UNIVERSITY OF CALIFORNIA SAN DIEGO

From Viruses to Sea Spray: Applications of All-Atom Molecular Dynamics Simulations to Environmental and Biological Aerosol Systems

A dissertation submitted in partial satisfaction of the

requirements for the degree

Doctor of Philosophy

in

Chemistry

by

Abigail Catherine Dommer

Committee in charge:

Professor Rommie E. Amaro, Chair
Professor Lihini Aluwihare
Professor Vicki H. Grassian
Professor Kimberly A. Prather
Professor Wei Wang

2021

The dissertation of Abigail Catherine Dommer is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2021

DEDICATION

To Maa

EPIGRAPH

"...Let us go," we said, "into the Sea of Cortez, realizing that we become forever a part of it; that our rubber boots slogging through a flat of eel-grass, that the rocks we turn over in a tide pool, make us truly and permanently a factor in the ecology of the region. We shall take something away from it, but we shall leave something too." And if we seem a small factor in a huge pattern, nevertheless it is of relative importance. We take a tiny colony of soft corals from a rock in a little water world. And that isn't terribly important to the tide pool. Fifty miles away the Japanese shrimp boats are dredging with overlapping scoops, bringing up tons of shrimps, rapidly destroying the species so that it may never come back, and with the species destroying the ecological balance of the whole region. That isn't very important in the world. And thousands of miles away the great bombs are falling and the stars are not moved thereby. None of it is important or all of it is.

John Steinbeck, *The Log from the Sea of Cortez* (1951)

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF SCHEMES

LIST OF TABLES

ACKNOWLEDGMENTS

I have been incredibly fortunate in my academic and personal journey which has led me to complete this body of work. Not only have countless teachers, professors, and mentors aided me with their time and expertise, but family, friends, and even serendipitous encounters with strangers have extended me encouragement, inspiration, grace, kindness, wisdom, and lessons, which continue to help me build a meaningful life. Completion of a PhD, like all things masquerading as purely academic endeavors, represents not just a mastery of analytical and technical skills and an extensive understanding of scientific literature in the field (among numerous other things); it is also a monument to deep personal growth through times often rife with uncertainty and isolation. Thus, I acknowledge not just my academic peers and mentors, but also those individuals and communities which have influenced and guided my personal and emotional development. I offer a few anecdotes with my acknowledgments, which my readers can feel free to ignore and skip to the end.

First, I acknowledge and thank the incredible teachers who served and molded me from childhood through adolescence: Mrs. McGlumphy (and the Thursday Thespians); Mrs. Kill (who told me I won "Best of Show" for my artwork in 2nd grade and I cried because it wasn't 1st place); Mr. Bennett (my AP Bio teacher, who incorrectly told us that electrons "disappear" at the end of the electron transport chain, inadvertently igniting my passion for research); Mr. D. (my AP Chemistry teacher who told us what *actually* happens to those electrons in the ETC); and Mrs. (Wilson) Porter (my high school math teacher to whom I owe many of my math comprehension skills).

I would like to thank my professors and mentors at Washington University in St. Louis who have similarly transformed my life: Mary Borgman (my freshman year drawing professor, a masterful artist and brilliant educator, whose lessons on portraiture profoundly changed me); Dr. Roshan Abraham (my wonderfully quirky Greek professor who helped us read the New Testament in the original Greek); Dr. Blankenship (my undergraduate research advisor, who opened a space in his lab for me to grow photosynthetic bugs); Dr. Sophia Hayes (whose classes thoroughly solidified my decision to pursue chemistry through point groups and MO-theory); Dr. Erica Majumder (my undergraduate research mentor, who taught, inspired, and encouraged me); and Dr. Steven Kinsley (my organic chemistry lab professor who believed in me harder than maybe anyone at that University).

I thank my countless mentors and colleagues at UCSD, SIO, and SDSC: Dr. Judy Kim, who advised me during my toughest year of graduate school *and* encouraged me to come back after I quit; Jeanine, Erica, Amy, Asmaa, and Jeff, in the Chem & Biochem Grad Student Services; Dr. Grant Deane, who co-advised me during my first year, introduced me to coding and statistics, and whose mentorship during that time still inspires me to produce rigorous, high-quality work *while* taking care of myself; Dr. Dale Stokes, a fantastic mentor who also taught me how to use tools; Dr. Andy Goetz, a brilliant scientist, mentor, and collaborator, always eager to review and improve my work; and Dr. Joe Patterson, who taught me well when MiniMART and I were still brand new. I also extend deepest gratitude to Dr. Jamie Schiffer, Dr. Jeff Wagner, Dr. Chris T. Lee, Dr. Bryn Taylor, Dr. Lorenzo Casalino, Dr. Fiona Kearns, Dr. Terra Sztain, Sasha Heyneman, Christian Seitz, Mia Rosenfeld, Clare Morris, and my wonderful undergrads, Andy, Donald, Aakash and Patiemma.

Thank you to all current and former CAICE members, especially Dr. Kimberly Carter-Fenk, Dr. Man Luo, and Mickey Rogers. Thank you to all current and former Amaro and McCammon lab members and to our *incredible* Gordon Bell ’20 and ’21 collaborators. Extra deep gratitude to John Stone and Jim Phillips, for helping us get set up on Blue Waters, Frontera, and Summit (which is just one sentence, but it doesn’t even remotely cover the amount of work they put in to help make our simulations happen). Super thank you to Bryan Hill and Kevin Smith for making our computers work.

I also thank: Laura Ryan (for her inspiring generosity of time, space, spirit, and love); Christine, Lydia, Margaret, David, and Melanie (for camaraderie); Kim Spearmon, Andrew Hillam, and Julie Choi Trepkau (for the gift of Ashtanga); and Maneli Jodat, Lise Marshall, LT Terris, and Joel Richey (for always holding space).

Special thank you to Nick Wauer, for being the best friend, colleague, officemate, confidant, and sounding board, for countless memes, puns, cat photos, conversations, and belly laughs. Thanks for joining CAICE and for being my cats’ favorite person besides me.

Thank you to Evan Kuester, for being my person for a little while and helping me fall in love with San Diego.

Thank you to Dr. Warren Chan, for being there from the beginning.

Deepest thank you to my amazing and supportive family, especially my parents and my stepfather. I thank our legion of strong women (Maa, Courtney, Aunt Tracey, Jordan, and Grandma Mary) who constantly inspire me to grow into the best version of myself. Thank you, Maa, for supporting me fiercely through every twist and turn and every rough patch, for believing in me harder than anyone else, and for being the goofiest, cat-lovingest, most generous derp of all time. That's right, I called my mother a derp in my dissertation. *finger guns*

I thank and acknowledge the members of my thesis committee, including Dr. Katja Lindenberg, Dr. Wei Wang, Dr. Vicki Grassian, and Dr. Lihini Aluwihare, who have provided valuable feedback, rooted for me, and kept watch over my progress throughout my graduate career.

Finally, I acknowledge our fearless leader Dr. Kim Prather and my incredible advisor Dr. Rommie Amaro. During our orientation week as first year grad students at UCSD in which we watched presentations from faculty looking for new students, Kim Prather gave a talk that changed the entire trajectory of my career. She described how she was able to "bring the ocean into the laboratory" to study sea spray aerosols, showing images of this massive wave channel and wave flume, located in this glorious warehouse space called the Hydraulics Lab. Back then, I thought I wanted to be an experimental physical chemist working on lasers and light harvesting as a basement gremlin, but this incredible talk encouraged me to email KP directly, in what I called a "wild card" move. I was entranced by the idea of studying the ocean and atmosphere and suddenly an entire world of possibilities broke open for me. A later, subsequent 20-minute conversation with her about Dobson's theory that SSA could have served as prebiotic cellular reactors in the origin of life story keeps me excited about my work to this day. Without Kim Prather's fierce dedication to the science, CAICE wouldn't exist, and I certainly would not be here.

After a battle with some health issues which led to me withdrawing from the program and taking two years away, I returned to UCSD with a renewed sense of self and a love of writing code (even longer story, I won't elaborate). Rommie Amaro, my amazing PhD advisor, welcomed me with open arms into her lab, where I and Jamie Schiffer launched the group's foray into CAICE research. I still remember the day I joined in 2017, before I'd even been admitted back into the grad program: I pointed to a 3D-printed model of the influenza virion on her desk and said, "I want to do this, but for sea spray aerosols," and she threw her hands up and said, "hired."

Rommie has fostered a group of some of the most talented, diverse, creative, and intelligent people I've ever met, and she is one of the greatest leaders I know. She advocates for us, believes in abundance, celebrates our accomplishments, and has allowed me to craft a PhD experience better than anything I could have dreamed up myself.

Much of my PhD was funded through NSF-CAICE grants (CHE-1305427 and CHE-1801971). I was supported for two years by the San Diego Fellowship (2018-2020) and in Winter 2021 by the UCSD Department of Chemistry and Biochemistry through the Distinguished Graduate Student Award. During the summer of 2020, I was funded through the Student Graduate Teaching Scholarship program. Additionally, I was supported by stipends via a 2018 NSF EAGER grant (NSF DUE-1645083) and the 2020 Bruno Zimm Award, funded by Larry Millstein.

Chapter 2, in full, is a modified reprint of the material as it appears in "Luo, M.[†] and Dommer, A. C.[†]; Schiffer, J. M.; Rez, D. J.; Mitchell, A. R.; Amaro, R. E.; Grassian, V. H. *Surfactant Charge Modulates Lipase-Embedded Monolayer Stability at Marine-Relevant Aerosol Surfaces*. Langmuir. 2019, 35 (27), 9050-9060." The dissertation author was a primary co-investigator and co-author of this work.

Chapter 3, in full, is a modified reprint of the material as it appears in "Carter-Fenk, K. A.; Dommer, A. C.; Fiamingo, M. E.; Kim, J.; Amaro, R. E.; Allen, H. C. *Calcium bridging drives polysaccharide co-adsorption to a proxy sea surface microlayer.* Phys. Chem. Chem. Phys. 2021, 23, 16401-16416." The dissertation author was a co-investigator and co-author of this work.

Chapter 4, in full, is a modified reprint of the material as it appears in "Lee, C.[†] and Dommer, A. C.[†]; Schiffer, J. M.; Amaro, R. E.; Grassian, V. H.; Prather, K. A. *Cation-driven lipopolysaccharide morphological changes impact nitric acid heterogeneous reactivity in sea spray aerosols*. J. Phys. Chem. Lett. 2021, 12 (20), 5023-5029." The dissertation author was a co-investigator and co-author of this work.

Chapter 5, in full, is in preparation for submission for publication and is presented in draft form containing mainly computational findings. It may appear as "Dommer, A. C.; Wauer, N. A.; Angle, K. J.; Luo, M.; Grassian, V. H.; Amaro, R. E. *All-Atom Molecular Dynamics Simulations Reveal Microemulsion-Like Phase Behavior in Model Submicron Marine Aerosols*. 2022." The dissertation author was the primary investigator and co-author of this work.

Chapter 6, in full, is a modified reprint of the material as it appears in "Casalino, L. †, Dommer, A. C. †, and Gaieb, Z. †; Barros, E. P.; Sztain, T.; Ahn, S.; Trifan, A.; Brace, A.; Ma, H.; Lee, H.; Bogetti, A.; Khalid, S.; Chong, L.; Simmerling, C.; Hardy, D. J.; Maia, J. D.; Phillips, J. C.; Kurth, T.; Stern, A.; Huang, L.; McCalpin, J.; Tatineni, M.; Gibbs, T.; Stone, J. E.; Jha, S.; Ramanathan, A.; Amaro, R. E. *AI-Driven Multiscale Simulations Illuminate Mechanisms of SARS-CoV-2 Spike Dynamics.* Int. J. High Perform. Comput. Appl. 2021, 35 (5), 432-451." The dissertation author was a primary co-investigator and co-author of this work. **Notably, this work won the 2020 ACM Gordon Bell Special Prize for COVID-19 Research.**

Chapter 7, in full, has been accepted for publication and is presented as it will appear in "Dommer, A., † Casalino, L.†, and Kearns, F.†; Rosenfeld, M.; Wauer, N.; Ahn, S-H.; Russo, J.; Oliveira, S.; Morris, C.; Bogetti, A.; Trifan, A.; Brace, A.; Sztain, T.; Clyde, A.; Ma, H.; Chennubhotla, C.; Lee, H.; Turilli, M.; Khalid, S.; Tamayo-Mendoza, T.; Welborn, M.; Christensen, A.; Smith, D.; Qiao, Z.; Sirumalla, S.; O'Connor, M.; Manby, F.; Anandkumar, A.; Hardy, D.; Phillips, J.; Stern, A.; Romero, J.; Clark, D.; Dorrell, M.; Maiden, T.; Huang, L.; McCalpin, J.; Woods, C.; Williams, M.; Barker, B.; Rajapaksha, H.; Pitts, R.; Gibbs, T.; Stone, J.; Zuckerman, D.; Mulholland, A.; Miller, T.; Jha, S.; Ramanathan, A.; Chong, L.; Amaro, R. *#COVIDisAirborne: AI-Enabled Multiscale Computational Microscopy of Delta SARS-CoV-2 in a Respiratory Aerosol.* Int. J. High Perform. Comput. Appl. 2022." The dissertation author was a primary co-investigator and co-author of this work. **Notably, this work was selected as a Finalist for the 2021 ACM Gordon Bell Special Prize for COVID-19 Research.**

VITA

2014	Bachelor of Arts, Washington University in St. Louis
2014 – 2015	Teaching Assistant, University of California San Diego
2015 – 2016	STEAM Teacher, Pacific Academy, Encinitas
2016	Master of Science, University of California San Diego
2016 – 2017	Adjunct Chemistry Professor, Palomar College
2016 – 2017	Adjunct Chemistry Professor, San Diego Mesa College
2016 – 2017	Teaching Assistant, University of California San Diego
2020	Instructional Assistant, University of California San Diego
2021	Doctor of Philosophy, University of California San Diego

PUBLICATIONS

Patterson, J. P.; Collins, D. B.; Michaud, J. M.; Axson, J. L.; Sultana, C. M., Moser, T.; **Dommer, A.** C.; Conner, J.; Grassian, V. H.; Stokes, M. D.; Deane, G. B.; Evans, J. E.; Burkart, M. D.; Prather, K. A.; Gianneschi, N. C. Sea Spray Aerosol Structure and Composition Using Cryogenic Transmission Electron Microscopy, *ACS Cent. Sci*. 2016, 2 (1), 40-47. DOI: 10.1021/acscentsci.5b0034

Stokes, M. D.; Deane, G.; Collins, D. B.; Cappa, C.; Bertram, T.; **Dommer, A.**; Schill, S.; Forestieri, S.; Survilo, M. A Miniature Marine Aerosol Reference Tank (MiniMART) as a Compact Breaking Wave Analogue, *Atmos. Meas. Tech*. 2016**,** 9 (9), 4257-4267. DOI: 10.5194/amt-2016-149

Schiffer, J. M.; Luo, M.; **Dommer, A. C**.; Thoron, G.; Pendergraft, M.; Santander, M. V.; Lucero, D.; De Barros, E. P.; Prather, K. A.; Grassian, V. H., Amaro, R. E. Impacts of Lipase Enzyme on the Surface Properties of Marine Aerosols. *J. Phys. Chem. Lett*. 2018, 9 (14), 3839-3849. DOI: 10.1021/acs.jpclett.8b0136

Amaro, R. E.; Ieong, P.; Huber, G., **Dommer, A**.; Alasdair, S.; Bush, R.; Durrant, J.; Votapka, L. A Computational Assay that Explores the Hemagglutinin/Neuraminidase Functional Balance Reveals the Neuraminidase Secondary Site as a Novel Anti-Influenza Target. *ACS Cent. Sci*. 2018, 4 (11), 1570-1577. DOI: 10.1021/acscentsci.8b00666

Luo, M.[†] and **Dommer, A. C**.[†]; Schiffer, J. M.; Rez, D. J.; Mitchell, A. R.; Amaro, R. E.; Grassian, V. H. Surfactant Charge Modulates Lipase-Embedded Monolayer Stability at Marine-Relevant Aerosol Surfaces. *Langmuir*. 2019, 35 (27), 9050-9060. DOI: 10.1021/acs.langmuir.9b00689

Durrant, J.[†] and Kochanek, S.[†]; Casalino, L.; Ieong, P.; **Dommer, A**.; Amaro, R. E. Mesoscale All-Atom Influenza Virus Simulations Suggest New Substrate Binding Mechanism. *ACS Cent. Sci*. 2020, 6 (2), 189-196. DOI: 10.1021/acscentsci.9b01071

Luo, M.; Wauer, N. A.; Angle, K. J.; **Dommer, A**. C; Song, M.; Novak, C. M.; Amaro, R. E.; Grassian, V. H. Insights into the Behavior of Nonanoic Acid and Its Conjugate Base at the Air/Water Interface Through a Combined Experimental and Theoretical Approach. *Chem. Sci.* 2020, 11, 10647-10656. DOI: 10.1039/D0SC02354J

Casalino, L. † and Gaieb, Z. †; Goldsmith, J. A.; Hjorth, C.; **Dommer, A. C**.; Harbison, A. M.; Fogarty, C. A.; Barros, E. P.; Taylor, B.C.; McLellan, J. S.; Fadda, E.; Amaro, R. E. Beyond Shielding: The Roles of Glycans in SARS-CoV-2 Spike Protein. *ACS Cent. Sci.* 2020, 6, (10), 1722-1734. DOI: 10.1021/acscentsci.0c01056

Barros, E. P.; Casalino, L.; Gaieb, Z.; **Dommer, A. C.**; Wang, Y.; Fallon, L.; Raguette, L.; Belfon, K.; Simmerling, C.; Amaro, R. E. The flexibility of ACE2 in the context of SARS-CoV-2 Infection. *Biophys. J.* 2020, 120 (6), 1072-1084. DOI: 10.1016/j.bpj.2020.10.036

Yu, A.; Pak, A. J.; He, P; Monje-Galvan, V.; Casalino, L.; Gaieb, Z.; **Dommer, A. C**.; Amaro, R. E.; Voth, G. A. A Multiscale Coarse-grained Model of the SARS-CoV-2 Virion. *Biophys. J.* 2020**,** 120 (6), 1097-1104. DOI: 10.1016/j.bpj.2020.10.048

Casalino, L. †, **Dommer, A. C.** †, and Gaieb, Z. †; Barros, E. P.; Sztain, T.; Ahn, S.; Trifan, A.; Brace, A.; Ma, H.; Lee, H.; Bogetti, A.; Khalid, S.; Chong, L.; Simmerling, C.; Hardy, D. J.; Maia, J. D.; Phillips, J. C.; Kurth, T.; Stern, A.; Huang, L.; McCalpin, J.; Tatineni, M.; Gibbs, T.; Stone, J. E.; Jha, S.; Ramanathan, A.; Amaro, R. E. AI-Driven Multiscale Simulations Illuminate Mechanisms of SARS-CoV-2 Spike Dynamics. *Int. J. High Perform. Comput. Appl.* 2021, 35 (5), 432-451. DOI:10.117/10943420211006452

Lee, C. † and **Dommer, A. C.** †; Schiffer, J. M.; Amaro, R. E.; Grassian, V. H.; Prather, K. A. Cation-driven lipopolysaccharide morphological changes impact nitric acid heterogeneous reactivity in sea spray aerosols. *J. Phys. Chem. Lett.* 2021, 12 (20), 5023-5029. DOI: 10.1021/acs.jpclett.1c00810

Carter-Fenk, K. A.; **Dommer, A. C**.; Fiamingo, M. E.; Kim, J.; Amaro, R. E.; Allen, H. C. Calcium bridging drives polysaccharide co-adsorption to a proxy sea surface microlayer. *Phys. Chem. Chem. Phys.* 2021, 23, 16401-16416. DOI: 10.1039/D1CP01407B

Dommer, A., † Casalino, L.†, and Kearns, F. †; Rosenfeld, M.; Wauer, N.; Ahn, S-H.; Russo, J.; Oliveira, S.; Morris, C.; Bogetti, A.; Trifan, A.; Brace, A.; Sztain, T.; Clyde, A.; Ma, H.; Chennubhotla, C.; Lee, H.; Turilli, M.; Khalid, S.; Tamayo-Mendoza, T.; Welborn, M.; Christensen, A.; Smith, D.; Qiao, Z.; Sirumalla, S.; O'Connor, M.; Manby, F.; Anandkumar, A.; Hardy, D.; Phillips, J.; Stern, A.; Romero, J.; Clark, D.; Dorrell, M.; Maiden, T.; Huang, L.; McCalpin, J.; Woods, C.; Williams, M.; Barker, B.; Rajapaksha, H.; Pitts, R.; Gibbs, T.; Stone, J.; Zuckerman, D.; Mulholland, A.; Miller, T.; Jha, S.; Ramanathan, A.; Chong, L.; Amaro, R. #COVIDisAirborne: AI-Enabled Multiscale Computational Microscopy of Delta SARS-CoV-2 in a Respiratory Aerosol. *Int. J. High Perform. Comput. Appl.* 2022. (accepted)

† Shared authorship.

