indicate radial and axial nearest neighbors. For instance, $N_{he,r}$ is the number of hydroxyl group neighbors in the radial direction that are ethyl groups. The parameters describing the strength of these interactions are: $\mathbf{J} = (J_{hh}, J_{ee}, J_{he}) \in [-1, 1]$ and $\lambda \in [0, 1]$. More positive J_{ij} values favor more i - j nearest-neighbor interactions. The final parameter, λ , determines the strength of axial versus radial interactions. When $\lambda = 0$, only radial nearest-neighbors contribute to the energy, and when $\lambda = 1$, only axial nearest-neighbors contribute to the energy.

The optimization now proceeds over four parameters but otherwise follows the same

basic procedure as the optimization for methyl-hydroxyl patterns. We first run the optimization fixing the composition of the pattern to 16 hydroxyl groups, 16 ethyl groups, and (the remainder) 96 methyl groups. For the first two steps, random parameter sets are generated using Latin Hypercube Sampling, while the remaining steps generate parameters by maximizing the pseudo expected improvement (equation 3.3). Fig. 3.4a shows the trajectory of the optimization. As with the 2-headgroup case, the optimization is able to rapidly identify parameter sets giving patterns with much larger flux difference than discovered during the two “random” steps. Fig. 3.4b shows some of the patterns that are explored by the algorithm. The optimal pattern, among those explored, contains separate rows of ethyl and hydroxyl groups, consistent with our previous finding that pores with rings of ethyl groups seem to uniquely slow boric acid (Fig. 3.2). Interestingly, the pattern with two separate rings of homogeneous composition outperforms the other ringed patterns. Similar to the methyl-hydroxyl case, the worst-performing patterns contain large, vertical patches, here of alternating rows of ethyl and hydroxyl groups. We hypothesize that boric acid is able to more rapidly transport down the large, accessible patch of methyl groups along the pore wall.

Finally, we run an optimization with methyl, hydroxyl, and ethyl groups while allowing the composition of the pattern to change. This optimization now occurs over six parameters, f_m , f_h , J_{hh} , J_{ee} , J_{he} , and λ . Two parameters, f_m and f_h , describe the composition of the pattern. Specifically, f_m describes the fraction of groups that are methyl groups, and f_h defines the fraction of non-methyl groups that are hydroxyls. These pa-

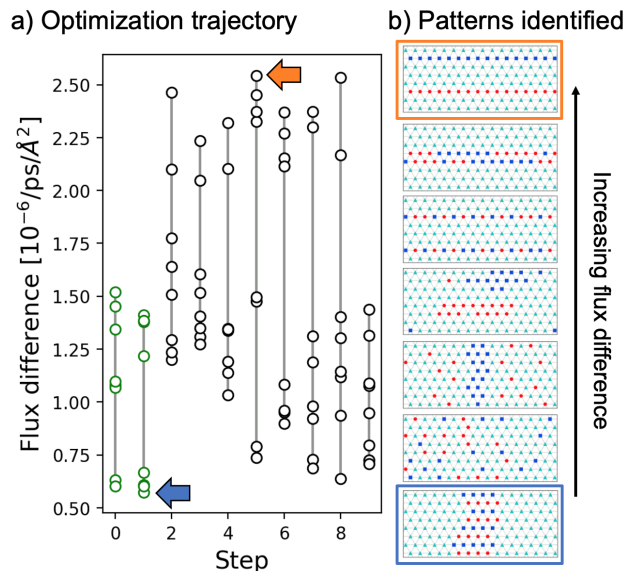


Figure 3.4: (a) The flux difference for patterns explored in each step (represented by individual markers) for the optimization over methyl-hydroxyl-ethyl group arrangements, fixing the composition. Parameters for the first two steps are randomly drawn (green markers). (b) A sample of the patterns identified, ordered by flux difference, and including the patterns with the overall lowest (blue border) and highest (orange border) flux difference, corresponding to the arrows of the same color on the trajectory.

parameters are bounded by $[0, 1]$. To generate patterns, we use the same energy function in equation 3.4 with the four parameters described earlier that define the strength of different types of interactions (J_{hh} , J_{ee} , J_{he}) and the strength of radial versus axial interactions (λ).

Fig. 3.5 shows the result of running this optimization for twenty steps, where the parameters for the first four steps are randomly drawn. The algorithm again rapidly finds an optimal in 7 steps with a flux difference of $4.4 \times 10^{-6}/\text{ps}/\text{\AA}^2$; the pattern contains two rows of methyl groups, spaced a single row apart, and no hydroxyl groups. Interestingly, this pattern is superior to that where the methyl rows are next to each other (another pattern suggested in step 6), which has a flux difference of $3.7 \times 10^{-6}/\text{ps}/\text{\AA}^2$. This

optimal pattern is also superior to the pattern in which the methyl rows are arranged evenly through the pore ($4.0 \times 10^{-6}/\text{ps}/\text{\AA}^2$). The placement of these barriers to transport modulates flux.

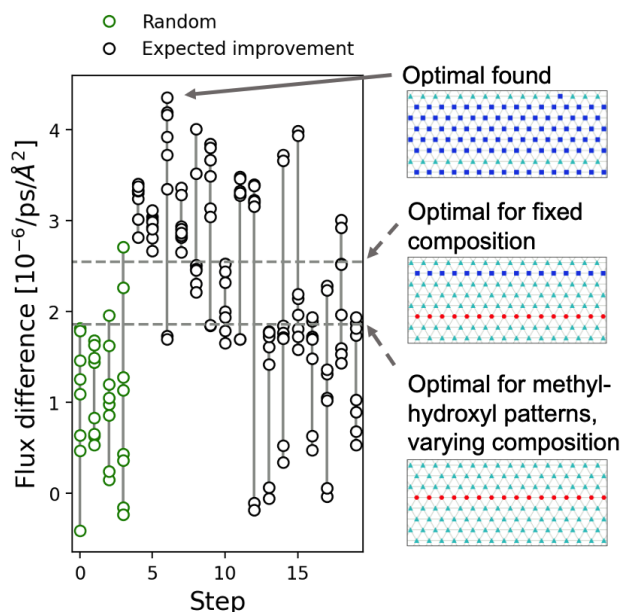


Figure 3.5: The flux difference for patterns explored in each step (represented by individual markers) for the optimization over methyl-hydroxyl-ethyl group arrangements, allowing the composition to vary. Parameters for the first four steps are randomly drawn (green markers).

The optimization results also suggest that inclusion of hydroxyl groups in the pattern is unfavorable. The algorithm discovered multiple patterns that are superior to the optimal discovered with fixed composition (Fig. 3.4). Moreover, these patterns also show enhanced solute rejection compared to the optimal discovered over all methyl-hydroxyl patterns, even considering varied composition (Fig. 2.2b). The best-performing of the patterns discovered that contain hydroxyl groups has a flux difference of $3.4 \times 10^{-6}/\text{ps}/\text{\AA}^2$, almost 25% lower than the overall optimal. Incorporation of hydroxyl groups significantly reduces water flux, compared to incorporation of ethyl groups in the same patterns

(compare Fig. 2.3a to Fig. 3.2a), without a sufficiently large reduction in boric acid flux to compensate. Thus, introduction of a rough free energy landscape for boric acid through modulation of pore topology seems to more effectively enhance selectivity compared to modulation of pore polarity.

We propose that the rings of methyl and ethyl groups create troughs from which boric acid has to desorb in order to transport down the pore wall. Similar to rings of hydroxyl groups, these methyl-ethyl rings thus create a free energy barrier, reducing boric acid’s dynamics in the pore. This energetic barrier is comparable in magnitude to the barrier created by the hydroxyl groups, since boric acid’s preference for the fully hydroxylated pore is similar to its preference for the bulk (the partition coefficient is 0.74, Table A.5). Thus, the reduction in boric acid dynamics in ringed methyl-ethyl pores is comparable to similar methyl-hydroxyl pores. However, the nonpolar groups have a much smaller effect on water’s dynamics compared to the hydroxyl groups, which can hydrogen bond to water and slow its dynamics. As a result, the optimization over all three (methyl, ethyl, and hydroxyl) groups finds that patterns with methyl and ethyl rings give higher flux differences compared to patterns with hydroxyl groups.

3.4 Conclusions

In this chapter, we show that surface roughness, in addition to chemistry, can modulate local water and solute dynamics, and that an active learning-based inverse design algorithm can suggest chemical and topological patterns that tune these dynamics to op-

timize solute rejection. In nonpolar pores, increasing surface roughness by adding ethyl groups decreases water and boric acid flux. We propose that the mechanisms for this reduction are unique. Rough pores reduce water flux by creating Å-scale protrusions in the surface around which water can structure, enhancing water’s tetrahedrality and reducing its dynamics. In contrast, rough pores reduce water flux by creating cavities for which boric acid has enhanced affinity, and must desorb from in order to transport down the pore. In contrast, surface roughness in polar pores has a comparatively smaller effect on transport of both water and boric acid. Water can hydrogen bond to both hydroxyl and hydroxymethyl groups, and so would not form tetrahedral structures near them. Boric acid is excluded from the pore walls by the hydrogen-bonded water, and thus the effect of the wall’s roughness is minimal.

To identify patterns of methyl, hydroxyl, and ethyl groups that optimize boric acid rejection, we develop an inverse design algorithm that optimizes parameters governing interactions between the functional groups, using an active learning approach to suggest new parameter sets. We show that this algorithm more rapidly identifies the optimal pattern of methyl and hydroxyl groups, compared to the genetic algorithm. With the inclusion of ethyl groups, the algorithm again suggests that patterns with the functional groups arranged in rings along the pore wall are the most optimal for inhibiting boric acid transport. Furthermore, nonpolar patterns are favored over patterns with hydroxyl groups, since rings of ethyl groups provide comparable reduction to boric acid dynamics without an equivalent reduction in water’s dynamics as they do not hydrogen bond with

water. As the optimization progresses, the design algorithm also generates numerous patterns with diverse arrangements, and thus serves to provide broad insight regarding this multifunctional design space, while targeting an optimum. The design approach presented in this chapter offers a more straightforward connection to synthetic strategies, where choosing chemistries with specific cross- and co- interaction strengths controlling their assembly into a surface is likely easier than precise placement of specific functionalities at a surface. This approach could be adapted to optimize over other conditions governing the synthesis of functionalized surfaces.

Acknowledgments

This work was supported by the Center for Materials for Water and Energy Systems (M-WET), an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Basic Energy Sciences under Award DE-SC0019272. Use was made of computational facilities purchased with funds from the National Science Foundation (CNS-1725797) and administered by the Center for Scientific Computing (CSC). The CSC is supported by the California NanoSystems Institute and the Materials Research Science and Engineering Center (MRSEC; NSF DMR 2308708) at UC Santa Barbara. S.J. acknowledges support by the National Science Foundation Graduate Research Fellowship (DGE 1650114).

Chapter 4

Quantifying polypeptoid conformational landscapes through integrated experiment and simulation

Adapted with permission from *Macromolecules* 2021, 54, 11, 5011–5021. Copyright 2023 American Chemical Society.

4.1 Introduction

In the previous chapters, we have explored the inverse design of highly idealized nanopores. Though carbon nanotube functionalization is an active area of interest, there are cur-

rently no routes to the direct and precise functionalization of nanotube sites. An alternative method for tuning interfacial chemical patterning is to decorate the surface with sequence-specific polymers. The monomer sequence along the chain can be easily and exactly set during polymer synthesis. One class of sequence-specific molecules that has attracted significant attention for surface modification comprise polypeptoids, or N-substituted glycines, which possess both the robustness of synthetic polymers and the tunability of polypeptides and other biomolecules.⁷³ Polypeptoids can now be routinely synthesized at gram scale, sequence-specifically up to 50 monomers in length in a process that eliminates the need for protection/deprotection procedures common in polypeptide synthesis, allowing access to a wide variety of chemistries.⁷³ As such, they have been extensively explored for a broad range of applications, including as antifoulants,⁷⁴ antimicrobials,^{157–159} other therapeutics,¹⁶⁰ delivery agents,^{161,162} sensing applications¹⁶³ such as stimuli-responsive materials,¹⁶⁴ binder materials for batteries,¹⁶⁵ crystallization modulators,¹⁶⁶ and programmable self-assembly.^{167–169} Many of these applications leverage the sequence-specificity of polypeptoids that has been shown to significantly modulate secondary^{170–173} and self-assembled structures,^{174,175} and therefore function. Thus, advancement of polypeptoid-containing materials and the fundamental design rules that they follow require elucidation of sequence-structure-function relationships.

Many polypeptoids used in these applications are structurally disordered and adopt a wide range of conformations. Sequence has been shown to tune functional properties even in these structurally-disordered cases. For example, variations in the location and number

of hydrophobic or charged monomers can modify antifouling and fouling release,^{176,177} loading and release in drug delivery,¹⁷⁸ gene transfection efficiency,¹⁷⁹ and the degree of melting point depression,¹⁸⁰ among other properties. Further evidence suggests that shifts in the conformational distributions of these disordered polypeptoids, modulated by the polypeptoid sequence, contribute to changes in function.^{180–182} However, the difficulty of probing the structure of disordered polypeptoids both experimentally and computationally has hindered the development of fundamental design rules to guide the discovery of novel and efficacious polypeptoids. Experiments typically report on dominant or average structures, obscuring variations within the ensemble of conformations. On the other hand, modeling provides direct access to the conformational landscape through insight into atomic-scale structures and structural correlations. However, polypeptoid simulations encounter major sampling challenges due to the long timescales associated with conformational transitions, which has also hampered the validation of polypeptoid simulation models.

In this chapter, we pursue a novel experimental-computational cross-validation approach involving measuring and comparing structural ensembles, as opposed to dominant or average structures. Specifically, the probability distribution of the distance between the ends (end-to-end distance) of a hydrophilic polypeptoid of varying length (Fig. 4.1a, Table 4.1) is measured using Double Electron-Electron Resonance (DEER) spectroscopy and Continuous Wave Electron Paramagnetic Resonance (cw-EPR). Recently developed techniques are leveraged to accurately extract these broad distance distributions. We

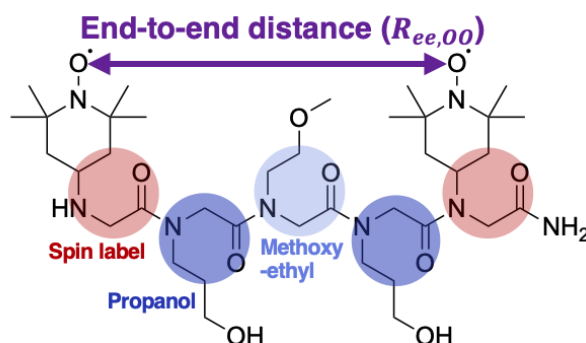
measure the same distributions computationally using an expanded-ensemble simulation technique that directly speeds up the sampling of the slow backbone bond rotations to resolve the full equilibrium conformational ensemble. The hydrophilic polypeptoid we choose to study (Fig. 4.1a) comprises an alternating sequence of methoxyethyl and propanol sidechains. The former have been previously studied as promising antifoulants⁷⁴ and as model polar sidechains in self-assembling systems¹⁷⁵ and the latter enhance the polypeptoid’s solubility in water. For the model polypeptoid studied here, the combined simulation and DEER workflows show excellent agreement. The distance distribution extracted using cw-EPR for a shorter polypeptoid also agrees with simulation, offering an important supplemental experimental probe of short polypeptoids.

N	Structure	Sim	Exp
1		Yes	None
2		Yes	None
3		Yes	cw-EPR
5		Yes	None
7		Yes	None
9		Yes	DEER
11		Yes	DEER
13		Yes	None
15		Yes	DEER
19		Yes	DEER

Table 4.1: Polypeptoids of varying length are probed both through DEER and cw-EPR experiments and through MD simulation. Note that N is the number of hydrophilic monomers, not including the terminal spin labels.

DEER is an electron paramagnetic resonance technique that provides unique access to conformational ensemble information by measuring the full distance distribution from a composite time domain signal that reports on distances between pairs of spin-labels.¹⁸³

a) Hydrophilic polypeptoid sequence



b) Disarcosine

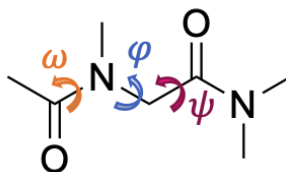


Figure 4.1: Hydrophilic polypeptoids used here to cross-validate the experimental and computational workflows are synthesized, characterized, and simulated. (a) The polypeptoids comprise an alternating sequence of hydrophilic monomers terminated on both ends by spin labels for Electron Paramagnetic Resonance, which measures the distance between the oxygen radicals, $R_{ee,OO}$, shortened to R_{ee} in the text. (b) Disarcosine is simulated to validate the expanded ensemble sampling. The arrows show rotation around the ω , ϕ , and ψ dihedrals in the polypeptoid backbone.

It is especially advantageous for comparison with simulations because its probes are relatively small,¹⁸⁴ minimally perturb the conformational ensemble,¹⁸⁴ and can be directly simulated.¹⁸⁵ DEER has traditionally been used to probe distances in structured biomacromolecules¹⁸³ or average distances in unstructured systems. In recent years, its ability to detail the full probability distribution over accessible distance ranges has garnered significant interest. The application of DEER to characterize disordered biological and synthetic polymeric systems through this distance distribution is emerging,^{186–192} and methodological developments are continually improving the accurate resolution of the

shape as well as the width of broad distributions.^{193–197} Recent work on aqueous solutions of polyethylene oxide, a canonical disordered, synthetic polymer, demonstrates consistency between the broad distance distributions determined by DEER and simulation,¹⁹² suggesting that DEER is well-positioned to characterize the conformational landscape of disordered macromolecules. Single-molecule Förster resonance energy transfer (FRET) also probes conformational distributions,¹⁹⁸ but DEER requires fewer assumptions^{199,200} and captures far more molecules, resulting in more accurate distance distributions.

The accessible distance range for DEER is generally considered to be between 20 Å and 80 Å.^{183,201} Supplementing DEER at shorter distances, we use the analysis of cw-EPR lineshapes where the dipolar interaction between two spin labels with distances between 8–25 Å leads to broadening of the spectrum that can be related to the distances between them by spectral simulation. Unlike DEER, cw-EPR cannot fully resolve multi-featured distributions but is a complementary probe of the existence and approximate width and character of the distribution in the short-distance region.

Simulations of polypeptoids present unique challenges because the polypeptoid amide bond isomerizes on long timescales (seconds²⁰²) relative to achievable simulation lengths due to a large (≈ 17 kcal/mol²⁰³) barrier to isomerization.²⁰⁴ Polypeptoid forcefields have a correspondingly large barrier in the ω amide bond dihedral potential, which hinders simulation studies of appreciable chain lengths since the number of isomerization states expands combinatorically with degree of polymerization. To accurately probe conformational space, the sampling technique must accelerate cis-trans isomerizations for each

amide bond. Here, we present and validate a technique to sample amide-bond dihedral isomerization in long polypeptoids by simulating in an expanded ensemble of states in which the amide bond dihedral potentials are scaled down, after which properties are rigorously reweighted back to the original state. This approach, which has previously been used to enhance conformational sampling in small molecules²⁰⁵ but has not yet been applied to polypeptoids, addresses many of the challenges of select earlier simulation strategies that do not extend well to longer chains in explicit solvent.^{206–208} More recently, the metadynamics technique has provided an efficient and targeted solution to specifically enhance sampling of the slow degrees of freedom.^{209–211} However, successful tuning of key algorithm convergence parameters can be difficult and a significant amount of simulation time is spent exploring regions of phase space inconsequential to the equilibrium ensemble. The approach presented here similarly targets specific degrees of freedom—the amide-bond isomerizations—and demonstrates excellent convergence properties while remaining extensible to longer chains.

4.2 Methods

Methods associated with polypeptoid synthesis and experimental characterization are described in Ref. [212].

4.2.1 Polypeptoid forcefield parameterization and simulation

In this study, we primarily utilize the MFTOID²⁰⁶ forcefield to simulate polypeptoids, but also compare results with other peptoid models: GAFF- ϕ ,²⁰⁷ a modified AMBER99SB-ILDN forcefield,²⁰⁹ and a modified CGenFF forcefield.²¹⁰ Section C.1.1 gives details of the forcefield parameterization. With any given forcefield, we create and geometry-minimize initial polypeptoid structures in the fully extended conformation using Avogadro 1.2.0.²¹³ We place each polypeptoid in a box of length at least 20 Å longer than its end-to-end distance when fully extended (Table C.19) and solvate the boxes using GROMACS 2016.1.^{214,215} We perform simulations using the OpenMM simulation engine.⁷¹ Electrostatic interactions are computed using particle-mesh Ewald. Lennard-Jones nonbonded interactions are cut off at 10 Å. Hydrogen bonds are constrained with SHAKE¹²³ and water is kept rigid with SETTLE.¹²⁴ We simulate Langevin dynamics with an integration timestep of 2 fs, a temperature of 300 K, and a friction coefficient of 0.01 ps⁻¹. The pressure is set to 1 bar using a Monte Carlo (MC) barostat with 500 fs between MC barostat moves.

4.2.2 Expanded-ensemble sampling

To sample across slow polypeptoid isomerization timescales, we develop a new sampling approach based on an expanded ensemble,^{216,217} allowing the system to visit states in which the dihedral potentials for the ω dihedral (Fig. 4.1b) are “softened,” i.e. the force constant is scaled by a factor, λ_{ee} , that varies between 0 and 1. When simulating with

the modified CGenFF forcefield, we multiply all backbone dihedral potentials (ϕ and ψ in Fig. 4.1b, in addition to ω) by λ_{ee} since isomerization of the ϕ dihedrals is also slow (section C.1.2). To simulate in the expanded ensemble, we perform Monte Carlo (MC) moves to change the λ_{ee} -state every 20 ps. For each MC move, we use Metropolized independence sampling¹²⁶ to compute the proposal and acceptance probabilities for each λ_{ee} -state. These probabilities are biased by weights for each state that are modified in the equilibration period to achieve a flat sampling of states, and then fixed in the production period. Based on the probabilities, a new λ_{ee} -state is proposed and accepted or rejected. For each system, there are a discrete set of λ_{ee} -states, distributed so that the self-transition probabilities are nearly flat. We reweight the properties computed during the expanded ensemble production period back to the original, unsoftened state using MBAR.^{125,218}

4.2.3 Comparison of polypeptoid forcefields

Simulation forcefield models for polypeptoids have seen far less development and refinement than their peptide counterparts. Several polypeptoid forcefields have been developed starting from general or protein forcefields and reparameterizing to fit either experimental or quantum mechanical measurements of short polypeptoids. Previous validation of forcefields have shown that they accurately reproduce the configurations of very short²¹⁹ or highly structured polypeptoids.^{207,220} Many studies, for instance, compare simulations with crystal structures,^{206,208,219,221} so that only sequences that show a

single dominant conformation can be evaluated. However, to date, there has been little work testing whether these polypeptoid forcefields accurately reproduce the broader and multi-featured conformational distributions of longer and disordered systems.

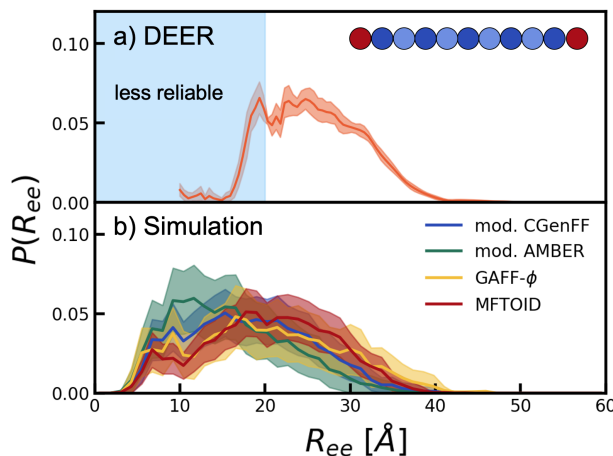


Figure 4.2: Experiments validate models for disordered polypeptoids. For the $N = 11$ polypeptoid, (a) the end-to-end distribution measured using DEER and fit with SF-SVD. The shaded region around the distribution gives 95% confidence intervals. The shaded blue box denotes the region below the accessible region for DEER. (b) The end-to-end distributions computed from simulations using various polypeptoid forcefields. Shaded regions show a 68% confidence interval computed using bootstrapping (section C.1.4).

Fig. 4.2 shows the end-to-end distance distribution of the hydrophilic polypeptoid of length $N = 11$ (a) measured by DEER and (b) computed from simulations using various forcefields. In the accessible region, the MFTOID and GAFF- ϕ forcefields show better agreement with experiment than the modified AMBER and modified CGenFF forcefields. Moving beyond visual assessment, to quantify the agreement between the experimentally (P_{DEER}) and computationally (P_{Sim}) derived end-to-end distance distributions, we compute the Jensen-Shannon distance (JS).²²² Smaller distances reflect greater overlap between two probability distributions, and thus better quantitative agreement between the experiment and simulation:

$$JS(P_{\text{DEER}}, P_{\text{Sim}}) = \sqrt{\frac{1}{2} \left[S_{\text{rel}}(P_{\text{DEER}}, \frac{P_{\text{Sim}} + P_{\text{DEER}}}{2}) + S_{\text{rel}}(P_{\text{Sim}}, \frac{P_{\text{Sim}} + P_{\text{DEER}}}{2}) \right]} \quad (4.1)$$

Here, S_{rel} is the relative entropy:

$$S_{\text{rel}}(P_1, P_2) = \int P_1(R_{ee}) \ln \frac{P_1(R_{ee})}{P_2(R_{ee})} dR_{ee} \quad (4.2)$$

We limit computations to the DEER-accessible region ($> 20\text{\AA}$) for both distributions. Distances between the experimental distribution and simulated distributions using the MFTOID, GAFF- ϕ , modified AMBER, and modified CGenFF forcefields are 0.16 ± 0.06 , 0.19 ± 0.07 , 0.30 ± 0.09 , and 0.22 ± 0.07 , respectively (ranges give a 95% confidence interval). An alternative metric gives similar trends (Table C.20). The modified AMBER and modified CGenFF forcefields therefore show the weakest agreement with the experimental distribution, while MFTOID and GAFF- ϕ seem comparable in accuracy. The latter, however, enforces an asymmetry in the ϕ dihedral angle distribution that is not physical for the achiral polypeptoids studied here, and thus the MFTOID forcefield is the model we adopt for the remainder of this work.

We attribute these differences to the polypeptoid forcefield parameters themselves and not the different water models used as Fig. C.12b shows that the $P(R_{ee})$ for $N = 11$ with the MFTOID forcefield is the same between the mTIP3P and SPC/E water models. Below the DEER-accessible region ($< 20\text{\AA}$), the forcefields also show differences in the height of the low-distance peak, suggesting that the strength of spin-label aggregation varies among the forcefields. These differences could be due to the polypeptoid forcefield

or the spin-label forcefield, but the latter is less likely because forcefields implemented with the same spin-label parameters (MFTOID and modified CGenFF) still give different $P(R_{ee})$'s in this region. Furthermore, the backbone end-to-end distance distributions for the labeled and unlabeled $N = 11$ polypeptoid, using the MFTOID forcefield, are consistent (Fig. C.18), suggesting that the presence of the spin-label does not significantly perturb the polypeptoid backbone conformational distribution. Thus, small deviations in the spin-label parameters, at least for the MFTOID forcefield, should not have a significant effect on the overall computed conformational landscape of the polypeptoid, especially for longer lengths.

4.3 Results and Discussion

4.3.1 Expanded-ensemble simulation accurately and efficiently samples isomerization states

Fig. 4.3 shows that the expanded ensemble simulation method reproduces the free energy difference between the cis and trans states computed from umbrella sampling for disarcosine (Fig. 4.1b). The 12 kcal/mol isomerization barrier also agrees with the value reported by Zhao *et al.*, who used parallel-bias metadynamics to sample across the isomerization barrier.²¹¹ The uncertainty in the free energy landscape computed using the expanded ensemble is larger at the barriers because the expanded ensemble only circumvents the high free energy barrier for isomerization through modulation of the dihedral

potential, so that the sampling of the high free energy regions may still be poor. On the other hand, approaches such as umbrella sampling and metadynamics directly bias the dihedrals, encouraging flat sampling of the space of dihedral angles. However, the equilibrium conformational landscape will be dominated by configurations within the cis or trans basins, not at the high free energy barriers. The expanded ensemble technique, then, allows crossing of high free energy barriers without spending unnecessary time thoroughly sampling these regions.

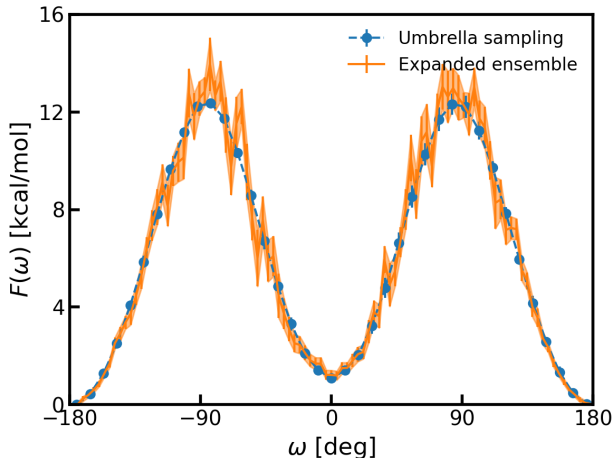


Figure 4.3: Validation of the expanded-ensemble methodology. The free energy landscape for rotation around the isomerizable amide bond of disarcosine computed using umbrella sampling and using the expanded ensemble method show excellent agreement, demonstrating that the expanded ensemble method is able to reproduce the correct cis/trans populations. These simulations utilize the MFTOID forcefield, as an example, but we would expect agreement for any forcefield. Both plots give a 95% confidence interval computed with MBAR^{125,218} (section C.1.4).

As previously reported for small molecules,²⁰⁵ the expanded ensemble significantly enhances the efficiency of the simulation. Scaling the dihedral potential by a factor λ_{ee} provides orders of magnitude increase in the isomerization rate, comparing the rate of transitions between cis and trans states when the ω dihedral potential is completely

removed ($\lambda_{ee} = 0$) to when it is unsoftened ($\lambda_{ee} = 1$) (Fig. C.7a); extremely high temperatures are required to achieve the same isomerization rates (Fig. C.7b), rendering temperature replica exchange impractical. The ability to access states at or close to $\lambda_{ee} = 0$ then makes the correlation time for the ω dihedrals tractable (Fig. C.8), even accounting for the additional computational cost of simulating the system in λ_{ee} -states that are far from the unbiased state (Fig. C.10). As a result, equilibration and production runs on the order of microseconds are sufficient to sample the end-to-end distance distribution of these polypeptoids (section C.1.2). This means that achieving an unbiased sampling of conformational space, assuming that all other degrees of freedom are fully sampled over, is now computationally tractable. The following results from MD simulation all rely on this sampling approach.

4.3.2 End-to-end distance distributions from experiments and simulations

The computational workflow presented here enables detailed characterization of the conformational distribution of polypeptoids with increasing chain length. The polypeptoids with alternating hydrophilic methoxyethyl (NME) and propanol groups (Fig. 4.1a) are synthesized, varying the number of hydrophilic monomers: $N = 9, 11, 15, 19$. Spin labels are attached to both ends of all polypeptoids for DEER measurements. End-to-end distance distributions, $P(R_{ee})$, are obtained from 4-pulse DEER experiments through singular value decomposition,^{196,197} as described in detail in Ref. [212]. Unless otherwise

noted, we use “end-to-end distance” to refer to the distance between the oxygen atoms on the spin labels. We simulate the same polypeptoids (including spin labels) using the expanded-ensemble sampling approach and the MFTOID forcefield (see section 4.2.3 for justification).

Fig. 4.4 shows the end-to-end distance distributions for lengths $N = 9, 11, 15, 19$. In the accessible region, simulations and DEER show excellent agreement, which offers strong cross-validation of both workflows. The locations of the peaks and the shapes of the long-distance tails of the distributions are especially well reproduced. Both methods consistently show not only the expected shift in the mean end-to-end distance, but also a broadening of the distribution with length, characteristic of disordered polymers. Below the DEER-accessible region ($< 20\text{\AA}$), the simulated distributions show a smaller peak at around $10\text{\AA}$ that decreases in height with increasing polypeptoid length. This peak suggests the spin-labels on both ends have a slight tendency to aggregate with each other (Fig. C.16, C.19), as was previously observed in simulations of end-labeled polyethylene oxide.¹⁹² Recalculating the $P(R_{ee})$ by removing the contribution from these short-distance conformations, which are not captured by DEER, further improves the agreement between DEER and simulation (Fig. C.20, C.21).

In the DEER-accessible region ($> 20\text{\AA}$), Fig. 4.4 demonstrates a few small inconsistencies between DEER and simulation. To determine whether these inconsistencies are reproduced in the time-domain or stem from instabilities in the fitted distance distribution due to the ill-posed nature of the inversion, we back-calculate the DEER time-

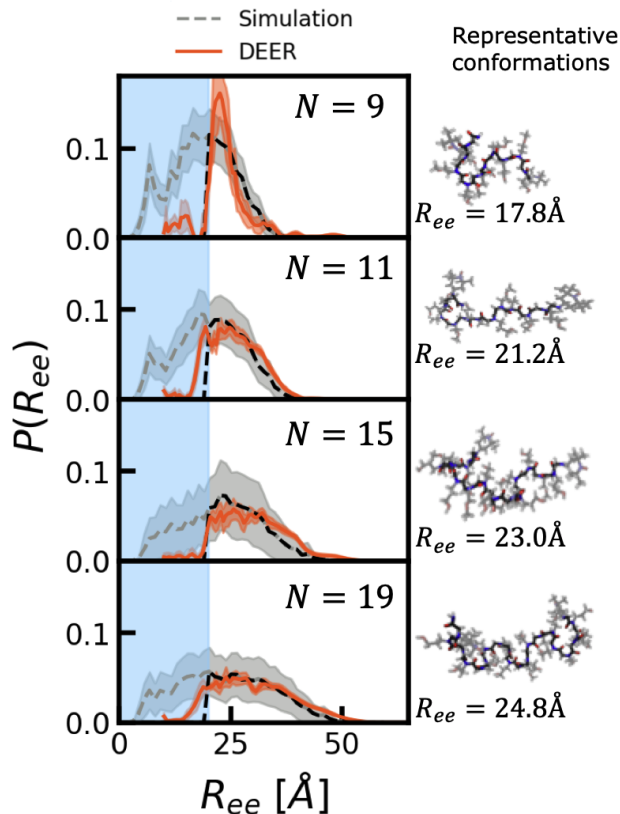


Figure 4.4: End-to-end distance distributions measured with DEER and computed from simulation are in excellent agreement. The gray lines are the simulated distributions (using the MFTOID forcefield) and the gray shaded regions give a 68% confidence interval computed from bootstrapping (section C.1.4). The orange lines and shaded regions are the fitted DEER distributions and uncertainty. All distributions are normalized such that the integral above 20 Å is 1. The light blue shaded box denotes the region below the accessible region for DEER. The snapshots to the right of the figures are representative conformations chosen to be close to the mean end-to-end distances and are visualized in VMD.¹⁵⁵

domain signal, $F(t)$, from the MD-simulated and SF-SVD-fitted distance distributions, performing the forward convolution that the fitting procedures attempt to invert (section C.2).

Fig. 4.5 shows the time-domain signals computed through convolution (section C.2) using the distributions in Fig. 4.4 (“back-calculated MD” and “back-calculated SF-SVD” signals), along with the background-corrected experimental time-domain data

from DEER (“raw” signal). The shifts in the back-calculated MD signal with increasing polypeptoid length are qualitatively consistent with that of the raw signal. In particular, the slower decay and broadening of the minimum at larger lengths are well reproduced. For all lengths, the back-calculated MD signal shows a steeper decay near $t = 0$. This feature reflects the presence of shorter end-to-end distances in the MD-simulated $P(R_{ee})$ that are not present in the DEER-derived $P(R_{ee})$. As discussed earlier, these short distances are inaccessible to DEER. Recomputing the back-calculated MD signal using only the portion of the MD-simulated $P(R_{ee})$ above the lower accessible limit of DEER ($20\text{\AA}$) leads to better agreement in the short-time decay (Fig. C.15). For $N = 15$ and $N = 19$, the raw signal displays a low-frequency oscillation (Fig. 4.5 inset) that arises from longer distances in the DEER-derived $P(R_{ee})$ and that is not reproduced in the back-calculated MD signal shown in Fig. 4.5. Thus, the differences between the simulation and DEER end-to-end distance distributions at longer lengths in Fig. 4.4 are real and not artifacts of the inversion. This discrepancy at longer lengths could be due to inaccuracies in the forcefield that lead to oversampling of shorter distances. For intrinsically disordered polypeptides, it was previously reported that peptide forcefields bias conformational distributions towards more compact structures than is realistic.^{223,224}

Cw-EPR is then used to probe the distance distribution of the $N = 3$ polypeptoid, whose end-to-end distance distribution largely lies below $20\text{\AA}$, and thus below the accessible region of DEER. Fig. 4.6 shows that the Gaussian approximation of the distance distribution from the cw-EPR experiment agrees with the longer-distance peak in the

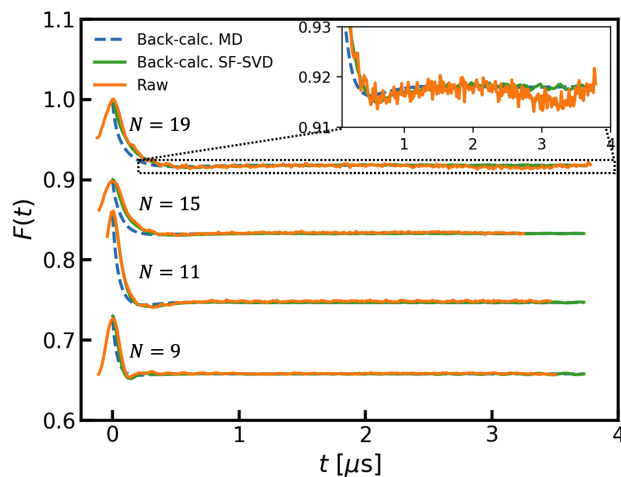


Figure 4.5: The raw DEER signals are consistent with signals back-calculated from simulated and fit end-to-end distance distributions. Orange lines show the raw DEER signals, while the blue and green lines show the signals back-calculated from the MD-simulated $P(R_{ee})$ (simulated using the MFTOID forcefield) and the SF-SVD-fit $P(R_{ee})$, respectively. For all signals, $V(0) = 1$, but the plots are shifted vertically (Table C.21) for each polypeptoid length for clarity. (Inset) Low-frequency oscillations in the raw signal for the $N = 19$ polypeptoid.

simulated distribution. The simulated distribution also contains a shorter-distance peak at around $8 \text{ \AA}$ that is also seen in Fig. 4.4 for longer polypeptoids, and here similarly indicates spin-label aggregation. The mean end-to-end distance from cw-EPR lineshape analysis is $16.5 \pm 0.1 \text{ \AA}$. Cw-EPR cannot resolve distances below $\approx 8 \text{ \AA}$, so a fraction of the low-distance population predicted from the simulations would not contribute to the cw-EPR-resolved distribution. However, cw-EPR clearly extends our insight into the experimental conformational distributions at distances below the DEER-accessible regime.

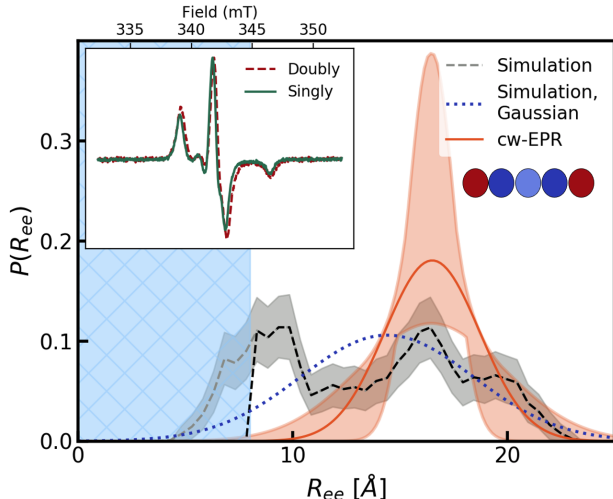


Figure 4.6: Continuous wave electron paramagnetic resonance allows access to the low-distance region and agrees with simulations. For the $N = 3$ polypeptoid, the Gaussian approximation of the end-to-end distance distribution from cw-EPR (orange) agrees with the simulated distribution (using the MFTOID forcefield) in the accessible region for cw-EPR (black dashed line gives the full distribution; the blue line is a Gaussian with the same mean and standard deviation as the simulated distribution in the cw-EPR accessible region). Shaded gray regions show a 68% confidence interval computed using bootstrapping (section C.1.4). The shaded orange region encompasses all Gaussian distributions with mean and standard deviation within the 95% confidence intervals obtained from the covariance matrix of $\langle R_{ee} \rangle$ and σ . All distributions are normalized such that the integral above $8 \text{ \AA}$ is 1. The shaded and hatched blue box denotes the region below the accessible region for cw-EPR. (Inset) The raw cw-EPR spectra for the doubly labeled 3mer (dashed red) and singly labeled 3mer (solid green) normalized by maximum peak height.

4.4 Conclusions

In this chapter, we validate a workflow to simulate long, disordered polypeptoid molecules. Specifically, we develop a state-of-the-art enhanced sampling expanded ensemble approach to overcome long isomerization timescales in polypeptoid simulations. We then demonstrate excellent agreement between distance distributions resolved through simulations and dipolar EPR spectroscopy for varying lengths of a model hydrophilic polypeptoid. The large widths of the distributions indicate that these hydrophilic polypeptoids

are indeed disordered and would thus be difficult to structurally characterize solely by examination of dominant conformations, instead requiring the resolution of full distributions. Validation of the expanded ensemble sampling technique and the MFTOID forcefield for polypeptoids allows us to computationally probe the conformational distribution of polypeptoids of appreciable length (here, tested up to $N = 19$, while we estimate up to $N \approx 30$ is reasonable for a single GPU).

These results establish highly transferable experimental and computational workflows to characterize the conformational landscapes of disordered polypeptoids. Though polypeptoid sequences have already been shown to modulate structure and function, the workflows developed here will enable more controlled investigations of sequence-structure-property relationships in the highly tunable, sequence-programmable polypeptoid platform. We will investigate such relationships in the next three chapters. For instance, the determination of changes in the conformational ensemble allows us to isolate the effects of backbone conformation and sidechain chemistries on functional properties (Chapter 5). This is especially important because changes such as movement of the polypeptoid from the solution to a surface (Chapter 6) or significant variations in sequence, such as incorporating an irregular sequence of hydrophobic sidechains (Chapter 7), will lead to significant perturbations in structure that can complicate the sequence-property relationship and cannot be fully captured by existing polymer physics models.

Acknowledgments

This work was supported by the Center for Materials for Water and Energy Systems (M-WET), an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Basic Energy Sciences under Award DE-SC0019272. Use was made of computational facilities purchased with funds from the National Science Foundation (OAC-1925717) and administered by the Center for Scientific Computing (CSC). The CSC is supported by the California NanoSystems Institute and the Materials Research Science and Engineering Center (MRSEC; NSF DMR 1720256) at UC Santa Barbara. S.J., J.I.M., and M.B. acknowledge support by the National Science Foundation Graduate Research Fellowship (DGE 1650114, DGE 1144085). M.B. also recognizes support from the Office of Naval Research (ONR) under Award N00014-17-1-2047. A.D. acknowledges support by the National Defense Science & Engineering Graduate (NDSEG) Fellowship Program. The authors gratefully acknowledge access to the WavPDS and SVDRReconstruction softwares through the National Biomedical Center for Advanced ESR Technology (ACERT). ACERT is hosted by the Cornell Center for Advanced Computing and supported by the National Institute of General Medical Sciences. Conversations with M. Srivastava regarding use of WavPDS and SVDRReconstruction were particularly helpful. We acknowledge use of MRL Shared Experimental Facilities which are supported by the MRSEC Program of the NSF under Award No. DMR 1720256; a member of the NSF-funded Materials Research Facilities Network (www.mrfn.org). We also acknowledge helpful discussions regarding DEER acquisition and analysis with T. Keller and K.

Tsay and conversations with R. Zuckermann regarding polypeptoid synthesis. We thank A. Prakash and J. Pfaendtner for providing modified AMBER99SB-ILDN forcefield files for peptoids and helpful discussions concerning their implementation. We thank R. Jain and E. Santiso for helpful discussions concerning the implementation of the modified CGenFF forcefield.

Chapter 5

Sequence modulates polypeptoid hydration water structure and dynamics

Adapted with permission from *Biomacromolecules* 2022, 23, 4, 1745-1756. Copyright 2023 American Chemical Society.

5.1 Introduction

In the last chapter, we developed and validated a workflow for simulating polypeptoids, a highly programmable platform that is useful for exploring sequence-structure-property relationships. In this chapter, we will use the workflow to investigate the effect of molecular chemical sequence and hydration water structure and dynamics. This is a key relation-

ship to understand because hydrated surfaces play essential roles in many applications, including protein folding, self-assembly, water filtration membranes, marine antifouling and drug delivery.^{225–227} Characterizing the behavior of water near surfaces and the manner by which such surfaces perturb water’s structure and dynamics can provide valuable insight into functional properties^{52,228–230} and thus guide the design of novel materials. Uncovering these links has been a longstanding area of research, and in particular has motivated many molecular simulation studies of model systems that have offered nanoscopic insight into the mechanisms by which surfaces modulate water behavior.^{55,61,141,145,231,232} While much progress has been made in understanding water near macroscopic surfaces, it remains challenging to disentangle hydration behavior near chemically heterogeneous and conformationally flexible macromolecules. In such cases, a complex interplay exists between polymer composition (e.g., hydrophobic versus hydrophilic content), sequence (primary patterning of different chemical groups), and dynamic conformational changes, all of which can influence water’s response and the effective water-mediated interactions.

In this chapter, we examine families of peptoid oligomers with precisely varied composition and sequence. In simulations, we are then able to separate the effects of sequence and conformation and reveal how these factors affect hydration water structure and dynamics. Importantly, we find measures of water structure that are consistently correlated with dynamics. This suggests the viability of probing water’s response to a molecular surface through experimental measurements of local hydration water dynamics, which we in turn demonstrate.

Previous works have shown that small nonpolar solutes enhance their hydration waters' structure^{233–236} and slow their dynamics,^{237–240} the reverse effect of nonpolar groups at extended, idealized surfaces. The observed differences in water's structural responses to small and large hydrophobes is consistent with the theoretical picture given by Lum, Chandler, and Weeks⁵⁶ in which waters around a small ($\approx < 1$ nm) nonpolar body are more structured to preserve their hydrogen-bonding networks ("small-lengthscale behavior"), whereas waters around a large, smooth nonpolar body form a soft interface with dangling hydrogen bonds ("large-lengthscale behavior"). However, the interplay of chemical and topological heterogeneity can complicate water's response.^{145,148}

Chemically heterogeneous but conformationally dynamic molecules epitomize the complex relationship between small and large lengthscale behaviors due to conformational fluctuations in addition to topological heterogeneity. Biswas and coworkers found that disordered protein regions possess distinct hydration behaviors from water near structured regions.²⁴¹ They also found differences between the hydration of nonpolar and polar residues of these partially disordered proteins,²⁴² but few studies have systematically compared water behavior near disordered molecules of varying chemistries as both simulations and experiments face tremendous challenges in probing this behavior. Surface-sensitive experimental techniques developed for extended surfaces cannot probe dynamics near molecular surfaces. Additionally, shifts in water's response due to conformational change or surface heterogeneity may be obscured by experiments that only report on averaged behaviors.

Another experimental challenge is that chemical heterogeneity can be difficult to precisely program. Polypeptoids are highly tunable molecules as they can be synthesized sequence-specifically *via* a process that allows access to a wide variety of chemistries.^{73,243} Previous computational efforts have begun to examine water behavior near polypeptoids in solution and at surfaces,^{209,244} but have been hampered by slow isomerization timescales (seconds²⁰²) in polypeptoids that are inaccessible to molecular dynamics simulation. Here, we leverage the previously developed computational workflow that accurately samples polypeptoid conformational landscapes in solution²¹² to probe the effects of polypeptoid sequence on local water behavior. We first examine the effect of an entirely polar peptoid sequence on hydration water behavior, making comparisons to bulk water (section 5.3.1). We then compare these findings to water’s local dynamics near peptoids containing mixtures of nonpolar and polar residues, varying both their numbers and sequence (i.e. arrangement along the backbone) (section 5.3.2). We demonstrate significant correlations between water structure and dynamics modulated by the peptoid sequence (section 5.3.3). Finally, we assess the feasibility of using Overhauser Dynamic Nuclear Polarization, a technique that can measure local dynamics, to probe water dynamics in the peptoid hydration shell (section 5.3.4).

5.2 Methods

Methods associated with peptoid synthesis and characterization using Overhauser Dynamic Nuclear Polarization are described in Ref. [245].

5.2.1 Peptoid simulation

Generating representative conformations

The peptoid sequences chosen are discussed at the beginning of section 5.3. We use the MFTOID forcefield,²⁰⁶ which we have previously shown accurately captures the experimentally measured structural distributions of disordered polypeptoids, such as the ones we study here.²¹² To prepare the system, we create and geometry-minimize initial peptoid structures in the fully extended conformation using Avogadro 1.2.0.²¹³ We perform simulations using the OpenMM simulation engine.⁷¹ Electrostatic interactions are computed using particle-mesh Ewald. Lennard-Jones nonbonded interactions are cut off at 10Å. Hydrogen bonds are constrained with SHAKE¹²³ and water is kept rigid with SETTLE.¹²⁴ We first run a simulation in an expanded ensemble over softened ω dihedral states, a technique we previously developed and validated that samples across slow isomerizations in the polypeptoid backbone,²¹² using Langevin dynamics with an integration timestep of 2 fs, a temperature of 300 K, and a friction coefficient of 0.01 ps⁻¹. The pressure is set to 1 bar using a Monte Carlo (MC) barostat with 500 fs between MC barostat moves. We save conformations every 200 ps. System and simulation details for each sequence are given in Table D.1. We then use MBAR^{125,218} to compute the weights for each conformation, and obtain a random sample of 100 conformations using these weights. To calculate dynamical, structural, and other quantities, we then perform 100 simulations in the microcanonical (constant NVE) ensemble starting from each of these conformations. During these “production” simulations, we note that the conformations

of the peptoids do not vary significantly (Fig. D.2) due to slow backbone isomerizations, such that sampling of conformation space is accomplished *via* the 100 representative starting samples generated from the expanded ensemble pre-run (Fig. D.2).

Peptoids in solution

We place each peptoid in a box of length at least $20\text{\AA}$ longer than its end-to-end distance when fully extended and solvate the boxes using GROMACS 2016.1^{214,215} with the modified TIP3P water model.²⁴⁶ We run the expanded ensemble with this water model and then, after choosing the set of 100 starting conformations, resolute conformations in OPC water, which has been shown to more faithfully reproduce experimental measurements of water’s bulk diffusivity,^{247,248} and perform a local energy minimization. For each NVE simulation, we first run an NVT equilibration for 100 ps using Langevin dynamics with a timestep of 2 fs, a temperature of 300 K, and a friction coefficient of 1 ps^{-1} . We then perform an NPT simulation for 100 ps at 1 bar using Langevin dynamics with the same conditions and an MC barostat with 200 fs between MC moves, saving configurations every 2 ps. We obtain the average box length from the second half of the NPT simulation and run another NVT equilibration for 100 ps using Langevin dynamics with the same conditions, saving configurations every 2 ps. We then initialize particle velocities so that the total energy matches the average total NVT energy (computed from the second half of the last NVT equilibration) and run a simulation in the microcanonical ensemble for 250 ps using the Verlet integrator with a timestep of 2 fs and saving conformations every 50 fs. Unless otherwise noted, all conformations are used in the

computation of structural and dynamical metrics.

Peptoids on silica

We prepare a slab of amorphous silica with silanol density of 4.0 nm^{-2} as in Ref. [249] using forcefield parameters from Ref. [250]. To mimic the experimental tethering of the peptoid to the silica surface, we add a glycine residue, with the same peptoid backbone parameters, to the N-terminus of the peptoid and add an acetyl capping group (Fig. D.4). We initialize the peptoid in the extended conformation oriented perpendicular to the silica surface with its N-terminus closest to the surface. We initialize the box length perpendicular to the silica surface to 7 nm, which, when the peptoid is fully extended, leaves approximately 1.5 nm between the peptoid's C-terminus and the other side of the silica surface. We solvate the box with OPC water using GROMACS 2016.1.^{214,215} During the simulation, the distance between a carbon in the acetyl capping group and a hydrogen on the silica surface is fixed using a harmonic restraint with force constant 1000 kJ/mol/nm². We run the expanded ensemble simulation, allowing all amide bonds to isomerize except that between the acetyl capping group and the glycine residue. We choose 100 starting conformations for the NVE simulations, weighted by the MBAR-computed conformation weights, and set the velocities to match the average total energy in the expanded ensemble. We then run a production simulation in the microcanonical ensemble for 250 ps using the Verlet integrator with a timestep of 2 fs and saving conformations every 50 fs.

Free spin label

Parameterization details for 4-OH TEMPO are given in section D.1.5. A single 4-OH TEMPO molecule is placed in a box of initial length 3.07 nm and solvated with 930 water molecules. 4-OH TEMPO simulations are performed following the same workflow as bulk water, described in section D.1.3.

5.2.2 Computational hydration water characterization

For each peptoid and property, we perform independent computations for each of the 100 NVE simulations. Then, to obtain uncertainties, we bootstrap, choosing a random sample of 100 simulations, averaging the property (a distribution or a trajectory) and repeating for a total of 1000 samples. We then take the 2.5th and 97.5th percentiles to get bounds for the reported 95% confidence intervals.

Radial distribution function

We compute the density of water molecules as a function of distance from each peptoid alpha carbon and then normalize by the average bulk water density at the same conditions. We then compute the radial distribution functions (RDFs) using a bin width of $0.125\text{\AA}$ up to $15\text{\AA}$ for the water oxygens as a function of distance from each alpha carbon. The RDFs for each alpha carbon are then averaged.

Three-body angle distribution

The three-body angle distribution is the distribution of angles formed by a central water oxygen and the oxygens of any pair of its neighboring waters, and has been shown to reveal signatures of water structuring in response to different types of solutes and interfaces.⁶² Here, we compute the distribution for peptoid hydration waters, i.e., central waters within $5\text{\AA}$ of any heavy atom on the peptoid. The $5\text{\AA}$ cutoff corresponds to the first minimum in the RDF between the peptoid alpha carbons and water oxygen atoms. Neighboring waters are defined as those whose oxygens lie within $3.4\text{\AA}$ of a central water. This $3.4\text{\AA}$ central-neighbor cutoff is the location of the first minimum in the RDF for bulk OPC water (Fig. D.3). The three-body angle distributions are then computed with a bin width of 4.86 degrees using conformations saved every 5 ps.

Mean-squared displacement

To calculate hydration water diffusivity, we track the mean-squared displacement (MSD) of water oxygens within $5\text{\AA}$ of an alpha carbon for a total time of 10 ps. This time scale is sufficient to reach the linear regime of the MSD while not allowing a significant fraction of hydration waters to move away from the peptoid.

Survival probability

We also characterize translational dynamics through survival probability calculations, which track the rate at which hydration waters leave a pre-defined cutoff around the peptoid. Every 50 ps, the waters within $5\text{\AA}$ of the alpha carbon are tagged and we track

the number remaining within the cutoff as a function of time, up to 50 ps. We average these 50 ps windows to obtain a within-cutoff time correlation function, which allows us to then extract an effective diffusivity by fitting to the solution to the diffusion equation in a spherical region with a constant initial condition and absorbing boundary condition:

$$P(t) = \frac{C_0\sqrt{Dt}}{\pi R} \left(e^{-R^2/Dt} - 3 \right) + \frac{2C_0(Dt)^{3/2}}{\pi R^3} \left(1 - e^{-R^2/Dt} \right) + C_0 \operatorname{erf} \left(\frac{R}{\sqrt{Dt}} \right) \quad (5.1)$$

where R is the radius of the cutoff region, D is the diffusivity, and C_0 is the initial concentration (section D.2.1). Here, we fit both D and C_0 over the region 5-50 ps.

Dipole dynamics

To quantify orientational water dynamics, we measure the rate of water dipole reorientation. Every 50 ps, the waters within $5\text{\AA}$ of the alpha carbon are tagged and then we track the cosine of the angle between the initial and instantaneous dipole vectors over the subsequent 50 ps and average the trajectories over these 50 ps windows. We extract a timescale by fitting the decay to a stretched exponential: $\cos \theta = \exp(-(t/T_{\text{orient}})^\beta)$ where we fit both T_{orient} , the timescale, and β , over the region 1-50 ps.

5.3 Results

To systematically explore the effect of sequence on local water behavior, we simulate a series of peptoid oligomer sequences built from polar and nonpolar residues, as shown in Fig. 5.1. Because peptoid chemistry allows access to a wide diversity of sidechains,

we can select polar and nonpolar sidechains that are comparable in size, in contrast to canonical amino acids. The polar methoxyethyl sidechain or “NME” residue (“E” in peptoid sequence labels used here) has been studied previously as a model polar group in self-assembly systems¹⁷⁵ and as a promising antifoulant,⁷⁴ and the nonpolar n-butyl sidechain or “alkyl” residue (“N” in peptoid sequence labels used here) has the same number of heavy atoms and same linear geometry. We fix the total number of residues in the peptoids at a length that makes systematic exploration of sequence effects feasible. We also fix the location of the nitroxide spin label residue (“T” in peptoid sequence labels used here), used for experimental measurements of local water translational dynamics using Overhauser Dynamic Nuclear Polarization (section 5.3.4). In one series, we explore the effect of the number of alkyl residues, keeping their placement (at the C-terminus) constant, and in another series, we explore the effect of the placement of the alkyl residues, keeping their number (3) constant. We will refer to the homopolymer of polar residues, EEEEE-T, as the NME homopolymer peptoid, and to the homopolymer of nonpolar residues, NNNNN-T, as the alkyl homopolymer peptoid.

5.3.1 Peptoids perturb water structure and dynamics

To obtain a baseline picture of how peptoid hydration water differs from bulk water, we quantify shifts in water structure and dynamics from bulk water near the most hydrophilic peptoid oligomer, the NME homopolymer (EEEEEE-T in Fig. 5.1). Near that peptoid’s backbone, we find that water is less dense than bulk water. Fig. 5.2a shows the backbone-

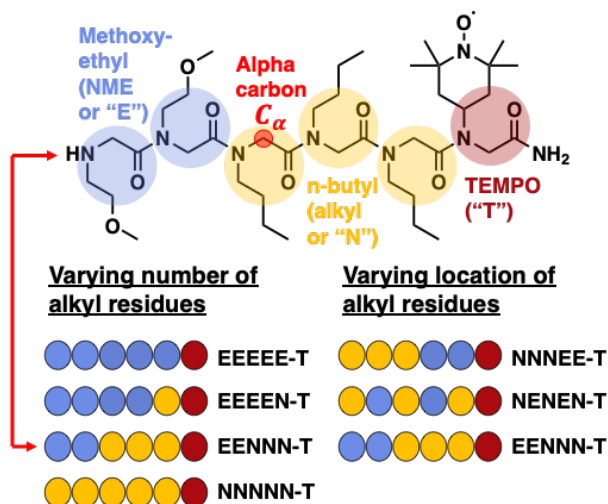


Figure 5.1: Peptoid oligomers in solution. We vary the positions of the polar “NME” residue and the nonpolar “alkyl” residue, keeping the overall peptoid length constant. In one series, we vary the number of alkyl residues keeping the placement (at the C-terminus) constant and in another we vary the placement of the alkyl residues keeping the number constant. The C-terminal TEMPO residue allows experimental probes of local water translational dynamics using Overhauser Dynamic Nuclear Polarization (section 5.3.4). The circled carbon denotes one of the peptoid’s alpha carbons or C_α (each residue contains one). Only waters within $5\text{\AA}$ of these alpha carbons are considered for the mean-squared displacement, survival probability, and reorientational time calculations (section 5.2.2).

water radial distribution function, which as one might expect is less than unity at close distances (63% of bulk water density at the first peak at $4.7\text{\AA}$), suggesting a dehydrated backbone. The shape of the RDF also indicates that the first hydration shell extends to approximately $5.3\text{\AA}$, where the local minimum is, and that the subsequent hydration shells are less well-defined, with only shoulders at $7.3\text{\AA}$ and $9.4\text{\AA}$. This suggests the peptoid does not structure water strongly past approximately $5\text{\AA}$, which we use as a cutoff for subsequent measures.

The RDF does not provide direct insight into collective aspects of water structure, such as its hydrogen bonding-induced tetrahedrality, known to be fundamental to water’s unique behavior.⁴⁸ Dhabal *et al.* previously reported that three-body measurements

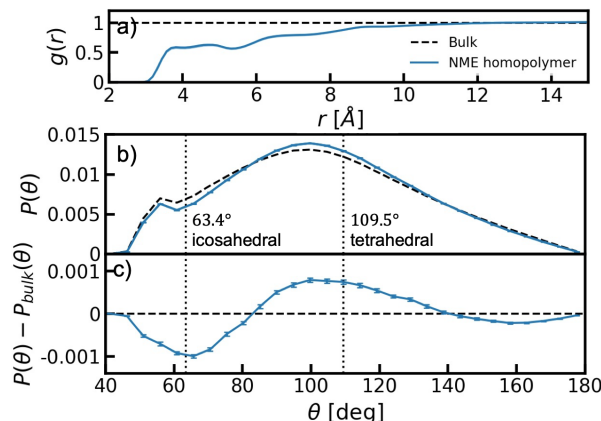


Figure 5.2: The NME homopolymer peptoid perturbs water from its bulk structure. (a) The radial distribution function (solid line) shows that the peptoid backbones are dehydrated compared to bulk (dashed line). (b) Compared to bulk water (dashed line), the three-body angle distribution for the peptoid hydration waters (solid line) are slightly enhanced in the tetrahedral (100° - 120°) region. (c) Subtracting out the bulk distribution more clearly shows the enhancement in the tetrahedral region for the peptoid’s hydration waters.

are needed to probe tetrahedrality.²⁵¹ Monroe *et al.* showed that the three-body angle distribution in particular can capture water’s collective behavior near a solute, and is predictive of thermodynamic properties of solvation.⁶² The three-body distribution for bulk water has two distinctive peaks: one near 63.4° characteristic of a simple, close-packed fluid in an icosahedral configuration, and one near 109.5° characteristic of tetrahedral configurations⁶² (Fig. 5.2b, dashed line). Near the NME homopolymer’s heavy atoms, the tetrahedral peak is enhanced and the icosahedral peak diminished (Fig. 5.2b, solid line), compared to bulk water, as more clearly illustrated in Fig. 5.2c, which shows the difference between the distributions of the peptoid hydration water and bulk water.

The observed shifts in water structure accompany reductions in water’s translational and orientational dynamics. Fig. 5.3a shows the mean-squared displacement (MSD) of water in the bulk and near the peptoid backbone alpha carbons (an alpha carbon

is circled in the chemical structure presented in Fig. 5.1). The hydration water MSD and hence its diffusivity is lower compared to bulk: the computed diffusivities from the MSDs are $0.246 \text{ \AA}^2/\text{ps}$ for bulk and $0.146 \text{ \AA}^2/\text{ps}$ for the peptoid's hydration water, showing a significant decrease of roughly 40% for water near the peptoid surface.

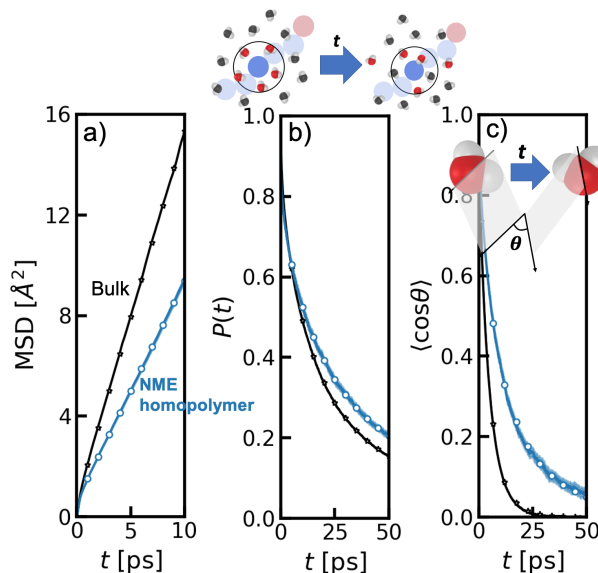


Figure 5.3: The NME homopolymer peptoid slows down its hydration water. (a) The mean-squared displacement, (b) survival probability, and (c) dipole correlation all show slower water near the peptoid's backbone (blue lines) compared to bulk behavior (black lines). The fits to the models (circular and star markers for peptoid hydration water and bulk water, respectively) are good and the fitted dynamical quantities show the same trend.

The MSD calculation may fail to capture highly localized behavior, as some water molecules can move out of the cutoff region during the 10 ps observation interval. The survival probability, $P(t)$, shown in Fig. 5.3b, provides an alternate, strictly localized approach to report on translational dynamics. This metric gives the probability that a water molecule within a cutoff region at time $t = 0$ will remain in that region after time t . Bulk water's survival probability (computed within a $5 \text{ \AA}$ radius sphere in a box of bulk water, simulation details in section D.1.3) decreases faster with time than that of water

near the peptoid backbone alpha carbons, suggesting slower dynamics in the latter case. By fitting survival probability time traces to a diffusion model based on the diffusion equation in a spherical geometry (section 5.2.2), we obtain a diffusivity for hydration water ($0.170 \text{ \AA}^2/\text{ps}$) that is 30% lower than that of bulk ($0.248 \text{ \AA}^2/\text{ps}$), mirroring the trend seen by the MSDs. The two computed metrics of translational diffusivity are generally correlated (Fig. D.7), but the diffusivities obtained from the survival probability are consistently higher than those from the MSD, as the diffusivity of the peptoid itself also contributes to the decay of the survival probability (section D.2.2).

Water’s orientational dynamics similarly slow near the peptoid backbone compared to the bulk. Fig. 5.3c shows the decay of $\langle \cos \theta \rangle$ for the angles θ between initial and later-time dipoles made by hydration waters. As with the survival probability, the correlation for bulk water decays faster than for hydration water near the peptoid backbone, indicating slower orientational dynamics in the latter. By fitting to a stretched exponential (section 5.2.2), we find characteristic rotational correlation times, T_{orient} , of 5.1 ps for bulk and 15.4 ps for hydration water, respectively. These collective findings are qualitatively consistent with previous computational studies showing that water is slower, both translationally and rotationally, near short peptide surfaces,^{252,253} though quantitative differences in the magnitude of the reduction exist, likely due to differences in sidechain chemistry.

5.3.2 Sequence and conformation affect water dynamics

The results discussed so far demonstrate that the NME homopolymer significantly perturbs its hydration water relative to bulk water. Now, we examine whether changes in the number and sequence of nonpolar alkyl residues in the peptoid can tune that perturbation. We compare local water dynamics around the backbone with varying sequences of the polar NME residue (“E”) and the nonpolar alkyl residue (“N”) (see Fig. 5.1 for chemistries). Fig 5.4 shows water diffusivities extracted from the survival probability and rotational correlation times extracted from the dipole correlation (full trajectories in Fig. D.16-D.17), while varying both the number and placement of the alkyl residues (EEEEEE-T, EEEEEN-T, EENNN-T, NNNNN-T, NNNEE-T, NENEN-T). When comparing the set of sequences with increasing number of alkyl residues at the C-terminus of the peptoid (EEEEEE-T, EEEEEN-T, EENNN-T, NNNNN-T), we observe that the addition of alkyl residues leads to slower translational and rotational dynamics of hydration water. The decreases do not scale linearly with the number of alkyl residues, suggesting that the effect is non-additive, i.e. that it cannot be separated into distinct contributions from each residue. The three sequences with the same number of alkyl residues, but varying placement (EENNN-T, NNNEE-T, NENEN-T), also show differences in their hydration water dynamics, which further reveals the non-additive effects of peptoid chemistry on water dynamics; the sequence, in addition to total composition, modulates the hydration water dynamics. Specifically, the peptoid oligomer with its alkyl residues at the C-terminus (EENNN-T) shows slower dynamics than the ones with these residues at the

N-terminus or distributed (NNNEE-T, NENEN-T).

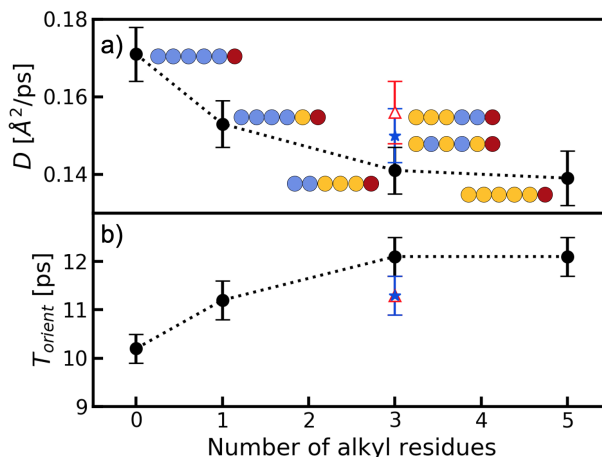


Figure 5.4: Both the number and arrangement of nonpolar residues affect local translational and orientational dynamics. Both the (a) translational diffusivity and (b) rotational correlation time show that water is slower near peptoids with more nonpolar residues at the C-terminus. Variations in the placement of the nonpolar residues can also affect local dynamics: water is faster when the nonpolar residues are moved further away from the C-terminus in the sequence.

Measuring water dynamics locally around individual residues provides additional evidence of the non-additive effect of peptoid sequence on hydration water behavior. The translational diffusivity measured around all residues of the alkyl homopolymer (Fig. 5.5a, orange stars) is consistently lower than the diffusivity around the position-equivalent residues in the NME homopolymer (Fig. 5.5a, blue circles). Slower dynamics around the alkyl homopolymer is consistent with the reduction in dynamics reported in Fig. 5.4. However, for the sequences containing both alkyl and NME residues (Fig. 5.5b, 5.5c, 5.5d), the dynamics around the individual residues do not match those of the homopolymer residues of the same chemistry and position. There are, instead, significant deviations between the diffusivity around the NME residues in the heteropolymers (circles) and the diffusivity around the NME homopolymer (upper bound of the shaded region), as well

as between the diffusivity around the alkyl residues (stars) and the diffusivity around the alkyl homopolymer (lower bound of the shaded region). The dynamics are clearly influenced by the chemistry of the neighboring residues. For instance, water near the N-terminal NME residues in EENNN-T (Fig. 5.5b) is significantly slower than water near the N-terminal residue in the NME homopolymer (compare left-most marker with the upper bound of the shaded region in Fig. 5.5b). These trends are also reflected in the rotational dynamics (Fig. D.18).

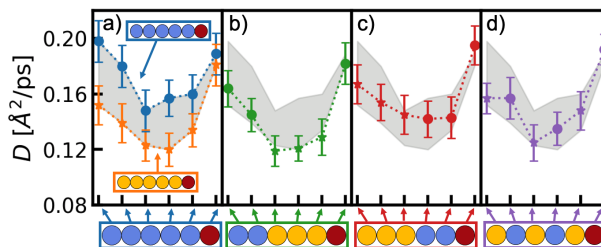


Figure 5.5: The effect of sequence on water dynamics is non-additive. The translational diffusivity measured around each alpha carbon of the peptoids (instead of averaged over all alpha carbons, as is reported in Fig. 5.4a) show complicated trends that reflect the chemistry of the residue’s neighbors, in addition to the residue itself. Trends for the homopolymers are plotted in panel (a) and are shown as a reference in all other panels as gray, shaded regions. For the heteropolymers, the sequences corresponding to each panel are shown below the panels. The arrows indicate the residue around which the diffusivity is measured. Star-shaped markers represent alkyl residues while circular markers represent NME residues. Thus, for the star-shaped markers, the position-equivalent diffusivity values for the homopolymer of the same chemistry are the lower boundary of the shaded region, while for the circular markers, the position-equivalent diffusivity values for the homopolymer of the chemistry are the upper boundary of the shaded region.

For conformationally dynamic molecules, residue sequence can also affect hydration water dynamics by inducing different conformational ensembles. To quantify such effects here, we calculate end-to-end distance distributions, $P(R_{ee})$, which give the probability of the terminal nitrogens on the peptoid backbones adopting a defined separation dis-

tance, R_{ee} . The end-to-end distance distribution for peptoids with more alkyl residues is shifted to lower distances (Fig. D.1), suggesting that these sequences are slightly more compact. To determine whether the observed effect of sequence on water dynamics in Fig. 5.4 is influenced by this slight shift in conformation, we separate out the effect of conformational changes by comparing hydration dynamics between peptoids with similar conformation. For this analysis, we focus on the NME homopolymer and the alkyl homopolymer, since they show the largest differences in dynamics as well as conformation. We identify common structural states of NME and alkyl homopolymers through joint conformational clustering (section D.3.1). We then compare the water dynamics around the most extended (average end-to-end distance, $\langle R_{ee}^2 \rangle^{1/2}$, is $12.8\text{\AA}$) NME homopolymers and similarly extended alkyl homopolymers, as well as around the most compact ($\langle R_{ee}^2 \rangle^{1/2}$, is $7.1\text{\AA}$) NME homopolymers and similarly compact alkyl homopolymers.

Fig. 5.6 compares the translational diffusivity for these four structural groups. When comparing the extended conformations, the NME homopolymer shows faster water dynamics than the alkyl one. The trend is the same for the groups of compact conformations. Again, the same trends are reflected in the rotational dynamics (Fig. D.19). Thus, the effect of sequence remains qualitatively the same whether we compare water around NME and alkyl homopolymers with similar conformations, or average over all conformations, even though the conformational distributions themselves shift between the two sequences. This suggests that the effect of sequence on hydration water dynamics here is primarily due to a change in the local chemical environment, and not due to variations

in the peptoid conformational landscape.

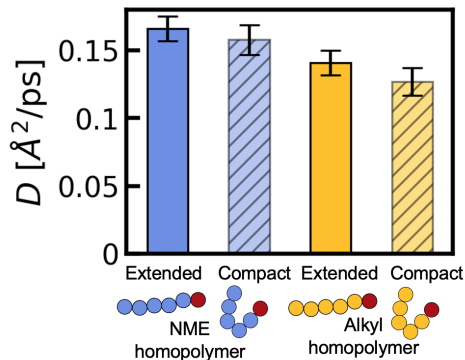


Figure 5.6: Comparison of water dynamics around different sequence with similar conformations shows the effect of sequence is qualitatively independent of conformation. Comparing groups of similar conformations from both the NME homopolymer and alkyl homopolymer sequences (solid bars are the extended conformations, hatched bars are the compact conformations), water remains slower near the alkyl homopolymer peptoid. Between groups for a given sequence (blue bars are the NME homopolymer sequence, yellow bars are the alkyl homopolymer), water is slower near conformations with lower end-to-end distance.

5.3.3 Sequence-induced shifts in water dynamics and structure are correlated

Section 5.3.1 showed that water around the peptoid is slower than in bulk, as well as less dense and more tetrahedral, suggesting that slower dynamics may be linked to more structured water. To confirm this, we computed the three-body angle distributions for all sequences studied (Fig. D.15). As the number of nonpolar residues increases, the three-body angle distributions shift further from bulk, exemplified by the distribution for the alkyl homopolymer shown in Fig. 5.7. Specifically, the characteristic peaks for the tetrahedral (100-120°) regions increase in magnitude, at the expense of the icosahedral (50-70°) regions, indicating a larger population of structured, tetrahedral water. To

quantify this increase in tetrahedrality, we integrate the distribution in the tetrahedral region: $P_{\text{tet}} = \int_{100}^{120} (P(\theta) - P_{\text{bulk}}(\theta)) d\theta$. These values are shown in the inset of Fig. 5.7. The trends clearly demonstrate the increase in tetrahedrality with the number of nonpolar residues, as well as the dependence on sequence, independent of the number of residues. The trends indeed mirror the effects of sequence on water dynamics shown in Fig. 5.4.

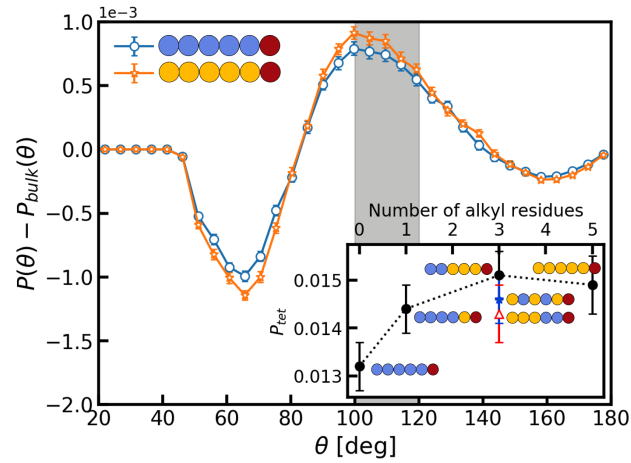


Figure 5.7: The three-body angle distribution shows enhanced tetrahedrality for the alkyl homopolymer compared to the NME homopolymer. The circular markers show the same distribution as Fig. 5.2c for the NME homopolymer, where the bulk water three-body angle distribution has been subtracted out. The star-shaped markers show the same distribution for the alkyl homopolymer, and there is a clear enhancement of the distribution in the tetrahedral region (shaded gray) and diminishment in the icosahedral region. (Inset) The integrated tetrahedral region, $P_{\text{tet}} = \int_{100}^{120} (P(\theta) - P_{\text{bulk}}(\theta)) d\theta$ for all the sequences studied shows that the enhancement in tetrahedrality increases with the number of nonpolar residues and is also sequence-dependent independent of composition.

Remarkably, this structural measure of tetrahedrality, P_{tet} , is quantitatively correlated with the dynamical quantities shown in Fig. 5.4. To illustrate, Fig. 5.8 gives cross-plots of the translational diffusivity (D), rotational correlation time (T_{orient}), and P_{tet} , along with another structural metric, N_{water} , which quantifies local water density by reporting

the total number of waters within a $15\text{\AA}$ radius of a peptoid heavy atom (obtained from integrating the RDF; full distributions in Fig. D.14). Note that Fig. 5.8 reveals strong relationships between tetrahedrality and dynamics in particular, with weaker but still apparent correlations between N_{water} and the dynamic measures. Similar relationships have been found in simulation studies of bulk water,^{254–257} small molecules,²⁵⁸ and extended surfaces.^{61,259} We comment more on this connection in section 5.4.

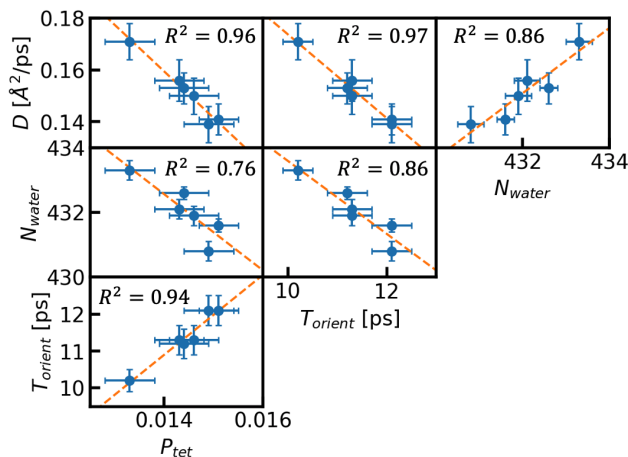


Figure 5.8: Local water structure and dynamics are correlated. The strongest correlations are between diffusivity from the survival probability (D), the timescale of dipole reorientation (T_{orient}), and the integrated tetrahedral region from the three-body angle distribution (P_{tet}). There are weaker correlations between these quantities and the number of waters computed from integrating the RDF (N_{water}).

5.3.4 The spin label captures changes in water dynamics

Overhauser Dynamic Nuclear Polarization (ODNP) can directly probe local water dynamics near a spin label that can be placed at different locations in the system.^{260,261} Previous work on a globular protein has shown that ODNP can measure the subtle site-specific differences in local water dynamics that are due to the folded protein surface’s

chemical and topological heterogeneity.²⁶² A crucial question is whether it is possible to detect sequence-specific or site-specific differences in local water dynamics near conformationally dynamic spin-labeled biomacromolecules. As a proof of concept for using ODNP to probe distinctions in peptoid hydration water dynamics, we study three systems: a 4-OH TEMPO spin label in solution, an NME-homopolymer peptoid in solution, and an NME-homopolymer peptoid tethered at its N-terminus to a hydrated silica surface. The peptoids are spin-labeled at the C-terminus. Details regarding the peptoid synthesis and ODNP measurements are given in Ref. [245].

Fig. 5.9 shows translational diffusivities computed from simulations (dark, solid bars; details in section D.1.5) and measured from ODNP (light, hatched bars). The results demonstrate that waters near the spin label on a peptoid are slower than water near a spin label in bulk solution. Both ODNP and MD simulation show an additional reduction in water dynamics around a peptoid tethered to a silica surface compared to the free peptoid. Notably, for both the free and tethered peptoid, ODNP registers a larger decrease in water diffusivity than MD simulations, also seen in work by Schrader *et al.* on silica surfaces.²⁴⁹ This difference in retardation may result from a variety of factors including differences in temperature (300 K for MD vs. 291.15 K for ODNP) or differences in hydrogen bond fluctuations and hydrogen hopping (accounted for in ODNP,²⁶¹ but not MD). Regardless, qualitative agreement between ODNP and MD findings demonstrate ODNP's ability to probe site-specific shifts in local water dynamics.

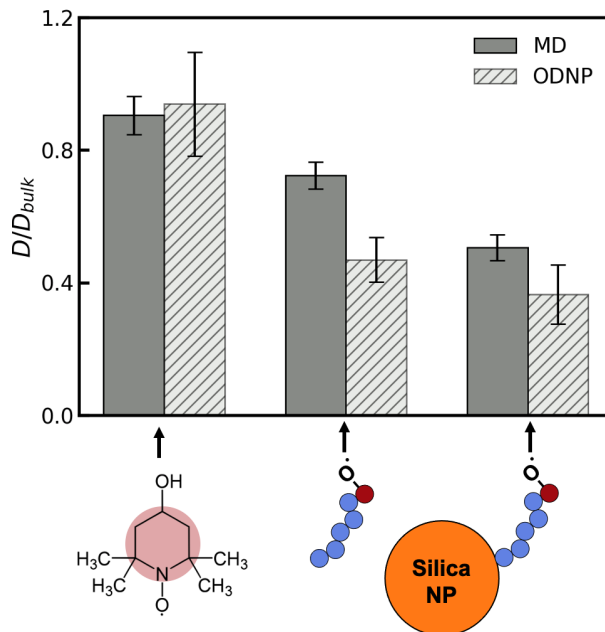


Figure 5.9: Water dynamics near the spin label shows clear decreases when the spin label is tethered to a peptoid and again when the peptoid is tethered to a silica nanoparticle. The MD diffusivities (dark, solid bars) are computed from the survival probability of water within $7\text{\AA}$ of the spin label’s oxygen radical atom, approximately encompassing the waters captured by ODNP.²⁶¹ The experimental ODNP measurements (light, hatched bars) show similar qualitative trends. Both measurements are normalized by an appropriate measure of bulk water diffusivity: $D_{\text{bulk,MD}} = 0.243\text{\AA}^2/\text{ps}$ is computed as described in section 5.3.1 from the survival probability in bulk water at 300 K and $D_{\text{bulk,ODNP}} = 0.19162\text{\AA}^2/\text{ps}$ is interpolated from a series of experimental results at varying temperatures^{263–271} where the interpolation is performed in Ref. [272] to 291.15 K.

5.4 Discussion

The present results demonstrate that peptoid oligomers with both polar and nonpolar side chains perturb their hydration waters from bulk behavior and that the perturbation is modulated by sequence. Further, we find that shifts in water dynamics are tightly correlated with structural metrics. Taken together, and in the context of previous literature on extended surfaces, these results address major points related to the hydration of disordered molecules, namely: At what lengthscale does water respond to disordered

molecules? Are these responses due to changes in conformation or local chemical environment? Are these responses correlated?

5.4.1 Water shows small-lengthscale responses to disordered polypeptoids

The reduction in dynamics with increasing nonpolar content shown in Fig. 5.4 is consistent with previous work for nonpolar sites on small-molecule surfaces that suggests changes in water’s local chemical environment, and in particular the unavailability of hydrogen-bonding partners, directly affect water dynamics.^{252,273} However, this behavior is opposite of the trend observed for extended surfaces, where it has been shown that the chemistry of the surface—both the types of chemical groups and their spatial arrangement—affects local water structure, dynamics, and thermodynamics.^{61,70,130,132,274–276} In particular, water near extended surfaces comprising polar groups exhibits increased structuring,⁷⁰ smaller density fluctuations,⁷⁰ and slower dynamics,⁶¹ especially when those polar groups are dispersed evenly across the surface.^{61,131} Near extended hydrophobic surfaces, on the other hand, water shows reduced structuring, larger density fluctuations, and faster dynamics.

Our finding suggests that even though the polypeptoid is generally larger than molecules observed to induce large-lengthscale hydrophobic behaviors, such as neopentane,²⁷⁷ water nevertheless responds to the polypeptoid surface as if it were a smaller molecule. This is further supported by the observation shown in Fig. 5.6 that changes in conformation

that straddle the 1 nm crossover lengthscale have little impact on the shift in water dynamics with sequence variation. The effective lengthscale, from the perspective of water, is thus that of the local peptoid residue and adjacent residues, as opposed to the entire molecule. This phenomenon could be due to the “rough” molecular surface that polypeptoids present, preserving small-lengthscale behavior. Furthermore, the relatively polar peptoid backbone that can hydrogen-bond with water may also reduce water dynamics and hinder the emergence of responses characteristic for large-lengthscale hydrophobic behavior.

This picture of small-lengthscale response is further reinforced by the analysis of water’s three-body angle distributions around the peptoids compared to bulk, showing that for all sequences the hydration water manifests enhanced tetrahedrality, and the addition of nonpolar sidechains enhances tetrahedrality further (Fig. 5.7 and Fig. D.15). Water in the peptoid system thus shows small-lengthscale responses, despite the fact that the peptoid is more comparable in size to larger colloids around which water shows shifts away from tetrahedrality.⁶²

5.4.2 In disordered polypeptoids, sequence has a larger effect on water behavior than conformation

For extended, idealized surfaces, topology or surface “roughness” does affect local water dynamics, in addition to chemistry.^{141,145} For realistic surfaces, both factors have also been shown to affect dynamics.^{141,148} However, Persson *et al.* argue that differences

in topology, not polarity, account for shifts in water dynamics at protein surfaces and present a model for the reduction in dynamics based solely on a “confinement index” that is independent of protein chemical identity.²⁷⁸ On the other hand, Sterpone *et al.* presented an “extended jump model” describing water reorientational dynamics that has contributions from both topology, in the form of an excluded volume contribution, and chemistry, in the form of solute-water hydrogen-bonding strength.^{273,279}

Here, comparison of water around the most extended and compact conformations of NME homopolymers (Fig. 5.6) shows a small effect of conformation, as water near peptoids with smaller mean end-to-end distances show consistently slower dynamics. The same trend is seen for water around extended and compact alkyl homopolymers. This suggests that more compact polypeptoids lead to more structured and slower water. The peptoid oligomers explored in this work, however, are still relatively small (near the 1 nm crossover), so future work could explore whether longer polypeptoids that collapse into globular structures would recover large-lengthscale hydrophobic responses.

The effect of conformation on water dynamics is smaller than that of sequence for disordered polypeptoids, as the reduction in dynamics going from an extended NME homopolymer to an extended alkyl homopolymer is larger than the reduction from an extended NME homopolymer to a compact NME homopolymer (Fig. 5.6). This finding does not necessarily contradict that of Persson *et al.*’s analysis for globular proteins²⁷⁸ as larger, structured molecules introduce the possibility of buried water molecules, leading to significant confinement effects. In the types of small and disordered systems here,

confinement effects and buried water molecules are unlikely to occur; for instance, Fig. 5.3c shows that the dipole reorientation decays to below 0.1 within 50 ps, signaling no significant contribution from exceptionally slow water molecules near the peptoid.

5.4.3 Water dynamics and tetrahedrality are connected

Fig. 5.8 reveals significant relationships between the structural and dynamical metrics characterizing water behavior near the peptoid. In particular, we find strong correlations between both the dynamical metrics and the enhancement in tetrahedrality in the peptoid’s hydration waters, with $R^2 = 0.96$ for translational and $R^2 = 0.94$ for orientational dynamics. This finding is consistent with correlations between diffusivity and other collective measures of water structuring, such as configurational or excess entropy for bulk and hydration water,^{61,254–257,259} as well as previous work specifically relating water dynamics to tetrahedrality. Galamba observed joint increases in tetrahedrality and reductions in dynamics near small-molecule hydrophobes, though they did not quantify a correlation.²⁵⁸ Dahanayake *et al.* showed a correlation between diffusivity and tetrahedral order over different sites at the surface of an enzyme.²⁵⁹ This present work shows that a similar correlation also applies to shifts in water behavior at disordered molecules due to changes in molecular chemistry. In contrast, the local density, quantified by N_{water} , shows weaker correlations with translational ($R^2 = 0.86$) and rotational ($R^2 = 0.86$) motion. Thus, the enhancement in tetrahedrality, a metric that captures collective, network-like structuring in water, is more closely related to the reduction in water dynamics compared

to the reduction in local density, a single-body metric.

Based on these findings, we hypothesize that the strong correlation between water dynamics and structural rearrangement with enhanced hydrogen-bonding networks can be used to infer shifts in collective water structure from experimental measurements of local dynamics. This ability will be important to future studies experimentally probing shifts in water behavior near molecules that are more difficult to interrogate computationally or with other experimental techniques, allowing us to reach a deeper understanding of water behavior in realistic systems and to relate such findings to simulation and theory.

5.5 Conclusions

In this chapter, we establish that polypeptoids perturb their hydration water from bulk behavior and that polypeptoid sequence affects the magnitude of these perturbations. Specifically, we find that a peptoid oligomer comprised of polar residues makes its hydration water less dense, more tetrahedral, and slower, both translationally and orientationally, compared to bulk and that these effects become stronger with increasing number of nonpolar residues. Furthermore, the contribution from nonpolar residues in this small-molecule system is non-additive and dependent on the sequence, i.e. the context of the inserted nonpolar residues. Specifically, nonpolar residues enhance water dynamics near adjacent polar residues, and vice versa, confirming that residue hydrophobicity in polypeptoids is context-dependent as has been observed for protein surfaces.²⁸⁰ Comparison of subsets of extended and compact conformations for the homopolymers shows that

the observed effect of sequence is qualitatively independent of changes in the dynamic conformations of disordered polypeptoids.

The overall effect of nonpolar groups in these systems is to reduce water’s dynamics. These findings provide insight into the effective polypeptoid lengthscale to which water responds, which appears more localized than the entire molecule. Thus, this polypeptoid system does not behave like extended surfaces, while it also shows non-additive behavior more complex than small-lengthscale hydrophobicity, inviting future investigation into both longer polypeptoids and more diverse polypeptoid chemistries and sequences to determine the generalizability of this picture of polypeptoid hydration water response.

We have also demonstrated that a collective measure of water structure—specifically, of tetrahedrality—is strongly correlated with the non-additive responses observed for polypeptoid hydration water dynamics. This suggests that the dynamical response is linked in origin to collective structural rearrangements of water, and that experimental techniques probing water dynamics may be an effective approach to report on collective structural behavior of hydration water, enabling further investigation into structure-dynamics-thermodynamics correlations of water hydrating conformationally dynamic systems. An understanding of how polypeptoids mediate water response allows for the tuning of water-mediated interactions, which is crucial for designing both self-assembling polypeptoid systems as well as polypeptoids that modulate assemblies of other materials. For instance, other groups have proposed mechanisms by which peptoid sequence control of water structure at a surface can tune crystallization processes.^{166,169} Thus,

a systematic study of the effect of the chemistry of surface-tethered peptoids on water behavior, expanding on the single silica-tethered peptoid system studied here, is the subject of the next chapter. More broadly, the design of novel chemistries and materials for applications aiming to modulate water and solute responses will rely on the elucidation of the effect of heterogeneous chemical functionalities on the water-mediated interaction of the surface with solutes.

Acknowledgments

This work was supported by the Center for Materials for Water and Energy Systems (M-WET), an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Basic Energy Sciences under Award DE-SC0019272. Use was made of computational facilities purchased with funds from the National Science Foundation (OAC-1925717) and administered by the Center for Scientific Computing (CSC). The CSC is supported by the California NanoSystems Institute and the Materials Research Science and Engineering Center (MRSEC; NSF DMR 1720256) at UC Santa Barbara. We acknowledge use of MRL Shared Experimental Facilities which are also supported by the MRSEC Program of the NSF under Award No. DMR 1720256; a member of the NSF-funded Materials Research Facilities Network (www.mrfn.org). Support for the ODNP studies was provided by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy—EXC-2033—Project number 390677874. S.J. acknowledges support by the National Science Foundation Graduate

Research Fellowship (DGE 1650114). D.R.M. acknowledges support from the MRSEC Program of the National Science Foundation under Award No. DMR 1720256. A.D. acknowledges support by the National Defense Science & Engineering Graduate (NDSEG) Fellowship Program.

Chapter 6

Relationships between water's structure and solute affinity at polypeptoid brush surfaces

6.1 Introduction

In the previous chapter, we determined that peptoid sequence can tune its hydration behavior in dilute solution. In this chapter, we show that peptoids tethered to a surface similarly affect the surface's hydration behavior, and furthermore that the changes in hydration behavior have consequences for the affinities of small-molecule solutes for the surface. Such effects are meaningful because a key desirable characteristic of an aqueous interface is the ability to resist unwanted solute physisorption. Much of the work in de-

signing such antifouling surfaces for applications ranging from water filtration membranes to medical devices is currently based on heuristics: studies suggest that to minimize affinity for solutes such as proteins and colloids, surfaces should be smooth, hydrophilic, and non-hydrogen-bonding,^{44,45} although these relationships are highly system-dependent.²⁸¹ Poly(ethylene oxide)^{282,283} and zwitterionic surfaces^{284–286} have been reported as particularly promising for antifouling. These chemistries are hypothesized to repel solutes *via* their effect on water structure and thermodynamics by creating a “tightly bound” layer of hydration water.^{28,287–290} However, a nanoscopic understanding of the structure of this water layer’s properties and scope, and how it is influenced by the surface chemistry, remains lacking. Understanding these relationships is key to designing antifouling interfaces and moving from heuristics to molecularly-informed strategies.

At the same time, a wide body of literature has developed a molecular picture of water and solutes at idealized surfaces. Importantly, this work has elucidated water’s distinct structuring around small (< 1 nm) versus large (> 1 nm) lengthscale hydrophobes and the consequences of this structuring for their solvation and aggregation properties.⁵⁶ Water has an even more complex response to surfaces and solutes with polar moieties, as prior work on systematically varied model systems has demonstrated.^{131,225,231,291,292} For instance, the arrangement of hydrophilic groups can significantly alter water’s structure, dynamics, and thermodynamics, and consequently can vary the affinity of solutes to such surfaces. As a result, simple additive models based solely on surface chemical composition cannot capture solute affinity for such chemically heterogeneous surfaces.

Instead, models that account for these spatial correlations in surface and solute chemical structure^{152,293,294} are required. Realistic surfaces further complicate this picture, as they contain both polar moieties and some degree of topological heterogeneity, leading to both small and large-lengthscale responses. Prior work has sought to understand and model how both surface chemistry and topology affect hydration water behavior, solvation thermodynamics, and surface-solute interactions.^{51,55,61,79,141,144,145,148,295–299}

One roadblock to elucidating the relationship between water behavior and solute adsorption for a realistic surface is the difficulty of characterizing water’s structure and dynamics at a molecular level. Sum frequency generation (SFG) vibrational spectroscopy is hypothesized to distinguish between strongly and weakly hydrogen-bonded water molecules at interfaces³⁰⁰ and has been used to characterize water hydrogen bonding at antifouling materials and changes in surface hydration upon adsorption of proteins or other solutes.^{301–303} These SFG studies have found that water at antifouling zwitterionic and poly (ethylene glycol) (PEG) surfaces is “strongly hydrogen-bonded,” supporting the hypothesis that antifouling surfaces have “tightly bound” hydration layers. However, definitive assignment of SFG features to specific water structures remains an open problem.³⁰⁴

Simulations can provide molecular-level mechanistic insight relating reduced solute adsorption to hydration water behavior at a variety of materials. For instance, Jiang and coworkers found that interfacial water is responsible for the repulsive force between model proteins and antifouling brush surfaces such as PEG, zwitterionic phosphorylcholine, and

zwitterionic poly-trimethylamine N-oxide (pTMAO).^{305–308} Similarly, Leng and coworkers found that hydration water repels solutes for the cases of alginate binding to both short zwitterionic sulfobetaine brush surfaces and PEG-decorated polyamide surfaces.^{309,310} Other efforts have found relationships between properties of a surface’s hydration water (e.g. hydrogen bonding energy, orientational structure) and affinity of smaller molecule solutes such as alkanes³¹¹ and surfactants.³¹² Conversely, some studies have focused on the relationship between the solute’s hydration shell structure and surface affinity, finding that small ionic and neutral solutes that have a larger hydration shell with more attractive water-solute energetic interactions show lower sorption in a polyamide membrane.³¹³ However, it is less clear how the interplay between surface and solute hydration water affects solute affinity for the surface.

To systematically study the effect of chemistry on surface hydration water and consequent solute affinity for realistic antifouling chemistries, we use the previously experimentally-validated simulation models and methods²¹² to examine water behavior near a series of chemically varying polypeptoid brush surfaces. Polypeptoids are highly tunable and thus provide an attractive platform for studying effects of surface chemistry in a controllable manner. Prior computational studies of water behavior near peptoids in dilute solution^{209,245} and comprising a brush surface²⁴⁴ found that peptoid chemistry significantly affects both water structure and dynamics. Furthermore, both uncharged and zwitterionic polypeptoid brushes demonstrate excellent protein antifouling properties,^{177,314} and studies have related these properties to features of the SFG spectra, suggesting more

strongly hydrogen bonded water.^{182,315} Thus, polypeptoid surfaces represent an excellent model system to investigate the relationships between water structure and solute affinity.

Here, we examine a series of surfaces with polypeptoid chemistries ranging from non-polar to polar (uncharged) to zwitterionic and compute the affinities of small-molecule solutes with representative chemistries similar to those found in natural organic matter (NOM) foulants^{316,317} along with other common contaminants.⁹² Through a decomposition of the free energy of binding, we show that an entropic restructuring term drives reduced affinity of these solutes for the zwitterionic surfaces. We examine several structural metrics that characterize water in the hydration shell and show that water’s triplet angle distribution in particular reveals significant variations between the uncharged and zwitterionic surfaces. Finally, we demonstrate how shifts in water’s tetrahedral structure near the surface in the presence of a solute reflect the entropic restructuring contribution. Collectively, these results offer a new perspective for understanding the structural and thermodynamic features of surface hydration water, distinct from water-surface binding strengths, that underpin solute-surface affinity and fouling.

6.2 Methods

6.2.1 Peptoid and poly (ethylene oxide) surface construction

We create a peptoid brush interface by tethering short peptoids (6 monomers long) to a self-assembled monolayer (SAM) comprising methyl-terminated chains (Fig. 6.1a). We

use the same SAM parameters and geometry as in Ref. [79] such that the surface has dimensions of $2.98 \text{ nm} \times 3.44 \text{ nm}$. The positions of the sulfur atoms in the SAM chains are kept fixed during the simulation. The peptoids are capped by a hydrogen (N-terminus) and amino group (C-terminus) and are modeled using the MFTOID forcefield.²⁰⁶ Our previous work has shown that MFTOID accurately reproduces peptoid conformation experimentally measured by Double Electron Electron Resonance.²¹² We model water using the OPC4 model^{247,248} and solutes following Ref. [79] except for boric acid, for which we use the model developed in Ref. [122].

We simulate five different surfaces (Fig. 6.1c), each with a homogeneous brush composition of the same peptoid sequence. Three surfaces are “uncharged,” where each peptoid is a homopolymer of net-neutral monomers. The “uncharged” sidechains include a “hydrophobic” n-butyl group and “hydrophilic” methoxy-ethyl and n-propanol groups (Fig. 6.1b). Two of the surfaces are zwitterionic, where each peptoid comprises four methoxy-ethyl residues, one positively charged residue (a lysine analog) and one negatively charged residue (a glutamic acid analog, Fig. 6.1b). In the “separated-charge zwitterionic” peptoid, the anionic residue is at the C-terminus (closest to the peptoid-water interface) while the cationic residue is in the middle of the peptoid, separated by two methoxy-ethyl residues. In the “consecutive-charge zwitterionic” peptoid, the anionic residue is again at the C-terminus with the cationic one immediately preceding it. Parameters for the uncharged sidechains and capping groups follow Refs. [212] and [245], while those for the charged sidechains follow Ref. [244]. We tether the peptoids at a

surface density of 1.17 nm^{-2} , or 12 peptoids per SAM surface (details in section E.1.2). For better statistics, we tether peptoids to both sides of the SAM surface as in Fig. 6.1a.

We also address the long timescale amide-bond isomerization that peptoids can undergo. Accurately sampling isomerization of all 120 (5 bonds per peptoid $\times$ 12 peptoids per surface $\times$ 2 surfaces) amide bonds in the simulation is intractable. Instead, we simulate a surface with a single tethered peptoid using the expanded-ensemble methodology developed in Ref. [212] to determine the percentage of cis and trans states for each amide bond in a given sequence (details in section E.1.1). For each peptoid brush simulation, we then independently and randomly assign cis and trans states to each peptoid amide bond, weighted by the computed equilibrium populations (details in section E.1.2).

As a comparative antifouling surface mimicking PEO brush chemistries, a poly (ethylene oxide) (PEO) brush surface is also generated by grafting methyl-terminated ethylene oxide (EO) 6-mers on each side of a grid of sulfur atoms with Au-hcp spacing as described in Ref. [318] such that the surface has dimensions $3.90 \text{ nm} \times 4.00 \text{ nm}$. The EO parameters developed in Ref. [192] are applied. As with the peptoid systems, the sulfur atom positions are fixed during simulation.

6.2.2 Surface simulations and analysis

We perform simulations using the OpenMM simulation engine.⁷¹ Electrostatic interactions are computed using particle-mesh Ewald. Lennard-Jones nonbonded interactions are cut off at $10\text{\AA}$. Hydrogen bonds are constrained with SHAKE¹²³ and water is kept

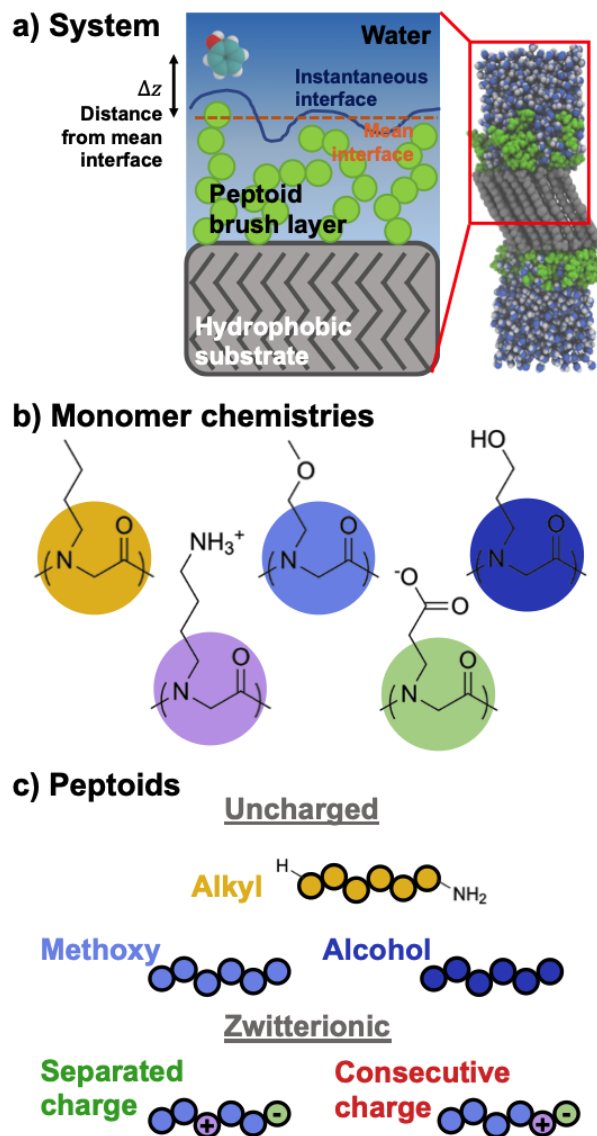


Figure 6.1: (a) Geometry of the peptoid brush surface system comprising the hydrophobic substrate (methyl-terminated self-assembled monolayer), peptoids tethered to it, solvation and reservoir water, and a solute. The wavy, blue line depicts the instantaneous peptoid-water interface (details given in section E.1.2), and the orange line depicts the mean over all snapshots of that interface in the direction perpendicular to the surface. The distance from the mean interface to the surface is denoted Δz . The snapshot on the right was created in Visual Molecular Dynamics, version 1.9.4.¹⁵⁵ (b) Monomer chemistries for the (c) peptoid sequences studied here. The sequences are depicted from the N- (left) to C-terminus (right), where the former is tethered to the substrate.

rigid with SETTLE.¹²⁴ We run all simulations using Langevin dynamics with an integration timestep of 2 fs, a temperature of 300 K and a friction coefficient of 0.1 ps^{-1} . For simulations in the NPT ensemble, we set the pressure to 1 bar using a Monte Carlo barostat with 200 fs between MC barostat moves. Simulations without a solute are run for $1 \mu\text{s}$ in the NPT ensemble, after equilibration (details in section E.1.2). For simulations with a solute, we conduct umbrella-sampling simulations to sample all solute-surface distances, simulating for 20 ns at each distance. We harmonically restrain the distance perpendicular to the surface between the centroid of the solute’s heavy atoms and the centroid of the SAM sulfur atoms (details in section E.1.2). We use MBAR^{125,218} to compute the potential of mean force in the unrestrained state. Uncertainties are computed from the standard error from independent repeats, as described further in section E.1.2.

To determine the instantaneous peptoid-water interface for a given snapshot, we follow the procedure of Willard and Chandler (details in section E.1.2).⁶⁹ We identify hydration waters as those with oxygen atoms within $7 \text{ \AA}$ above (in the direction towards the bulk water) any point of the instantaneous interface. The mean interface is the average position, in the z -direction (perpendicular to the surface) of all the points in the instantaneous interface, over all snapshots, and is reasonably constant over the different peptoid chemistries (see section E.1.2 for more details).

The PEO surface is simulated using similar procedures with minor differences described in more detail in section E.1.3.

6.3 Results and Discussion

6.3.1 Solute affinities to surfaces

We first determine whether the zwitterionic peptoid surfaces show enhanced antifouling properties compared to the uncharged surfaces. To quantify solute-surface affinity, we compute potentials of mean force (PMFs) or free energies $G(z)$ between each peptoid surface and a variety of solutes of varying polarity and hydrogen-bonding capabilities: methane, methanol, isopropanol, phenol, benzene, boric acid, and capped glycine. Fig. 6.2 shows PMFs for exemplary polar (isopropanol) and nonpolar (benzene) solutes, while additional solutes are in section E.3.1. As exemplified in these calculations, the PMF for all solutes and surfaces asymptotes to zero around $10\text{\AA}$, while deep into the surface ($\Delta z < 0\text{\AA}$), the free energies increase sharply as solutes are less able to penetrate the increasing peptoid brush density, with a corresponding increase in uncertainty in the PMF. At intermediate distances for both isopropanol and benzene, the PMFs with uncharged surfaces display a local minimum at around $\Delta z = 4\text{\AA}$, just above the peptoid-water interface, indicating favorable surface-solute interactions and preferable solute adsorption. In contrast, the PMFs with the zwitterionic surfaces show solute repulsion.

Benzene consistently shows more favorable interactions with all surfaces compared to isopropanol. One factor could be the loss of isopropanol-water hydrogen bonds upon binding. A decomposition of the PMF discussed in section 6.3.2 also suggests that benzene experiences a smaller entropic penalty associated with water restructuring upon

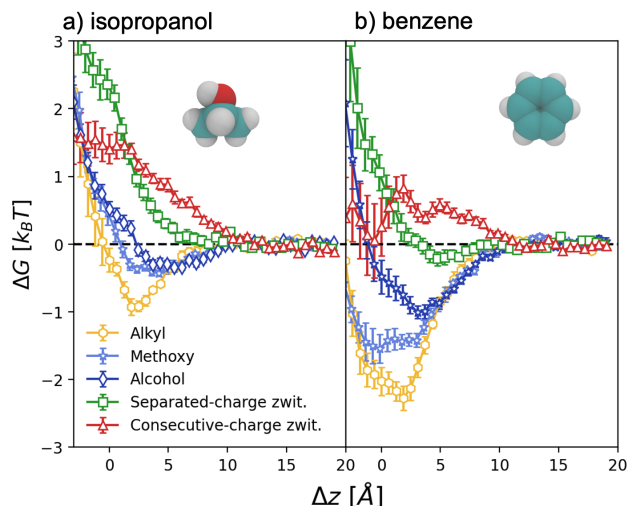


Figure 6.2: Potential of mean force (PMF), ΔG , between the peptoid surfaces and (a) isopropanol or (b) benzene as a function of distance from the surface-water interface, Δz . Here, Δz is computed with respect to the mean location of the surface-water instantaneous interface (Table E.1), as described in section 6.2.2. The PMFs are shifted such that they reach 0 far away from the surface. The error bars give the standard error computed from 10 independent simulations, as described in section E.1.2.

binding to the uncharged surfaces (Fig. E.11). However, the qualitative trends with respect to surface chemistry for the two solutes are the same: the wells in the PMF for the alkyl surface are consistently the deepest, suggesting that both nonpolar and polar solutes have the highest affinity for the most hydrophobic surface, agreeing with earlier calculations for small solutes near more idealized surfaces.⁷⁹ The methoxy surface is the second most attractive, followed by the alcohol one, which shows the shallowest, though still extant, well. Finally, both the separated-charge and consecutive-charge zwitterionic surfaces uniquely repel the solutes, with the consecutive-charge surface showing stronger repulsion.

To quantify solute affinity, we take the minimum value of the PMF within the $5\text{\AA}$ window above the peptoid-water interface, $0\text{\AA} < \Delta z < 5\text{\AA}$, for each combination of

solute and surface (Fig. 6.3), including for the zwitterionic surfaces, where the PMFs are largely positive in this region. Notably, in Ref. [79], the same solutes consistently have higher affinities for a hydrophobic methyl-terminated self-assembled monolayer (SAM) compared to a hydrophilic hydroxyl-terminated SAM. As in that work, all solutes here exhibit favorable affinity for the uncharged surfaces, and only with the addition of charged chemistries do any of the surfaces display solute repulsion. Also, like the uncharged SAM surfaces, the uncharged peptoid surfaces generally have the highest affinity for phenol, followed by benzene, and have the lowest affinity for methanol. Thus, despite the more complex topological landscape presented by the peptoid surfaces compared to much smoother and conformationally rigid SAM surfaces, the basic trends with respect to surface and solute chemistries hold.

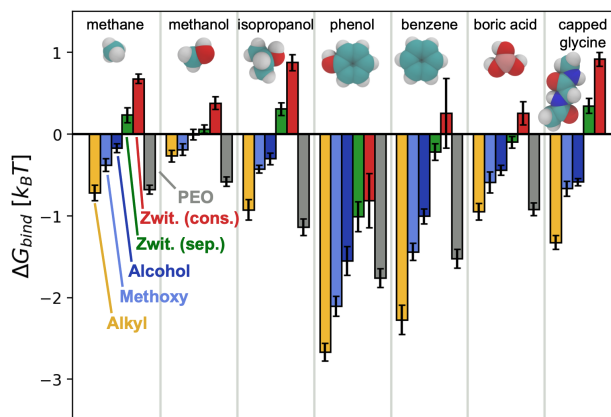


Figure 6.3: Affinities of a variety of solutes for the peptoid and PEO surfaces, computed from the minima in the potential of mean force within $5\text{\AA}$ above the surface-water interface. The error bars give the standard error computed from independent simulations, as described in section E.1.2.

To put affinities in the context of typical antifouling surface chemistries, a similar calculation is performed for a poly (ethylene oxide) (PEO) brush surface, a system more

widely studied and employed as an antifouling coating.^{28,44,289,301,319} Gray bars in Fig. 6.3 show the affinities of each solute for the PEO surface. Protein fouling measurements by Ostuni *et al.*⁴⁴ suggest comparable antifouling capacity for methoxy-terminated (PEO) and hydroxyl-terminated poly (ethylene glycol) (PEG) SAMs. We leave it to future efforts to determine the effect of terminal chemistry on the solute-surface affinities of the small solutes considered in the present study.

Interestingly, the affinities of all solutes to the PEO surface are similar to those for the uncharged peptoid cases, suggesting that uncharged peptoid brush surfaces may have similar antifouling properties as PEO-based brush surfaces. While the overall trend with respect to solute chemistry is also similar for the PEO and peptoid surfaces (e.g. among the solutes, phenol shows the highest affinity while methanol the lowest), the quantitative trends in affinities differ slightly. For some solutes, the PEO surface affinity is the closest to that of the alkyl peptoid surface (e.g. methane, isopropanol, boric acid), while for others, the affinity is closer to that of the methoxy peptoid case (e.g. benzene). While the chemical functionality of the PEO surface is similar to the methoxy (ether groups) and alkyl (terminal methyl groups) peptoid sidechain chemistries, other differences in surface geometry could lead to these differences in the trends of solute affinity. The PEO brush has a higher surface density compared to the peptoid brushes (4.62 chains/nm² versus 1.17 chains/nm², an approximately 4-fold difference). The peptoid chains adopt more diverse conformations, resulting in approximately equal average number density of the surface's heavy atoms (Fig. E.3 and E.7). However, the relative flexibility of the

peptoid chains leads to a rougher interface. The width of the instantaneous interface is approximately $4\text{\AA}$ for peptoids versus $1\text{\AA}$ for PEO (Fig. E.2) and water density decreases more sharply for PEO: the transition from bulk density to half of bulk density is over approximately $8\text{\AA}$ for peptoids (Fig. E.6) versus $4\text{\AA}$ for PEO (Fig. E.8). The increased roughness of the peptoid surface could induce a more complex dependence on solute size and chemistry compared to the PEO surface.

The alcohol and zwitterionic peptoid surfaces exhibit enhanced solute repulsion – in other words, weaker solute affinities – compared to the model PEO brush surface. While previous work has suggested that surfaces with hydrogen-bond-donating groups may lead to poorer fouling resistance, these studies were conducted with much larger solutes such as proteins⁴⁴ and algae³¹⁵ that can have numerous hydrogen bonding interactions with the surface. The smaller molecule solutes studied here would have much weaker interactions with surface hydrogen-bond donors. The enhancement in solute repulsion for the zwitterionic cases is consistent with prior experimental³²⁰ and computational studies³²¹ showing stronger protein adsorption at PEG surfaces compared to zwitterionic ones. Reduced affinity near the consecutive-charge zwitterionic surface, compared to the separated-charge case is also consistent with experimental work showing that an increase in charge spacing correlates with an increase in protein adsorption.^{322,323} On the other hand, the observed effect of charge spacing contrasts the work of Lau and coworkers showing that the separation distance between these same peptoid sidechain chemistries does not significantly affect protein adsorption;¹⁷⁷ we note that the longer

(20-mer) peptoid chains in that work may have enabled greater peptoid flexibility, such that charges on distinct chains could rearrange to be in close proximity, masking the effect of intrachain charge spacing. Indeed a computational study of similar peptoid brushes showed significant chain mobility and charge complexation.²⁴⁴

6.3.2 Decomposition of solute free energy of binding

We also find that thermodynamic contributions to surface binding affinities reveal important contributions from and distinct signatures of water restructuring. Our prior work showed that enhanced solute affinity for nonpolar SAM surfaces is due to a reduction in water restructuring entropic penalties when the solute is at the interface versus in the bulk.⁷⁹ Specifically, we rigorously decompose the PMF (Ref. [62]’s Supplementary Material contains the complete derivation) into (1) a direct solute-surface energetic interaction and (2) a relative entropy capturing water restructuring^{62,79}:

$$G = \langle U_{sw} \rangle + TS_{rel} \quad (6.1)$$

where G is the free energy, $\langle U_{sw} \rangle$ is the average energetic interaction between the solute and the rest of the system, and S_{rel} is the relative entropy between the states with and without the solute; all of these terms are implicitly dependent on the distance from the surface, Δz . Here, the entropic term can be shown to be strictly positive such that TS_{rel} always serves as an entropic penalty in this decomposition; this fact and the positive sign for that term are the consequence of its role as a relative entropy. Moreover, for rigid solutes and surfaces, S_{rel} only depends on variations in configurations of hydration layer

waters of the surface and solute and hence always measures the unfavorable restructuring of this water relative to bulk. The role of such restructuring in solute-surface affinity is revealed by calculating:

$$\Delta G = \Delta \langle U_{sw} \rangle + T\Delta S_{rel} \quad (6.2)$$

where the Δ 's report differences in bringing a solute from bulk to adjacent to a surface. For the earlier study on methylated SAM surfaces,⁷⁹ for example, the $T\Delta S_{rel}$ term was large and negative, offsetting the energetic term to result in overall favorable surface-solute affinity for a range of small solutes, thus highlighting the importance of water restructuring as a driving force for affinity. In the present study, the surfaces are not rigid, and thus the relative entropy term also includes the effects of peptoid surface restructuring. However, we estimate that this contribution is not significant as the configurational fluctuations within any peptoid surface vary little as the solute moves close it (Fig. E.14), suggesting that the distribution of surface conformations is not significantly different in the presence of the solute. Thus, we expect that the changes captured by $T\Delta S_{rel}$ can be ascribed to differences in the hydration water's structural rearrangement in the presence of a solute.

Fig. 6.4 shows that this thermodynamic decomposition for phenol reveals distinct driving forces for its affinity at the peptoid and PEO surfaces. For all peptoid cases, the solute's energetic interactions with its environment become more favorable (negative) as it moves closer to the surface. Even for the alkyl surface, this contribution remains slightly favorable, signaling that either the solute does not undergo complete desolvation

as it approaches the surface or, that this change is energetically offset by interactions with the surface itself. Interestingly, the favorability of the energetic contribution follows the opposite order of the affinities, and it is the relative entropy contribution that ultimately determines the ranking in the overall surface affinity. Specifically for the alcohol and zwitterionic surfaces, the relative entropy component becomes increasingly repulsive as the solute moves closer to the surface, suggesting that the surface and solute hydration water undergo an increasing degree of restructuring to accommodate the solute near the interface. In contrast, this contribution is negative for the alkyl, methoxy, and PEO surfaces, therefore enhancing the favorability of binding. This decomposition gives similar qualitative trends for the other small molecule solutes examined (Fig. E.11). Unlike the peptoid surfaces, phenol sorption at the PEO surface is energetically unfavorable, suggesting weaker surface-solute energetics; this could result from the solute’s weaker interactions with the outward facing methyl terminal groups of the PEO chains.

6.3.3 Structure of surface hydration water

The results above show that the model peptoid surfaces capture trends in antifouling behavior observed in prior work on model membrane surfaces—zwitterionic peptoids enhance solute repulsion compared to uncharged PEG/PEO-like (alcohol, methoxy, alkyl) peptoids. We therefore expect this peptoid platform to provide insights into relationships between antifouling behavior and hydration layer properties, and to identify molecular features that distinguish water behavior near zwitterionic surfaces from water behavior

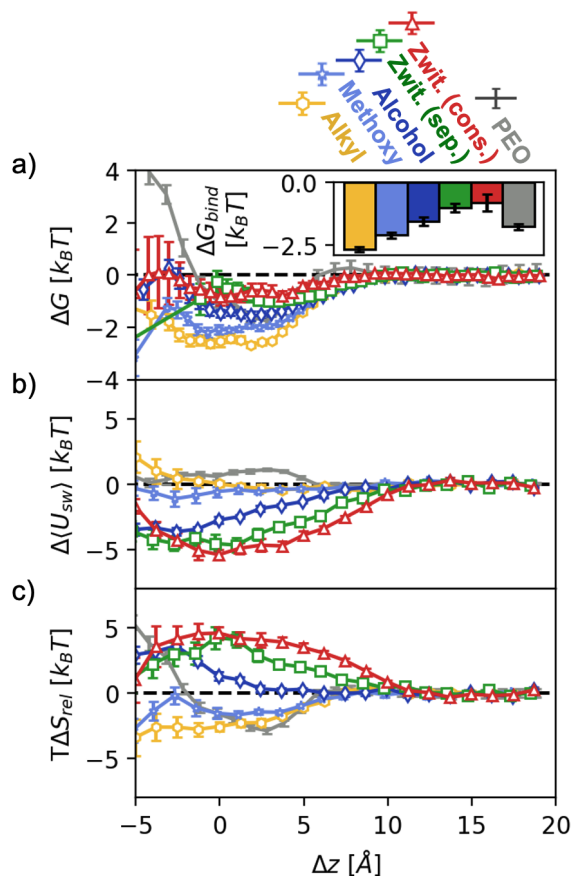


Figure 6.4: Decomposition of the (a) potential of mean force, ΔG , between the solute phenol and the peptoid and PEO surfaces (legend at top) into the contributions from (b) the energetic interaction between phenol and the system (surface and water) and (c) the relative entropy describing restructuring of the system to accommodate the phenol solute. The error bars give the standard error computed from 10 independent simulations, as described in section E.1.2.

near uncharged surfaces. Much discussion has hypothesized a tightly bound hydration water layer as a mechanism for antifouling. At the molecular level, this has been shown to be related to reduced water dynamics.³²¹ Consistent with earlier work, here the translational and orientational dynamics of water near the zwitterionic peptoid surfaces are slower compared to the uncharged surfaces. Table 6.1 shows the diffusivity of water near the surface (E.2.1 contains details of the calculation and diffusivities as a function of distance from the surface), which is up to 25% lower approaching zwitterionic surfaces. Fig.

E.5 shows the O-H bond orientational correlation function for waters in the hydration layer, which decays up to 30% slower for zwitterionic surfaces. However, slow water dynamics does not necessarily provide insight into the thermodynamics of solute solvation. Instead, characterization of water structure at these interfaces is necessary to develop a mechanistic understanding of the relationship between water behavior and solute affinity, and in particular the entropic restructuring contribution that drives changes in the free energy of binding.

Surface	D [$\text{\AA}^2/\text{ps}$]	ρ [$\text{\AA}^{-3}$]	$N_{HB,w-w}$	$N_{HB,w-s}$	N_{HB}
Alkyl	0.158 ₂	0.0173 ₃	3.433 ₅	0.116 ₅	3.549 ₂
Methoxy	0.157 ₂	0.0182 ₃	3.375 ₈	0.152 ₇	3.527 ₁
Alcohol	0.150 ₂	0.0185 ₄	3.27 ₁	0.28 ₁	3.548 ₃
Separated-charge zwitterionic	0.135 ₁	0.0182 ₆	3.20 ₂	0.30 ₁	3.498 ₆
Consecutive-charge zwitterionic	0.118 ₂	0.0183 ₄	3.08 ₁	0.399 ₉	3.475 ₄

Table 6.1: Hydration properties of the peptoid surfaces: diffusivity (D) at $\Delta z = 5\text{\AA}$ from the water-surface interface, average water density (ρ), and average number of hydrogen bonds formed by waters in the hydration layer (N_{HB}) with other waters ($N_{HB,w-w}$) and the surface ($N_{HB,w-s}$). The subscript gives the standard error computed from 5 independent simulations as described in section E.1.2.

One of the simplest and long-studied measures of water structure is its density. Could enhanced density be indicative of “tightly bound” water? Table 6.1 shows the average density in the region $10\text{\AA}$ from the mean interface. The zwitterionic surfaces do not show a pronounced increase in average hydration water density relative to the methoxy surface. However, water is less dense near the alkyl surface compared to the methoxy one, consistent with the behavior around peptoids in dilute solution,²⁴⁵ while water around the alcohol surface is slightly more dense. The spatial density profiles (Fig. E.6) show more significant chemistry-dependent differences, with stronger oscillations for the

zwitterionic and alcohol surfaces compared to the others. This structuring is correlated with oscillations in the peptoid density profile (Fig. E.7), suggesting that solvation of peptoids drives changes in local water density. However, no clear distinction between the zwitterionic and alcohol surfaces emerges. Other local measures of water density, such as peptoid-water coordination numbers (Table E.2), are similarly ambiguous. These initial results suggest that density, both average and local, is not related to the unique features of hydration water at antifouling surfaces.

Another metric that could suggest a “tightly bound” hydration layer is hydrogen bonding in the hydration layer. Table 6.1 shows that the average number of hydrogen bonds formed by hydration waters with the peptoid surface (calculation details in section E.2.3) increases with the surface’s ostensible hydrophilicity, from the alkyl to methoxy to alcohol to zwitterionic surfaces. These trends are generally consistent with prior simulation studies showing more water-surface hydrogen bonding²⁴⁴ near zwitterionic surfaces compared to uncharged ones. There is a further slight enhancement in surface-water hydrogen bonding in the consecutive-charge zwitterionic surface compared to the separated-charge one, likely due to the enhanced accessibility of both charged groups when they are co-located at the hydrated end of the peptoid. The enhancement in surface-water hydrogen bonds is offset by a reduction in water-water ones, so that the total number of hydrogen bonds formed by a hydration water decreases from the alkyl to the alcohol to the zwitterionic surfaces. These calculations reveal that the zwitterionic surfaces do not enhance the total density of water’s hydrogen bond network.

The density and hydrogen bonding of water near these surfaces, while showing some small variations, does not reveal significant, differential effects of the surface chemistries, and specifically, does not show any qualitatively unique features of the zwitterionic cases. To provide an alternate perspective, we characterize how the surfaces perturb water's structural orientational correlations. Specifically, we investigate the water triplet or three-body angle distribution $P(\theta)$ that has previously been shown to correlate with measures of water dynamics^{63,245} and hydration free energy.^{62,64} This distribution characterizes the populations of angles θ between all triplets of neighboring water molecules (angles measured between oxygen centers), and can show shifts in orientational structure in several key regimes: a tetrahedral population near $\theta = 109.5^\circ$, an icosahedral population typical of simple liquids around $\theta = 63.4^\circ$, and an ideal gas like population centered near $\theta = 90^\circ$.^{62,63} The tetrahedral and icosahedral populations are hypothesized to reflect changes in small-lengthscale hydrophobic responses, as water around small solutes has been shown to possess enhanced tetrahedrality and reduced icosahedrality.⁶³ On the other hand, the ideal gas population reflects changes in large-lengthscale hydrophobic responses, as water around a large hydrophobic surface has been shown to be closer to its vapor-liquid equilibrium and therefore exhibits larger density fluctuations, spontaneously forming larger voids.^{324,325} These small and large lengthscale behaviors correspond to qualitative changes in the thermodynamics of solvation.⁵⁶ Thus, variations in the three-body angle distribution can provide key structural insight into water's thermodynamic behavior.

Here, to show how surfaces perturb this distribution, we compute it for all hydration waters (within $7\text{\AA}$ above the instantaneous peptoid-water interface) and subtract off the “reference” triplet distribution for bulk water, as described in section E.2.4. Remarkably, the triplet angle distributions in Fig. 6.5a show significant, qualitative differences between the zwitterionic and uncharged surfaces, unlike the prior metrics. Compared to bulk water, water is more tetrahedral near the uncharged surfaces, as evidenced by population enhancement in the tetrahedral ($\theta = 109.5^\circ$) and reduction in the icosahedral ($\theta = 63.4^\circ$) regions. This behavior is consistent with water near peptoids in dilute solution as studied in Ref. [245], where it was argued that the enhancement in tetrahedrality signaled a small-lengthscale hydrophobic response. The similar enhancement seen here suggests that even near the peptoid brush surfaces, water retains a small-lengthscale response. Notably, however, for the alkyl surface, the enhanced tetrahedral peak is slightly shifted towards lower angles, near the disordered (ideal gas) region ($\theta = 90^\circ$); this behavior is not present near the same peptoid in dilute solution (Fig. E.9). As such, the response suggests the emergence of some large-lengthscale hydrophobic behavior due to the brush surface geometry, in concert with features of small-lengthscale hydrophobicity.

In contrast to the nonionic surfaces, the differential three-body-angle distributions near the zwitterionic interfaces exhibit either no enhancement in the tetrahedral peak (for the separated-charge surface) or a reduction in the tetrahedral peak (for the consecutive-charge surface). This suggests that water’s tetrahedral network is weaker at the zwitterionic compared to the uncharged surfaces, which is consistent with the formation of fewer

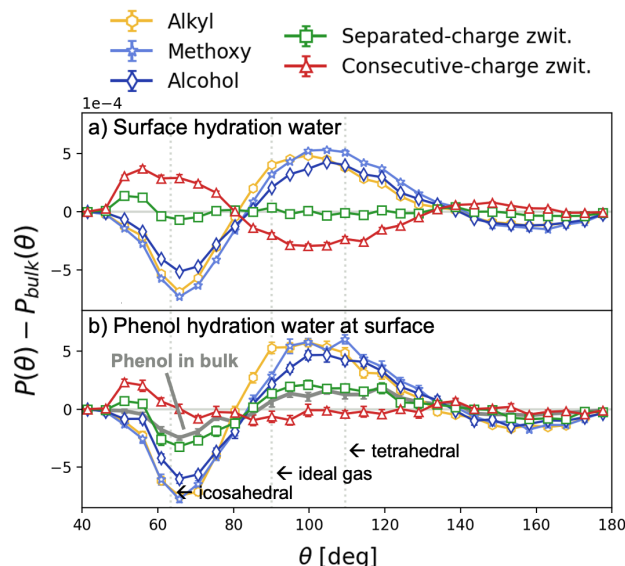


Figure 6.5: The three body angle distributions, $P(\theta)$ of the hydration water of (a) the surface and (b) phenol, both at the surface (colored) and far away from the surface (gray). All distributions are shown relative to bulk (i.e., the distribution for bulk water, $P_{bulk}(\theta)$, is subtracted). The dotted lines indicate populations of icosahedral ($\theta = 63.4^\circ$), ideal gas-like ($\theta = 90^\circ$), and tetrahedral ($\theta = 109.5^\circ$) water. The error bars give the standard error computed from (a) 5 and (b) 10 independent simulations, as described in section E.1.2.

water-water hydrogen bonds (Table 6.1). Instead, water near these surfaces exhibits a peak or shoulder around $\theta = 50^\circ$ that has been related to highly coordinated water (i.e., a fifth water in the first coordination shell),^{62,64} possibly leading to the disruption of the tetrahedral network. At the same time, water density in the hydration layer is not significantly different in the zwitterionic surfaces, suggesting that the distortion of the first coordination shell is not simply due to increased numbers of waters at the interface. The origin of these water structures that are distinct from both small and large-lengthscale responses is unclear and warrants additional systematic investigation into water's triplet angle distribution near zwitterionic molecules and surfaces that will be the subject of future work. However, we can conclude that this distribution clearly distinguishes be-

tween the zwitterionic and uncharged surfaces, showing opposing shifts from bulk water behavior for the two classes of surfaces.

The three-body angle distribution may also provide insight into the relevant shifts in water structure that affect the entropic restructuring contribution, $T\Delta S_{rel}$, to the binding free energy. The relative entropy, $\Delta S_{rel} = S_{rel,surface} - S_{rel,bulk}$, contains contributions from water’s structuring at the surface and in bulk (far away from the surface) with and without a solute. To quantify restructuring, we compute the three-body angle distribution in the solute hydration shell both near the surface and in bulk (details in S1DIII). Fig. 6.5b shows that far away from the surface (“bulk”), phenol induces a small-lengthscale response in its hydration water, leading to an enhanced tetrahedral peak and a depressed icosahedral peak. At the surface, phenol’s hydration water distribution shifts towards that of the surface’s hydration water. Near the uncharged surfaces, water around phenol becomes significantly more tetrahedral, with a slight shift towards the ideal gas peak for the alkyl surface. Near the separated-charge zwitterionic surface, phenol’s hydration water’s distribution is the same as phenol in bulk water except for a small peak in the highly-coordinated region (around $\theta = 50^\circ$), while near the consecutive-charge zwitterionic surface, phenol’s hydration water resembles bulk water except for a larger peak in the highly-coordinated region. The distribution of phenol’s hydration water is thus a combination of that around phenol far from the surface and the distribution around the surface without phenol present.

To quantify and simplify these shifts in the 3-body angle distributions, we compute

their projections onto the principal directions along which they change. Specifically, our prior work⁶³ examined shifts in triplet distributions for a wide range of aqueous mixtures across different temperatures, compositions, and chemistries, and using principal component analysis, found that the distributions only varied along three principal components as shown in Fig. 6.6d-6.6f. This work hypothesized that the first principal component, with peaks in the tetrahedral and icosahedral regions, captures changes in water structure around small-lengthscale hydrophobic solutes; for instance, the 3-body angle distribution around methane looks very similar to this component. The second principal component, with a peak in the disordered region, captures changes in water structure around large-lengthscale hydrophobic solutes, such as an idealized, smooth hydrophobic interface. Finally, the third principal component, which has a peak around $\theta = 50^\circ$, has been hypothesized to capture highly coordinated water structures (e.g., a fifth water in the first coordination shell).^{62,64} Earlier work has also shown that this low-angle peak is predictive of hydration free energies.⁶⁴

Variations in hydration water structure as captured by these three principal components naturally quantify structural correlations with distinct and well-defined physical interpretations: small-lengthscale, large-lengthscale, and over-coordinated water structures. Fig. 6.6 shows projections of the triplet distributions onto the (a) first, (b) second, and (c) third principal components for the hydration water of the surface (without phenol present) and phenol when it is near the surface. Projections for other solutes show similar qualitative trends and are shown in Fig. E.15. We first focus on the projections

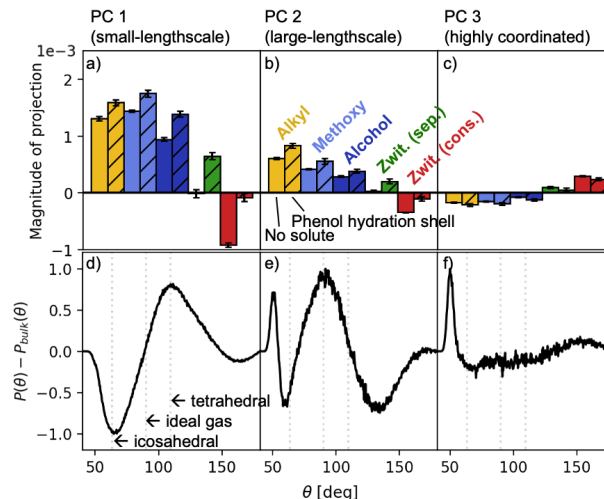


Figure 6.6: (a-c) Projections of the differential three-body angle distributions of water onto (d-f) the three principal components (PC 1, PC 2, PC 3) for hydration water of the surface (solid bars) and phenol (hatched bars) when it is close to the surface. The faint dotted lines in (d-f) indicate populations of icosahedral ($\theta = 63.4^\circ$), ideal gas-like ($\theta = 90^\circ$), and tetrahedral ($\theta = 109.5^\circ$) water. The error bars in (a-c) give the standard error computed from 10 independent simulations, as described in section E.1.2.

for systems without solutes (solid bars). For the uncharged surfaces, the largest shifts occur in the first principal component, showing that the behavior of water is primarily small-lengthscale as it approaches uncharged surfaces. The alkyl surface shows slightly more small-lengthscale behavior compared to the alcohol surface. Positive projections onto the second principal component suggest the emergence of some large-lengthscale behavior near the uncharged surfaces as well, with the alkyl surface the most prominent. Finally, the projection onto the third principal component is negative for the uncharged surfaces, suggesting a slight diminishment of highly coordinated water structures near them. Interestingly, the consecutive-charge zwitterionic surface shows opposite behaviors for projections onto all components, while the projection for the separated-charge zwitterionic surface is close to zero except for the third component. For these zwitterionic surfaces, the projection onto the third component is positive, suggesting a slight increase in highly coordinated water structures near them.

terionic surfaces, small-lengthscale and large-lengthscale behaviors are diminished and the positive projections onto the third principal component suggest higher populations of highly coordinated water. These highly coordinated waters may be related to the observed enhancement in surface-water hydrogen bonding (Table 6.1). Phenol’s hydration water exhibits similar trends (hatched bars), showing enhanced small-lengthscale behavior near the uncharged surfaces and depressed small-lengthscale behavior near the zwitterionic ones.

The difference in the magnitude of the projections for phenol near the surface and for the surface without phenol measures how much the hydration water at the surface must rearrange to accommodate phenol, and thus should roughly correlate with $S_{rel,surface}$. For the first principal component, this difference is the largest for the zwitterionic surfaces, and furthermore follows the trend in the contribution of $T\Delta S_{rel} = T(S_{rel,surface} - S_{rel,bulk})$. (We note that $S_{rel,bulk}$ is constant over varying surface chemistries.) In fact, for the zwitterionic surfaces, these principal components are predictive of $T\Delta S_{rel}$ (Fig. E.16) using a linear model (details in section E.4). Thus, the change in the triplet angle distribution that is due to the presence of the solute appears closely related to the restructuring contribution that drives the free energy of solute binding. One factor that contributes to water’s ease of restructuring may be the magnitude of the small-lengthscale component of the triplet angle distribution. We hypothesize that here the formation of more hydrogen bonds by water with the zwitterionic surfaces disrupts the water-water hydrogen bonds in the hydration layer, therefore discouraging the emergence of small-lengthscale water

structures. The hydration water of the zwitterionic surfaces therefore must restructure to accommodate small solutes, which encourage small-lengthscale behaviors in bulk. This restructuring penalizes surface-solute binding, leading to repulsion of small molecules from the zwitterionic surfaces.

6.4 Conclusions

The work in this chapter provides molecular level insight into small solute-surface affinity as a step towards understanding surface chemistries with antifouling properties. Our main finding is that the reduced solute-surface affinity shown by zwitterionic surfaces, compared to uncharged surfaces, is related to the rearrangement in water structure in the hydration layer as captured by the water triplet angle distribution. Molecular dynamics simulations of polypeptoid brush surfaces with varying sidechain chemistries show that the free energy of solute binding is more repulsive for zwitterionic peptoid surfaces compared to uncharged ones, where the latter show similar solute affinity compared to PEO brush surfaces, a more commonly used antifouling chemistry. The decomposition of the free energy of binding into direct energetic and entropic restructuring compositions shows that the penalty due to water restructuring is higher for the zwitterionic surfaces and is what determines the trends in solute binding. We then compare several order parameters for the surface’s hydration water to demonstrate that this restructuring is most clearly captured by the three-body angle distribution, which shows shifts from bulk behavior near the zwitterionic surfaces that are dramatically different from that near uncharged

surfaces. These zwitterionic surfaces disrupt water’s tetrahedral network and instead induce the formation of crowded coordination shells.

This work offers a molecular perspective on the “tightly bound” water that is hypothesized to be key for antifouling surfaces. We find that the water structure induced by the surface, and in particular the ease with which it restructures to accommodate a solute, determines solute affinity. However, importantly, such structures are dramatically different for zwitterionic and uncharged surfaces, suggesting that umbrella terms used to describe hydration layers of antifouling surfaces likely fail to capture this more nuanced picture, which requires further molecular level exploration. The triplet angle distribution of a surface’s hydration water could additionally provide insight into solute affinity in more complex systems with multiple solutes or added salt, where changes to surface hydration are hypothesized to modulate affinity.

The restructuring penalty observed here is independent of solute chemistry, therefore making it generally applicable to a wide range of small solutes, as demonstrated by the zwitterionic surfaces in this study. Design of surfaces that resist small-molecule fouling in aqueous systems could focus on strengthening water-surface hydrogen bonds and disrupting the naturally tetrahedral structure of the hydration water of small molecules. A general antifouling strategy may thus be to design surfaces to induce water structures distinct from the structure of a solute’s hydration water in bulk. Here, the functional groups are of comparable size to the solutes examined, such that the uncharged chemistries and small molecules both induce small-lengthscale water structure with enhanced tetrahe-

drality. This method may also be applicable to larger solutes, although solute-surface interactions may be more dominant in those cases and compete with the effect of water restructuring, as previously discussed. This strategy highlights the importance of nanoscopic characterization of surface and solute hydration water structures.

Chapter 7

Effect of polypeptoid sequence on conformation

Partially adapted with permission from *Macromolecules* 2021, 54, 11, 5011–5021. Copyright 2023 American Chemical Society.

7.1 Introduction

In the previous chapter, we demonstrated that changes in peptoid chemistries at a surface modulate the affinity of small-molecule solutes primarily by perturbing the surface’s hydration water’s tetrahedral structuring. This finding supports the idea that tuning local water behavior is key to designing antifouling surfaces. However, peptoids in solution can tune their hydration water behavior *via* both changes in sidechain chemistry or changes in conformation, as shown in Chapter 5. Experimental sum frequency gen-

eration and ambient pressure X-ray photoelectron spectroscopy measurements have also shown that polypeptoid chemistry can influence conformation at the surface.^{182,326} Thus, understanding the relationship between peptoid sequence and conformation is a key step towards rational design of peptoid-tethered surfaces.

The space of peptoid sequences is vast. Furthermore, the flexibility of a peptoid chain leads to potential interactions between sidechain chemistries that are far apart along the chain. These “long-range” interactions make it difficult to determine, *a priori*, a set of simplified sequence features that parameterize conformation. Work on aggregation of hydrophobic-hydrophilic sequences suggests that such low-dimensional parameters may exist, but are not trivial and that their identification relies on a large set of existing sequence-structure data.^{327,328} To navigate this large design space, we apply the same strategy we used in Chapters 2 and 3 in this dissertation and hypothesize that an inverse design approach, in which we pair simulations with an optimization routine to identify peptoid sequences that possess specified conformational behaviors, will help uncover sequence-conformation relationships.

The inverse problem of designing sequences to achieve specific conformations is generally a difficult one, and has been widely considered in the protein design community. One challenge is the forward problem of predicting the conformational ensemble of a given sequence. Detailed experimental characterization of folded structures is laborious,³²⁹ precluding large-scale studies. Faster but less detailed methods have been developed. For instance, Knight and coworkers used a dye that is selective for hydrophobic envi-

ronments to distinguish different peptoid conformations.³³⁰ This experimental method does not provide insight into specific conformations, as it can only detect gradations in the overall hydrophobic environment created by the chain, but the resulting sequences identified in a high-throughput screening can be simulated to gather additional conformational insight.³³⁰ Double Electron Electron Resonance (DEER) spectroscopy, previously discussed in Chapter 4, provides the distance distribution between two labeled positions in a chain,¹⁸³ offering quantitative, though partial, conformational information.

Predicting structure from sequence using computational models also poses challenges. Simulations of long chains are expensive and often require coarse-grained models to achieve the relevant timescales. Other work has fit surrogate models from more expensive simulations in order to rapidly explore design space. Models utilizing long short-term memory (LSTM) recurrent neural networks, which have also been used for language processing, have been successful in predicting chain shape metrics, such as radius of gyration.^{331,332} The success of LSTM in this present context may be due to their ability to capture associated longer-range sequence correlations that is also characteristic of interpreting language. AlphaFold2, a deep neural network model is impressively accurate at predicting protein folded structures,³³³ though work has suggested that it extrapolates poorly to non-evolved protein sequences³³⁴ as well as sequences with broader conformational ensembles.^{335–337} AlphaFold2, which was fit on a large database of sequence-structure data for proteins, along with related deep learning methods have leveraged decades of extensive work in protein structure resolution. Such databases are

much smaller for non-evolved sequences and intrinsically disordered proteins, limiting AlphaFold2’s accuracy in these areas. Furthermore, these databases are non-existent for other sequence-specific polymers, and will not emerge in the near future. Thus, AlphaFold2 may be limited in the insight it can offer for sequence-conformation relationships, and more physics-based approaches are needed.

Another challenge is how to pair the forward problem with an efficient optimization routine. Khokhlov and Khalatur proposed a workflow for sequences in which monomers are assigned hydrophobic or hydrophilic identities based on their location in a folded structure, followed by re-minimization of the structure, after which the process repeats.³³⁸ Application of this method to peptoids suggested that such designed sequences had more cooperative collapse behaviors compared to a sequence comprising short repeat units.³³⁹ The RosettaDesign algorithm similarly alternates between modification of the sequence in a fixed structure and optimization of the structure for a fixed sequence, and has been used to design a protein with a novel folded structure.³⁴⁰ The sequence optimization utilizes a Monte Carlo optimization in which replacement of an amino acid at a specific location is either accepted or rejected depending on the change in energy.³⁴¹ A similar approach instead used a genetic algorithm to identify new sequences.³⁴² More recent work has directly inverted neural network-based surrogate functions describing sequence-conformation relationships.^{343,344} However, such surrogate functions remain elusive for other sequence-specific polymers.