FIELD OF STUDY

Major Field: Chemistry and Biochemistry

Studies in Computational and Theoretical Chemistry
Professor Rommie E. Amaro

ABSTRACT OF THE DISSERTATION

From Viruses to Sea Spray: Applications of All-Atom Molecular Dynamics Simulations to Environmental and Biological Systems

by

Abigail Catherine Dommer

Doctor of Philosophy in Chemistry

University of California San Diego, 2021

Professor Rommie E. Amaro, Chair

Aerosols are solid or liquid particles suspended in the air that can have far-reaching impacts on climate and human health. Aerosols impact climate through their radiative properties and their ability to seed cloud droplets or ice crystals. They also provide surfaces at which heterogeneous multiphase reactions can occur and serve as sinks for atmospheric sulfur, carbon and nitrogen. From a human health perspective, the physical and chemical properties of aerosols including their size, shape, and composition, can impact their transfer and deposition into the lungs. Smaller particles in particular can contain pollutants and pathogens and are able to travel deeper into the bronchioles to trigger irritation and infection. This body of work applies molecular dynamics simulations to understand aerosol systems, investigating their morphologies, impacts on climate, and ultimately their role in transporting the airborne SARS-CoV-2 virus. Molecular simulation and analysis methods are integrated with experiment to first probe surfactant interfaces with varying levels of chemical complexity, then to explore whole aerosol dynamics and phase within the context of understanding impacts of sea spray

aerosols (SSA) on climate. This work shows that 1) surfactant charge modulates the surface activity of *Burkholderia cepacia* lipase at lipid monolayer interfaces; 2) calcium enhances polysaccharide adsorption to fatty acid monolayers; and 3) divalent cations induce morphological changes in LPS-containing aerosols, hindering the reactive uptake of atmospheric nitric acid. This dissertation also describes methods for building large-scale, intact SSA models with full chemical complexity and shows how organic components distribute throughout the aerosol, suggesting that SSA may adopt microemulsion-like morphologies. Finally, a workflow is developed to build ultra-large systems for the study of airborne disease, demonstrating the successful construction and simulation of 1) the SARS-CoV-2 wild type envelope, and 2) a billion-atom respiratory aerosol containing the full breadth of chemical complexity, including the first all-atom model of the Delta SARS-CoV-2 envelope and never-before-modeled pulmonary mucins. The latter project presents the first atomic-level views of the SARS-CoV-2 virus within a respiratory aerosol and represents a novel approach to investigating the infection mechanisms of airborne pathogens.

Chapter 1
Introduction

1.1 Overview

This dissertation is organized into three parts. Part I, entitled "Building Up Chemical Complexity at The Air/Sea Interface," contains material from two publications[1,2] as well as ongoing work to understand interfacial dynamics and processes at the molecular level that impact the climate-relevant properties of marine aerosols. With a simple lipid monolayer as a baseline, we add new chemical components piece by piece to increase the interfacial complexity and then monitor the molecular conformational and dynamical response. In Part II, entitled "Sea Spray Aerosol Morphology and Dynamics," we build on elements in Part I to explore the impacts of complexity on whole aerosol properties. Whereas Part I isolates just the interface, Part II[3] expands this view to incorporate the intact aerosol, which includes an exploration of interfacial curvature, as well as the impacts of chemical complexity on physical elements such as shape, phase, viscosity, and molecular diffusion, which are then validated by experimental observations. Finally, Part III, entitled "Special Applications to Airborne Disease," applies the lessons learned from large-scale atomic modeling to building and analyzing the SARS-CoV-2[4] viral envelope in Chapter 6. Chapter 7 then describes the culmination of this body of work to model the SARS-CoV-2 virion in a respiratory aerosol, significantly advancing the state of the art of molecular virology.[5] To note: Chapters 6 and 7 are formatted as submissions to the ACM Gordon Bell Special Prize for COVID-19 Research, corresponding to the 2020 and 2021 awards cycles. Chapter 6 was the winning paper for the 2020 Gordon Bell Prize; Chapter 7 was a finalist for the Prize in 2021.

This body of work breaks new ground on multiple fronts. First, all-atom molecular dynamics simulations containing multiple millions of atoms are relatively untenable without the use of leadership-class petascale computing facilities. The systems described in Chapters 5, 6, and 7 contain each 3 million, 300 million, and 1 billion atoms, respectively. Reaching scientifically relevant timescales with systems of these sizes requires the use of the most powerful supercomputing centers available; chiefly, UIUC Blue Waters, TACC Frontera,

and ORNL Summit were employed in these works through generous NSF-funded and/or director's discretionary allocations. This work would simply not have been feasible *via* computing resources accessed through traditional routes, i.e., through the XSEDE[6] allocation program. Further development of the code in the NAMD molecular dynamics engine was required to achieve optimal performance on these systems, which involved close collaboration with the NAMD developers.[7–10] These simulations represent some of the first all-atom, explicit-solvent simulations to run at these ultra-large scales.

Secondly, experimental techniques and instrumentation for atmospheric environmental applications have advanced less rapidly compared to biological and medical applications. For example, we are not only able to generate and image cellular membranes, bilayers, micelles, and other such samples in solution, but we can also probe their properties using various methods. Conversely, effectively generating and collecting aerosol samples as well as imaging them at the single-particle level is still an active area of research and development. The modeling methods used in this dissertation can be considered a "computational microscope,"[11–14] serving to complement, augment, and build upon the current capacity of experiments to understand and characterize atmospheric aerosols. All-atom molecular dynamics simulations can visualize a system and its dynamics at the atomic level with nanosecond to microsecond timescale resolution, giving us information that cannot be attained using experiments alone. Thus, computation can significantly enhance experimental investigations and inform upon and drive hypothesis development. This is not to say, however, that computational methods should be used alone.

This work is especially forward-thinking in that it harnesses the power of interdisciplinary and collaborative science. Just as experiments can be enhanced by theory, so too can theory be enhanced by experimental observation. By weaving together experiment and computation, the work becomes more than the sum of its parts. An experiment can give us bulk, macroscopic properties that describe, for example, the extent of nitric acid reactive uptake by aerosols containing LPS-Ca^{2+} or LPS-Na^{+} mixtures. Although it is an important finding alone, it is much more powerful when integrating the computational observation that Ca^{2+} cations induce dramatic structural changes at the molecular level that dramatically reduce the diffusion of water throughout the particle.[3] Not only is the complementary finding more complete in that it provides both bulk observation with atomic-level, theoretically rigorous calculations, but the time-to-solution is largely cut in half. By collaborating

across disciplines to solve a problem from all perspectives rather than publishing each story independently, we reduce the time that data is withheld from the community and accelerate scientific progress.

With that said, the chapters in this dissertation represent highly collaborative studies. None of the chapters given here describes work completed in isolation. The dissertation author claims the computational components, the discussions that drive the experimental hypotheses, and the development and writing of the manuscript. Experimental findings are given here not only because they complete the scientific story, but also because they validate and support the theoretical findings. Specific author contributions are detailed in the Acknowledgments sections of each chapter.

The remainder of this Introduction provides a brief overview of aerosol systems as well as the molecular dynamics (MD) simulation method.

1.2 Aerosols

Aerosols are microscopic solid, liquid, or semi-solid particles of natural or anthropogenic origin that are suspended in the atmosphere and, in some cases, can be picked up by wind currents and dispersed across the globe.[15–19] Aerosols are categorized by their origins and chemical properties, but are further typed according to size, which is typically dependent upon their generation mechanisms. Small, submicron aerosols can have long atmospheric lifetimes, while larger, supermicron ones settle to the ground. Of the many aerosol types, the two types discussed in this dissertation are sea spray and respiratory aerosols.

Sea spray aerosols (SSA) are generated through bubble bursting mechanisms at the ocean surface.[20] They make up the largest portion of natural atmospheric aerosols by mass and their cloud-interaction properties represent the greatest source of uncertainty in radiative forcing models.[21–24] SSA can influence climate directly by interacting with incoming solar radiation, or indirectly by serving as surfaces for multiphase heterogeneous atmospheric chemistry.[21,24–31] Submicron SSA are generally produced by the bursting of a film cap at the sea surface and largely contain aliphatic organic material.[20,32] Numerous experimental studies have been able to identify the major classes of organic molecules found in SSA, which include surfactants, alkanes, protein, saccharides, and other biogenic macromolecules.[22,29,32–36] In terms of inorganic species, it has been demonstrated that cations such as sodium, magnesium, and calcium, are enriched at the sea surface, which may enhance their

transfer into SSA.[34,37–39] The work presented here will focus specifically on building up complexity of marine-relevant systems, starting with simple fatty acid and phospholipid surfactants, then adding more complex species like protein and oligosaccharides. The impact of calcium on these systems is also explored in this work.

Respiratory aerosols (RA) are particles generated in the lungs that are exhaled during breathing, speaking, singing, and coughing.[15,16,40–42] The recent COVID-19 pandemic has brought attention to the production and airborne dispersion of RA as a significant mode of infection between individuals, prompting our interest in studying the SARS-CoV-2 virus in an RA environment.[5,16,40] RA is known to contain, like SSA, a variety of organic material, from lung surfactants which include fatty acids, phospholipids and cholesterol, to proteins and glycoproteins like albumin, immunoglobulins and mucus.[43,44] If an individual is infected with a respiratory virus, such as influenza or SARS-CoV-2, the aerosols emitted contain the virus as well.[45–47] According to recent literature, submicron aerosols, such as those emitted by breathing, are most likely to contain active virus.[45] There is also growing evidence that increased viral affinity for pulmonary mucus may provide a possible explanation for the increased airborne transmissibility and lifetime of Delta SARS-CoV-2 compared to the wild type, and that this mechanism may similarly apply to other respiratory pathogens.[48] Part III of this dissertation details the construction and simulation of SARS-CoV-2 in a RA environment.

1.3 All-Atom Molecular Dynamics Simulations

Molecular dynamics (MD) is a numerical method that solves the classical equations of motion to generate trajectories of atom positions, first demonstrated in the 1950's by Alder & Wainwright.[49] When evaluated over enough timesteps, we can effectively run *in silico*, statistically rigorous chemistry experiments at microsecond timescales without stepping foot in a wet lab.

For a given system of N interacting particles (e.g., atoms), we calculate how those particles will move over a specific timestep Δt, and then continue to perform this calculation over many timesteps. We call upon classical Newtonian mechanics to calculate the next position of an atom with mass m from acceleration, a:

$$m_i a_i = m_i \frac{d^2 \boldsymbol{r_i}}{dt^2} = \boldsymbol{F}_i \quad i = 1 \ldots N, \tag{1.1}$$

where $\mathbf{r}_i(t)$ and $\mathbf{F}_i(t)$ are the position and force, respectively, of atom i at time t. The force on the atom, calculated from all the interactions of the atom in the system, can be defined by the gradient (∇) of the potential energy function V,

$$\boldsymbol{F}_i = -\nabla_{\boldsymbol{r}_i} V(\boldsymbol{r}_i, \ldots, \boldsymbol{r}_N), \tag{1.2}$$

which is determined by all the initial atomic positions in the system. The negative sign in this equation indicates that the integration will attempt to minimize the total energy of the system. In other words, the force acting upon the atoms will cause them to move to new positions that are lower in energy than before, much like a ball rolling down a hill minimizes its potential energy.

The potential energy function can be evaluated for a system *via* the summation of all the individual potentials of each atom, which are in turn, summations of all the energetic contributions from each atomic interaction with surrounding atoms. Evaluating this function accurately for all atoms in a system can be extremely computationally expensive and slow, so approximations have been developed and tested that are able to balance the computational cost of calculation and achieve reasonable accuracy.

The energetic contributions to the potential are divided into bonded and non-bonded interactions. Bonded interactions are those that arise from atoms connected to one another through chemical bonds and are dictated by the bond type (approximated to a harmonic spring), angle, and rotation. Non-bonded interactions are those arising between pairs of unconnected atoms that are near one another in space and are defined by attractive and repulsive components (Lennard-Jones), as well as electrostatics (Coulombic). **Table 1.1** outlines each of these interactions and the numerical approximation used to calculate their energies. A detailed description of these terms and their derivations can be found elsewhere.[50]

Table 1.1: Bonded and non-bonded energetic contributions to the potential energy function.

Potential Term	Equation	Definitions	Eq.
$V_{bond}(r_{ij})$	$\frac{k_{ij}^{b}}{2}\left(r_{ij}-r_{ij}^{0}\right)^{2}$	k_{ij}^{b} stiffness force constant r_{ij} distance between atoms r_{ij}^{0} equilibrium bond length	1.3
$V_{angle}(\vartheta_{ijk})$	$\frac{k_{ijk}^{\vartheta}}{2}\left(\vartheta_{ijk}-\vartheta_{ijk}^{0}\right)^{2}$	k_{ijk}^{ϑ} angle distortion force constant ϑ_{ijk} angle ϑ_{ijk}^{0} equilibrium bond angle	1.4
$V_{dihedral}(\phi_{ijkl})$	$k_{jk}^{\phi}[1+\cos(n\phi-\phi_{0})]$	ϕ dihedral angle ϕ_0 equilibrium dihedral angle n number of minima over full rotation	1.5
$V_{Lennard-Jones}(r_{ij})$	$\frac{4\varepsilon\sigma^{12}}{r_{ij}^{12}}-\frac{4\varepsilon\sigma^{6}}{r_{ij}^{6}}$	ε dispersion energy σ distance where V=0, ~'atom size'	1.6
$V_{Coulomb}(r_{ij})$	$\frac{q_i q_j}{4\pi\epsilon_0\epsilon_\tau r_{ij}}$	q particle charge ϵ_0 permittivity of free space ϵ_τ relative permittivity	1.7

The collection of constants associated with each bond, angle, rotation, and atom pair, and the summation of the parameters for a simulation, make up the *force field*. These values are often, preferably, derived from experimental observation. In many cases, however, experimental data is not available, or there are additional features of the system that are necessary to reproduce more precisely (as in the case of water) such that the approximations given by the potential terms are inadequate to describe the behavior of the whole system. Critical limitations like these exist and are kept in mind when carefully evaluating MD simulations. Force field development to address issues such as these in detail is out of the scope of this work, but the dissertation author includes brief discussions of force field selection in the methods of the following chapters.

Expanding upon this, ensemble properties of a system, such as temperature and pressure, are carefully controlled in MD simulations such that the system proceeds in agreement with statistical mechanics. Temperature and pressure coupling methods are continually being tested and developed for various systems in order to rigorously maintain the statistical mechanical ensemble.[51–53] Ultimately, these techniques allow us to extract many qualitative macroscale observables from our microscale simulations with reasonable accuracy.

Quantitative comparison to experimental observations remains a challenge, however, due to the many approximations made in the force field and throughout the MD integration. All comparisons to experiment made in this dissertation are done so considering these limitations in the discussion.

References

(1) Luo, M.; Dommer, A. C.; Schiffer, J. M.; Rez, D. J.; Mitchell, A. R.; Amaro, R. E.; Grassian, V. H. Surfactant Charge Modulates Structure and Stability of Lipase-Embedded Monolayers at Marine-Relevant Aerosol Surfaces. *Langmuir* **2019**, *35* (27), 9050–9060. https://doi.org/10.1021/acs.langmuir.9b00689.

(2) Carter-Fenk, K. A.; Dommer, A. C.; Fiamingo, M. E.; Kim, J.; Amaro, R. E.; Allen, H. C. Calcium Bridging Drives Polysaccharide Co-Adsorption to a Proxy Sea Surface Microlayer. *Phys. Chem. Chem. Phys.* **2021**, *23* (30), 16401–16416. https://doi.org/10.1039/D1CP01407B.

(3) Lee, C.; Dommer, A. C.; Schiffer, J. M.; Amaro, R. E.; Grassian, V. H.; Prather, K. A. Cation-Driven Lipopolysaccharide Morphological Changes Impact Heterogeneous Reactions of Nitric Acid with Sea Spray Aerosol Particles. *The Journal of Physical Chemistry Letters* **2021**, *12* (20), 5023–5029. https://doi.org/10.1021/acs.jpclett.1c00810.

(4) Casalino, L.; Dommer, A. C.; Gaieb, Z.; Barros, E. P.; Sztain, T.; Ahn, S.-H.; Trifan, A.; Brace, A.; Bogetti, A. T.; Clyde, A.; Ma, H.; Lee, H.; Turilli, M.; Khalid, S.; Chong, L. T.; Simmerling, C.; Hardy, D. J.; Maia, J. D.; Phillips, J. C.; Kurth, T.; Stern, A. C.; Huang, L.; McCalpin, J. D.; Tatineni, M.; Gibbs, T.; Stone, J. E.; Jha, S.; Ramanathan, A.; Amaro, R. E. AI-Driven Multiscale Simulations Illuminate Mechanisms of SARS-CoV-2 Spike Dynamics. *The International Journal of High Performance Computing Applications* **2021**, 109434202110064. https://doi.org/10.1177/10943420211006452.

(5) Dommer, A.; Casalino, L.; Kearns, F.; Rosenfeld, M.; Wauer, N.; Ahn, S.-H.; Russo, J.; Oliveira, S.; Morris, C.; Bogetti, A.; Trifan, A.; Brace, A.; Sztain, T.; Clyde, A.; Ma, H.; Chennubhotla, C.; Lee, H.; Turilli, M.; Khalid, S.; Tamayo-Mendoza, T.; Welborn, M.; Christiansen, A.; Smith, D. G. A.; Qiao, Z.; Sirumalla, S. K.; O'Connor, M.; Manby, F.; Anandkumar, A.; Hardy, D.; Phillips, J.; Stern, A.; Romero, J.; Clark, D.; Dorrell, M.; Maiden, T.; Huang, L.; McCalpin, J.; Woods, C.; Gray, A.; Williams, M.; Barker, B.; Rajapaksha, H.; Pitts, R.; Gibbs, T.; Stone, J.; Zuckerman, D.; Muholland, A.; Miller, T.; Jha, S.; Ramanathan, A.; Chong, L.; Amaro, R. E. *#COVIDisAirborne: AI-Enabled Multiscale Computational Microscopy of Delta SARS-CoV-2 in a Respiratory Aerosol*; 2021; p 2021.11.12.468428. https://doi.org/10.1101/2021.11.12.468428.

(6) Towns, J.; Cockerill, T.; Dahan, M.; Foster, I.; Gaither, K.; Grimshaw, A.; Hazlewood, V.; Lathrop, S.; Lifka, D.; Peterson, G. D.; Roskies, R.; Scott, J. R.; Wilkens-Diehr, N. XSEDE: Accelerating Scientific Discovery. *Computing in Science & Engineering* **2014**, *16* (5), 62–74. https://doi.org/10.1109/MCSE.2014.80.

(7) Acun, B.; Hardy, D. J.; Kale, L.; Li, K.; Phillips, J. C.; Stone, J. E. Scalable Molecular Dynamics with NAMD on the Summit System. *IBM\ J.\ Res.\ Dev.* **2019**, *62*, 4:1-4:9.

(8) Phillips, J. C.; Hardy, D. J.; Maia, J. D. C.; Stone, J. E.; Ribeiro, J. V.; Bernardi, R. C.; Buch, R.; Fiorin, G.; Hénin, J.; Jiang, W.; McGreevy, R.; Melo, M. C. R.; Radak, B.; Skeel, R. D.; Singharoy, A.; Wang, Y.; Roux, B.; Aksimentiev, A.; Luthey-Schulten, Z.; Kalé, L. V.; Schulten, K.; Chipot, C.; Tajkhorshid,

E. Scalable Molecular Dynamics on CPU and GPU Architectures with NAMD. *J. Chem.\ Phys.* **2020**, *153*, 044130. https://doi.org/10.1063/5.0014475.

(9) Phillips, J.; Zheng, G.; Kumar, S.; Kale, L. NAMD: Biomolecular Simulation on Thousands of Processors. In *Proceedings of the IEEE/ACM SC2002 Conference, Technical Paper 277*; IEEE Press: Baltimore, Maryland, 2002; pp 1–18.

(10) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *Journal of Computational Chemistry*. 2005. https://doi.org/10.1002/jcc.20289.

(11) Dror, R. O.; Dirks, R. M.; Grossman, J. P.; Xu, H.; Shaw, D. E. Biomolecular Simulation: A Computational Microscope for Molecular Biology. *Annu Rev Biophys* **2012**, *41*, 429–452. https://doi.org/10.1146/annurev-biophys-042910-155245.

(12) Lee, E. H.; Hsin, J.; Sotomayor, M.; Comellas, G.; Schulten, K. Discovery through the Computational Microscope. *Structure* **2009**, *17*, 1295–1306.

(13) Malmstrom, R. D.; Kornev, A. P.; Taylor, S. S.; Amaro, R. E. Allostery through the Computational Microscope: CAMP Activation of a Canonical Signalling Domain. *Nat Commun* **2015**, *6* (1), 7588. https://doi.org/10.1038/ncomms8588.

(14) Mehandzhiyski, A. Y.; Zozoulenko, I. Computational Microscopy of PEDOT:PSS/Cellulose Composite Paper. *ACS Appl. Energy Mater.* **2019**, *2* (5), 3568–3577. https://doi.org/10.1021/acsaem.9b00307.

(15) Wang, C. C.; Prather, K. A.; Sznitman, J.; Jimenez, J. L.; Lakdawala, S. S.; Tufekci, Z.; Marr, L. C. Airborne Transmission of Respiratory Viruses. *Science 373* (6558), eabd9149. https://doi.org/10.1126/science.abd9149.

(16) Miller, S. L.; Nazaroff, W. W.; Jimenez, J. L.; Boerstra, A.; Buonanno, G.; Dancer, S. J.; Kurnitski, J.; Marr, L. C.; Morawska, L.; Noakes, C. Transmission of SARS-CoV-2 by Inhalation of Respiratory Aerosol in the Skagit Valley Chorale Superspreading Event. *Indoor Air* **2021**, *31* (2), 314–323. https://doi.org/10.1111/ina.12751.

(17) Hodzic, A.; Campuzano-Jost, P.; Bian, H.; Chin, M.; Colarco, P. R.; Day, D. A.; Froyd, K. D.; Heinold, B.; Jo, D. S.; Katich, J. M.; Kodros, J. K.; Nault, B. A.; Pierce, J. R.; Ray, E.; Schacht, J.; Schill, G. P.; Schroder, J. C.; Schwarz, J. P.; Sueper, D. T.; Tegen, I.; Tilmes, S.; Tsigaridis, K.; Yu, P.; Jimenez, J. L. Characterization of Organic Aerosol across the Global Remote Troposphere: A Comparison of ATom Measurements and Global Chemistry Models. *Atmospheric Chemistry and Physics* **2020**, *20* (8), 4607–4635. https://doi.org/10.5194/acp-20-4607-2020.

(18) Chakraborty, S.; Guan, B.; Waliser, D. E.; da Silva, A. M.; Uluatam, S.; Hess, P. Extending the Atmospheric River Concept to Aerosols: Climate and Air Quality Impacts. *Geophysical Research Letters* **2021**, *48* (9), e2020GL091827. https://doi.org/10.1029/2020GL091827.

(19) Eck, T. F.; Holben, B. N.; Dubovik, O.; Smirnov, A.; Goloub, P.; Chen, H. B.; Chatenet, B.; Gomes, L.; Zhang, X.-Y.; Tsay, S.-C.; Ji, Q.; Giles, D.; Slutsker, I. Columnar Aerosol Optical Properties at AERONET Sites in Central Eastern Asia and Aerosol Transport to the Tropical Mid-Pacific. *Journal of Geophysical Research: Atmospheres* **2005**, *110* (D6). https://doi.org/10.1029/2004JD005274.

(20) Wang, X.; Deane, G. B.; Moore, K. A.; Ryder, O. S.; Stokes, M. D.; Beall, C. M.; Collins, D. B.; Santander, M. V.; Burrows, S. M.; Sultana, C. M.; Prather, K. A. The Role of Jet and Film Drops in Controlling the Mixing State of Submicron Sea Spray Aerosol Particles. *Proceedings of the National Academy of Sciences of the United States of America* **2017**, *114* (27), 6978–6983. https://doi.org/10.1073/pnas.1702420114.

(21) Cochran, R. E.; Ryder, O. S.; Grassian, V. H.; Prather, K. A. Sea Spray Aerosol: The Chemical Link between the Oceans, Atmosphere, and Climate. *Accounts of Chemical Research* **2017**, *50* (3), 599–604. https://doi.org/10.1021/acs.accounts.6b00603.

(22) Wang, X.; Sultana, C. M.; Trueblood, J.; Hill, T. C. J.; Malfatti, F.; Lee, C.; Laskina, O.; Moore, K. A.; Beall, C. M.; McCluskey, C. S.; Cornwell, G. C.; Zhou, Y.; Cox, J. L.; Pendergraft, M. A.; Santander, M. V.; Bertram, T. H.; Cappa, C. D.; Azam, F.; DeMott, P. J.; Grassian, V. H.; Prather, K. A. Microbial Control of Sea Spray Aerosol Composition: A Tale of Two Blooms. *ACS Central Science* **2015**, *1* (3), 124–131. https://doi.org/10.1021/acscentsci.5b00148.

(23) Prather, K. A.; Bertram, T. H.; Grassian, V. H.; Deane, G. B.; Stokes, M. D.; DeMott, P. J.; Aluwihare, L. I.; Palenik, B. P.; Azam, F.; Seinfeld, J. H.; Moffet, R. C.; Molina, M. J.; Cappa, C. D.; Geiger, F. M.; Roberts, G. C.; Russell, L. M.; Ault, A. P.; Baltrusaitis, J.; Collins, D. B.; Corrigan, C. E.; Cuadra-Rodriguez, L. A.; Ebben, C. J.; Forestieri, S. D.; Guasco, T. L.; Hersey, S. P.; Kim, M. J.; Lambert, W. F.; Modini, R. L.; Mui, W.; Pedler, B. E.; Ruppel, M. J.; Ryder, O. S.; Schoepp, N. G.; Sullivan, R. C.; Zhao, D. Bringing the Ocean into the Laboratory to Probe the Chemical Complexity of Sea Spray Aerosol. *Proceedings of the National Academy of Sciences* **2013**, *110* (19), 7550–7555. https://doi.org/10.1073/PNAS.1300262110.

(24) Quinn, P. K.; Collins, D. B.; Grassian, V. H.; Prather, K. A.; Bates, T. S. Chemistry and Related Properties of Freshly Emitted Sea Spray Aerosol. *Chemical Reviews* **2015**, *115* (10), 4383–4399. https://doi.org/10.1021/cr500713g.

(25) Forestieri, S. D.; Staudt, S. M.; Kuborn, T. M.; Faber, K.; Ruehl, C. R.; Bertram, T. H.; Cappa, C. D. Establishing the Impact of Model Surfactants on Cloud Condensation Nuclei Activity of Sea Spray Aerosols. *Atmospheric Chemistry and Physics Discussions* **2018**, 1–44. https://doi.org/10.5194/acp-2018-207.

(26) DeMott, P. J.; Mason, R. H.; McCluskey, C. S.; Hill, T. C. J.; Perkins, R. J.; Desyaterik, Y.; Bertram, A. K.; Trueblood, J. V.; Grassian, V. H.; Qiu, Y.; Molinero, V.; Tobo, Y.; Sultana, C. M.; Lee, C.; Prather, K. A. Ice Nucleation by Particles Containing Long-Chain Fatty Acids of Relevance to Freezing by Sea Spray Aerosols. *Environmental Science: Processes and Impacts* **2018**, *20* (11), 1559–1569. https://doi.org/10.1039/c8em00386f.

(27) Mccluskey, C. S.; Hill, T. C. J.; Sultana, C. M.; Laskina, O.; Trueblood, J.; Santander, M. V.; Beall, C. M.; Michaud, J. M.; Kreidenweis, S. M.; Prather, K. A.; Grassian, V.; Demott, P. J. A Mesocosm Double Feature: Insights into the Chemical Makeup of Marine Ice Nucleating Particles. *Journal of the Atmospheric Sciences* **2018**, *75* (7), 2405–2423.

(28) Ault, A. P.; Guasco, T. L.; Baltrusaitis, J.; Ryder, O. S.; Trueblood, J. V.; Collins, D. B.; Ruppel, M. J.; Cuadra-Rodriguez, L. A.; Prather, K. A.; Grassian, V. H. Heterogeneous Reactivity of Nitric Acid with Nascent Sea Spray Aerosol: Large Differences Observed between and within Individual Particles. *Journal of Physical Chemistry Letters* **2014**, *5* (15), 2493–2500. https://doi.org/10.1021/jz5008802.

(29) Cochran, R. E.; Laskina, O.; Trueblood, J. V.; Estillore, A. D.; Morris, H. S.; Jayarathne, T.; Sultana, C. M.; Lee, C.; Lin, P.; Laskin, J.; Laskin, A.; Dowling, J. A.; Qin, Z.; Cappa, C. D.; Bertram, T. H.;

Tivanski, A. V.; Stone, E. A.; Prather, K. A.; Grassian, V. H. Molecular Diversity of Sea Spray Aerosol Particles: Impact of Ocean Biology on Particle Composition and Hygroscopicity. *Chem* **2017**, *2* (5), 655–667. https://doi.org/10.1016/j.chempr.2017.03.007.

(30) Ryder, O. S.; Campbell, N. R.; Morris, H.; Forestieri, S.; Ruppel, M. J.; Cappa, C.; Tivanski, A.; Prather, K.; Bertram, T. H. Role of Organic Coatings in Regulating N2O5 Reactive Uptake to Sea Spray Aerosol. *Journal of Physical Chemistry A* **2015**, *119* (48), 11683–11692. https://doi.org/10.1021/acs.jpca.5b08892.

(31) Shrestha, M.; Luo, M.; Li, Y.; Xiang, B.; Xiong, W.; Grassian, V. H. Let There Be Light: Stability of Palmitic Acid Monolayers at the Air/Salt Water Interface in the Presence and Absence of Simulated Solar Light and a Photosensitizer. *Chemical Science* **2018**, *9* (26), 5716–5723. https://doi.org/10.1039/C8SC01957F.

(32) Bertram, T. H.; Cochran, R. E.; Grassian, V. H.; Stone, E. A. Sea Spray Aerosol Chemical Composition: Elemental and Molecular Mimics for Laboratory Studies of Heterogeneous and Multiphase Reactions. *Chemical Society Reviews* **2018**, *47* (7), 2374–2400. https://doi.org/10.1039/C7CS00008A.

(33) Cochran, R. E.; Laskina, O.; Jayarathne, T.; Laskin, A.; Laskin, J.; Lin, P.; Sultana, C.; Lee, C.; Moore, K. A.; Cappa, C. D.; Bertram, T. H.; Prather, K. A.; Grassian, V. H.; Stone, E. A. Analysis of Organic Anionic Surfactants in Fine and Coarse Fractions of Freshly Emitted Sea Spray Aerosol. *Environmental Science and Technology* **2016**, *50* (5), 2477–2486. https://doi.org/10.1021/acs.est.5b04053.

(34) Jayarathne, T.; Sultana, C. M.; Lee, C.; Malfatti, F.; Cox, J. L.; Pendergraft, M. A.; Moore, K. A.; Azam, F.; Tivanski, A. V.; Cappa, C. D.; Bertram, T. H.; Grassian, V. H.; Prather, K. A.; Stone, E. A. Enrichment of Saccharides and Divalent Cations in Sea Spray Aerosol during Two Phytoplankton Blooms. *Environmental Science and Technology* **2016**. https://doi.org/10.1021/acs.est.6b02988.

(35) Cochran, R. E.; Jayarathne, T.; Stone, E. A.; Grassian, V. H. Selectivity Across the Interface: A Test of Surface Activity in the Composition of Organic-Enriched Aerosols from Bubble Bursting. *Journal of Physical Chemistry Letters* **2016**, *7* (9), 1692–1696. https://doi.org/10.1021/acs.jpclett.6b00489.

(36) Rastelli, E.; Corinaldesi, C.; Dell'anno, A.; Lo Martire, M.; Greco, S.; Facchini, M. C.; Rinaldi, M.; O'dowd, C.; Ceburnis, D.; Danovaro, R. Transfer of Labile Organic Matter and Microbes from the Ocean Surface to the Marine Aerosol: An Experimental Approach OPEN. https://doi.org/10.1038/s41598-017-10563-z.

(37) Salter, M. E.; Hamacher-Barth, E.; Leck, C.; Werner, J.; Johnson, C. M.; Riipinen, I.; Nilsson, E. D.; Zieger, P. Calcium Enrichment in Sea Spray Aerosol Particles. *Geophysical Research Letters* **2016**, *43* (15), 8277–8285. https://doi.org/10.1002/2016GL070275.

(38) Xu, G.; Chen, L.; Xu, T.; He, S.; Gao, Y. Distributions of Water-Soluble Ions in Size-Aggregated Aerosols over the Southern Ocean and Coastal Antarctica. *Environ. Sci.: Processes Impacts* **2021**, *23* (9), 1316–1327. https://doi.org/10.1039/D1EM00089F.

(39) Mukherjee, P.; Reinfelder, J. R.; Gao, Y. Enrichment of Calcium in Sea Spray Aerosol in the Arctic Summer Atmosphere. *Marine Chemistry* **2020**, *227*, 103898. https://doi.org/10.1016/j.marchem.2020.103898.

(40) Coleman, K. K.; Tay, D. J. W.; Sen Tan, K.; Ong, S. W. X.; Son, T. T.; Koh, M. H.; Chin, Y. Q.; Nasir, H.; Mak, T. M.; Chu, J. J. H.; Milton, D. K.; Chow, V. T. K.; Tambyah, P. A.; Chen, M.; Wai, T. K. Viral

Load of SARS-CoV-2 in Respiratory Aerosols Emitted by COVID-19 Patients While Breathing, Talking, and Singing. *Clinical Infectious Diseases* **2021**. https://doi.org/10.1093/cid/ciab691.

(41) Fennelly, K. P. Particle Sizes of Infectious Aerosols: Implications for Infection Control. *The Lancet Respiratory Medicine* **2020**, *8* (9), 914–924. https://doi.org/10.1016/S2213-2600(20)30323-4.

(42) Morawska, L.; Johnson, G. R.; Ristovski, Z. D.; Hargreaves, M.; Mengersen, K.; Corbett, S.; Chao, C. Y. H.; Li, Y.; Katoshevski, D. Size Distribution and Sites of Origin of Droplets Expelled from the Human Respiratory Tract during Expiratory Activities. *Journal of Aerosol Science* **2009**, *40* (3), 256–269. https://doi.org/10.1016/j.jaerosci.2008.11.002.

(43) Reifenrath, R. Chemical Analysis of the Lung Alveolar Surfactant Obtained by Alveolar Micropuncture. *Respiration Physiology* **1973**, *19* (1), 35–46. https://doi.org/10.1016/0034-5687(73)90088-1.

(44) Walker, J. S.; Archer, J.; Gregson, F. K. A.; Michel, S. E. S.; Bzdek, B. R.; Reid, J. P. Accurate Representations of the Microphysical Processes Occurring during the Transport of Exhaled Aerosols and Droplets. *ACS Central Science* **2021**. https://doi.org/10.1021/acscentsci.0c01522.

(45) Port, J. R.; Yinda, C. K.; Owusu, I. O.; Holbrook, M.; Fischer, R.; Bushmaker, T.; Avanzato, V. A.; Schulz, J. E.; Martens, C.; van Doremalen, N.; Clancy, C. S.; Munster, V. J. SARS-CoV-2 Disease Severity and Transmission Efficiency Is Increased for Airborne Compared to Fomite Exposure in Syrian Hamsters. *Nat Commun* **2021**, *12* (1), 4985. https://doi.org/10.1038/s41467-021-25156-8.