Because of the limited conformational data available for sequence-specific polymers

that are not peptides, sequence design of these molecules cannot rely on big-data-based strategies. Instead, accelerated physics-based models can be used to quickly predict conformational structure from sequence, and paired with sequence optimization methods to design novel sequences and learn sequence-structure relationships. In this chapter, we use polypeptoids to demonstrate such a workflow. We first demonstrate that the atomistic polypeptoid model captures sequence-dependent conformational behavior. We then show that a coarse-grained (CG) model with orders-of-magnitude enhancement in computational efficiency is able to reproduce those conformational behaviors. We pair the CG simulations with a genetic algorithm-surrogate model approach, similar to that used for design of nanopore chemical patternings in Chapter 2, in order to identify sequences that show extremal conformational behaviors.

7.2 Methods

7.2.1 Atomistic polypeptoid simulations

The atomistic simulations are run using the workflow outlined in Chapter 4 and Ref. [212]. Simulations of the polypeptoids with alternating hydrophilic groups are described in section 4.2 and details regarding box sizes and simulation lengths are provided in Table C.19. In this chapter, we also simulate polypeptoids with varying number of aromatic hydrophobic groups. These sequences are shown in Fig. 7.1. The chemical structure of the methoxyethyl, propanol, and TEMPO residues are given in Fig. 4.1 and the chemical

structure of the aromatic residue is given in Fig. 6.1. We use the MFTOID forcefield, with additional parameters for the alcohol and methoxy sidechains given in Ref. [212] and the TIP3P water model, modified for the CHARMM forcefield.²⁴⁶ Details regarding the simulation setup and equilibration times are given in Table F.1. After equilibration of the expanded ensemble weights, each polypeptoid is simulated for 10 μ s.

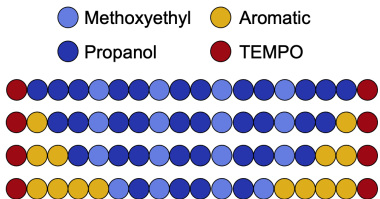


Figure 7.1: Peptoid sequences with varying number of aromatic groups.

7.2.2 Coarse-grained polypeptoid simulations

We develop a coarse-grained (CG) polypeptoid model using the relative entropy coarse graining approach developed and implemented by Shell and coworkers.^{345–349} The CG model comprises two beads for each monomer: one backbone bead (the same for all chemistries) and one sidechain bead (distinct for each chemistry). The model defines harmonic bond interactions between the backbone and sidechain beads of each individual monomer, as well as consecutive backbone beads. Angle and dihedral interactions are modeled as spline potentials. Non-bonded interactions are also modeled as spline potentials. Details of the polypeptoid CG model will be given in a subsequent publication. Briefly, we run atomistic simulations of short polypeptoids containing four types of sidechain chemistries: the methoxy, aromatic, and alcohol sidechains shown in Fig. 6.1

along with the alkyl chemistry shown in Fig. 5.1. We fit parameters for CG beads where each polypeptoid residue is represented by a backbone bead bonded to a sidechain bead. The backbone bead parameters are constant across different residue types.

Simulations of CG polypeptoids are run in the NVT ensemble in boxes of length 10 nm using Velocity Verlet integration with a timestep of 2 fs. The temperature is set to 300 K using a Langevin thermostat with a friction coefficient of 0.1 ps^{-1} . The system is equilibrated for 1 ns before saving configurations, which are then saved every 20 ps. For simulations run by the genetic algorithm, the production run is 100 ns. For all other simulations, the production run is 500 ns.

7.2.3 Genetic algorithm workflow

We follow the basic structure of the genetic algorithm outlined in Ref. [61]. The sequences are represented by binary arrays (0 = methoxy and 1 = aromatic) where the composition of the sequence is kept fixed at 6 aromatic monomers and 18 methoxy monomers. In each generation, 8 new sequences are generated and simulated using the workflow described in section 7.2.2. We run the algorithm with 5 generations where new sequences are generated randomly, followed by 55 generations where new sequences are generated from combining parents chosen from the most optimal of the previously examined sequences. We randomly generate sequences of fixed composition by randomly reordering the entire sequence. Parents are chosen through a tournament selection procedure⁶¹ and combined using an order crossover³⁵⁰ in order to preserve composition, after which the new se-

quences are mutated by swapping one pair of residues (equivalent to an 8% mutation rate).

For the first 20 generations, we compute the objective function by running a simulation of the CG model (section 7.2.2). We then fit a simple linear model that predicts the objective function (or simulation observables that are in the objective function) from features of the sequence. The features are:

1. The mean location of the aromatic groups, which measures the symmetry of the placement of the aromatic groups.
2. The square root of the second moment of the location of the aromatic groups about the center of the sequence, which measures the spread of the aromatic groups around the center.
3. The number of consecutive aromatic groups, which measures the degree of clustering of the aromatic groups.

We use this surrogate function for the next 30 generations, and then run 10 additional generations with the CG model.

7.3 Results and Discussion

7.3.1 Hydrophilic polypeptoids show excluded volume scaling

As a first step to developing sequence-conformation relationships, we examine the conformational behavior of fully hydrophilic polypeptoids. We simulate the same polypeptoid

sequences described in Chapter 4 with alternating methoxyethyl and propanol sidechains and labeled at the ends with TEMPO spin probes. In order to investigate the scaling behavior of these peptoid chains, we vary the number of hydrophilic monomers: $N = 1, 2, 3, 5, 7, 9, 11, 13, 15, 19$. We also extract mean end-to-end distances from the experimental Double Electron-Electron Resonance (DEER)-derived distributions in Fig. 4.4. Fig. 7.2 shows that the mean end-to-end distances from simulations and DEER experiments follow excluded volume scaling, characteristic of polymers in a good solvent.³⁵¹

The mean end-to-end distances along the backbone, between terminal nitrogens instead of the spin-label oxygen radicals (see Fig. 7.2 inset), computed from simulations and fit to $\langle R_{ee,NN}^2 \rangle^{1/2} = bN^\alpha$, give a scaling exponent and prefactor of $\alpha = 0.61 \pm 0.02$ and $b = 4.0 \pm 0.2 \text{ \AA}$, respectively. The value of the exponent is consistent with excluded volume scaling ($\alpha = 0.588$) and reasonably consistent with Zhao *et al.* who reported a scaling exponent of 0.66 ± 0.01 for the radius of gyration of polysarcosine,²¹¹ another relatively hydrophilic polypeptoid, computed from molecular dynamics simulations.

However, the mean distance that is computed from the simulated and DEER distributions in Fig. 4.4 is not the backbone distance, but the spin-label end-to-end distance, which incurs an additional contribution (relative to the backbone) from the spin labels themselves. To account for this effect, we fit mean spin-label distances to a shifted power law, $\langle R_{ee,OO}^2 \rangle^{1/2} = b(N + \gamma)^\alpha$ where $b = 4.0 \text{ \AA}$ is the prefactor fit from the backbone end-to-end distances and γ accounts for the effective increase in the number of monomers due to the addition of the spin labels and the distance between the backbone and the

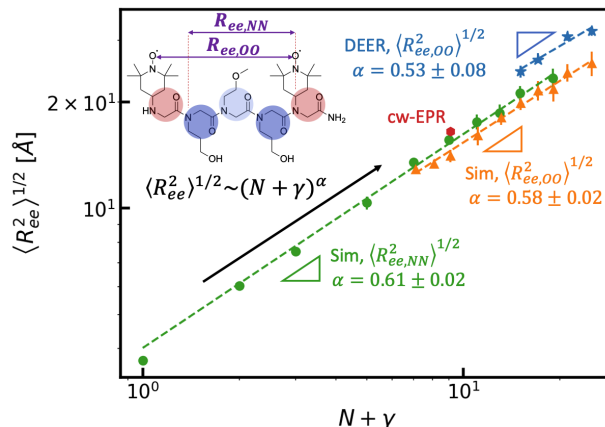


Figure 7.2: Scaling of the mean end-to-end distance shows reasonable agreement between experiments and simulations and suggests excluded volume scaling. The mean end-to-end distance computed from the DEER distributions (blue stars) and from simulation (orange triangles) give the mean distance between the oxygen radicals and green circles give the mean distance along the backbone) using the MFTOID forcefield. The red hexagon shows the mean end-to-end distance from cw-EPR. Dashed lines show fits to a power law, shifted by an effective number of additional monomers γ for the spin-label end-to-end distances, as described in the text. The labels report the fitted exponents. (Inset) The spin-label end-to-end distance, $R_{ee,OO}$, is measured by DEER and cw-EPR and is computed from simulations, while the backbone end-to-end distance, $R_{ee,NN}$, is only accessible in simulations.

oxygen radical on the spin label. Previous work has used a similar offset parameter to fit mean end-to-end distances from FRET.^{352,353} The fitted exponent for the simulated mean spin-label distances is $\alpha = 0.58 \pm 0.02$, corroborating excluded-volume scaling. The fitted effective increase in the number of monomers is $\gamma = 6.0$, suggesting that each spin label effectively behaves as three additional monomers.

We then fit the DEER-derived mean spin-label distances using this shift factor, i.e. to $\langle R_{ee,OO}^2 \rangle^{1/2} = b(N + 6.0)^\alpha$. The resulting exponent is $\alpha = 0.53 \pm 0.08$, and again agrees with excluded volume scaling within uncertainty. Thus, the scaling exponents computed from simulations and DEER are consistent. However, the mean end-to-end distances from simulations are systematically smaller, consistent with previous work comparing mean

distances from DEER and simulation for poly(ethylene oxide).¹⁹² This is partly due to the challenges in resolving the region below 20 Å for DEER. Computing the mean of the portion above the lower accessible limit for DEER (20 Å) for both the DEER-resolved and simulated distributions reduces the discrepancy (Fig. C.21). As far as we know, no experimental works have previously reported scaling exponents for polypeptoids.

If we furthermore include the mean end-to-end distance from the Continuous Wave Electron Paramagnetic Resonance measurement (Fig. 4.6) with the DEER data and fit to the power law, $\langle R_{ee,OO}^2 \rangle^{1/2} = b(N+6.0)^\alpha$, we obtain a scaling exponent of $\alpha = 0.63 \pm 0.06$, remaining consistent with excluded volume scaling.

To confirm excluded volume scaling, leveraging the rich information provided by full conformational distributions as opposed to average values, Fig. F.2 compares simulated distance distributions to those for a theoretical excluded volume polymer, showing reasonable agreement that improves with chain length.

7.3.2 Hydrophobic monomers shift the conformational distribution

As shown above, the conformational behavior of hydrophilic peptoids is straightforward to predict using excluded volume statistics. However, sequences with hydrophobic sidechains will deviate from excluded volume scaling, introducing more varied conformational behavior that is not yet well characterized. As an example, we perform atomistic simulations of peptoids of constant length (16 residues, doubly labeled with spin labels)

with increasing number of aromatic residues at the ends of the sequence (Fig. 7.1). Fig. 7.3 shows the end-to-end distance distributions, $P(R_{ee})$, computed from these simulations (solid lines). Here, in order to compare to the distributions computed from the coarse-grain model (dashed lines), we measure the end-to-end distance as the distance between the center of mass of the heavy atoms in the backbone of the first and last non-spin-label residues. As the total number of aromatic residues, N_{ARO} , increases, the peak of the distribution shifts to lower end-to-end distances. The shape of the distribution also changes and exhibits an increasingly asymmetrical tail at high end-to-end distances at higher N_{ARO} . These behaviors suggest that the peptoid samples both extended and compact conformations (Fig. 7.3 snapshots), with the proportion of the former decreasing (but remaining finite) at higher N_{ARO} .

7.3.3 Coarse-grained model reproduces conformations

To check that the coarse-grained (CG) polypeptoid model reproduces the conformational behavior of the atomistic model, we simulate the same sequences examined atomistically, without the TEMPO spin probe end-labels. The CG model has low barriers of dihedral rotation, such that rotation around backbone-backbone bonds is orders of magnitude faster compared to the slow rotation around the amide bond that is observed in atomistic simulations (section F.1.2). As a result, end-to-end distances converge quickly (Fig. F.1) without enhanced sampling. We are thus able to obtain converged distance distributions in 1-2 orders of magnitude less simulation time (500 ns versus 10 μ s). Meanwhile, the CG

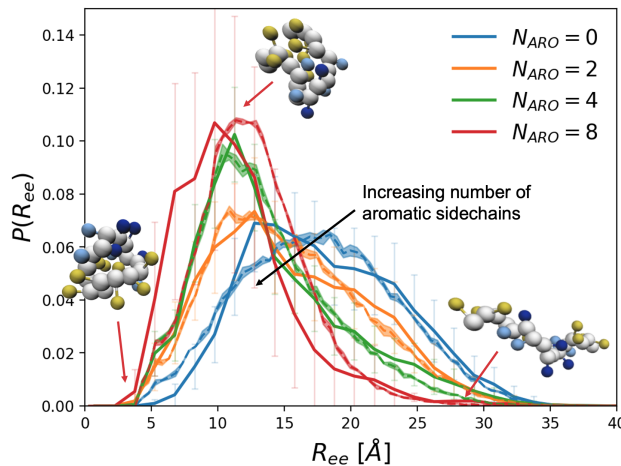


Figure 7.3: End-to-end distance distributions, $P(R_{ee})$, from simulations for peptoids varying the number of aromatic residues in the sequence, N_{ARO} (sequences in Fig. 7.1). The solid lines are computed from atomistic simulations and the dashed lines are computed from simulations of the coarse-grained (CG) peptoid model. Uncertainties for the atomistic simulations (error bars) are computed as described in section C.1.4 and uncertainties for the CG simulations (shaded regions) are computed from three independent simulations. Snapshots are of the sequence with $N_{ARO} = 8$ at various end-to-end distances and are created in VMD.^{155,354} The white beads are CG backbone beads and the colored beads are CG alcohol (darker blue), methoxy (lighter blue), and aromatic (gold) sidechain beads.

simulations themselves are much faster to run because they treat water implicitly. As an example, each of the 10 μs atomistic simulations of the 16-mer sequences in Fig. 7.3 took over a month to run on a single NVIDIA V100 GPU, while the 500 ns CG simulation took approximately 61.5 hours on a single CPU core of an Intel Xeon L5410 Processor, representing at least a $12\times$ speedup, without fully accounting for the reduced statistical uncertainty in the CG simulation.

We compute the end-to-end distance as the distance between the first and last backbone beads. For the alternating hydrophilic sequences, we simulate lengths $N = 5, 7, 9, 11, 13, 15, 17, 19$. We then fit the root-mean-squared end-to-end distance to a power law, $\langle R_{ee,NN}^2 \rangle^{1/2} = bN^\alpha$, and obtain $b = 3.7 \pm 0.2 \text{ \AA}$ and $\alpha = 0.58 \pm 0.02$, within error of

the fitted parameters from the atomistic simulations, $b = 4.0 \pm 0.2 \text{ \AA}$ and $\alpha = 0.61 \pm 0.02$ and consistent with excluded volume scaling.

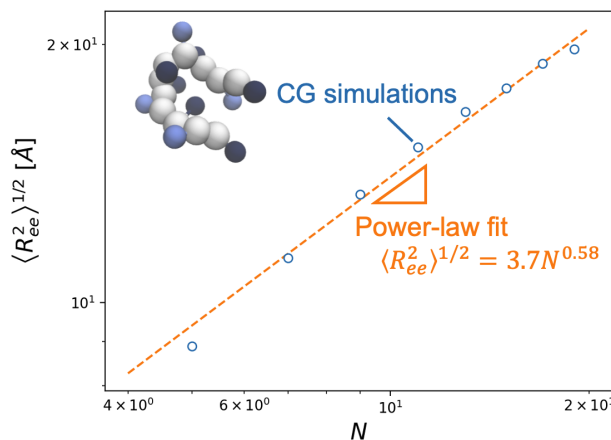


Figure 7.4: Scaling of the mean end-to-end distance with peptoid length, where N is the number of monomers, from CG simulations. The markers show the simulated mean distances (error bars from three independent simulations are smaller than the marker size) and the dashed line shows the fit to the power law relationship. The snapshot, created in VMD,^{155,354} shows an example conformation. The white beads are CG backbone beads and the blue beads are CG alcohol (darker blue) and methoxy (lighter blue) sidechain beads.

We also use the CG model to simulate the sequences with aromatic monomers (Fig. 7.1) and compute the end-to-end distances, again as the distance between the first and last backbone beads. Fig. 7.3 contains the distance distributions from these CG simulations (dashed lines), which show excellent agreement with the distributions from the atomistic simulations (solid lines). The CG model was fit using simulations of shorter peptoids (5 to 6 residues in length) but is able to capture the sequence-dependent conformational behavior of longer peptoids with more significant conformational diversity.

7.3.4 Genetic algorithm optimization identifies extremal sequences

In section 7.3.1, we showed that the end-to-end distances of hydrophilic polypeptoids obey excluded volume scaling. However, addition of hydrophobic monomers shifts the end-to-end distance to lower distance. In order to determine whether hydrophobic monomers can induce deviations away from scaling laws, we use the genetic algorithm approach introduced in Chapter 2, coupled with the CG simulations and a surrogate model. We first fix the composition of the sequence (6 aromatic monomers and 18 methoxy monomers) and choose a simple objective function: $|\alpha^2 \langle R_g^2 \rangle - \langle R_{ee}^2 \rangle|$, which is a measure of deviation from the Flory relationship, $\langle R_{ee}^2 \rangle = \alpha^2 \langle R_g^2 \rangle$. Here, R_{ee}^2 is the squared end-to-end distance or distance between first and last backbone beads, and R_g^2 is the squared radius of gyration for the backbone beads. The brackets indicate that both quantities are averaged over conformations sampled in a CG simulation. We choose $\alpha^2 = 4.231$ such that $|\alpha^2 \langle R_g^2 \rangle - \langle R_{ee}^2 \rangle| = 0$ for the peptoid with aromatic monomers evenly spaced.

Fig. 7.5 shows the genetic algorithm trajectory for the runs maximizing and minimizing this objective function. We run 20 generations in which the sequences are simulated using the CG model, 50 generations where the objective function is computed by a surrogate function, and 10 additional generations in which the sequences are simulated. In the first 5 generations, sequences are generated randomly; in the remaining generations, they are generated *via* crossover. The surrogate model is a linear function of three “features” of the sequence, as described in more detail in section 7.2.3 and reasonably predicts the $\langle R_{ee}^2 \rangle$ and $\langle R_g^2 \rangle$ for a given sequence (Fig. F.3).

During the maximization run, the algorithm finds that the sequence with the aromatic groups clustered at the ends will give the largest deviation from the Flory relationship. For this sequence, the aromatic groups lead to aggregation of the ends of the polypeptoid, reducing the mean end-to-end distance without a comparable reduction in the radius of gyration. During the minimization run, the algorithm finds numerous sequences for which the Flory relationship approximately holds, with deviations of less than 0.01. For the optimal sequence found, the aromatic groups are dispersed throughout the chain, although not perfectly evenly. However, the uncertainty associated with the estimates of $\langle R_{ee}^2 \rangle$ and $\langle R_g^2 \rangle$ from the simulations is larger than the computed value of $|\alpha^2 \langle R_g^2 \rangle - \langle R_{ee}^2 \rangle| = 3.5 \times 10^{-4}$ for this sequence. This result suggests that there are multiple sequences where aromatic groups are separated with varied spacing along the chain that have approximately the same relationship between $\langle R_{ee}^2 \rangle$ and $\langle R_g^2 \rangle$ as the sequence with aromatic groups evenly dispersed.

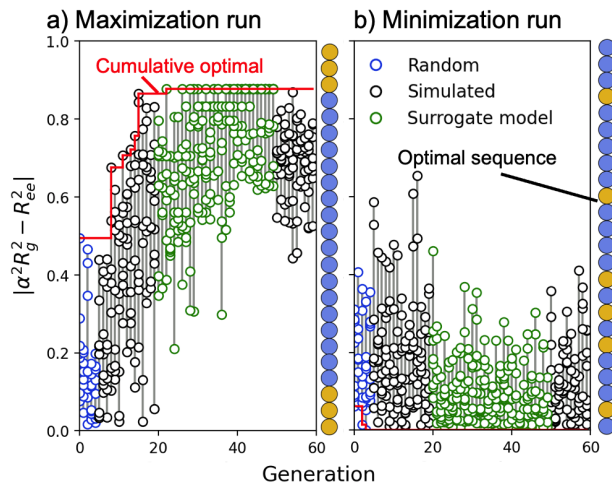


Figure 7.5: Genetic algorithm trajectory for (a) maximizing and (b) minimizing deviation from the Flory relationship. Each marker represents a sequence generated by the algorithm, where the blue markers are sequences generated randomly, the black markers are sequences generated by crossover and simulated, and the green markers are sequences generated by crossover and whose properties are predicted by a surrogate model. The red line indicates the cumulative optimal. The optimal sequences discovered are shown to the right of each trajectory.

7.4 Conclusions

In this chapter, we show that polypeptoid conformational behavior can be captured by a coarse-grained (CG) model, which can then be coupled with a genetic algorithm approach to identify sequences that display extremal conformational behaviors. Atomistic simulations of polypeptoids show that hydrophilic polypeptoids display excluded volume scaling of their mean end-to-end distances, while placing aromatic groups at the ends of the sequence induces deviation from excluded volume behavior and shifts the distribution to lower distances. We show that a CG model developed through relative entropy minimization accurately reproduces both the scaling of mean end-to-end distances and shifts in the distance distributions for hydrophobic sequences. The genetic algorithm suggests that sequences with hydrophobic groups at the ends maximize deviation from the

relationship, $\langle R_{ee}^2 \rangle = \alpha^2 \langle R_g^2 \rangle$, while sequences with aromatic groups separated adhere closely to that relationship.

This workflow can be extended to optimization of other more complex conformational behaviors with less predictable optimal sequences, such as temperature-dependent conformational changes. These optimizations will suggest non-intuitive conformational behaviors, which can then be verified by longer CG or atomistic simulations, or even experimental synthesis and characterization. This approach provides a targeted, design-oriented method for elucidation of sequence-conformation relationships that may be generalizable beyond polypeptoids to other sequence-specific polymers.

Acknowledgments

This work was supported by the Center for Materials for Water and Energy Systems (M-WET), an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Basic Energy Sciences under Award DE-SC0019272. Use was made of computational facilities purchased with funds from the National Science Foundation (CNS-1725797) and administered by the Center for Scientific Computing (CSC). The CSC is supported by the California NanoSystems Institute and the Materials Research Science and Engineering Center (MRSEC; NSF DMR 2308708) at UC Santa Barbara. S.J. acknowledges support by the National Science Foundation Graduate Research Fellowship (DGE 1650114).

Chapter 8

Conclusions

8.1 Summary

In this dissertation, we use molecular dynamics simulations to investigate the nanoscopic mechanisms relevant to the design of membrane surfaces that can modulate the transport of small, neutral solutes and resist fouling. We show that water’s structure and dynamics are significantly perturbed by variations in chemical and topological patterning at both extended and molecular surfaces. These shifts in water structure affect the behavior of local solutes, leading to tunable affinity and transport of small-molecule solutes. We furthermore introduce workflows for the simulation and computational inverse design of multifunctional surfaces and molecules, and show that this approach identifies non-intuitive chemistries that can inspire further investigation of fundamental mechanisms and structure-property relationships. These relationships suggest design rules that can be used to propose other new materials (Fig. 8.1).

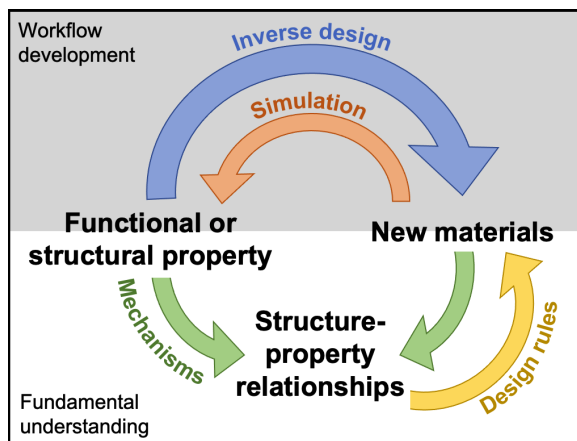


Figure 8.1: In this dissertation, we develop new workflows for the simulation and computational inverse design of new materials (top). We then use these workflows to understand nanoscopic mechanisms controlling surface-water-solute interactions, elucidating relationships between surface structure (chemistry and topology) and functional properties (bottom). These relationships suggest general design rules for new membrane materials.

In Chapter 2, we use a model nanopore, to which functional groups can be directly tethered and precisely patterned, to show that changes in the patterns of polar hydroxyl and nonpolar methyl functional groups can modulate the transport of water and a small, neutral solute (boric acid) through the pore. Importantly, the effects on water and boric acid are distinct, potentially driving selectivity for water. A genetic algorithm coupled to molecular dynamics simulations and machine-learned surrogate models discovers that wall chemistries with rings of hydroxyl groups optimally hinder boric acid transport over water transport. We propose that these rings create an energetic barrier that boric acid must “hop” over in order to traverse down the pore, and show that the effect generalizes to other small, neutral solutes. The genetic algorithm inverse design procedure for a particular system thus inspires a more general design approach to improving rejection of small, neutral molecules: creating a “rough” free energy surface.

We investigate an alternative method of creating rough free energy surfaces *via* oscillations in the surface topology in Chapter 3, where we simulate nanopore wall patterns comprising longer polar hydroxymethyl and nonpolar ethyl groups, in addition to shorter hydroxyl and methyl groups. We find that patterning of ethyl groups in methyl pores affects boric acid and water transport, and that rows of ethyl groups have a similar qualitative effect compared to rows of hydroxyl groups, hindering boric acid transport without a similar reduction in water transport. In order to explore the combined effects of chemistry and topology, we improve on the inverse design workflow in Chapter 3 by performing the optimization in a lower-dimensional space, such that it is computationally tractable to search through patterns with three or more headgroups. We find that fully nonpolar patterns that hinder boric acid transport *via* surface roughness are superior to patterns with hydroxyl groups.

We then transition from model systems to a realistic and synthetically accessible platform, with similarly high tunability: polypeptoid molecules, which are closely related to polypeptides. In Chapter 4, we establish a workflow for computationally characterizing these polypeptoids. We first develop an enhanced sampling molecular dynamics procedure, which efficiently samples over the long timescales associated with backbone conformational changes. Detailed information about polypeptoid conformational ensembles, namely the end-to-end distance distribution obtained by experimental colleagues using Double Electron-Electron Resonance spectroscopy, provides a unique validation of the computational workflow, compared to previous efforts.

Prior work also lacked an understanding of how polypeptoid chemical heterogeneity (i.e. sequence) tunes local water behavior, a surprising gap since polypeptoids have been used to modulate properties of aqueous interfaces. In Chapter 5, we show that in dilute solution, nonpolar and polar peptoid sidechains have distinct effects on the structure and dynamics of the peptoid’s hydration water. Specifically, replacement of polar sidechains by nonpolar sidechains slows down water and results in enhanced tetrahedrality. These two effects are highly correlated, useful for extending the information obtained from experimental measurements of dynamics. Notably, the trends in water’s behavior are distinct from those near nonpolar versus polar extended, smooth surfaces. Water’s response to polypeptoids, in fact, more closely resembles its response to small-lengthscale solutes compared to its response to extended surfaces, suggesting that even when the polypeptoids are tethered to surfaces, the effect of changes in polypeptoid chemistry will be distinct from the behavior of idealized surfaces.

We therefore next investigate water behavior near polypeptoid-tethered surfaces in Chapter 6, finding that the response of water has contributions from both its response to molecular surfaces as well as its response to extended, smooth surfaces. For both these “small-lengthscale” and “large-lengthscale” contributions, water’s structure does show clear distinctions between uncharged and charged, zwitterionic peptoid surfaces, which correlate with the reduction in solute affinity near these zwitterionic surfaces. We relate water structure and solute affinity through decomposition of the free energy of binding into direct energetic and relative entropic components, the latter of which character-

izes the penalty associated with changes in water structure upon solute binding. This molecular-level insight provides a mechanistic hypothesis for the improved antifouling capabilities observed for zwitterionic surfaces.

We close by revisiting the inverse design ideas from the beginning of this dissertation and demonstrate how they can be applied to the design of polypeptoid sequences in Chapter 7. Here, we establish workflows that tune sequence to achieve specific single-chain conformational behaviors, inspired by the protein-folding problem. Simulations of the atomistic polypeptoid model show that polypeptoid sequence can shift conformational distributions. Importantly, we also show that these shifts can be accurately captured by a computationally cheaper coarse-grained model, allowing us to run inverse design workflows in reasonable amounts of time. Finally, we use the genetic algorithm approach from Chapter 2 to identify sequences that induce specific conformational behaviors.

8.2 Outlook

Further advancements to a nanoscopic understanding of surface-water-solute interactions *via* an inverse design approach can be made by improving the workflows laid out in this dissertation or applying them to new systems.

8.2.1 Improvements to computational workflows

While the active learning approach in Chapter 3 is a significant improvement over the genetic algorithm approach in Chapter 2 in terms of efficiently identifying diverse and

optimal patterns, there remain many directions for additional enhancements.

One existing inefficiency is that the algorithm can still simulate the same pattern multiple times. For instance, when the composition of the pattern is allowed to vary, the algorithm explores regions where a certain headgroup type is either not present or present at very low fractions (i.e. only 1 group). As a result, the pattern is independent of any interaction parameters governing interaction between types of that group. However, the algorithm does not know this and therefore suggests multiple sets of parameters that result in the same pattern. A relatively easy fix for this is to check each suggested pattern against the set of patterns that have already been simulated, and to avoid simulating it if it is a repeat (a randomly generated parameter set can be used instead, to generate a new pattern). One could even avoid simulating patterns that are sufficiently similar to previously simulated patterns, though this risks missing out on patterns that improve on the objective function. A more complex but potentially more powerful approach is to encode some restrictions in the bounds for each parameter that account for these and other degeneracies (e.g. two sets of seemingly different parameter sets may generate the same pattern). This approach requires an understanding of the space of parameters that may be more difficult to obtain for larger parameter sets used to generate patterns with more than three headgroups, limiting generalization.

Performing the optimization in a lower dimensional space necessarily limits the number of patterns that the algorithm can generate, such that it could miss the actual optimal. For instance, annealing a pattern energy function generates patterns that are

more regular (lower “pattern entropy,” as defined in section A.2.1). The conditions of the simulated annealing procedure can be modified (or even included in the optimization) in order to generate less regular patterns (e.g. by annealing at a higher temperature). Alternatively, the active learning optimization in parameter space could be followed by an additional short genetic algorithm optimization in pattern space starting from the optimal discovered by the first stage. This combination would benefit from the ability of the optimization in lower dimensions to efficiently discover diverse patterns, as shown in Fig. 3.4, as well as the ability of the genetic algorithm to explore smaller changes to pattern structure that cannot be enforced by the pattern energy function. This tradeoff between a simplified design space and potentially missing actual optima will be especially important for sequence design problems, where optimal sequences may be difficult to generate from a low dimensional set of parameters. This problem may benefit from a shorter initial optimization in parameter space and longer genetic algorithm runs.

The genetic algorithm optimization in sequence space can be accelerated by using a surrogate model that predicts the objective function based on the sequence alone. Chapters 2 and 7 used a simple model that was linear in “features” of the pattern, where the features are generated based on existing intuition. Recent work has shown promise for bypassing physics-based models in favor of black-box ones.³⁵⁵ In Chapter 2, we found that such black-box neural network-based surrogate models that bypass intuited features do not perform significantly better than simpler models, similar to the result in Ref. [293], particularly for cases where datasets are small. However, useful features for

sequence design may be difficult to intuit, due to potential “long-range” interactions between chemistries along the chain. Neural network approaches such as long short-term memory recurrent neural networks have been shown to have good predictive properties for sequence design problems,³³¹ and could be tested for peptoid sequences. The peptoid coarse-grained (CG) model makes generation of larger datasets possible, which could improve the accuracy of neural network models. While it is more difficult to interpret neural network models compared to feature-based linear models, examination of the optimized sequences can provide physical insight.

The peptoid model can also be improved through better sampling of peptoid conformations. The CG model has shown impressive accuracy for single-chain conformational behavior. However, other systems, such as peptoid-tethered surfaces, may require additional atomistic simulations to either fit other CG models or investigate local water behavior. While the expanded ensemble method efficiently samples over amide bond isomerization, other long timescales associated with conformational changes could extend the requisite simulation time. For instance, peptoids that are tethered to a surface can fully adsorb to the surface, slowing down conformational changes. Combining the expanded ensemble method with another enhanced sampling technique would accelerate sampling along other slow degrees of freedom. Parallel bias metadynamics, in particular, has shown promise in efficiently accelerating multiple degrees of freedom or collective variables.³⁵⁶ It could be easily combined with the expanded ensemble technique by including the expanded ensemble parameter that softens dihedral potentials, λ_{ee} , in the set

of collective variables, along with the peptoid-surface distance, for instance.

8.2.2 Application of workflows to new systems

Since membrane surfaces must perform multiple objectives, such as resisting fouling and enhancing water-solute selectivity, a natural use of the inverse design workflow is to optimize multiple objectives. One particular interesting set of objectives would be the permeability-selectivity behavior. In Chapter 2, we computed permeability and selectivity metrics and found that the genetic-algorithm-discovered pores showed more optimal properties (higher permeability and selectivity). However, the computational expense of accurately sampling boric acid partitioning into the nanopore precluded a full optimization in the permeability-selectivity space. The modified inverse design workflow presented in Chapter 3 could make such an optimization possible, by reducing the number of iterations needed. The multi-objective optimization could be done using Pareto front optimization techniques such as those presented in Refs. [149] and [78].

In addition to membrane surfaces, the inverse design workflows can be applied to optimization of other aqueous interfaces whose properties are tuned by modulating chemical functionality, such as antimicrobial and antifouling surface coatings. In particular, it would be interesting to apply these design approaches to the peptoid brush surfaces examined in Chapter 6 to design peptoid sequences that minimize solute adsorption. The work in 6 found that zwitterionic surfaces showed enhanced antifouling capabilities and furthermore that the sequence of the charged residues modulated fouling and water be-

havior. However, only a couple different zwitterionic sequences were explored, and many other tunable factors were not varied, including peptoid surface grafting density, peptoid length, and identity of the charged groups. Since these factors are discrete, a genetic algorithm approach may be the most useful in searching this space for an optimal peptoid design. To accelerate this optimization, we can predict the free energy of solute adsorption using the three-body angle distribution of the surface's hydration water, a relationship that was established in Chapter 6, allowing for faster computations. The optimal candidates can suggest peptoid sequences to synthesize and test experimentally.

This study could be paired with a detailed and systematic investigation of the effect of zwitterionic chemistries on the response of water and solutes. In Chapter 6, we observed that zwitterionic chemistries at a surface induce unique water structures that are opposite those created by uncharged chemistries. Analysis of water structure near zwitterionic peptoids in dilute solution will reveal whether such effects are characteristic of all zwitterions or only those located at an extended surface. By systematically changing the spacing between positively and negatively charged groups along the peptoid chain, we can also determine the effect on water structure and develop a nanoscopic mechanism underlying the hypothesis that such spacing affects antifouling behavior.^{322,323}

Synthesis of peptoid brush surfaces that are similar to what is simulated (peptoids tethered to a nonpolar surface) could also enable sum frequency generation (SFG) measurements that can be compared with computationally predicted spectra. SFG measurements have been used to study water structure at antifouling surfaces, but the specific

arrangements and behaviors of water implied by peaks in the SFG spectra are poorly understood. Computational workflows exist for predicting SFG spectra from simulations,^{357,358} though the computational expense may be significant since simulations need to be run with flexible water models. In addition, the structural trends predicted by the flexible water model will need to be compared with the rigid OPC4 model used in this dissertation. Successful computation of the SFG spectra, however, will allow for direct assignment of SFG spectra and a molecular picture of the water characteristics associated with “tightly bound” or “strongly hydrogen-bonded” water versus weakly bound and weakly hydrogen-bonded water.

These workflows could also be used to examine fouling and transport of other small, neutral solutes that are also relevant to water treatment efforts, such as arsenous acid and 1,4-dioxane. This effort would also involve development and/or validation of accurate classical forcefield parameters applicable for dilute aqueous solutions for less frequently modeled solutes such as arsenous acid. A recent study, for instance, has conducted such validation for 1,4-dioxane by comparison against experimental properties.³⁵⁹ Another route for validation is comparison against density functional theory calculations, as was done for boric acid in Ref. [122].

Appendix A

Supporting Information for Inverse Design of Pore Wall Chemistry

A.1 Simulation models and property calculations

A.1.1 Pore geometry

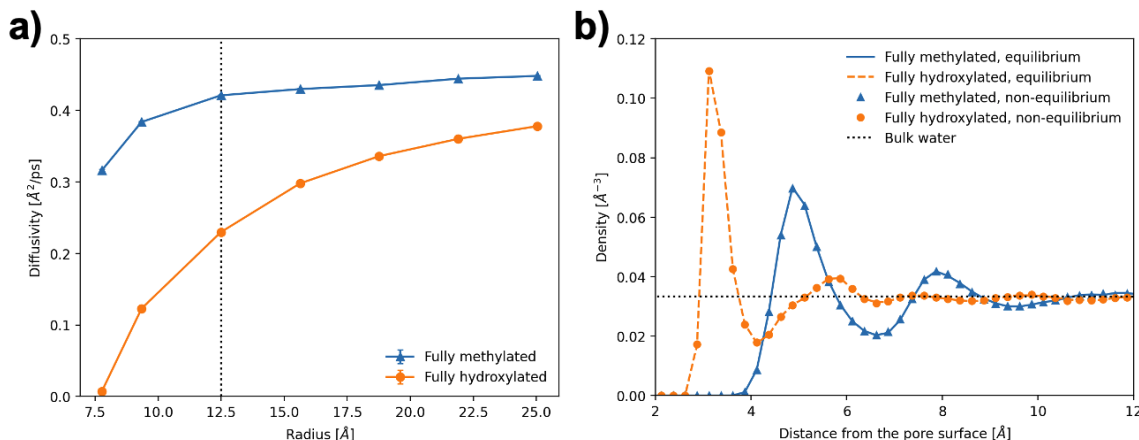


Figure A.1: (a) As the pore radius increases, water diffusivity in fully methylated (triangles) and fully hydroxylated (circles) pores increases and the difference between the diffusivity in the fully methylated and fully hydroxylated pores shrinks. The dotted, black line marks the pore radius used in the rest of this work. We use the same workflow described in section 2.2 to simulate these pores and measure the diffusivity and uncertainty (smaller than the markers in this plot) but omit the boric acid solute. (b) The radial density profiles illustrate that at a pore radius of 1.25 nm, the water density reaches that of the bulk in the middle of the pore, but the pore wall chemical functionality has a significant effect on water structuring inside the pore for the fully methylated (blue) and fully hydroxylated pores (orange). The profiles from the equilibrium (lines) and nonequilibrium (markers) simulations are consistent, suggesting that the applied pressure gradient is not large enough to perturb water structure from equilibrium. We use the same workflow described in section 2.2 to simulate these pores but omit the boric acid solute.

A.1.2 Functional group layout

A.1.3 Forcefield details

Bonded parameters

The CNT carbon-carbon bond lengths are 0.142 nm. Harmonic bond and angle parameters for the functional groups are taken from the OPLS-AA/L forcefield¹⁵³ as described in Tables A.1-A.2. Angle potentials and dihedral potentials between the functional group and carbon nanotube atoms are set to 0. All 1-3 LJ and electrostatic interactions are set to 0. 1-4 LJ interactions between the carbon nanotube atoms and functional group heavy atom are computed using the Lorentz-Berthelot combining rules, while 1-4 LJ interactions between the carbon nanotube atoms and functional group hydrogens are set

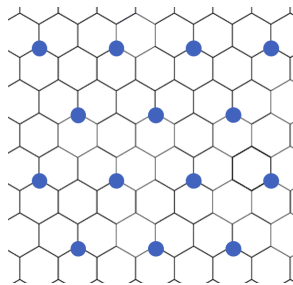


Figure A.2: A schematic showing how functional groups (blue circles) are arranged on the carbon nanotube lattice. The functional groups are placed every 4 carbons to achieve a surface density of $0.21 \text{ nm}^2/\text{group}$.

to 0. 1-4 electrostatic interactions are multiplied by 0.8333. This keeps the functional groups inside the nanotube while allowing them to rotate freely.

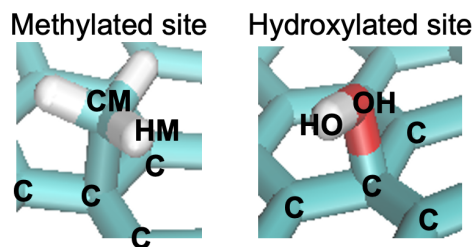


Figure A.3: Labeling of atoms in methylated and hydroxylated site for Tables A.1-A.2

Bond	b_0 [nm]	K_b [kJ/mol/nm ²]	Source
C-CM	0.1529	224262.4	OPLS-AA/L CT-CT
CM-HM	0.109	284512	OPLS-AA/L CT-HC
C-OH	0.141	267776	OPLS-AA/L CT-OH
OH-HO	0.0945	462750.4	OPLS-AA/L OH-HO

Table A.1: Harmonic bond parameters for nanopore functional groups.

Angle	θ_0 [deg]	K_θ [kJ/mol/rad ²]	Source
C-C-CM	0	0	—
C-C-OH	0	0	—
C-CM-HM	110.7	313.8	OPLS-AA/L CT-CT-HC
HM-CM-HM	107.8	276.144	OPLS-AA/L HC-CT-HC
C-OH-HO	108.5	460.240	OPLS-AA/L CT-OH-HO

Table A.2: Harmonic angle parameters for nanopore functional groups.

Boric acid forcefield

Simulations with the boric acid dihedral potential (H-O-B-O) from Ref. [122] result in long rotation times around the oxygen-boron bond, resulting in poor sampling of different dihedral states (Fig. A.4). For this work, we reduce the force constant by half to enhance sampling (Fig. A.4) in lieu of running separate simulations for both dihedral states. The force constant remains significant (over $3k_B T$ even after reduction) and thus still enforces planarity in the molecule. This reduction does not affect the transport properties (Fig. A.5).

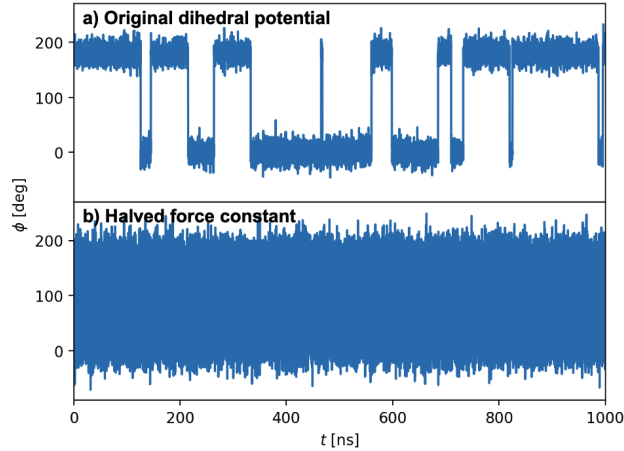


Figure A.4: Boric acid dihedral (H-O-B-O) angle with (a) the original dihedral potential from Ref. [122] and (b) the force constant halved. With the original potential, isomerization of the dihedral occurs on the order of hundreds of ns, resulting in poor sampling of the dihedral states in the 100 ns production simulation used by the genetic algorithm to compute boric acid flux. Reducing the force constant significantly enhances the sampling of dihedral states. The workflow is the same as described in section 2.2 for the infinite pore system with applied pressure gradient, except that the production simulations are run for 1 μ s.

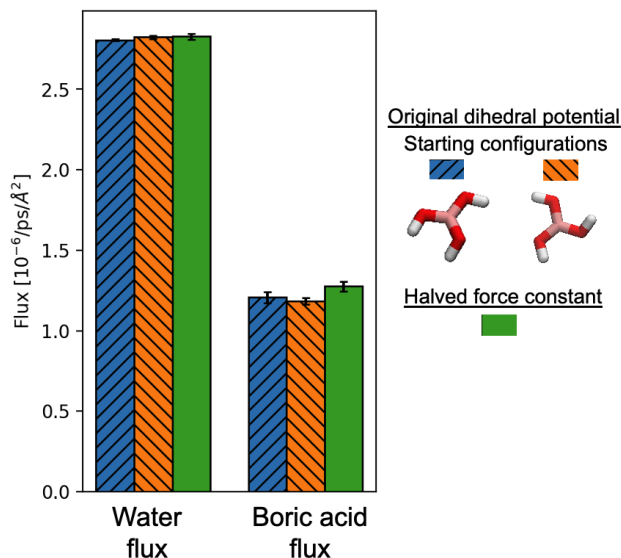


Figure A.5: The measured water and boric acid flux does not change when the dihedral force constant is altered. The workflow is the same as described in section 2.2 for the infinite pore system with applied pressure gradient, except that the production simulations are run for 1 μ s. Furthermore, an additional harmonic potential with force constant 20 kJ/mol/rad² is applied to the dihedral during the initial NVT equilibrations and is set to achieve the starting configurations in the figure. The reported fluxes are computed by averaging from three independent simulations. The reported uncertainty is the standard error of the mean.

A.1.4 Overview of pore simulation steps

For each pattern, the following main calculations are performed:

1. **Simulate a finite pore between two reservoirs (Fig. 2.1b in main text) with a single boric acid molecule to compute the average water density inside the pore.** We run a 100 ps NVT equilibration, then a 10 ns NPT equilibration, and finally a 20 ns NPT production simulation.
2. **Prepare an infinite pore (Fig. 2.1b in main text) with the same water density as the finite pore.**
3. **Compute the flux of boric acid and water in the infinite pore under an applied pressure gradient *via* a nonequilibrium simulation.** We run a 2 ns NVT equilibration followed by a 100 ns NVT production simulation (in the genetic algorithm) or a 200 ns NVT production simulation (to examine the rationally designed patterns). The first 10 ns are discarded when computing properties.

For each rationally designed pattern (i.e. beyond those explored by the genetic algorithm), the following additional calculations are performed:

5. **Run an expanded ensemble simulation turning on the boric acid non-bonded parameters to compute the free energy of solvation in the pore, and thus the partition coefficient of boric acid inside the pore.** We run a 500 ps NVT equilibration and then a 500 ps NPT equilibration with all non-bonded and partial charges turned on. Subsequently, we run the expanded ensemble simulation in the NPT ensemble (500 fs between MC barostat moves) for 10 ns to equilibrate state weights and then for 410 ns with fixed weights, changing expanded ensemble states every 1 ps. We use the final 410 ns when computing properties.
6. **Compute the diffusivity of boric acid and water in the infinite pore without an applied pressure gradient *via* an equilibrium simulation.** We run a 2 ns NVT equilibration followed by a 1 μ s NVT production simulation.

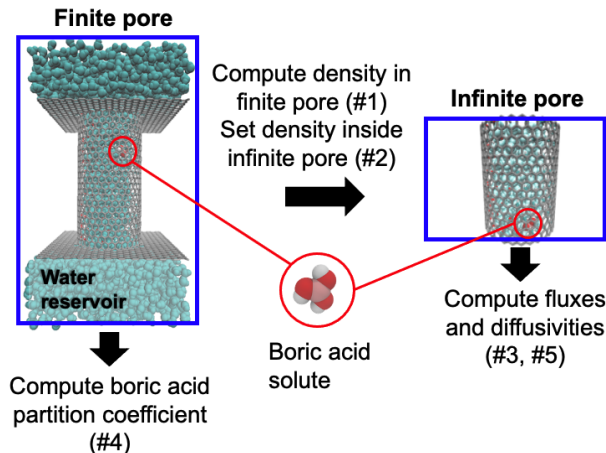


Figure A.6: Visualization of the finite pore and infinite pore systems (reproduction of the snapshots from Fig. 2.1b in the main text at a larger scale), annotated with the steps outlined in section A.1.4 for simulating the pores and computing the transport properties.

A.1.5 Finite pore setup

To eliminate end effects in the finite pore, we first repeat the pore pattern for 4 additional rows in both directions, creating a pore of length of 5.0 nm. Calculations of density and partition coefficients are performed excluding this added region.

To bound the water reservoirs, we create two graphene sheets with width and length equal to 5 nm (4 times the pore radius), with openings for the pore created by (1) deleting all carbon atoms within the convex hull defined by carbon atoms in the graphene sheets that are closest to CNT carbon atoms and then (2) bonding pairs of graphene and CNT carbon atoms that are within 0.242 nm of each other. The graphene and CNT positions are fixed during the simulation; these bonds are needed only so that the barostat recognizes them as a single molecule. Nonbonded forcefield parameters for the carbon

atoms in the graphene sheet are the same as those for the CNT. We solvate the entire box using GROMACS²¹⁴ and then remove all waters in between the graphene sheets (region laterally external to the pore). We then insert a single boric acid molecule using GROMACS.

For simulations with the finite pore, unless otherwise noted, we restrain the center of mass of the single boric acid so that it remains within the nanopore using a flat-bottom harmonic restraint, $0.5k(dz - dz_0)^2\text{step}(dz - dz_0)$ where dz is the distance along the pore axis between the boric acid heavy atoms' center of mass and a carbon atom near the center of the pore, $dz_0 = 1.70384$ nm (half the pore length, minus the added region length), $\text{step}(x)=1$ when $x \geq 0$ and is 0 otherwise, and $k = 1000$ kJ/mol/nm².

For the expanded ensemble simulation, we additionally apply a flat-bottom harmonic restraint to the boric acid solute in the radial direction to keep it inside the pore. The restraint has the form $0.5k_r(r - R)^2\text{step}(r - R)$ where r is the distance in the xy -plane between the boric acid heavy atoms' center of mass and the center of the pore, R is the pore radius, $\text{step}(x) = 1$ when $x \geq 0$ and is 0 otherwise, and $k_r = 1000$ kJ/mol/nm².

A.1.6 Partition coefficient calculation

The partition coefficients of boric acid between bulk water and the pore are computed as described in the main text. The reference values $K_0 = 8.0$ and $\Delta G_0 = -3.4k_B T$ are computed for the fully methylated pore, where the partition coefficient is computed from an unbiased MD simulation in the finite pore system with the same setup as described previously but with 20 boric acid molecules (placed initially in the water reservoir using GROMACS). For the unbiased MD simulation, we perform a 100 ps NVT equilibration followed by a 100 ns NPT equilibration where we save configurations every 1 ns. We set the box length in the axial direction to the average of the second half of the NPT equilibration. We then perform an additional 100 ns NVT equilibration. Finally, we perform a 400 ns NVT production. The friction coefficient for Langevin dynamics is 0.1 ps⁻¹. We do not restrain the boric acid molecules in these simulations and compute the partition coefficient as the ratio of the average boric acid density inside the pore by the average boric acid density in the water reservoir (excluding the region within 5 Å from the graphene sheets). We do this for each 80-ns block in the production simulation and then average the resulting partition coefficients. Fig. A.7 shows the agreement between the partition coefficients computed from the expanded ensemble and unbiased simulations.

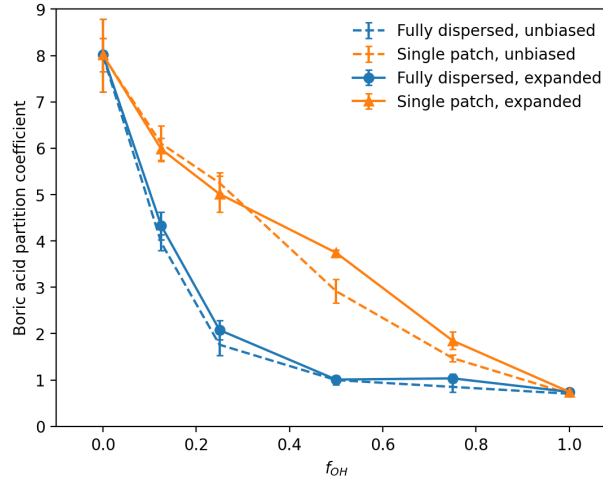


Figure A.7: Partition coefficients computed from the free energy of solvating the boric acid inside the pore (solid), as described in the main text, compared to those computed from the unbiased MD simulations described above (dashed), show good agreement. For the former, three independent runs are averaged together and the uncertainty is the standard error of the mean. For the latter, the standard error of the average partition coefficient from 5 80-ns blocks, as described earlier, is reported. The orange markers represent patterns with a “single patch” of hydroxyl groups, spanning the pore axis, while the blue markers represent “fully dispersed” patterns.

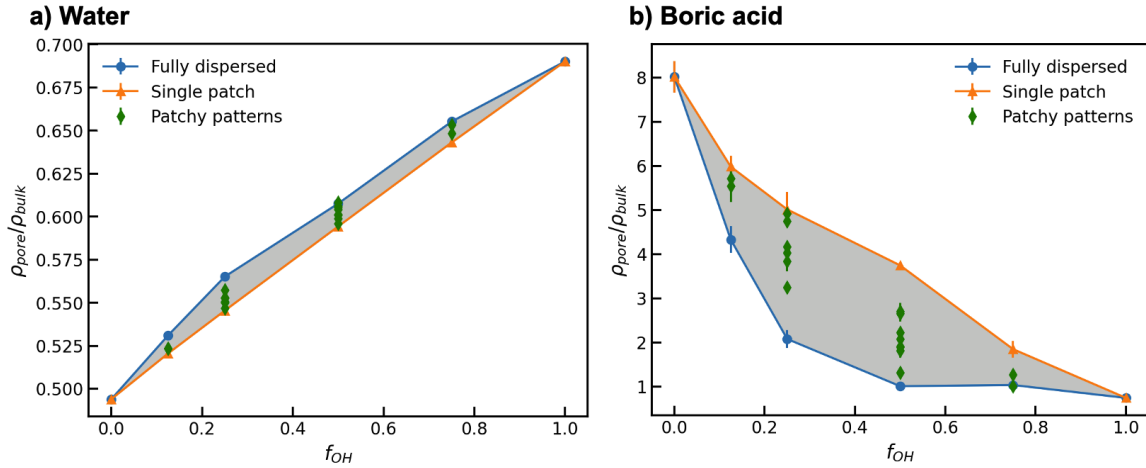


Figure A.8: Partition coefficients for a) water and b) boric acid for all rationally designed patterns. Uncertainties are the standard error of the mean from three independent simulations (for the water partition coefficient, they are smaller than the marker size).

A.1.7 Infinite pore setup

We generate the initial positions of the CNT carbons and functional groups using an in-house script. The positions of the CNT carbon atoms are then fixed during the simulation.

We then place a single boric acid molecule into the pore using GROMACS. Finally, we solvate the pore by placing waters in one by one, randomly drawing their center of mass positions and orientations. We continue to adjust waters until their number in the pore equals that of a finite pore with the same pattern connected to a water reservoir (without the end regions).

A.1.8 Flux calculation

To compute the flux of water and boric acid under an applied pressure gradient, we add a constant acceleration to each water and boric acid atom. The added acceleration is

$$a = \frac{dP}{dz} \frac{R^2 L \pi}{M} \quad (\text{A.1})$$

where dP/dz is the desired pressure gradient, R is the radius of the nanopore, L is the length of the nanopore, and M is the total mass of the water molecules and boric acid solute. The pressure gradient is set to $dP/dz = 12$ MPa/nm, which is higher than pressure gradients in typical membranes. However, measurements of water flux over different pressure gradients reveal that 12 MPa/nm is still within the range where water flux is linear with pressure gradient, suggesting that the applied force is small enough that the transport trends can be extrapolated to lower gradients.

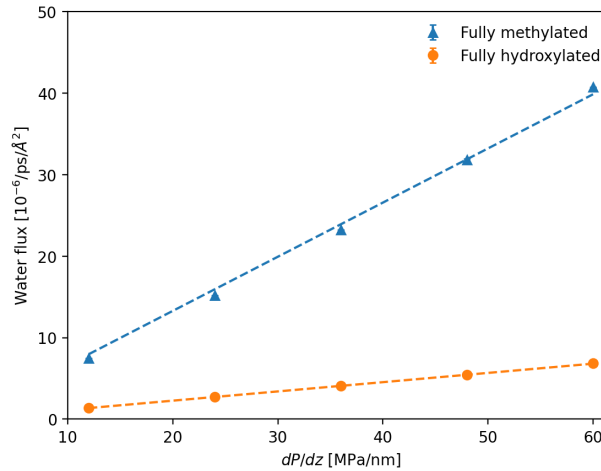


Figure A.9: The average molar flux of water is linear with respect to the pressure gradient. We use the same workflow described in section 2.2 to simulate these pores and compute the water flux and uncertainty (smaller than the markers in this plot) but omit the boric acid solute. The dashed lines show the lines of best fit, with the intercept fixed at the origin.

To compute the flux, we divide the production simulation into 5 blocks. In each block, the total net displacements of the water oxygen atom and boric acid boron atom are computed and divided by the length of the block, the volume of the pore, and the

number of molecules (water or boric acid, depending on the calculation). The fluxes are averaged over the 5 blocks.

A.2 Genetic algorithm workflow

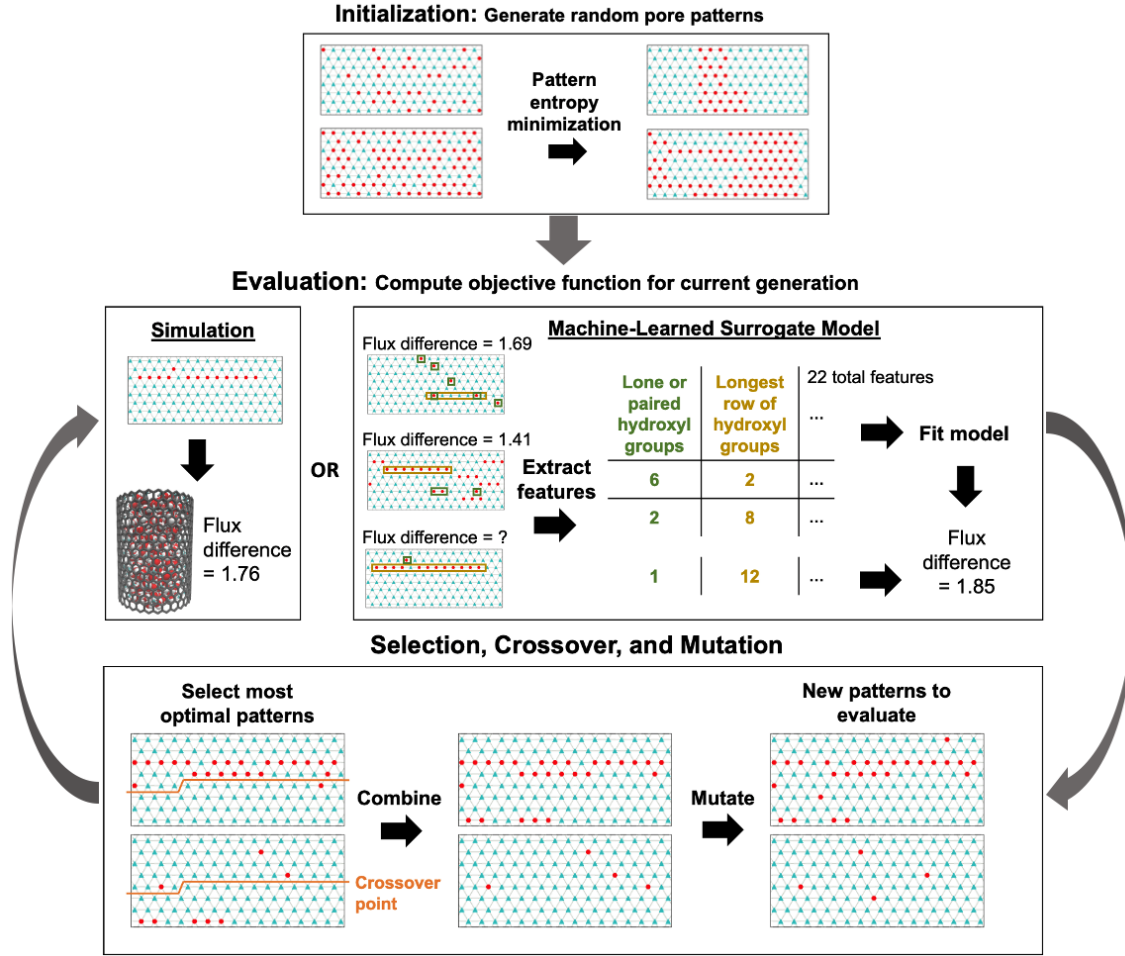


Figure A.10: The genetic algorithm workflow.

A.2.1 Pattern entropy minimization

The pattern entropy is computed as $S = -\sum_{i=1}^{512} P_i \log P_i$ where i is an index over all $2^9 = 512$ possible 3×3 arrangements of methyl and hydroxyl groups and P_i is the fraction of 3×3 subsections of the pattern (128 subsections in total) that have the i th arrangement. For instance, the index corresponding to the 3×3 arrangement in the subsections marked in the pattern in Fig. A.11 is $i = 447$ and the 3×3 arrangement appears twice in the pattern, as marked in Fig. A.11. Thus, $P_{447} = 2/128$.

The pattern entropy is minimized *via* quenching. In each step of this procedure, the positions of one hydroxyl and one methyl group are exchanged. The change in pattern entropy, ΔS , is computed and the move is accepted with probability $\min[1, \exp(-\Delta S/T)]$.

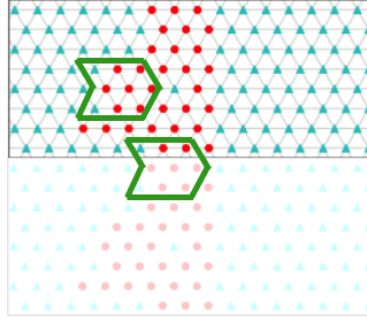


Figure A.11: Example pattern with 3x3 subsections marked to illustrate pattern entropy calculation. The calculation accounts for periodic boundaries, as depicted here.

where T is a (“temperature”) parameter that adjusts the rate at which moves that increase pattern entropy are accepted. For the random generations, we run this quenching procedure on each generated pattern for 1000 steps at a fixed temperature of 0.01. Fig. A.12 shows example quenching trajectories. This pattern entropy minimization procedure creates more feature-diverse patterns (sampling a wider range of pattern features) as shown in Fig. A.13. Later in the genetic algorithm, we also use the pattern entropy quenching minimization to refine the first generation of patterns created by the second set of MD-driven generations; this time, the algorithm is run for 10,000 steps at a temperature of 0.05.

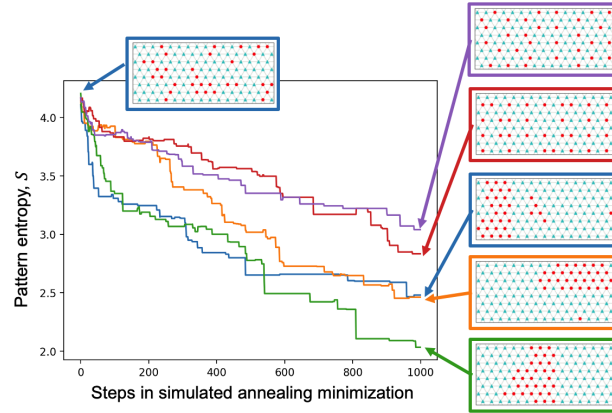


Figure A.12: Example quenching trajectories. Patterns with 32 hydroxyl groups are generated by randomly choosing 32 sites. Each pattern is then entropy-minimized through the quenching procedure for 1000 steps at a temperature of 0.01.

A.2.2 Maximization generations

In the genetic algorithm, the random generations are followed by a “maximization” generation, which follow the procedures described in Ref. [61]. Briefly, in each maximization

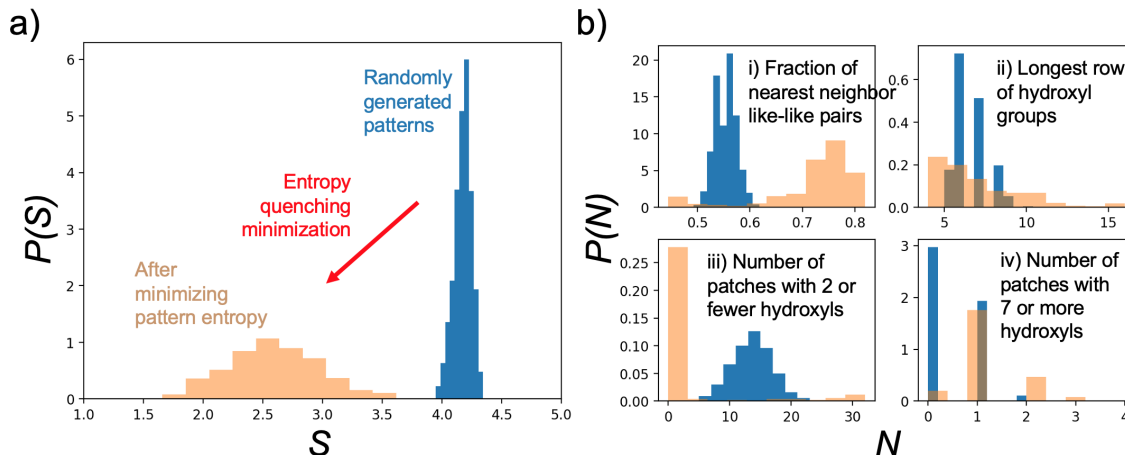


Figure A.13: Entropy minimization leads to more diverse patterns. To illustrate this, we create 1000 “randomly generated” patterns by randomly choosing 32 sites to be hydroxyl groups. We also create 1000 “entropy minimized” patterns by again randomly choosing 32 sites to be hydroxyl groups and then minimizing pattern entropy by running the quenching algorithm for 1000 steps at a temperature of 0.01. (a) The quenching algorithm leads to lower pattern entropies in the entropy minimized patterns (tan) versus the randomly generated patterns (blue). (b) Histograms over various pattern features show that the entropy minimized patterns (tan) cover a larger range of pattern feature values compared to the randomly generated patterns (blue).

generation, a set of 8 optimal parent patterns is chosen *via* a tournament selection procedure. These parent patterns are randomly grouped in pairs and combined to form new child patterns. The patterns are then mutated. The mutation rate used here is 0.04.

A.2.3 Model-driven generations

The only difference between the model-driven generations and the MD-simulated generations is the calculation of the flux difference. Models of the water and boric acid flux are fit based on 22 features of the pattern (section A.3.1 lists the features used). The sklearn package is used to scale the features and the log of the water and boric acid fluxes (fitting the log of the fluxes was found to yield more accurate models). Then, using multiple linear regression in sklearn, linear models are fit that predict the scaled logarithm of the water and boric acid fluxes from the set of pattern features. These models are then used in the model-driven generations to predict the flux difference. When the model is refit based on the second set of simulated generations, we recompute the predictions from the previous model-predicted maximization generations using the new model.

We hypothesize that the surrogate model enhances the efficiency of the genetic algorithm by making it more robust to uncertainty in the MD simulation. To demonstrate this, we test a mock objective function that is the “correlation” with an “ideal” pattern. Here, the correlation is the number of functional groups of the same type in the same

position (e.g. the maximum overlap is 128) and accounts for all translations and reflections of the pattern. Since the correlation can be computed quickly (in contrast to an MD simulation), this mock objective function provides a way to rapidly test the genetic algorithm workflow. We can also artificially introduce “error” to the correlation calculation by adding Gaussian noise. Fig. A.14 shows that when there is error in the objective function, the genetic algorithm finds, on average, more optimal patterns when it uses the surrogate function to predict the objective function. We note that more complex models with a larger number of fitted parameters compared to the linear model may be more prone to overfitting and thus may not enhance the robustness of the genetic algorithm, one of the motivations for choosing a linear model.

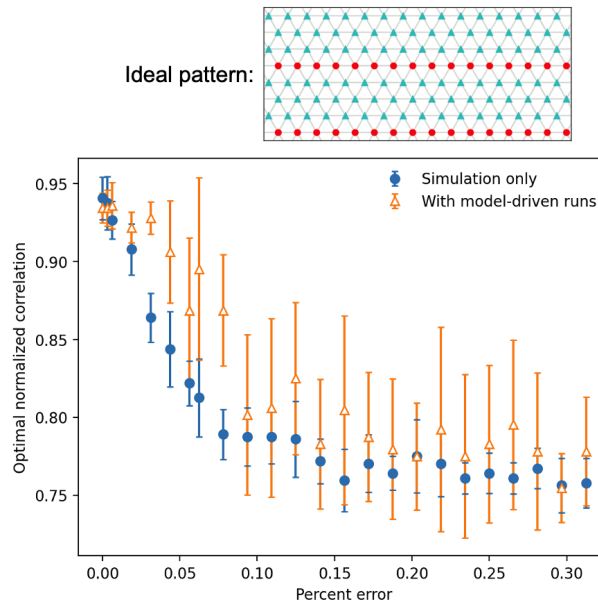


Figure A.14: The genetic algorithm with the simpler objective function (the “ideal” pattern is shown at the top) is run for 200 generations. For the “simulation only” case (blue circles), the objective function evaluation is performed only by directly computing the pattern correlation and adding an error generated from a normal random variable centered at 0 with varying standard deviation, σ . For the “with model-driven runs” case (orange triangles), the algorithm begins with 20 generations where the objective function evaluation is performed as described for the “simulation only” case followed by 180 generations where the objective function is predicted from a model fit from the data from the initial 20 generations. Ten runs for both the “simulation only” and “with model-driven runs” cases are performed at varying values of σ (plotted on the x-axis normalized by 128). The optimal true correlation with an ideal surface found in each run is averaged over the 10 runs (plotted on the y-axis normalized by 128) and the error bars show the standard deviation over those 10 runs. The rest of the genetic algorithm workflow is the same as described, except that the initial random generations do not use the entropy minimization procedure. At non-zero values of error in the objective function, the runs with the surrogate model identify more optimal patterns, on average, than the runs without it.

A.2.4 Generations in each stage

Stage	Fixed f_{OH}	Varying f_{OH}
Random	5	5
First maximization with MD simulations	10	18
First maximization with surrogate model	1000	1000
Second maximization with MD simulations	15	18
Second maximization with surrogate model	1000	2000

Table A.3: The number of generations per stage of the genetic algorithm optimization for the runs fixing and varying the hydroxyl group fraction, f_{OH} . Since the run varying hydroxyl group fraction explores a much larger portion of the pattern space, we increase the number of generations in some of the stages compared to the fixed hydroxyl group fraction run.

A.3 Surrogate function

A.3.1 Features

The surrogate function models the flux difference as a linear function of 22 features of the pattern, listed below:

1. Fraction of nearest neighbors of a functional group that are the same type of group. The nearest neighbors of a functional group are those surrounding it in the hexagonal lattice. Each functional group has 6 nearest neighbors.
2. Fraction of nearest neighbors of hydroxyl groups that are also hydroxyl groups
3. Fraction of nearest neighbors of methyl groups that are also methyl groups
4. Maximum number of hydroxyl groups in a row
5. Maximum number of methyl groups in a row
6. Number of patches comprising 1 or 2 hydroxyl groups. Two hydroxyl groups are in the same patch if they are nearest neighbors.
7. Number of patches of 3, 4, 5, or 6 hydroxyl groups
8. Number of patches of 7 or more hydroxyl groups
9. Number of times two hydroxyl groups are within 1 spacing apart along the pore circumference. Here, a spacing is the distance between two functional groups in hexagonal lattice.
10. Number of times two hydroxyl groups are between 1 and 2 spacings apart along the pore circumference
11. Number of times two hydroxyl groups are between 2 and 3 spacings apart along the pore circumference
12. Number of times two hydroxyl groups are between 3 and 4 spacings apart along the pore circumference
13. Number of times two hydroxyl groups are between 4 and 5 spacings apart along the pore circumference
14. Number of times two hydroxyl groups are between 5 and 6 spacings apart along the pore circumference
15. Number of times two hydroxyl groups are between 6 and 7 spacings apart along the pore circumference

16. Number of times two hydroxyl groups are within 1 spacing apart along the pore axis.
17. Number of times two hydroxyl groups are between 1 and 2 spacings apart along the pore axis.
18. Number of times two hydroxyl groups are between 2 and 3 spacings apart along the pore axis.
19. Number of times two hydroxyl groups are within 1 spacing apart.
20. Number of times two hydroxyl groups are between 1 and 2 spacings apart.
21. Number of times two hydroxyl groups are between 2 and 3 spacings apart.
22. Pattern entropy (calculation described in section A.2.1).

A.3.2 Surrogate model performance

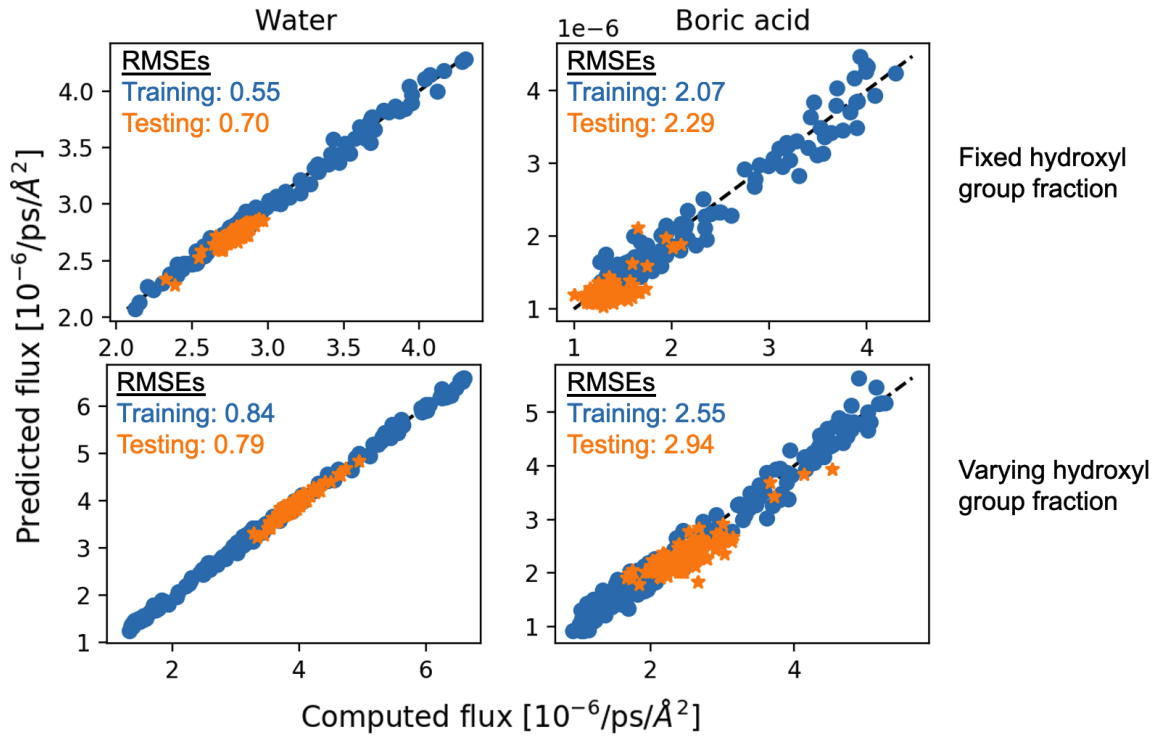


Figure A.15: The surrogate model reasonably predicts water and boric acid fluxes when trained on data from the first set of simulated generations (120 patterns for fixed hydroxyl group fraction, 184 patterns for varying hydroxyl group fraction) and tested on the second set of simulated generations (120 patterns for fixed hydroxyl group fraction, 144 patterns for varying hydroxyl group fraction).

A.3.3 Alternative surrogate models

We compare the model used here (linear, with the set of 22 pattern features as predictors) with alternative surrogate models: LASSO (predictors: pattern features), fully connected artificial neural network (predictors: pattern features), convolutional neural network (predictors: explicit patterns), fully connected artificial neural network (predictors: explicit patterns). We use the dataset generated from the genetic algorithm run fixing hydroxyl group fraction (240 patterns and MD-computed flux differences). Instead of predicting the log of water and boric acid fluxes separately, we perform a single prediction of the flux difference. We use the sklearn package for the models where the predictors are the feature set and keras when the predictors are the patterns. For the latter, we augment the dataset with translational variants of each pattern (128 translational variants per original pattern).

For the LASSO regression, we use a regularization parameter of $\alpha = 0.04$. The artificial neural network (ANN) with feature predictors uses two fully connected layers with 25 nodes per layer. The convolutional neural network (CNN) uses 1 convolutional layer with 4 5x5 filters, 1 dropout layer with a dropout rate of 0.5, and 2 fully connected layers with 10 nodes per layer. The ANN model with pattern predictors uses 4 fully connected layers with 50 nodes per layer.

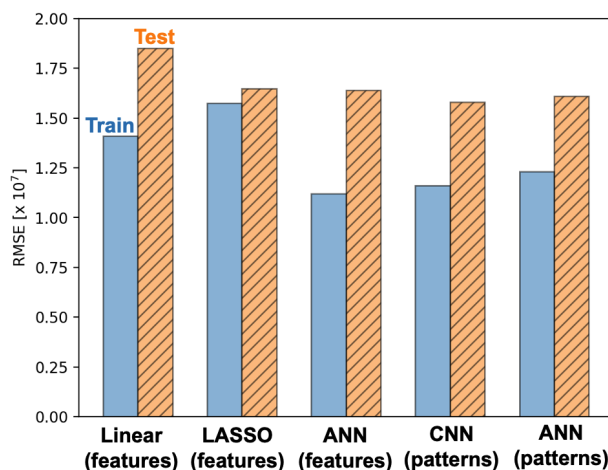


Figure A.16: Comparison of alternative surrogate models shows that performance of neural-network-based models is approximately equal to that of simpler, regularized linear models. For each model, the bars show the average RMSEs computed from 5-fold cross-validation of the training (solid) and testing (hatched) sets.

A.3.4 Feature importance

We use all data from the simulated generations of the fixed hydroxyl group fraction run (240 patterns in total) to evaluate feature importance. We identify the most important features two different ways: (1) random forest regression and (2) LASSO regression. For

both, we randomly divide the dataset into 5 different folds (48 patterns per fold), and run the regression 5 times, each time removing a different fold from the dataset and using the remaining data to fit the models. For random forest regression, we use the sklearn package to fit a random forest regressor and compute feature importance values. Features with an importance over 0.01 are listed in Table A.4. For LASSO regression, we use the sklearn package to fit a LASSO model with $\alpha = 0.05$. The non-zeroed features are listed in Table A.4.

Fit	Random forest	LASSO
1	4, 9, 16, 17, 18, 22	4, 5, 9, 11, 16, 21
2	4, 9, 13, 16, 17, 18, 20, 22	4, 5, 8, 9, 15, 16, 18
3	4, 9, 10, 16, 17, 18, 19, 22	4, 5, 8, 9, 11, 15, 16, 18, 21
4	4, 9, 16, 17, 18, 22	4, 5, 9, 11, 16
5	4, 9, 13, 15, 16, 17, 18, 22	4, 5, 9, 16, 17, 18, 21

Table A.4: Most predictive features, as identified by random forest regression and LASSO regression, for each fit (a different, random fold of the dataset is removed for each fit). Feature numbers correspond to the list in section A.3.1. Bolded numbers are those that appear for each fit for a given regression method.

A.4 Rationally designed patterns

A.4.1 Pattern clustering

We perform spectral clustering of the 240 patterns from the simulated generations of the fixed hydroxyl group fraction run. We pre-compute a 240x240 correlation matrix for all patterns where the correlation between two patterns is the number of hydroxyl groups in the same position (accounting for all translations and reflections of the pattern). We then cluster the patterns using sklearn's spectral clustering implementation, fixing the number of clusters as 6.

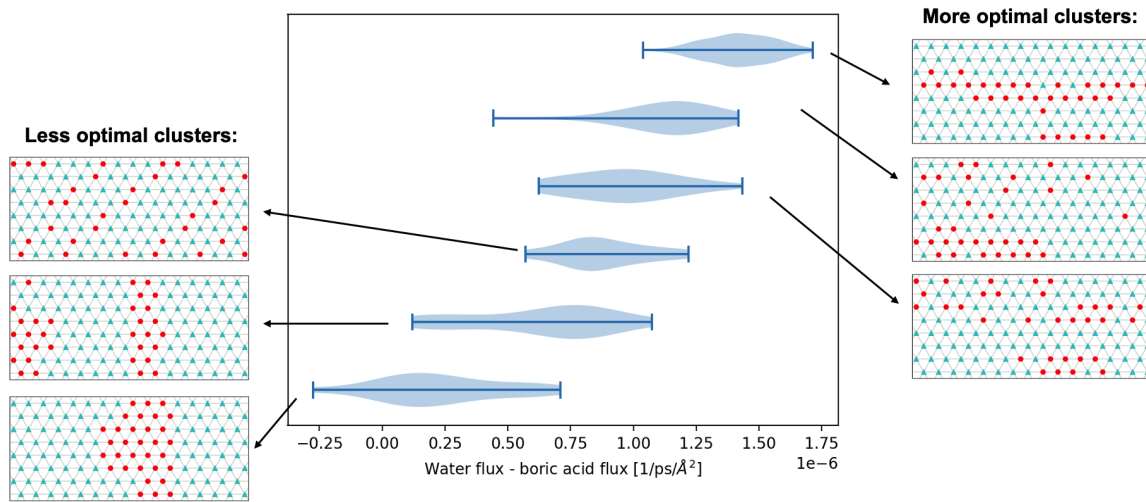
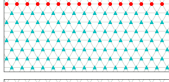
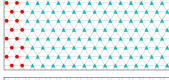
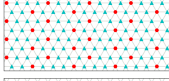
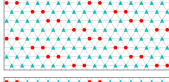
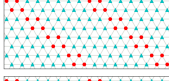
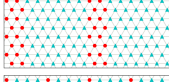
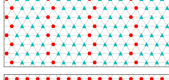
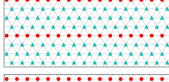
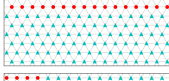
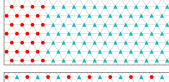
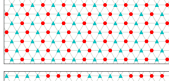
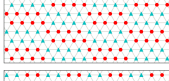
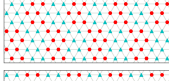
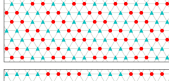
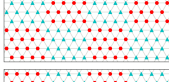
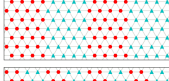
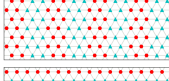
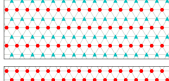
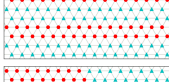
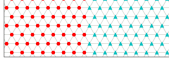


Figure A.17: Spectral clustering of the 240 simulated patterns from the genetic algorithm run with fixed hydroxyl group fraction suggests that patterns with rows of hydroxyl groups and axial patches of hydroxyl groups may maximize and minimize, respectively, the difference between water flux and boric acid flux. Each violin plot represents a separate cluster. One pattern snapshot is shown for each cluster.

A.4.2 Patterns and computed transport properties

Pattern	f_{OH}	K_w	K_b	D_w	D_b	F_w	F_b	$F_w - F_b$
	0	0.4940 ₁	8.0 ₄	0.421 ₁	0.24 ₂	7.61 ₂	5.76 ₆	1.85 ₈
	0.125	0.5310 ₃	4.3 ₃	0.3192 ₄	0.17 ₁	3.44 ₂	2.71 ₅	0.73 ₇
	0.125	0.5234 ₂	5.5 ₄	0.350 ₁	0.199 ₉	4.57 ₂	3.96 ₃	0.61 ₄

	0.125	0.5234 ₂	5.7 ₂	0.357 ₁	0.121 ₂	4.44 ₂	2.58 ₈	1.86 ₇
	0.125	0.5205 ₂	6.0 ₂	0.3676 ₆	0.241 ₉	5.385 ₅	4.69 ₆	0.69 ₆
	0.25	0.5652 ₂	2.1 ₂	0.2528 ₆	0.100 ₆	1.970 ₆	1.51 ₇	0.46 ₇
	0.25	0.5574 ₂	3.2 ₁	0.2822 ₆	0.103 ₇	2.454 ₈	1.55 ₈	0.91 ₉
	0.25	0.5500 ₂	4.2 ₁	0.307 ₁	0.131 ₅	3.191 ₆	2.33 ₁	0.86 ₂
	0.25	0.5469 ₂	4.7 ₁	0.3232 ₄	0.216 ₈	3.71 ₁	3.58 ₆	0.13 ₆
	0.25	0.55259 ₈	3.8 ₂	0.2939 ₆	0.17 ₁	2.79 ₃	2.47 ₅	0.32 ₅
	0.25	0.5530 ₂	4.03 ₉	0.299 ₁	0.081 ₅	2.83 ₂	1.27 ₃	1.55 ₄
	0.25	0.5507 ₃	4.92 ₈	0.3255 ₈	0.065 ₄	3.45 ₂	1.61 ₄	1.84 ₆
	0.25	0.5454 ₂	5.0 ₄	0.3404 ₉	0.22 ₁	4.39 ₃	4.3 ₁	0.13 ₉
	0.5	0.60762 ₇	1.01 ₆	0.2206 ₆	0.092 ₄	1.489 ₃	1.24 ₆	0.25 ₆
	0.5	0.6054 ₁	1.89 ₇	0.2415 ₅	0.068 ₄	1.70 ₁	0.89 ₃	0.82 ₄
	0.5	0.60115 ₆	2.1 ₂	0.2436 ₉	0.096 ₄	1.796 ₅	1.265 ₂	0.532 ₉
	0.5	0.6041 ₂	1.91 ₇	0.2358 ₇	0.073 ₄	1.657 ₉	0.99 ₆	0.66 ₆
	0.5	0.60085 ₆	2.7 ₂	0.257 ₁	0.077 ₅	1.967 ₂	1.10 ₄	0.86 ₄
	0.5	0.5961 ₃	2.7 ₂	0.2762 ₄	0.191 ₃	2.39 ₂	2.56 ₃	-0.18 ₅
	0.5	0.5989 ₄	1.8 ₂	0.2474 ₅	0.150 ₇	1.878 ₄	1.75 ₂	0.13 ₃
	0.5	0.6084 ₂	1.31 ₇	0.2297 ₈	0.105 ₇	1.576 ₆	1.16 ₂	0.41 ₂
	0.5	0.6061 ₂	2.22 ₈	0.2533 ₅	0.061 ₄	1.85 ₁	0.83 ₁	1.010 ₂
	0.5	0.5943 ₂	3.74 ₇	0.2884 ₈	0.22 ₁	2.87 ₁	3.52 ₈	-0.65 ₈

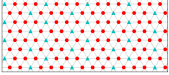
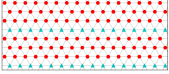
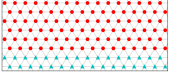
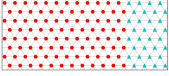
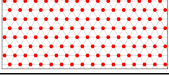
	0.75	0.6553 ₁	1.03 ₉	0.2125 ₇	0.091 ₆	1.302 ₄	1.17 ₂	0.14 ₂
	0.75	0.6534 ₄	1.01 ₁	0.2265 ₄	0.095 ₅	1.388 ₆	1.12 ₄	0.27 ₄
	0.75	0.6484 ₂	1.27 ₅	0.2420 ₆	0.071 ₅	1.582 ₇	1.04 ₁	0.54 ₁
	0.75	0.64309 ₅	1.8 ₂	0.2503 ₈	0.16 ₁	1.793 ₆	2.16 ₄	-0.37 ₄
	1	0.6902 ₃	0.743 ₇	0.2297 ₄	0.109 ₅	1.337 ₇	1.56 ₄	-0.23 ₃

Table A.5: List of all rationally designed patterns and computed transport properties for water and boric acid, including partition coefficients K_w and K_b , diffusivities D_w and D_b , and fluxes F_w and F_b . The subscripts w and b stand for water and boric acid, respectively. The units for diffusivity are $\text{\AA}^2/\text{ps}$ and the units for fluxes are $10^{-6}/\text{ps}/\text{\AA}^2$.

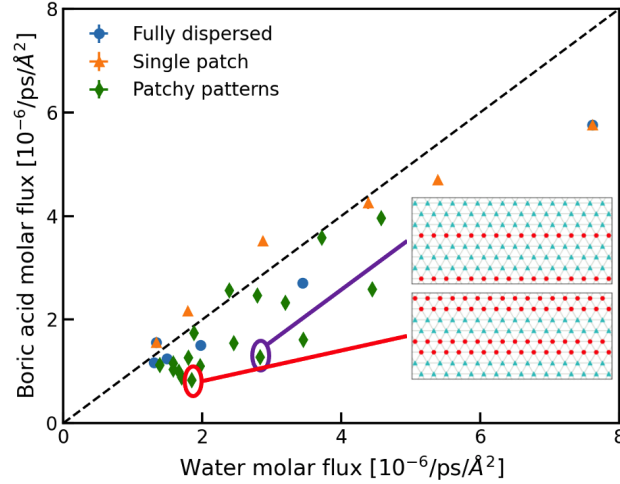


Figure A.18: Unlike evenly dispersed and single-patch patterns, other non-intuitive patterns can reduce boric acid flux without affecting water flux to the same extent.

A.4.3 Diffusivity of water and boric acid

To compute diffusivities, we divide the $1\ \mu\text{s}$ NVT production simulation into 5 blocks of 200 ns each. Within each block, we compute the mean squared displacement trajectories of the water oxygens and boron atoms over 1 ns intervals. We fit a line to the region above 100 ps and compute the diffusivity as half of the slope. We take the mean of the diffusivities fit for each of the 5 blocks to get the reported diffusivity. The reported uncertainty is the standard error of the mean.

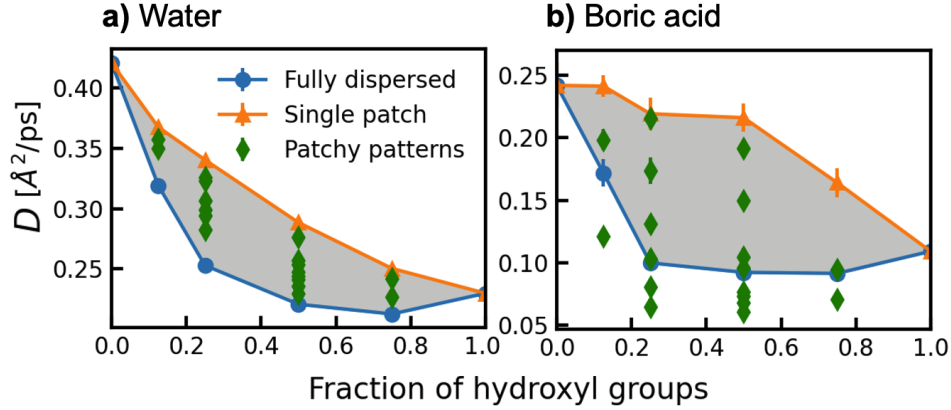


Figure A.19: Diffusivities of (a) water and (b) boric acid for all rationally designed patterns. Note that some rationally designed patterns give lower boric acid diffusivities compared to the fully dispersed surfaces, in contrast to the case for water, producing differential effects on the behavior of boric acid relative to water.

A.4.4 Analytical model for boric acid transport

We use the model derived in Ref. [133] for the 1-D diffusion coefficient, D , of a Brownian particle in an arbitrary periodic potential, $V(x)$, with no tilting force:

$$\frac{D}{D_0} = \frac{\int_{x_0}^{x_0+L} \frac{dx}{L} I_+^2(x) I_-(x)}{\left[\int_{x_0}^{x_0+L} \frac{dx}{L} I_+(x) \right]^3} \quad (\text{A.2})$$

$$I_{\pm}(x) = \int_0^L \frac{dy}{D_0} \exp([\pm V(x) \mp V(x \mp y)]/k_B T) \quad (\text{A.3})$$

Where D_0 is the diffusion coefficient when $V(x)$ is constant and L is the period of $V(x)$, so that $V(x) = V(x + L)$.

For simplicity, we model $V(x)$ as a stepwise function, as depicted in Fig. A.20, taking on a constant value V_0 in the methylated region of length a , and another constant value V_1 in the hydroxylated region of length b . The lengths of the regions are proportional to the number of rows of each group (e.g. in Fig. A.20, $a/b = 3$). We furthermore assume that $\exp(-(V_0 - V_1)/k_B T) = K_0/K_1$, where K_0 is the partition coefficient in the methylated region and K_1 is the partition coefficient in the hydroxylated region. Then, the diffusivity relationship simplifies to:

$$\frac{D}{D_0} = \frac{\gamma(\gamma + (1 - \gamma)\phi)^2(\gamma + (1 - \gamma)/\phi) + (1 - \gamma)(1 - \gamma + \gamma/\phi)^2(1 - \gamma + \gamma\phi)}{[\gamma(\gamma + (1 - \gamma)\phi) + (1 - \gamma)(1 - \gamma + \gamma/\phi)]^3} \quad (\text{A.4})$$

Where $\gamma = a/(a + b)$ and $\phi = K_0/K_1$.

Finally, to empirically account for the reduction in diffusivity with increasing fraction

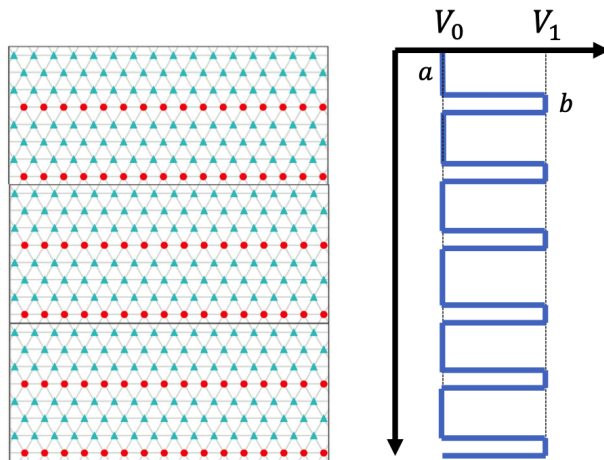


Figure A.20: $V(x)$ takes on two distinct values in the methylated and hydroxylated regions.

of hydroxyl groups, we model D_0 as an exponential: $D_0 = A \exp(-B f_{OH}) + C$ where f_{OH} is the fraction of hydroxyl groups and the constants $A = 0.1803$, $B = 1.3380$, and $C = 0.0618$ are fit from the computed diffusivities as $f_{OH} = 0$, $f_{OH} = 1$, and the average of diffusivities of the single-patch and dispersed patterns at $f_{OH} = 0.5$.

A.4.5 Other solutes

Parameters for solutes besides boric acid and arsenous acid are from the GAFF2 force-field³⁶⁰ with AM1-BCC charges.³⁶¹ For isopropanol and phenol charges are scaled according to the method described in Ref. [362], as was performed in Ref. [79]. For arsenous acid, non-bonded LJ parameters are from the GAFF2 forcefield, while non-bonded LJ parameters and bonded parameters are from the DREIDING forcefield,³⁶³ with unscaled 1-4 interactions and the harmonic angle potential functional form. We compute partial charges for arsenous acid by running a geometry optimization and then a single-point electrostatic potential (ESP) calculation in Gaussian 16³⁶⁴ using the B3LYP functional with the 6-311++G(d,p) basis set. We set the spin multiplicity to 1 for this calculation. We then use AmberTools16³⁶⁰ to perform a restrained electrostatic potential fitting to fit partial charges for each arsenous acid atom.

Solute	N_H	$\Delta G_{solv}^{CH_3} - \Delta G_{solv}^{OH} [k_B T]$	$\Delta F_{ring} - \Delta F_{methyl} [10^{-6}/ps/\text{\AA}^2]$
Benzene	0	-3.22	2.64
Phenol	1.15	-3.39	2.35
Isopropanol	1.49	-2.65	0.54
Boric acid	1.27	-0.86	0.011
Ammonia	0.76	-0.73	-0.21
Arsenous acid	1.53	N/A	-0.24

Table A.6: For each solute simulated here, the average number of hydrogen bonds formed with the nanopore hydroxyl groups in the fully hydroxylated pore N_H , the difference in affinity for methylated and hydroxylated self-assembled monolayer (SAM) surfaces as measured in Ref. [79], and the enhancement in solute rejection with the ringed pore $\Delta F_{ring} - F_{methylated}$. The average number of hydrogen bonds with hydroxyl groups in the fully hydroxylated pore is computed from an equilibrium simulation in an infinite pore with the given solute (same workflow as described in the main text for boric acid) and is computed over snapshots for which the solute is within $5\text{\AA}$ of the pore wall. Hydrogen bonds are determined *via* a geometric criteria: the distance between the donor and acceptor oxygen atoms must be less than $3.5\text{\AA}$ and the angle formed by the acceptor oxygen – donor oxygen – donor hydrogen must be less than 30° . The enhancement in solute rejection with the ringed pore is computed from subtracting the flux difference (water flux minus solute flux) in the methylated pore, F_{methyl} , from the flux difference in the pore with a single ring of hydroxyl groups (bottom-right pore depicted in Fig. 2.4), F_{ring} . The chosen solutes show a large range of hydrogen-bonding and surface affinity behavior. The difference between their affinities for the methylated and hydroxylated SAMs is correlated with the enhancement in rejection by the ringed pores, as is consistent with the proposed diffusive hopping mechanism.

A.5 Permeability-selectivity trade-off

As described in section 2.2, water permeability is computed from:

$$P = \frac{K_w F_w V_{\text{pore}} \rho_{\text{bulk}}}{dP/dz} \quad (\text{A.5})$$

where K_w is the partition coefficient of water, F_w is the in-pore flux of water, V_{pore} is the pore volume, ρ_{bulk} is the bulk water density, and dP/dz is the pressure gradient (the relationship between the pressure gradient and the applied acceleration is given in section A.1.8).

Selectivity is computed from:

$$S = \frac{K_w F_w}{K_b F_b} \frac{1}{1.4 \times 10^{-6}} \quad (\text{A.6})$$

where K_b is the partition coefficient of boric acid, F_b is the in-pore flux of boric acid, and 1.4×10^{-6} is the approximate mole fraction of boric acid in seawater from Ref. [92].

In Fig. A.21, the Pareto front (shaded gray) shows the set of all optimal combinations of permeability and selectivity from 20 new randomly generated pores (without pattern entropy minimization, i.e. distinct from those used in the genetic algorithm optimization). Here, the Pareto front, an approach to quantify improvement in multi-objective optimization,^{78,149} defines pore patterns that lie on the extremes of higher permeability and selectivity. The front crosses from the top-left corner (high selectivity and low permeability) to the bottom right (low selectivity and high permeability), showing that, for randomly generated pores, improvements in selectivity are at the expense of permeability. We compute the permeability and selectivity for the optimal patterns along the genetic algorithm optimization trajectories (Fig. 2.2, red lines) and determine new Pareto fronts, adding each of the optimal patterns to the original set of 20 randomly generated patterns. For the optimization runs, the best surfaces in terms of flux difference also expand the Pareto front, indicating that the genetic algorithm optimization of a single objective function discovers surfaces that are more optimal in a related, multi-dimensional space, i.e. with more optimal permeability-selectivity behavior. While not directly related, the log of the selectivity contains a term that is the difference of the log of the in-pore fluxes, along with a term related to the difference in partition coefficients and a constant term: $\log S = (\log F_w - \log F_b) + (\log K_w - \log K_b) - \log(1.4 \times 10^{-6})$. Thus, larger flux differences will lead to improved selectivity, as long as the resulting difference in partition coefficients does not offset it.

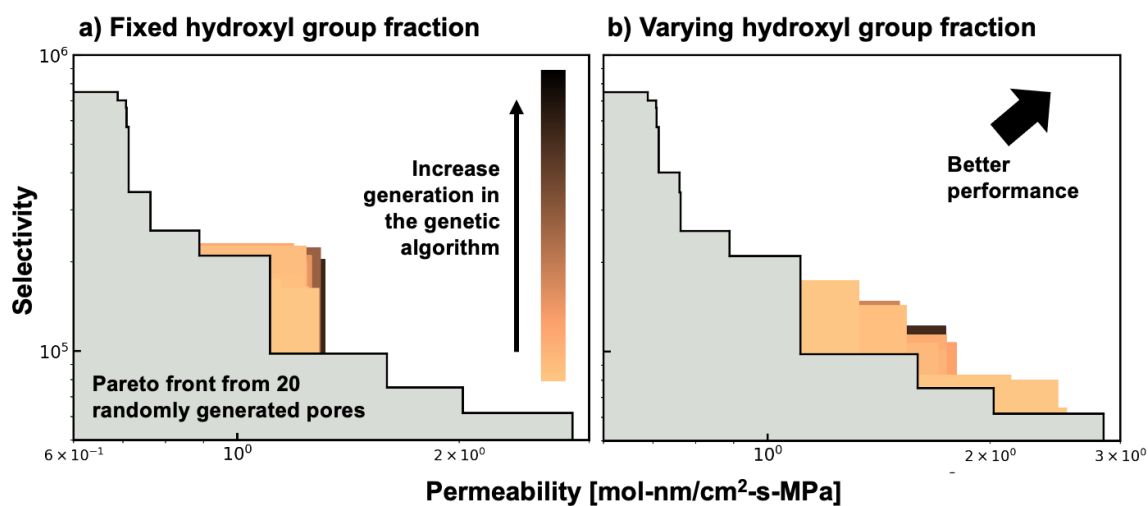


Figure A.21: For both the genetic algorithm runs with (a) hydroxyl group fraction fixed at 0.25 and (b) varying hydroxyl group fraction, the patterns along the running optimal trajectory (red line in Fig. 2.2) extend the Pareto front (borders of tan rectangles) beyond that created from randomly generated pores (black line with gray shading). The darker shades correspond to patterns discovered later in the optimization.

Appendix B

Supporting Information for Inverse Design of Chemistry and Topology

B.1 Simulation details

Tables B.1 and B.2 give extra bond and angle parameters required to simulate the ethyl and hydroxymethyl groups, derived from the OPLS-AA/L forcefield. The remaining bond and angle parameters can be found in Tables A.1 and A.2. As with the methyl and hydroxyl groups, we set all dihedrals involving the functional groups to zero.

The charges for the hydroxymethyl O-H group are the same as the hydroxyl group, with the remaining charge split between the CH₂ carbons (no charge for the CH₂ carbon). For the ethyl group, all hydrogens have charge 0.06 (as with the methyl group), which are balanced by the charges on the carbons bonded to the hydrogens (the CH₂ carbon has charge −0.12 and the CH₃ carbon has charge −0.18).

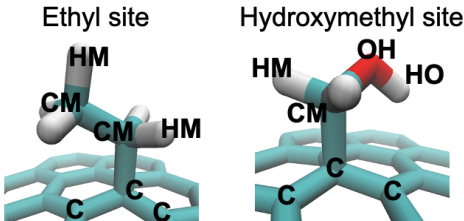


Figure B.1: Labeling of atoms in ethyl and hydroxymethyl sites for Tables B.1 and B.2.

Bond	b_0 [nm]	K_b [kJ/mol/nm ²]	Source
CM-CM	0.1529	224262.4	OPLS-AA/L CT-CT
CM-OH	0.141	267776	OPLS-AA/L CT-OH

Table B.1: Harmonic bond parameters for nanopore functional groups. The bond potential has the form $U(r_{ij}) = \frac{1}{2}K_b(r_{ij} - b_0)^2$.

Angle	θ_0 [deg]	K_θ [kJ/mol/rad ²]	Source
C-CM-CM	112.7	488.273	OPLS-AA/L CT-CT-CT
CM-CM-HM	110.7	313.8	OPLS-AA/L CT-CT-HC
CM-OH-HO	108.5	460.24	OPLS-AA/L CT-OH-HO
HM-CM-OH	109.5	292.88	OPLS-AA/L HC-CT-OH
C-CM-OH	109.5	418.4	OPLS-AA/L CT-CT-OH

Table B.2: Harmonic angle parameters for nanopore functional groups. The angle potential has the form $U(r_{ijk}) = \frac{1}{2}K_\theta(\theta_{ijk} - \theta_0)^2$.

B.2 Visualizing solute locations in pore

One helpful visualization of solute transport through the pore is to project the locations of the solute (boric acid) onto the plane of the pore wall (pore circumference versus pore axis). Fig. B.2a shows examples of this projection for the nonpolar pores with (i) only methyl and (ii) only ethyl groups. Here, we only include boric acid locations when boric acid is at least $10\text{\AA}$ from the carbon nanotube wall. We use the location of the boron atom. The projected coordinates along the pore circumference, x' , and pore axis, y' are (assuming the center of the pore is located at $(x_c, y_c, z_c) = (0, 0, 0)$):

$$x' = R \cos^{-1} \left(\frac{x}{\sqrt{x^2 + y^2}} \right) \frac{y}{|y|} \quad (\text{B.1})$$

$$y' = z \quad (\text{B.2})$$

where x , y , and z are the x -, y -, and z -coordinates of the boron atom and R is the radius of the nanopore. We also compute projected coordinates of the functional groups (specifically, the CH_3 carbon or hydroxyl oxygen atoms) using the same equations for the first saved frame of the simulation.

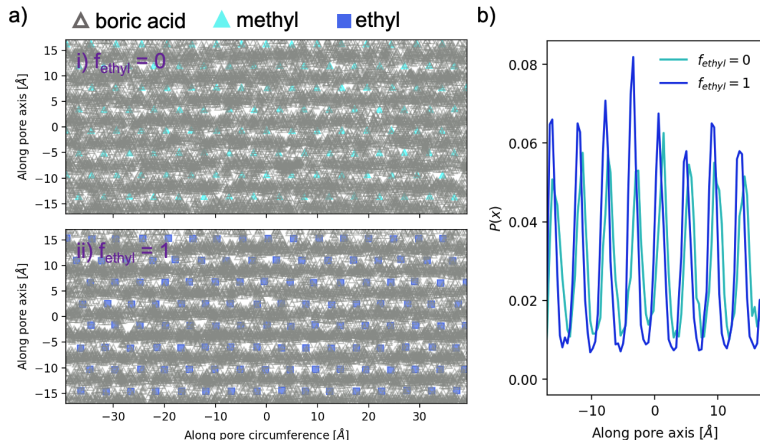


Figure B.2: (a) Projection of locations of boric acid solute within the pores with (i) only methyl groups ($f_{\text{ethyl}} = 0$) and (ii) only ethyl groups ($f_{\text{ethyl}} = 1$) onto the pore wall. Gray triangles are boric acid, cyan triangles are methyl group locations, and blue squares are ethyl group locations. (b) Histograms of boric acid locations along the pore axis.

These projected coordinates reveal precise locations along the pore wall where boric acid preferentially adsorbs. For instance, Fig. B.2a suggests that boric acid prefers to reside in between the rows of functional groups, and that there is a slightly stronger preference for being between rows of ethyl groups compared to rows of methyl groups. The histogram of boric acid locations along the pore axis, shown in Fig. B.2b confirms this, with higher peaks and lower wells for the pore with only ethyl groups. This enhanced

preference for residing in between ethyl groups could be due to the increased length and thus flexibility of the ethyl groups, allowing for rearrangement to accommodate the solute. Thus, there is a slightly higher energetic barrier for boric acid to transport down the axis of the pore with rows of ethyl groups compared to rows of methyl groups, which may account for the slight reduction in flux in Fig. 3.2.

Fig. B.3 are similar projections for pores with both ethyl and methyl groups. Boric acid generally prefers to reside near methyl groups, compared to ethyl groups. In the pore with 50% ethyl groups, the arrangement of the methyl groups creates “channels” of boric acid groups, one possible reason that flux in these pores is unexpectedly high (Fig. 3.2). In contrast, with larger fractions of ethyl groups, the boric acid resides in pockets and needs to hop between pockets to transport down the pore.

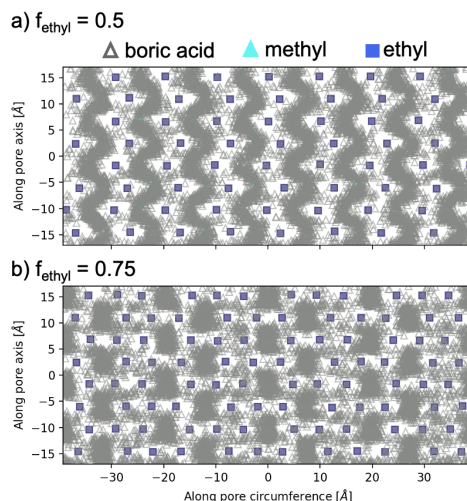


Figure B.3: Projection of locations of boric acid solute onto the pore wall for pores with (a) 50% ethyl groups and (b) 75% ethyl groups, arranged in a fully dispersed pattern. Gray triangles are boric acid, cyan triangles are methyl group locations, and blue squares are ethyl group locations.

B.3 Water properties

Fig. B.4 shows the water density profiles along the pore radius for homogeneous pores (i.e. pores with a single type of functional group). All functional groups induce some change in the structure of water from bulk. The pores with shorter groups (compare methyl (cyan) to ethyl (blue) or compare hydroxyl (red) to hydroxymethyl (magenta)) lead to a higher initial peak in the density profile, indicating stronger structuring of water. A possible reason for this is that the longer groups are more flexible and therefore allow water to reside at more locations with respect to the pore wall.

Fig. B.5a shows the three-body angle distributions for water in the pores. The three-body angle is formed by the oxygen atoms of a central water molecule and two of its

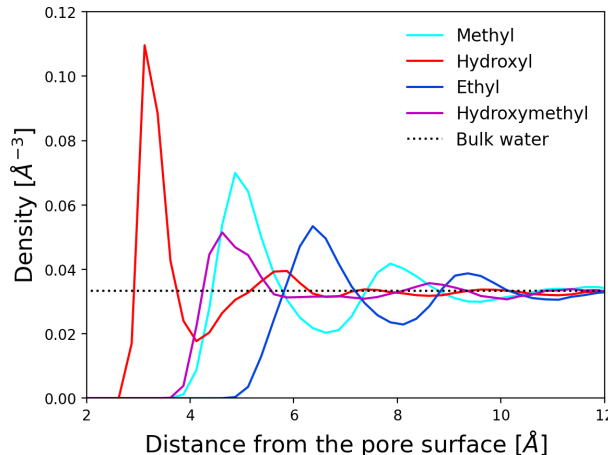


Figure B.4: Density of water as a function of distance from the carbon nanotube wall for various homogeneous pores. The label indicates the type of functional group in the pore. The profile is computed with a bin width of $0.25\text{\AA}$. The dotted black line shows the bulk density of water, $0.0334\text{\AA}^{-3}$.

nearest neighbors (the cutoff for a nearest neighbor used here is $3.4\text{\AA}$). The distributions are computed for all waters within $7\text{\AA}$ of a heavy atom on the pore wall. They are shifted by the distribution for bulk TIP3P water, computed from a simulation with 884 TIP3P water molecules. For the bulk water simulation, we first run an NVT equilibration for 100 ps in a box of length 3 nm. We then perform an NPT equilibration for 100 ps at 1 bar. We perform a second NVT equilibration for 100 ps using the average box length from the second half of the NPT equilibration. Finally, we run an NVT simulation for 1 μs , saving configurations every 20 ps.

In order to characterize shifts in the distribution with respect to the fraction of ethyl groups in the pore, f_{ethyl} , we compute the projections onto the principal components shown in Fig. 6.6. Fig. B.5b shows the projections as a function of f_{ethyl} . As discussed further in section 6.3.3, the first principal component measures the “small-lengthscale” tetrahedral behavior of water, the second component measures the “large-lengthscale” ideal-gas-like behavior, and the third component measures the degree to which the hydration shells are over-coordinated. The projections show that water’s tetrahedrality reaches at maximum at $f_{\text{ethyl}} = 0.5$.

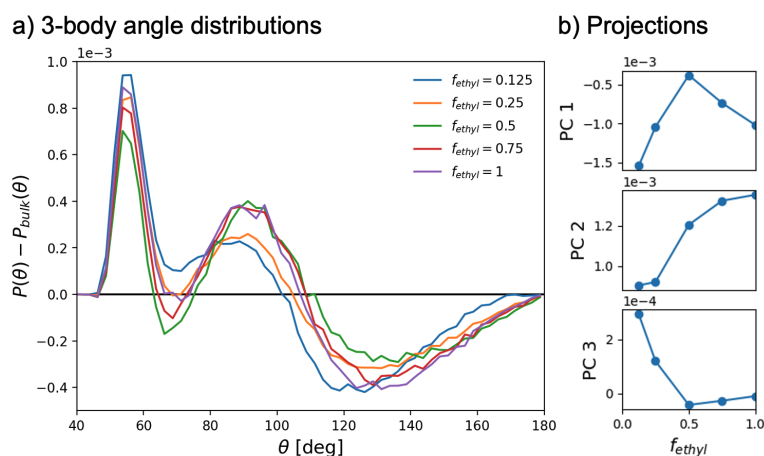


Figure B.5: (a) Three-body angle distributions for water in pores with varying fractions of ethyl groups (f_{ethyl}) dispersed evenly across the pore wall. The distributions are shifted by the distribution for bulk TIP3P water. (b) Projections of the distributions onto the first, second, and third principal components shown in Fig. 6.6, as a function of the fraction of groups that are ethyl groups.

Appendix C

Supporting Information for Quantifying Polypeptoid Conformational Landscapes

C.1 Simulation workflow details

C.1.1 Parameterization

Partial charge assignment

For the partial charges computed using ESP calculations (hydrophilic monomers for the modified AMBER forcefield and spin-label monomers for the modified AMBER and GAFF- ϕ forcefields), we create the structures in Avogadro 1.2.0²¹³ and perform a geometry minimization. Then, we perform a geometry optimization and then a single-point calculation in Gaussian 16³⁶⁴ using the default SCF procedure and outputting charges derived from the ESP to an Antechamber file using 8 points per unit area and 7 layers. The spin multiplicity is set to 1 for homopolymers of the hydrophilic monomers (a spin multiplicity of 3 results in higher energies) and 2 for calculations with the spin label (a spin multiplicity of 4 results in higher energies).

To obtain charges for the hydrophilic monomers, we then use RESP fitting to fit partial charges on 8-mer homopolymers of the hydrophilic monomers. Then, we use the prepgen tool in AmberTools16³⁶⁰ to adjust the charges so that each monomer is net neutral. We average the charges for all monomers except the N-terminal monomer. This procedure is different from the charging procedure for the modified AMBER forcefield in Ref. [209], which involves performing ESP and RESP calculations on an entire polypeptoid and is intractable for longer polypeptoids. We repeat this calculation varying the length of the homopolymer and find that the average charges converge by 8 monomers (Fig. C.1 and C.2). We also repeat the calculation for shorter homopolymers using the B3LYP functional with the 6-311++G(d,p) basis set and find convergence to similar values (Fig. C.1 and C.2). Charges for the C-terminal amino group are the charges fit for the group in the homopolymer of 8 NME monomers.

For the N-terminal spin-label monomer, charges for a twomer with the spin-label monomer at the N-terminus and the NME monomer at the C-terminus are derived from an ESP calculation using the procedure above with the B3LYP functional and the 6-311++G(d,p) basis set and then fit with RESP. Then, prepgen is used to adjust charges so that each monomer is net neutral. The same procedure is followed for the C-terminal spin-label monomer, except with a twomer with the NME monomer at the N-terminus and the spin-label monomer at the C-terminus.

For the hydrophilic monomers for the GAFF- ϕ forcefield, we create twomers with an N-terminal alcohol monomer and the monomer of interest at the C-terminus in Avogadro and perform a geometry minimization. Partial charges for the twomer are assigned with the AM1-BCC model using AmberTools16. Then, we use the prepgen tool in AmberTools16 to adjust the charges so that each monomer is net neutral. Charges for the C-terminal amino group are fit by fitting charges for a twomer of the alcohol monomer, after which charges are adjusted for the amino group so that it is net neutral.

The MFTOID forcefield and the modified CGenFF forcefield share the same partial charges. The charge on the backbone nitrogens are set to -0.42, the value fit in Ref.

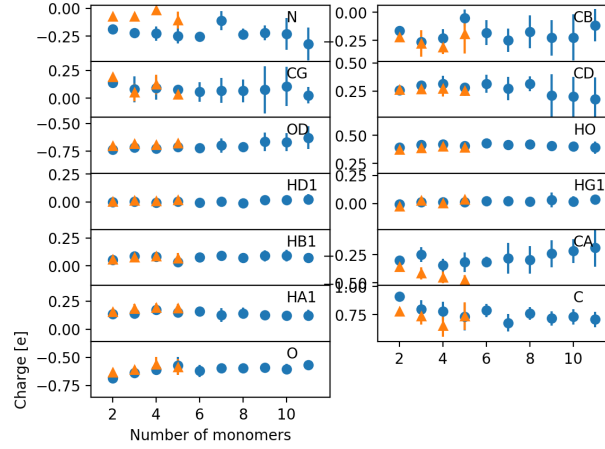


Figure C.1: Each plot shows the average partial charge on an atom (names correspond to those in Fig. C.3) in homopolymers of the alcohol monomer of varying length. The blue circles correspond to calculations using Hartree-Fock level of theory with the 6-31G(d) basis set, while the orange triangles correspond to calculations using the B3LYP functional with the 6-311++G(d,p) basis set. The error bars show the standard deviation in the partial charge over the homopolymer. As the number of monomers increases, the average converges, but the longer homopolymers have access to more conformations, therefore increasing the standard deviation. Here, we use the partial charges computed for a homopolymer with 8 monomers. The y-axis for each plot is $0.5e$ long. The x-axis is the total number of monomers in the homopolymer, but averages are taken excluding the N-terminal monomer.

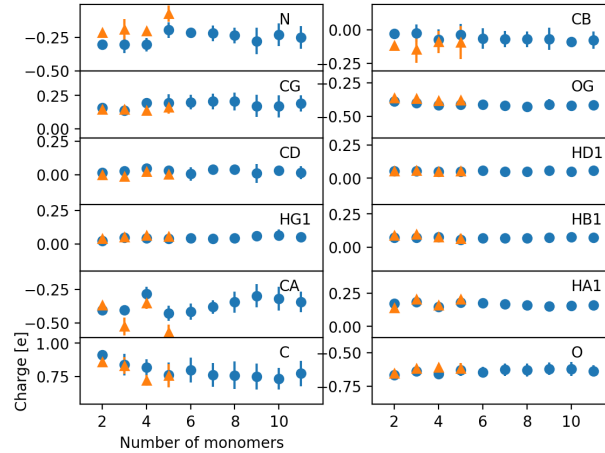


Figure C.2: The same as Fig. C.1, but for the NME monomer.

[²⁰⁶]. For the alcohol monomer, charges for the alcohol group, adjacent methylene group, and backbone carbonyl group are set to the same as those for serine in the CHARMM22 forcefield. Charges for the next methylene group (CG in Fig. C.3) are set to charges for the methylene group in the same position in proline in the CHARMM22 forcefield.

Charges for the remaining hydrogens are set to 0.09, to be consistent with charges on other methylene hydrogens in the amino acids for the CHARMM22 forcefield, and the remaining charge is divided evenly between the backbone methylene and first sidechain methylene (CA and CB, respectively, in Fig. C.3) to make the monomer net neutral. For the NME monomer, the backbone charges and first sidechain methylene (CB in Fig. C.3) charges are set to the same as for the alcohol monomer. The ether group charges (CG, OG, CD and associated hydrogens in Fig. C.3) are set to those for linear ether groups defined in Ref. [365]. For the spin label monomer, the backbone charges are set to the same as for the alcohol monomer. Charges for the ring atoms, excluding the one bonded to the backbone nitrogen (CB in Fig. C.3), are the same as in Ref. [185]. Charges for all hydrogens are 0.09, as in Ref. [185]. The net charge in the sidechain ring, excluding CB and its hydrogen (HB1 in Fig. C.3), are divided evenly among the methyl groups, giving a charge of -0.3305. The net charge for the backbone CB is assigned to CB, giving a charge of 0.12. For the N-terminal monomers, the charge for the N-terminal hydrogen is set to 0.33, to be consistent with the charge for an N-terminal hydrogen for peptides in CHARMM22 (defined for NH_3^+ terminal groups). The charge on the backbone nitrogen is reduced to -0.75 to keep the monomer net neutral. For the C-terminal amino group, the charges on the hydrogen are set to 0.33, the same as the N-terminal hydrogen, and the charge on the nitrogen is set to 0.66 to make the group net neutral.

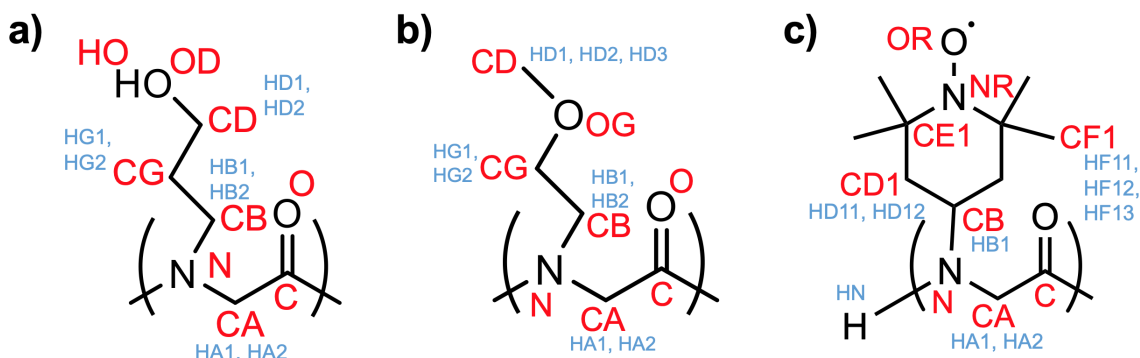


Figure C.3: Atom names for the (a) alcohol (b) NME and (c) spin label monomer, corresponding to entries in Tables C.1-C.5.

Atom	Modified AMBER	MFTOID / modified CGenFF	GAFF- ϕ
C	0.760	0.51	0.641
O	-0.597	-0.51	-0.605
CA	-0.294	0.03	0.018
HA*	0.143	0.09	0.090
N	-0.233	-0.42	-0.469
CB	-0.174	0.03	0.094
HB*	0.073	0.09	0.061
CG	0.067	-0.18	-0.144
HG*	0.018	0.09	0.059
CD	0.316	0.05	0.131
HD*	-0.010	0.09	0.047
OD	-0.712	-0.66	-0.583
HO	0.420	0.43	0.403

Table C.1: Partial charges for atoms on the alcohol monomer used with each polypeptoid forcefield. Atom names are shown in Fig. C.3. Asterisks signal that there are multiple atoms with the same charge (e.g. HA1 and HA2).

Atom	Modified AMBER	MFTOID / modified CGenFF	GAFF- ϕ
C	0.759	0.51	0.644
O	-0.629	-0.51	-0.605
CA	-0.344	0.03	0.017
HA*	0.160	0.09	0.091
N	-0.235	-0.42	-0.466
CB	-0.069	0.03	0.087
HB*	0.071	0.09	0.073
CG	0.206	-0.01	0.106
HG*	0.043	0.09	0.038
OG	-0.429	-0.34	-0.417
CD	0.039	-0.1	0.120
HD*	0.051	0.09	0.036

Table C.2: Same as Table C.1 but for the NME monomer.

Atom	Modified AMBER / GAFF- ϕ	MFTOID / modified CGenFF
C	0.820	0.51
O	-0.635	-0.51
CA	-0.351	0.03
HA*	0.130	0.09
N	0.004	-0.42
CB	0.015	0.12
HB1	0.105	0.09
CD*	-0.484	-0.237
HD*	0.106	0.09
CE*	0.633	0.284
CF*	-0.330	-0.3305
HF*	0.073	0.09
NR	-0.079	0.167
OR	-0.413	-0.379

Table C.3: Same as Table C.1 but for the C-terminal spin-label.

Atom	Modified AMBER / GAFF- ϕ	MFTOID / modified CGenFF
C	0.698	0.51
O	-0.565	-0.51
CA	-0.094	0.03
HA*	0.082	0.09
N	0.049	-0.75
HN	0.334	0.33
CB	0.361	0.12
HB1	0.033	0.09
CD*	-0.672	-0.237
HD*	0.181	0.09
CE*	0.557	0.284
CF*	-0.257	-0.3305
HF*	0.056	0.09
NR	0.049	0.167
OR	-0.432	-0.379

Table C.4: Same as Table C.1 but for the N-terminal spin-label.

Atom	Modified AMBER	MFTOID / modified CGenFF	GAFF- ϕ
N	-0.974	-0.66	-0.644
H*	0.487	0.33	0.321

Table C.5: Same as Table C.1 but for the C-terminal amino group.

Atom type assignment

For the modified AMBER and GAFF- ϕ forcefields, we perform atom type assignment to types in the AMBER99SB-ILDN and GAFF v1.8 forcefields, respectively, using AmberTools16.³⁶⁰ All atom types listed in Tables C.6-C.9 are from these forcefields, unless otherwise noted. Stendardo *et al.*³⁶⁶ provides atom types for the spin labels for the AMBER forcefield. For the MFTOID²⁰⁶ forcefield, we perform atom type assignment to types in the CHARMM22²⁴⁶ and CGenFF3.0.1³⁶⁷ (downloaded 12/10/2018) forcefields (and one new type, TC, from MFTOID²⁰⁶) using Mirijanian *et al.*,²⁰⁶ analogous amino acids, ParamChem,^{368,369} and Sezer *et al.*¹⁸⁵ for the spin label. For the modified CGenFF forcefield, we perform atom type assignment to types in the CGenFF4.1 (downloaded 10/24/2019) or CHARMM22 forcefields (and one new type, NTOID, from Weiser *et al.*²¹⁰) using ParamChem,^{368,369} Weiser *et al.*,²¹⁰ and Sezer *et al.*¹⁸⁵ for the spin label.

Atom	Modified AMBER	GAFF- ϕ	MFTOID	modified CGenFF
C	C	c	TC	CG2O1
O	O	o	O	OG2D1
CA	CT	c3	CT2	CG321
HA*	H1	h1	HB	HGA2
N	N	n	NH1	NTOID
CB	CT	c3	CT2	CG321
HB*	H1	h1	HA	HGA2
CG	CT	c3	CT2	CG321
HG*	HC	hc	HA	HGA2
CD	CT	c3	CT2	CG321
HD*	H1	h1	HA	HGA2
OD	OH	oh	OH1	OG311
HO	HO	ho	H	HGP1

Table C.6: Atom types for atoms on the alcohol monomer used with each polypeptoid forcefield. Atom names are shown in Fig. C.3. Asterisks signal that there are multiple atoms with the same assignment (e.g. HA1 and HA2). All atom types for the MFTOID forcefield are from the CHARMM22 forcefield, except for TC, which is a new atom type from Ref. [206]. For the MFTOID forcefield, we assigned the types of the backbone atoms (C, O, CA, HA*, and N) using the analogous atom in disarcosine in Ref. [206] and the sidechain atoms using the analogous atoms in the serine residue of CHARMM22. All atom types for the modified CGenFF forcefield are from the CGenFF forcefield and are assigned using ParamChem, except for NTOID, which is a new atom type from Ref. [210].

Atom	Modified AMBER	GAFF- ϕ	MFTOID	modified CGenFF
CG	CT	c3	CT2	CG321
HG*	H1	h1	HA	HGA2
OG	OS	os	OG301	OG301
CD	CT	c3	CT3	CG331
HD*	H1	h1	HA	HGA3

Table C.7: Same as Table C.6 but for the NME monomer. The backbone atom types (atoms C, O, CA, HA*, N) along with atoms CB and HB* are the same as in Table C.6 and so are not shown. All atom types for the MFTOID forcefield are from the CHARMM22 forcefield, except for OG301, which is an atom type from the CGenFF forcefield and is assigned using ParamChem. The other types are assigned to be consistent with other sidechain CH₂ groups. All atom types for the modified CGenFF forcefield are from the CGenFF forcefield and are assigned using ParamChem.

Atom	Modified AMBER	GAFF- ϕ	MFTOID	modified CGenFF
N	NT/N	n3/n	NH1	NTOID
HN	H	hn	HC	HGPAM1
CB	CT	c3	CT2	CG311
HB1	H1	h1	HA	HGA1
CD*	CT	c3	CTL2	CTL2
HD*	HC	hc	HAL2	HAL2
CE*	CT	c3	CTL1	CTL1
CF*	CT	c3	CTL3	CTL3
HF*	HN	hc	HAL3	HAL3
NR	NN	n3	NN	NN
OR	ON	o	ON	ON

Table C.8: Same as Table C.6 but for the spin-label monomer. The HN atom only appears in the N-terminal spin label. For the modified AMBER and GAFF- ϕ forcefields, the atom type for the nitrogen N in the N-terminal spin label is distinct (NT and n3, respectively) from the atom type for the nitrogen in the C-terminal spin label (N and n, respectively). The atom types for C, O, CA, and HA* are the same as in Table C.6 and so are not shown. The atom types for the MFTOID forcefield are from the CHARMM22 forcefield, except for NN and ON which are from Ref. [185]. N is assigned NH1 to be consistent with other peptoid nitrogens and HN is assigned HC because that is the CHARMM22 type for hydrogens in the N-terminal group. CB and HB1 are assigned CT2 and HA to be consistent with other sidechains, and the rest of the atoms are assigned using the analogous atoms in Ref. [185]. For the CGenFF forcefield, N is assigned NTOID to be consistent with other peptoid nitrogens and HN is assigned HGPAM1 (which is an atom type from the CGenFF forcefield) using ParamChem. CB and HB1 are assigned CG311 and HGA1 using ParamChem. The rest of the atoms share assignments with the MFTOID forcefield.

Atom	Modified AMBER	GAFF- ϕ	MFTOID	modified CGenFF
N	N	n	NH1	NTOID
H*	H	hn	HC	HGP1

Table C.9: Same as Table C.6 but for the C-terminal amino group. The atom types for the MFTOID forcefield are from the CHARMM22 forcefield. N is assigned NH1 to be consistent with other peptoid nitrogens and H* is assigned HC because that is the CHARMM22 type for hydrogens in the N-terminal group. For the CGenFF forcefield, N is assigned NTOID to be consistent with other peptoid nitrogens and H* is assigned HGP1 (which is an atom type from the CGenFF forcefield) using ParamChem.

Forcefield parameters

An overview of the forcefield parameterization are as follows:

MFTOID. We use MFTOID with the TIP3P water model modified for the CHARMM forcefield²⁴⁶ (mTIP3P). We assign backbone atom types from Ref. [206]. We assign sidechain atom types and partial charges based on parameters for amino acids^{246,370} or using ParamChem^{368,369} (sections C.1.1 and C.1.1). We take nitroxide spin label parameters and charging from Ref. [185], a parameterization compatible with CHARMM forcefields. Partial charges and atom types for disarcosine are from Mirijanian *et al.*²⁰⁶

GAFF- ϕ . We use the GAFF- ϕ forcefield with the SPC/E³⁷¹ water model. GAFF- ϕ was originally parameterized with implicit solvent, but implicit solvent simulations show larger deviation from the DEER results than explicit solvent simulations (Fig. C.12a). We perform GAFF atom type assignment using AmberTools16³⁶⁰ and GAFF version 1.8 and derive partial charges using the AM1-BCC model³⁷² for the hydrophilic monomers. For the spin-labels we run electrostatic potential (ESP) calculations in Gaussian 16³⁶⁴ for dimers of the NME sidechain and the spin-label monomer with the B3LYP functional at the 6-311++G(d,p) level of theory and then use restrained electrostatic potential (RESP) fitting³⁶¹ in AmberTools to localize the charge on each atom.

Modified AMBER99SB-ILDN. We use the modified AMBER99SB-ILDN forcefield with the SPC/E water model, as in the original study.²⁰⁹ We perform AMBER atom type assignment using AmberTools16 and derive partial charges for each hydrophilic monomer by computing the average partial charge on each atom of the monomer in homopolymers of length 8 after performing an ESP calculation in Gaussian 16 using Hartree-Fock level of theory with the 6-31G(d) basis set followed by RESP fitting. We take nitroxide spin label parameters from Ref. [366], a parameterization compatible with AMBER forcefields, omitting the lone-pair parameters, and assign the same partial charges as those used with the GAFF- ϕ forcefield.

Modified CGenFF. We use the modified CGenFF forcefield with the mTIP3P water model. We assign backbone atom types from Ref. [210]. For the hydrophilic monomers, we assign sidechain atom types using ParamChem^{368,369} and assign the same partial charges as those used with MFTOID. We take nitroxide spin label parameters and charging from Ref. [185] and we assign remaining atom types using ParamChem.

In all cases, any additional missing parameters listed in Tables C.10-C.18 are derived based on analogous parameters in the forcefield. These missing parameters were derived from analogous parameters in the CGenFF4.1,³⁶⁷ CHARMM22,²⁴⁶ MFTOID,²⁰⁶ Weiser *et al.*,²¹⁰ and Sezer *et al.*¹⁸⁵ forcefields.

Atoms	b_{ij}	k_{ij}	Source	Analogous atoms
CT2-OG301	0.1415	301248	CGenFF	CG321-OG301
CT3-OG301	0.1415	301248	CGenFF	CG331-OG301
CT2-CTL2	0.1538	186188	CHARMM22	CT2-CT1

Table C.10: Additional parameters for harmonic bond potential in the MFTOID-based forcefield. The potential in GROMACS has the form $V(r_{ij}) = \frac{1}{2}k_{ij}(r_{ij} - b_{ij})^2$. The units for the parameters are: b_{ij} [nm] and k_{ij} [kJ/mol/nm²].

Atoms	θ_{ijk}^0	k_{ijk}	r_{ik}^0	k_{ijk}^{UB}	Source	Analogous atoms
HC-NH1-TC	123	284.512	0	0	CGenFF	CG201-NG2S1-HGP1
CT2-CT2-OG301	111.5	376.56	0	0	CGenFF	CG321-CG321-OG301
CT2-OG301-CT3	109.7	794.96	0	0	CGenFF	CG321-OG301-CG331
OG301-CT2-HA	108.89	384.0912	0	0	CGenFF	OG301-CG321-HGA2
OG301-CT3-HA	108.89	384.0912	0	0	CGenFF	OG301-CG331-HGA3
HAL2-CTL2-CT2	110.1	221.752	0.2179	18853.104	CHARMM22	HA-CT2-CT2
CTL2-CT2-CTL2	113.5	488.2728	0.2561	9338.688	CHARMM22	CT1-CT2-CT1
CTL1-CTL2-CT2	113.5	488.2728	0.2561	9338.688	CHARMM22	CT1-CT2-CT1
HA-CT2-CTL2	110.1	288.696	0.2179	18853.104	CHARMM22	HA-CT1-CT2
NH1-CT2-CTL2	113.5	585.76	0	0	CHARMM22	NH1-CT1-CT2

Table C.11: Additional parameters for the Urey-Bradley angle potential in the MFTOID-based forcefield. The potential in GROMACS has the form $V(\theta_{ijk}) = \frac{1}{2}k_{ijk}(\theta_{ijk} - \theta_{ijk}^0)^2 + \frac{1}{2}k_{ijk}^{UB}(r_{ik} - r_{ik}^0)^2$. The units for the parameters are: θ_{ijk}^0 [deg], k_{ijk} [kJ/mol/rad²], r_{ik}^0 [nm], and k_{ijk}^{UB} [kJ/mol/nm²].

Atoms	ϕ_s	k	n	Source	Analogous atoms
TC-CT2-NH1-CT2	0	7.5312	1	Ref. [206]	TC-CT2-NH1-CT3
CT2-TC-NH1-CT2	180	7.5312	2	Ref. [206]	CT2-TC-NH1-CT3
OG301-CT2-CT2-HA	0	0.79496	3	CGenFF	OG301-CG321-CG321-HGA2
CT2-CT2-OG301-CT3	0	2.38488	1	CGenFF	CG321-CG321-OG301-CG331
CT2-CT2-OG301-CT3	0	1.21336	2	CGenFF	CG321-CG321-OG301-CG331
CT2-CT2-OG301-CT3	0	1.79912	3	CGenFF	CG321-CG321-OG301-CG331
HA-CT2-OG301-CT3	0	1.18826	3	CGenFF	HGA-CG321-OG31-CG331
HA-CT3-OG301-CT2	0	1.18826	3	CGenFF	HGA-CG321-OG31-CG321
NH1-CT2-CT2-OG301	180	0.66944	1	CGenFF	CG331-CG321-CG321-OG301
NH1-CT2-CT2-OG301	0	1.63176	2	CGenFF	CG331-CG321-CG321-OG301
TC-NH1-CT2-HA	0	0	3	Ref. [206]	HA-CT2-NH1-TC
CT2-NH1-CT2-HA	0	0	1	Ref. [206]	HB-CT2-NH1-CT3
CT2-NH1-CT2-HB	0	0	1	Ref. [206]	HB-CT2-NH1-CT3
CT2-NH1-CT2-CT2	0	0.4184	3	Ref. [206]	X-CT3-NH1-X
CT2-TC-NH1-HC	0	0	3	N/A	None
O-TC-NH1-HC	0	0	3	N/A	None
NN-CTL1-CTL2-CT2	0	2.481112	3	Ref. [185]	NN-CTL1-CTL2-CTL1
HA-CT2-CTL2-CTL1	0	0.8368	3	CHARMM22	X-CTL1-CTL2-X
CTL2-CT2-CTL2-CTL1	0	0.8368	3	CHARMM22	X-CTL1-CTL2-X
NH1-CT2-CTL2-CTL1	0	0.8368	3	CHARMM22	X-CTL1-CTL2-X
CTL2-CT2-CTL2-HAL2	0	0.8368	3	CHARMM22	X-CTL1-CTL2-X
HA-CT2-CTL2-HAL2	0	0.8368	3	CHARMM22	X-CTL1-CTL2-X
CT2-NH1-CT2-CTL2	0	0.4184	3	Ref. [206]	X-CT3-NH1-X
NH1-CT2-CTL2-HAL2	0	0	3	N/A	None

Table C.12: Additional parameters for the proper dihedral potential in the MFTOID-based forcefield. The potential in GROMACS has the form $V(\theta_{ijkl}) = k(1 + \cos(n\phi - \phi_s))$. The units for the parameters are: ϕ_s [deg] and k [kJ/mol]. For dihedrals with no obvious analogy in other forcefields, we set the potential to 0 because they involve hydrogen atoms.

Atoms	ξ_0	k	Source	Analogous atoms
NH1-X-X-CT2	0	167.36	Ref. [206]	NH1-X-X-CT3

Table C.13: Additional parameters for the improper dihedral potential in the MFTOID-based forcefield. The potential in GROMACS has the form $V(\xi_{ijkl}) = \frac{1}{2}k(\xi_{ijkl} - \xi_0)^2$. The units for the parameters are: ξ_0 [deg] and k [kJ/mol/rad²].

Atoms	θ_{ijk}^0	k_{ijk}	Analogous atoms
NT-CT-C	107	50	N-CT-C
O-C-CT	120.4	80	O-C-CT
O-C-N	122.9	80	O-C-N

Table C.14: Additional parameters for the harmonic angle potential in the modified AMBER forcefield. The ones listed here are those that were added to the input file (NT-CT-C) or those that differed between the input file sent by Arushi Prakash (corresponding to the modified AMBER99SB-ILDN forcefield in Ref. [209]) and the original MFTOID forcefield from Mirijanian *et al.* (O-C-CT, O-C-N). In these cases, we used the same parameters as the input file from Arushi Prakash. The potential in AMBER has the form $V(\theta_{ijk}) = k_{ijk}(\theta_{ijk} - \theta_{ijk}^0)^2$. The units for the parameters are: θ_{ijk}^0 [deg] and k_{ijk} [kcal/mol/rad²].

Atoms	ϕ_s	k	n	Analogous atoms
X-CT-N-X	0	0	0	X-CT-N-X

Table C.15: Additional parameters for the dihedral potential in the modified AMBER forcefield. The one listed here is the only one that differed between the input file sent by Arushi Prakash and the original MFTOID forcefield. For this work, we use the same parameters as the input file from Arushi Prakash. The potential in AMBER has the form $V(\phi_{ijkl}) = k(1 + \cos(n\phi - \phi_s))$. The units for the parameters are: ϕ_s [deg] and k [kcal/mol].

Atoms	b_{ij}	k_{ij}	Source	Analogous atoms
CG311-CTL2	0.1538	186188	CHARMM22	CT2-CT1
NTOID-HGPAM1	0.098	338904	CHARMM22	NH1-HC
NTOID-HGP1	0.098	338904	CHARMM22	NH1-HC
CG311-NTOID	0.143	267776	CGenFF	CG311-NG2S1

Table C.16: Same as Table C.10 but for the modified CGenFF forcefield.

Atoms	θ_{ijk}^0	k_{ijk}	r_{ik}^0	k_{ijk}^{UB}	Source	Analogous atoms
HAL2-CTL2-CG311	110.1	221.752	0.2179	18853.104	CHARMM22	HA-CT2-CT2
CTL2-CG311-CTL2	113.5	488.2728	0.2561	9338.688	CHARMM22	CT1-CT2-CT1
CTL1-CTL2-CG311	113.5	488.2728	0.2561	9338.688	CHARMM22	CT1-CT2-CT1
HGA1-CG311-CTL2	110.1	288.696	0.2179	18853.104	CHARMM22	HA-CT1-CT2
NTOID-CG311-CTL2	113.5	585.76	0	0	CHARMM22	NH1-CT1-CT2
HGPAM1-NTOID-CG321	109.5	251.04	0.2074	16736	CHARMM22	HC-NH3-CT2
HGPAM1-NTOID-CG311	109.5	251.04	0.2074	16736	CHARMM22	HC-NH3-CT1
HGP1-NTOID-HGP1	106.5	326.352	0	0	CHARMM22	HC-NH2-HC
HTP1-NTOID-CG2O1	123	284.512	0	0	CGenFF	HGP1-NG2S1-CG2O1

Table C.17: Same as Table C.11 but for the modified CGenFF forcefield.

Atoms	ϕ_s	k	n	Source	Analogous atoms
NN-CTL1-CTL2-CG311	0	2.481112	3	Ref. [185]	NN-CTL1-CTL2-CTL1
HGA1-CG311-CTL2-CTL1	180	0.8368	3	CHARMM22	X-CTL1-CTL2-X
CTL2-CG311-CTL2-CTL1	0	0.8368	3	CHARMM22	X-CTL1-CTL2-X
NTOID-CG311-CTL2-CTL1	0	0.8368	3	CHARMM22	X-CTL1-CTL2-X
CTL2-CG311-CTL2-HAL2	0	0.8368	3	CHARMM22	X-CTL1-CTL2-X
CG321-NTOID-CG311-CTL2	0	0.4184	3	Ref. [206]	X-CT3-NH1-X
HGA1-CG311-CTL2-HAL2	0	0.8368	3	CHARMM22	X-CTL1-CTL2-X
NTOID-CG311-CTL2-HAL2	0	0	3	N/A	None
CG321-CG2O1-NTOID-HGP1	180	0	3	N/A	None
OG2D1-CG2O1-NTOID-HGP1	0	0	3	N/A	None
HGPAM1-NTOID-CG321-CG321	0	0.4184	3	Ref. [206]	X-CT3-NH1-X
HGPAM1-NTOID-CG321-HGA2	0	0.4184	3	Ref. [206]	X-CT3-NH1-X
HGPAM1-NTOID-CG321-CG2O1	0	0.4184	3	Ref. [206]	X-CT3-NH1-X
HGPAM1-NTOID-CG311-HGA1	0	0.4184	3	Ref. [206]	X-CT3-NH1-X
HGPAM1-NTOID-CG311-CTL2	0	0.4184	3	Ref. [206]	X-CT3-NH1-X
CG2O1-NTOID-CG311-HGA1	0	0.4184	3	Ref. [206]	X-CT3-NH1-X
CG2O1-NTOID-CG311-CTL2	0	0.4184	3	Ref. [206]	X-CT3-NH1-X
CG311-NTOID-CG321-CG2O1	0	3.3472	2	Ref. [210]	CG331-NTOID-CG321-CG2O1
CG321-NTOID-CG321-CG2O1	0	3.3472	2	Ref. [210]	CG331-NTOID-CG321-CG2O1
CG311-NTOID-CG321-HGA2	0	0	3	Ref. [210]	HGA2-CG321-NTOID-CG331
CG321-NTOID-CG321-HGA2	0	0	3	Ref. [210]	HGA2-CG321-NTOID-CG331
CG2O1-NTOID-CG321-CG321	0	0.4184	3	Ref. [206]	X-CT3-NH1-X
CG321-NTOID-CG321-CG321	0	0.4184	3	Ref. [206]	X-CT3-NH1-X
CG321-NTOID-CG311-CG321	0	0.4184	3	Ref. [206]	X-CT3-NH1-X
NTOID-CG321-CG321-OG301	180	0.66944	1	CGenFF	CG331-CG321-CG321-OG301
NTOID-CG321-CG321-OG301	0	1.63176	2	CGenFF	CG331-CG321-CG321-OG301

Table C.18: Same as Table C.12 but for the modified CGenFF forcefield. For dihedrals with no obvious analogy in other forcefields, we set the potential to 0 because they involve hydrogen atoms.

C.1.2 Expanded-ensemble sampling of dihedral rotations

Expanded-ensemble workflow

For the equilibration period, we initialize all state weights to 0 and then use Self-Adjusted Mixture Sampling³⁷³ (SAMS) to sample across the λ_{ee} -states and a global weights update scheme³⁷³ to adjust state weights. When the histogram of states sampled is within 30% of flat, we estimate the free energy differences between each state and the unbiased state using the multistate Bennett acceptance ratio (MBAR)¹²⁵ and set the state weights to the free energy differences. After this equilibration period during which the weights are continuously updated, we fix the weights to those most recently computed with MBAR. We alternate between continuous weight updates with SAMS and simulating with fixed weights until the fixed weights are able to sample across all λ_{ee} -states. With the final fixed weights, we then run a production simulation with Metropolized independence sampling to sample states, saving configurations every 0.2 ns.