(46) Rabaan, A. A.; Al-Ahmed, S. H.; Al-Malkey, M. K.; Alsubki, R. A.; Ezzikouri, S.; Al-Hababi, F. H.; Sah, R.; Mutair, A. A.; Alhumaid, S.; Al-Tawfiq, J. A.; Al-Omari, A.; Al-Qaaneh, A. M.; Al-Qahtani, M.; Tirupathi, R.; Hamad, M. A. A.; Al-baghli, N. A.; Sulaiman, T.; Alsubait, A.; Mehta, R.; Abass, E.; Alawi, M.; Alshahrani, F.; Shrestha, D. B.; Karobari, M. I.; Pecho-Silva, S.; Arteaga-Livias, K.; Aldana, D. K. B.; Rodriguez-Morales, A. J. Airborne Transmission of SARS-CoV-2 Is the Dominant Route of Transmission: Droplets and Aerosols. 10.

(47) Chirizzi, D.; Conte, M.; Feltracco, M.; Dinoi, A.; Gregoris, E.; Barbaro, E.; La Bella, G.; Ciccarese, G.; La Salandra, G.; Gambaro, A.; Contini, D. SARS-CoV-2 Concentrations and Virus-Laden Aerosol Size Distributions in Outdoor Air in North and South of Italy. *Environment International* **2021**, *146*, 106255. https://doi.org/10.1016/j.envint.2020.106255.

(48) Wallace, L. E.; Liu, M.; van Kuppeveld, F. J. M.; de Vries, E.; de Haan, C. A. M. Respiratory Mucus as a Virus-Host Range Determinant. *Trends Microbiol* **2021**, *29* (11), 983–992. https://doi.org/10.1016/j.tim.2021.03.014.

(49) Alder, B. J.; Wainwright, T. E. Studies in Molecular Dynamics. I. General Method. *J. Chem. Phys.* **1959**, *31* (2), 459–466. https://doi.org/10.1063/1.1730376.

(50) Tuckerman, M. *Statistical Mechanics: Theory and Molecular Simulation*, Illustrated edition.; OUP Oxford, 2010.

(51) Davidchack, R. L.; Handel, R.; Tretyakov, M. V. Langevin Thermostat for Rigid Body Dynamics. *Journal of Chemical Physics* **2009**, *130* (23). https://doi.org/10.1063/1.3149788.

(52) Evans, D. J.; Holian, B. L. The Nose–Hoover Thermostat. *J. Chem. Phys.* **1985**, *83* (8), 4069–4074. https://doi.org/10.1063/1.449071.

(53) Hünenberger, P. H. Thermostat Algorithms for Molecular Dynamics Simulations. In *Advanced Computer Simulation: Approaches for Soft Matter Sciences I*; Dr. Holm, C., Prof. Dr. Kremer, K., Eds.; Advances in Polymer Science; Springer: Berlin, Heidelberg, 2005; pp 105–149. https://doi.org/10.1007/b99427.

(54) Murray, B. S.; Nelson, P. V. A Novel Langmuir Trough for Equilibrium and Dynamic Measurements on Air–Water and Oil–Water Monolayers. *Langmuir* **1996**, *12* (25), 5973–5976. https://doi.org/10.1021/la960748o.

Part I
Building Up Chemical Complexity at the Air/Sea Interface

Chapter 2

Surfactant Charge Modulates Structure and Stability of Lipase-Embedded Monolayers at Marine-Relevant Surfaces

2.1 Abstract

Lipases, as well as other enzymes, are present and active within the sea surface microlayer (SSML). Upon bubble bursting, lipases partition into sea spray aerosol (SSA) along with surface active molecules such as lipids. Lipases are likely to be embedded in the lipid monolayer at the SSA surface and thus have potential to influence SSA interfacial structure and chemistry. Elucidating the structure of the lipid monolayer at SSA interfaces and how this structure is altered upon interaction with a protein system like lipase is of interest, given the importance of how aerosols interact with sunlight, influence cloud formation and provide surfaces for chemical reactions. Herein, we report an integrated experimental and molecular dynamics (MD) study of *Burkholderia cepacia* lipase (BCL) embedded in a lipid monolayer and highlight the importance and role of electrostatic interactions, rather than hydrophobic interactions, as a driver for monolayer stability. Specifically, we combine Langmuir film experiments and computational methods to examine the interactions between the zwitterionic dipalmitoylphosphatidylcholine (DPPC) monolayer and BCL. Upon insertion of BCL from the underlying subphase into the lipid monolayer, it is shown that BCL permeates and largely disorders the monolayer while strongly interacting with zwitterionic DPPC molecules, as experimentally observed by Langmuir adsorption curves and Infrared Reflectance Absorbance Spectroscopy (IRRAS). All-atom MD is then used to provide insights into inter- and intramolecular interactions that drive these observations, with specific attention to the formation of salt bridges, or ionic bonding interactions. We show that after insertion into the DPPC monolayer, lipase is maintained at high surface pressures and in large BCL concentrations by forming a salt-bridge-stabilized lipase-DPPC complex. In comparison, when embedded in an anionic monolayer at low surface pressures, BCL preferentially forms intramolecular salt bridges, reducing its total favorable interactions with the surfactant and partitioning out of the monolayer shortly after injection. Overall, this study shows that the structure and dynamics of lipase-embedded SSA surfaces vary based on surface charge and pressure, and that these variations have the potential to differentially modulate the properties of marine aerosols.

2.2 Introduction

The lipid monolayer interface has largely been studied for its biological relevance—monolayers occur in the body as lung surfactants and tear films—but is also of great interest to the atmospheric chemistry community for its role in modulating SSA particle reactions and dynamics.[1–9] Lipid monolayers are known to coat the ocean surface and the surface of nascent SSA and can significantly impact SSA climate-relevant properties.[3–5,7–11] It has been shown that the lipid type and composition of SSA surfaces have all been found to influence hygroscopicity,[3,6] cloud condensation nucleation activity,[3,5,6] and ice nucleation activity of SSA.[4,11] Amphiphilic lipids can impact interfacial properties of SSA primarily because they are surface active. Surfactants such as phospholipids participate in photooxidation and ozonolysis,[12] their anionic headgroups selectively bind and concentrate trace metal cations at the surface,[13–15] and their carbon chain length and headgroup charge(s) influence the transport and reactivity of gases at the interface.[7–9] While the chemical composition of SSA and their effects on climate are becoming more resolved, the molecular structure and dynamics of SSA surfaces remain difficult to probe experimentally.[16–23]

Computational methods such as all-atom MD have therefore contributed largely to the existing body of work on the lipid monolayer interface.[24–27] Since computational methods can resolve chemical systems at the atomic level, the integration of computation with experimental aerosol techniques is expected to significantly advance our ability to model the SSA interface.[28] The present study thus takes an integrated computational and experimental approach to understand how the lipid monolayer surface of SSA is affected by the presence of lipase.

Of particular importance to this work, lipases, as well as other enzymes, have been previously found in nascent SSA and remain active after aerosolization with consequences for aerosol reactivity.[19] Lipases catalyze the breakdown of high order glycerides into their respective fatty acids.[29–31] BCL, the lipase used in this study, is produced by a complex of bacterial species commonly found in marine environments.[32,33] In SSA, free lipases can either have been secreted by marine bacteria or released as their bacterial cells lyse during aerosolization.[34,35] A recent study on BCL embedded in DPPA, palmitic acid, and palmitate bilayers shows that BCL structure is highly dependent on its immediate chemical environment, suggesting that subtle variations in SSA lipid

composition and surface pressure could have drastic effects on its activity.[36–38] Although it is currently unknown whether lipases embedded in lipid monolayers, rather than bilayers, retain their activity, BCL in particular is known to be structurally and functionally robust; it is able to retain its conformation and catalytic activity in various aqueous environments, across a broad pH range, and in the presence of high organic solvent concentrations.[39–41] In the context of SSA surface chemistry, the presence of an active lipase enzyme at the surface would mean that the surface properties of the particle are dynamic and change over time. If lipases process lipids at the surface of SSA, there would be a flux in surface pressure, charge distribution, and even overall stability and solubility of the monolayer as the particle ages. [13,25,42–45] Furthermore, bioaerosols containing proteinaceous material provide a host of additional reactive sites that readily participate in processes such as radical photochemistry, nitration by gaseous NO_3, and ozonolysis.[46–48] The biotransformation of SSA by lipase is thus of direct relevance to atmospheric chemistry.

In addition to directly modifying the chemical composition of SSA, lipases are also likely to alter interfacial chemistry through intermolecular interactions with surrounding surfactants.[49] The lipid population at SSA surfaces is not necessarily dominated by the enzyme substrate. Rather, at the surface is a complex mixture of fatty acids, phospholipids, triglycerides, and many other organic components, meaning that folded and functional lipases will not always chemically process their surroundings. [5,11,16,20,23] For example, BCL induces lipid curvature when embedded in DPPA bilayers, a finding which is unsurprising given that membrane-embedded proteins are well-known to structurally modulate their surrounding chemical environment and *vice versa*.[36,50–52] The lipid-lipase interactions are likely to be impacted by the lipid type and headgroup charge, as these influence hydrogen-bonding, Van der Waal's forces, and salt-bridges. One of the goals of the present study is to characterize the dominant intermolecular interactions between lipase and both charge-balanced zwitterionic and negatively-charged surfactant monolayers, and to understand how these interactions might affect SSA surface chemistry.

Here we integrate IRRAS combined with Langmuir trough techniques with all-atom MD simulations to characterize the surface properties of model lipase-embedded monolayer systems. We investigate BCL interactions with two atmospherically-relevant lipids, zwitterionic DPPC and anionic DPPA, both of which form stable monolayer films at air-aqueous interfaces. In SSA, these lipids originate in bacterial cell membranes and

are released when bacterial cells lyse due to cell death or through the aerosolization process.[34,35] While nascent SSA surfaces are likely to be heterogeneous mixtures of phospholipids and fatty acids,[16] homogenous monolayers were selected for the present study to reduce the complexity of real world systems and allow for careful examination of lipase-surfactant interactions. DPPA has a similar chemical structure to DPPC; both are glycerophospholipids containing two 16-carbon, saturated hydrocarbon chains, with the exception that DPPC contains a positively-charged quaternary ammonium moiety attached to the negatively-charged phosphate headgroup. BCL dynamics in DPPA at low surface pressure have been studied previously; DPPC was selected for this work because the effects of zwitterionic lipids on lipase is not well understood.[36] IRRAS is used to extract information about surfactant organization and ordering, coupled with a film balance and a Langmuir trough to assess the effects of varying surface pressure on BCL-embedded monolayers. All-atom MD is then used to model protein-lipid interactions, lipid interactions, and protein structure and dynamics to inform the analysis of the experimental data. Understanding the reciprocal interactions between lipid monolayers and lipase can help us better understand the structural and chemical properties of nascent SSA.

2.3 Experimental Section

2.3.1 Materials and Sample Preparation

Lipase from *Pseudomonas cepacia* (≥ 30 U/mg) was purchased from Sigma Aldrich. Lipids (DPPA and DPPC) (> 99%) were from Avanti Polar Lipids. Chloroform (> 99.9%) was from Fisher Scientific. The agents above were used without further purification. The DPPC solution was prepared in chloroform with a concentration of 1 mg/ml and the DPPA solution was prepared in a 4 : 1 chloroform : methanol solution with a concentration of 0.42 mg/ml.[53,54] Milli-Q water with an electric resistance of 18.2 MΩ was used for the aqueous subphase. NaCl salt purchased from Fisher Scientific was purified by baking at 200°C overnight to remove organic contaminants and was prepared as a 0.4 M salt water solution. The concentration of NaCl solution was chosen to be near the sea water concentration. The lipase solution was made in 0.4 M NaCl with a concentration of 4.55 mg/ml. The pH of the subphase was around 5.5, which is within the range of aerosol acidity.[45,54]

2.3.2 Langmuir Surface Adsorption Curve

A computer-controlled film balance with a Langmuir trough (KSV NIMA LB, S/N AAAA100505) was used for performing the surface adsorption experiments while holding surface pressure constant. A water circulator (Beckman Geneline Cooler) controlled the temperature of the subphase in the trough at 22 ℃. The lipid solution was first spread onto the 55 ml subphase in the trough using a microsyringe. The monolayer was then left for 20 minutes in order for the solvent to evaporate. After that, the monolayer was compressed to a desired surface pressure and was held at this surface pressure. A 0.2 mL lipase solution was injected with a syringe coupled with Teflon tubing on the needle underneath the lipid monolayer as described previously.[36] The final concentrations of lipase in the subphase were 250 nM, 500 nM, 1000 nM, or 1350 nM. The change in molecular area was monitored throughout the experiment while holding the surface pressure constant.

We performed holding area experiments using the computer-controlled film balance with a slightly modified side hole petri dish, designed by Allen and co-workers.[15] The lipid monolayer was first formed on the surface of 55 ml subphase in the petri dish. The lipase solution was then injected underneath the monolayer through the side hole after 30 minutes and the final concentration of lipase in the subphase was 500 nM. While the area was kept constant, the surface pressure was monitored with the computer-controlled film balance throughout the experiment.

2.3.3 Infrared Reflection Absorption Spectroscopy (IRRAS) System

The IRRAS set up has been described previously.[12] The IR beam from an infrared spectrometer (Bruker Tensor 37) is directed onto the aqueous solution surface in the Langmuir trough or the side hole petri dish at a 30° angle from the surface. At an air-water interface, the optimal angle for the incident beam is 0º to 40º when an unpolarized beam is used.[55,56] The reflected beam is collected and sent to an MCT detector (Infrared Associates Inc., mid-band with ZnSe window). Each IRRAS spectrum is an average of 300 scans with a spectral resolution of 4 cm^{-1}. Due to the low signal and water vapor interference in other regions, here we focus only on the C-H stretching region. For some of the experiments, the IRRAS spectrum was taken every 30 minutes throughout the experiment and some were taken once at the beginning and once at the end. The reflectance-absorbance (RA) for the IRRAS spectra were plotted as a function of wavenumber where

$$RA = -\log (R/R_0) \quad (2.1)$$

and R is the reflectivity of the film covered surface while R_0 is the reflectivity of the aqueous subphase.

2.3.4 Molecular Dynamics System Preparation

All monolayer systems were set up using CHARMM-GUI in 0.4 M NaCl at 298.15 K and were composed of a rectangular box with periodic boundary conditions.[57] Two lipid monolayer leaflets containing molecules of either DPPA or DPPC were placed above and below a TIP3P water slab, creating two interfaces in the xy-plane, with air (vacuum) on either side. The xy dimensions of the box were 100 x 100 Å, while the total length of the box in the z direction was 300 Å. A diagram of the system set-up is given in Figure 3.7. TIP3P was selected due to its high compatibility with the CHARMM36 force field for lipids.[58] Following routine protocols for monolayer simulations, vacuum was used at the air side of the air-aqueous interfaces to approximate the density of gases in the atmosphere.[24–26,59] The lipase starting structure used for protein-lipid systems was based on the PDB entry *3LIP*, with protonation states for all amino acids assigned for a pH 6 system through the Schrödinger Protein Preparation Wizard.[60,61] For a mean molecular area (MMA) of 47 $Å^2$/lipid, 212 DPPC or DPPA molecules were placed in each leaflet. For leaflets containing lipase, the lipase was placed in the center of the leaflet at an orientation calculated using the Orientation of Proteins in Membranes (OPM) database and server and surrounded by 179 lipid molecules.[62] Each system built consisted of approximately 100,000-120,000 atoms. All systems were prepared for simulation with GROMACS and parametrized using the CHARMM36 forcefield.[58,63,64] The monolayers were energy-minimized and equilibrated in 6 equilibration steps using GROMACS 2018 on an NVIDIA GTX 1080Ti GPU (GeForce GTX Titan, NVIDIA, Santa Clara, CA). The energy minimization and equilibration steps are described in the Supporting Information.

2.3.5 Molecular Dynamics Simulations and Data Analysis

Production steps for each system were run for approximately 400 ns total with 3 replicates under an *NVT* ensemble, an ensemble commonly used for monolayer simulations.[25,26,59] For lipase in each DPPA and DPPC, replicates 1, 2 and 3 were run for 100, 100 and 200 ns, respectively. The systems were assumed to be fully equilibrated when energy, RMSD, and solvent-accessible surface areas (SASA) converged (Figure 3.8). All production runs utilized the Extreme Science and Engineering Discovery Environment (XSEDE), which is

supported by National Science Foundation grant number ACI-1548562.[65] Specifically, simulations were performed on the Bridges supercomputer, which is supported by NSF award number ACI-1445606 at the Pittsburgh Supercomputing Center (PSC).[66] The production run script is given in full in Figure 3.9.

Molecular simulation data was analyzed in the IPython Jupyter Notebook environment, and all IPython notebooks for this work will be made available as part of the UCSD Library Digital Collections: Center for Aerosol Impacts on Chemistry of the Environment (CAICE) (https://library.ucsd.edu/dc/collection/bb96275693).[67]

2.4 Results and Discussion

2.4.1 Lipase insertion disrupts the lipid monolayer as evidenced by experimental and computational methods

A Langmuir trough coupled with IRRAS spectroscopy was used to interrogate the impact of lipase on the surface properties of a DPPC monolayer. The isotherm of DPPC can be seen in Figure 3.10a. This isotherm has been reported in the literature.[14,68,69]. We performed our holding surface pressure experiments at a surface pressure of 10, 15 and 25 mN/m. In **Figure 2.1**, the Langmuir curve of the DPPC monolayer at 10 mN/m after lipase injection shows a significant increase in area (90%) within the first 5 minutes. Figure 3.11 shows that when holding the DPPC monolayer at a certain area, the surface pressure increased after lipase injection. Based on the fact that the area relaxation and pressure relaxation experiments are done with different methods, the pattern of the increasing curves between these two kinds of experiments are not comparable. Before lipase injection, the IRRAS spectra of the C-H stretching region between 2800 cm^{-1} to 3000 cm^{-1} included three peaks: the methylene symmetric stretch (2850 cm^{-1}), methylene asymmetric stretch (2920 cm^{-1}) and methyl asymmetric stretch (2958 cm^{-1}). From the IRRAS spectra before and 5 minutes after lipase injection, we observed a significant decrease in the peak intensity associated with the methylene stretch after lipase injection, as well as a broadening of the methylene stretching peaks and a shift to a higher wavenumber (**Figure 2.1** insertion). The peak associated with the methylene asymmetric stretch shifted from 2920 cm^{-1} to 2924 cm^{-1}. The methyl asymmetric stretch does not show a significant shift in frequency or peak broadening, but instead has a small decrease in intensity. The peak at 2870 cm^{-1} after lipase injection could be associated with the methyl symmetric

stretch[55] which shows a small shoulder near the methylene symmetric stretch at 2850 cm^{-1} before lipase injection. The methyl stretch only shows a small decrease in intensity while the methylene stretch shows a significant decrease in intensity; the peak at 2850 cm^{-1} before lipase injection disappears into the methyl stretch peak at 2870 cm^{-1} upon lipase injection. The shift of the methylene groups to a higher wavenumber indicates a disordered structure of the DPPC monolayer after lipase insertion, as previous studies have found that the methylene stretching peaks shift to a higher wavenumber when entering a less ordered state.[70] Additional evidence for increased disorder for DPPC structure is seen in the peak broadening. Previous studies have found that peaks in the C-H stretching region in Raman spectra broaden when going from a more ordered crystalline state toward a more disordered state.[71–73] Conformational order information can also be obtained from the peak intensity ratio of the methylene asymmetric and methylene symmetric peaks.[71–74] This ratio changed from 1.9 to 1.5 after lipase injection, which indicates a more disordered system induced by lipase insertion. The intensity loss can be attributed to either the increase in mean molecular area of the DPPC molecules, indicating a lower density of DPPC molecules on the surface, or a loss of DPPC molecules through dissolution into the subphase. The IRRAS spectra collected at longer time (greater than 5 minutes) remained almost the same (data not shown).

We compared the lipase-DPPC interactions at different surface pressures by injecting lipase into the DPPC monolayer and monitoring the change in surface area increase (**Figure 2.1b**). After lipase injection, there was a fast increase followed by a slower increase in surface area before reaching a plateau at every surface pressure studied. The total area expansion decreased with an increase in surface pressure of DPPC, which indicates either a suppression of lipase insertion or a smaller surface area of lipase embedded in the surface at higher surface pressures. This is consistent with previous experimental studies showing that, when injected into lipid monolayers, the area increase induced by lipase is reduced at high surface pressures compared to lower surface pressures.[75,76] The IRRAS spectra before and after lipase injection for the DPPC monolayer at 15 mN/m and 25 mN/m are shown in **Figure 2.12a** and **b**, respectively. These spectra indicate that lipase has a smaller effect on DPPC structure at a higher surface pressure of 25 mN/m in comparison to the lower surface pressures investigated.

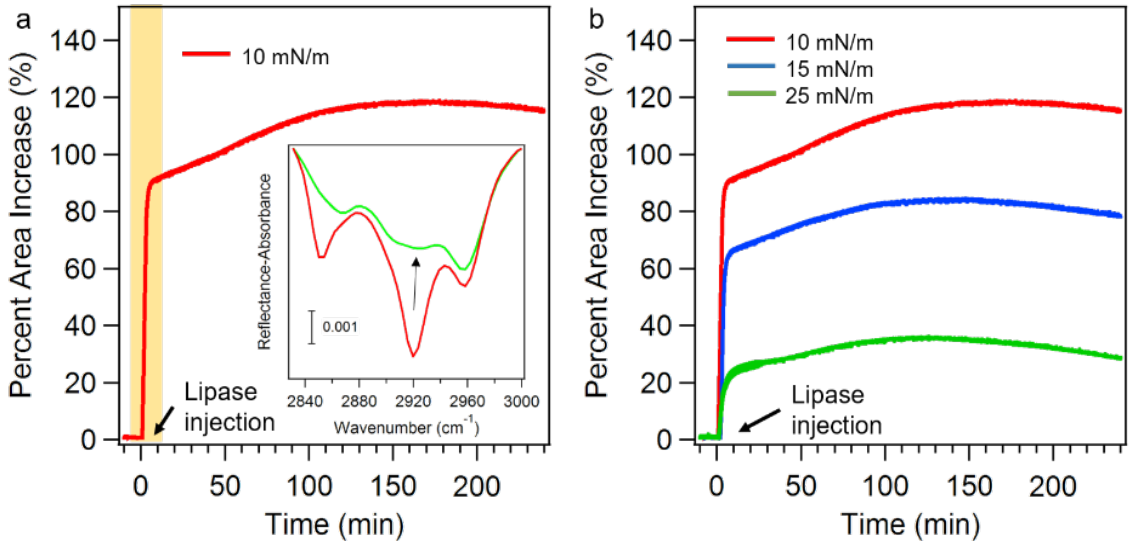


Figure 2.1: Percent area increase of DPPC monolayers after lipase injection and associated IRRAS spectra. A) Increase in area of DPPC monolayer after holding at 10 mN/m induced by injection of lipase in the underlying subphase. Insert: IRRAS spectra before (red) and 5 mins after (green) lipase injection (final concentration of lipase of 500 nM). The yellow box indicates the time when the IRRAS spectra are taken. B) Percent area increase of DPPC monolayers at 10, 15, and 25 mN/m induced by injection of lipase into the underlying subphase (final concentration of lipase of 500 nM).

Molecular dynamics simulations of BCL in DPPC show that not only does lipase induce molecular disordering of DPPC, but it becomes embedded in the monolayer such that the surface of the enzyme is exposed to the atmosphere. In comparison, BCL in DPPA shows similar behavior; it fully embeds into the surfactant monolayer and disorders the surrounding lipids (**Figure 2.2**). Figure 3.2b gives plots of the total atmosphere-exposed surface area over the final 50 ns of each production step for lipase in DPPC and in DPPA at MMA 47 $Å^2$—a surface pressure of approximately 25 mN/m. This suggests a molecular mechanism to explain the intensity decrease and peak broadening in the IRRAS spectra. Lipase expands and embeds itself into the monolayer, decreasing the concentration of DPPC molecules at the surface and lowering the intensity of the IRRAS signal. The lipase could also contribute to the broadening of the IRRAS peaks by inducing curvature of the monolayer film. **Figure 2.2a** shows the monolayer curvature around the enzyme over the final 50 ns of simulation in DPPC and DPPA. The lipids surrounding lipase are warped, i.e., disrupted, from their hexagonal packing structure and reoriented around the lipase surface. Warping the monolayer out of the interfacial plane disrupts the C-H stretching signal, causing a decrease in and broadening of signal intensity. [77,78] Finally, **Figure 2.2c** shows

comparisons between the nematic ordering of the lipids with and without embedded lipase. The nematic order parameter, a value between 0 and 1, is an indication of unidirectional packing of the lipids, with 1.0 representing perfect ordering. Figure3.2c shows that the nematic ordering decreases in the presence of lipase both in DPPC and DPPA. These findings are consistent with previous studies indicating that the surrounding interfacial environment is key to lipase structure, function and activity; however, these show this phenomenon occurring at monolayer, rather than bilayer or aqueous-organic, interfaces.[36–38,51,52]

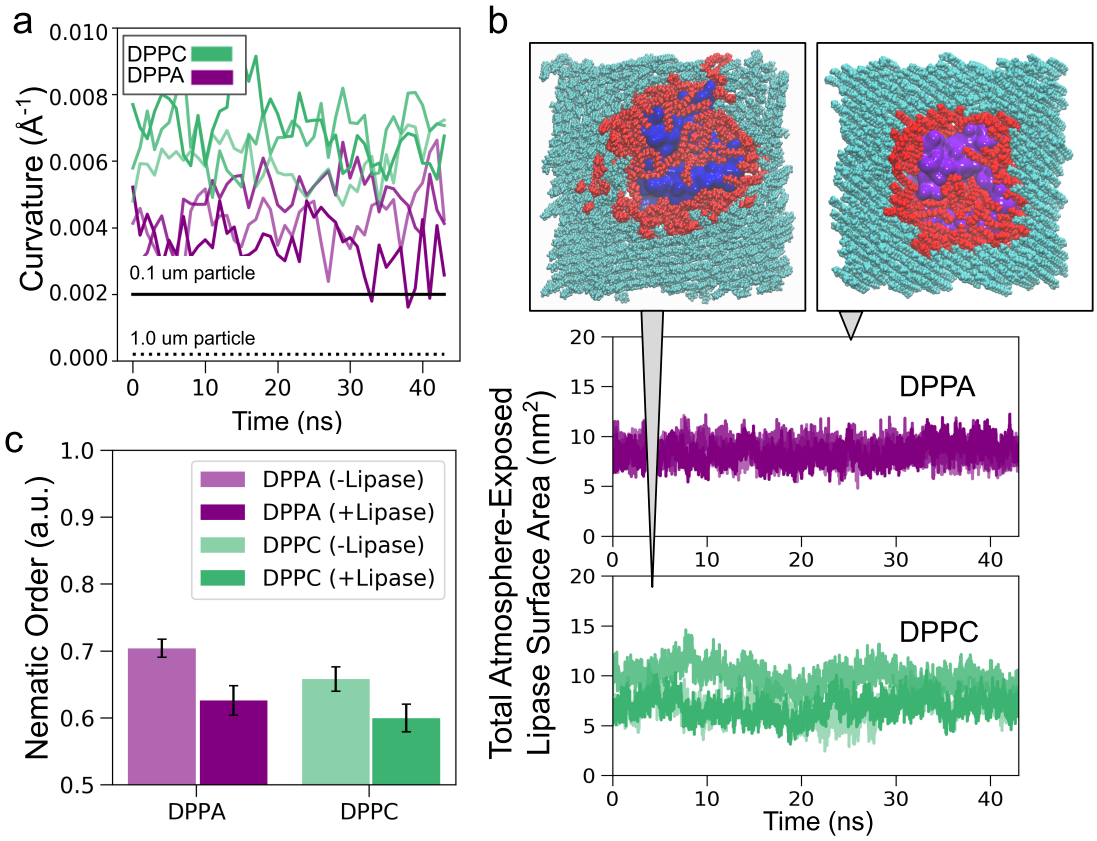


Figure 2.2: Data pulled from MD simulations of lipase in DPPC and comparison to DPPA at MMA 47 Å^2, or approximately 25 mN/m in surface pressure. A) Quantification of monolayer curvature over the final 50 ns of each simulation replicate. Curvature was calculated using SciPy's least squares fit; after fitting a surface to the points given by the phosphate headgroups, the inverse radius of each osculating circle in the xz- and yz-planes was extracted. The values representing the largest curvatures, in units of Å^{-1}, are plotted. For reference, lines corresponding to the curvatures of a 0.1- and 1.0-micron particle are shown in black (solid and dotted lines, respectively). B) Total atmosphere-exposed surface area of lipase over the final 50 ns of simulation for each replicate. Insets: Top views of lipase (blue) in DPPC (left) and lipase (purple) in DPPA (right). DPPC and DPPA molecules within 5 Å of lipase are colored red and outside of 5 Å, cyan. C) Average nematic order parameters for lipid leaflets with and without lipase embedded. DPPA is given in purple, DPPC in green.

These results indicate that even though the lipase does not directly act upon on the surrounding lipids, it significantly alters the aerosol surface morphology with potential climate-relevant effects. For example, the insertion of lipase into SSA surfaces could increase particle hygroscopicity due to the disordering of the lipids at the interface, evidenced by the decrease in nematic order parameter and increase in monolayer curvature. Tight packing conformations of lipids in a monolayer film have previously been cited as a barrier to water uptake.[3,43,79,80] Early studies on mixed lipid monolayer films show a well-characterized drop in monolayer resistance in the presence of low concentrations of impurities.[81–83] At high surface pressures with highly-aligned lipid packing conformations, it can be expected that a small defect such as a disruption to the monolayer microstructure by lipase, will lead to a large drop in monolayer resistance and a corresponding increase in water permeation across the interface. Because of this, it may be reasonable to suspect that lipase-containing SSA could account for a portion of the cloud condensation nucleation activity of organic-enriched aerosols. Additionally, the exposure of lipase at the surface could have implications for atmospheric chemistry: proteins contain charged amino acids that easily react upon collision with atmospheric gases. For example, BCL tyrosine Y129 is exposed at the surface in our MD simulations. Tyrosine reacts readily with atmospheric ozone and NO_x species, providing one mechanism by which the biotransformation of aerosols could play a role in the atmospheric nitrogen cycle.[46]

2.4.2 A comparison of lipase interactions with anionic DPPA and zwitterionic DPPC monolayers: experimental data

When lipase was injected under DPPC and DPPA monolayers at 25 mN/m, while lipase remained inserted in the DPPC monolayer, it crashed out of the DPPA monolayer after 85 minutes, suggesting lower lipase stability in DPPA monolayer than in DPPC monolayer (**Figure 2.3**). The isotherm for DPPA can be found in Figure 3.10b. In previous work, BCL was reported to be stable in the DPPA monolayer at 5 mN/m, which indicates that lipase can bind with DPPA, but the protein-lipid interaction is weak compared to that of lipase with DPPC.[36] The percent area increase for DPPA before lipase crashes out is smaller than in DPPC at the same time, which also indicates that DPPA binds weakly with lipase in comparison to DPPC. **Figure 2.3** shows a slow decrease in area at the end of the DPPC curve, as well as a decrease in area below the starting point in the DPPA curve. This decrease can be partially attributed to the spontaneous dissolution of the lipids into the monolayer as

previously seen for studies of long chain fatty acid monolayers at the air-water interface.[12,84,85] **Figure 2.12** shows that the background area decreases when the pure lipid monolayer is held at a constant 25 mN/m surface pressure. However, the area loss due to the lipid background is still less than the area loss shown in **Figure 2.3**. Therefore, this area loss can be attributed to two additional mechanisms: (i) a slight loss of lipase-lipid complex, as simulations indicate that lipids can not only be incorporated into the lipase active site, but they also can adhere to its surface; and (ii) the structural realignment of the lipids towards higher nematic orders (straighter hydrophobic tails), which decreases their individual molecular areas.

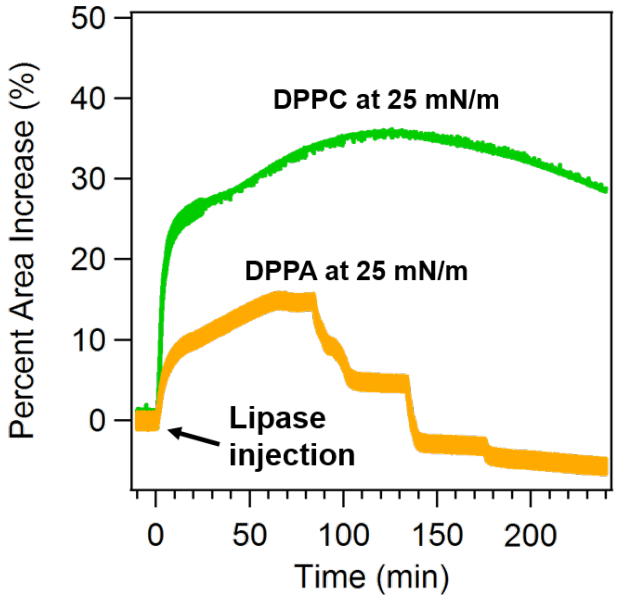


Figure 2.3: Percent area increase and decrease of DPPA and DPPC monolayers at 25 mN/m, induced by injection of lipase in the underlying subphase (final lipase concentration 500 nM).

To test the effects of lipase concentration on its insertion to DPPC and DPPA monolayers at higher surface pressures, we performed similar experiments at 25 mN/m with different lipase concentrations. Figure 3.4a shows that lipase was stable in DPPC monolayers even at high lipase concentrations. The final area increase of the DPPC monolayer was dependent on the concentration of lipase used in our study (**Figure 2.4b**). In contrast, lipase was stable in DPPA monolayers only at low lipase concentrations (**Figure 2.4c**); at higher concentrations, it crashed out from the monolayer after a brief time. The "crashing out" phenomenon is further supported by IRRAS spectra provided in Figure 3.4d. At high lipase concentrations (1000 nM), the area increased rapidly in the first 10 minutes after lipase injection, followed by a slower increase until 110 minutes

lapsed. During this time, the IRRAS spectra (**Figure 2.4d**, left IRRAS panel) show a decrease in signal intensity and a broadening of the peaks in the C-H stretching region, indicating that lipase decreased the chain ordering of the DPPA monolayer, similar to the behavior of lipase in DPPC (**Figure 2.1a**). This disordering is confirmed by MD simulations (**Figure 2.2c**). However, after 110 minutes, the area gradually decreased back to its initial state, and the IRRAS spectra (**Figure 2.4d**, middle IRRAS panel) indicated an increase in signal, in which the peaks gradually recovered. For comparison, the IRRAS spectra before and after 240 minutes of lipase injection are presented in **Figure 2.4d** (right IRRAS panel) and they appear almost the same. These IRRAS spectra indicate that almost all of the lipases inserted into the DPPA monolayer were eventually squeezed out from the interface. This "crashing out" phenomenon has been observed previously where Meister et al. found that the farnesylated and hexadecylated N-Ras proteins were squeezed out from the lipid monolayer at 30 mN/m, but doubly hexadecylated N-Ras proteins at the same surface pressure remained embedded.[76]

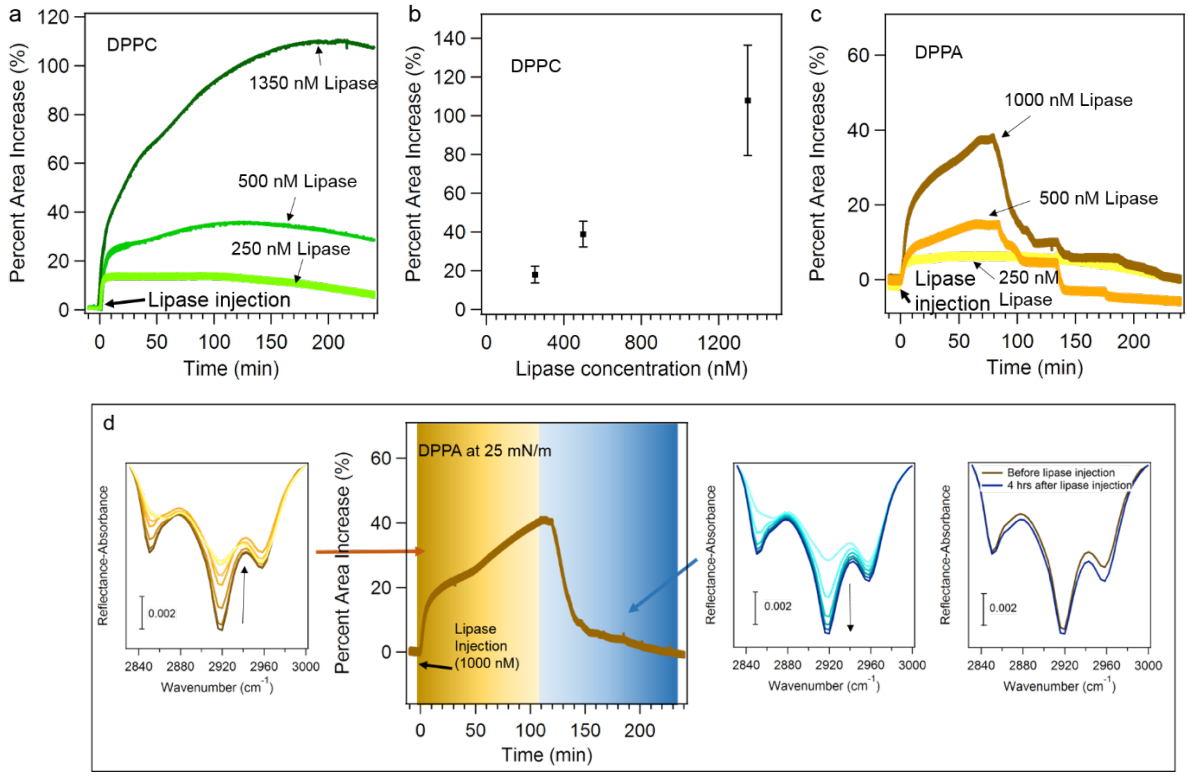


Figure 2.4: Percent area increase upon lipase injection and associated spectra: comparison of DPPC to DPPA monolayers. A) Percent area increase of the DPPC monolayer at 25 mN/m, induced by injection of different concentrations of lipase to the underlying subphase. B) Percent area increase of the DPPC monolayer vs. concentration of lipase in the underlying subphase. C) Percent area increase of the DPPA monolayer at 25 mN/m induced by injection of different concentrations of lipase into the underlying subphase. D) Area increase and decrease of the DPPA monolayer at 25 mN/m, induced by injection of lipase into the underlying subphase (final concentration 1000 nM). IRRAS spectra are: (left) before and 5, 10, 30, 60, 90 and 120 minutes after lipase injection (dark yellow to light yellow), (middle) 120, 135, 150, 180, 210 and 240 minutes after lipase injection (light blue to dark blue) and (right) before lipase injection (dark yellow) and 240 minutes after lipase injection (dark blue).