Validation

We validate the expanded ensemble approach by computing the free energy of rotation around the ω dihedral in disarcosine using umbrella sampling. We run forty simulations of disarcosine restraining the ω dihedral to some target value, evenly spaced between -180° and 180° . The parameters and partial charges are from Ref. [206] and we simulate with mTIP3P water using the number of water molecules reported in Table C.19. A harmonic potential, $U = k(\omega - \omega_0)^2$ with force constant $k = 100\text{kJ/mol}$ and ω in radians restrains the dihedral. Prior to the umbrella sampling run, we run an NVT equilibration for 500 ps, followed by an NPT equilibration for 500 ps, using a Monte Carlo (MC) barostat with 500 fs between MC moves. For the production runs, we use the same conditions (temperature, pressure, thermostat, barostat) as the expanded ensemble simulation. Each umbrella simulation is run for 20 ns, with configurations saved every 20 ps. The free energy around the ω dihedral is reconstructed using MBAR.

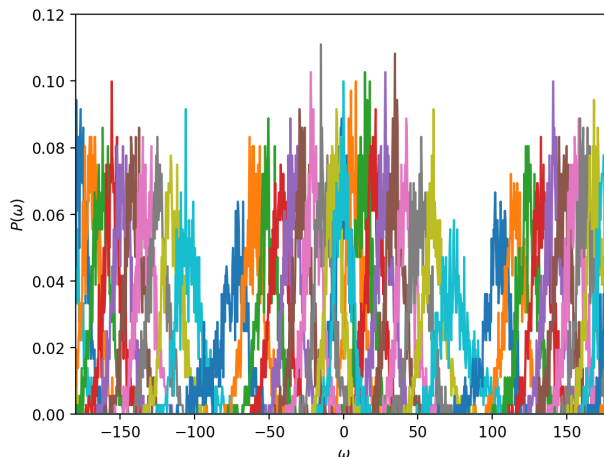


Figure C.4: The distribution of ω dihedrals accessed in each separate umbrella sampling simulation.

Convergence

The weights for each λ_{ee} -state converge to within $\approx 0.5k_B T$ of the free energy difference between the λ_{ee} -states (computed from the longer, production simulation) for $N = 11$ within 600 ns and for $N = 19$ within 800 ns (the trajectories of the weights for these two cases are shown as examples in Fig. C.5, but we would expect this to be the typical behavior across all simulations). We iteratively adjust the weights using the two-stage SAMS scheme proposed in Ref. [373] with $\beta = 0.6$ and the burn-in size, which determines the rate at which the weight updates decay, set to the number of trial state sampling steps for the histogram of states sampled to be within 30% of flat. We note that in the other equilibration simulations used to generate the weights used in this work, we typically set the burn-in size to infinity to allow for larger fluctuations in the weight updates to achieve final weights that enable a non-flat sampling of λ_{ee} -states that slightly oversamples the states closer to $\lambda_{ee} = 1$ in order to achieve better statistics. This could be more rigorously accomplished by adjusting the mixture weights, π , as defined in Ref. [373].

Dihedral correlation times

Fig. C.7a shows the isomerization rate of the ω dihedral as a function of λ_{ee} . The isomerization rate is computed by simulating disarcosine (after an NVT equilibration for 500 ps, followed by an NPT equilibration for 500 ps, using a Monte Carlo (MC) barostat with 500 fs between MC moves) at each λ_{ee} state for 2 ns, saving configurations every 0.2 ps, and counting the number of times the ω dihedral transitions between $-60 < \omega < 60$ and $\omega > 120$ or $\omega < -120$ and dividing by the total time of the production simulation. All settings (temperature, pressure, thermostat, barostat) are the same as for the expanded ensemble simulation. The parameters and partial charges are from Ref. [206] and we simulate with mTIP3P water using the number of water molecules reported in Table

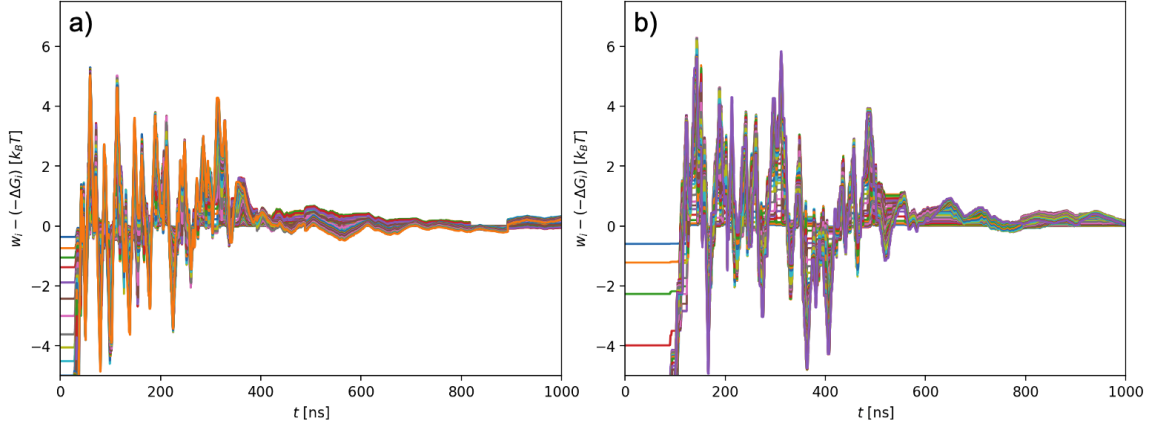


Figure C.5: The difference between the estimated weight for each state, w_i , and the ideal weights, $-\Delta G_i$ (where $\Delta G_i = G(\lambda_{ee} = \lambda_{ee,i}) - G(\lambda_{ee} = 1)$, the free energy difference between the unbiased state and state i , estimated using MBAR¹²⁵ from the production simulation) converges to within $0.5k_B T$ within 600 ns for (a) $N = 11$ and (b) $N = 19$ with the MFTOID forcefield. We do not expect significant differences between forcefields besides the modified CGenFF for which convergence may be slower because more dihedrals are softened, leading to greater free energy differences between the unsoftened and fully softened states. The different colors correspond to the different states.

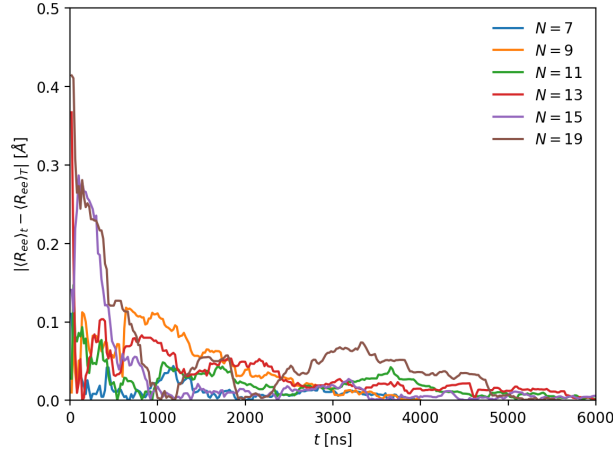


Figure C.6: The convergence of the mean end-to-end distance to within $0.1\text{\AA}$ of its final value takes on the order of μs . The plotted values are the trajectories of the absolute value differences between the running and final mean end-to-end distances for the different lengths with the MFTOID forcefield (we do not expect significant differences between forcefields besides the modified CGenFF for which convergence may be slower because more dihedrals are softened).

C.19. The measured isomerization rates, especially at small λ_{ee} , are underestimates, limited by the frequency of saving configurations, but the timescale associated with the maximum isomerization rate ($\approx 1/180 \text{ ns}^{-1} = 5.6 \text{ ps}$) is more than 20 times the save

frequency, so that the save frequency should not have a significant effect. The definition of an isomerization is also arbitrary and the measured isomerization rates change with the cutoffs used to define the cis and trans states which, at small λ_{ee} , are not the dominant ω states.

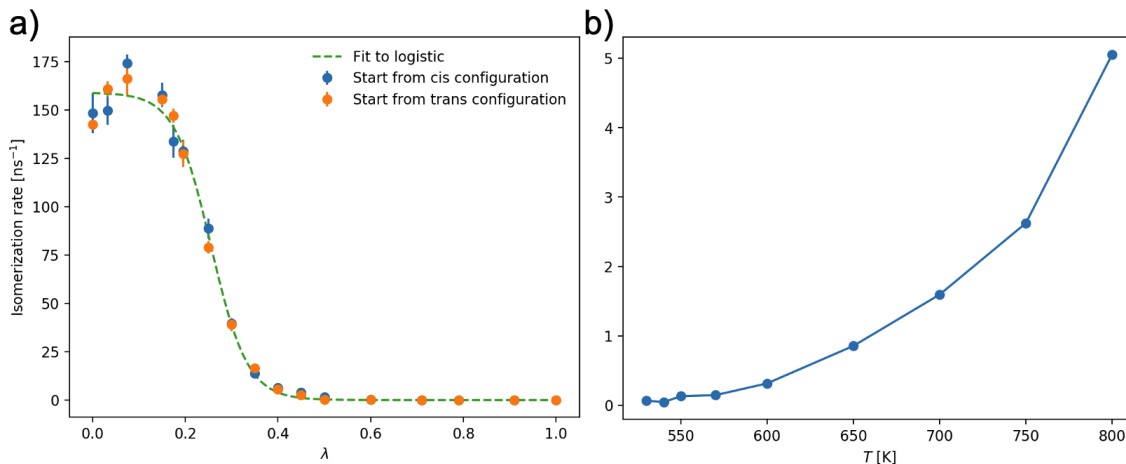


Figure C.7: The isomerization rate of the ω dihedral as a function of (a) expanded ensemble state, λ_{ee} and (b) temperature. (a) The error bars show the standard error computed from three independent runs from the same initial cis or trans configurations and the points give the average over those three runs. The green dashed line shows a fit to a logistic relationship: $r = a/(1 + \exp(-k(\lambda - \lambda_0)))$.

The isomerization rate increases by orders of magnitude as λ_{ee} decreases, regardless of whether the ω dihedral is initialized in the cis or trans configuration. When the ω dihedral potential is completely off ($\lambda_{ee} = 0$), the isomerization rate for the isomerizable ω dihedral in disarcosine is $\approx 150 \text{ ns}^{-1}$. While measuring the isomerization rate in the unsoftened ($\lambda_{ee} = 1$) state is not tractable with the accessible simulation timescales, the isomerization rate as a function of λ_{ee} follows approximately a logistic relationship (Fig. C.7a). Fitting the trend and extrapolating to $\lambda_{ee} = 1$ leads to an estimated isomerization rate of 890 s^{-1} , so that simulating in the fully softened ($\lambda_{ee} = 0$) state provides a speedup of 9 orders of magnitude. Experiments report even slower isomerization rates of 0.1 s^{-1} to 0.2 s^{-1} ,²⁰² which would indicate a speedup of 12 orders of magnitude.

In contrast, increasing the isomerization rate by increasing the temperature, as done in temperature replica exchange, requires very high temperatures. For a series of temperatures, we equilibrate disarcosine in implicit solvent 100 ps in the NVT ensemble and then run for 200 ns in the NVT ensemble, saving configurations every 0.2 ps. Implicit solvent simulations were run with the GBn2 solvation model³⁷⁴ without non-bonded interaction cutoffs using a Langevin thermostat with a friction coefficient of 1 ps^{-1} . All other settings are the same as those in the expanded ensemble. At 800 K, the highest temperature used in Ref. [207], the isomerization rate is 5 ns^{-1} (Fig. C.7b), still slower than the isomerization rate at the fully softened state here, demonstrating the utility of

a more targeted approach.

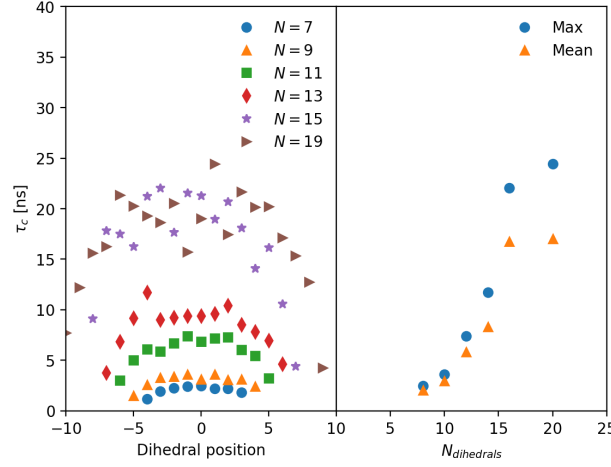


Figure C.8: (Left) The correlation times for each dihedral in the hydrophilic polypeptoids, computed from the expanded-ensemble simulations using the MFTOID forcefield as the time constant from an exponential fit of the angle autocorrelation function defined in Ref. [375], are on the order of nanoseconds. Generally, correlation times increase with the length of the polypeptoid and distance of the dihedral from the end of the polypeptoid. Correlation times also depend on the distribution of λ_{ee} -states (having more states closer to $\lambda_{ee} = 0$ reduces it, at the expense of simulating longer at more biased and thus less statistically significant states). (Right) The maximum and average ω correlation times for each polypeptoid length.

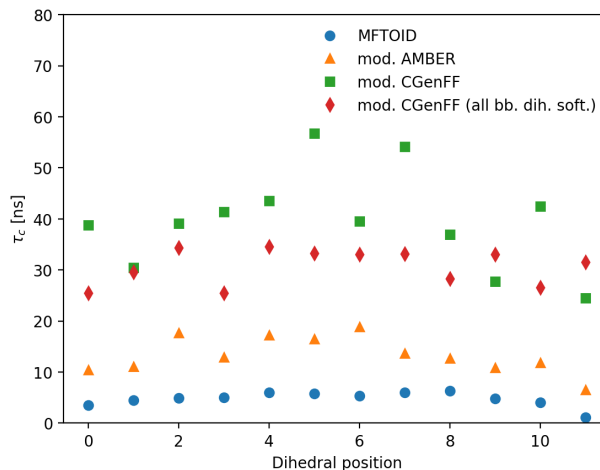


Figure C.9: The correlation times for each ϕ dihedral in the $N = 11$ polypeptoid, computed from the expanded-ensemble simulations as the time constant from an exponential fit of the angle autocorrelation function defined in Ref. [375]. For all points except the red diamonds, only the ω dihedral is softened. For the modified CGenFF forcefield, this results in poor sampling of isomerizations of the ϕ dihedral. Weiser and Santiso observed fewer than 10 transitions within 50 ns at 450 K for a trimer with methyl sidechains, and a decrease in frequency with sidechain size and polypeptoid length.²¹⁰ Softening all backbone dihedrals (the ω , ϕ , and ψ dihedrals shown in Fig. 4.1) reduces the correlation time for the modified CGenFF forcefield.

C.1.3 Simulation settings

System	Forcefield	$R_{ee,max}$	l_b	N_w	T_{equil}	T_{prod}	N_s
$N = 1$	MFTOID	9.2	36.3752	1,490	400	1,400	737
$N = 1$	Modified CGenFF	9.2	36.3752	1,490	200	2,000	990
$N = 2$	MFTOID	15.9	39.8024	2,091	400	2,000	417
$N = 2$	Modified CGenFF	15.9	39.8024	2,091	420	2,400	628
$N = 3$	MFTOID	16.9	44.3112	2,809	200	4,000	263
$N = 3$	Modified CGenFF	16.9	44.3112	2,809	420	3,600	600
$N = 5$	MFTOID	23.1	51.0089	4,284	200	2,800	101
$N = 5$	Modified CGenFF	23.1	51.0089	4,284	220	3,600	227
$N = 7$	MFTOID	30.3	58.3857	6,552	600	4,000	198
$N = 7$	Modified CGenFF	30.3	58.3857	6,552	220	4,800	369
$N = 9$	MFTOID	37.5	65.2181	8,908	600	4,000	253
$N = 9$	Modified CGenFF	37.5	65.2181	8,908	220	11,000	429
$N = 11$	MFTOID	45.4	72.6911	12,341	600	6,000	180
$N = 11$	MFTOID, implicit solvent	45.4	72.6911	N/A	600	4,000	666
$N = 11$	MFTOID, SPC/E	45.4	72.6911	12,341	1,400	8,800	142
$N = 11$	MFTOID, OPC4	45.4	72.6911	12,576	600	7,600	106
$N = 11$	Modified AMBER	45.4	72.6911	12,341	600	10,000	116
$N = 11$	Modified AMBER, implicit solvent	45.4	72.6911	N/A	600	6,000	750
$N = 11$	GAFF- ϕ	45.4	72.6911	12,341	1,400	8,000	95
$N = 11$	GAFF- ϕ , implicit solvent	45.4	72.6911	N/A	1,000	6,000	1,764
$N = 11$	Modified CGenFF	45.4	72.6911	12,341	300	5,200	158
$N = 11$	Modified CGenFF, implicit solvent	45.4	72.6911	N/A	600	4,000	625
$N = 13$	MFTOID	53.0	81.7029	17,754	600	6,200	134
$N = 13$	Modified CGenFF	53.0	81.7029	17,754	300	7,600	104
$N = 15$	MFTOID	62.5	89.4396	23,293	600	6,600	158
$N = 15$	Modified CGenFF	62.5	89.4396	23,293	820	6,400	46
$N = 19$	MFTOID	80.0	106.0111	38,639	500	5,600	114
$N = 19$	Modified CGenFF	80.0	106.0111	38,639	720	3,000	30
$N = 11$, unlabeled	MFTOID	N/A	64.2859	8,644	400	2,000	152
Disarcosine	MFTOID	N/A	40.0	2,156	200	5,000	5,000

Table C.19: Besides disarcosine, all systems correspond to varying lengths of the hydrophilic sequence shown in Fig. 4.1. l_b is the initial box length [$\text{\AA}$], N_w is the number of water molecules, T_{equil} and T_{prod} [ns] give the lengths of the equilibration and production simulations, N_s is the number of significant and independent samples (section C.1.4), and $R_{ee,max}$ [$\text{\AA}$] is the maximum end-to-end distance between spin label oxygen radicals, estimated from the initial, fully extended configuration. Simulations use the water models specified in section C.1.1 unless otherwise stated.

C.1.4 Uncertainty calculations

Free energy of rotation around disarcosine’s amide bond

For the expanded-ensemble calculation, the uncertainties shown in Fig. 4.3 are those returned from MBAR¹²⁵ when reweighting the potential of mean force (PMF) for rotation around the amide bond, multiplied by 1.96 to give a 95% confidence interval. Note that for the MBAR-computed uncertainties to be accurate, the energies input must be from uncorrelated configurations. The statistical inefficiency estimated from the potential energies using the pymbar timeseries module is 304 ps and configurations are saved every 200 ps, and so only every other configuration is used for the calculation of the PMF.

For the umbrella-sampling calculation, the uncertainties shown in Fig. 4.3 are also those returned from MBAR when reweighting the potential of mean force for rotation around the amide bond, multiplied by 1.96 to give a 95% confidence interval. Here, MBAR takes in all configurations, saved every 200 ps, so that the uncertainty is likely underestimated.

End-to-end distance distributions

First, the statistical inefficiency, g , is computed using the pymbar timeseries module^{125,218} using the maximum g computed for the trajectories of the potential energy, backbone end-to-end distance, and spin-label end-to-end distance. The number of independent samples, $N_{s,uncorr}$ is estimated to be the total number of configurations divided by the statistical inefficiency. MBAR is used to estimate the weights for each configuration. Then, 1000 distributions are generated by repeatedly uniformly sampling $N_{s,uncorr}$ configurations from the entire trajectory with replacement and computing the distance distribution with each configuration weighted according to the MBAR weights. These 1000 distributions are averaged to produce a mean distance distribution and the uncertainties plotted in Figs. 4.2, 4.4, and 4.6 are a 68% confidence interval. The uncertainty in the MBAR-derived weights is assumed to be small relative to the uncertainty due to limited sampling of the end-to-end distance.

Number of independent and statistically significant samples

We estimate the average length of simulation time between independent and statistically significant samples using two different methods. Results from both are shown in Fig. C.10 and are reasonably consistent. In the first method, we compute the effective sample number, N_{eff} , the number of samples that contribute to an average at the unbiased state, or the number of “significant” samples, for the entire dataset using MBAR. Then, N_{eff}/N is the fraction of samples that are significant. If we assume that each sample is equally likely to be significant, then the time in between independent and significant samples can be approximated as gN/N_{eff} , where g is the statistical inefficiency defined as above. For the second method, we repeatedly draw a random sample of size $N_{s,uncorr}$, defined as above. For each sample, which should now contain, on average, only independent

configurations, we compute $N_{\text{eff},\text{sample}}$. Then, on average, the simulation time between independent and significant samples is $T/N_{\text{eff},\text{sample}}$ where T is the total length of the simulation.

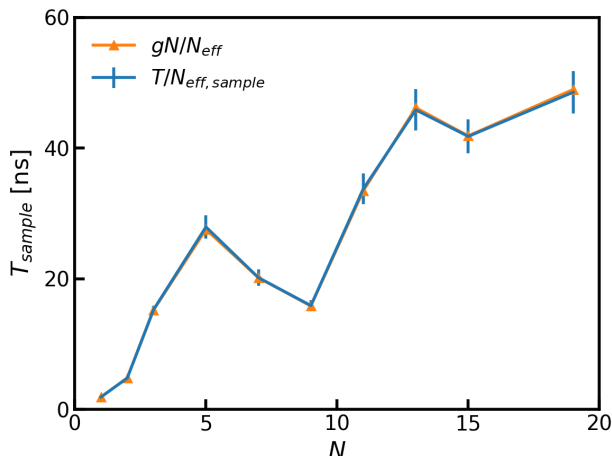


Figure C.10: Time in between independent and statistically significant samples. Estimates using two different methods described above for different polypeptoid lengths with the MFTOID forcefield (we do not expect significant differences between forcefields besides the modified CGenFF where more dihedrals are softened) are consistent and increase with the number of monomers N . However, the quantity also depends on the λ_{ee} states and state weights and thus will not necessarily be monotonic with N .

Distances between probability distributions

The same procedure as used in the computation of the uncertainty in the end-to-end distance distribution is used to generate 1000 distance distributions. The Shannon-Jensen distance and the Wasserstein distance are computed between each simulated distance distribution and the DEER distribution, yielding 1000 distances. The reported value is the mean of these values and the reported intervals give 95% confidence intervals. The contribution to the uncertainty due to uncertainty in the DEER distribution is assumed to be small and not computed.

Mean end-to-end distances

As in the computation of the uncertainty in the end-to-end distance distribution, the number of independent samples is computed from the statistical inefficiency. MBAR is used to estimate the weights for each configuration. Then, 1000 means are computed by repeatedly uniformly sampling $N_{\text{s,uncorr}}$ configurations from the entire trajectory with replacement and computing the mean distance with each configuration weighted according to the MBAR weights. The 1000 means are averaged and the error bars in Fig. 7.2 show a 95% confidence interval.

The same procedure is followed to compute the uncertainty in the mean within the cw-EPR optimal region for the $N = 3$ polypeptoid, whose distribution is shown in Fig. 4.6. The only change is that the weights for the configurations with end-to-end distances outside the optimal region are set to 0.

Scaling exponents

The reported uncertainties of the scaling exponent, prefactors, and shift factors are one standard deviation estimated from the `curve_fit` function in `scipy.optimize`. These uncertainties do not account for the uncertainty in the estimates of the individual values or the propagated uncertainties for values held constant during the fit (e.g. the prefactor in the scaling of the spin-label end-to-end distance or the shift factor in the scaling of the DEER distances); they reflect only the uncertainty of the fit.

C.1.5 Forcefield selection

Forcefield	JS SF-SVD	JS LD	Wasserstein SF-SVD	Wasserstein LD
MFTOID	0.16 (0.11, 0.22)	0.16 (0.11, 0.21)	1.5 (0.7, 2.3)	1.5 (0.7, 2.3)
GAFF- ϕ	0.19 (0.13, 0.26)	0.19 (0.13, 0.27)	1.1 (0.6, 2.1)	1.2 (0.6, 2.1)
Modified AMBER	0.30 (0.21, 0.39)	0.30 (0.21, 0.40)	3.0 (1.6, 4.3)	3.0 (1.6, 4.3)
Modified CGenFF	0.22 (0.15, 0.29)	0.22 (0.15, 0.29)	2.2 (1.2, 3.2)	2.3 (1.2, 3.2)

Table C.20: The first Wasserstein distance or the earth mover’s distance³⁷⁶ measures the distance between two probability distributions. The calculation gives similar trends compared to the Jensen-Shannon distance (the GAFF- ϕ and MFTOID forcefields are flipped, but this difference is well within the uncertainty). Computing both distances using the DEER distribution fit from LongDistances (LD) (as opposed to SF-SVD) gives very similar results, as the two DEER distributions are very similar above 20Å (Fig. C.11). These metrics are computed in the DEER-optimal ($> 20\text{\AA}$) region only. Prior to computation, in order to eliminate any unphysical artifacts, the DEER distribution is set to 0 above the maximum possible R_{ee} (45.4Å, see Table C.19) and negative values in the distributions are also set to 0. The DEER distribution is then renormalized. Parentheses give 95% confidence intervals, computed as described in section C.1.4.

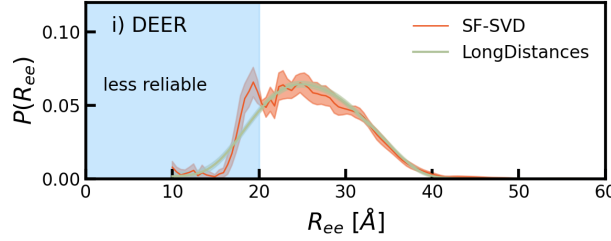


Figure C.11: Same as Fig. 4.2a, but with the DEER distribution fit using LongDistances (regularized inversion) as well.

C.1.6 Water model selection

Implicit solvent simulations (Fig. C.12a) generally give more collapsed conformations compared to the explicit solvent simulations (Fig. 4.2). Implicit solvent simulations were run with the GBn2 solvation model³⁷⁴ without non-bonded interaction cutoffs using a Langevin thermostat with a friction coefficient of 1 ps^{-1} . We also compare different explicit water models (Fig. C.12b) and find that there is little deviation, only a slight shift to lower spin-label end-to-end distances for the SPC/E model.

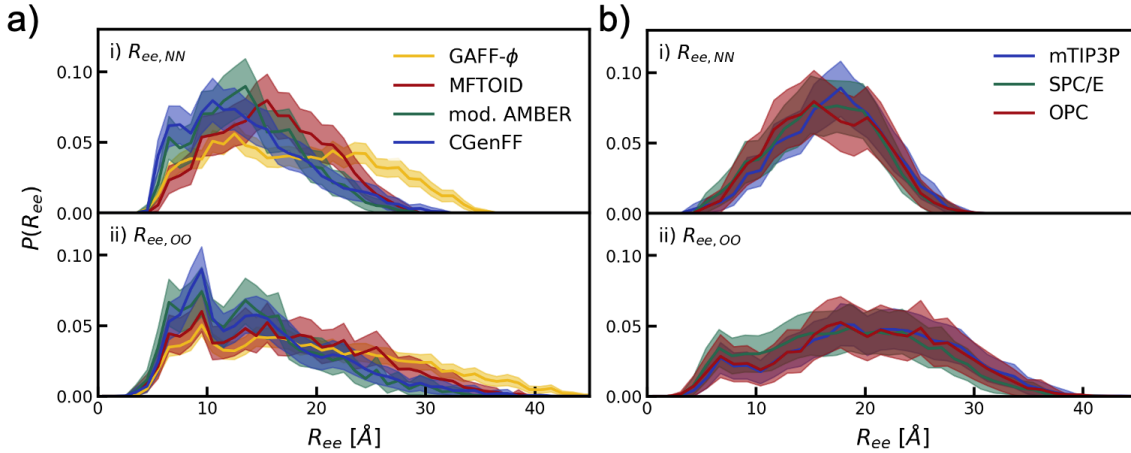


Figure C.12: Backbone (top) and spin-label (bottom) end-to-end distance distributions for $N = 11$ computed for (a) each polypeptoid forcefield in implicit solvent and (b) the MFTOID forcefield with the modified TIP3P, SPC/E, and OPC4^{247,248} water models.

C.2 Time-domain simulations

Simulated background-corrected DEER time-domain data, $F(t)$, can be computed,¹⁹⁴ given a distance distribution $P(r)$, through convolution with the kernel, $\kappa(r, t)$:

$$F(t) = 1 - \lambda \left(1 - \sum_i^{N_r} \kappa(r_i, t) P(r_i) \Delta r \right) \quad (\text{C.1})$$

$$\kappa(r, t) = \sqrt{\frac{\pi r^3}{6C_0 t}} \left[\cos\left(\frac{C_0 t}{r^3}\right) \mathcal{C}\left(\sqrt{\frac{6C_0 t}{\pi r^3}}\right) + \sin\left(\frac{C_0 t}{r^3}\right) \mathcal{S}\left(\sqrt{\frac{6C_0 t}{\pi r^3}}\right) \right] \quad (\text{C.2})$$

Here, λ is the inversion efficiency (given by the DEER experiment), N_r is the number of points in the discretized distance distribution, Δr is the interval between distances in the distribution, C_0 is a constant and is equal to $3.27 \times 10^5 \text{ \AA}^3/\mu\text{s}$, and $\mathcal{C}$ and $\mathcal{S}$ are the Fresnel cosine and sine integrals. To compute the time-domain data from the simulated distance distributions, the latter are first linearly interpolated to reduce Δr , since insufficient points in the distance domain can lead to artificial high-frequency oscillations in the time-domain (Fig. C.14). This interpolation should not have a significant impact on the computed time-domain signal (Fig. C.14). For all simulated signals, the inversion efficiency, λ , in equation C.1 is set to the experimentally derived value fitted in LongDistances (Table C.21). The fitted distance distributions from SF-SVD are first modified by setting all negative values to zero and all values above the maximum end-to-end distance (Table C.19) to zero, and renormalizing before computing the time-domain signal, in order to eliminate any obviously unphysical artifacts of the fit.

N	λ	Vertical shift in Fig. 4.5
9	0.072	-0.27
11	0.112968	-0.14
15	0.067	-0.10
19	0.082	0.0

Table C.21: The modulation depth, λ , in equation C.1 is fitted in LongDistances. Here, it does not show a clear trend with respect to polypeptoid length, number of hydrophilic monomers N , in contrast to Kaminker *et al.*¹⁸⁶ who showed that λ decreases with an increase in the population of short distances. Here, our use of different observer pulse lengths for each experiment (in order to optimize the signal-to-noise ratio) results in different effective excitation bandwidths, which complicate the dependence of modulation depth on dipolar coupling.²⁰¹

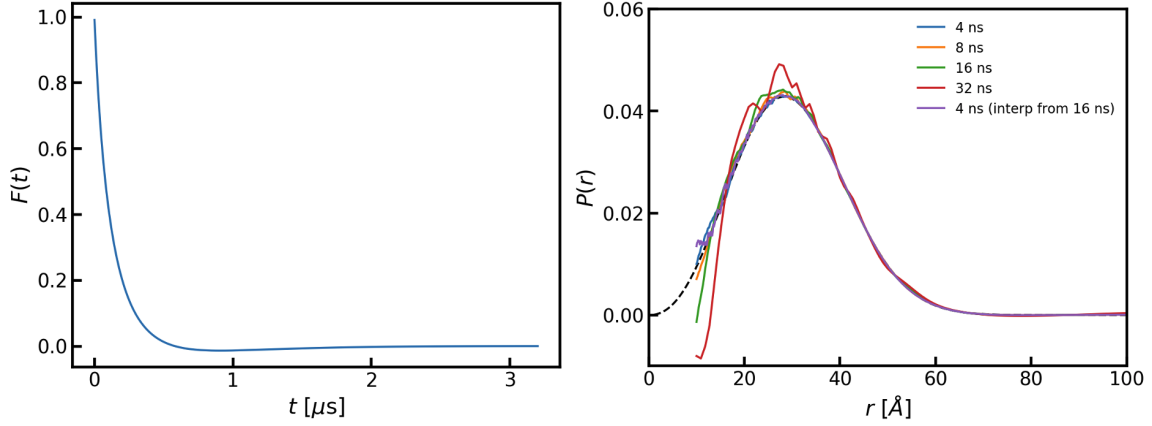


Figure C.13: A time-domain signal (left) was computed at 1 ns resolution using equation C.1 with $P(r) = a_1(r/\tilde{r})^{\epsilon_1} \exp(-a_2(r/\tilde{r})^{\epsilon_2}) 4\pi(r/\tilde{r})^2$ with $\epsilon_1 = 0.275$, $\epsilon_2 = 2.427$, $a_1 = 0.239$, $a_2 = 1.076$, $\tilde{r} = 30\text{\AA}$ for an excluded volume polymer distribution with mean end-to-end distance $30\text{\AA}$ (black dashed line on right). The time-domain signal is then subsampled at intervals at varying lengths to give signals at different resolution (4 ns, 8 ns, 16 ns, 32 ns). These signals are then fit using SF-SVD to produce the distance distributions at the right. As the time-resolution worsens, the deviation from the original distribution at shorter distances increases. Notably, fitting a signal with 4 ns resolution interpolated (linearly) from the 16 ns recovers the shorter distances, though we note that interpolation likely would not perform as well for a more realistic, noisy time-domain signal.

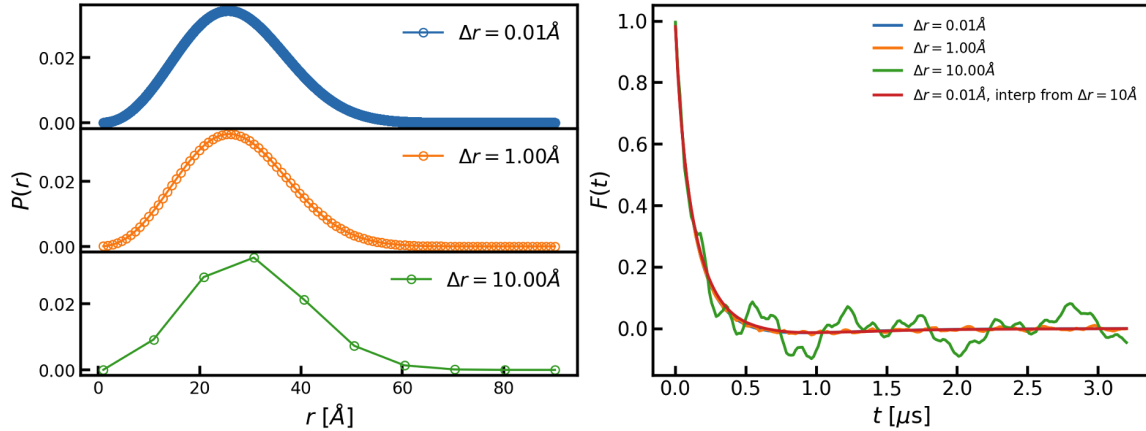


Figure C.14: The same excluded volume polymer distribution as in Fig. C.13 was used to generate distance distributions at varying distance resolution, Δr (left). The time-domain signals were computed using equation C.1. Large Δr values lead to high-resolution frequencies in the time-domain signals, as the $\Delta r = 10\text{\AA}$ case demonstrates dramatically, and as is still visible for $\Delta r = 1\text{\AA}$. Linear interpolation of the distance distribution to reduce Δr to $0.01\text{\AA}$ removes these oscillations (red), and gives a time-domain signal approximately equal to the signal computed from original distance distribution with resolution $\Delta r = 0.01\text{\AA}$.

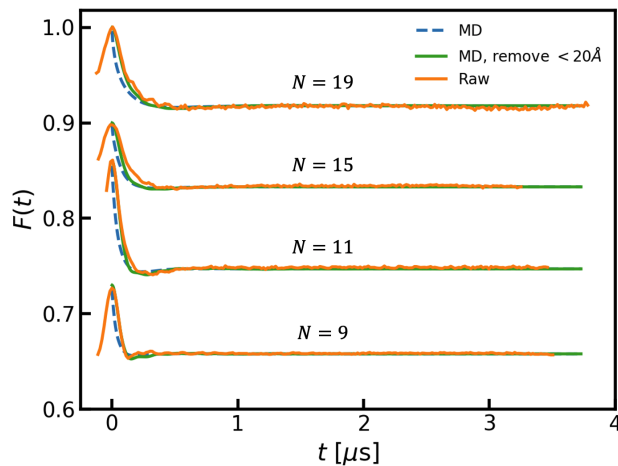


Figure C.15: Removing the population below $20\text{\AA}$ (green) shifts the low-time region of the back-calculated MD signal (derived from the $P(R_{ee})$ computed using the MFTOID forcefield) closer to the raw signal, suggesting that the differences between the simulated and raw signals at short times could be due to DEER’s lack of resolution in the $< 20\text{\AA}$ regime. However, even if the population below $20\text{\AA}$ is removed from the $N = 9$ distance distribution before computation of the time-domain signal, the simulated signal still shows a slightly broader minimum than the raw signal, which is reflected in the narrower SF-SVD distribution in Fig. 4.4.

C.3 Polypeptoid conformational behaviors

C.3.1 Visualization of the conformational distribution

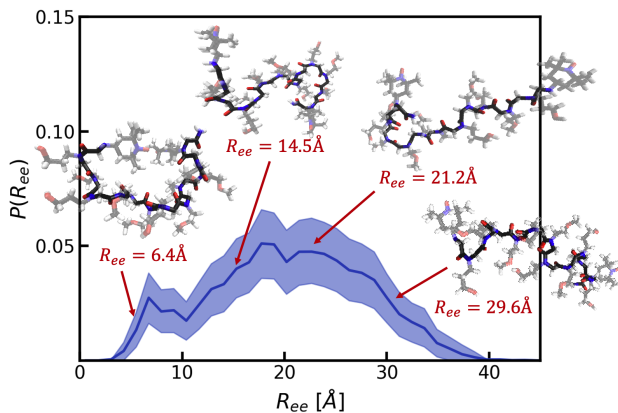


Figure C.16: Snapshots from various regions of the end-to-end distribution shows that higher end-to-end distances between the spin-labels correspond with more extended backbone conformations. Snapshots are visualized using VMD.¹⁵⁵

C.3.2 Dihedral preferences

The probability distribution for rotation around each ω dihedral in the backbone of the hydrophilic polypeptoids and also around the ω dihedral in disarcosine are reweighted from the simulations to the unbiased state using MBAR. The fraction of equilibrium conformations in the cis state is computed as the fraction of the distribution between $\omega > 90$ and $\omega < 90$. These calculations are from the expanded ensemble simulations using the MFTOID forcefield. For the $N = 11$ alternating hydrophilic polypeptoid, the cis preference for each dihedral with the nitrogen in a hydrophilic monomer is, on average, 0.44 (Fig. C.17). The cis preference for the dihedral with the nitrogen in the C-terminal spin-label monomer is higher, at 0.68, consistent with evidence that bulky groups directly attached to the backbone nitrogen lead to a stronger cis preference.⁷³ For the same reason, the cis preference for the isomerizable ω dihedral in disarcosine, whose sidechain is a methyl group, is much lower, at 0.12 (Fig. C.17). Since the most extended conformation places all dihedrals in the trans configuration, a stronger trans preference in disarcosine could bias the conformations towards longer end-to-end distances. As chain length increases, the trends in the cis preferences do not change significantly, with dihedrals in the middle of the polypeptoid remaining close to 0.5 and the last dihedral in the polypeptoid (the one with the nitrogen in the C-terminal spin-label monomer) showing stronger cis preference.

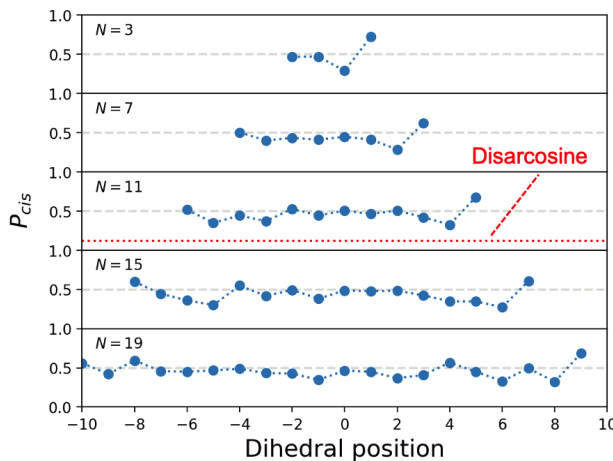


Figure C.17: The fraction of equilibrium conformations in the cis state, P_{cis} , for the hydrophilic polypeptoids of varying length (blue markers) and disarcosine (dotted red line). The dashed gray lines are at $P_{cis} = 0.5$.

C.3.3 Comparison of labeled and unlabeled molecules

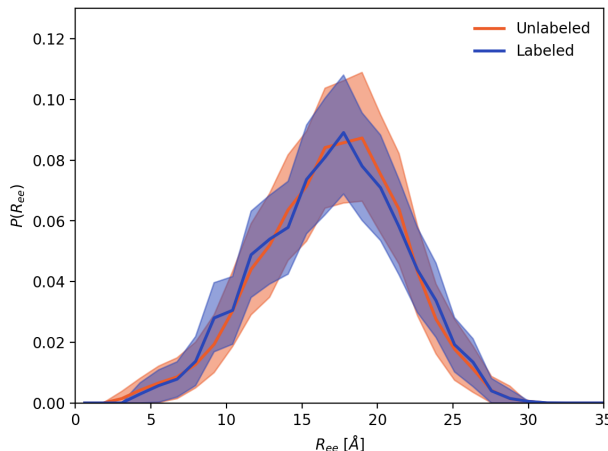


Figure C.18: The backbone end-to-end distance distributions, $R_{ee,NN}$, for the $N = 11$ polypeptoid terminated with spin labels (“Labeled”) and without spin labels (“Unlabeled”), both simulated using the MFTOID forcefield. Their agreement suggests that the spin labels do not significantly perturb the underlying backbone conformational distribution of the polypeptoids.

C.3.4 Comparison in DEER-accessible region

The simulated distributions reveal a peak at low end-to-end distance (approximately $8\text{\AA}$) that decreases with polypeptoid length and that is likely due to spin-label interaction. This short-distance feature is below the accessible region for DEER (within the shaded blue box in Fig. 4.4) and has been previously observed in simulations of polyethylene oxide labeled with the same spin-label, but using a different spin-label model,¹⁹² suggesting the aggregation occurs independently of the forcefield parameters. To eliminate the contribution from spin-label interaction that is not captured by DEER, we remove the conformations at short distances in which the spin labels are oriented towards each other. To separate out this contribution, we cluster the configurations to generate groups with the same spin-label conformations. This clustering is performed only on the spin-label heavy atom positions for conformations with a spin-label end-to-end distance below $20\text{\AA}$ and is based on the optimally aligned root-mean-squared distance⁵⁰ between conformations of only the spin-label heavy atoms, using an RMSD cutoff of $2.5\text{\AA}$. Fig. C.19 shows an example of this clustering for the $N = 11$ polypeptoid from the expanded ensemble simulation with the MFTOID forcefield.

We recompute the end-to-end distance distribution without the configurations in the clusters at the lowest end-to-end distance, effectively subtracting out the peak at small end-to-end distances ($\approx 8\text{\AA}$) but preserving the distributions at higher end-to-end distances.

Calculation of the mean end-to-end distance using these modified distributions gives the same scaling exponents for the backbone and spin-label mean end-to-end distances, compared to the means of the unmodified, simulated distributions. The fitted prefactor

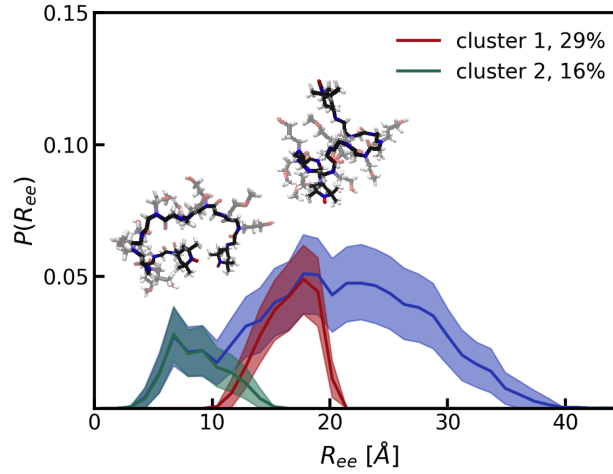


Figure C.19: Clustering the spin-label conformations allows identification of the conformations in which the spin labels have aggregated. Here, in the cluster centered at the lowest end-to-end distance (green), the spin labels are oriented towards each other.

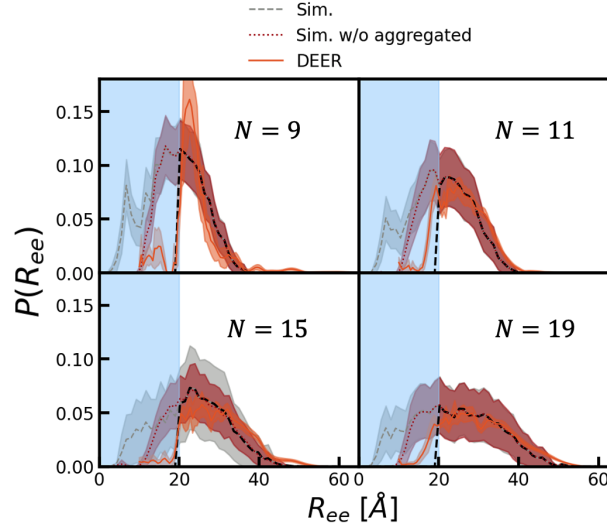


Figure C.20: The short-distance peak in the unmodified distribution (dashed lines, same as those in Fig. 4.4 in the main text, computed from expanded ensemble simulations with the MFTOID forcefield) has been subtracted out of the modified distribution (dotted lines) to approximately account for the inability of DEER to resolve shorter distances where dipolar coupling occurs. Both the unmodified and modified end-to-end distributions show a shift to longer distances and broadening with increasing number of monomers. All distributions are normalized so that their area above 20 Å is 1. Shaded regions show a 68% confidence interval, computed as described in section C.1.4.

for the simulated distances is also within uncertainty of the prefactor for the unmodified distributions. The largest difference is in the fitted shift factor, which is 1.8 “monomers”

larger for the modified distributions. In other words, removing the states in which the spin labels are oriented towards each other increases their contribution to the effective number of monomers.

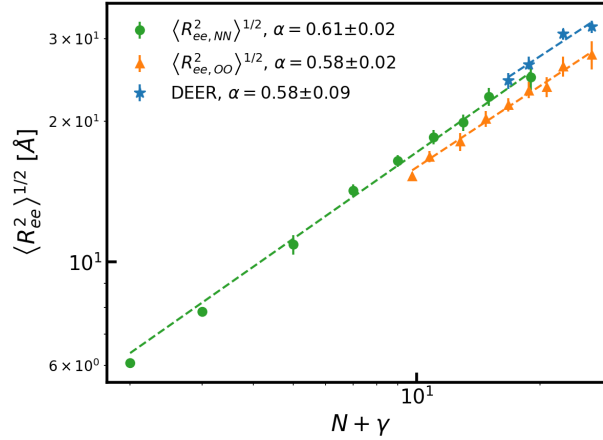


Figure C.21: The scaling exponents are fit the same way as reported in section 4.3. The simulated distances are the mean end-to-end distances for the distributions shown as dotted lines in Fig. C.20 along with the corresponding backbone distance distributions (not shown).

The backbone end-to-end distances, $\langle R_{ee,NN}^2 \rangle^{1/2}$ are fit to the power law bN^α , resulting in a scaling exponent of $\alpha = 0.61 \pm 0.02$ and a prefactor of $b = 4.2 \pm 0.2$. The spin-label end-to-end distances, $\langle R_{ee,OO}^2 \rangle^{1/2}$, are then fit to a shifted power law, $b(N + \gamma)^\alpha$ with b set to 4.2, yielding scaling exponent of $\alpha = 0.58 \pm 0.02$ and a shift factor of $\gamma = 7.8$. When the DEER data is fit to $b(N + \gamma)^\alpha$ with this shift factor, the resulting scaling exponent is $\alpha = 0.58 \pm 0.09$ and the prefactor is $b = 4.9 \pm 1.4$.

Another way to remove the contribution of configurations at shorter distances is to simply remove all distances below the DEER optimal region.

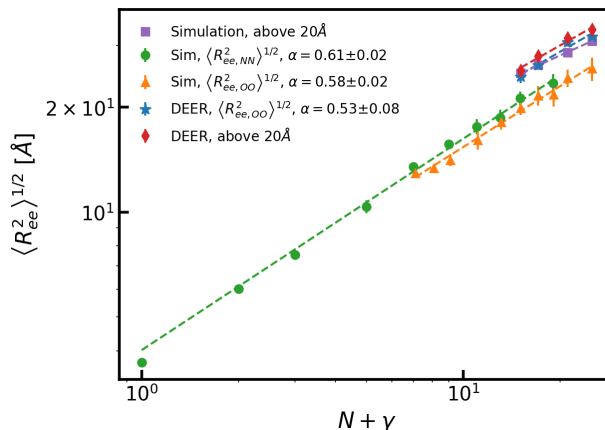


Figure C.22: Same as Fig. 7.2 but also showing the mean end-to-end distances for the DEER and simulation distributions after cutting off all distances below 20 Å. These are much closer together than the means computed from the full distributions, suggesting that the main source of the inconsistency between the mean distances from DEER and the simulations is the lack of resolution at low end-to-end distances.

C.4 Dependence of predicted distribution on cw-EPR optimal region bounds

The cw-EPR calculation approximates the distance distribution as a Gaussian and fits a mean and standard deviation, which should approximately correspond to the mean and standard deviation of the distribution within the optimal region for cw-EPR. The lower bound of the cw-EPR region corresponds to the distance at which exchange coupling becomes comparable to dipolar coupling, while the upper bound corresponds to the distance at which broadening becomes negligible.²⁰¹ As both these definitions rely on order-of-magnitude comparisons, there is some uncertainty in the exact bounds for each, which result in variations in the predicted mean and width over the optimal region from the simulated distribution.

Between 8 Å and 25 Å, the mean and standard deviation of the simulated distribution are 14.4 ± 0.5 Å (95% CI) and 4.0 Å, respectively. Both show some deviation from the corresponding cw-EPR measurements, which are 16.5 ± 0.1 Å and 2.2 ± 1.2 Å, respectively. This discrepancy could result from uncertainty in the estimate of the lower and upper bounds for cw-EPR; while 8 to 25 Å, as first proposed by Rabenstein and Shin,³⁷⁷ is the most generally accepted,^{377–379} other bounds have also been reported in the literature^{201,380–382} (Table C.22), and the predicted mean and standard deviation for this relatively narrow distribution are highly sensitive to this choice (Fig. C.23). Furthermore, this analysis assumes that all distances within the bounds contribute to the cw-EPR estimates according to their actual population, whereas in reality their contributions may decay close to the boundaries.²⁰¹

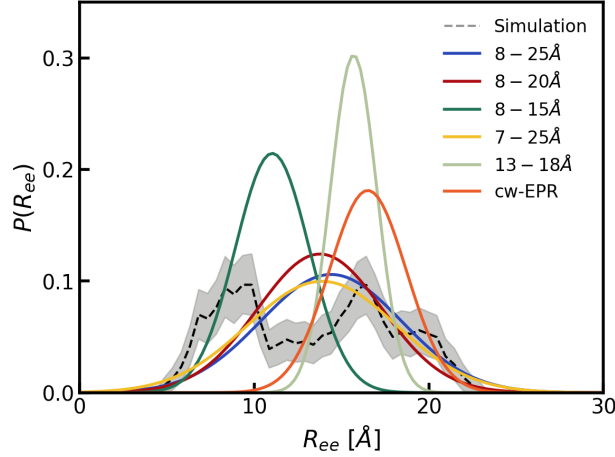


Figure C.23: The black, dashed line shows the simulated end-to-end distribution for the $N = 3$ polypeptoid (computed from the expanded ensemble simulation with the MFTOID forcefield), the orange line shows the Gaussian approximation of the end-to-end distance distribution from cw-EPR, and the other colored lines show Gaussian distributions with the same mean and standard deviation of the simulated distribution within various proposed cw-EPR bounds (Table C.22).

Lower bound [$\text{\AA}$]	Upper bound [$\text{\AA}$]	$\langle R_{ee} \rangle$ [$\text{\AA}$]	$\sigma_{R_{ee}}$ [$\text{\AA}$]	Source
8	25	14.4(13.9-14.9)	4.0	377-379
8	20	13.8(13.3-14.2)	3.6	201,380,381,383
8	15	11.0(10.7-11.4)	2.1	201
7	25	13.9(13.4-14.4)	4.3	382
13	18	15.7(15.4-15.9)	1.3	384

Table C.22: Different definitions of the lower and upper bounds yield different estimates of the mean, $\langle R_{ee} \rangle$, and standard deviation, $\sigma_{R_{ee}}$, of the simulated distribution (computed from the expanded ensemble simulation of the $N = 3$ polypeptoid using the MFTOID forcefield) within those bounds. Numbers inside the parentheses give a 95% confidence interval of the mean.

Appendix D

Supporting Information for Sequence Modulates Polypeptoid Hydration Water Structure and Dynamics

D.1 Simulation details

D.1.1 Expanded ensemble simulations

The simulations are run following the procedure outlined in Ref. [212] using λ_{ee} states of 1, 0.975, 0.95, 0.93, 0.91, 0.88, 0.85, 0.82, 0.79, 0.77, 0.75, 0.73, 0.71, 0.68, 0.65, 0.625, 0.6, 0.575, 0.55, 0.525, 0.5, 0.475, 0.45, 0.425, 0.4, 0.375, 0.35, 0.325, 0.3, 0.275, 0.25, 0.225, 0.195, 0.185, 0.175, 0.167, 0.15, 0.11, 0.075, 0.055, 0.033, 0.017, 0.0.

Sequence	l_b [$\text{\AA}$]	$R_{ee,max}$ [$\text{\AA}$]	N_w	T_{equil} [ns]	T_{prod} [ns]	N_s
EEEEEE-T	44.9476	22.536	2922	1000	7400	772
EEEEEN-T	44.7935	22.487	2890	1000	6000	618
EENNN-T	46.0174	22.905	3125	1000	9600	905
NNNEE-T	44.8044	22.592	2887	1000	8400	859
NENEN-T	44.3347	22.148	2805	1000	6000	768
NNNNN-T	44.3071	22.184	2801	1000	8000	576
EEEEEE	40.5223	18.509	2177	1000	9000	2142
NNNNN	41.0963	18.606	2250	1000	7000	1346

Table D.1: The initial box length, l_b , and number of waters, N_w , are set as described in section 5.2.1. The maximum end-to-end distance along the backbone, $R_{ee,max}$ is estimated from distance between the N-terminal residue’s nitrogen and the C-terminal capping group’s nitrogen in the initial, fully extended configuration, which is created in Avogadro 1.2.0 and then geometry minimized.²¹³ T_{equil} and T_{prod} give the lengths of the equilibration (varying weights) and production (fixed weights) simulations, respectively. To compute the number of independent and statistically significant conformations, N_s , we repeatedly draw a random sample of size $N_{s,uncorr}$. $N_{s,uncorr}$, the number of independent samples, is estimated to be the total number of configurations divided by the statistical inefficiency. The statistical inefficiency, g , is computed using the pymbar timeseries module using the maximum g computed for the trajectories of the potential energy and backbone end-to-end distance (taken between the N-terminal monomer’s nitrogen and the spin-label monomer’s nitrogen). For each sample of size $N_{s,uncorr}$, which should now contain, on average, only independent configurations, we compute the effective sample number, N_s , the number of samples that contribute to an average at the unbiased state, or the number of “significant” samples, for the entire dataset using MBAR.^{125,218}

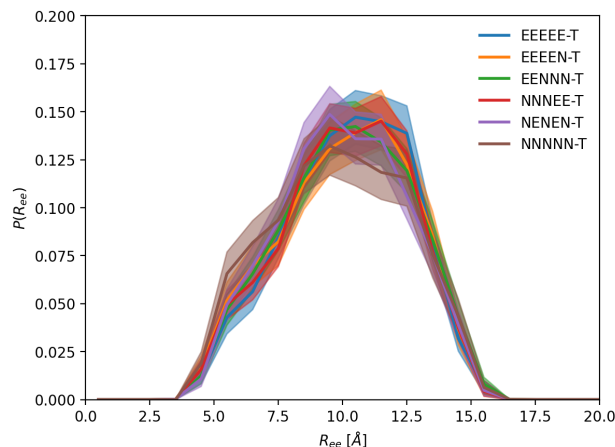


Figure D.1: End-to-end distance distributions for the peptoid sequences studied from the expanded ensemble simulations, reweighted to the unbiased ensemble as described in Ref. [212]. The end-to-end distance is taken between the N-terminal monomer’s nitrogen and the spin-label monomer’s nitrogen. The peptoids with more nonpolar monomers show slightly more compact conformations.

D.1.2 NVE simulations

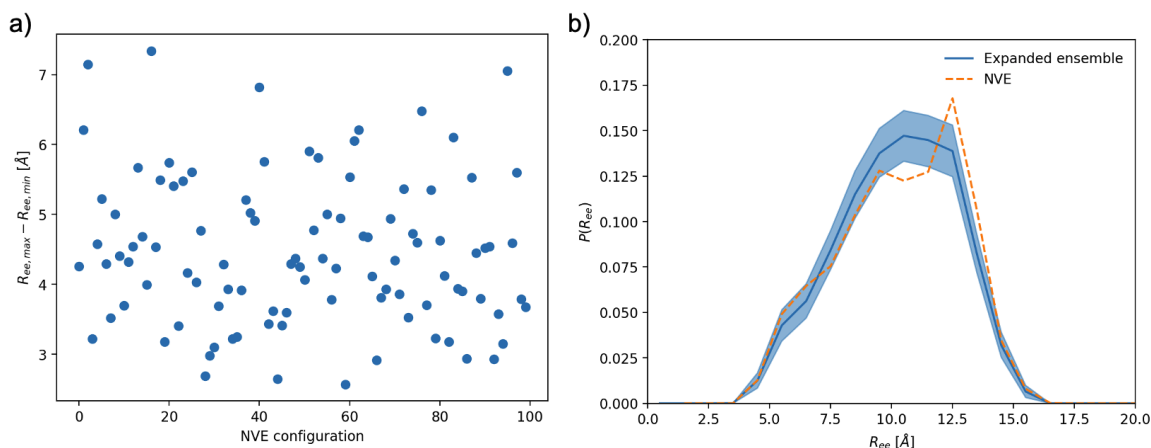


Figure D.2: (a) For each of the 100 NVE configurations of EEEEE-T, the range in end-to-end distances achieved is less than $7.5\text{\AA}$. In contrast, the typical range achieved in an expanded ensemble simulation is larger than $10\text{\AA}$. This suggests the peptoid conformations do not change significantly during the NVE simulations. (b) The end-to-end distance distributions from the 100 NVE simulations averaged together (dashed line) agrees with the reweighted distribution from the expanded ensemble simulation (solid line with shading) for EEEEE-T. This suggests the 100 NVE simulations reasonably reproduce the conformational distribution from the expanded ensemble.

D.1.3 Bulk water simulation

To compute dynamical properties of bulk water, we simulate 2854 water molecules. We use the OPC water model (the same model is used for the peptoid NVE simulations) with the same treatment of nonbonded interactions as described in section 5.2.1 in the main text. We first run an NVT equilibration for 100 ps in a box of length 4.42 nm using Langevin dynamics with a timestep of 2 fs, a temperature of 300 K, and a friction coefficient of 0.1 ps^{-1} . We then perform an NPT equilibration for 100 ps at 1 bar using Langevin dynamics with the same conditions and an MC barostat with 200 fs between MC moves, saving configurations every 2 ps. We obtain the average box length from the second half of the NPT simulation and run another NVT equilibration for 100 ps using Langevin dynamics with the same conditions. We then run an NVT simulation for 100 ns, saving configurations every 20 ps. After each 1 ns of the NVT simulation, we launch a simulation in the microcanonical ensemble, adjusting particle velocities so that the total energy matches the average total NVT energy (computed from the second half of the NVT simulation up to this point), for 250 ps using the Verlet integrator with a timestep of 2 fs and saving conformations every 50 fs. This gives us a total of 100 NVE simulations, starting from configurations separated by 1 ns in the NVT simulation. Uncertainties for bulk water quantities are computed *via* bootstrapping as described in section 5.2.2 of the main text.

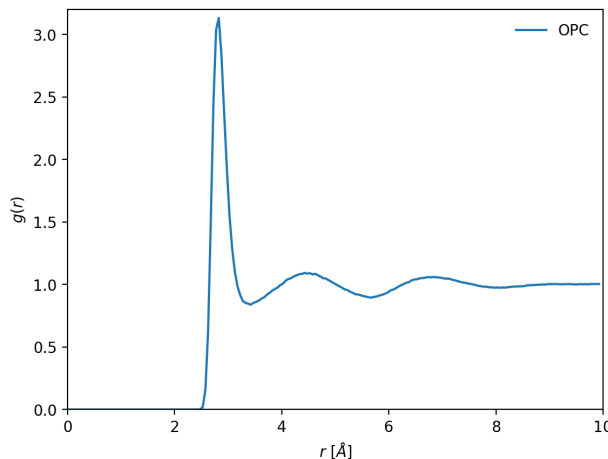


Figure D.3: The radial distribution function (RDF) of bulk OPC water has a first minimum around $3.4\text{\AA}$.

D.1.4 Silica-tethered peptoid systems

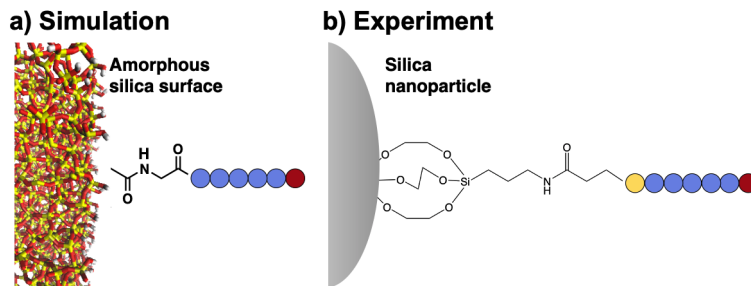


Figure D.4: (a) The simulated “silica-tethered” peptoid comprises an amorphous silica surface, as described in the main text, and a peptoid whose terminal carbon on the acetyl group is restrained to be near the silica surface. (b) The experimental system comprises a silica nanoparticle with a relatively longer linking group to the peptoid. The yellow circle represents a glycine residue.

D.1.5 Calculation of diffusivity near spin labels

Ref. [385] gives a partial parameterization of the 4-OH TEMPO molecule but omits parameters associated with the hydroxyl group. As such, we use parameters for the comparable atoms in the spin label in Ref. [212] (a full set of parameters are given in the SI of Ref. [212] and are partially based on the parameterization in Ref. [385]) combined with a standard CHARMM22²⁴⁶ hydroxyl group. We use partial charges for cyclohexanol from ParamChem^{368,369} as the partial charges for the hydroxyl group, set the partial charge of the adjacent carbon’s hydrogen to 0.09 and set the carbon’s charge to neutralize the group. Table D.2 gives the atom types and partial charges used.

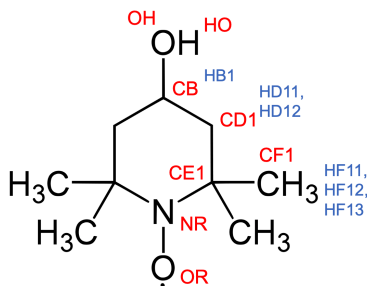


Figure D.5: Atom names for the 4-OH TEMPO molecule corresponding to entries in Table D.2.

The parameters associated with the OH1-CT2-CTL2 angle were set to the parameters for the OG311-CG311-CG321 in the CGenFF4.1 forcefield³⁶⁷ (downloaded 10/24/2019). The parameters associated with the CTL1-CTL2-CT2-OH1 and HAL2-CTL2-CT2-OH1 dihedrals were set to the parameters for the X-CTL1-CTL2-X wildcard in the CHARMM22 forcefield. The diffusivities reported in Fig. 5.9 in the main text are computed by extracting a diffusivity from the survival probability of water near the spin label’s oxygen radical with a cutoff of $7\text{\AA}$ and then correcting it for the motion of the spin-label (as

Atom	Type	Partial charge
OH	OH1	-0.648
HO	H	0.419
CB	CT2	0.139
HB1	HA	0.09
CD*	CTL2	-0.237
HD*	HAL2	0.09
CE*	CTL1	0.284
CF*	CTL3	-0.3305
HF*	HAL3	0.09
NR	NN	0.167
OR	ON	-0.379

Table D.2: Atom types and partial charges for the 4-OH TEMPO molecule. Atom names are shown in Fig. D.5. Asterisks signal that there are multiple atoms with the same assignment (e.g. CD1 and CD2).

discussed further in section D.2.2) by subtracting off the diffusivity of the spin label derived from the mean-squared displacement of the spin label. The uncertainty shown in Fig. 5.9 in the main text is computed by propagating the uncertainties reported for each of the quantities in the difference, and then dividing by the bulk water diffusivity, $0.243 \text{ \AA}^2/\text{ps}$.

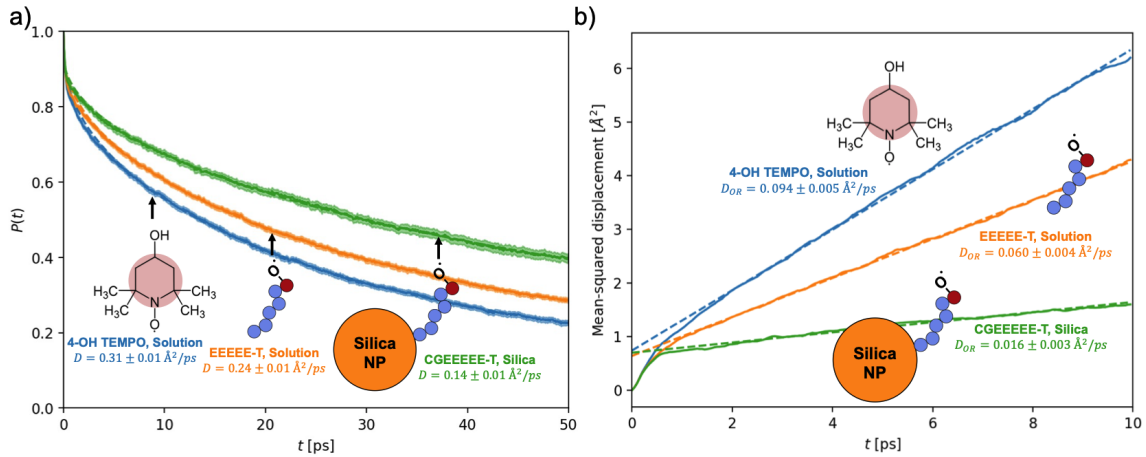


Figure D.6: (a) Survival probabilities for water within $7 \text{ \AA}$ of the spin label oxygen radical for the 4-OH TEMPO molecule, NME homopolymer, and silica-tethered NME homopolymer. The diffusivities are fit using equation 5.1. (b) Mean-squared displacements of the oxygen radicals for the 4-OH TEMPO molecule, NME homopolymer, and silica-tethered NME homopolymer. The diffusivities are the slope of a line fit to the region 1-10 ps (dashed lines show this fit), divided by 6. The uncertainties are computed *via* bootstrapping, as described in section 5.2.2.

D.2 Extracting diffusivity from survival probability

D.2.1 Derivation of survival probability model

We want to solve:

$$\frac{\partial C(r, t)}{\partial t} = D \nabla^2 C \quad (\text{D.1})$$

With boundary condition $C(r \rightarrow \infty, t) = 0$ and initial condition $C(r \leq R, 0) = C_0$, where $r = \sqrt{x^2 + y^2 + z^2}$, R is the radius of the cutoff region, D is the diffusivity, and C_0 is the initial concentration.

The solution for the case where the initial condition is a delta function is a Gaussian, so we can write the solution for this case as a superposition of Gaussians:

$$C(\mathbf{x}, 0) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} C(\boldsymbol{\xi}, 0) \delta(\mathbf{x} - \boldsymbol{\xi}) d\boldsymbol{\xi} \quad (\text{D.2})$$

$$C(\mathbf{x}, 0) = \frac{1}{(2\sqrt{\pi Dt})^3} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} C(\boldsymbol{\xi}, 0) \exp(-|\mathbf{x} - \boldsymbol{\xi}|^2/4Dt) d\boldsymbol{\xi} \quad (\text{D.3})$$

Substituting in the initial condition and converting to spherical coordinates:

$$C(r, \theta, \phi, t) = \frac{1}{(2\sqrt{\pi Dt})^3} \int_0^R \int_0^{2\pi} \int_0^\pi C_0 \exp(-|\mathbf{x} - \boldsymbol{\xi}|^2/4Dt) r_\xi^2 \sin \phi_\xi d\phi_\xi d\theta_\xi dr_\xi \quad (\text{D.4})$$

Assuming, without loss of generality, that $\mathbf{x}$ is on the z-axis, so that $x = 0$, $y = 0$, and $z = r$, then, letting $\boldsymbol{\xi} = (x_\xi, y_\xi, z_\xi)$:

$$|\mathbf{x} - \boldsymbol{\xi}|^2 = x_\xi^2 + y_\xi^2 + (r - z_\xi)^2 = r_\xi^2 \sin^2 \phi_\xi + (r - r_\xi \cos \phi_\xi)^2 \quad (\text{D.5})$$

Substituting this in:

$$C(r, t) = \frac{C_0}{(2\sqrt{\pi Dt})^3} \int_0^R \int_0^{2\pi} \int_0^\pi \exp(-(r_\xi^2 \sin^2 \phi_\xi + (r - r_\xi \cos \phi_\xi)^2)/4Dt) r_\xi^2 \sin \phi_\xi d\phi_\xi d\theta_\xi dr_\xi \quad (\text{D.6})$$

Integrating over θ_ξ :

$$C(r, t) = \frac{2\pi C_0}{(2\sqrt{\pi Dt})^3} \int_0^R \int_0^\pi \exp(-(r_\xi^2 \sin^2 \phi_\xi + (r - r_\xi \cos \phi_\xi)^2)/4Dt) r_\xi^2 \sin \phi_\xi d\phi_\xi dr_\xi \quad (\text{D.7})$$

Integrating over r_ξ and ϕ_ξ :

$$C(r, t) = \frac{2\pi C_0}{(2\sqrt{\pi Dt})^3} \left(\frac{4D^2 t^2}{r} \left(e^{-(r+R)^2/4Dt} - e^{-(r-R)^2/4Dt} \right) + 2D^{3/2} \sqrt{\pi} t^{3/2} \left(\operatorname{erf} \left(\frac{-r+R}{2\sqrt{Dt}} \right) + \operatorname{erf} \left(\frac{r+R}{2\sqrt{Dt}} \right) \right) \right) \quad (\text{D.8})$$

To get the average concentration in the spherical volume from $r = 0$ to $r = R$ at time t , we integrate:

$$\langle C(t) \rangle_R = \frac{1}{V} \int_0^R C(r, t) 4\pi r^2 dr = \frac{3}{R^3} \int_0^R C(r, t) r^2 dr \quad (\text{D.9})$$

Performing the integral:

$$\begin{aligned} \langle C(t) \rangle_R = \frac{3}{R^3} \frac{2\pi C_0}{(2\sqrt{\pi Dt})^3} & \left(\frac{4D^2 t^2}{3} e^{-R^2/Dt} (R^2(1 - 3e^{R^2/Dt}) + 2Dt(-1 + e^{R^2/Dt})) \right. \\ & \left. + \frac{4D^{3/2} \sqrt{\pi} R^3 t^{3/2}}{3} \operatorname{erf} \left(\frac{R}{\sqrt{Dt}} \right) \right) \quad (\text{D.10}) \end{aligned}$$

Simplifying:

$$\langle C(t) \rangle_R = \frac{C_0 \sqrt{Dt}}{\pi R} (e^{-R^2/Dt} - 3) + \frac{2C_0 (Dt)^{3/2}}{\pi R^3} (1 - e^{-R^2/Dt}) + C_0 \operatorname{erf} \left(\frac{R}{\sqrt{Dt}} \right) \quad (\text{D.11})$$

D.2.2 Survival probability around a moving center

We compute the survival probability as the fraction of waters originally within some cutoff radius of an atom at some time t_0 that are still within the cutoff radius after some time $t_0 + t$. The motion of the atom around which the cutoff radius is centered can thus affect the computed survival probability. This is shown prominently if we compare survival probabilities in bulk water around: (1) randomly chosen fixed points in the box, (2) randomly chosen water molecules, (3) water molecules that are fixed during the simulation (new simulations with single water molecules fixed are run in order to compute this last quantity). Fig. D.8 shows the survival probabilities using a radial cutoff of $5\text{\AA}$ (solid lines) along with the fits (dashed lines) to the model in equation 5.1. The fitted diffusivities are given in the legend. Measuring the survival probability around a water molecule gives an apparently higher diffusivity than measuring the survival probability around a fixed point. When the water molecule is fixed, then the fitted diffusivity is lower than the diffusivity around a fixed point.