2.4.3 Molecular dynamics calculations to further probe and compare interactions between lipase DPPC and DPPA

To further investigate and compare the interactions between lipase and DPPA and DPPC, molecular dynamics simulations run for each system at a surface pressure of 25 mN/m were analyzed for indicators of intermolecular electrostatic interactions. To describe electrostatic interactions between the lipase and surrounding charges on the lipids, we measured instances of salt bridge formation across all trajectories. Salt bridges, defined here as ionic interactions between a positively charged nitrogen and a negatively charged oxygen, are determined to have formed if the two participating atoms enter within 3.5 Å of one another. Since

salt bridge distances can be highly variable, we have selected 3.5 Å as an intermediate value within previously reported ranges (2.8-5.0 Å) as a cutoff.[86–89] To understand the sensitivity of our distance cutoff selection, we provide error bars representing higher and lower cutoff values, 4.0 and 3.2 Å, respectively. **Figure 2.5a** shows, in color, salt bridge formation between DPPC or DPPA headgroups and lipase. We have also included in grey competing intramolecular salt bridges ("INT") that form between protein residues. This figure gives a depiction of the likelihood that specific interactions will occur.

The additional charge on zwitterionic DPPC as compared to anionic DPPA provides for more total interactions between the lipid headgroups and lipase (**Figure 2.13**). Both headgroups contain a negatively-charged phosphate, which can interact with positively-charged residues such as lysine (K, green) and arginine (R, red). However, DPPC also contains a positively-charged ammonium that can interact with negatively-charged surface residues such as aspartate (D, orange) and glutamate (E, purple). **Figure 2.5a** shows that not only does lipase in DPPC contain significantly more favorable electrostatic interactions than lipase in DPPA, it contains more interactions that are sustained for more than 40% of the simulation time. Additionally, DPPA coordinates strongly to sodium ions in solution (**Figure 2.14**). The negatively-charged DPPA surface is stabilized by recruiting sodium cations, a process that does not occur in the neutral DPPC monolayer; thus, the charged sites on BCL compete with sodium ions for coordination with DPPA. This provides some molecular evidence to explain why lipase is more stable in DPPC than DPPA at high lipase concentrations, and why the percent area increase after lipase injection in DPPA is significantly smaller than in DPPC.

Figure 2.5b shows all intramolecular salt bridges that occur between the two lipase systems. For BCL in DPPA, the high prevalence of competing intramolecular salt bridges (**Figure 2.5a**), as well as the number and duration of total intramolecular salt bridges (**Figure 2.5b**), suggests that when BCL is surrounded by negative charges, it favors more intramolecular electrostatic interactions than BCL in DPPC. Compared to BCL embedded in DPPC, not only is there a larger percentage of available enzyme sites that have competing interactions, but those interactions appear to be stabilized for longer portions of the simulation time. This suggests that in anionic surfactants, BCL derives its structural stability from internal electrostatics, supporting experimental evidence that shows decreased BCL-embedded monolayer lifetimes and a decreased BCL-induced percent area increase

for DPPA. Conversely, BCL readily complexes with zwitterionic DPPC surfactants, which allows the DPPC surface to sustain the embedded lipase at high concentrations and for longer time periods.

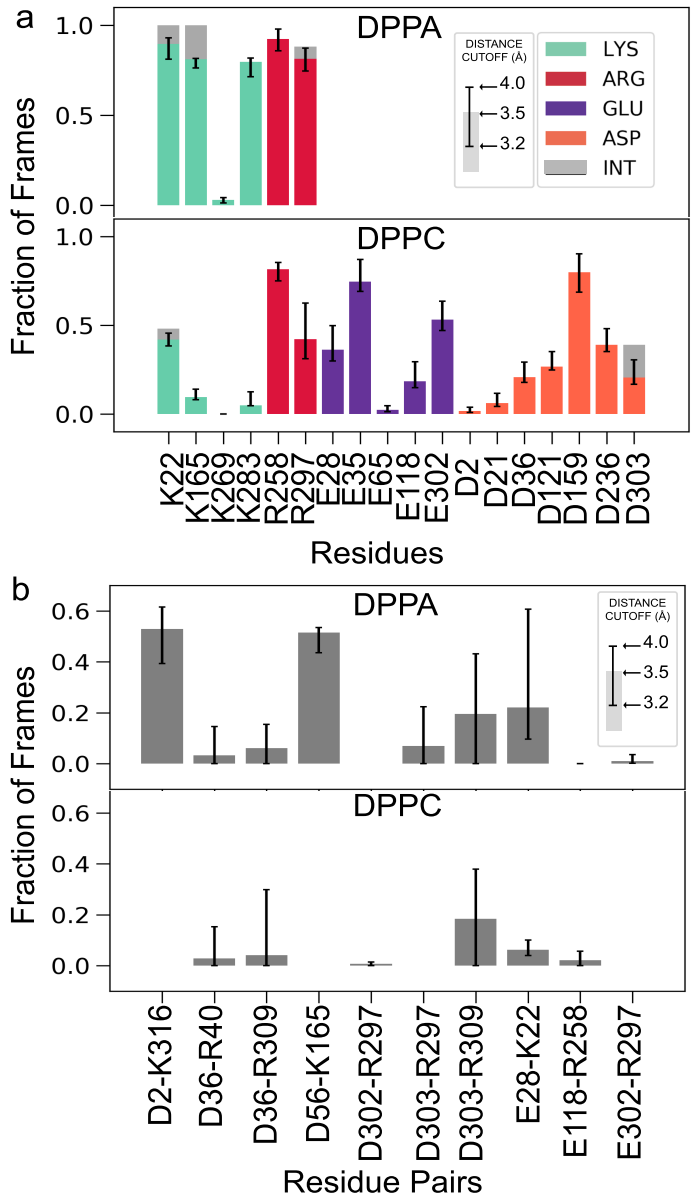


Figure 2.5: Lipase forms more intermolecular salt bridges with DPPC than DPPA but forms more intramolecular salt bridges in DPPA than in DPPC. Error bars are provided to indicate the sensitivity of the analysis to the distance cutoff selection. A) Bar plot of salt bridge formation between DPPC or DPPA molecules and lipase. Salt bridges were defined as sustained intermolecular interactions between a positively-charged nitrogen group and a negatively-charged oxygen group. Internal salt bridges ("INT", gray) are competing intramolecular salt bridges occurring between lipase residues. B) Intramolecular salt bridges forming between lipase residues.

Our results suggest that electrostatics modulate the ability of lipase to embed in a lipid monolayer. Previous all-atom MD studies have attributed the orientation of BCL at aqueous-organic interfaces to the degree of hydrophobicity of the organic phase, which functionally stabilizes the enzyme in its open conformation as lipids are oriented into the active site.[72,73] The mechanism of BCL orientation in a lipid monolayer is similar. The hydrophobic lid helices (α9, 4 and 5) that flank the catalytic pocket (**Figure 2.15**) are immersed into the organic layer, and the remaining hydrophilic portion of the lipase is exposed to aqueous solvent (**Figures 2.6b**, **2.6d**, and **2.15**). However, the phospholipid monolayers used in this work are modified aqueous-organic interfaces; the headgroups are charged and the length of the hydrophobic tails limit the depth of the organic "phase." At low surface pressures, it is possible that lipase stability at a monolayer interface is hindered by decreased organic phase depth and enhanced only in the presence of both positive and negative charges.

Considering our experimental and simulation data together, we provide a hypothesis to explain the rapid and discontinuous loss of area from the lipase-embedded DPPA monolayer seen in **Figure 2.3**. Due to the low intermolecular interactions between BCL and DPPA, it is possible that the lipases form small islands in the DPPA monolayer and then desorb from the monolayer together as aggregates. At low surface pressures, the hydrophobic lid residues on BCL drive the lipases to the surface. Initially, the lipases are likely to be evenly distributed throughout the monolayer. However, the insertion of the lipases causes a disruption to the lipid ordering (**Figures 2.1a**, **2.2c** and **2.4d**), decreasing the Van der Waal's interactions between the hydrophobic tails. Since DPPA favors higher nematic orders (**Figure 2.2c**), over time, the monolayer undergoes structural rearrangement to align the lipids back together (**Figures 2.1a** and 2.**4d**), leading to the slow decrease in area; the rearrangement also contributes to the clustering of the lipases into islands in order to reduce the overall free energy of the system. Ultimately, since BCL forms fewer salt bridges with DPPA and more internal salt bridges (**Figure 2.5b**), the lipase islands desorb from the surface in discrete groups until the monolayer returns to its original, lipase-free state. There is also evidence of structural realignment after lipase insertion to DPPC monolayers that causes the broad hump in **Figure 2.3**, but the salt bridges between lipase and DPPC stabilize BCL in the monolayer.

2.4.4 Differences in lipase structure and dynamics based on monolayer surface charge and reciprocal impacts on the lipase-embedded surface

Molecular dynamics results show that lipase adopts substantial structural changes when introduced to charged surfactants to maximize favorable intermolecular interactions. Taken with the results given above, structural alignment techniques, clustering, and SASA analyses are used to interrogate the major structural and dynamic differences between lipase systems.

Table 2.1 shows the results of surface area analyses on BCL categorized by polar and nonpolar surface areas as well as lipid-embedded and solvent-accessible surface areas. In DPPA, the solvent-accessible surface area increases from its starting structure by 19%, compared to 0% for BCL in DPPC. This can be attributed to partial unraveling of the lipase in response to exposure to the negatively charged interface, coordinating its positive charges with negatively-charged surfactants, and reducing negative-to-negative charge interactions. Insights into this structural change can be seen **in Figure 2.6e**. Compared to BCL in DPPC, BCL in DPPA achieves a final conformation with 11 nm^2 more solvent-exposed polar residues and 6 nm^2 more solvent-exposed nonpolar residues. While the total percentage of solvent-exposed surface area increases, the percentage of lipid-embedded surface area stays the same, differing from the starting structure by at most 4% in the case of DPPC, with negligible differences in total lipid-embedded surface area between the two final lipase conformations. Surface areas for lipids interacting with lipase are given in **Table 2.2**. **Figure 2.6a** shows the total lipase surface area over the final 50 ns of each replicate.

Gromos clustering was used to highlight the major structural and dynamic differences between the two lipase-embedded monolayer systems (**Figure 2.6e**).[74] In total, BCL embedded in DPPC attained 33 clusters, while BCL in DPPA attained 164. The top 80% of contributing clusters are given in **Table 2.3**. For BCL in DPPC, approximately 60% of simulation time is dominated by two BCL conformations, shown in **Figure 2.6e**, with the dominant cluster contributing to 43% of frames. In contrast, in DPPA, the same percentage of simulation time is dominated by 11 different conformations, with the highest contributing conformation representing only 10% of frames. For simplicity, **Figure 2.6e** only shows 5 of these conformations, representing 34% of total simulation frames. It is clear that the major contributing clusters differ significantly from each other in each surfactant type. Clustering results indicate that BCL is highly flexible in DPPA and rigid in DPPC, which is

supported by our salt-bridge analysis. Due to the high number of salt bridges formed between BCL and DPPC, we would expect the enzyme to be more rigid than in DPPA, forming a stable DPPC-lipase complex. Conversely, the lack of salt bridges formed between BCL in DPPA allows the enzyme to be more dynamic, sampling approximately 5-fold more conformational states than in DPPC. Clustering results, taken together with a thorough analysis of electrostatic interactions, provide molecular support for the destabilization of lipase at anionic surfaces and stabilization at zwitterionic surfaces.

Table 2.1: Surface area analysis on hydrophobic and polar residues exposed to solvent (*exposed)* and lipid (*embedded)* for MD simulations at surface areas of 47 $Å^2$/lipid.

Lipase Residues	Initial Structure[a] (nm^2)	DPPA (nm^2)	DPPC (nm^2)	Δ (nm^2) (DPPA-DPPC)
Polar + Charged *exposed*	54 ± 2	63 ± 3 (+16%)[b]	52 ± 2 (-4%)	+ 11 ± 2
Hydrophobic *exposed*	34 ± 2	43 ±33 (+26%)	37 ±3 (+8%)	+ 6 ± 2
Total Exposed	89 ± 4	106 ± 4 (+19%)	89 ± 4 (+0%)	+ 17 ± 3
Polar + Charged *embedded*	19 ± 1	19 ±32 (+0%)	21 ± 2 (+10%)	- 2 ± 2
Hydrophobic *embedded*	38 ± 1	37 ± 2 (-3%)	34 ±2 (-11%)	+ 3 ± 2
Total Embedded	57 ± 1	56 ± 2 (-2%)	55 ±3 (-4%)	+ 1 ± 2

[a] Surface areas of initial post-equilibration structures differed from each other by < 1%
[b] Difference between average and initial structure

Finally, we investigated how electrostatics govern the overall structure of the protein-embedded monolayer (**Figure 2.6**, **2.14**). **Figure 2.6b-d** shows a comparison between orientations of the lipase at the surface. **Figure 2.6c** shows the evolution of lipase tilt angle over the final 50 ns of each simulation, and **Figures 2.6b** and **2.6d** show MD snapshots of the lipase oriented in their respective lipids. In DPPC, BCL adopts a more tilted structure compared to BCL in DPPA, with the lipase tilting to a 45º angle with respect to the z-axis in DPPC *versus* only 20º in DPPA. In both cases, the hydrophobic active site helices are embedded in the lipids, but they achieve varying orientations in the monolayer plane. In DPPA, the loop formed by helices 4 and 5 lies in the plane of the headgroups, while in DPPC, the loop twists out of the plane (**Figure 2.6b**, **d**, and **Figure**

2.15). The surfactant headgroups in DPPC appear to scatter across the surface of lipase, increasing monolayer curvature, while the headgroups in DPPA remain more planar (Figure 3.2a).

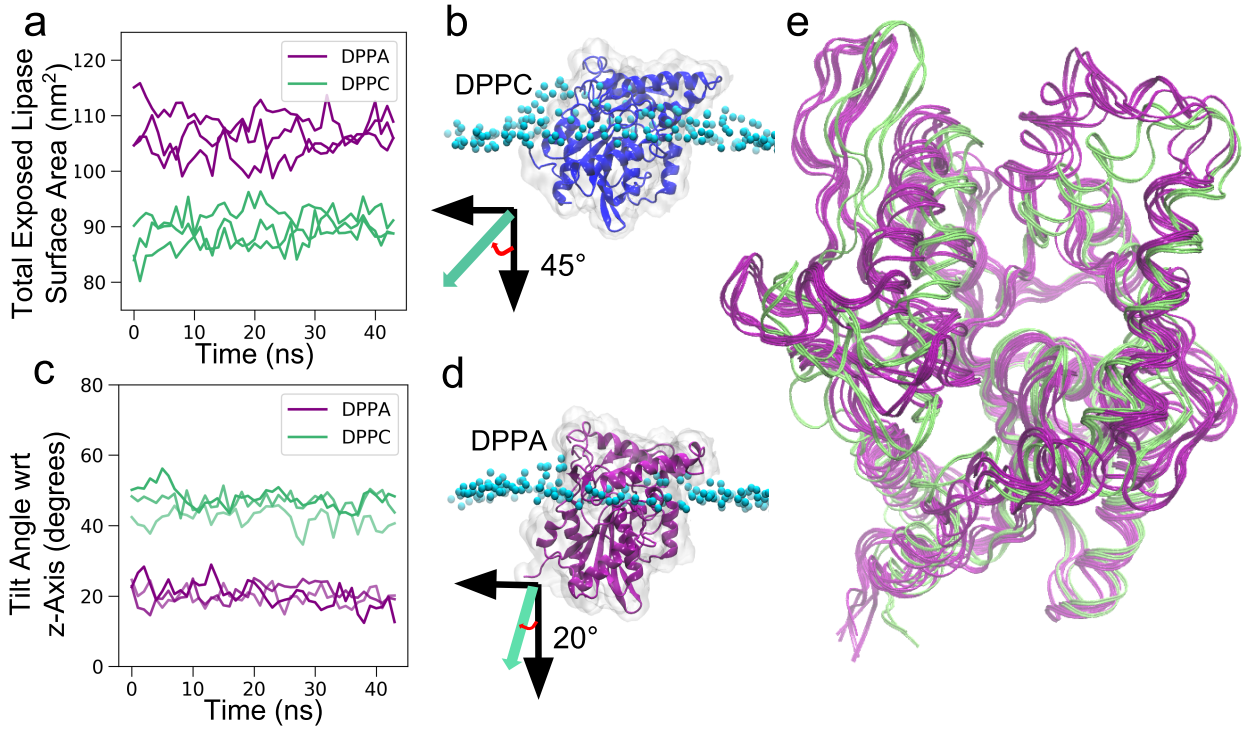


Figure 2.6: The structure of lipase and the lipase-monolayer complex changes based on surfactant charge, demonstrated by notable changes in lipase solvent-accessible surface area, angle of embedment into the monolayer, and structural clustering. A) Total solvent-exposed lipase surface area by surfactant; B) Orientation of lipase (blue) in DPPC with the phosphate headgroups represented as cyan spheres; C) The tilt angle, defined by the vectors associated with internal beta sheets with respect to the z-axis. D) Orientation of lipase (purple) in DPPA with the phosphate headgroups represented as cyan spheres; E) Results of Gromos clustering analyses: two lipase structures from DPPC simulations (green) representing the top 50% of contributing clusters are overlaid with 5 lipase structures from DPPA simulations (purple) representing the top 35% of contributing clusters.

These results indicate that monolayer surface charges govern large conformational changes in lipase to maximize favorable intermolecular interactions, which in turn impacts the structure of the surface. Since the total embedded surface area given in **Table 2.1** does not change, it is likely that while Van der Waals interactions may impact the partitioning of lipase to an aqueous-organic interface, they only play a small role in its overall stability in monolayers.[72,73] As suggested previously, this is to be expected, given that the hydrocarbon chain length limits the organic phase depth at the phospholipid interface. The unwinding of the lipase in DPPA can be

attributed to significant repulsive electrostatic interactions between the negatively-charged residues on BCL and the negatively-charged phosphate headgroups. The few sites on BCL available for binding to DPPA—charged residues LYS 22, ARG297 and ARG258—coordinate with the phosphate headgroups in the monolayer plane. In DPPC, because there are cationic and anionic BCL residues available for binding, the lipase is stabilized in its open crystal structure conformation, and the phosphate headgroups disperse across the lipase surface, increasing local curvature and decreasing lipid packing order. Since the surfactants in real SSA surfaces are likely to be variable in charge as well as in lipid composition, we postulate that these results will extend to real SSA systems.

We suggest here that the stability of lipase-embedded monolayer surfaces is dictated by electrostatic interactions, and that the interactions induce changes in the folded structure and dynamics of the enzyme, the packing of the lipids, and the overall structure and dynamics of the resulting protein-embedded surface. Because the structure of the enzyme changes with surfactant charge, it is likely that the activity of lipase also changes, making it possible that, depending on details of the SSA surface composition, lipase could differentially act on and transform its surrounding lipid environment while in aerosol form. Furthermore, the dynamics of the surface are likely to modulate multiphase and heterogeneous atmospheric chemistry, where rigidity may enhance the ability of a surface to template into ice crystals, and flexibility could enhance gas or water transport across the surface.[75-78]

2.5 Conclusion

The results of this study suggest that electrostatic interactions between the external, charged amino acids on lipase and the individual charges on the surfactants play a significant role in the stability of lipases at the monolayer surface, as well as in the resulting structural properties of that surface. We present MD studies of BCL embedded in a lipid monolayer as an aqueous-organic interface, where organic phase depth is dependent on the length of the lipid tail, and stability at the interface is dependent on coordinated intermolecular interactions between protein and surfactant. We demonstrate that, while the negative charge on DPPA is not enough to stabilize lipase at high concentrations or high surface pressures, the presence of a zwitterionic headgroup as in DPPC causes a significant increase in the monolayer-lipase lifetime. Although real SSA surfaces are unlikely to

be homogenous matrices of either DPPC or DPPA, they are likely to vary in surface charge distribution based on their lipid profiles. In the context of sea spray aerosol chemistry, where the aerosol surface modulates atmospheric multiphase and heterogeneous reactions, understanding the surfactant-dependent dynamics and fluctuation of protein to and from the surface could bring us closer to accurately depicting SSA properties in atmospheric chemistry and climate models.

2.6 Acknowledgments

This work was supported by the National Science foundation through the Center for Aerosol Impacts on Chemistry of the Environment (CAICE), an NSF funded center for Chemical Innovation (CHE-1801971). We want to thank Dr. Mona Shrestha and Dr. Liubin Huang for helpful discussions on the experimental designs. We would also like to thank Nicholas Wauer for insightful suggestions in the MD analyses. ACD did the computational studies and ML did experimental studies. ACD and ML jointly planned the experiment, discussed analysis, and wrote the paper.

Chapter 2, in full, is a modified reprint of the material as it appears in "Luo, M.† and Dommer, A. C.†; Schiffer, J. M.; Rez, D. J.; Mitchell, A. R.; Amaro, R. E.; Grassian, V. H. *Surfactant Charge Modulates Lipase-Embedded Monolayer Stability at Marine-Relevant Aerosol Surfaces*. Langmuir. 2019, 35 (27), 9050-9060." The dissertation author was a primary co-investigator and co-author of this work.

2.7 Supporting Information

The Supporting Information is available free of charge on the ACS Publications website and contains: The simulation system set-up; simulation convergence analyses, including: dihedral energy, total energy, SASA, and RMSD; production run script; the isotherms of DPPC and DPPA; the surface pressure increase after lipase injected underneath DPPC monolayer with a certain area and a certain starting surface pressure; the IRRAS spectrum of C-H stretching region for DPPC monolayer holding at 15 mN/m and 25 mN/m before and 5 minutes after lipase injection; the structure of DPPC and DPPA; density profiles of DPPC and DPPA phosphates coordinated with sodium ions; percent area change of pure DPPA and DPPC monolayers; lipase labeled; SASA analysis on total surface areas of hydrophobic and polar residues exposed to solvent and lipid; and results of Gromos clustering analysis.

2.7.1 All-Atom MD System Minimization & Equilibration Procedure

The equilibration settings used were unchanged from the default scripts provided by CHARMM-GUI. The equilibration restraints were held on the hydrogen bonds and constrained using the LINCS algorithm with a Verlet cutoff scheme.[90,91] Equilibration steps 1-2 were done under an *NVT* ensemble with temperature-coupling using the Berendsen thermostat for 25 ps each.[92] In step 1, restraints on the atomic positions and dihedral angles were 1000 kJ mol^{-1}; dihedral restraints were lowered to 400 kJ mol^{-1} for step 2. Equilibration steps 3-6 were under an *NPT* ensemble with a Berendsen barostat (subsequent simulations used the Nose Hoover thermostat for accuracy). Restraint forces on step 3 were reduced to 400 kJ mol^{-1} for position and 200 kJ mol^{-1} for dihedrals over 25 ps, then 200 and 200 kJ mol^{-1} for step 4 for 50 ps, 40 and 100 kJ mol^{-1} for step 5 for 50 ps, and finally, 0 and 0 kJ mol^{-1} for step 6 for 50 ps. The system equilibrated for 225 ps total.

2.7.2 Gromos Clustering Analysis

Gromos clustering was implemented with the built-in Gromacs cluster function in Gromacs version 5.0.4. The default clustering settings were used, which includes an RMSD cut-off of 0.1 nm. The RMSD ranges for BCL in DPPC was 0.045 to 0.250, with an average RMSD of 0.009. The RMSD ranges for BCL in DPPA was 0.045 to 0.419, with an average RMSD of 0.013.

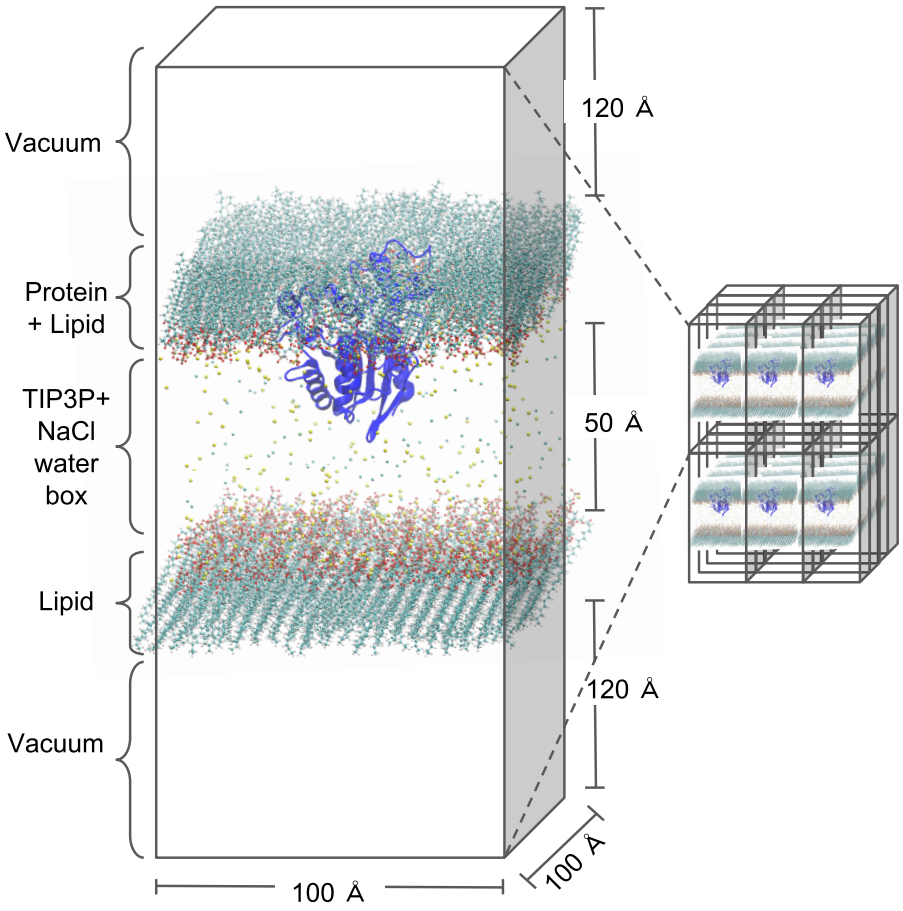


Figure 2.7: Simulation system set-up. The center water block contains TIP3 water molecules and 0.4 M NaCl. The upper leaflet contains either DPPC or DPPA, embedded with a BCL lipase (3LIP). The lower leaflet contains only DPPC or DPPA molecules. There is an average of 47 $Å^2$ per lipid. The system is set up with periodic boundary conditions (given by the boxes on the right).

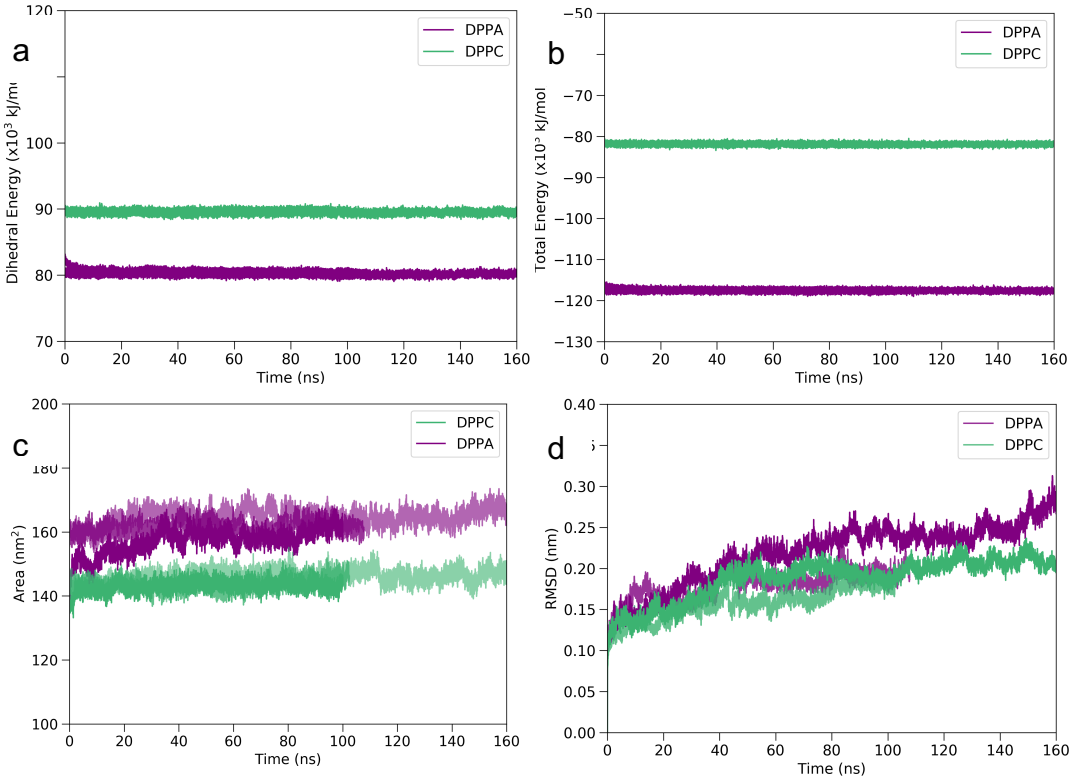


Figure 2.8: Simulation convergence analyses, including: A) dihedral energy, B) total energy, C) solvent-exposed surface area of lipase, and D) RMSD

```
integrator    = md
dt            = 0.002
nsteps        = 50000000
nstlog        = 1000
nstxout       = 5000
nstvout       = 5000
nstfout       = 5000
nstcalcenergy = 100
nstenergy     = 1000
cutoff-scheme = Verlet
nstlist       = 20
rlist         = 1.2
coulombtype   = pme
rcoulomb      = 1.2
vdwtype       = Cut-off
vdw-modifier  = Force-switch
rvdw_switch   = 1.0
rvdw          = 1.2
tcoupl        = Nose-Hoover
```

```
ref_t               = 298.15 298.15 298.15
pcoupl               = no
pcoupltype           = semiisotropic
DispCorr             = EnerPres
tau_p                = 5.0
compressibility      = 4.5e-5  4.5e-5
ref_p                = 1.0     1.0
constraints          = h-bonds
constraint_algorithm = LINCS
continuation         = yes
nstcomm              = 100
comm_mode            = linear
comm_grps           = PROT   MEMB   SOL_ION
```

Figure 2.9: Production run script: step7_production.mdp

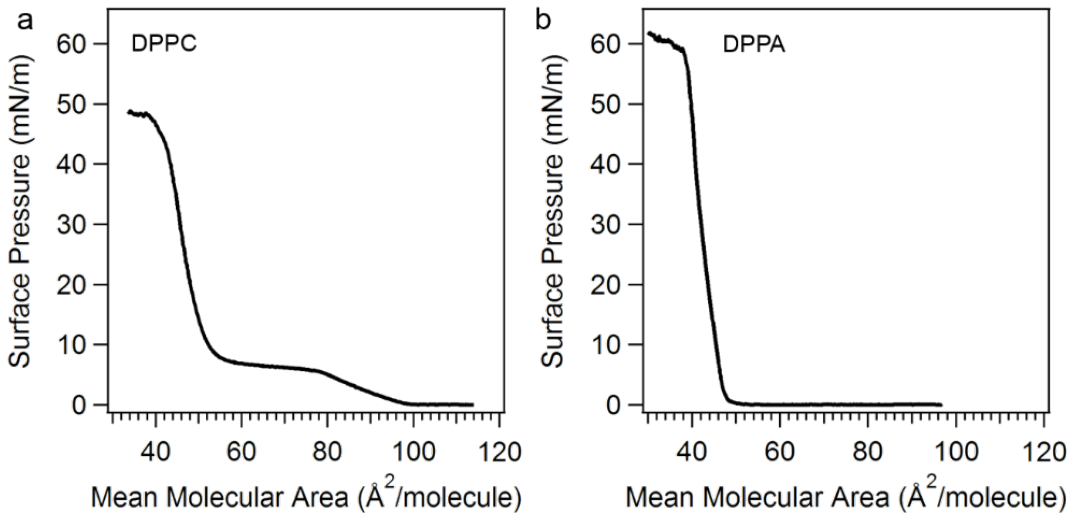


Figure 2.10: Isotherm for DPPC (a) and DPPA (b) on 0.4 M NaCl subphase.

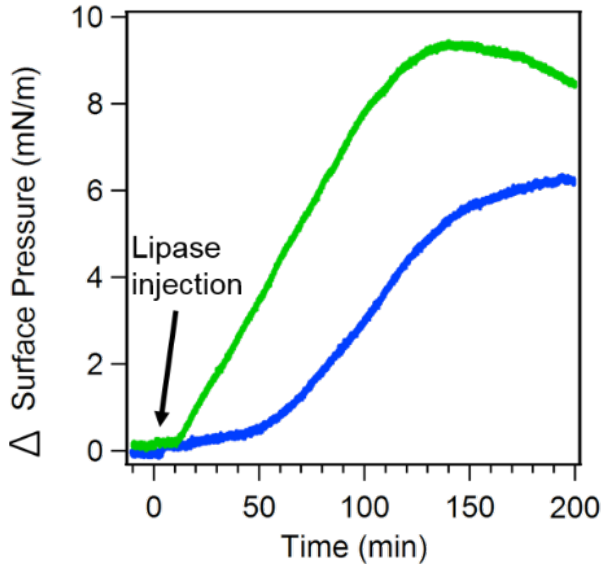


Figure 2.11: Increase of surface pressure vs. time when lipase was injected under a DPPC monolayer at a constant surface area with a starting surface pressure of 15 mN/m (green) and 7.5 mN/m (blue)

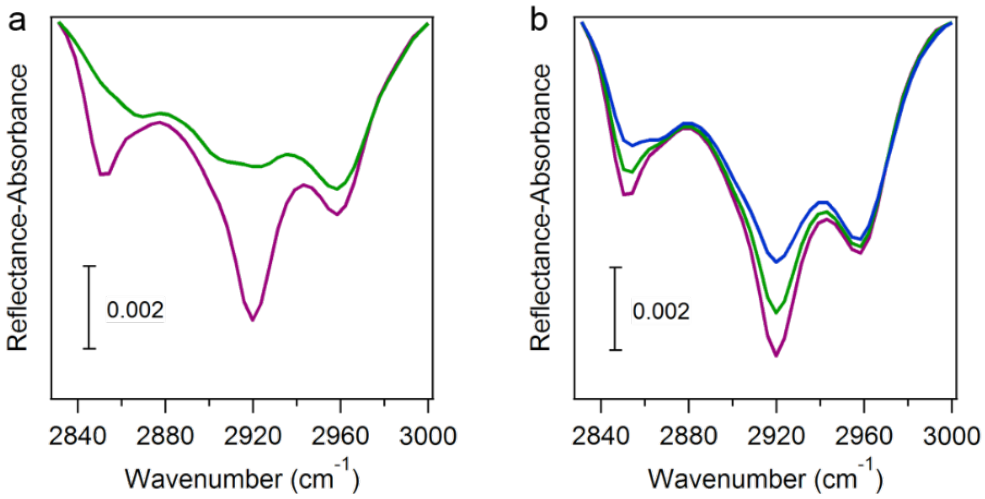


Figure 2.12: IRRAS spectrum of C-H stretching region for a DPPC monolayer while holding at 15 mN/m (a) and 25 mN/m (b) before (purple) and 5 minutes after (green) lipase injection (final concentration 500 nM). We also show the spectra at 30 minutes after lipase injection for the 20 mN/m trial (blue) as the change of spectra in this one is slower than the other. IRRAS spectra later than 5 minutes for (a) and later than 10 minutes for (b) stay almost the same (data not shown).

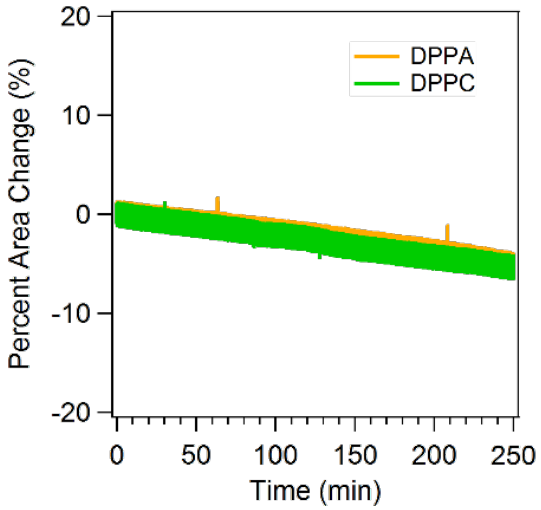


Figure 2.13: Background area decreases when DPPC and DPPA monolayers are held at 25 mN/m surface pressure.