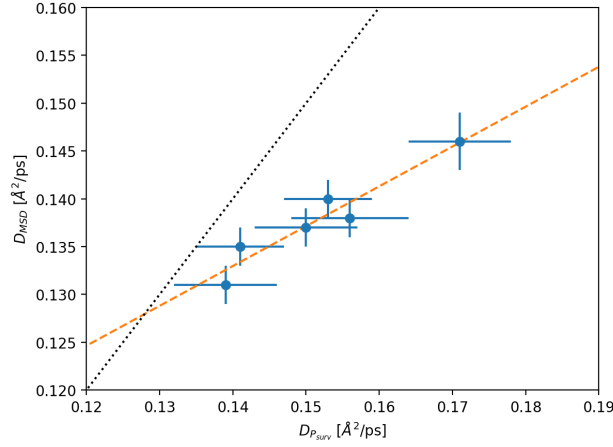


Figure D.7: The diffusivities from fitting the survival probability (x -axis) and the mean-squared displacement (y -axis) for the six sequences plotted in Fig. 5.4 are correlated with an R -squared value of 0.92.

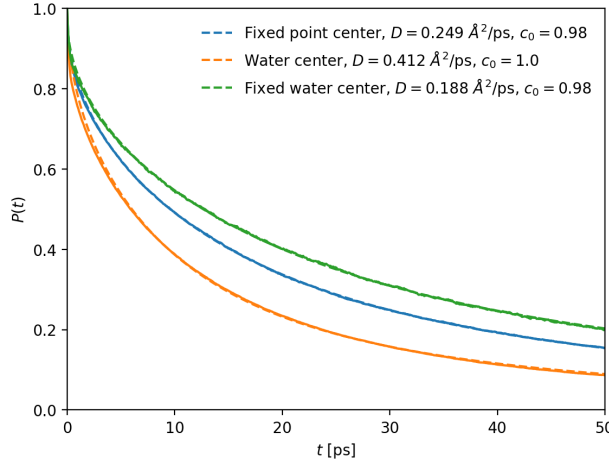


Figure D.8: The “fixed point center” trajectory is computed in a box of water by taking the survival probabilities for 20 spheres where their centers are chosen uniformly at random inside the simulation box. The “water center” trajectory is computed in a box of water by taking the survival probabilities around 100 randomly chosen water oxygens. The “fixed water center” trajectory is computed by choosing a water molecule to fix (all its atoms are prevented from moving) during the simulation, computing the survival probability around this fixed water oxygen, repeating this a total of 5 times, and averaging the survival probabilities.

To model this, we can begin to solve the diffusion equation in the same way we did in section D.2.1 except with a center that starts at the origin and diffuses with diffusion constant D' . We can write the survival probability $S(t)$ as the weighted average of concentrations, $S(t) = \int_0^\infty \tilde{c}(r') 4\pi r'^2 dr'$ where $P(r')$ is the probability that the center is at r' , which is the solution to the diffusion equation with initial condition a delta function

at the origin:

$$P(r') = \frac{1}{(2\sqrt{\pi D't})^3} \exp\left(-\frac{r'^2}{4D't}\right) \quad (\text{D.12})$$

And $\tilde{c}(r')$ is the concentration of water that was originally within the sphere of radius R around the origin that is within a sphere of radius R around the new center position, r' . If we assume that the movement of the center does not change the evolution of the population of waters originally in the sphere of radius R around the origin, then we already know what the concentration of these waters is as a function of distance r from the origin from the previous derivation for the fixed center case, equation D.8. Now, we need to average $C(r, t)$ over the sphere of radius R around r' to get $\tilde{c}(r')$. When $r' > R$:

$$\tilde{c}_1(r') = \frac{1}{\frac{4}{3}\pi R^3} \int_{r'-R}^{r'+R} C(r_0, t) A(r_0, R, r') dr_0 \quad (\text{D.13})$$

Where $A(r_0, R, r')$ is the volume of the slice of a sphere centered at r' with radius R that is r_0 away from the origin. When $r' < R$:

$$\tilde{c}_2(r') = \frac{1}{\frac{4}{3}\pi R^3} \left(\int_0^{R-r'} C(r_0, t) 4\pi r_0^2 dr_0 + \int_{R-r'}^{r'+R} C(r_0, t) A(r_0, R, r') dr_0 \right) \quad (\text{D.14})$$

So if we split $S(t)$ up into contributions from $r' < R$ and $r' > R$, we get:

$$S(t) = \int_0^R \tilde{c}_2(r') P(r') 4\pi r'^2 dr' + \int_R^\infty \tilde{c}_1(r') P(r') 4\pi r'^2 dr' \quad (\text{D.15})$$

Now, we just need an expression for $A(r_0, R, r')$:

$$A(r_0, R, r') = \int_0^{2\pi} d\theta \int_0^\alpha d\phi r_0^2 \sin \phi = -2\pi r_0^2 (\cos \alpha - 1) \quad (\text{D.16})$$

Where α is the angle in the triangle formed by side lengths r_0 , R , and r' between sides r_0 and r' . Using the Law of Cosines:

$$\cos \alpha = \frac{R^2 - r'^2 - r_0^2}{-2r'r_0} \quad (\text{D.17})$$

Then we can write:

$$A(r_0, R, r') = \pi r_0^2 \left(\frac{R^2 - r'^2 - r_0^2}{r'r_0} + 2 \right) \quad (\text{D.18})$$

The derived model requires numerically integrating to get the survival probability, $S(t)$. If we use this model to generate survival probabilities with varying D and D' values, and then fit the resulting survival probability to the fixed-center model (equation

5.1) to extract an “apparent” diffusivity, we see that the fitted diffusivities are equal to $D + D'$.

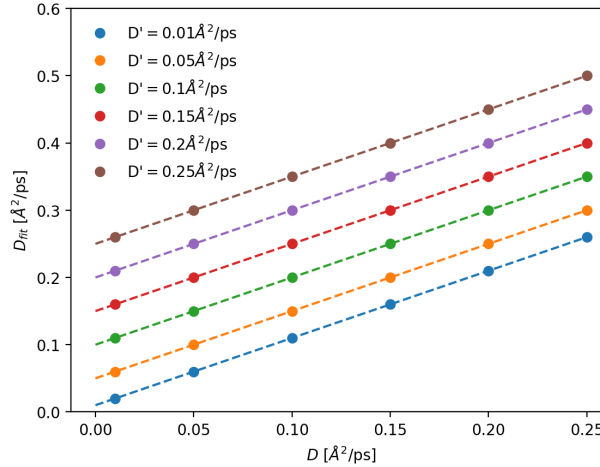


Figure D.9: For each point, $S(t)$ is computed according to the model above given the values of D and D' . Then, it is fit to the fixed-center model (equation 5.1) to yield D_{fit} . The dashed lines are $D + D'$. They intersect with the points, indicating that the apparent diffusivity extracted with the fixed-center model is the sum of the diffusivity of the moving center and the diffusivity of the surrounding waters.

This suggests that we can “correct” the diffusivities extracted from the survival probabilities around the peptoid atoms by subtracting out the diffusivity of the peptoid atoms. We compute this latter diffusivity by computing the mean-squared displacement for peptoid alpha carbons over 50 ps and fitting a straight line to the region past 10 ps. Then, we subtract these diffusivities from the survival-probability-fitted diffusivities. The correlation of these “corrected” diffusivities with those fitted from the MSDs of the peptoid’s hydration water is about the same (R -squared of 0.93 instead of 0.92 as in Fig. D.7), but now the diffusivities from the survival probability are lower than those from the hydration water MSD. This could be because the hydration water MSD calculation is not perfectly local and captures the dynamics of waters that move outside of the original cutoff radius in the 10 ps over which the MSD is measured.

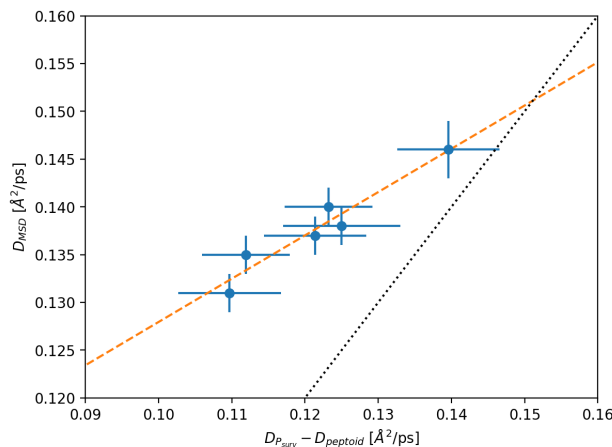


Figure D.10: The diffusivities from fitting the survival probability and then corrected this calculation by subtracting off the diffusivity of the peptoid (x -axis) and the mean-squared displacement (y -axis) for the six sequences plotted in Fig. 5.4 are correlated with an R -squared value of 0.93.

D.3 Conformational clustering

D.3.1 Clustering conformations of NME and alkyl homopolymers

To identify conformations of both the NME homopolymer (EEEEEE-T) and alkyl homopolymer (NNNNN-T) peptoids that are similar, we cluster configurations from the expanded ensemble simulations of the two peptoids based on the optimally aligned³⁸⁶ root-mean-squared distance (RMSD) of the positions of the backbone and side-chain heavy atoms (not including the spin-label heavy atoms), using an RMSD cutoff of $2.5\text{\AA}$. The conformations are weighted by the MBAR-determined weights. When a distance is computed between EEEEE-T and NNNNN-T peptoid configurations, we take the distances between carbons on the sidechain with the same positions, as well as between the oxygen in the methoxyethyl sidechain and the second-to-terminal carbon on the n-butyl sidechain. Table D.3 and Fig. D.11 show the results of this clustering for both sequences. For each set of clustered conformations, we choose a random sample of 50 conformations (approximately the number of statistically independent conformations in each cluster; see tables D.1 and D.3) and run NVE simulations starting from each set of 50 to determine cluster-specific quantities, plotted in Fig. 5.6.

D.3.2 Separating out effect of conformation for all sequences

Conformations from the expanded ensemble simulations for all sequences studied were clustered following a similar procedure as described in section D.3.1 for clustering only the homopolymers again using an RMSD cutoff of $2.5\text{\AA}$. The conformations are weighted by

Cluster	$f_{\text{EEEE-T}}$	$f_{\text{NNNN-T}}$	$R_{\text{EEEE-T}} [\text{\AA}]$	$R_{\text{NNNN-T}} [\text{\AA}]$
1	12.4	10.4	12.4	12.8
2	10.5	11.2	10.0	9.9
3	11.2	10.0	12.6	12.9
4	10.7	10.1	10.6	10.2
5	8.5	12.1	7.3	6.8
6	10.4	10.0	10.6	10.4
7	9.3	10.1	10.2	10.1
8	9.3	8.9	7.4	7.3

Table D.3: Information about each cluster when the fully hydrophilic (EEEE-T) and fully hydrophobic (NNNN-T) peptoid conformations from the expanded ensemble simulations are clustered together. The “end-to-end distance” for the EEEEE-T and NNNNN-T conformations in each cluster, $R_{\text{EEEE-T}}$ and $R_{\text{NNNN-T}}$, respectively, is taken between the N-terminal monomer’s nitrogen and the spin-label monomer’s nitrogen. The weighted percentages of conformations of EEEEE-T and NNNNN-T in each cluster, $f_{\text{EEEE-T}}$ and $f_{\text{NNNN-T}}$, respectively, are computed using MBAR-computed weights from the expanded ensemble. While the weighted fraction of conformations for each cluster is around 10%, the weights for more extended clusters (higher mean end-to-end distance) are slightly higher for EEEEE-T than NNNNN-T, while the weights for the more collapsed clusters (lower mean end-to-end distance) are slightly higher for NNNNN-T. This is consistent with the slight overall shift towards more collapsed conformations going from EEEEE-T to NNNNN-T. The bolded rows are the clusters chosen for analysis of conformation-independent sequence effects (Fig. 5.6 of the main text).

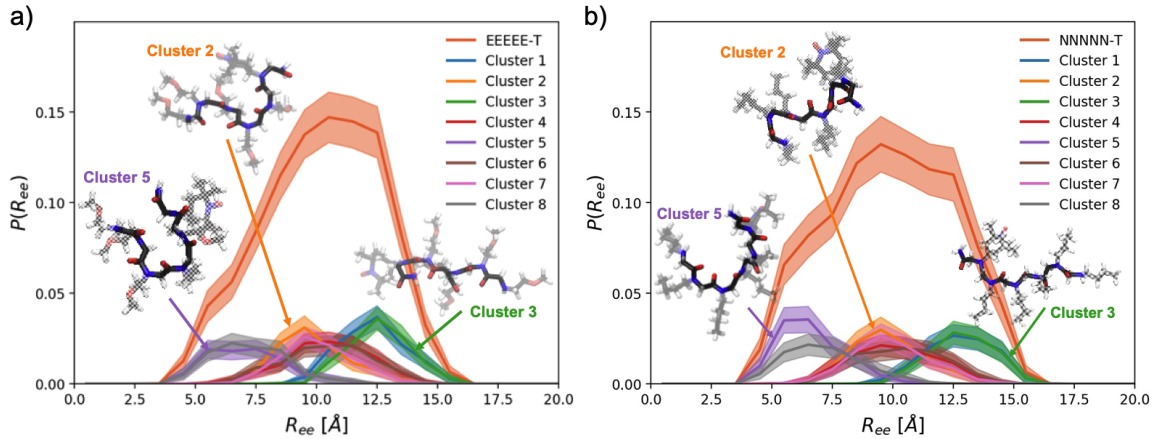


Figure D.11: End-to-end distance distributions (where the end-to-end distance is defined as for Table D.3) for the clusters of (a) EEEEE-T and (b) NNNNN-T conformations, along with the full distribution for each peptoid sequence, when EEEEE-T and NNNNN-T conformations are clustered together. Representative snapshots from each cluster are visualized using Visual Molecular Dynamics¹⁵⁵; the sidechain atoms and backbone hydrogens are transparent.

the MBAR-determined weights and distances are computed between position-equivalent atoms on the NME and alkyl sidechains. For each set of clustered conformations, we choose a random sample of 50 conformations and run NVE simulations starting from each set of 50.

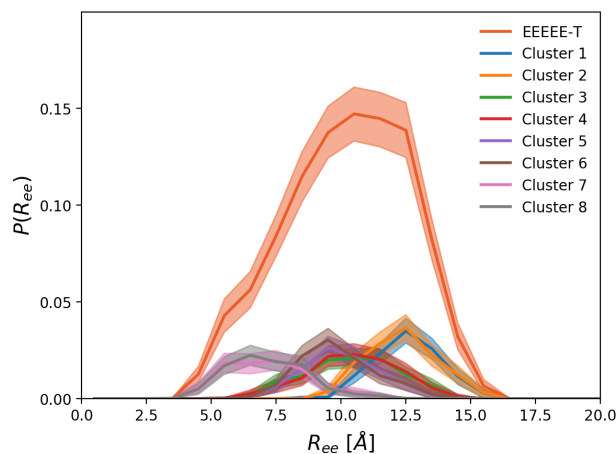


Figure D.12: End-to-end distance distributions (where the end-to-end distance is defined as for Table D.3) for the NME homopolymer and the clusters of the NME homopolymer conformations when conformations for all sequences are clustered together.

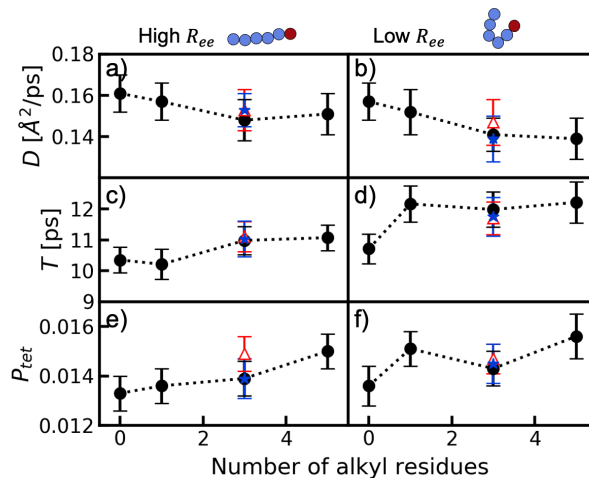


Figure D.13: (a,b) Diffusivity, (c,d) dipole correlation time, and (e,f) tetrahedrality for the group of (a,c,e) most extended conformations (cluster 1 in Fig. D.12) and (b,d,f) most compact conformations (cluster 7 in Fig. D.12). Consistent with the conclusion drawn from Fig. 5.6, the qualitative trend of water’s structural and dynamical behaviors with the number of nonpolar residues remains even when controlling for the effect of conformation. For both the most extended and most compact conformations, increasing the number of nonpolar residues leads to reduced dynamics (lower translational diffusivity and larger rotational correlation time) and enhanced tetrahedrality. Additional simulations are needed to resolve whether the effect of changing the position of the nonpolar residues, independent of their number (compare black circles, blue stars, and red triangles at number of alkyl residues = 3) is the same when controlling for the effect of conformation as the current results are within uncertainty.

D.4 Additional hydration water measurements

D.4.1 Structural distributions and dynamical trajectories

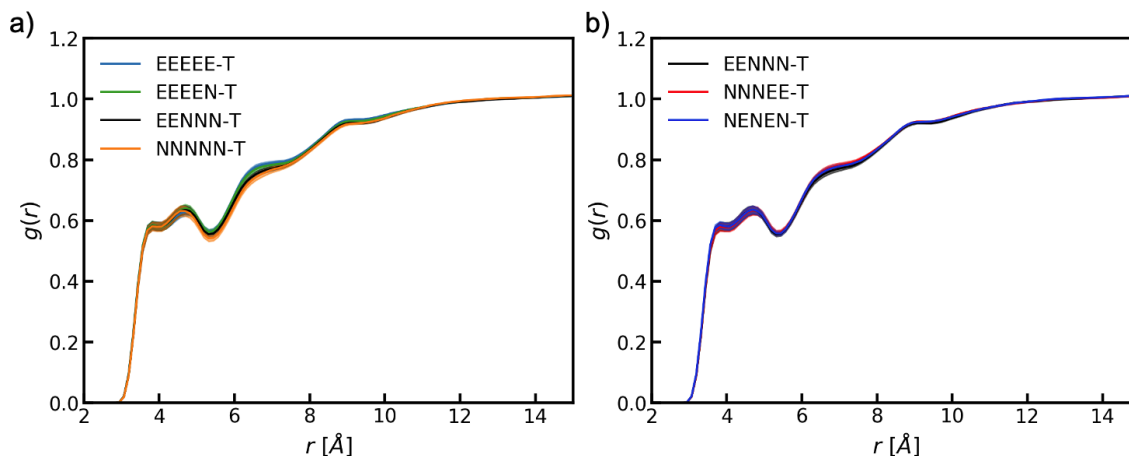


Figure D.14: Radial distribution function (RDF) for the set of peptoids varying the (a) number of alkyl residues at the C-terminus and (b) placement of three nonpolar residues in the sequence. With increasing number of nonpolar residues, the RDF decreases, especially around 6-7 Å, indicating that water is less dense.

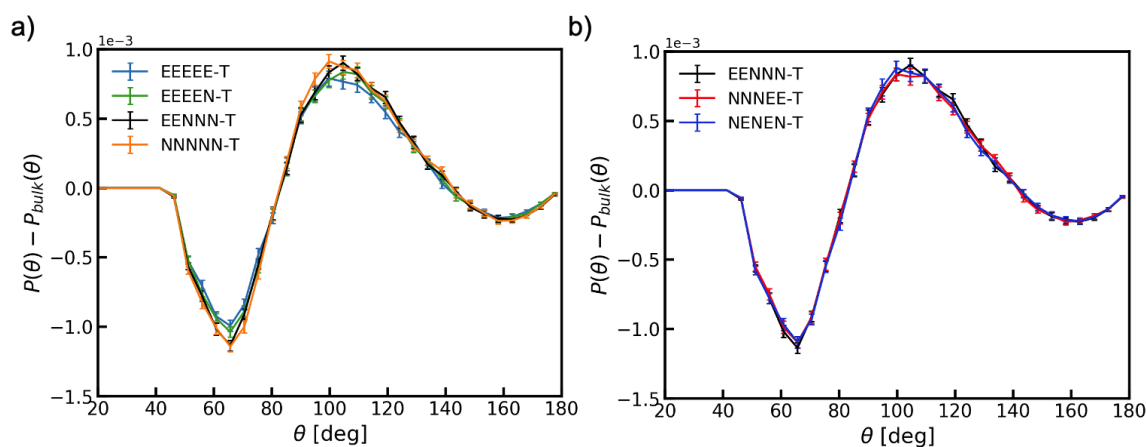


Figure D.15: Deviation from bulk three-body angle distributions for the set of peptoids varying the (a) number of alkyl residues at the C-terminus and (b) placement of three nonpolar residues in the sequence. With increasing number of nonpolar residues, there is a further diminishment of the icosahedral peak and an enhancement of the tetrahedral peak.

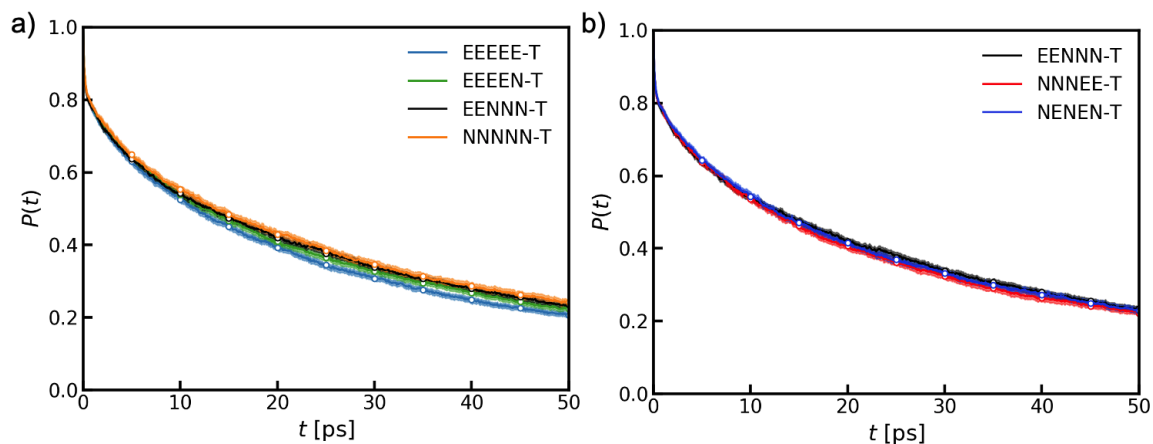


Figure D.16: Survival probability for the set of peptoids varying the (a) number of alkyl residues at the C-terminus and (b) placement of three nonpolar residues in the sequence. With increasing number of nonpolar residues, the survival probability increases (water is slower).

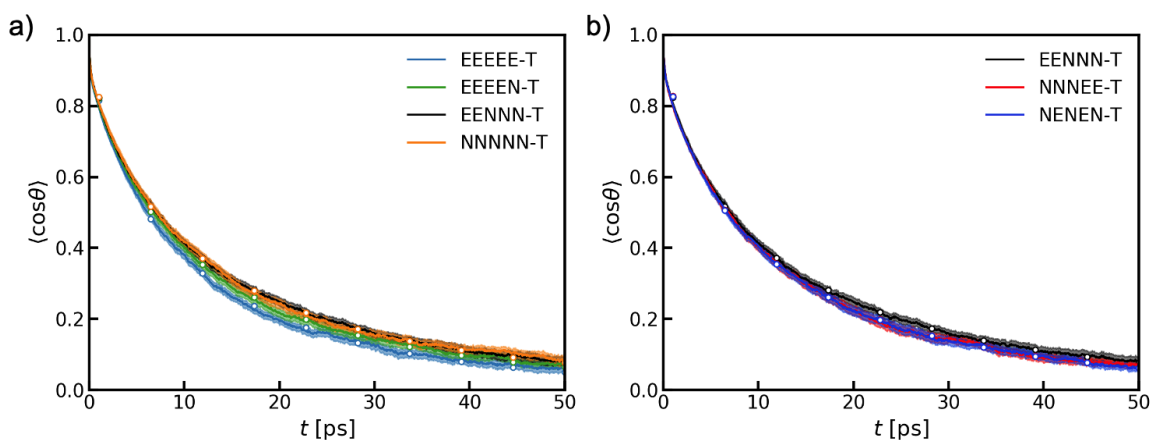


Figure D.17: Dipole correlation for the set of peptoids varying the (a) number of alkyl residues at the C-terminus and (b) placement of three nonpolar residues in the sequence. With increasing number of nonpolar residues, dipole correlation increases (reorientation is slower).

D.4.2 Trends in rotational correlation time

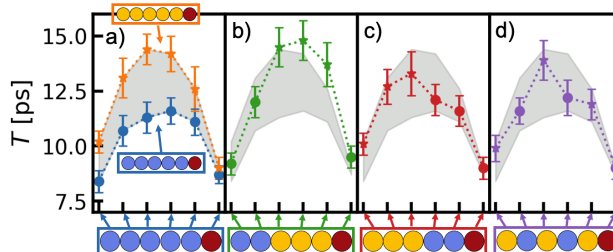


Figure D.18: Rotational correlation times measured locally around individual residues. Panel (a) contains the rotational correlation times around residues of the NME homopolymer (blue circles) and alkyl homopolymer (orange stars) which denote the lower and upper boundaries, respectively, of the shaded regions. As with the translational diffusivity, there are significant deviations between the rotational correlation times around the NME residues in the heteropolymers (circles) and the rotational correlation times around the position-equivalent residues in the NME homopolymer (lower bound of the shaded region), as well as for the alkyl residues (stars vs. upper bound of the shaded region).

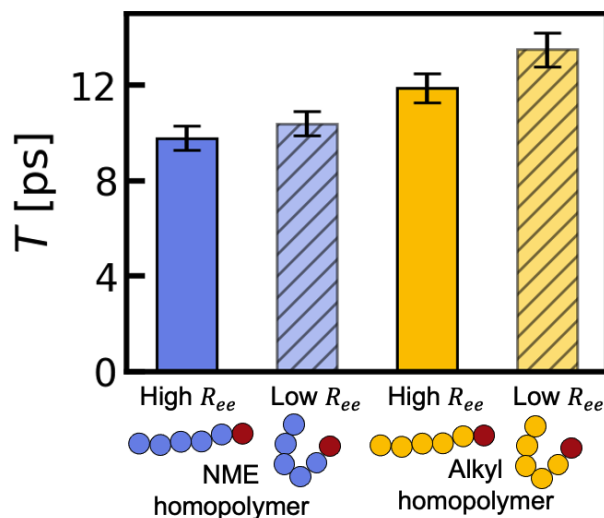


Figure D.19: Rotational correlation times for the sets of compact and extended NME homopolymers and alkyl homopolymers. As with the translational diffusivity, comparing groups of common conformation (extended NME vs. extended alkyl; compact NME vs. compact alkyl) water is slower (larger rotational correlation time) near the alkyl homopolymer. Also, as with the translational diffusivity, comparing groups of conformations for a given sequence (extended NME vs. compact NME; extended alkyl vs. compact alkyl) water is slower near the more compact conformations.

D.4.3 Scaling down NME charges

There are slight topological differences between the alkyl and NME monomers. To confirm that changes in the local chemical environment alone can lead to changes in dynamics, we measure water dynamics near the NME homopolymer peptoid with charges scaled down to a quarter of their original values (the original charges can be found in Table C.2) so that the sidechain is less polar, but approximately preserving the bond lengths and van der Waals radii. We again observe shifts towards reduced local density, enhanced tetrahedrality, and slower translational and orientational dynamics of water (Fig. D.20), suggesting that the slight conformational change between alkyl and NME monomers does not affect the overall conclusion that sequence affects water dynamics independent of conformation.

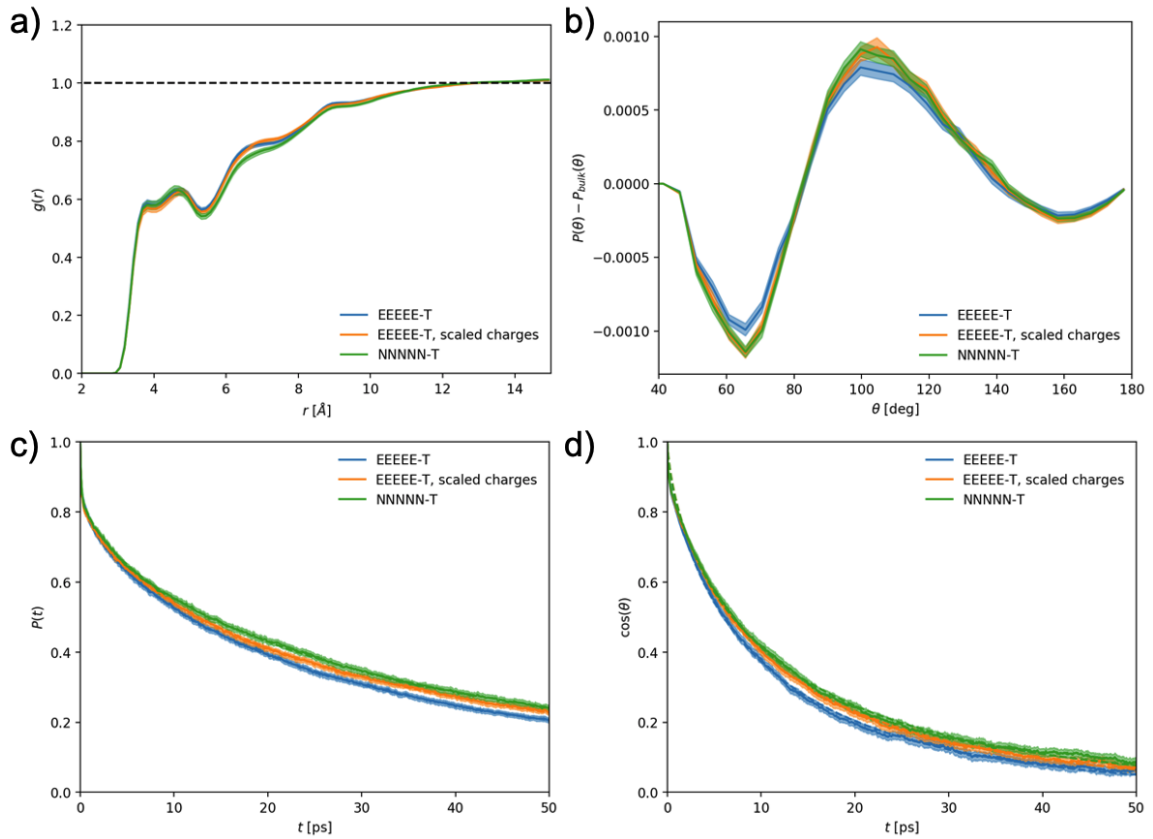


Figure D.20: The (a) radial distribution function, (b) difference in the three-body angle distribution compared to that of bulk water, (c) survival probability, and (d) dipole orientation correlation for the EEEEE-T peptoid with charges scaled down, along with that for the unmodified EEEEE-T and NNNNN-T for reference.

D.4.4 Comparison of water behavior around unlabeled peptoids

To check that the spin label does not qualitatively affect the observed trends in water structure and dynamics with respect to sequence, we ran the same calculations (expanded ensemble simulation, followed by NVE simulations, followed by water structure and dynamics calculations) for the unlabeled NME homopolymer (EEEEEE according to the labeling in Fig. 5.1) and the unlabeled alkyl homopolymer (NNNNN according to the labeling in Fig. 5.1). Fig. D.21 shows that water around the alkyl homopolymer still shows a reduction in dynamics and enhancement in tetrahedrality, consistent with the results for the labeled peptoids.

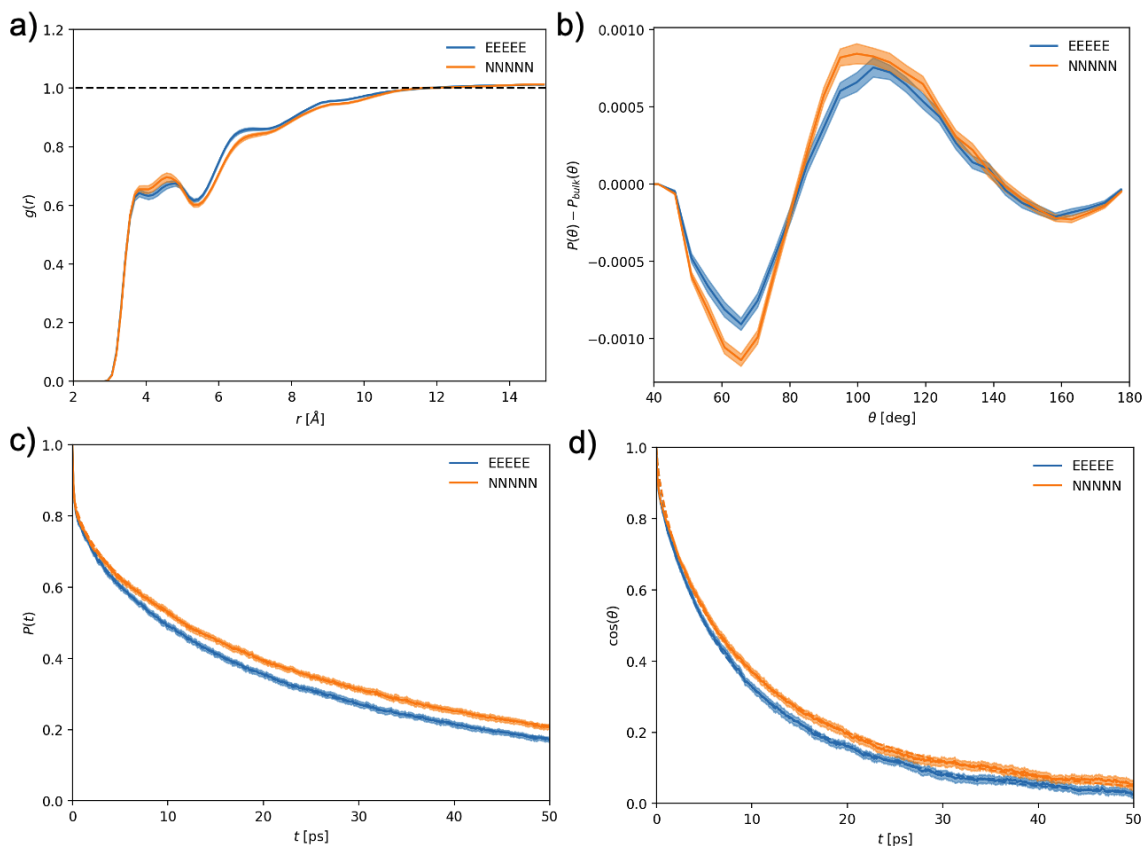


Figure D.21: The (a) radial distribution function, (b) difference in the three-body angle distribution compared to that of bulk water, (c) survival probability, and (d) dipole orientation correlation near the unlabeled, alkyl homopolymer, compared to that around the unlabeled, NME homopolymer. These trends (less dense, more tetrahedral, and slower water near the alkyl homopolymer compared to the NME homopolymer) are consistent with the trends in the labeled peptoids.

D.4.5 Varying the cutoff for water behavior calculations

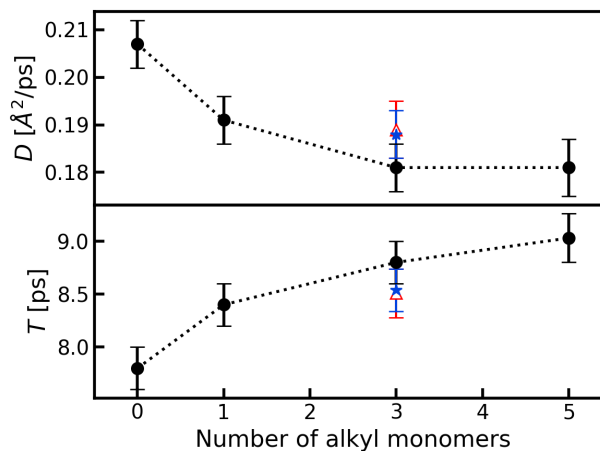


Figure D.22: Same as Fig. 5.4 except using a $7\text{\AA}$ cutoff around the peptoid for the dynamics calculations, instead of a $5\text{\AA}$ cutoff. Even with the larger cutoff, the trends in dynamics with respect to sequence are unchanged.

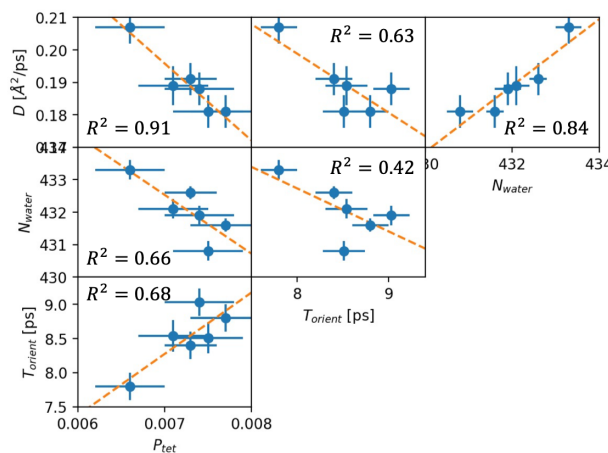


Figure D.23: Same as Fig. 5.8 except using a $7\text{\AA}$ cutoff around the peptoid for the dynamics calculations, instead of a $5\text{\AA}$ cutoff. The correlations are weaker with the larger cutoff, suggesting that the range of the peptoid's effect differs for the various structural and dynamical quantities, consistent with what is found in Ref. [387].

D.4.6 Additional structure-dynamics correlations

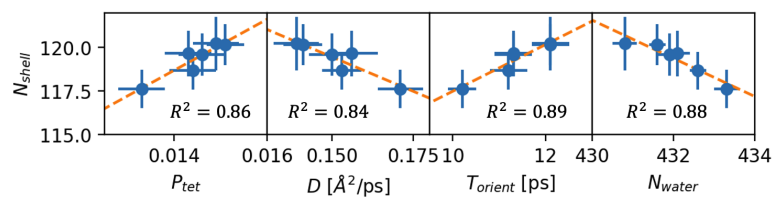


Figure D.24: Correlation between hydration metrics plotted in Fig. 5.8 and N_{shell} , the number of waters within $5\text{\AA}$ of a peptoid heavy atom. As with the other single-body density metric (N_{water} , computed by integrating the radial distribution function), N_{shell} also shows some correlation with the reduction in dynamics and enhancement in tetrahedrality, but it is weaker than the correlations between the dynamical metrics and enhancement in tetrahedrality.

Appendix E

Supporting Information for Relationships between Water's Structure and Solute Affinity

E.1 Additional simulation details

E.1.1 Expanded ensemble simulations

For each peptoid surface, we simulate a single peptoid of the same length and chemistry tethered to the self-assembled monolayer (SAM) surface to obtain the equilibrium fraction of cis/trans for each amide bond. We use the expanded ensemble methodology developed in Ref. [212] to sample over slow amide bond isomerizations. In order to enhance sampling over additional peptoid conformational degrees of freedom, we apply a bias based on the distance in the z -direction (perpendicular to the surface) between the C-terminal end of the peptoid and the sulfur atoms in the SAM. The bias has the form:

$$U_{\text{rep}} = \begin{cases} k(\Delta z_{\text{max}} - \Delta z)^2 & \Delta z_{\text{max}} \geq \Delta z \\ 0 & \Delta z_{\text{max}} < \Delta z \end{cases} \quad (\text{E.1})$$

where $\Delta z = z_{\text{peptoid}} - z_{\text{SAM}}$, z_{peptoid} is the z -coordinate of the nitrogen in the C-terminal amino group of the peptoid, and z_{SAM} is the mean z -coordinate of the SAM sulfur atoms. The force constant is $k = 500 \text{ kJ/mol/nm}^2$ and we conduct separate simulations for $\Delta z_{\text{max}} = 0.2, 0.7, 1.2, 1.7, 2.2, 2.7 \text{ nm}$. This potential, U_{rep} , biases the peptoid to adopt conformations in which the C-terminal end is more extended from the surface.

We perform a 100-ps NVT equilibration followed by a 100-ps NPT equilibration. We then run an expanded ensemble simulation over softened amide-bond dihedral states where the dihedral potentials controlling rotation around the amide bond are scaled by a factor λ_{ee} that takes on the values 1, 0.95, 0.91, 0.81, 0.71, 0.60, 0.50, 0.45, 0.41, 0.3875, 0.35, 0.3375, 0.3, 0.2875, 0.25, 0.2375, 0.2, 0.175, 0.1, 0.033, 0.016, 0. We let the expanded ensemble weights equilibrate over 1 μs and then simulate with fixed weights for 4 μs . If the expanded ensemble states are not sampled evenly with the equilibrated weights, we run the equilibration stage for another 1 μs . We use MBAR^{125,218} to reweight back to the unbiased ($U_{\text{rep}} = 0$ and $\lambda_{ee} = 1$) state and compute the free energy difference between the cis and trans states for each amide bond, where we define the cis state as $-90^\circ < \omega < 90^\circ$, with ω the value of the backbone $\text{C}_{\alpha}\text{-C-N-C}_{\alpha}$ dihedral, and the trans states as all other values of ω . These free energies are shown in Fig. E.1.

Generally, the amide bonds at the ends of the peptoid favor the trans state while those in the center prefer the cis state. However, the strength of these preferences depends on peptoid chemistry. We compute the fraction of cis states with:

$$f_{\text{cis}} = 1 - \frac{1}{1 + \exp(G_{\text{trans}} - G_{\text{cis}})} \quad (\text{E.2})$$

E.1.2 Peptoid brush surface simulations

We solvate the system using GROMACS 2016.^{214,215} with OPC water for an initial box length of 11 nm. For simulations with a solute, we initialize the solute above one of the

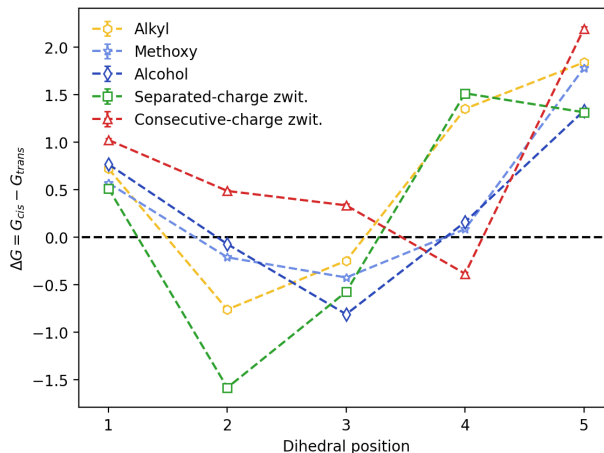


Figure E.1: The free energy difference between the cis and trans states for each isomerizable amide bond (dihedral position of 1 corresponds to the amide bond closest to the surface, 5 corresponds to that furthest from the surface). The dotted lines are meant to guide the eye.

peptoid surfaces, also using GROMACS. We initialize peptoids in extended conformations, created and geometry-minimized in Avogadro 1.2.0.²¹³

To tether the peptoids to the SAM, we apply a flat-bottom harmonic potential between the N-terminus nitrogen of the peptoid and a methyl-headgroup carbon of the SAM, chosen so that the peptoids are evenly spaced on the SAM. The potential has the form $U = \text{step}(r - r_0)0.5k(r - r_0)^2$, where r is the distance between the two atoms, $r_0 = 5\text{\AA}$, and $k = 1000\text{ kJ/mol/nm}^2$.

To ensure that the peptoids adopt the assigned cis/trans states, we apply a dihedral restraint of the form $U_{dih} = k(\theta - \theta_0)^2$, where $k = 100\text{ kJ/mol/rad}^2$, $\theta_0 = 0$ for cis conformations and $\theta_0 = 180$ for trans ones during the first equilibration stage of the simulation. The restraint is removed for subsequent equilibration and production steps, but the barrier for isomerization is sufficiently high that the peptoids typically remain in their initial isomerization states.

For simulations without a solute, we first perform a 100-ps NVT equilibration, followed by a 100-ns NPT equilibration and a 1- μ s NPT production simulation with configurations saved every 100 ps, which we use to compute water structural properties. For dynamics calculations, we replace the last NPT run with a 10-ns NVT run (the box length is the average from the second half of the NPT equilibration), with configurations saved every 1 ps.

For umbrella sampling simulations, we use a potential of the form $U = k(|z_{\text{SAM}} - z_{\text{solute}}| - z_0)^2$ where z_{SAM} is the position of the sulfur atoms along the z -axis (perpendicular to the surface), z_{solute} is the location of the centroid of the solute heavy atoms, $k = 100\text{ kJ/mol/nm}^2$, and z_0 is the restrained distance. We first restrain the solute far from the surface, $z_0 = 4.6\text{ nm}$ (which translates to around 2 nm from the peptoid-water interface) and perform a 100-ps NVT equilibration followed by a 20-ns NPT equilibration. We then

sample 40 states equally spaced in the region $1.5 \text{ nm} \leq z_0 \leq 4.6 \text{ nm}$, starting from the state where the solute is restrained farthest from the surface and moving the solute closer to it. We simulate each state for 20 ns, saving configurations every 200 ps. We discard the first 2 ns of each state before analysis.

We use the standard error of an observable over independent repeats of a simulation in order to compute the reported uncertainties. For simulations without a solute, we run 5 repeats of each surface. For umbrella sampling simulations, we run 10 repeats of each solute-surface combination for the peptoid surfaces and 5 repeats for the PEO surface.

Instantaneous interface

To compute the instantaneous interface, we follow the procedure described in Ref. [69]. We first compute a coarse-grained density field, $\bar{\rho}(\mathbf{r}, t) = \sum_i \phi(|\mathbf{r} - \mathbf{r}_i(t)|; \xi)$ where $\phi(\mathbf{r}; \xi) = (2\pi\xi^2)^{-3/2} \exp(-|\mathbf{r}|^2/2\xi^2)$. Here, we set $\xi = 4.8\text{\AA}$ and compute the density field by discretizing the system into cubes of length $2\text{\AA}$. To identify the location of the interface we then determine the points, r , for which $\rho(\mathbf{r}, t) = 0.016\text{\AA}^{-3}$.

To compute the average location of the surface-water interface, we first compute the instantaneous interface for each snapshot. We then take the average of the z positions (in the direction perpendicular to the surface) of the points in the instantaneous interface and compute the distance from the z position of the sulfur atoms in the self-assembled monolayer (SAM). We average this distance over all snapshots to get the average location of the surface-water interface with respect to the SAM sulfur atoms. These locations are listed in Table E.1 and used to shift the relationships shown in Fig. 6.2 and Fig. 6.4 (along with SI Fig. E.4, E.6, E.10-E.14, E.17-E.19) so that the x -axis is distance from the surface-water interface, or $\Delta z = z - z_{\text{surf}}$, where z_{surf} is the mean interface.

Surface	$z_{\text{surf}} [\text{\AA}]$
Alkyl	26.91
Methoxy	25.81
Alcohol	25.71
Separated-charge zwit	25.88
Consecutive-charge zwit	25.68

Table E.1: Locations of the surface-water interface, z_{surf} , for each surface chemistry.

To obtain the interfacial distribution, for each snapshot we take the z positions of all points in the instantaneous interface, subtract the mean z position, and bin them over all snapshots to create a histogram with bin width $0.33\text{\AA}$. Fig. E.2 shows these distributions.

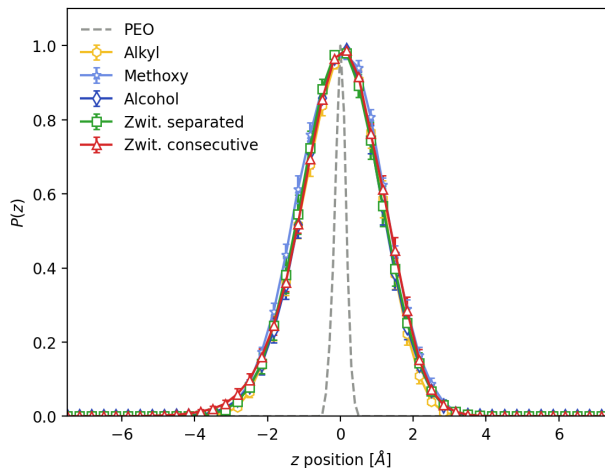


Figure E.2: The distribution of the instantaneous surface-water interface in the direction perpendicular to the surface. For the peptoid surfaces, the error bars give the standard error over 5 independent runs (with 2 interfaces per run—one on each side of the surface). All distributions are normalized so that the maximum is 1.

E.1.3 PEO brush surface simulations

The PEO chains are arranged with Au-hcp spacing, as in Ref. [318] on an 8x9 grid. We add OPC waters to the PEO system using the Packmol³⁸⁸ software with an initial box length of 12 nm. To initialize simulations with a solute, we place the solute above the PEO surface again using Packmol.

Prior to conducting umbrella sampling calculations, we equilibrate the hydrated PEO surface for 5 ns in the NVT ensemble at 500 K to relax the PEO chain configurations more rapidly. We follow this melting simulation with a 5-ns NPT ensemble simulation at 298.15 K and 1 bar to cool the system to the desired production temperature. Finally, we run a production simulation in the NPT ensemble at 298 K and 1 bar for 100 ns, saving the system coordinates every 10 ps. We generate solute-free structural information such as the density profile and three-body angle distributions using this trajectory.

The umbrella sampling procedure for the PEO surface is essentially identical to the peptoid surface procedure discussed in section E.1.2. While we apply the same harmonic centroid potential between the centroid of the SAM sulfur atoms and the centroid of the given solute molecule as the peptoid surfaces, we implement a force constant of $k = 1000$ kJ/mol/nm². The simulation steps are the same but with different simulation equilibrium and production times. Using Packmol,³⁸⁸ we insert a solute molecule at a random position 4 nm above the sulfur z -position of the equilibrated and hydrated PEO system (without a solute). We equilibrate the solute at this farthest position for 2 ns in the NVT ensemble followed by 10 ns in the NPT ensemble. We perform umbrella simulations at 30 umbrella states evenly spaced in the region $1.0 \text{ nm} \leq dz_0 \leq 4.0 \text{ nm}$. We briefly equilibrate each of these states in the NVT ensemble for 100 ps followed by the NPT ensemble for 500 ps. Finally, we perform each umbrella state production run

for 12 ns and save the system coordinates every 1 ps. The first 2 ns of each state is discarded as in the peptoid surface simulations.

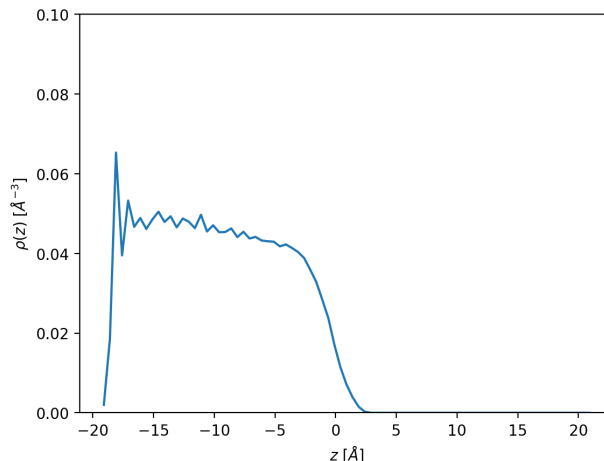


Figure E.3: The density profile for PEO heavy atoms shows that the number density in the surface is similar to that for the peptoid brush surface.

E.1.4 Bulk water simulations

We performed simulations of bulk water as described in section D.1.3. These simulations are used to compute water density, the three-body-angle distribution, diffusivity, and the average number of hydrogen bonds formed by a water molecule in bulk.

E.2 Water properties

We include details of the calculations of water's dynamical and structural properties.

E.2.1 Water dynamics

To compute the diffusivity perpendicular to the surface, D_z , we first divide the box into rectangular regions, along the axis perpendicular to the surface, of width $5.5\text{\AA}$ and track the probability that a water molecule in the region at some initial time $t = 0$ will be in the region at time t over a region $0 < t < 100$ ps. To obtain the diffusivity, we fit this probability over the entire region to that computed from solving the 1D diffusion equation in an infinite domain with initial concentration c_0 in a region of length L :

$$P(t) = c_0 \left[\text{erf} \left(\frac{L}{2\sqrt{D_z t}} \right) + \sqrt{\frac{4D_z t}{\pi L^2}} \left[\exp \left(-\frac{L^2}{4D_z t} \right) - 1 \right] \right] \quad (\text{E.3})$$

Here, c_0 and D_z are the fitted parameters, while $L = 5.5\text{\AA}$. We interpolate the distance-dependent diffusivity at a distance of $5\text{\AA}$ to obtain the value in Table 6.1.

To compute the diffusivity parallel to the surface, we divide the box into the same rectangular regions used for the survival probability calculation. In each region, we compute the mean-squared displacement in the xy -plane (parallel to the surface) of water molecules that are inside the region at $t = 0$, over intervals of 100 ps. We then fit a line to the mean-squared displacement over the interval $20 < t < 100$ ps and divide the slope by 4 to obtain the diffusivity parallel to the surface.

The trends in both diffusivities are qualitatively similar, as shown in Fig. E.4.

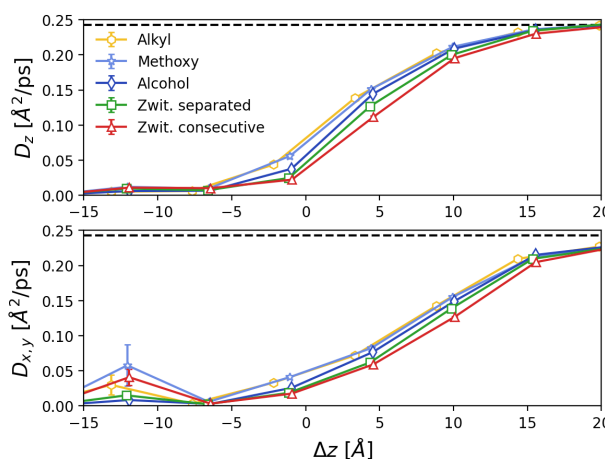


Figure E.4: Diffusivity as a function of distance from the surface-water interface in the directions (top) perpendicular to the surface, computed from a survival probability fit to a diffusion equation and (bottom) parallel to the surface, computed from a mean-squared displacement. The dashed black line shows the diffusivity of bulk water, $D = 0.243\text{\AA}^2/\text{ps}$ (section E.1.4). The two calculations give diffusivities that follow the same qualitative trends. Far away from the surface, water approaches its bulk diffusivity. However, the perturbation to water’s dynamics is long-ranged, slowing water even at distances where water’s density is at bulk ($\Delta z = 15\text{\AA}$), consistent with earlier work showing that a protein surface has a longer-ranged effect on translational dynamics compared to translational structure.³⁸⁷ The error bars give the standard error computed from 5 independent simulations, as described in section E.1.2.

E.2.2 Density profiles

E.2.3 Hydrogen bonding

Hydrogen bonds are determined *via* geometric criteria: the distance between the donor and acceptor oxygen/nitrogen atoms must be less than $3.5\text{\AA}$ and the angle formed by the acceptor oxygen/nitrogen – donor oxygen/nitrogen – donor hydrogen must be less than 30° . All peptoid oxygens are considered possible acceptors, along with primary

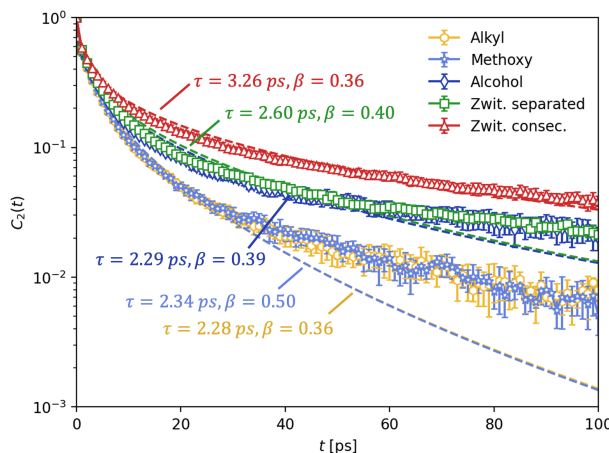


Figure E.5: The O-H bond reorientation correlation function for waters in the hydration layer (identified as described in sections 6.2.2 and E.1.2). The correlation function³⁸⁹ is computed as $C_2(t) = \langle P_2[e(0)\dot{e}(t)] \rangle$, where P_2 is the Legendre polynomial of rank 2 and $e(t)$ is the unit vector of an O-H bond at time t . The timescales of this decay are similar to those observed for model protein surfaces.³⁹⁰ The error bars give the standard error computed from 5 independent simulations, as described in section E.1.2. The dashed lines show the fits to a stretched exponential function, $C_2(t) = \exp(-(t/\tau)^\beta)$, and the labels show the fitted time constants τ and exponents β . The time constants are larger for the zwitterionic surfaces compared to the uncharged surfaces.

and secondary amine nitrogens. All peptoid hydrogens in alcohol and amine groups are considered possible donors.

E.2.4 Three-body angle distribution

The three-body angle is formed by a hydration water molecule and two of its neighbors. Neighboring waters are those within $3.4\text{\AA}$ of the central water. We compute the histogram of angles with a bin width of $\Delta\theta = 4.86^\circ$, normalize the distribution such that the integrated area is 1 and then subtract off the distribution for bulk water.

For the distributions in Fig. 6.5a, we perform this calculation for all hydration waters (waters within $7\text{\AA}$ above the instantaneous interface). For the distributions in Fig. 6.5b and those used in the projections in Fig. 6.6 (along with SI Fig. E.15, E.18, E.19) when a solute is present we consider only waters within $7\text{\AA}$ of a solute heavy atom when the solute is within $5\text{\AA}$ of the surface-water interface (around the same location where we compute the affinity). For the distributions used in the projections in Fig. 6.6 (along with SI Fig. E.15, E.18, E.19) when a solute is not present we choose random points within $5\text{\AA}$ of the surface-water interface and consider only waters within $7\text{\AA}$ of those randomly chosen points. This last procedure is a different way of computing the distribution for the surface's hydration water designed to be more comparable to how it is computed around solutes when they are near the surface.

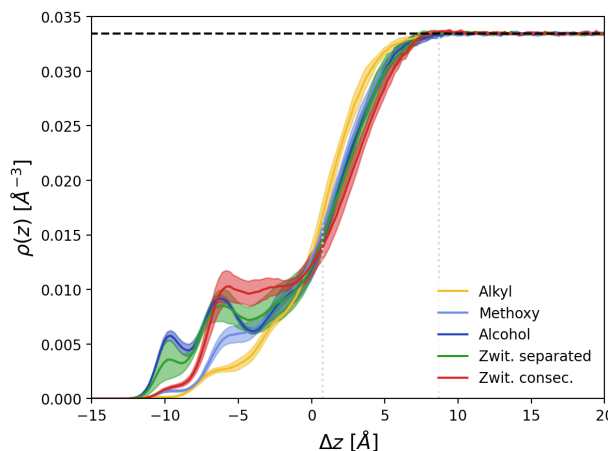


Figure E.6: The density of water as a function of distance from the water-peptoid mean interface (computed as described in section E.1.2). The dashed line indicates the density of bulk water. Positive distances are in the direction of bulk water. Water adopts its bulk density at approximately $\Delta z > 10\text{\AA}$ for all systems; moving closer to the surface, the density decreases sharply but remains finite for roughly 1 nm into the peptoid surface, defining a hydrated brush layer, and finally reaching zero before the peptoid-SAM interface (Fig. E.7). However, the zwitterionic surface’s hydration water is not significantly more dense than the alcohol surface: a small density enhancement occurs around $\Delta z = -3\text{\AA}$, but this is offset by a relatively lower density farther away at $\Delta z = 3\text{\AA}$ (Table 6.1). The zwitterionic surfaces do show distinct local density responses from each other: the consecutive-charge surface shows a single peak at $\Delta z = -6\text{\AA}$, while the separated-charge one shows two peaks, one at $z = -6\text{\AA}$ and the other at $z = -10\text{\AA}$, suggesting that water strongly hydrates both charged groups, forming separate areas of enhanced density when the charges are separated. The dotted light gray lines show the $8\text{\AA}$ region over which the density at the alkyl surface falls from that of bulk water to half its bulk value, for comparison with the PEO surface. The error bars give the standard error computed from 5 independent simulations, as described in section E.1.2.

Surface	N_{coord}
Alkyl	0.475_1
Methoxy	0.514_3
Alcohol	0.574_5
Separated-charge zwit	0.567_6
Consecutive-charge zwit	0.599_5

Table E.2: The peptoid-water coordination number, N_{coord} , is the number of waters within $7\text{\AA}$ of a peptoid heavy atom, divided by the number of peptoid heavy atoms. There is not a clear distinction between the coordination number for the zwitterionic and uncharged alcohol surfaces.

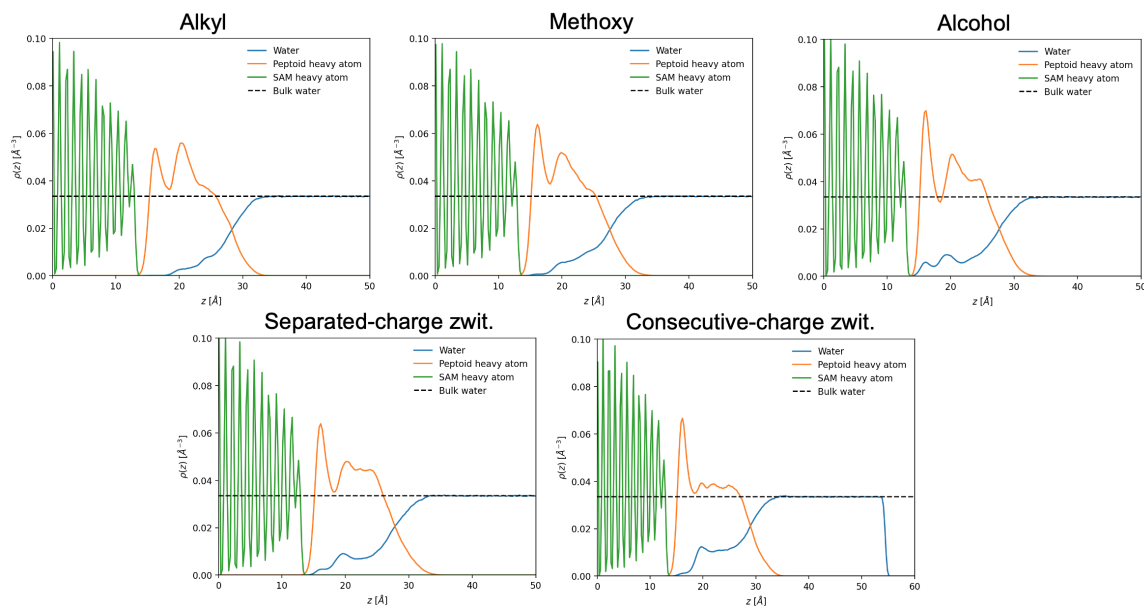


Figure E.7: Density profiles for water and the heavy atoms of the peptoid brush surface and self-assembled monolayer (SAM) for each peptoid chemistry. The dashed line shows bulk water's density. Here, $z = 0\text{\AA}$ is the location of the sulfur atoms in the SAM. The density profiles are computed with a bin width of $0.25\text{\AA}$. The peaks and shoulders in the water density profile at low z occur along with peaks in the peptoid density profile.

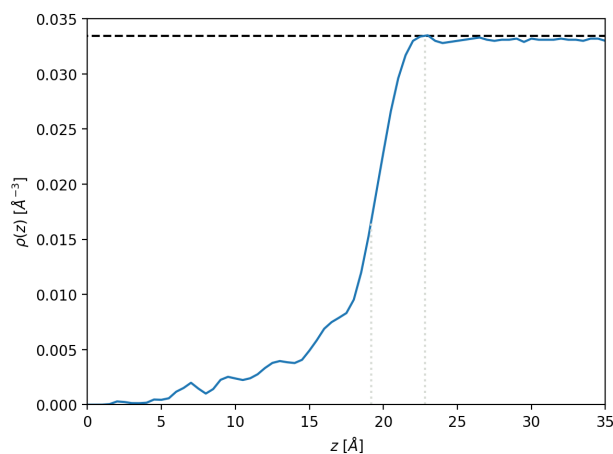


Figure E.8: The density profile of water at the PEO surface shows a sharper decline compared to the water density profile at peptoid surfaces (Fig. E.6). The dashed line indicates the density of bulk water. The dotted light gray lines show the $4\text{\AA}$ region over which the density at the PEO surface falls from that of bulk water to half its bulk value, for comparison with the peptoid surfaces.

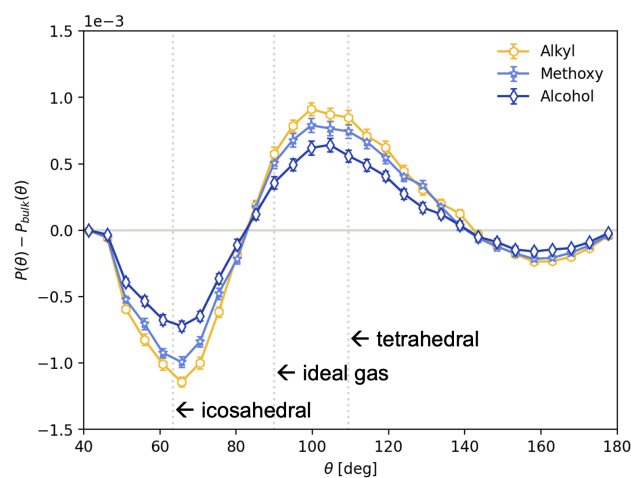


Figure E.9: The three-body angle distributions around peptoids in dilute solution, simulated as described in Ref. [245]. These peptoids have the same length, termini, and sidechain chemistries as the peptoids comprising the brush surfaces in this study, except that the last residue on the C-terminal end is a 4-amino TEMPO group. The three-body angle distribution is computed for all waters within $5\text{\AA}$ of a peptoid heavy atom.

E.3 Solute-surface interactions

E.3.1 Potential of mean force for other solutes

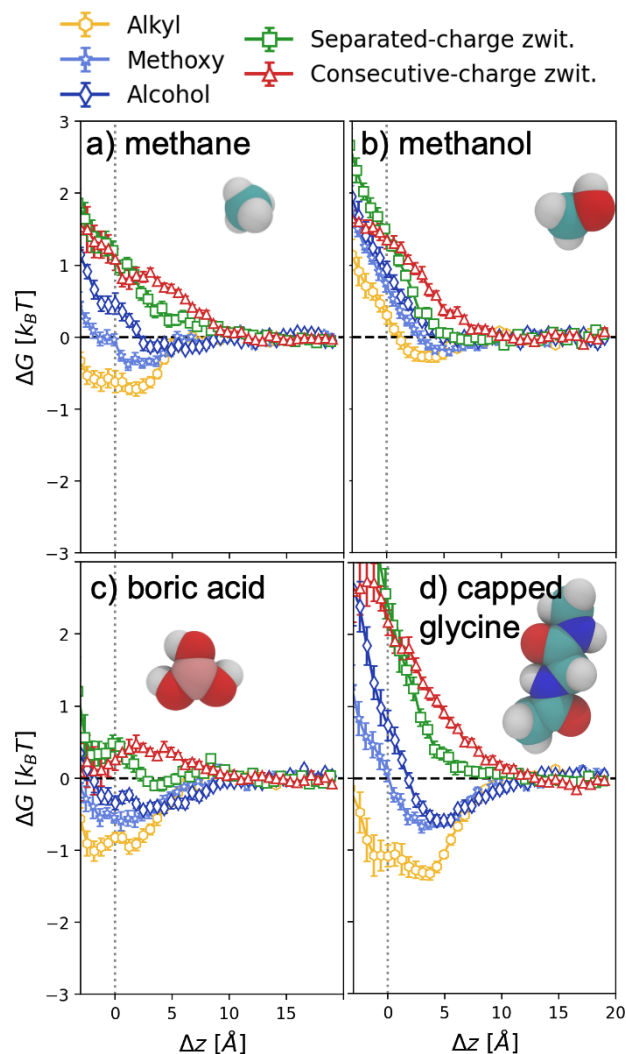


Figure E.10: Potentials of mean force (PMF), ΔG , between the peptoid surfaces and (a) methane, (b) methanol, (c) boric acid, and (d) capped glycine (not shown in the main text). As in Fig. 6.2, the PMFs are presented as a function of distance from the surface-water interface and shifted so that they reach 0 far away from the surface. The error bars give the standard error computed from 10 independent simulations, as described in section E.1.2.

We can further decompose the change in the solute-system direct energetic interaction, $\Delta \langle U_{sw} \rangle$, into electrostatic and Van der Waals components. Fig. E.12-E.13 show the electrostatic ($\Delta \langle U_{sw,Q} \rangle$) and VDW ($\Delta \langle U_{sw,VDW} \rangle$) components as a function of solute distance from the surface-water interface, shifted so that the components are 0 far away

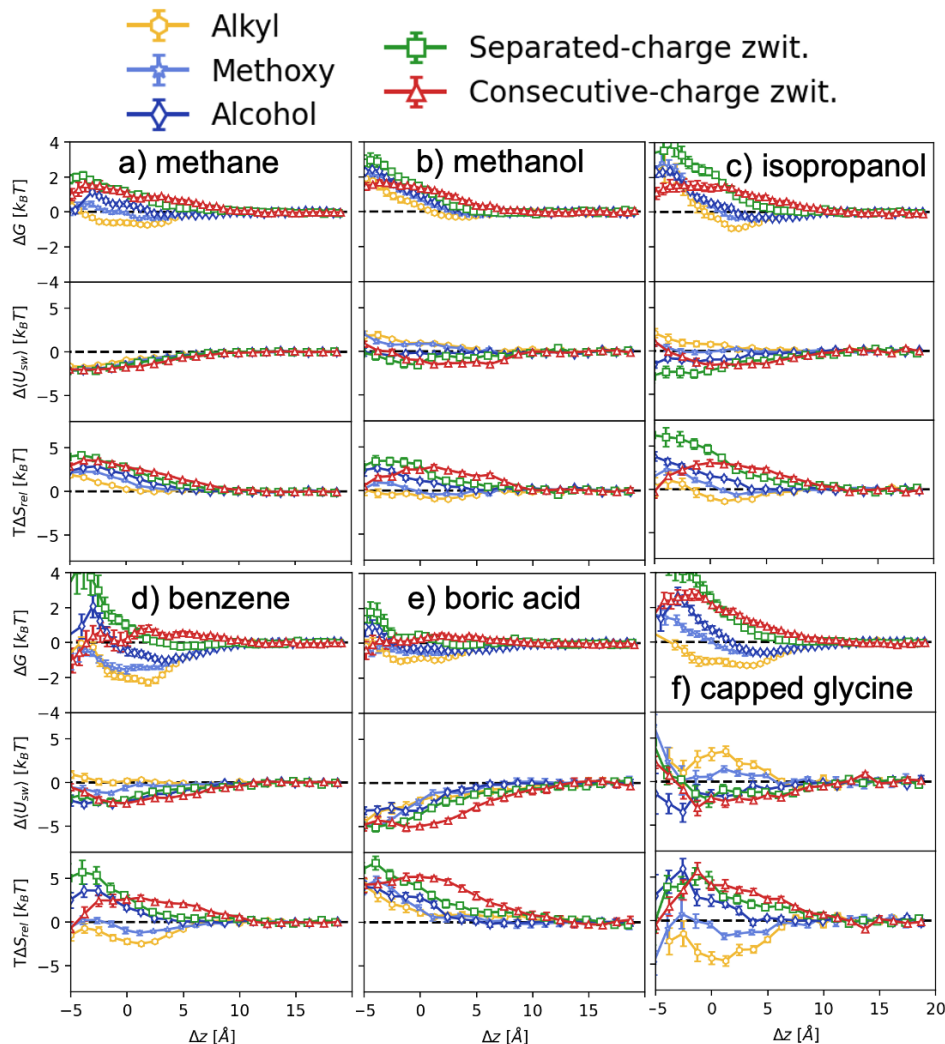


Figure E.11: Decomposition of the potential of mean force, ΔG , into direct energetic, $\Delta \langle U_{sw} \rangle$, and relative entropic, $T\Delta S_{rel}$, terms (the decomposition for phenol is shown in the main text, Fig. 6.4). The error bars give the standard error computed from 10 independent simulations, as described in section E.1.2.

from the surface. With the exception of boric acid, the solutes with hydrogen bonding groups show an unfavorable change in the solute-system electrostatic interaction (Fig. E.12) as they move towards the surface, possibly reflecting a loss of hydrogen bonds with water. Near the alcohol and zwitterionic surfaces, these hydrogen bonds are partly recovered, resulting in a smaller change in the electrostatic interaction energy. Methane has weak electrostatic interactions with both the solvent and surface, and so shows little change in the electrostatic interaction as it moves towards the surface, regardless of surface chemistry. Boric acid interestingly also shows little change in the electrostatic energy as it moves towards the surface; it is possible that its multiple donors allow it to

form more hydrogen bonds with the surface.

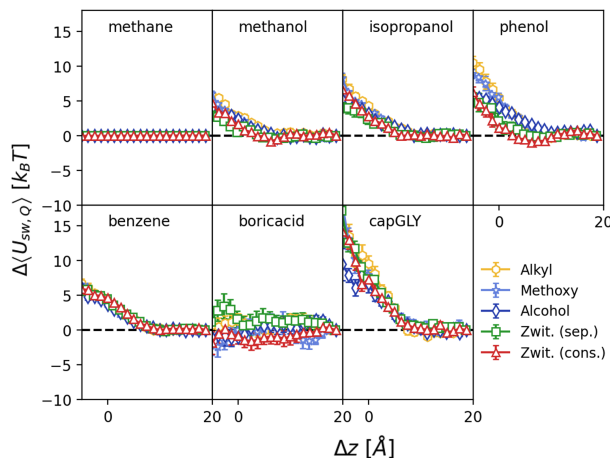


Figure E.12: Electrostatic component, $\Delta \langle U_{sw,Q} \rangle$, of the direct energetic contribution to the free energy of binding, $\Delta \langle U_{sw} \rangle$, for each surface-solute combination. The error bars give the standard error computed from 10 independent simulations, as described in section E.1.2

In contrast, the VDW component of the solute-system interaction energy (Fig. E.13) becomes more negative as the solute moves closer to the surface. This change is larger for the bulkier solutes, as they can form stronger VDW interactions with the surface. Moreover, these trends are less sensitive to surface chemistry than the trends in the electrostatic energy.

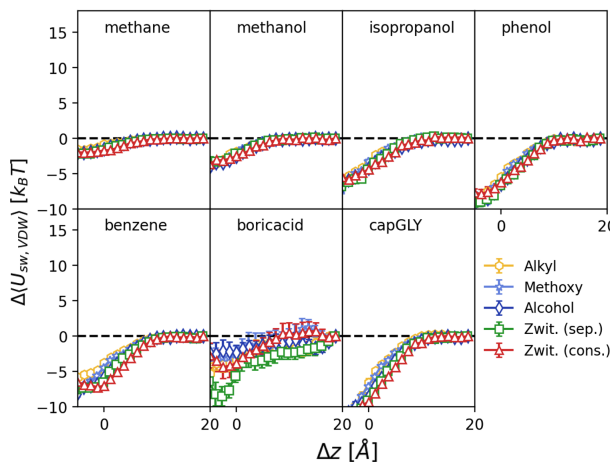


Figure E.13: Van der Waals component, $\Delta \langle U_{sw,VDW} \rangle$, of the direct energetic contribution to the free energy of binding, $\Delta \langle U_{sw} \rangle$, for each surface-solute combination. The error bars give the standard error computed from 10 independent simulations, as described in section E.1.2

E.3.2 Rearrangement of peptoid surface

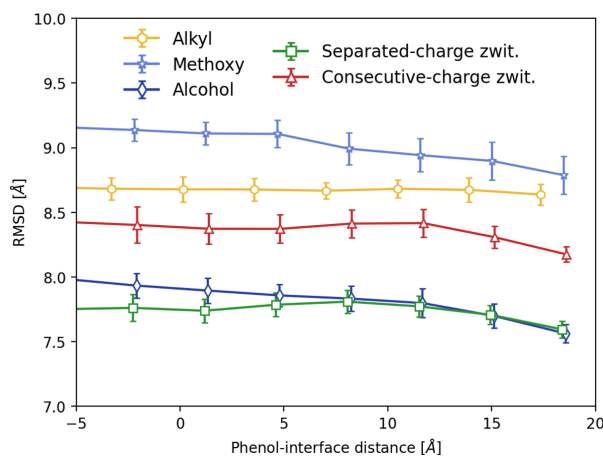


Figure E.14: The root mean squared displacement between the peptoids in the surface and a reference configuration as a function of solute-interface distance for phenol. There are no large shifts in the RMSD as the solute moves closer to the surface, suggesting no large surface rearrangements for any of the surface chemistries. The error bars give the standard error computed from 10 independent simulations, as described in section E.1.2.

E.3.3 Projections of 3-body distributions

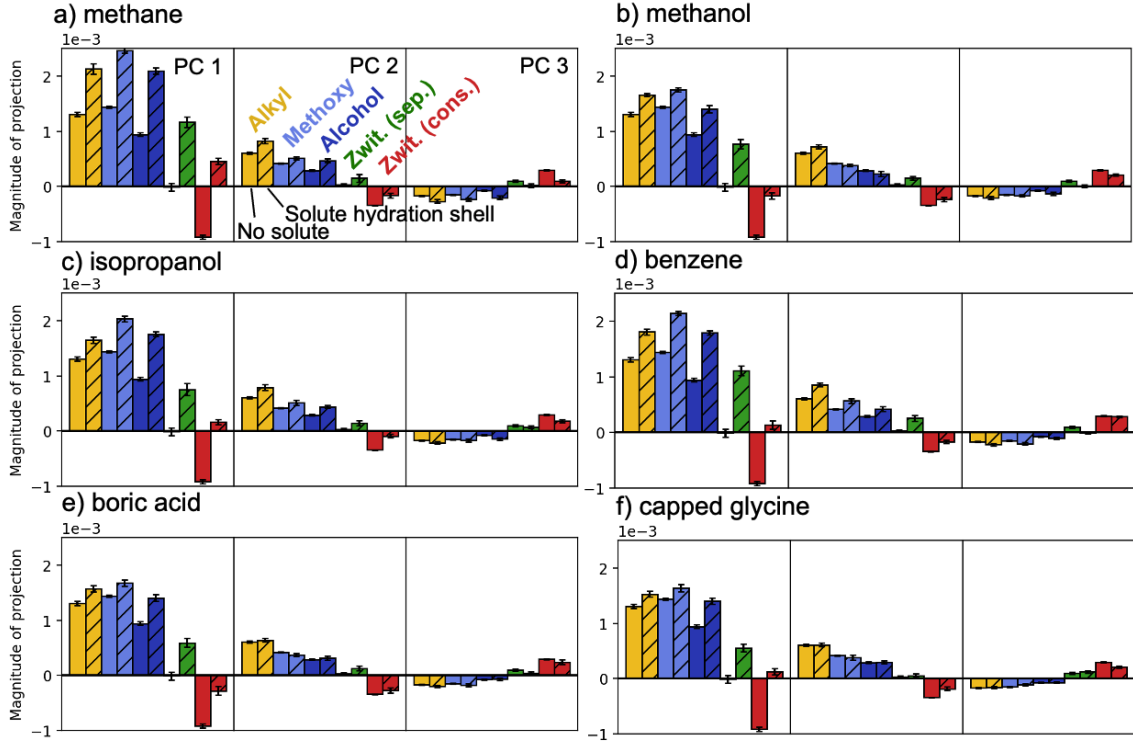


Figure E.15: Projections of the differential three-body angle distributions of water onto the first, second, and third principal components (PC1, PC2, and PC3) for the hydration water of the surface (solid bars) and in the hydration water of the solute (hatched bars) when the solute is within $5\text{\AA}$ of the surface-water interface (computed as described in section E.1.2). Details of this calculation are in section E.2.4. The values of the hydration water of the surface are independent of solute identity and are repeated in each subfigure (a-f). The projections for phenol are shown in Fig. 6.6 of the main text. The error bars give the standard error computed from 10 independent simulations, as described in section E.1.2.

E.4 Prediction of $T\Delta S_{\text{rel}}$

Inspired by the prediction of TS_{rel} for colloidal solvation using projections of the three-body angle distribution of the hydration water onto principal components in Ref. [62], we develop a model for $T\Delta S_{\text{rel}}(z) = T(S_{\text{rel,surface}}(z) - S_{\text{rel,bulk}})$. Here, z is the distance from the surface-water interface (this quantity is referred to as Δz instead of z elsewhere in the text, but we use z here to keep the equations concise).

The model for $TS_{\text{rel,bulk}}$ is a straightforward application of the model in Ref. [62], which describes the relative entropy contribution to the free energy of solvation for a solute in bulk water:

$$\frac{TS_{\text{rel,bulk}}}{N_b} = s_1 p_{1,b} + s_3 p_{3,b} + s_{13} p_{1,b} p_{3,b} \quad (\text{E.4})$$

Here, s_1 , s_3 , and s_{13} are the coefficients of the model, $p_{1,b}$ is the magnitude of the projection of the differential three-body angle distribution for waters in the solute hydration shell onto the first principal component when the solute is in bulk water, $p_{3,b}$ is the projection onto the third principal component, and N_b is the number of waters in the solute hydration shell when the solute is in bulk water. Here, the differential three-body angle distribution is the change from the three-body angle distribution of bulk water. The solute hydration shell is defined as waters within $7\text{\AA}$ of the solute. We compute these bulk quantities when the solute is furthest away from the surface (shown in Fig. E.17-E.19).

Here, we use the projections onto the first and third principal component instead of the first and second principal component used in Ref. [62] because that choice gave a lower root mean square error (RMSE) for the final prediction. We also use the number of waters in the solute hydration shell, N_b , instead of the number of three-body angles in the solute hydration shell used in Ref. [62]; this change does not significantly affect the RMSE.

To model $TS_{\text{rel,surface}}(z)$, we propose that there are two contributions, the change in water structure from bulk due to the solvation of the solute and surface and the change in the water structure from bulk due to the solvation of the surface alone:

$$\frac{TS_{\text{rel,surface}}(z)}{N(z)} = [s_1 p_1(z) + s_3 p_3(z) + s_{13} p_1(z) p_3(z)] - [s'_1 p'_1(z) + s'_3 p'_3(z) + s'_{13} p'_1(z) p'_3(z)] \quad (\text{E.5})$$

Here, s_1 , s_3 , and s_{13} are the same coefficients as before, capturing the effect of shifts in the solute hydration shell on the water restructuring contribution, while s'_1 , s'_3 , and s'_{13} are coefficients capturing the effect of shifts in the surface hydration shell on the water restructuring contribution. The terms $p_1(z)$ and $p_3(z)$ are the magnitudes of the projections of the differential three-body angle distribution for waters in the solute hydration shell onto the first and third principal component for varying solute-surface distances, z . The terms $p'_1(z)$ and $p'_3(z)$ are the magnitudes of the projections of the differential three-body angle distribution for waters at varying distances from the surface, z . Details of the three-body angle distribution calculations are in section E.2.4 and the z -dependent values are shown in Fig. E.18-E.19. Finally, $N(z)$ is the number of waters in the solute hydration shell for varying solute-surface distances; these values are shown in Fig. E.17.

Therefore, the full model for $T\Delta S_{\text{rel}}(z)$ is:

$$T\Delta S_{\text{rel}} = N[s_1 p_1 + s_3 p_3 + s_{13} p_1 p_3 - s'_1 p'_1 - s'_3 p'_3 - s'_{13} p'_1 p'_3] - N_b[s_1 p_{1,b} + s_3 p_{3,b} + s_{13} p_{1,b} p_{3,b}] \quad (\text{E.6})$$

For conciseness, we have omitted the explicit dependence on z . We fit a single model

using data from all surface-solute combinations. The comparison between the fitted and actual values of $T\Delta S_{\text{rel}}$ is shown in Fig. E.16, which shows that this model reasonably predicts $T\Delta S_{\text{rel}}$ for the zwitterionic surfaces but performs less well for the uncharged surfaces. The RMSE for this dataset is $0.58k_B T$ and the values of the fitted coefficients are: $s_1 = 0.20$, $s_3 = 0.22$, $s_{13} = 0.20$, $s'_1 = -1.16$, $s'_3 = 1.39$, $s'_{13} = -0.34$.

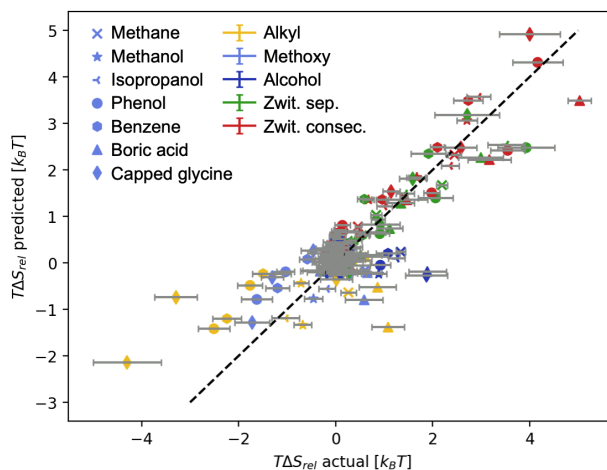


Figure E.16: Comparison of the relative entropy contribution, $T\Delta S_{\text{rel}}$, to the solute-surface PMF predicted using the model in equation E.6 to the actual computed values for each solute (varying marker shape) and surface (varying colors) combination, over varying solute-surface distances. The error bars give the standard error in $T\Delta S_{\text{rel}}$ computed from 10 independent simulations, as described in section E.1.2. The dashed line is $x = y$.

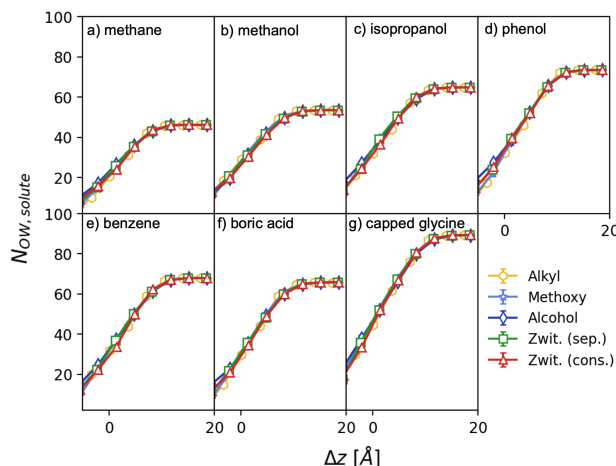


Figure E.17: Number of waters within $7\text{\AA}$ of the solute heavy atoms for each solute at varying distances from each surface. These values are used for $N(\Delta z)$ in the model for $T\Delta S_{\text{rel}}$ in equation E.6. The values at the largest Δz are used for N_b . The error bars, which are smaller than the marker size here, give the standard error computed from 10 independent simulations, as described in section E.1.2.

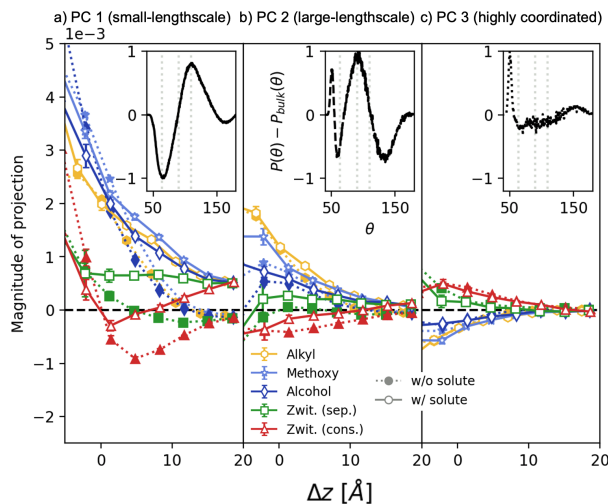


Figure E.18: Projections of the differential three-body angle distributions of water onto the (a) first, (b) second, and (c) third principal components for water in the hydration shell of phenol (open markers, solid lines) and water at randomly chosen locations without phenol (filled markers, dotted lines) as a function of distance from the surface-water interface. These values are used in the model for $T\Delta S_{\text{rel}}$ in equation E.6. The values at the largest Δz are used as bulk values in the model. The insets show each principal component and the faint dotted lines indicate the populations of icosahedral ($\theta = 63.4^\circ$), ideal gas-like ($\theta = 90^\circ$), and tetrahedral ($\theta = 109.5^\circ$) water. The error bars give the standard error computed from 10 independent simulations, as described in section E.1.2.

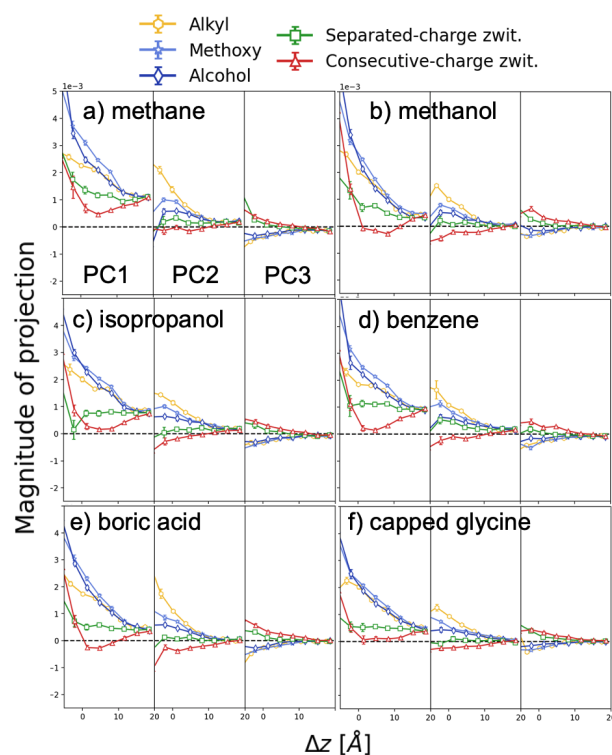


Figure E.19: Same as Fig. E.18 except for the water in the hydration shell of other solutes. These values are used in the model for $T\Delta S_{\text{rel}}$ in equation E.6. The values at the largest Δz are used as bulk values in the model. The error bars give the standard error computed from 10 independent simulations, as described in section E.1.2.

Appendix F

Supporting Information for Effect of Polypeptoid Sequence on Conformation

F.1 Simulation details

F.1.1 Atomistic simulations

We create initial polypeptoid configurations in the all-trans, fully extended conformation using the mBuild package.³⁹¹ We set the box length to be at least 20Å larger than the end-to-end distance in this fully extended conformation and solvate the system using GROMACS.^{214,215} The simulations are run following the procedure outlined in Ref. [212] using λ_{ee} states of 1, 0.975, 0.95, 0.91, 0.85, 0.79, 0.75, 0.71, 0.65, 0.6, 0.575, 0.55, 0.525, 0.5, 0.475, 0.4625, 0.45, 0.425, 0.4125, 0.4, 0.3875, 0.375, 0.3625, 0.35, 0.3375, 0.325, 0.3125, 0.3, 0.2875, 0.275, 0.2625, 0.25, 0.2375, 0.225, 0.21, 0.195, 0.185, 0.175, 0.167, 0.15, 0.11, 0.075, 0.055, 0.033, 0.017, 0.0.

Sequence	l_b [Å]	$R_{ee,max}$ [Å]	N_w	T_{equil} [ns]	T_{prod} [ns]	N_s
$N_{ARO} = 0$	85.14	60.3	20,139	1000	10,000	135
$N_{ARO} = 1$	87.79	62.9	21,936	1000	10,000	120
$N_{ARO} = 2$	85.93	61.2	20,641	3000	10,000	173
$N_{ARO} = 4$	82.43	57.3	18,168	4000	10,000	27

Table F.1: The initial box length, l_b , and number of waters, N_w , are set as described in section 5.2.1. The maximum end-to-end distance along the backbone, $R_{ee,max}$ is estimated from distance between the N-terminal residue’s nitrogen and the C-terminal capping group’s nitrogen in the initial, fully extended configuration. T_{equil} and T_{prod} give the lengths of the equilibration (varying weights) and production (fixed weights) simulations, respectively. To compute the number of independent and statistically significant conformations, N_s , we follow the procedure described in section C.1.4.

F.1.2 Coarse-grained simulations

The coarse grained (CG) model has a much lower barrier to rotation around the backbone amide bond (i.e. the sidechain-backbone-backbone-sidechain bead dihedral) compared to the atomistic model. Specifically, the potential controlling this rotation has a barrier of around 1 kcal/mol, as opposed to over 15 kcal/mol in the MFTOID model.²⁰⁶ As a result, the timescale associated with isomerization rate is short in simulations with the CG model. We compute this timescale for each such dihedral in the alternating sequence with $N = 11$ monomers by dividing the length of the simulation by the number of times the dihedral changes from “cis” ($-\pi/2 < \omega < \pi/2$) to “trans” (all other values of ω). The average timescale is 0.05 ns, which is approximately two orders of magnitude faster than the timescale associated with isomerization in the expanded ensemble atomistic simulation of the same peptoid (Fig. C.7). Consequently, the end-to-end distances rapidly decorrelate in the CG simulation (Fig. F.1).

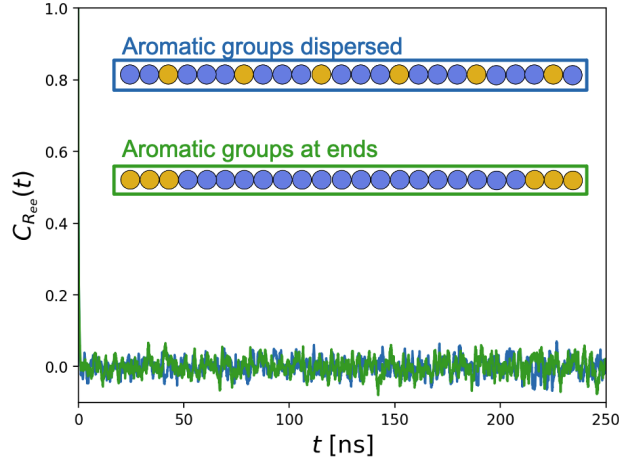


Figure F.1: The autocorrelation of the end-to-end distances, $C_{R_{ee}}$, decays quickly in the coarse-grained simulation, indicating rapid decorrelation of these distances.

F.2 Comparison to theoretical form of $P(R_{ee})$

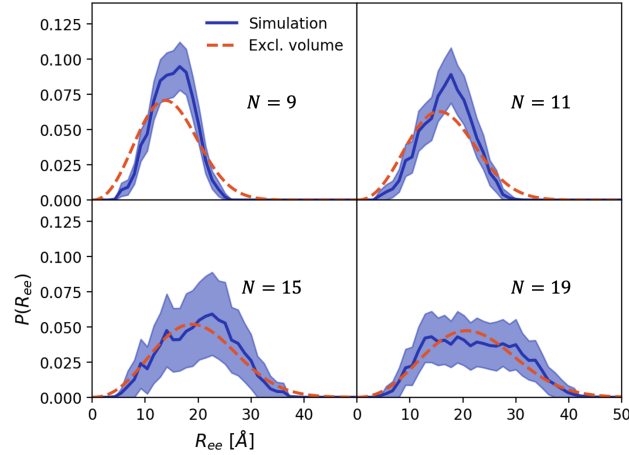


Figure F.2: The blue lines give the $P(R_{ee,NN})$, the end-to-end distance distribution measured from simulation (with the MFTOID forcefield) in between terminal nitrogens on the polypeptoid backbone (Fig. 7.2 inset) with shaded regions showing a 68% confidence interval, computed as described in section C.1.4. The orange, dashed lines show a theoretical $P(R_{ee})$ given by: $P(R_{ee}) = a_1 x^{\epsilon_1} \exp(-a_2 x^{\epsilon_2}) 4\pi x^2$ where $x = R_{ee} / \langle R_{ee}^2 \rangle^{1/2}$ (distribution from Ref. [392]) and $\epsilon_1 = 0.269$, $\epsilon_2 = 2.427$, $a_1 = 0.299$, and $a_2 = 1.269$. At longer lengths, the theoretical distribution is consistent with the simulated distribution, suggesting that the polypeptoid backbone follows excluded volume scaling.

F.3 Surrogate model

The surrogate models are fit to the data from the first 20 generations (160 sequences). Separate models are fit for predicting $\langle R_{ee}^2 \rangle$ and $\langle R_g^2 \rangle$. Fig. F.3 shows the comparison of the predicted and simulated metrics for the sequences used to fit the model. The coefficients suggest that the first feature is generally less important than the second and third features in predicting the conformational metrics. Both end-to-end distance and the radius of gyration are negatively correlated with the degree of spreading of aromatic groups about the center of the sequence as well as the clustering of the aromatic groups. The spread about the center is slightly more important for the end-to-end distance (consistent with Fig. 7.3 showing that hydrophobic groups at ends significantly shift the end-to-end distance distribution), while the clustering is slightly more important for the radius of gyration.

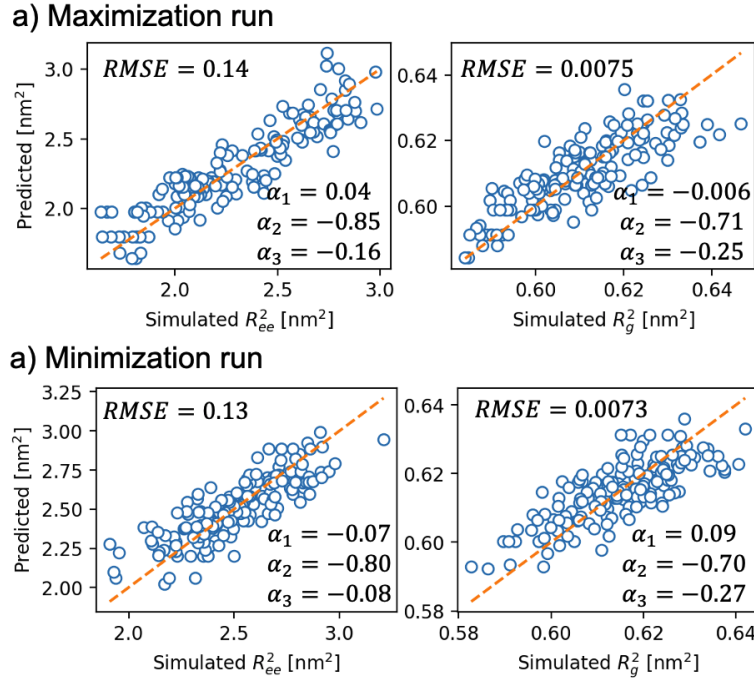


Figure F.3: Mean squared end-to-end distances and radii of gyration predicted from the surrogate model and simulations for the genetic algorithm runs (a) maximizing and (b) minimizing deviation from the Flory relationship. The markers indicate individual sequences and the dashed line is $y = x$. The labels give the root mean squared error and coefficients of the linear model.

References

- [1] A. Ludwig. Discovery of new materials using combinatorial synthesis and high-throughput characterization of thin-film materials libraries combined with computational methods. *npj Computational Materials*, 5(1):70, July 2019. doi:10.1038/s41524-019-0205-0.
- [2] I. G. Clayson, D. Hewitt, M. Hutereau, T. Pope, and B. Slater. High Throughput Methods in the Synthesis, Characterization, and Optimization of Porous Materials. *Advanced Materials*, 32(44):2002780, Nov. 2020. doi:10.1002/adma.202002780.
- [3] A. Zunger. Inverse design in search of materials with target functionalities. *Nature Reviews Chemistry*, 2(4):0121, Mar. 2018. doi:10.1038/s41570-018-0121.
- [4] J. G. Freeze, H. R. Kelly, and V. S. Batista. Search for Catalysts by Inverse Design: Artificial Intelligence, Mountain Climbers, and Alchemists. *Chemical Reviews*, 119(11):6595–6612, June 2019. doi:10.1021/acs.chemrev.8b00759.
- [5] S. Molesky, Z. Lin, A. Y. Piggott, W. Jin, J. Vucković, and A. W. Rodriguez. Inverse design in nanophotonics. *Nature Photonics*, 12(11):659–670, Nov. 2018. doi:10.1038/s41566-018-0246-9.
- [6] B. Kuhlman and P. Bradley. Advances in protein structure prediction and design. *Nature Reviews Molecular Cell Biology*, 20(11):681–697, Nov. 2019. doi:10.1038/s41580-019-0163-x.
- [7] K. Sattari, Y. Xie, and J. Lin. Data-driven algorithms for inverse design of polymers. *Soft Matter*, 17(33):7607–7622, 2021. doi:10.1039/D1SM00725D.
- [8] G. Chen, Z. Shen, A. Iyer, U. F. Ghumman, S. Tang, J. Bi, W. Chen, and Y. Li. Machine-Learning-Assisted De Novo Design of Organic Molecules and Polymers: Opportunities and Challenges. *Polymers*, 12(1):163, Jan. 2020. doi:10.3390/polym12010163.
- [9] T. K. Patra. Data-Driven Methods for Accelerating Polymer Design. *ACS Polymers Au*, 2(1):8–26, Feb. 2022. doi:10.1021/acspolymersau.1c00035.

- [10] K. R. Gadelrab, A. F. Hannon, C. A. Ross, and A. Alexander-Katz. Inverting the design path for self-assembled block copolymers. *Molecular Systems Design & Engineering*, 2(5):539–548, 2017. doi:10.1039/C7ME00062F.
- [11] M. T. H. Van Vliet, E. R. Jones, M. Flörke, W. H. P. Franssen, N. Hanasaki, Y. Wada, and J. R. Yearsley. Global water scarcity including surface water quality and expansions of clean water technologies. *Environmental Research Letters*, 16(2):024020, Feb. 2021. doi:10.1088/1748-9326/abbfc3.
- [12] C. He, Z. Liu, J. Wu, X. Pan, Z. Fang, J. Li, and B. A. Bryan. Future global urban water scarcity and potential solutions. *Nature Communications*, 12(1):4667, Aug. 2021. doi:10.1038/s41467-021-25026-3.
- [13] P. Greve, T. Kahil, J. Mochizuki, T. Schinko, Y. Satoh, P. Burek, G. Fischer, S. Tramberend, R. Burtscher, S. Langan, and Y. Wada. Global assessment of water challenges under uncertainty in water scarcity projections. *Nature Sustainability*, 1(9):486–494, Sept. 2018. doi:10.1038/s41893-018-0134-9.
- [14] S. N. Gosling and N. W. Arnell. A global assessment of the impact of climate change on water scarcity. *Climatic Change*, 134(3):371–385, Feb. 2016. doi:10.1007/s10584-013-0853-x.
- [15] N. Fitton, P. Alexander, N. Arnell, B. Bajzelj, K. Calvin, J. Doelman, J. Gerber, P. Havlik, T. Hasegawa, M. Herrero, T. Krisztin, H. Van Meijl, T. Powell, R. Sands, E. Stehfest, P. West, and P. Smith. The vulnerabilities of agricultural land and food production to future water scarcity. *Global Environmental Change*, 58:101944, Sept. 2019. doi:10.1016/j.gloenvcha.2019.101944.
- [16] M. Trnka, S. Feng, M. A. Semenov, J. E. Olesen, K. C. Kersebaum, R. P. Rötter, D. Semerádová, K. Klem, W. Huang, M. Ruiz-Ramos, P. Hlavinka, J. Meitner, J. Balek, P. Havlík, and U. Büntgen. Mitigation efforts will not fully alleviate the increase in water scarcity occurrence probability in wheat-producing areas. *Science Advances*, 5(9):eaau2406, Sept. 2019. doi:10.1126/sciadv.aau2406.
- [17] A. M. Hamiche, A. B. Stambouli, and S. Flazi. A review of the water-energy nexus. *Renewable and Sustainable Energy Reviews*, 65:319–331, Nov. 2016. doi:10.1016/j.rser.2016.07.020.
- [18] D. Bauer. The Water-Energy Nexus: Challenges and Opportunities. Technical report, U.S. Department of Energy, June 2014.
- [19] Transforming Our World: The 2030 Agenda for Sustainable Development. General Assembly Resolution 70/1, United Nations, 2015.
- [20] P. Chowdhary, R. N. Bharagava, S. Mishra, and N. Khan. Role of Industries in Water Scarcity and Its Adverse Effects on Environment and Human Health. In

- V. Shukla and N. Kumar, editors, *Environmental Concerns and Sustainable Development*, pages 235–256. Springer Singapore, Singapore, 2020. doi:[10.1007/978-981-13-5889-0_2](https://doi.org/10.1007/978-981-13-5889-0_2).
- [21] A. Fakhru'l-Razi, A. Pendashteh, L. C. Abdullah, D. R. A. Biak, S. S. Madaeni, and Z. Z. Abidin. Review of technologies for oil and gas produced water treatment. *Journal of Hazardous Materials*, 170(2-3):530–551, Oct. 2009. doi:[10.1016/j.jhazmat.2009.05.044](https://doi.org/10.1016/j.jhazmat.2009.05.044).
- [22] M. A. Al-Ghouti, M. A. Al-Kaabi, M. Y. Ashfaq, and D. A. Da'na. Produced water characteristics, treatment and reuse: A review. *Journal of Water Process Engineering*, 28:222–239, Apr. 2019. doi:[10.1016/j.jwpe.2019.02.001](https://doi.org/10.1016/j.jwpe.2019.02.001).
- [23] S. Ghafoori, M. Omar, N. Koutahzadeh, S. Zendehboudi, R. N. Malhas, M. Mohamed, S. Al-Zubaidi, K. Redha, F. Baraki, and M. Mehrvar. New advancements, challenges, and future needs on treatment of oilfield produced water: A state-of-the-art review. *Separation and Purification Technology*, 289:120652, May 2022. doi:[10.1016/j.seppur.2022.120652](https://doi.org/10.1016/j.seppur.2022.120652).
- [24] N. Ungureanu, V. Vlăduț, and G. Voicu. Water Scarcity and Wastewater Reuse in Crop Irrigation. *Sustainability*, 12(21):9055, Oct. 2020. doi:[10.3390/su12219055](https://doi.org/10.3390/su12219055).
- [25] A. G. Capodaglio. Fit-for-purpose urban wastewater reuse: Analysis of issues and available technologies for sustainable multiple barrier approaches. *Critical Reviews in Environmental Science and Technology*, 51(15):1619–1666, Aug. 2021. doi:[10.1080/10643389.2020.1763231](https://doi.org/10.1080/10643389.2020.1763231).
- [26] M. R. Landsman, R. Sujanani, S. H. Brodfuehrer, C. M. Cooper, A. G. Darr, R. J. Davis, K. Kim, S. Kum, L. K. Nalley, S. M. Nomaan, C. P. Oden, A. Paspureddi, K. K. Reimund, L. S. Rowles, S. Yeo, D. F. Lawler, B. D. Freeman, and L. E. Katz. Water Treatment: Are Membranes the Panacea? *Annual Review of Chemical and Biomolecular Engineering*, 11(1):559–585, June 2020. doi:[10.1146/annurev-chembioeng-111919-091940](https://doi.org/10.1146/annurev-chembioeng-111919-091940).
- [27] J. Schwinge, P. Neal, D. Wiley, D. Fletcher, and A. Fane. Spiral wound modules and spacers. *Journal of Membrane Science*, 242(1-2):129–153, Oct. 2004. doi:[10.1016/j.memsci.2003.09.031](https://doi.org/10.1016/j.memsci.2003.09.031).
- [28] D. J. Miller, D. R. Dreyer, C. W. Bielawski, D. R. Paul, and B. D. Freeman. Surface Modification of Water Purification Membranes. *Angewandte Chemie International Edition*, 56(17):4662–4711, Apr. 2017. doi:[10.1002/anie.201601509](https://doi.org/10.1002/anie.201601509).
- [29] W. J. Koros and R. P. Lively. Water and beyond: Expanding the spectrum of large-scale energy efficient separation processes. *AIChE Journal*, 58(9):2624–2633, Sept. 2012. doi:[10.1002/aic.13888](https://doi.org/10.1002/aic.13888).

- [30] J. R. Werber, A. Deshmukh, and M. Elimelech. The Critical Need for Increased Selectivity, Not Increased Water Permeability, for Desalination Membranes. *Environmental Science & Technology Letters*, 3(4):112–120, Apr. 2016. doi:10.1021/acs.estlett.6b00050.
- [31] R. Sujanani, M. R. Landsman, S. Jiao, J. D. Moon, M. S. Shell, D. F. Lawler, L. E. Katz, and B. D. Freeman. Designing Solute-Tailored Selectivity in Membranes: Perspectives for Water Reuse and Resource Recovery. *ACS Macro Letters*, 9(11):1709–1717, Nov. 2020. doi:10.1021/acsmacrolett.0c00710.
- [32] I.-S. Lee, W.-J. Sim, C.-W. Kim, Y.-S. Chang, and J.-E. Oh. Characteristic occurrence patterns of micropollutants and their removal efficiencies in industrial wastewater treatment plants. *Journal of Environmental Monitoring*, 13(2):391–397, 2011. doi:10.1039/C0EM00130A.
- [33] S. Mohammadi, A. Kargari, H. Sanaeepur, K. Abbassian, A. Najafi, and E. Mofarrah. Phenol removal from industrial wastewaters: a short review. *Desalination and Water Treatment*, 53(8):2215–2234, Feb. 2015. doi:10.1080/19443994.2014.883327.
- [34] D. K. Stepien, P. Diehl, J. Helm, A. Thoms, and W. Püttmann. Fate of 1,4-dioxane in the aquatic environment: From sewage to drinking water. *Water Research*, 48:406–419, Jan. 2014. doi:10.1016/j.watres.2013.09.057.
- [35] M. H. Isa, E. H. Ezechi, Z. Ahmed, S. F. Magram, and S. R. M. Kutty. Boron removal by electrocoagulation and recovery. *Water Research*, 51:113–123, Mar. 2014. doi:10.1016/j.watres.2013.12.024.
- [36] M. J. López-Muñoz, J. M. Arsuaga, and A. Sotto. Separation of phenols and their advanced oxidation intermediate products in aqueous solution by NF/RO membranes. *Separation and Purification Technology*, 71(2):246–251, Feb. 2010. doi:10.1016/j.seppur.2009.12.004.
- [37] W. Raza, J. Lee, N. Raza, Y. Luo, K.-H. Kim, and J. Yang. Removal of phenolic compounds from industrial waste water based on membrane-based technologies. *Journal of Industrial and Engineering Chemistry*, 71:1–18, Mar. 2019. doi:10.1016/j.jiec.2018.11.024.
- [38] S. Sonawane, K. Fedorov, M. P. Rayaroth, and G. Boczkaj. Degradation of 1,4-dioxane by sono-activated persulfates for water and wastewater treatment applications. *Water Resources and Industry*, 28:100183, Dec. 2022. doi:10.1016/j.wri.2022.100183.
- [39] W. Li, S. Patton, J. M. Gleason, S. P. Mezyk, K. P. Ishida, and H. Liu. UV Photolysis of Chloramine and Persulfate for 1,4-Dioxane Removal in Reverse-Osmosis Permeate for Potable Water Reuse. *Environmental Science & Technology*, 52(11):6417–6425, June 2018. doi:10.1021/acs.est.7b06042.

- [40] R. Epsztein, R. M. DuChanois, C. L. Ritt, A. Noy, and M. Elimelech. Towards single-species selectivity of membranes with subnanometre pores. *Nature Nanotechnology*, 15(6):426–436, June 2020. doi:10.1038/s41565-020-0713-6.
- [41] A. Marshall, P. Munro, and G. Trägårdh. The effect of protein fouling in micro-filtration and ultrafiltration on permeate flux, protein retention and selectivity: A literature review. *Desalination*, 91(1):65–108, Mar. 1993. doi:10.1016/0011-9164(93)80047-Q.
- [42] R. Zhang, Y. Liu, M. He, Y. Su, X. Zhao, M. Elimelech, and Z. Jiang. Antifouling membranes for sustainable water purification: strategies and mechanisms. *Chemical Society Reviews*, 45(21):5888–5924, 2016. doi:10.1039/C5CS00579E.
- [43] X. Du, J. Meng, R. Xu, Q. Shi, and Y. Zhang. Polyol-grafted polysulfone membranes for boron removal: Effects of the ligand structure. *Journal of Membrane Science*, 476:205–215, Feb. 2015. doi:10.1016/j.memsci.2014.11.042.
- [44] E. Ostuni, R. G. Chapman, R. E. Holmlin, S. Takayama, and G. M. Whitesides. A Survey of StructureProperty Relationships of Surfaces that Resist the Adsorption of Protein. *Langmuir*, 17(18):5605–5620, Sept. 2001. doi:10.1021/la010384m.
- [45] E. M. Vrijenhoek, S. Hong, and M. Elimelech. Influence of membrane surface properties on initial rate of colloidal fouling of reverse osmosis and nanofiltration membranes. *Journal of Membrane Science*, 188(1):115–128, June 2001. doi:10.1016/S0376-7388(01)00376-3.
- [46] J. R. Werber, C. O. Osuji, and M. Elimelech. Materials for next-generation desalination and water purification membranes. *Nature Reviews Materials*, 1(5):16018, May 2016. doi:10.1038/natrevmats.2016.18.
- [47] E. A. Eugene, W. A. Phillip, and A. W. Dowling. Data science-enabled molecular-to-systems engineering for sustainable water treatment. *Current Opinion in Chemical Engineering*, 26:122–130, Dec. 2019. doi:10.1016/j.coche.2019.10.002.
- [48] E. Brini, C. J. Fennell, M. Fernandez-Serra, B. Hribar-Lee, M. Lukšič, and K. A. Dill. How Water’s Properties Are Encoded in Its Molecular Structure and Energies. *Chemical Reviews*, 117(19):12385–12414, Oct. 2017. doi:10.1021/acs.chemrev.7b00259.
- [49] E. Xi and A. J. Patel. The hydrophobic effect, and fluctuations: The long and the short of it. *Proceedings of the National Academy of Sciences*, 113(17):4549–4551, Apr. 2016. doi:10.1073/pnas.1603014113.
- [50] A. Chaimovich and M. S. Shell. Length-scale crossover of the hydrophobic interaction in a coarse-grained water model. *Physical Review E*, 88(5):052313, Nov. 2013. doi:10.1103/PhysRevE.88.052313.

- [51] T. Young, R. Abel, B. Kim, B. J. Berne, and R. A. Friesner. Motifs for molecular recognition exploiting hydrophobic enclosure in protein–ligand binding. *Proceedings of the National Academy of Sciences*, 104(3):808–813, Jan. 2007. doi:10.1073/pnas.0610202104.
- [52] N. B. Rego, E. Xi, and A. J. Patel. Identifying hydrophobic protein patches to inform protein interaction interfaces. *Proceedings of the National Academy of Sciences*, 118(6):e2018234118, Feb. 2021. doi:10.1073/pnas.2018234118.
- [53] N. Choudhury and B. M. Pettitt. On the Mechanism of Hydrophobic Association of Nanoscopic Solutes. *Journal of the American Chemical Society*, 127(10):3556–3567, Mar. 2005. doi:10.1021/ja0441817.
- [54] X. Huang, R. Zhou, and B. J. Berne. Drying and Hydrophobic Collapse of Paraffin Plates. *The Journal of Physical Chemistry B*, 109(8):3546–3552, Mar. 2005. doi:10.1021/jp045520l.
- [55] L. Wang, R. A. Friesner, and B. J. Berne. Hydrophobic interactions in model enclosures from small to large length scales: non-additivity in explicit and implicit solvent models. *Faraday Discussions*, 146:247–262, 2010. doi:10.1039/b925521b.
- [56] K. Lum, D. Chandler, and J. D. Weeks. Hydrophobicity at Small and Large Length Scales. *The Journal of Physical Chemistry B*, 103(22):4570–4577, June 1999. doi:10.1021/jp984327m.
- [57] S. Vaikuntanathan and P. L. Geissler. Putting Water on a Lattice: The Importance of Long Wavelength Density Fluctuations in Theories of Hydrophobic and Interfacial Phenomena. *Physical Review Letters*, 112(2):020603, Jan. 2014. doi:10.1103/PhysRevLett.112.020603.
- [58] R. C. Remsing and A. J. Patel. Water density fluctuations relevant to hydrophobic hydration are unaltered by attractions. *The Journal of Chemical Physics*, 142(2):024502, Jan. 2015. doi:10.1063/1.4905009.
- [59] A. J. Patel, P. Varilly, D. Chandler, and S. Garde. Quantifying Density Fluctuations in Volumes of All Shapes and Sizes Using Indirect Umbrella Sampling. *Journal of Statistical Physics*, 145(2):265–275, Oct. 2011. doi:10.1007/s10955-011-0269-9.
- [60] E. Xi, R. C. Remsing, and A. J. Patel. Sparse Sampling of Water Density Fluctuations in Interfacial Environments. *Journal of Chemical Theory and Computation*, 12(2):706–713, Feb. 2016. doi:10.1021/acs.jctc.5b01037.
- [61] J. I. Monroe and M. S. Shell. Computational discovery of chemically patterned surfaces that effect unique hydration water dynamics. *Proceedings of the National Academy of Sciences*, 115(32):8093–8098, Aug. 2018. doi:10.1073/pnas.1807208115.

- [62] J. I. Monroe and M. S. Shell. Decoding signatures of structure, bulk thermodynamics, and solvation in three-body angle distributions of rigid water models. *The Journal of Chemical Physics*, 151(9):094501, Sept. 2019. doi:10.1063/1.5111545.
- [63] D. C. Robinson Brown, T. R. Webber, S. Jiao, D. M. Rivera Mirabal, S. Han, and M. S. Shell. Relationships between Molecular Structural Order Parameters and Equilibrium Water Dynamics in Aqueous Mixtures. *The Journal of Physical Chemistry B*, 127(20):4577–4594, May 2023. doi:10.1021/acs.jpcb.3c00826.
- [64] B. C. Dallin, A. S. Kelkar, and R. C. Van Lehn. Structural features of interfacial water predict the hydrophobicity of chemically heterogeneous surfaces. *Chemical Science*, 14(5):1308–1319, 2023. doi:10.1039/D2SC02856E.
- [65] A. S. Kelkar, B. C. Dallin, and R. C. Van Lehn. Predicting Hydrophobicity by Learning Spatiotemporal Features of Interfacial Water Structure: Combining Molecular Dynamics Simulations with Convolutional Neural Networks. *The Journal of Physical Chemistry B*, 124(41):9103–9114, Oct. 2020. doi:10.1021/acs.jpcb.0c05977.
- [66] M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids*. Oxford University PressOxford, 2 edition, June 2017. doi:10.1093/oso/9780198803195.001.0001.
- [67] R. Qiao and N. R. Aluru. Ion concentrations and velocity profiles in nanochannel electroosmotic flows. *The Journal of Chemical Physics*, 118(10):4692–4701, Mar. 2003. doi:10.1063/1.1543140.
- [68] S. Faucher, N. Aluru, M. Z. Bazant, D. Blankschtein, A. H. Brozena, J. Cumings, J. Pedro De Souza, M. Elimelech, R. Epsztein, J. T. Fourkas, A. G. Rajan, H. J. Kulik, A. Levy, A. Majumdar, C. Martin, M. McEldrew, R. P. Misra, A. Noy, T. A. Pham, M. Reed, E. Schwegler, Z. Siwy, Y. Wang, and M. Strano. Critical Knowledge Gaps in Mass Transport through Single-Digit Nanopores: A Review and Perspective. *The Journal of Physical Chemistry C*, 123(35):21309–21326, Sept. 2019. doi:10.1021/acs.jpcc.9b02178.
- [69] A. P. Willard and D. Chandler. Instantaneous Liquid Interfaces. *The Journal of Physical Chemistry B*, 114(5):1954–1958, Feb. 2010. doi:10.1021/jp909219k.
- [70] R. Godawat, S. N. Jamadagni, and S. Garde. Characterizing hydrophobicity of interfaces by using cavity formation, solute binding, and water correlations. *Proceedings of the National Academy of Sciences*, 106(36):15119–15124, Sept. 2009. doi:10.1073/pnas.0902778106.
- [71] P. Eastman, J. Swails, J. D. Chodera, R. T. McGibbon, Y. Zhao, K. A. Beauchamp, L.-P. Wang, A. C. Simmonett, M. P. Harrigan, C. D. Stern, R. P. Wiewiora, B. R. Brooks, and V. S. Pande. OpenMM 7: Rapid development of high performance

- algorithms for molecular dynamics. *PLOS Computational Biology*, 13(7):e1005659, July 2017. doi:10.1371/journal.pcbi.1005659.
- [72] D. Yokelson, N. V. Tkachenko, R. Robey, Y. W. Li, and P. A. Dub. Performance Analysis of CP2K Code for Ab Initio Molecular Dynamics on CPUs and GPUs. *Journal of Chemical Information and Modeling*, 62(10):2378–2386, May 2022. doi:10.1021/acs.jcim.1c01538.
- [73] A. S. Knight, E. Y. Zhou, M. B. Francis, and R. N. Zuckermann. Sequence Programmable Peptoid Polymers for Diverse Materials Applications. *Advanced Materials*, 27(38):5665–5691, Oct. 2015. doi:10.1002/adma.201500275.
- [74] V. Saxena, M. G. L. Merrilees, and K. H. A. Lau. Antifouling Peptoid Biointerfaces. In P. Chandra and L. M. Pandey, editors, *Biointerface Engineering: Prospects in Medical Diagnostics and Drug Delivery*, pages 55–73. Springer Singapore, Singapore, 2020. doi:10.1007/978-981-15-4790-4_3.
- [75] R. J. Simon, R. S. Kania, R. N. Zuckermann, V. D. Huebner, D. A. Jewell, S. Banville, S. Ng, L. Wang, S. Rosenberg, and C. K. Marlowe. Peptoids: a modular approach to drug discovery. *Proceedings of the National Academy of Sciences*, 89(20):9367–9371, Oct. 1992. doi:10.1073/pnas.89.20.9367.
- [76] N. D. Austin, N. V. Sahinidis, and D. W. Trahan. Computer-aided molecular design: An introduction and review of tools, applications, and solution techniques. *Chemical Engineering Research and Design*, 116:2–26, Dec. 2016. doi:10.1016/j.cherd.2016.10.014.
- [77] B. Mohr, K. Shmilovich, I. S. Kleinwächter, D. Schneider, A. L. Ferguson, and T. Bereau. Data-driven discovery of cardiolipin-selective small molecules by computational active learning. *Chemical Science*, 13(16):4498–4511, 2022. doi:10.1039/D2SC00116K.
- [78] J. P. Janet, S. Ramesh, C. Duan, and H. J. Kulik. Accurate Multiobjective Design in a Space of Millions of Transition Metal Complexes with Neural-Network-Driven Efficient Global Optimization. *ACS Central Science*, 6(4):513–524, Apr. 2020. doi:10.1021/acscentsci.0c00026.
- [79] J. I. Monroe, S. Jiao, R. J. Davis, D. Robinson Brown, L. E. Katz, and M. S. Shell. Affinity of small-molecule solutes to hydrophobic, hydrophilic, and chemically patterned interfaces in aqueous solution. *Proceedings of the National Academy of Sciences*, 118(1):e2020205118, Jan. 2021. doi:10.1073/pnas.2020205118.
- [80] Y. Ma, J. C. Aulicino, and A. L. Ferguson. Inverse Design of Self-Assembling Diamond Photonic Lattices from Anisotropic Colloidal Clusters. *The Journal of Physical Chemistry B*, 125(9):2398–2410, Mar. 2021. doi:10.1021/acs.jpcc.0c08723.

- [81] V. Meenakshisundaram, J.-H. Hung, T. K. Patra, and D. S. Simmons. Designing Sequence-Specific Copolymer Compatibilizers Using a Molecular-Dynamics-Simulation-Based Genetic Algorithm. *Macromolecules*, 50(3):1155–1166, Feb. 2017. doi:10.1021/acs.macromol.6b01747.
- [82] Z. M. Sherman, M. P. Howard, B. A. Lindquist, R. B. Jadrich, and T. M. Truskett. Inverse methods for design of soft materials. *The Journal of Chemical Physics*, 152(14):140902, Apr. 2020. doi:10.1063/1.5145177.
- [83] F. R. Rijsberman. Water scarcity: Fact or fiction? *Agricultural Water Management*, 80(1-3):5–22, Feb. 2006. doi:10.1016/j.agwat.2005.07.001.
- [84] E. Jones, M. Qadir, M. T. van Vliet, V. Smakhtin, and S.-m. Kang. The state of desalination and brine production: A global outlook. *Science of The Total Environment*, 657:1343–1356, Mar. 2019. doi:10.1016/j.scitotenv.2018.12.076.
- [85] B. Peñate and L. García-Rodríguez. Current trends and future prospects in the design of seawater reverse osmosis desalination technology. *Desalination*, 284:1–8, Jan. 2012. doi:10.1016/j.desal.2011.09.010.
- [86] R. C. Scholes, A. N. Stiegler, C. M. Anderson, and D. L. Sedlak. Enabling Water Reuse by Treatment of Reverse Osmosis Concentrate: The Promise of Constructed Wetlands. *ACS Environmental Au*, 1(1):7–17, Nov. 2021. doi:10.1021/acsenvironau.1c00013.
- [87] C. E. Clark and J. A. Veil. Produced water volumes and management practices in the United States. Technical Report ANL/EVS/R-09-1, 1007397, Sept. 2009. doi:10.2172/1007397.
- [88] C. M. Cooper, J. McCall, S. C. Stokes, C. McKay, M. J. Bentley, J. S. Rosenblum, T. A. Blewett, Z. Huang, A. Miara, M. Talmadge, A. Evans, K. A. Sitterley, P. Kurup, J. R. Stokes-Draut, J. Macknick, T. Borch, T. Y. Cath, and L. E. Katz. Oil and Gas Produced Water Reuse: Opportunities, Treatment Needs, and Challenges. *ACS ES&T Engineering*, 2(3):347–366, Mar. 2022. doi:10.1021/acsestengg.1c00248.
- [89] R. P. Sabie, L. Pillsbury, and P. Xu. Spatiotemporal Analysis of Produced Water Demand for Fit-For-Purpose Reuse—A Permian Basin, New Mexico Case Study. *Water*, 14(11):1735, May 2022. doi:10.3390/w14111735.
- [90] J. S. Guest, S. J. Skerlos, J. L. Barnard, M. B. Beck, G. T. Daigger, H. Hilger, S. J. Jackson, K. Karvazy, L. Kelly, L. Macpherson, J. R. Mihelcic, A. Pramanik, L. Raskin, M. C. M. Van Loosdrecht, D. Yeh, and N. G. Love. A New Planning and Design Paradigm to Achieve Sustainable Resource Recovery from Wastewater. *Environmental Science & Technology*, 43(16):6126–6130, Aug. 2009. doi:10.1021/es9010515.

- [91] World Health Organization. *A global overview of national regulations and standards for drinking-water quality*. World Health Organization, Geneva, 2nd ed edition, 2021. URL <https://apps.who.int/iris/handle/10665/350981>. Section: iv, 101 p.
- [92] N. Hilal, G. Kim, and C. Somerfield. Boron removal from saline water: A comprehensive review. *Desalination*, 273(1):23–35, June 2011. doi:[10.1016/j.desal.2010.05.012](https://doi.org/10.1016/j.desal.2010.05.012).
- [93] M. Zaman, S. A. Shahid, and L. Heng. *Guideline for Salinity Assessment, Mitigation and Adaptation Using Nuclear and Related Techniques*. Springer International Publishing, Cham, 2018. doi:[10.1007/978-3-319-96190-3](https://doi.org/10.1007/978-3-319-96190-3).
- [94] P. Agre. Aquaporin Water Channels (Nobel Lecture). *Angewandte Chemie International Edition*, 43(33):4278–4290, Aug. 2004. doi:[10.1002/anie.200460804](https://doi.org/10.1002/anie.200460804).
- [95] R. Latorre and C. Miller. Conduction and selectivity in potassium channels. *The Journal of Membrane Biology*, 71(1-2):11–30, Feb. 1983. doi:[10.1007/BF01870671](https://doi.org/10.1007/BF01870671).
- [96] G. Hummer, J. C. Rasaiah, and J. P. Noworyta. Water conduction through the hydrophobic channel of a carbon nanotube. *Nature*, 414(6860):188–190, Nov. 2001. doi:[10.1038/35102535](https://doi.org/10.1038/35102535).
- [97] A. Kalra, S. Garde, and G. Hummer. Osmotic water transport through carbon nanotube membranes. *Proceedings of the National Academy of Sciences*, 100(18):10175–10180, Sept. 2003. doi:[10.1073/pnas.1633354100](https://doi.org/10.1073/pnas.1633354100).
- [98] A. W. Thornton, A. Ahmed, M. Mainak, H. B. Park, and A. J. Hill. Ultrafast Transport in Nanotubes and Nanosheets. In Y. Chen, editor, *Nanotubes and Nanosheets*, pages 271–304. CRC Press, 1st edition, Feb. 2015. doi:[10.1201/b18073-15](https://doi.org/10.1201/b18073-15).
- [99] R. H. Tunuguntla, R. Y. Henley, Y.-C. Yao, T. A. Pham, M. Wanunu, and A. Noy. Enhanced water permeability and tunable ion selectivity in sub-nanometer carbon nanotube porins. *Science*, 357(6353):792–796, Aug. 2017. doi:[10.1126/science.aan2438](https://doi.org/10.1126/science.aan2438).
- [100] M. H. Köhler and J. R. Bordin. Surface, Density, and Temperature Effects on the Water Diffusion and Structure Inside Narrow Nanotubes. *The Journal of Physical Chemistry C*, 122(12):6684–6690, Mar. 2018. doi:[10.1021/acs.jpcc.8b00112](https://doi.org/10.1021/acs.jpcc.8b00112).
- [101] A. Striolo, A. A. Chialvo, P. T. Cummings, and K. E. Gubbins. Simulated water adsorption in chemically heterogeneous carbon nanotubes. *The Journal of Chemical Physics*, 124(7):074710, Feb. 2006. doi:[10.1063/1.2171349](https://doi.org/10.1063/1.2171349).
- [102] M. Majumder, N. Chopra, and B. J. Hinds. Mass Transport through Carbon Nanotube Membranes in Three Different Regimes: Ionic Diffusion and Gas and Liquid Flow. *ACS Nano*, 5(5):3867–3877, May 2011. doi:[10.1021/nn200222g](https://doi.org/10.1021/nn200222g).

- [103] F. Fornasiero, J. B. In, S. Kim, H. G. Park, Y. Wang, C. P. Grigoropoulos, A. Noy, and O. Bakajin. pH-Tunable Ion Selectivity in Carbon Nanotube Pores. *Langmuir*, 26(18):14848–14853, Sept. 2010. doi:10.1021/la101943h.
- [104] B. Corry. Water and ion transport through functionalised carbon nanotubes: implications for desalination technology. *Energy & Environmental Science*, 4(3):751, 2011. doi:10.1039/c0ee00481b.
- [105] I. Sadeghi and A. Asatekin. Membranes with Functionalized Nanopores for Aromaticity-Based Separation of Small Molecules. *ACS Applied Materials & Interfaces*, 11(13):12854–12862, Apr. 2019. doi:10.1021/acsami.9b00090.
- [106] M. Thomas and B. Corry. Modifying Water Flow, Ion Selectivity, and Salt Rejection in Carbon Nanotubes via Surface Adsorption. *The Journal of Physical Chemistry C*, 124(6):3820–3826, Feb. 2020. doi:10.1021/acs.jpcc.9b08485.
- [107] Q. Lyu, D.-Y. Kang, S. Hu, and L.-C. Lin. Exploiting interior surface functionalization in reverse osmosis desalination membranes to mitigate permeability–selectivity trade-off: Molecular simulations of nanotube-based membranes. *Desalination*, 491:114537, Oct. 2020. doi:10.1016/j.desal.2020.114537.
- [108] X. Gong, J. Li, K. Xu, J. Wang, and H. Yang. A Controllable Molecular Sieve for Na^+ and K^+ Ions. *Journal of the American Chemical Society*, 132(6):1873–1877, Feb. 2010. doi:10.1021/ja905753p.
- [109] X. Zhao, R. Zhang, Y. Liu, M. He, Y. Su, C. Gao, and Z. Jiang. Antifouling membrane surface construction: Chemistry plays a critical role. *Journal of Membrane Science*, 551:145–171, Apr. 2018. doi:10.1016/j.memsci.2018.01.039.
- [110] H. B. Park, J. Kamcev, L. M. Robeson, M. Elimelech, and B. D. Freeman. Maximizing the right stuff: The trade-off between membrane permeability and selectivity. *Science*, 356(6343):eaab0530, June 2017. doi:10.1126/science.aab0530.
- [111] I. K. Meier, M. Langsam, and H. C. Klotz. Selectivity enhancement via photooxidative surface modification of polyimide air separation membranes. *Journal of Membrane Science*, 94(1):195–212, Sept. 1994. doi:10.1016/0376-7388(93)E0174-I.
- [112] L. E. Peng, Z. Yao, X. Liu, B. Deng, H. Guo, and C. Y. Tang. Tailoring Polyamide Rejection Layer with Aqueous Carbonate Chemistry for Enhanced Membrane Separation: Mechanistic Insights, Chemistry-Structure-Property Relationship, and Environmental Implications. *Environmental Science & Technology*, 53(16):9764–9770, Aug. 2019. doi:10.1021/acs.est.9b03210.
- [113] H. Guo, Y. Deng, Z. Tao, Z. Yao, J. Wang, C. Lin, T. Zhang, B. Zhu, and C. Y. Tang. Does Hydrophilic Polydopamine Coating Enhance Membrane Rejection of Hydrophobic Endocrine-Disrupting Compounds? *Environmental Science & Technology Letters*, 3(9):332–338, Sept. 2016. doi:10.1021/acs.estlett.6b00263.

- [114] S. Le Caër, S. Pin, S. Esnouf, Q. Raffy, J. P. Renault, J.-B. Brubach, G. Creff, and P. Roy. A trapped water network in nanoporous material: the role of interfaces. *Physical Chemistry Chemical Physics*, 13(39):17658, 2011. doi:10.1039/c1cp21980d.
- [115] A. W. Knight, N. G. Kalugin, E. Coker, and A. G. Ilgen. Water properties under nano-scale confinement. *Scientific Reports*, 9:8246, Dec. 2019. doi:10.1038/s41598-019-44651-z.
- [116] J. C. Rasaiah, S. Garde, and G. Hummer. Water in Nonpolar Confinement: From Nanotubes to Proteins and Beyond. *Annual Review of Physical Chemistry*, 59(1):713–740, May 2008. doi:10.1146/annurev.physchem.59.032607.093815.
- [117] S. Chakraborty, H. Kumar, C. Dasgupta, and P. K. Maiti. Confined Water: Structure, Dynamics, and Thermodynamics. *Accounts of Chemical Research*, 50(9):2139–2146, Sept. 2017. doi:10.1021/acs.accounts.6b00617.
- [118] S. Zhang, T. Hedtke, X. Zhou, M. Elimelech, and J.-H. Kim. Environmental Applications of Engineered Materials with Nanoconfinement. *ACS ES&T Engineering*, 1(4):706–724, Apr. 2021. doi:10.1021/acsestengg.1c00007.
- [119] J. C. Love, L. A. Estroff, J. K. Kriebel, R. G. Nuzzo, and G. M. Whitesides. Self-Assembled Monolayers of Thiolates on Metals as a Form of Nanotechnology. *Chemical Reviews*, 105(4):1103–1170, Apr. 2005. doi:10.1021/cr0300789.
- [120] S. Wang, L. Lu, X. Lu, W. Cao, and Y. Zhu. Adsorption of binary CO₂/CH₄ mixtures using carbon nanotubes: Effects of confinement and surface functionalization. *Separation Science and Technology*, 51(7):1079–1092, May 2016. doi:10.1080/01496395.2016.1150296.
- [121] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2):926–935, July 1983. doi:10.1063/1.445869.
- [122] F. Risplendi, F. Raffone, L.-C. Lin, J. C. Grossman, and G. Cicero. Fundamental Insights on Hydration Environment of Boric Acid and Its Role in Separation from Saline Water. *The Journal of Physical Chemistry C*, 124(2):1438–1445, Jan. 2020. doi:10.1021/acs.jpcc.9b10065.
- [123] H. C. Andersen. Rattle: A “velocity” version of the shake algorithm for molecular dynamics calculations. *Journal of Computational Physics*, 52(1):24–34, Oct. 1983. doi:10.1016/0021-9991(83)90014-1.
- [124] S. Miyamoto and P. A. Kollman. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of Computational Chemistry*, 13(8):952–962, Oct. 1992. doi:10.1002/jcc.540130805.

- [125] M. R. Shirts and J. D. Chodera. Statistically optimal analysis of samples from multiple equilibrium states. *The Journal of Chemical Physics*, 129(12):124105, Sept. 2008. doi:10.1063/1.2978177.
- [126] J. D. Chodera and M. R. Shirts. Replica exchange and expanded ensemble simulations as Gibbs sampling: Simple improvements for enhanced mixing. *The Journal of Chemical Physics*, 135(19):194110, Nov. 2011. doi:10.1063/1.3660669.
- [127] I. Moskowitz, M. A. Snyder, and J. Mittal. Water transport through functionalized nanotubes with tunable hydrophobicity. *The Journal of Chemical Physics*, 141(18):18C532, Nov. 2014. doi:10.1063/1.4897974.
- [128] E. Xi, V. Venkateshwaran, L. Li, N. Rego, A. J. Patel, and S. Garde. Hydrophobicity of proteins and nanostructured solutes is governed by topographical and chemical context. *Proceedings of the National Academy of Sciences*, 114(51):13345–13350, Dec. 2017. doi:10.1073/pnas.1700092114.
- [129] H. Acharya, S. Vembanur, S. N. Jamadagni, and S. Garde. Mapping hydrophobicity at the nanoscale: Applications to heterogeneous surfaces and proteins. *Faraday Discussions*, 146:353–365, 2010. doi:10.1039/b927019a.
- [130] N. Giovambattista, P. G. Debenedetti, and P. J. Rossky. Hydration Behavior under Confinement by Nanoscale Surfaces with Patterned Hydrophobicity and Hydrophilicity. *The Journal of Physical Chemistry C*, 111(3):1323–1332, Jan. 2007. doi:10.1021/jp065419b.
- [131] L. Hua, R. Zangi, and B. J. Berne. Hydrophobic Interactions and Dewetting between Plates with Hydrophobic and Hydrophilic Domains. *The Journal of Physical Chemistry C*, 113(13):5244–5253, Apr. 2009. doi:10.1021/jp8088758.
- [132] A. P. Willard and D. Chandler. Coarse-grained modeling of the interface between water and heterogeneous surfaces. *Faraday Discuss.*, 141:209–220, 2009. doi:10.1039/B805786A.
- [133] P. Reimann, C. Van den Broeck, H. Linke, P. Hänggi, J. M. Rubi, and A. Pérez-Madrid. Giant Acceleration of Free Diffusion by Use of Tilted Periodic Potentials. *Physical Review Letters*, 87(1):010602, June 2001. doi:10.1103/PhysRevLett.87.010602.
- [134] M. F. Harrach, F. Klameth, B. Drossel, and M. Vogel. Effect of the hydroaffinity and topology of pore walls on the structure and dynamics of confined water. *The Journal of Chemical Physics*, 142(3):034703, Jan. 2015. doi:10.1063/1.4905557.
- [135] D. Ding, P. Gao, Q. Ma, D. Wang, and F. Xia. Biomolecule-Functionalized Solid-State Ion Nanochannels/Nanopores: Features and Techniques. *Small*, 15(32):1804878, Aug. 2019. doi:10.1002/smll.201804878.

- [136] C. Tang, E. M. Lennon, G. H. Fredrickson, E. J. Kramer, and C. J. Hawker. Evolution of Block Copolymer Lithography to Highly Ordered Square Arrays. *Science*, 322(5900):429–432, Oct. 2008. doi:10.1126/science.1162950.
- [137] W. Barthlott and C. Neinhuis. Purity of the sacred lotus, or escape from contamination in biological surfaces. *Planta*, 202(1):1–8, Apr. 1997. doi:10.1007/s004250050096.
- [138] S. Latthe, C. Terashima, K. Nakata, and A. Fujishima. Superhydrophobic Surfaces Developed by Mimicking Hierarchical Surface Morphology of Lotus Leaf. *Molecules*, 19(4):4256–4283, Apr. 2014. doi:10.3390/molecules19044256.
- [139] J. P. Folkers, P. E. Laibinis, and G. M. Whitesides. Self-assembled monolayers of alkanethiols on gold: the adsorption and wetting properties of monolayers derived from two components with alkane chains of different lengths. *Journal of Adhesion Science and Technology*, 6(12):1397–1410, Jan. 1992. doi:10.1163/156856192X00700.
- [140] F. Müller-Plathe, S. Pal, H. Weiss, and H. Keller. Tutorial: Can Nanostructuring Improve the Properties of Hydrophobic Surfaces? *Soft Materials*, 3(1):21–43, Jan. 2005. doi:10.1080/15394450500408959.
- [141] N. Giovambattista, C. F. Lopez, P. J. Rossky, and P. G. Debenedetti. Hydrophobicity of protein surfaces: Separating geometry from chemistry. *Proceedings of the National Academy of Sciences*, 105(7):2274–2279, Feb. 2008. doi:10.1073/pnas.0708088105.
- [142] P. Liu, X. Huang, R. Zhou, and B. J. Berne. Observation of a dewetting transition in the collapse of the melittin tetramer. *Nature*, 437(7055):159–162, Sept. 2005. doi:10.1038/nature03926.
- [143] P. R. Pandey and S. Roy. Is it Possible to Change Wettability of Hydrophilic Surface by Changing Its Roughness? *The Journal of Physical Chemistry Letters*, 4(21):3692–3697, Nov. 2013. doi:10.1021/jz401946v.
- [144] B. C. Dallin, H. Yeon, A. R. Ostwalt, N. L. Abbott, and R. C. Van Lehn. Molecular Order Affects Interfacial Water Structure and Temperature-Dependent Hydrophobic Interactions between Nonpolar Self-Assembled Monolayers. *Langmuir*, 35(6):2078–2088, Feb. 2019. doi:10.1021/acs.langmuir.8b03287.
- [145] B. C. Dallin and R. C. Van Lehn. Spatially Heterogeneous Water Properties at Disordered Surfaces Decrease the Hydrophobicity of Nonpolar Self-Assembled Monolayers. *The Journal of Physical Chemistry Letters*, 10(14):3991–3997, July 2019. doi:10.1021/acs.jpcllett.9b01707.

- [146] H. Yeon, C. Wang, R. C. Van Lehn, and N. L. Abbott. Influence of Order within Nonpolar Monolayers on Hydrophobic Interactions. *Langmuir*, 33(19):4628–4637, May 2017. doi:10.1021/acs.langmuir.7b00226.
- [147] H. Moon, R. P. Collanton, J. I. Monroe, T. M. Casey, M. S. Shell, S. Han, and S. L. Scott. Evidence for Entropically Controlled Interfacial Hydration in Mesoporous Organosilicas. *Journal of the American Chemical Society*, 144(4):1766–1777, Feb. 2022. doi:10.1021/jacs.1c11342.
- [148] N. Giovambattista, P. G. Debenedetti, and P. J. Rossky. Enhanced surface hydrophobicity by coupling of surface polarity and topography. *Proceedings of the National Academy of Sciences*, 106(36):15181–15185, Sept. 2009. doi:10.1073/pnas.0905468106.
- [149] K. M. Jablonka, G. M. Jothiappan, S. Wang, B. Smit, and B. Yoo. Bias free multiobjective active learning for materials design and discovery. *Nature Communications*, 12(1):2312, Dec. 2021. doi:10.1038/s41467-021-22437-0.
- [150] C. Dai and S. C. Glotzer. Efficient Phase Diagram Sampling by Active Learning. *The Journal of Physical Chemistry B*, 124(7):1275–1284, Feb. 2020. doi:10.1021/acs.jpcb.9b09202.
- [151] C. Kim, A. Chandrasekaran, A. Jha, and R. Ramprasad. Active-learning and materials design: the example of high glass transition temperature polymers. *MRS Communications*, 9(3):860–866, Sept. 2019. doi:10.1557/mrc.2019.78.
- [152] S. Jiao, L. E. Katz, and M. S. Shell. Inverse Design of Pore Wall Chemistry To Control Solute Transport and Selectivity. *ACS Central Science*, 8(12):1609–1617, Dec. 2022. doi:10.1021/acscentsci.2c01011.
- [153] G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, and W. L. Jorgensen. Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. *The Journal of Physical Chemistry B*, 105(28):6474–6487, July 2001. doi:10.1021/jp003919d.
- [154] M. D. McKay, R. J. Beckman, and W. J. Conover. A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output From a Computer Code. *Technometrics*, 42(1):55–61, Feb. 2000. doi:10.1080/00401706.2000.10485979.
- [155] W. Humphrey, A. Dalke, and K. Schulten. VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, 14(1):33–38, Feb. 1996. doi:10.1016/0263-7855(96)00018-5.
- [156] D. Zhan, J. Qian, and Y. Cheng. Pseudo expected improvement criterion for parallel EGO algorithm. *Journal of Global Optimization*, 68(3):641–662, July 2017. doi:10.1007/s10898-016-0484-7.

- [157] R. M. Green and K. L. Bicker. Evaluation of peptoid mimics of short, lipophilic peptide antimicrobials. *International Journal of Antimicrobial Agents*, 56(2):106048, June 2020. doi:[10.1016/j.ijantimicag.2020.106048](https://doi.org/10.1016/j.ijantimicag.2020.106048).
- [158] J. S. Khara, B. Mojsoska, D. Mukherjee, P. R. Langford, B. D. Robertson, H. Jenssen, P. L. R. Ee, and S. M. Newton. Ultra-Short Antimicrobial Peptoids Show Propensity for Membrane Activity Against Multi-Drug Resistant Mycobacterium tuberculosis. *Frontiers in Microbiology*, 11:417, Mar. 2020. doi:[10.3389/fmicb.2020.00417](https://doi.org/10.3389/fmicb.2020.00417).
- [159] J. H. Kim, S. C. Kim, M. A. Kline, E. M. Grzincic, B. W. Tresca, J. Cardiel, M. Karbaschi, D. C. Dehigaspitiya, Y. Chen, V. Udumula, T. Jian, D. J. Murray, L. Yun, M. D. Connolly, J. Liu, G. Ren, C.-L. Chen, K. Kirshenbaum, A. R. Abate, and R. N. Zuckermann. Discovery of Stable and Selective Antibody Mimetics from Combinatorial Libraries of Polyvalent, Loop-Functionalized Peptoid Nanosheets. *ACS Nano*, 14(1):185–195, Jan. 2020. doi:[10.1021/acsnano.9b07498](https://doi.org/10.1021/acsnano.9b07498).
- [160] M. K. Mackwitz, E. Hespings, K. Eribez, A. Schöler, Y. Antonova-Koch, J. Held, E. A. Winzeler, K. T. Andrews, and F. K. Hansen. Investigation of the in vitro and in vivo efficacy of peptoid-based HDAC inhibitors with dual-stage antiplasmodial activity. *European Journal of Medicinal Chemistry*, 211:113065, Feb. 2021. doi:[10.1016/j.ejmech.2020.113065](https://doi.org/10.1016/j.ejmech.2020.113065).
- [161] K. G. Asfaw, Q. Liu, X. Xu, C. Manz, S. Purper, R. Eghbalian, S. W. Münch, I. Wehl, S. Bräse, E. Eiche, B. Hause, I. Bogeski, U. Schepers, M. Riemann, and P. Nick. A mitochondria-targeted coenzyme Q peptoid induces superoxide dismutase and alleviates salinity stress in plant cells. *Scientific Reports*, 10(1):11563, Dec. 2020. doi:[10.1038/s41598-020-68491-4](https://doi.org/10.1038/s41598-020-68491-4).
- [162] S. Mukudai, I. Kraja, R. Bing, D. M. Nalband, M. Tatikola, N. Hiwatashi, K. Kirshenbaum, and R. C. Branski. Implementing Efficient Peptoid-Mediated Delivery of RNA-Based Therapeutics to the Vocal Folds. *Laryngoscope Investigative Otolaryngology*, 4(6):640–644, Dec. 2019. doi:[10.1002/lio2.310](https://doi.org/10.1002/lio2.310).
- [163] M. Wang, Y. Song, P. Mu, X. Cai, Y. Lin, and C.-L. Chen. Peptoid-Based Programmable 2D Nanomaterial Sensor for Selective and Sensitive Detection of H₂ S in Live Cells. *ACS Applied Bio Materials*, 3(9):6039–6048, Sept. 2020. doi:[10.1021/acsbm.0c00657](https://doi.org/10.1021/acsbm.0c00657).
- [164] Y. Li, J. C. Tom, P. Biehl, J. Ling, and F. H. Schacher. Block Polypeptides: Synthesis, Characterization, and Response Toward Irradiation with UV Light and Temperature. *Macromolecules*, 53(13):5218–5226, July 2020. doi:[10.1021/acs.macromol.0c00654](https://doi.org/10.1021/acs.macromol.0c00654).

- [165] Q. Zhang, C. Zhang, W. Luo, L. Cui, Y. Wang, T. Jian, X. Li, Q. Yan, H. Liu, C. Ouyang, Y. Chen, C. Chen, and J. Zhang. Sequence-Defined Peptoids with -OH and -COOH Groups As Binders to Reduce Cracks of Si Nanoparticles of Lithium-Ion Batteries. *Advanced Science*, 7(18):2000749, Sept. 2020. doi:10.1002/advs.202000749.
- [166] C.-L. Chen, J. Qi, R. N. Zuckermann, and J. J. DeYoreo. Engineered Biomimetic Polymers as Tunable Agents for Controlling CaCO₃ Mineralization. *Journal of the American Chemical Society*, 133(14):5214–5217, Apr. 2011. doi:10.1021/ja200595f.
- [167] S. Zhang, J. Chen, J. Liu, H. Pyles, D. Baker, C. Chen, and J. J. De Yoreo. Engineering Biomolecular Self-Assembly at Solid–Liquid Interfaces. *Advanced Materials*, 33(23):1905784, July 2020. doi:10.1002/adma.201905784.
- [168] S. Xuan and R. N. Zuckermann. Engineering the atomic structure of sequence-defined peptoid polymers and their assemblies. *Polymer*, 202:122691, Aug. 2020. doi:10.1016/j.polymer.2020.122691.
- [169] B. Cai, Z. Li, and C.-L. Chen. Programming Amphiphilic Peptoid Oligomers for Hierarchical Assembly and Inorganic Crystallization. *Accounts of Chemical Research*, 54(1):81–91, Jan. 2021. doi:10.1021/acs.accounts.0c00533.
- [170] N. H. Shah, G. L. Butterfoss, K. Nguyen, B. Yoo, R. Bonneau, D. L. Rabenstein, and K. Kirshenbaum. Oligo(*N*-aryl glycines): A New Twist on Structured Peptoids. *Journal of the American Chemical Society*, 130(49):16622–16632, Dec. 2008. doi:10.1021/ja804580n.
- [171] K. Kirshenbaum, A. E. Barron, R. A. Goldsmith, P. Armand, E. K. Bradley, K. T. V. Truong, K. A. Dill, F. E. Cohen, and R. N. Zuckermann. Sequence-specific polypeptoids: A diverse family of heteropolymers with stable secondary structure. *Proceedings of the National Academy of Sciences*, 95(8):4303–4308, Apr. 1998. doi:10.1073/pnas.95.8.4303.
- [172] P. Armand, K. Kirshenbaum, R. A. Goldsmith, S. Farr-Jones, A. E. Barron, K. T. V. Truong, K. A. Dill, D. F. Mierke, F. E. Cohen, R. N. Zuckermann, and E. K. Bradley. NMR determination of the major solution conformation of a peptoid pentamer with chiral side chains. *Proceedings of the National Academy of Sciences*, 95(8):4309–4314, Apr. 1998. doi:10.1073/pnas.95.8.4309.
- [173] A. M. Rosales, H. K. Murnen, S. R. Kline, R. N. Zuckermann, and R. A. Segalman. Determination of the persistence length of helical and non-helical polypeptoids in solution. *Soft Matter*, 8(13):3673, 2012. doi:10.1039/c2sm07092h.
- [174] R. Kudirka, H. Tran, B. Sanii, K. T. Nam, P. H. Choi, N. Venkateswaran, R. Chen, S. Whitlam, and R. N. Zuckermann. Folding of a single-chain, information-rich

- polypeptoid sequence into a highly ordered nanosheet. *Biopolymers*, 96(5):586–595, 2011. doi:10.1002/bip.21590.
- [175] A. L. Patterson, S. P. O. Danielsen, B. Yu, E. C. Davidson, G. H. Fredrickson, and R. A. Segalman. Sequence Effects on Block Copolymer Self-Assembly through Tuning Chain Conformation and Segregation Strength Utilizing Sequence-Defined Polypeptoids. *Macromolecules*, 52(3):1277–1286, Feb. 2019. doi:10.1021/acs.macromol.8b02298.
 - [176] W. van Zoelen, H. G. Buss, N. C. Ellebracht, N. A. Lynd, D. A. Fischer, J. Finlay, S. Hill, M. E. Callow, J. A. Callow, E. J. Kramer, R. N. Zuckermann, and R. A. Segalman. Sequence of Hydrophobic and Hydrophilic Residues in Amphiphilic Polymer Coatings Affects Surface Structure and Marine Antifouling/Fouling Release Properties. *ACS Macro Letters*, 3(4):364–368, Apr. 2014. doi:10.1021/mz500090n.
 - [177] K. H. A. Lau, T. S. Sileika, S. H. Park, A. M. Sousa, P. Burch, I. Szleifer, and P. B. Messersmith. Molecular Design of Antifouling Polymer Brushes Using Sequence-Specific Peptoids. *Advanced Materials Interfaces*, 2(1):1400225, Jan. 2015. doi:10.1002/admi.201400225.
 - [178] E. Maron, Z. Kochovski, R. N. Zuckermann, and H. G. Börner. Peptide-Assisted Design of Peptoid Sequences: One Small Step in Structure and Distinct Leaps in Functions. *ACS Macro Letters*, 9:233–237, Jan. 2020. doi:10.1021/acsmacrolett.9b00977.
 - [179] L. Zhu, J. M. Simpson, X. Xu, H. He, D. Zhang, and L. Yin. Cationic Polypeptoids with Optimized Molecular Characteristics toward Efficient Nonviral Gene Delivery. *ACS Applied Materials & Interfaces*, 9(28):23476–23486, July 2017. doi:10.1021/acsami.7b06031.
 - [180] M. L. Huang, D. Ehre, Q. Jiang, C. Hu, K. Kirshenbaum, and M. D. Ward. Biomimetic peptoid oligomers as dual-action antifreeze agents. *Proceedings of the National Academy of Sciences*, 109(49):19922–19927, Dec. 2012. doi:10.1073/pnas.1212826109.
 - [181] W. van Zoelen, R. N. Zuckermann, and R. A. Segalman. Tunable Surface Properties from Sequence-Specific Polypeptoid–Polystyrene Block Copolymer Thin Films. *Macromolecules*, 45(17):7072–7082, Sept. 2012. doi:10.1021/ma3009806.
 - [182] C. Leng, H. G. Buss, R. A. Segalman, and Z. Chen. Surface Structure and Hydration of Sequence-Specific Amphiphilic Polypeptoids for Antifouling/Fouling Release Applications. *Langmuir*, 31(34):9306–9311, Sept. 2015. doi:10.1021/acs.langmuir.5b01440.
 - [183] G. Jeschke. DEER Distance Measurements on Proteins. *Annual Review of Physical Chemistry*, 63(1):419–446, 2012. doi:10.1146/annurev-physchem-032511-143716.

- [184] Y.-W. Chiang, P. P. Borbat, and J. H. Freed. The determination of pair distance distributions by pulsed ESR using Tikhonov regularization. *Journal of Magnetic Resonance*, 172(2):279–295, Feb. 2005. doi:10.1016/j.jmr.2004.10.012.
- [185] D. Sezer, J. H. Freed, and B. Roux. Parametrization, Molecular Dynamics Simulation, and Calculation of Electron Spin Resonance Spectra of a Nitroxide Spin Label on a Polyalanine -Helix. *The Journal of Physical Chemistry B*, 112(18):5755–5767, May 2008. doi:10.1021/jp711375x.
- [186] R. Kaminker, I. Kaminker, W. R. Gutekunst, Y. Luo, S. Lee, J. Niu, S. Han, and C. J. Hawker. Tuning conformation and properties of peptidomimetic backbones through dual *N* / *C* -substitution. *Chemical Communications*, 54(41):5237–5240, 2018. doi:10.1039/C8CC01356J.
- [187] J. Morimoto, Y. Fukuda, D. Kuroda, T. Watanabe, F. Yoshida, M. Asada, T. Nakamura, A. Senoo, S. Nagatoishi, K. Tsumoto, and S. Sando. A Peptoid with Extended Shape in Water. *Journal of the American Chemical Society*, 141(37):14612–14623, Sept. 2019. doi:10.1021/jacs.9b04371.
- [188] S. A. Kotler, V. Tugarinov, T. Schmidt, A. Ceccon, D. S. Libich, R. Ghirlando, C. D. Schwieters, and G. M. Clore. Probing initial transient oligomerization events facilitating Huntingtin fibril nucleation at atomic resolution by relaxation-based NMR. *Proceedings of the National Academy of Sciences*, 116(9):3562–3571, Feb. 2019. doi:10.1073/pnas.1821216116.
- [189] T. Schmidt, C. D. Schwieters, and G. M. Clore. Spatial domain organization in the HIV-1 reverse transcriptase p66 homodimer precursor probed by double electron-electron resonance EPR. *Proceedings of the National Academy of Sciences*, 116(36):17809–17816, Sept. 2019. doi:10.1073/pnas.1911086116.
- [190] T. Schmidt, J. Jeon, Y. Okuno, S. C. Chiliveri, and G. M. Clore. Submillisecond Freezing Permits Cryoprotectant-Free EPR Double Electron-Electron Resonance Spectroscopy. *ChemPhysChem*, 21(12):1224–1229, June 2020. doi:10.1002/cphc.202000312.
- [191] T. Schmidt, J. M. Louis, and G. M. Clore. Probing the Interaction between HIV-1 Protease and the Homodimeric p66/p66’ Reverse Transcriptase Precursor by Double Electron-Electron Resonance EPR Spectroscopy. *ChemBioChem*, 21(21):3051–3055, Nov. 2020. doi:10.1002/cbic.202000263.
- [192] N. Sherck, T. Webber, D. R. Brown, T. Keller, M. Barry, A. DeStefano, S. Jiao, R. A. Segalman, G. H. Fredrickson, M. S. Shell, and S. Han. End-to-End Distance Probability Distributions of Dilute Poly(ethylene oxide) in Aqueous Solution. *Journal of the American Chemical Society*, 142(46):19631–19641, Nov. 2020. doi:10.1021/jacs.0c08709.

- [193] L. Fábregas Ibáñez, G. Jeschke, and S. Stoll. DeerLab: a comprehensive software package for analyzing dipolar electron paramagnetic resonance spectroscopy data. *Magnetic Resonance*, 1(2):209–224, Oct. 2020. doi:10.5194/mr-1-209-2020.
- [194] T. H. Edwards and S. Stoll. A Bayesian approach to quantifying uncertainty from experimental noise in DEER spectroscopy. *Journal of Magnetic Resonance*, 270:87–97, Sept. 2016. doi:10.1016/j.jmr.2016.06.021.
- [195] T. H. Edwards and S. Stoll. Optimal Tikhonov regularization for DEER spectroscopy. *Journal of Magnetic Resonance*, 288:58–68, Mar. 2018. doi:10.1016/j.jmr.2018.01.021.
- [196] M. Srivastava and J. H. Freed. Singular Value Decomposition Method to Determine Distance Distributions in Pulsed Dipolar Electron Spin Resonance. *The Journal of Physical Chemistry Letters*, 8(22):5648–5655, Nov. 2017. doi:10.1021/acs.jpcllett.7b02379.
- [197] M. Srivastava and J. H. Freed. Singular Value Decomposition Method To Determine Distance Distributions in Pulsed Dipolar Electron Spin Resonance: II. Estimating Uncertainty. *The Journal of Physical Chemistry A*, 123(1):359–370, Jan. 2019. doi:10.1021/acs.jpca.8b07673.
- [198] E. S. de Souza, I. Y. Hirata, L. Juliano, and A. S. Ito. End-to-end distance distribution in bradykinin observed by Förster resonance energy transfer. *Biochimica et Biophysica Acta (BBA) - General Subjects*, 1474(2):251–261, Apr. 2000. doi:10.1016/S0304-4165(00)00004-0.
- [199] R. B. Best, K. A. Merchant, I. V. Gopich, B. Schuler, Bax, Ad, and Eaton, William A. Effect of flexibility and cis residues in single-molecule FRET studies of polyproline. *Proceedings of the National Academy of Sciences*, 104(48):18964–18969, 2007. doi:10.1073/pnas.0709567104.
- [200] J. A. Riback, M. A. Bowman, A. M. Zmyslowski, K. W. Plaxco, P. L. Clark, and T. R. Sosnick. Commonly used FRET fluorophores promote collapse of an otherwise disordered protein. *Proceedings of the National Academy of Sciences*, 116(18):8889–8894, Apr. 2019. doi:10.1073/pnas.1813038116.
- [201] J. Banham, C. Baker, S. Ceola, I. Day, G. Grant, E. Groenen, C. Rodgers, G. Jeschke, and C. Timmel. Distance measurements in the borderline region of applicability of CW EPR and DEER: A model study on a homologous series of spin-labelled peptides. *Journal of Magnetic Resonance*, 191(2):202–218, Apr. 2008. doi:10.1016/j.jmr.2007.11.023.
- [202] Q. Sui, D. Borchardt, and D. L. Rabenstein. Kinetics and Equilibria of Cis/Trans Isomerization of Backbone Amide Bonds in Peptoids. *Journal of the American Chemical Society*, 129(39):12042–12048, Oct. 2007. doi:10.1021/ja0740925.

- [203] F. P. Gasparro and N. H. Kolodny. NMR determination of the rotational barrier in N,N-dimethylacetamide. *Journal of Chemical Education*, 54(4):258, Apr. 1977. doi:10.1021/ed054p258.
- [204] L. J. Weiser and E. E. Santiso. Molecular modeling studies of peptoid polymers. *AIMS Materials Science*, 4(5):1029–1051, 2017. doi:10.3934/matersci.2017.5.1029.
- [205] A. S. Paluch, D. L. Mobley, and E. J. Maginn. Small Molecule Solvation Free Energy: Enhanced Conformational Sampling Using Expanded Ensemble Molecular Dynamics Simulation. *Journal of Chemical Theory and Computation*, 7(9):2910–2918, Sept. 2011. doi:10.1021/ct200377w.
- [206] D. T. Mirijanian, R. V. Mannige, R. N. Zuckermann, and S. Whitelam. Development and use of an atomistic CHARMM-based forcefield for peptoid simulation. *Journal of Computational Chemistry*, 35(5):360–370, Feb. 2014. doi:10.1002/jcc.23478.
- [207] S. Mukherjee, G. Zhou, C. Michel, and V. A. Voelz. Insights into Peptoid Helix Folding Cooperativity from an Improved Backbone Potential. *The Journal of Physical Chemistry B*, 119(50):15407–15417, Dec. 2015. doi:10.1021/acs.jpcb.5b09625.
- [208] G. L. Butterfoss, B. Yoo, J. N. Jaworski, I. Chorny, K. A. Dill, R. N. Zuckermann, R. Bonneau, K. Kirshenbaum, and V. A. Voelz. De novo structure prediction and experimental characterization of folded peptoid oligomers. *Proceedings of the National Academy of Sciences*, 109(36):14320–14325, Sept. 2012. doi:10.1073/pnas.1209945109.
- [209] A. Prakash, M. D. Baer, C. J. Mundy, and J. Pfaendtner. Peptoid Backbone Flexibility Dictates Its Interaction with Water and Surfaces: A Molecular Dynamics Investigation. *Biomacromolecules*, 19(3):1006–1015, Mar. 2018. doi:10.1021/acs.biomac.7b01813.
- [210] L. J. Weiser and E. E. Santiso. A CGenFF-based force field for simulations of peptoids with both *cis* and *trans* peptide bonds. *Journal of Computational Chemistry*, 40(22):1946–1956, Aug. 2019. doi:10.1002/jcc.25850.
- [211] M. Zhao, J. Sampath, S. Alamdari, G. Shen, C.-L. Chen, C. J. Mundy, J. Pfaendtner, and A. L. Ferguson. A MARTINI-Compatible Coarse-Grained Model for the Mesoscale Simulation of Peptoids. *The Journal of Physical Chemistry B*, 124(36):7745–7764, Aug. 2020. doi:10.1021/acs.jpcb.0c04567.
- [212] S. Jiao, A. DeStefano, J. I. Monroe, M. Barry, N. Sherck, T. Casey, R. A. Segalman, S. Han, and M. S. Shell. Quantifying Polypeptoid Conformational Landscapes through Integrated Experiment and Simulation. *Macromolecules*, 54(11):5011–5021, June 2021. doi:10.1021/acs.macromol.1c00550.

- [213] M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek, and G. R. Hutchison. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *Journal of Cheminformatics*, 4(1):17, Dec. 2012. doi:10.1186/1758-2946-4-17.
- [214] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1-2:19–25, Sept. 2015. doi:10.1016/j.softx.2015.06.001.
- [215] A. Bondi. van der Waals Volumes and Radii. *The Journal of Physical Chemistry*, 68(3):441–451, Mar. 1964. doi:10.1021/j100785a001.
- [216] A. P. Lyubartsev, A. A. Martsinovski, S. V. Shevkunov, and P. N. Vorontsov-Velyaminov. New approach to Monte Carlo calculation of the free energy: Method of expanded ensembles. *The Journal of Chemical Physics*, 96(3):1776–1783, Feb. 1992. doi:10.1063/1.462133.
- [217] F. A. Escobedo and F. J. Martínez-Veracoechea. Optimized expanded ensembles for simulations involving molecular insertions and deletions. I. Closed systems. *The Journal of Chemical Physics*, 127(17):174103, Nov. 2007. doi:10.1063/1.2800320.
- [218] J. D. Chodera, W. C. Swope, J. W. Pitera, C. Seok, and K. A. Dill. Use of the Weighted Histogram Analysis Method for the Analysis of Simulated and Parallel Tempering Simulations. *Journal of Chemical Theory and Computation*, 3(1):26–41, Jan. 2007. doi:10.1021/ct0502864.
- [219] V. A. Voelz, K. A. Dill, and I. Chorny. Peptoid conformational free energy landscapes from implicit-solvent molecular simulations in AMBER. *Biopolymers*, 96(5):639–650, 2011. doi:10.1002/bip.21575.
- [220] R. V. Mannige, T. K. Haxton, C. Proulx, E. J. Robertson, A. Battigelli, G. L. Butterfoss, R. N. Zuckermann, and S. Whitelam. Peptoid nanosheets exhibit a new secondary-structure motif. *Nature*, 526(7573):415–420, Oct. 2015. doi:10.1038/nature15363.
- [221] W. Brandt, T. Herberg, and L. Wessjohann. Systematic conformational investigations of peptoids and peptoid-peptide chimeras. *Biopolymers*, 96(5):651–668, 2011. doi:10.1002/bip.21620.
- [222] J. Lin. Divergence measures based on the Shannon entropy. *IEEE Transactions on Information Theory*, 37(1):145–151, Jan. 1991. doi:10.1109/18.61115.
- [223] S. Piana, A. G. Donchev, P. Robustelli, and D. E. Shaw. Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States. *The Journal of Physical Chemistry B*, 119(16):5113–5123, Apr. 2015. doi:10.1021/jp508971m.

- [224] J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. de Groot, H. Grubmüller, and A. D. MacKerell. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature Methods*, 14(1):71–73, Jan. 2017. doi:10.1038/nmeth.4067.
- [225] A. J. Patel, P. Varilly, S. N. Jamadagni, H. Acharya, S. Garde, and D. Chandler. Extended surfaces modulate hydrophobic interactions of neighboring solutes. *Proceedings of the National Academy of Sciences*, 108(43):17678–17683, Oct. 2011. doi:10.1073/pnas.1110703108.
- [226] K. Meister, S. Strazdaite, A. L. DeVries, S. Lotze, L. L. C. Olijve, I. K. Voets, and H. J. Bakker. Observation of ice-like water layers at an aqueous protein surface. *Proceedings of the National Academy of Sciences*, 111(50):17732–17736, Dec. 2014. doi:10.1073/pnas.1414188111.
- [227] M.-C. Bellissent-Funel, A. Hassanali, M. Havenith, R. Henchman, P. Pohl, F. Sterpone, D. van der Spoel, Y. Xu, and A. E. Garcia. Water Determines the Structure and Dynamics of Proteins. *Chemical Reviews*, 116(13):7673–7697, July 2016. doi:10.1021/acs.chemrev.5b00664.
- [228] S. Bonella, D. Raimondo, E. Milanetti, A. Tramontano, and G. Ciccotti. Mapping the Hydropathy of Amino Acids Based on Their Local Solvation Structure. *The Journal of Physical Chemistry B*, 118(24):6604–6613, June 2014. doi:10.1021/jp500980x.
- [229] A. Chowdhury, S. A. Kovalenko, I. V. Aramburu, P. S. Tan, N. P. Ernstring, and E. A. Lemke. Mechanism-Dependent Modulation of Ultrafast Interfacial Water Dynamics in Intrinsically Disordered Protein Complexes. *Angewandte Chemie International Edition*, 58(14):4720–4724, Mar. 2019. doi:10.1002/anie.201813354.
- [230] Y. Fichou, G. Schirò, F.-X. Gallat, C. Laguri, M. Moulin, J. Combet, M. Zamponi, M. Härtlein, C. Picart, E. Mossou, H. Lortat-Jacob, J.-P. Colletier, D. J. Tobias, and M. Weik. Hydration water mobility is enhanced around tau amyloid fibers. *Proceedings of the National Academy of Sciences*, 112(20):6365–6370, May 2015. doi:10.1073/pnas.1422824112.
- [231] S. H. Lee and P. J. Rossky. A comparison of the structure and dynamics of liquid water at hydrophobic and hydrophilic surfaces—a molecular dynamics simulation study. *The Journal of Chemical Physics*, 100(4):3334–3345, Feb. 1994. doi:10.1063/1.466425.
- [232] A. J. Patel, P. Varilly, and D. Chandler. Fluctuations of Water near Extended Hydrophobic and Hydrophilic Surfaces. *The Journal of Physical Chemistry B*, 114(4):1632–1637, Feb. 2010. doi:10.1021/jp909048f.

- [233] T. Head-Gordon. Is water structure around hydrophobic groups clathrate-like? *Proceedings of the National Academy of Sciences*, 92(18):8308–8312, Aug. 1995. doi:10.1073/pnas.92.18.8308.
- [234] N. Galamba. Water’s Structure around Hydrophobic Solutes and the Iceberg Model. *The Journal of Physical Chemistry B*, 117(7):2153–2159, Feb. 2013. doi:10.1021/jp310649n.
- [235] T. Head-Gordon, J. Sorenson, A. Pertsemlidis, and R. Glaeser. Differences in hydration structure near hydrophobic and hydrophilic amino acids. *Biophysical Journal*, 73(4):2106–2115, Oct. 1997. doi:10.1016/S0006-3495(97)78241-9.
- [236] K. Shiraga, T. Suzuki, N. Kondo, and Y. Ogawa. Hydration and hydrogen bond network of water around hydrophobic surface investigated by terahertz spectroscopy. *The Journal of Chemical Physics*, 141(23):235103, Dec. 2014. doi:10.1063/1.4903544.
- [237] D. Russo, R. K. Murarka, J. R. D. Copley, and T. Head-Gordon. Molecular View of Water Dynamics near Model Peptides. *The Journal of Physical Chemistry B*, 109(26):12966–12975, July 2005. doi:10.1021/jp051137k.
- [238] G. Niehues, M. Heyden, D. A. Schmidt, and M. Havenith. Exploring hydrophobicity by THz absorption spectroscopy of solvated amino acids. *Faraday Discussions*, 150:193–207, 2011. doi:10.1039/c0fd00007h.
- [239] T. Seto, T. Mashimo, I. Yoshiya, A. Shimizu, K. Fumino, and Y. Taniguchi. NMR Studies on Dynamic Structure of Hydrated Water Molecules in Alcohol-D₂O Solutions. *Journal of Molecular Liquids*, 70(1):1–9, Aug. 1996. doi:10.1016/0167-7322(96)00951-8.
- [240] J. Qvist and B. Halle. Thermal Signature of Hydrophobic Hydration Dynamics. *Journal of the American Chemical Society*, 130(31):10345–10353, Aug. 2008. doi:10.1021/ja802668w.
- [241] P. Rani and P. Biswas. Local Structure and Dynamics of Hydration Water in Intrinsically Disordered Proteins. *The Journal of Physical Chemistry B*, 119(34):10858–10867, Aug. 2015. doi:10.1021/jp511961c.
- [242] L. Aggarwal and P. Biswas. Hydration Water Distribution around Intrinsically Disordered Proteins. *The Journal of Physical Chemistry B*, 122(15):4206–4218, Apr. 2018. doi:10.1021/acs.jpcc.7b11091.
- [243] A. J. DeStefano, R. A. Segalman, and E. C. Davidson. Where Biology and Traditional Polymers Meet: The Potential of Associating Sequence-Defined Polymers for Materials Science. *JACS Au*, 1(10):1556–1571, Aug. 2021. doi:10.1021/jacsau.1c00297.

- [244] D. L. Cheung and K. H. A. Lau. Atomistic Study of Zwitterionic Peptoid Antifouling Brushes. *Langmuir*, 35(5):1483–1494, Feb. 2019. doi:10.1021/acs.langmuir.8b01939.
- [245] S. Jiao, D. M. Rivera Mirabal, A. J. DeStefano, R. A. Segalman, S. Han, and M. S. Shell. Sequence Modulates Polypeptoid Hydration Water Structure and Dynamics. *Biomacromolecules*, 23(4):1745–1756, Apr. 2022. doi:10.1021/acs.biomac.1c01687.
- [246] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *The Journal of Physical Chemistry B*, 102(18):3586–3616, Apr. 1998. doi:10.1021/jp973084f.
- [247] S. Izadi, R. Anandakrishnan, and A. V. Onufriev. Building Water Models: A Different Approach. *The Journal of Physical Chemistry Letters*, 5(21):3863–3871, Nov. 2014. doi:10.1021/jz501780a.
- [248] A. V. Onufriev and S. Izadi. Water models for biomolecular simulations: Water models for biomolecular simulations. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 8(2):e1347, Mar. 2018. doi:10.1002/wcms.1347.
- [249] A. M. Schrader, J. I. Monroe, R. Sheil, H. A. Dobbs, T. J. Keller, Y. Li, S. Jain, M. S. Shell, J. N. Israelachvili, and S. Han. Surface chemical heterogeneity modulates silica surface hydration. *Proceedings of the National Academy of Sciences*, 115(12):2890–2895, Mar. 2018. doi:10.1073/pnas.1722263115.
- [250] F. S. Emami, V. Puddu, R. J. Berry, V. Varshney, S. V. Patwardhan, C. C. Perry, and H. Heinz. Force Field and a Surface Model Database for Silica to Simulate Interfacial Properties in Atomic Resolution. *Chemistry of Materials*, 26(8):2647–2658, Apr. 2014. doi:10.1021/cm500365c.
- [251] D. Dhabal, M. Singh, K. T. Wikfeldt, and C. Chakravarty. Triplet correlation functions in liquid water. *The Journal of Chemical Physics*, 141(17):174504, Nov. 2014. doi:10.1063/1.4898755.
- [252] P. J. Rossky and M. Karplus. Solvation. A molecular dynamics study of a dipeptide in water. *Journal of the American Chemical Society*, 101(8):1913–1937, Apr. 1979. doi:10.1021/ja00502a001.
- [253] D. Russo, G. Hura, and T. Head-Gordon. Hydration Dynamics Near a Model Protein Surface. *Biophysical Journal*, 86(3):1852–1862, Mar. 2004. doi:10.1016/S0006-3495(04)74252-6.

- [254] A. Scala, F. W. Starr, E. L. Nave, F. Sciortino, and H. E. Stanley. Configurational entropy and diffusivity of supercooled water. *Nature*, 406(6792):166–169, July 2000. doi:10.1038/35018034.
- [255] M. Agarwal, M. Singh, R. Sharma, M. Parvez Alam, and C. Chakravarty. Relationship between Structure, Entropy, and Diffusivity in Water and Water-Like Liquids. *The Journal of Physical Chemistry B*, 114(20):6995–7001, May 2010. doi:10.1021/jp101956u.
- [256] R. Chopra, T. M. Truskett, and J. R. Errington. On the Use of Excess Entropy Scaling to Describe the Dynamic Properties of Water. *The Journal of Physical Chemistry B*, 114(32):10558–10566, Aug. 2010. doi:10.1021/jp1049155.
- [257] D. Nayar and C. Chakravarty. Water and water-like liquids: relationships between structure, entropy and mobility. *Physical Chemistry Chemical Physics*, 15(34):14162–14177, July 2013. doi:10.1039/C3CP51114F.
- [258] N. Galamba. Water Tetrahedrons, Hydrogen-Bond Dynamics, and the Orientational Mobility of Water around Hydrophobic Solutes. *The Journal of Physical Chemistry B*, 118(15):4169–4176, Apr. 2014. doi:10.1021/jp500067a.
- [259] J. N. Dahanayake and K. R. Mitchell-Koch. Entropy connects water structure and dynamics in protein hydration layer. *Physical Chemistry Chemical Physics*, 20(21):14765–14777, 2018. doi:10.1039/C8CP01674G.
- [260] J. Monroe, M. Barry, A. DeStefano, P. Aydogan Gokturk, S. Jiao, D. Robinson-Brown, T. Webber, E. J. Crumlin, S. Han, and M. S. Shell. Water Structure and Properties at Hydrophilic and Hydrophobic Surfaces. *Annual Review of Chemical and Biomolecular Engineering*, 11(1):523–557, June 2020. doi:10.1146/annurev-chembioeng-120919-114657.
- [261] J. M. Franck and S. Han. Overhauser Dynamic Nuclear Polarization for the Study of Hydration Dynamics, Explained. In *Methods in Enzymology*, volume 615, pages 131–175. Elsevier, 2019. doi:10.1016/bs.mie.2018.09.024.
- [262] R. Barnes, S. Sun, Y. Fichou, F. W. Dahlquist, M. Heyden, and S. Han. Spatially Heterogeneous Surface Water Diffusivity around Structured Protein Surfaces at Equilibrium. *Journal of the American Chemical Society*, 139(49):17890–17901, Dec. 2017. doi:10.1021/jacs.7b08606.
- [263] R. Mills. Self-diffusion in normal and heavy water in the range 1–45°. *The Journal of Physical Chemistry*, 77(5):685–688, Mar. 1973. doi:10.1021/j100624a025.
- [264] A. J. Easteal, W. E. Price, and L. A. Woolf. Diaphragm cell for high-temperature diffusion measurements. Tracer Diffusion coefficients for water to 363 K. *Journal of the Chemical Society, Faraday Transactions 1: Physical Chemistry in Condensed Phases*, 85(5):1091–1097, 1989. doi:10.1039/f19898501091.

- [265] K. R. Harris and L. A. Woolf. Pressure and temperature dependence of the self diffusion coefficient of water and oxygen-18 water. *Journal of the Chemical Society, Faraday Transactions 1: Physical Chemistry in Condensed Phases*, 76:377–385, 1980. doi:[10.1039/f19807600377](https://doi.org/10.1039/f19807600377).
- [266] P. Tofts, D. Lloyd, C. Clark, G. Barker, G. Parker, P. McConville, C. Baldock, and J. Pope. Test liquids for quantitative MRI measurements of self-diffusion coefficient in vivo. *Magnetic Resonance in Medicine*, 43(3):368–374, Mar. 2000. doi:[10.1002/\(SICI\)1522-2594\(200003\)43:3<368::AID-MRM8;3.0.CO;2-B](https://doi.org/10.1002/(SICI)1522-2594(200003)43:3<368::AID-MRM8;3.0.CO;2-B).
- [267] M. Holz, S. R. Heil, and A. Sacco. Temperature-dependent self-diffusion coefficients of water and six selected molecular liquids for calibration in accurate ^1H NMR PFG measurements. *Physical Chemistry Chemical Physics*, 2(20):4740–4742, 2000. doi:[10.1039/b005319h](https://doi.org/10.1039/b005319h).
- [268] K. T. Gillen, D. C. Douglass, and M. J. R. Hoch. Self-Diffusion in Liquid Water to 31°C . *The Journal of Chemical Physics*, 57(12):5117–5119, Dec. 1972. doi:[10.1063/1.1678198](https://doi.org/10.1063/1.1678198).
- [269] J. S. Murday and R. M. Cottis. Self-Diffusion in Liquids: H_2O , D_2O , and Na. *The Journal of Chemical Physics*, 53(12):4724–4725, Dec. 1970. doi:[10.1063/1.1674011](https://doi.org/10.1063/1.1674011).
- [270] C. R. Becker, L. R. Schad, and W. J. Lorenz. Measurement of diffusion coefficients using a quick echo split NMR imaging technique. *Magnetic Resonance Imaging*, 12(8):1167–1174, Jan. 1994. doi:[10.1016/0730-725X\(94\)90082-3](https://doi.org/10.1016/0730-725X(94)90082-3).
- [271] D. E. O'Reilly and E. M. Peterson. Self-Diffusion Coefficients and Rotational Correlation Times in Polar Liquids. II. *The Journal of Chemical Physics*, 55(5):2155–2163, Sept. 1971. doi:[10.1063/1.1676386](https://doi.org/10.1063/1.1676386).
- [272] Olaf Dietrich. Diffusion Coefficients of Water, 2002. URL <https://dtrx.de/od/diff/>.
- [273] F. Sterpone, G. Stirnemann, J. T. Hynes, and D. Laage. Water Hydrogen-Bond Dynamics around Amino Acids: The Key Role of Hydrophilic Hydrogen-Bond Acceptor Groups. *The Journal of Physical Chemistry B*, 114(5):2083–2089, Feb. 2010. doi:[10.1021/jp9119793](https://doi.org/10.1021/jp9119793).
- [274] M. Szöri, D. J. Tobias, and M. Roeselová. Microscopic Wetting of Mixed Self-assembled Monolayers: A Molecular Dynamics Study. *The Journal of Physical Chemistry B*, 113(13):4161–4169, Apr. 2009. doi:[10.1021/jp8074139](https://doi.org/10.1021/jp8074139).
- [275] M. Kanduč and R. R. Netz. Atomistic simulations of wetting properties and water films on hydrophilic surfaces. *The Journal of Chemical Physics*, 146(16):164705, Apr. 2017. doi:[10.1063/1.4979847](https://doi.org/10.1063/1.4979847).

- [276] C. Wang, C.-K. D. Ma, H. Yeon, X. Wang, S. H. Gellman, and N. L. Abbott. Non-additive Interactions Mediated by Water at Chemically Heterogeneous Surfaces: Nonionic Polar Groups and Hydrophobic Interactions. *Journal of the American Chemical Society*, 139(51):18536–18544, Dec. 2017. doi:[10.1021/jacs.7b08367](https://doi.org/10.1021/jacs.7b08367).
- [277] X. Huang, C. J. Margulis, and B. J. Berne. Do Molecules as Small as Neopentane Induce a Hydrophobic Response Similar to That of Large Hydrophobic Surfaces? *The Journal of Physical Chemistry B*, 107(42):11742–11748, Oct. 2003. doi:[10.1021/jp030652k](https://doi.org/10.1021/jp030652k).
- [278] F. Persson, P. Söderhjelm, and B. Halle. How proteins modify water dynamics. *The Journal of Chemical Physics*, 148(21):215103, June 2018. doi:[10.1063/1.5026861](https://doi.org/10.1063/1.5026861).
- [279] F. Sterpone, G. Stirnemann, and D. Laage. Magnitude and Molecular Origin of Water Slowdown Next to a Protein. *Journal of the American Chemical Society*, 134(9):4116–4119, Mar. 2012. doi:[10.1021/ja3007897](https://doi.org/10.1021/ja3007897).
- [280] S. N. Jamadagni, R. Godawat, and S. Garde. Hydrophobicity of Proteins and Interfaces: Insights from Density Fluctuations. *Annual Review of Chemical and Biomolecular Engineering*, 2(1):147–171, 2011. doi:[10.1146/annurev-chembioeng-061010-114156](https://doi.org/10.1146/annurev-chembioeng-061010-114156).
- [281] D. Rana and T. Matsuura. Surface Modifications for Antifouling Membranes. *Chemical Reviews*, 110(4):2448–2471, Apr. 2010. doi:[10.1021/cr800208y](https://doi.org/10.1021/cr800208y).
- [282] W. R. Gombotz, W. Guanghai, T. A. Horbett, and A. S. Hoffman. Protein adsorption to poly(ethylene oxide) surfaces. *Journal of Biomedical Materials Research*, 25(12):1547–1562, Dec. 1991. doi:[10.1002/jbm.820251211](https://doi.org/10.1002/jbm.820251211).
- [283] N. P. Desai and J. A. Hubbell. Biological responses to polyethylene oxide modified polyethylene terephthalate surfaces. *Journal of Biomedical Materials Research*, 25(7):829–843, July 1991. doi:[10.1002/jbm.820250704](https://doi.org/10.1002/jbm.820250704).
- [284] E. F. Murphy, J. R. Lu, J. Brewer, J. Russell, and J. Penfold. The Reduced Adsorption of Proteins at the Phosphoryl Choline Incorporated PolymerWater Interface. *Langmuir*, 15(4):1313–1322, Feb. 1999. doi:[10.1021/la9813580](https://doi.org/10.1021/la9813580).
- [285] R. E. Holmlin, X. Chen, R. G. Chapman, S. Takayama, and G. M. Whitesides. Zwitterionic SAMs that Resist Nonspecific Adsorption of Protein from Aqueous Buffer. *Langmuir*, 17(9):2841–2850, May 2001. doi:[10.1021/la0015258](https://doi.org/10.1021/la0015258).
- [286] Z. Zhang, T. Chao, S. Chen, and S. Jiang. Superlow Fouling Sulfobetaine and Carboxybetaine Polymers on Glass Slides. *Langmuir*, 22(24):10072–10077, Nov. 2006. doi:[10.1021/la062175d](https://doi.org/10.1021/la062175d).

- [287] S. Chen, F. Yu, Q. Yu, Y. He, and S. Jiang. Strong Resistance of a Thin Crystalline Layer of Balanced Charged Groups to Protein Adsorption. *Langmuir*, 22(19):8186–8191, Sept. 2006. doi:10.1021/la061012m.
- [288] P. Zhang, B. D. Ratner, A. S. Hoffman, and S. Jiang. Nonfouling Surfaces. In *Bio-materials Science*, pages 507–513. Elsevier, 2020. doi:10.1016/B978-0-12-816137-1.00034-9.
- [289] H. Ju, B. D. McCloskey, A. C. Sagle, Y.-H. Wu, V. A. Kusuma, and B. D. Freeman. Crosslinked poly(ethylene oxide) fouling resistant coating materials for oil/water separation. *Journal of Membrane Science*, 307(2):260–267, Jan. 2008. doi:10.1016/j.memsci.2007.09.028.
- [290] B. Li, P. Jain, J. Ma, J. K. Smith, Z. Yuan, H.-C. Hung, Y. He, X. Lin, K. Wu, J. Pfaendtner, and S. Jiang. Trimethylamine *N* -oxide–derived zwitterionic polymers: A new class of ultralow fouling bioinspired materials. *Science Advances*, 5(6):eaaw9562, June 2019. doi:10.1126/sciadv.aaw9562.
- [291] L. Wang, R. A. Friesner, and B. J. Berne. Competition of Electrostatic and Hydrophobic Interactions between Small Hydrophobes and Model Enclosures. *The Journal of Physical Chemistry B*, 114(21):7294–7301, June 2010. doi:10.1021/jp100772w.
- [292] N. Giovambattista, P. Rossky, and P. Debenedetti. Computational Studies of Pressure, Temperature, and Surface Effects on the Structure and Thermodynamics of Confined Water. *Annual Review of Physical Chemistry*, 63(1):179–200, May 2012. doi:10.1146/annurev-physchem-032811-112007.
- [293] N. B. Rego, A. L. Ferguson, and A. J. Patel. Learning the relationship between nanoscale chemical patterning and hydrophobicity. *Proceedings of the National Academy of Sciences*, 119(48):e2200018119, Nov. 2022. doi:10.1073/pnas.2200018119.
- [294] A. Lazaric, V. Pattni, K. Fuegner, A. Ben-Naim, and M. Heyden. Solvation free energy arithmetic for small organic molecules. *Journal of Computational Chemistry*, 44(13):1263–1277, May 2023. doi:10.1002/jcc.27081.
- [295] Y.-K. Cheng and P. J. Rossky. Surface topography dependence of biomolecular hydrophobic hydration. *Nature*, 392(6677):696–699, Apr. 1998. doi:10.1038/33653.
- [296] R. C. Harris and B. M. Pettitt. Effects of geometry and chemistry on hydrophobic solvation. *Proceedings of the National Academy of Sciences*, 111(41):14681–14686, Oct. 2014. doi:10.1073/pnas.1406080111.
- [297] A. J. Patel and S. Garde. Efficient Method To Characterize the Context-Dependent Hydrophobicity of Proteins. *The Journal of Physical Chemistry B*, 118(6):1564–1573, Feb. 2014. doi:10.1021/jp4081977.

- [298] I.-C. Yeh, J. L. Lenhart, and B. C. Rinderspacher. Molecular Dynamics Simulations of Adsorption of Catechol and Related Phenolic Compounds to Alumina Surfaces. *The Journal of Physical Chemistry C*, 119(14):7721–7731, Apr. 2015. doi:10.1021/jp512780s.
- [299] R. Abel, L. Wang, R. A. Friesner, and B. J. Berne. A Displaced-Solvent Functional Analysis of Model Hydrophobic Enclosures. *Journal of Chemical Theory and Computation*, 6(9):2924–2934, Sept. 2010. doi:10.1021/ct100215c.
- [300] L. F. Scatena, M. G. Brown, and G. L. Richmond. Water at Hydrophobic Surfaces: Weak Hydrogen Bonding and Strong Orientation Effects. *Science*, 292(5518):908–912, May 2001. doi:10.1126/science.1059514.
- [301] C. Leng, H.-C. Hung, S. Sun, D. Wang, Y. Li, S. Jiang, and Z. Chen. Probing the Surface Hydration of Nonfouling Zwitterionic and PEG Materials in Contact with Proteins. *ACS Applied Materials & Interfaces*, 7(30):16881–16888, Aug. 2015. doi:10.1021/acsami.5b05627.
- [302] C. Leng, S. Sun, K. Zhang, S. Jiang, and Z. Chen. Molecular level studies on interfacial hydration of zwitterionic and other antifouling polymers in situ. *Acta Biomaterialia*, 40:6–15, Aug. 2016. doi:10.1016/j.actbio.2016.02.030.
- [303] C. A. Del Grosso, C. Leng, K. Zhang, H.-C. Hung, S. Jiang, Z. Chen, and J. J. Wilker. Surface hydration for antifouling and bio-adhesion. *Chemical Science*, 11(38):10367–10377, 2020. doi:10.1039/D0SC03690K.
- [304] T. Ishiyama and A. Morita. Computational Analysis of Vibrational Sum Frequency Generation Spectroscopy. *Annual Review of Physical Chemistry*, 68(1):355–377, May 2017. doi:10.1146/annurev-physchem-052516-044806.
- [305] J. Zheng, L. Li, H.-K. Tsao, Y.-J. Sheng, S. Chen, and S. Jiang. Strong Repulsive Forces between Protein and Oligo (Ethylene Glycol) Self-Assembled Monolayers: A Molecular Simulation Study. *Biophysical Journal*, 89(1):158–166, July 2005. doi:10.1529/biophysj.105.059428.
- [306] Y. He, Y. Chang, J. C. Hower, J. Zheng, S. Chen, and S. Jiang. Origin of repulsive force and structure/dynamics of interfacial water in OEG–protein interactions: a molecular simulation study. *Physical Chemistry Chemical Physics*, 10(36):5539, 2008. doi:10.1039/b807129b.
- [307] Y. He, J. Hower, S. Chen, M. T. Bernards, Y. Chang, and S. Jiang. Molecular Simulation Studies of Protein Interactions with Zwitterionic Phosphorylcholine Self-Assembled Monolayers in the Presence of Water. *Langmuir*, 24(18):10358–10364, Sept. 2008. doi:10.1021/la8013046.

- [308] H. Huang, C. Zhang, R. Crisci, T. Lu, H.-C. Hung, M. S. J. Sajib, P. Sarker, J. Ma, T. Wei, S. Jiang, and Z. Chen. Strong Surface Hydration and Salt Resistant Mechanism of a New Nonfouling Zwitterionic Polymer Based on Protein Stabilizer TMAO. *Journal of the American Chemical Society*, 143(40):16786–16795, Oct. 2021. doi:10.1021/jacs.1c08280.
- [309] Y. Xiang, R.-G. Xu, and Y. Leng. Molecular Dynamics Simulations of a Poly(ethylene glycol)-Grafted Polyamide Membrane and Its Interaction with a Calcium Alginate Gel. *Langmuir*, 32(18):4424–4433, May 2016. doi:10.1021/acs.langmuir.6b00348.
- [310] Y. Xiang, R.-G. Xu, and Y. Leng. Molecular Simulations of the Hydration Behavior of a Zwitterion Brush Array and Its Antifouling Property in an Aqueous Environment. *Langmuir*, 34(6):2245–2257, Feb. 2018. doi:10.1021/acs.langmuir.7b03386.
- [311] V. T. Bui, A. Abdelrasoul, and D. W. McMartin. Investigation on the stability and antifouling properties of polyvinylidene fluoride (PVDF)-zwitterion mixed matrix membranes (MMMs) using molecular dynamics simulation (MDS). *Computational Materials Science*, 187:110079, Feb. 2021. doi:10.1016/j.commatsci.2020.110079.
- [312] Y. Kawabata, R. R. Gonzales, K. Nakagawa, T. Shintani, H. Matsuyama, Y. Fujimura, T. Kawakatsu, and T. Yoshioka. Molecular dynamics study on the elucidation of polyamide membrane fouling by nonionic surfactants and disaccharides. *Physical Chemistry Chemical Physics*, 23(36):20313–20322, 2021. doi:10.1039/D1CP01455B.
- [313] M. Shen, S. Keten, and R. M. Lueptow. Rejection mechanisms for contaminants in polyamide reverse osmosis membranes. *Journal of Membrane Science*, 509:36–47, July 2016. doi:10.1016/j.memsci.2016.02.043.
- [314] K. H. A. Lau, C. Ren, T. S. Sileika, S. H. Park, I. Szleifer, and P. B. Messersmith. Surface-Grafted Polysarcosine as a Peptoid Antifouling Polymer Brush. *Langmuir*, 28(46):16099–16107, Nov. 2012. doi:10.1021/la302131n.
- [315] M. E. Barry, E. C. Davidson, C. Zhang, A. L. Patterson, B. Yu, A. K. Leonardi, N. Duzen, K. Malaviya, J. L. Clarke, J. A. Finlay, A. S. Clare, Z. Chen, C. K. Ober, and R. A. Segalman. The Role of Hydrogen Bonding in Peptoid-Based Marine Antifouling Coatings. *Macromolecules*, 52(3):1287–1295, Feb. 2019. doi:10.1021/acs.macromol.8b02390.
- [316] R. Sutton and G. Sposito. Molecular Structure in Soil Humic Substances: The New View. *Environmental Science & Technology*, 39(23):9009–9015, Dec. 2005. doi:10.1021/es050778q.

- [317] J.-P. Croue, G. V. Korshin, M. M. Benjamin, and A. R. Foundation, editors. *Characterization of natural organic matter in drinking water*. AWWA Research Foundation and American Water Works Association, Denver, CO, 2000.
- [318] A. E. Ismail, G. S. Grest, and M. J. Stevens. Structure and Dynamics of Water near the Interface with Oligo(ethylene oxide) Self-Assembled Monolayers. *Langmuir*, 23(16):8508–8514, July 2007. doi:10.1021/la700829r.
- [319] T. McPherson, A. Kidane, I. Szleifer, and K. Park. Prevention of Protein Adsorption by Tethered Poly(ethylene oxide) Layers: Experiments and Single-Chain Mean-Field Analysis. *Langmuir*, 14(1):176–186, Jan. 1998. doi:10.1021/la9706781.
- [320] C.-M. Xing, F.-N. Meng, M. Quan, K. Ding, Y. Dang, and Y.-K. Gong. Quantitative fabrication, performance optimization and comparison of PEG and zwitterionic polymer antifouling coatings. *Acta Biomaterialia*, 59:129–138, Sept. 2017. doi:10.1016/j.actbio.2017.06.034.
- [321] Y. Liu, D. Zhang, B. Ren, X. Gong, L. Xu, Z.-Q. Feng, Y. Chang, Y. He, and J. Zheng. Molecular simulations and understanding of antifouling zwitterionic polymer brushes. *Journal of Materials Chemistry B*, 8(17):3814–3828, 2020. doi:10.1039/D0TB00520G.
- [322] Z. Zhang, H. Vaisocherová, G. Cheng, W. Yang, H. Xue, and S. Jiang. Nonfouling Behavior of Polycarboxybetaine-Grafted Surfaces: Structural and Environmental Effects. *Biomacromolecules*, 9(10):2686–2692, Oct. 2008. doi:10.1021/bm800407r.
- [323] J. Yang, M. Zhang, H. Chen, Y. Chang, Z. Chen, and J. Zheng. Probing the Structural Dependence of Carbon Space Lengths of Poly(*N*-hydroxyalkyl acrylamide)-Based Brushes on Antifouling Performance. *Biomacromolecules*, 15(8):2982–2991, Aug. 2014. doi:10.1021/bm500598a.
- [324] D. M. Huang, P. L. Geissler, and D. Chandler. Scaling of Hydrophobic Solvation Free Energies. *The Journal of Physical Chemistry B*, 105(28):6704–6709, July 2001. doi:10.1021/jp0104029.
- [325] A. P. Willard and D. Chandler. The molecular structure of the interface between water and a hydrophobic substrate is liquid-vapor like. *The Journal of Chemical Physics*, 141(18):18C519, Nov. 2014. doi:10.1063/1.4897249.
- [326] M. E. Barry, P. Aydogan Gokturk, A. J. DeStefano, A. K. Leonardi, C. K. Ober, E. J. Crumlin, and R. A. Segalman. Effects of Amphiphilic Polypeptoid Side Chains on Polymer Surface Chemistry and Hydrophilicity. *ACS Applied Materials & Interfaces*, 14(5):7340–7349, Feb. 2022. doi:10.1021/acsami.1c22683.
- [327] A. Statt, D. C. Kleeblatt, and W. F. Reinhart. Unsupervised learning of sequence-specific aggregation behavior for a model copolymer. *Soft Matter*, 17(33):7697–7707, 2021. doi:10.1039/D1SM01012C.

- [328] D. Bhattacharya, D. C. Kleeblatt, A. Statt, and W. F. Reinhart. Predicting aggregate morphology of sequence-defined macromolecules with recurrent neural networks. *Soft Matter*, 18(27):5037–5051, 2022. doi:10.1039/D2SM00452F.
- [329] M. Chruszcz, A. Wlodawer, and W. Minor. Determination of Protein Structures—A Series of Fortunate Events. *Biophysical Journal*, 95(1):1–9, July 2008. doi:10.1529/biophysj.108.131789.
- [330] E. C. Day, S. S. Chittari, K. C. Cunha, J. Zhao, J. N. Dodds, D. C. Davis, E. S. Baker, R. B. Berlow, J.-E. Shea, R. U. Kulkarni, and A. S. Knight. A High-Throughput Workflow to Analyze Sequence-Conformation Relationships and Explore Hydrophobic Patterning in Disordered Peptoids. preprint, Chemistry, Oct. 2023. doi:10.26434/chemrxiv-2023-b84wl.
- [331] M. A. Webb, N. E. Jackson, P. S. Gil, and J. J. De Pablo. Targeted sequence design within the coarse-grained polymer genome. *Science Advances*, 6(43):eabc6216, Oct. 2020. doi:10.1126/sciadv.abc6216.
- [332] J. M. Lotthammer, G. M. Ginell, D. Griffith, R. J. Emenecker, and A. S. Holehouse. Direct Prediction of Intrinsically Disordered Protein Conformational Properties From Sequence: Version 3 [2023-05-28]. preprint, Biophysics, May 2023. doi:10.1101/2023.05.08.539824.
- [333] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Židek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, and D. Hassabis. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, Aug. 2021. doi:10.1038/s41586-021-03819-2.
- [334] M. A. Pak, K. A. Markhieva, M. S. Novikova, D. S. Petrov, I. S. Vorobyev, E. S. Maksimova, F. A. Kondrashov, and D. N. Ivankov. Using AlphaFold to predict the impact of single mutations on protein stability and function. *PLOS One*, 18(3):e0282689, Mar. 2023. doi:10.1371/journal.pone.0282689.
- [335] N. J. Fowler and M. P. Williamson. The accuracy of protein structures in solution determined by AlphaFold and NMR. *Structure*, 30(7):925–933.e2, July 2022. doi:10.1016/j.str.2022.04.005.
- [336] T. Saldaño, N. Escobedo, J. Marchetti, D. J. Zea, J. Mac Donagh, A. J. Velez Rueda, E. Gonik, A. García Melani, J. Novomisky Nechcoff, M. N. Salas, T. Peters, N. Demitroff, S. Fernandez Alberti, N. Palopoli, M. S. Fornasari, and G. Parisi. Impact of protein conformational diversity on AlphaFold predictions. *Bioinformatics*, 38(10):2742–2748, May 2022. doi:10.1093/bioinformatics/btac202.

- [337] T. J. Lane. Protein structure prediction has reached the single-structure frontier. *Nature Methods*, 20(2):170–173, Feb. 2023. doi:10.1038/s41592-022-01760-4.
- [338] A. R. Khokhlov and P. G. Khalatur. Conformation-Dependent Sequence Design (Engineering) of AB Copolymers. *Physical Review Letters*, 82(17):3456–3459, Apr. 1999. doi:10.1103/PhysRevLett.82.3456.
- [339] H. K. Murnen, A. R. Khokhlov, P. G. Khalatur, R. A. Segalman, and R. N. Zuckermann. Impact of Hydrophobic Sequence Patterning on the Coil-to-Globule Transition of Protein-like Polymers. *Macromolecules*, 45(12):5229–5236, June 2012. doi:10.1021/ma300707t.
- [340] B. Kuhlman, G. Dantas, G. C. Ireton, G. Varani, B. L. Stoddard, and D. Baker. Design of a Novel Globular Protein Fold with Atomic-Level Accuracy. *Science*, 302(5649):1364–1368, Nov. 2003. doi:10.1126/science.1089427.
- [341] B. Kuhlman and D. Baker. Native protein sequences are close to optimal for their structures. *Proceedings of the National Academy of Sciences*, 97(19):10383–10388, Sept. 2000. doi:10.1073/pnas.97.19.10383.
- [342] D. T. Jones. De novo protein design using pairwise potentials and a genetic algorithm: De novo protein design. *Protein Science*, 3(4):567–574, Apr. 1994. doi:10.1002/pro.5560030405.
- [343] C. Norn, B. I. M. Wicky, D. Juergens, S. Liu, D. Kim, D. Tischer, B. Koepnick, I. Anishchenko, Foldit Players, D. Baker, and S. Ovchinnikov. Protein sequence design by conformational landscape optimization. *Proceedings of the National Academy of Sciences*, 118(11):e2017228118, Mar. 2021. doi:10.1073/pnas.2017228118.
- [344] C. A. Goverde, B. Wolf, H. Khakzad, S. Rosset, and B. E. Correia. De novo protein design by inversion of the AlphaFold structure prediction network. *Protein Science*, 32(6):e4653, June 2023. doi:10.1002/pro.4653.
- [345] M. S. Shell. The relative entropy is fundamental to multiscale and inverse thermodynamic problems. *The Journal of Chemical Physics*, 129(14):144108, Oct. 2008. doi:10.1063/1.2992060.
- [346] A. Chaimovich and M. S. Shell. Relative entropy as a universal metric for multiscale errors. *Physical Review E*, 81(6):060104, June 2010. doi:10.1103/PhysRevE.81.060104.
- [347] M. S. Shell. Coarse-Graining with the Relative Entropy. In S. A. Rice and A. R. Dinner, editors, *Advances in Chemical Physics*, volume 161, pages 395–441. Wiley, 1st edition, Oct. 2016. doi:10.1002/9781119290971.ch5.

- [348] T. Sanyal and M. S. Shell. Coarse-grained models using local-density potentials optimized with the relative entropy: Application to implicit solvation. *The Journal of Chemical Physics*, 145(3):034109, July 2016. doi:10.1063/1.4958629.
- [349] A. Chaimovich and M. S. Shell. Anomalous waterlike behavior in spherically-symmetric water models optimized with the relative entropy. *Physical Chemistry Chemical Physics*, 11(12):1901–1915, 2009. doi:10.1039/b818512c.
- [350] L. Davis, editor. *Handbook of genetic algorithms*. International Thomson Computer Press, London, 1996.
- [351] Michael Rubinstein. *Polymer Physics*. Oxford University Press, Oxford, 2003.
- [352] E. R. McCarney, J. H. Werner, S. L. Bernstein, I. Ruczinski, D. E. Makarov, P. M. Goodwin, and K. W. Plaxco. Site-specific Dimensions Across a Highly Denatured Protein; A Single Molecule Study. *Journal of Molecular Biology*, 352(3):672–682, Sept. 2005. doi:10.1016/j.jmb.2005.07.015.
- [353] G. Fuertes, N. Banterle, K. M. Ruff, A. Chowdhury, D. Mercadante, C. Koehler, M. Kachala, G. Estrada Girona, S. Milles, A. Mishra, P. R. Onck, F. Gräter, S. Esteban-Martín, R. V. Pappu, D. I. Svergun, and E. A. Lemke. Decoupling of size and shape fluctuations in heteropolymeric sequences reconciles discrepancies in SAXS vs. FRET measurements. *Proceedings of the National Academy of Sciences*, 114(31):E6342–E6351, Aug. 2017. doi:10.1073/pnas.1704692114.
- [354] A. Kohlmeyer, J. Vermaas, and E. Braun. akohlmey/topotools: Release 1.9, Sept. 2022. doi:10.5281/ZENODO.598373.
- [355] B. Sanchez-Lengeling and A. Aspuru-Guzik. Inverse molecular design using machine learning: Generative models for matter engineering. *Science*, 361(6400):360–365, July 2018. doi:10.1126/science.aat2663.
- [356] J. Pfendtner and M. Bonomi. Efficient Sampling of High-Dimensional Free-Energy Landscapes with Parallel Bias Metadynamics. *Journal of Chemical Theory and Computation*, 11(11):5062–5067, Nov. 2015. doi:10.1021/acs.jctc.5b00846.
- [357] T. Ishiyama and A. Morita. Analysis of anisotropic local field in sum frequency generation spectroscopy with the charge response kernel water model. *The Journal of Chemical Physics*, 131(24):244714, Dec. 2009. doi:10.1063/1.3279126.
- [358] T. Ishiyama, D. Terada, and A. Morita. Hydrogen-Bonding Structure at Zwitterionic Lipid/Water Interface. *The Journal of Physical Chemistry Letters*, 7(2):216–220, Jan. 2016. doi:10.1021/acs.jpclett.5b02567.
- [359] I. Bakulin, N. Kondratyuk, A. Lankin, and G. Norman. Properties of aqueous 1,4-dioxane solution via molecular dynamics. *The Journal of Chemical Physics*, 155(15):154501, Oct. 2021. doi:10.1063/5.0059337.

- [360] D.A. Case, R.M. Betz, D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke, A.W. Goetz, N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, C. Lin, T. Luchko, R. Luo, B. Madej, D. Mermelstein, K.M. Merz, G. Monard, H. Nguyen, H.T. Nguyen, I. Omelyan, A. Onufriev, D.R. Roe, A. Roitberg, C. Sagui, C.L. Simmerling, W.M. Botello-Smith, J. Swails, R.C. Walker, J. Wang, R.M. Wolf, X. Wu, L. Xiao, and P.A. Kollman. AMBER 2016, 2016.
- [361] C. I. Bayly, P. Cieplak, W. Cornell, and P. A. Kollman. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *The Journal of Physical Chemistry*, 97(40):10269–10280, Oct. 1993. doi:10.1021/j100142a004.
- [362] C. J. Fennell, K. L. Wymer, and D. L. Mobley. A Fixed-Charge Model for Alcohol Polarization in the Condensed Phase, and Its Role in Small Molecule Hydration. *The Journal of Physical Chemistry B*, 118(24):6438–6446, June 2014. doi:10.1021/jp411529h.
- [363] S. L. Mayo, B. D. Olafson, and W. A. Goddard. DREIDING: a generic force field for molecular simulations. *The Journal of Physical Chemistry*, 94(26):8897–8909, Dec. 1990. doi:10.1021/j100389a010.
- [364] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, and D. J. Fox. Gaussian 16, Revision C.01, 2016. Gaussian Inc. Wallingford CT.
- [365] I. Vorobyov, V. M. Anisimov, S. Greene, R. M. Venable, A. Moser, R. W. Pastor, and A. D. MacKerell. Additive and Classical Drude Polarizable Force Fields for Linear and Cyclic Ethers. *Journal of Chemical Theory and Computation*, 3(3):1120–1133, May 2007. doi:10.1021/ct600350s.
- [366] E. Stendardo, A. Pedone, P. Cimino, M. Cristina Menziani, O. Crescenzi, and V. Barone. Extension of the AMBER force-field for the study of large nitroxides in condensed phases: an ab initio parameterization. *Physical Chemistry Chemical Physics*, 12(37):11697, 2010. doi:10.1039/c001481h.

- [367] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, and A. D. Mackerell. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of Computational Chemistry*, 31(4):665–896, July 2009. doi:[10.1002/jcc.21367](https://doi.org/10.1002/jcc.21367).
- [368] K. Vanommeslaeghe and A. D. MacKerell. Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing. *Journal of Chemical Information and Modeling*, 52(12):3144–3154, Dec. 2012. doi:[10.1021/ci300363c](https://doi.org/10.1021/ci300363c).
- [369] K. Vanommeslaeghe, E. P. Raman, and A. D. MacKerell. Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges. *Journal of Chemical Information and Modeling*, 52(12):3155–3168, Dec. 2012. doi:[10.1021/ci3003649](https://doi.org/10.1021/ci3003649).
- [370] A. D. Mackerell, M. Feig, and C. L. Brooks. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *Journal of Computational Chemistry*, 25(11):1400–1415, Aug. 2004. doi:[10.1002/jcc.20065](https://doi.org/10.1002/jcc.20065).
- [371] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma. The missing term in effective pair potentials. *The Journal of Physical Chemistry*, 91(24):6269–6271, Nov. 1987. doi:[10.1021/j100308a038](https://doi.org/10.1021/j100308a038).
- [372] A. Jakalian, D. B. Jack, and C. I. Bayly. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *Journal of Computational Chemistry*, 23(16):1623–1641, Oct. 2002. doi:[10.1002/jcc.10128](https://doi.org/10.1002/jcc.10128).
- [373] Z. Tan. Optimally Adjusted Mixture Sampling and Locally Weighted Histogram Analysis. *Journal of Computational and Graphical Statistics*, 26(1):54–65, Jan. 2017. doi:[10.1080/10618600.2015.1113975](https://doi.org/10.1080/10618600.2015.1113975).
- [374] H. Nguyen, D. R. Roe, and C. Simmerling. Improved Generalized Born Solvent Model Parameters for Protein Simulations. *Journal of Chemical Theory and Computation*, 9(4):2020–2034, Apr. 2013. doi:[10.1021/ct3010485](https://doi.org/10.1021/ct3010485).
- [375] D. van der Spoel and H. Berendsen. Molecular dynamics simulations of Leu-enkephalin in water and DMSO. *Biophysical Journal*, 72(5):2032–2041, May 1997. doi:[10.1016/S0006-3495\(97\)78847-7](https://doi.org/10.1016/S0006-3495(97)78847-7).
- [376] E. Levina and P. Bickel. The Earth Mover’s distance is the Mallows distance: some insights from statistics. In *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, volume 2, pages 251–256, Vancouver, BC, Canada, 2001. IEEE Comput. Soc. doi:[10.1109/ICCV.2001.937632](https://doi.org/10.1109/ICCV.2001.937632).

- [377] M. D. Rabenstein and Y. K. Shin. Determination of the distance between two spin labels attached to a macromolecule. *Proceedings of the National Academy of Sciences*, 92(18):8239–8243, Aug. 1995. doi:[10.1073/pnas.92.18.8239](https://doi.org/10.1073/pnas.92.18.8239).
- [378] J. A. Cooke and L. J. Brown. Distance Measurements by Continuous Wave EPR Spectroscopy to Monitor Protein Folding. In A. F. Hill, K. J. Barnham, S. P. Bottomley, and R. Cappai, editors, *Protein Folding, Misfolding, and Disease*, volume 752, pages 73–96. Humana Press, Totowa, NJ, 2011. doi:[10.1007/978-1-60327-223-0_6](https://doi.org/10.1007/978-1-60327-223-0_6). Series Title: Methods in Molecular Biology.
- [379] Y.-S. Liu, P. Sompornpisut, and E. Perozo. Structure of the KcsA channel intracellular gate in the open state. *Nature Structural Biology*, 8(10):883–887, Oct. 2001. doi:[10.1038/nsb1001-883](https://doi.org/10.1038/nsb1001-883).
- [380] M. Persson, J. R. Harbridge, P. Hammarström, R. Mitri, L.-G. Mårtensson, U. Carlsson, G. R. Eaton, and S. S. Eaton. Comparison of Electron Paramagnetic Resonance Methods to Determine Distances between Spin Labels on Human Carbonic Anhydrase II. *Biophysical Journal*, 80(6):2886–2897, June 2001. doi:[10.1016/S0006-3495\(01\)76254-6](https://doi.org/10.1016/S0006-3495(01)76254-6).
- [381] L. J. Brown, K. L. Sale, R. Hills, C. Rouviere, L. Song, X. Zhang, and P. G. Fajer. Structure of the inhibitory region of troponin by site directed spin labeling electron paramagnetic resonance. *Proceedings of the National Academy of Sciences*, 99(20):12765–12770, Oct. 2002. doi:[10.1073/pnas.202477399](https://doi.org/10.1073/pnas.202477399).
- [382] F. Zhang, Y. Chen, D.-H. Kweon, C. S. Kim, and Y.-K. Shin. The Four-helix Bundle of the Neuronal Target Membrane SNARE Complex Is Neither Disordered in the Middle nor Uncoiled at the C-terminal Region. *Journal of Biological Chemistry*, 277(27):24294–24298, July 2002. doi:[10.1074/jbc.M201200200](https://doi.org/10.1074/jbc.M201200200).
- [383] Y.-W. Chiang, T.-Y. Zheng, C.-J. Kao, and J.-C. Horng. Determination of Interspin Distance Distributions by cw-ESR Is a Single Linear Inverse Problem. *Biophysical Journal*, 97(3):930–936, Aug. 2009. doi:[10.1016/j.bpj.2009.05.030](https://doi.org/10.1016/j.bpj.2009.05.030).
- [384] Gunnar Jeschke. Introduction to electron paramagnetic resonance spectroscopy. URL https://ethz.ch/content/dam/ethz/special-interest/chab/physical-chemistry/epr-dam/documents/education/slides_EPR_2016.pdf.
- [385] D. Sezer, M. J. Prandolini, and T. F. Prisner. Dynamic nuclear polarization coupling factors calculated from molecular dynamics simulations of a nitroxide radical in water. *Physical Chemistry Chemical Physics*, 11(31):6626–6637, 2009. doi:[10.1039/b905709a](https://doi.org/10.1039/b905709a).
- [386] W. Kabsch. A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A*, 32(5):922–923, Sept. 1976. doi:[10.1107/S0567739476001873](https://doi.org/10.1107/S0567739476001873).

- [387] M. Heyden. Heterogeneity of water structure and dynamics at the protein-water interface. *The Journal of Chemical Physics*, 150(9):094701, Mar. 2019. doi:10.1063/1.5081739.
- [388] L. Martínez, R. Andrade, E. G. Birgin, and J. M. Martínez. PACKMOL: A package for building initial configurations for molecular dynamics simulations. *Journal of Computational Chemistry*, 30(13):2157–2164, Oct. 2009. doi:10.1002/jcc.21224.
- [389] D. Laage, G. Stirnemann, F. Sterpone, R. Rey, and J. T. Hynes. Reorientation and Allied Dynamics in Water and Aqueous Solutions. *Annual Review of Physical Chemistry*, 62(1):395–416, May 2011. doi:10.1146/annurev.physchem.012809.103503.
- [390] A. C. Fogarty and D. Laage. Water Dynamics in Protein Hydration Shells: The Molecular Origins of the Dynamical Perturbation. *The Journal of Physical Chemistry B*, 118(28):7715–7729, July 2014. doi:10.1021/jp409805p.
- [391] C. Klein, J. Sallai, T. J. Jones, C. R. Iacovella, C. McCabe, and P. T. Cummings. A Hierarchical, Component Based Approach to Screening Properties of Soft Matter. In R. Q. Snurr, C. S. Adjiman, and D. A. Kofke, editors, *Foundations of Molecular Modeling and Simulation*, pages 79–92. Springer Singapore, Singapore, 2016. doi:10.1007/978-981-10-1128-3_5. Series Title: Molecular Modeling and Simulation.
- [392] L. Schäfer. *Excluded Volume Effects in Polymer Solutions*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1999. doi:10.1007/978-3-642-60093-7.