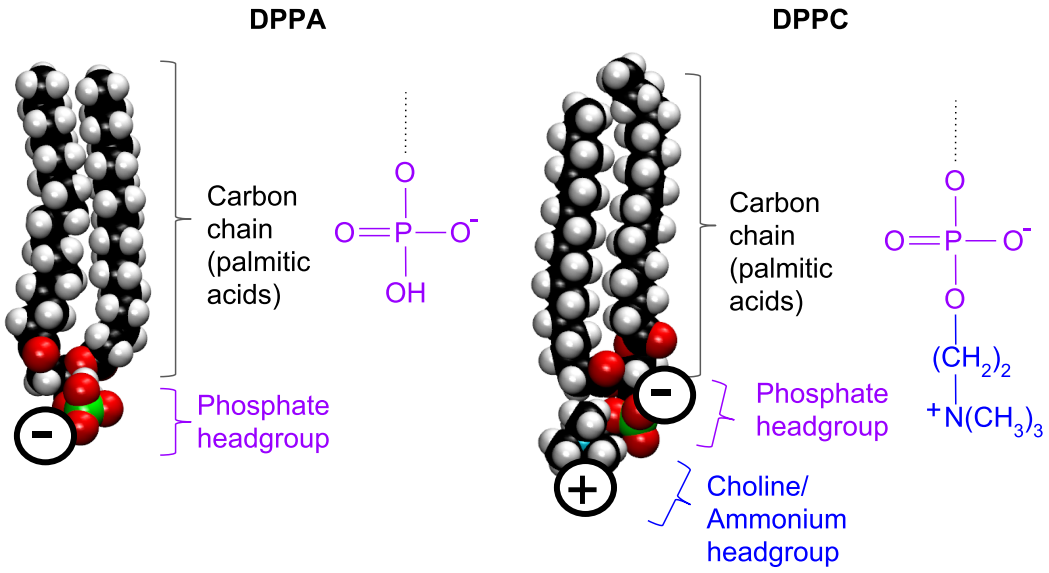


Figure 2.14: DPPA and DPPC chemical structures.

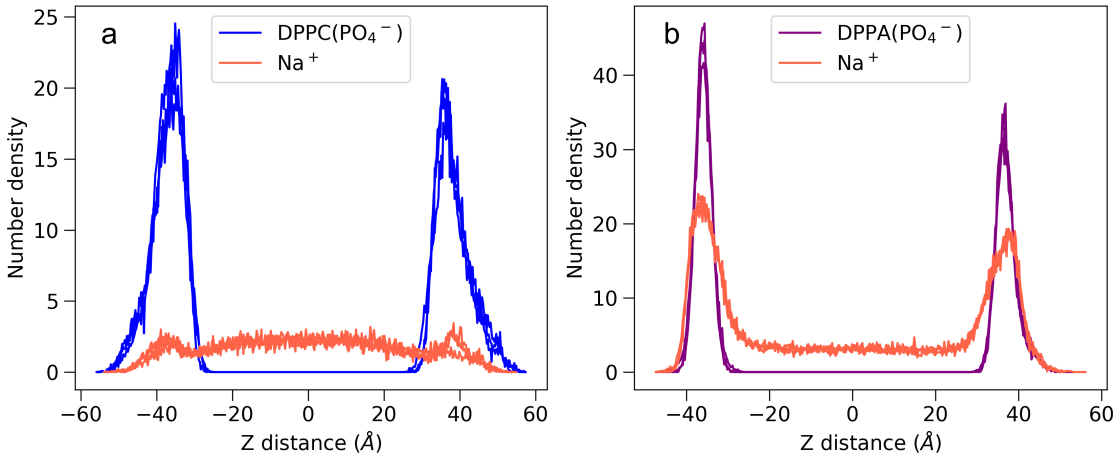


Figure 2.15: Density profiles of a) DPPC and b) DPPA phosphates coordinated with sodium ions.

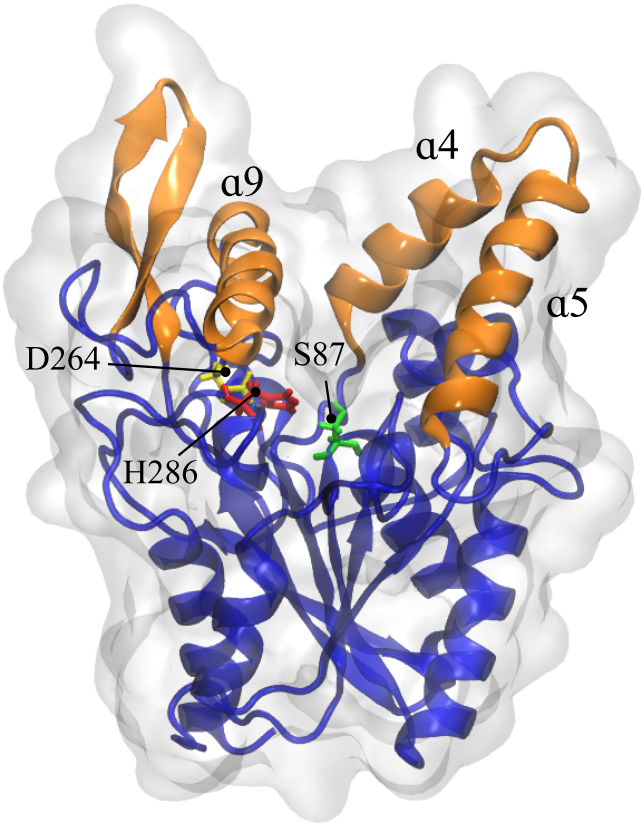


Figure 2.16: Lipase labeled. Hydrophobic lid helices and beta hairpin loop are given in orange. The catalytic triad residues are colored and labeled: D264 (yellow), S87 (green), and H286 (red).

Table 2.2: Surface area analyses on total surface areas on the surfactants either exposed to protein ("Interactive") or exposed to solvent ("Non-Interactive").

Lipid	Initial (nm^2)	Final (nm^2)	Interactive[a] (nm^2)	Non-Interactive[a] (nm^2)
DPPA	1000 ± 10	962 ± 9 (-4%)	51 ± 2	910 ± 10
DPPC	1120 ± 30	1100 ± 30 (-2%)	52 ± 2	1050 ± 30

[a]Lipids are interactive if they are exposed to protein residues or non-interactive if not exposed.

Table 2.3 Results of Gromos clustering analysis. In total, BCL in DPPC produced 33 clusters, while Lipase in DPPA produced 167 clusters. The top 80% of cluster conformations are given.

Lipase-DPPC			Lipase-DPPA					
Cluster ID	Frames	%	Cluster ID	Frames	%	ID	Frames	%
1	13270	43.0	**1**	3627	10.1	**12**	1027	2.9
2	4824	15.6	**2**	2495	7.0	**13**	870	2.4(3)
3	3058	9.9	**3**	2324	6.5	**14**	855	2.3(9)
4	2971	9.6	**4**	2002	5.6	**15**	797	2.2
			5	1819	5.1	**16**	733	2.1
			6	1672	4.7	**17**	680	1.9
			7	1505	4.2	**18**	645	1.8
			8	1346	3.8	**19**	508	1.4
			9	1309	3.7	**20**	423	1.2
			10	1211	3.4	**21**	398	1.1(1)
			11	1174	3.3	**22**	381	1.0(6)

References

(1) Reifenrath, R. Chemical Analysis of the Lung Alveolar Surfactant Obtained by Alveolar Micropuncture. *Respir. Physiol.* **1973**, *19* (1), 35–46. https://doi.org/10.1016/0034-5687(73)90088-1.

(2) Butovich, I. A. Tear Film Lipids. *Exp. Eye Res.* **2013**, *117*, 4–27. https://doi.org/10.1016/j.exer.2013.05.010.

(3) Forestieri, S. D.; Staudt, S. M.; Kuborn, T. M.; Faber, K.; Ruehl, C. R.; Bertram, T. H.; Cappa, C. D. Establishing the Impact of Model Surfactants on Cloud Condensation Nuclei Activity of Sea Spray Aerosols. *Atmospheric Chem. Phys. Discuss.* **2018**. https://doi.org/10.5194/acp-2018-207.

(4) DeMott, P. J.; Mason, R. H.; McCluskey, C. S.; Hill, T. C. J.; Perkins, R. J.; Desyaterik, Y.; Bertram, A. K.; Trueblood, J. V.; Grassian, V. H.; Qiu, Y.; Molinero, V.; Tobo, Y.; Sultana, C. M.; Lee, C.; Prather, K. A. Ice Nucleation by Particles Containing Long-Chain Fatty Acids of Relevance to Freezing by Sea Spray Aerosols. *Environ. Sci. Process. Impacts* **2018**. https://doi.org/10.1039/C8EM00386F.

(5) Collins, D. B.; Ault, A. P.; Moffet, R. C.; Ruppel, M. J.; Cuadra-Rodriguez, L. A.; Guasco, T. L.; Corrigan, C. E.; Pedler, B. E.; Azam, F.; Aluwihare, L. I.; Bertram, T. H.; Roberts, G. C.; Grassian, V. H.; Prather, K. A. Impact of Marine Biogeochemistry on the Chemical Mixing State and Cloud Forming Ability of Nascent Sea Spray Aerosol. *J. Geophys. Res. Atmospheres* **2013**, *118* (15), 8553–8565. https://doi.org/10.1002/jgrd.50598.

(6) Fuentes, E.; Coe, H.; Green, D.; McFiggans, G. On the Impacts of Phytoplankton-Derived Organic Matter on the Properties of the Primary Marine Aerosol – Part 2: Composition, Hygroscopicity and Cloud Condensation Activity. *Atmospheric Chem. Phys.* **2011**, *11* (6), 2585–2602. https://doi.org/10.5194/acp-11-2585-2011.

(7) Cosman, L. M.; Bertram, A. K. Reactive Uptake of N_2O_5 on Aqueous H_2SO_4 Solutions Coated with 1-Component and 2-Component Monolayers. *J. Phys. Chem. A* **2008**, *112* (20), 4625–4635. https://doi.org/10.1021/jp8005469.

(8) Ryder, O. S.; Campbell, N. R.; Shaloski, M.; Al-Mashat, H.; Nathanson, G. M.; Bertram, T. H. Role of Organics in Regulating $ClNO_2$ Production at the Air−Sea Interface. **2015**. https://doi.org/10.1021/jp5129673.

(9) Gord, J. R.; Zhao, X.; Liu, E.; Bertram, T. H.; Nathanson, G. M. Control of Interfacial Cl_2 and N_2O_5 Reactivity by a Zwitterionic Phospholipid in Comparison with Ionic and Uncharged Surfactants. *J. Phys. Chem. A* **2018**, *122* (32), 6593–6604. https://doi.org/10.1021/acs.jpca.8b04590.

(10) Tervahattu, Heikki. Hartonen, Kari. Kerminen, Veli-Matti, Kupiainen, Kaarle. Aarnio, Paivi. Koskentalo, Tarja. Tuck, Adrian and Vaida, Veronica. New Evidence of an Organic Layer on Marine Aerosols. *J. Geophys. Res.* **2002**, *107* (D7), 4053. https://doi.org/10.1029/2000JD000282.

(11) Mccluskey, C. S.; Hill, T. C. J.; Sultana, C. M.; Laskina, O.; Trueblood, J.; Santander, M. V; Beall, C. M.; Michaud, J. M.; Kreidenweis, S. M.; Prather, K. A.; Grassian, V.; Demott, P. J. A Mesocosm Double Feature: Insights into the Chemical Makeup of Marine Ice Nucleating Particles. *J. Atmospheric Sci.* **2018**, *75* (7), 2405–2423.

(12) Shrestha, M.; Luo, M.; Li, Y.; Xiang, B.; Xiong, W.; Grassian, V. H. Let There Be Light: Stability of Palmitic Acid Monolayers at the Air/Salt Water Interface in the Presence and Absence of Simulated Solar Light and a Photosensitizer. *Chem. Sci.* **2018**, *9* (26), 5716–5723. https://doi.org/10.1039/C8SC01957F.

(13) Tang, C. Y.; Huang, Z.; Allen, H. C. Binding of Mg 2+and Ca 2+to Palmitic Acid and Deprotonation of the COOH Headgroup Studied by Vibrational Sum Frequency Generation Spectroscopy. *J. Phys. Chem. B* **2010**, *114* (51), 17068–17076.

(14) Adams, E. M.; Casper, C. B.; Allen, H. C. Effect of Cation Enrichment on Dipalmitoylphosphatidylcholine (DPPC) Monolayers at the Air-Water Interface. *J. Colloid Interface Sci.* **2016**, *478*, 353–364. https://doi.org/10.1016/j.jcis.2016.06.016.

(15) Zhang, T.; Fiamingo, M.; Allen, H. C. Trace Metal Enrichment Driven by Phosphate Functional Group Binding Selectivity. *J. Geophys. Res. Oceans* **2018**, *123* (8), 5286–5297. https://doi.org/10.1029/2018JC013926.

(16) Cochran, R. E.; Laskina, O.; Jayarathne, T.; Laskin, A.; Laskin, J.; Lin, P.; Sultana, C.; Lee, C.; Moore, K. A.; Cappa, C. D.; Bertram, T. H.; Prather, K. A.; Grassian, V. H.; Stone, E. A. Analysis of Organic

Anionic Surfactants in Fine and Coarse Fractions of Freshly Emitted Sea Spray Aerosol. *Environ. Sci. Technol.* **2016**, *50* (5), 2477–2486. https://doi.org/10.1021/acs.est.5b04053.

(17) Michaud, J. M.; Thompson, L. R.; Kaul, D.; Espinoza, J. L.; Richter, R. A.; Xu, Z. Z.; Lee, C.; Pham, K. M.; Beall, C. M.; Malfatti, F.; Azam, F.; Knight, R.; Burkart, M. D.; Dupont, C. L.; Prather, K. A. Taxon-Specific Aerosolization of Bacteria and Viruses in an Experimental Ocean-Atmosphere Mesocosm. *Nat. Commun.* **2018**, *9* (1), 2017. https://doi.org/10.1038/s41467-018-04409-z.

(18) Lee, H. D.; Estillore, A. D.; Morris, H. S.; Ray, K. K.; Alejandro, A.; Grassian, V. H.; Tivanski, A. V. Direct Surface Tension Measurements of Individual Sub-Micrometer Particles Using Atomic Force Microscopy. *J. Phys. Chem. A* **2017**, *121* (43), 8296–8305. https://doi.org/10.1021/acs.jpca.7b04041.

(19) Wang, X.; Sultana, C. M.; Trueblood, J.; Hill, T. C. J.; Malfatti, F.; Lee, C.; Laskina, O.; Moore, K. A.; Beall, C. M.; McCluskey, C. S.; Cornwell, G. C.; Zhou, Y.; Cox, J. L.; Pendergraft, M. A.; Santander, M. V.; Bertram, T. H.; Cappa, C. D.; Azam, F.; DeMott, P. J.; Grassian, V. H.; Prather, K. A. Microbial Control of Sea Spray Aerosol Composition: A Tale of Two Blooms. *ACS Cent. Sci.* **2015**, *1* (3), 124–131. https://doi.org/10.1021/acscentsci.5b00148.

(20) Cochran, R. E.; Laskina, O.; Trueblood, J.; Estillore, A. D.; Morris, H. S.; Jayarathne, T.; Sultana, C. M.; Lee, C.; Lin, P.; Laskin, J.; Laskin, A.; Dowling, J.; Qin, Z.; Cappa, C. D.; Bertram, T. H.; Tivanski, A. V; Stone, E. A.; Prather, K. A.; Grassian, V. H. Molecular Characterization of Sea Spray Particles : Influence of Ocean Biology on Particle Composition and Interaction with Water. *Chem* **2017**, *2*, 655–667. https://doi.org/10.1016/j.chempr.2017.03.007.

(21) Keene, W. C.; Maring, H.; Maben, J. R.; Kieber, D. J.; Pszenny, A. A. P.; Dahl, E. E.; Izaguirre, M. A.; Davis, A. J.; Long, M. S.; Zhou, X.; Smoydzin, L.; Sander, R. Chemical and Physical Characteristics of Nascent Aerosols Produced by Bursting Bubbles at a Model Air-Sea Interface. *J. Geophys. Res. Atmospheres* **2007**. https://doi.org/10.1029/2007JD008464.

(22) Ault, A. P.; Moffet, R. C.; Baltrusaitis, J.; Collins, D. B.; Ruppel, M. J.; Cuadra-Rodriguez, L. A.; Zhao, D.; Guasco, T. L.; Ebben, C. J.; Geiger, F. M.; Bertram, T. H.; Prather, K. A.; Grassian, V. H. Size-Dependent Changes in Sea Spray Aerosol Composition and Properties with Different Seawater Conditions. *Environ. Sci. Technol.* **2013**, *47* (11), 5603–5612. https://doi.org/10.1021/es400416g.

(23) Jayarathne, T.; Sultana, C. M.; Lee, C.; Malfatti, F.; Cox, J. L.; Pendergraft, M. A.; Moore, K. A.; Azam, F.; Tivanski, A. V.; Cappa, C. D.; Bertram, T. H.; Grassian, V. H.; Prather, K. A.; Stone, E. A. Enrichment of Saccharides and Divalent Cations in Sea Spray Aerosol during Two Phytoplankton Blooms. *Environ. Sci. Technol.* **2016**. https://doi.org/10.1021/acs.est.6b02988.

(24) Duncan, S. L.; Larson, R. G. Comparing Experimental and Simulated Pressure-Area Isotherms for DPPC. *Biophys. J.* **2008**. https://doi.org/10.1529/biophysj.107.114215.

(25) Baoukina, S.; Mendez-Villuendas, E.; Tieleman, D. P. Molecular View of Phase Coexistence in Lipid Monolayers. *J. Am. Chem. Soc.* **2012**, *134* (42), 17543–17553. https://doi.org/10.1021/ja304792p.

(26) Habartová, A.; Roeselová, M.; Cwiklik, L. Investigation of Mixed Surfactant Films at Water Surface Using Molecular Dynamics Simulations. *Langmuir* **2015**, *31* (42), 11508–11515. https://doi.org/10.1021/acs.langmuir.5b02854.

(27) Lin, W.; Clark, A. J.; Paesani, F. Effects of Surface Pressure on the Properties of Langmuir Monolayers and Interfacial Water at the Air-Water Interface. *Langmuir* **2015**, *31* (7), 2147–2156. https://doi.org/10.1021/la504603s.

(28) Schiffer, J. M.; Mael, L. E.; Prather, K. A.; Amaro, R. E.; Grassian, V. H. Sea Spray Aerosol: Where Marine Biology Meets Atmospheric Chemistry. **2019**, *19*, 46. https://doi.org/10.1021/acscentsci.8b00674.

(29) Chapus, C.; Rovery, M.; Sarda, L.; Verger, R. Minireview on Pancreatic Lipase and Colipase. *Biochimie* **1988**, *70* (9), 1223–1233. https://doi.org/10.1016/0300-9084(88)90188-5.

(30) Sánchez, D. A.; Tonetto, G. M.; Ferreira, M. L. Burkholderia Cepacia Lipase: A Versatile Catalyst in Synthesis Reactions. *Biotechnol. Bioeng.* **2018**, *115* (1), 6–24. https://doi.org/10.1002/bit.26458.

(31) Svendsen, A. Lipase Protein Engineering. *Biochim. Biophys. Acta BBA - Protein Struct. Mol. Enzymol.* **2000**, *1543* (2), 223–238. https://doi.org/10.1016/S0167-4838(00)00239-9.

(32) Vial, L.; Chapalain, A.; Groleau, M.-C.; Déziel, E. The Various Lifestyles of the *Burkholderia Cepacia* Complex Species: A Tribute to Adaptation. *Environ. Microbiol.* **2011**, *13* (1), 1–12. https://doi.org/10.1111/j.1462-2920.2010.02343.x.

(33) Maravić, A.; Skočibušić, M.; Šprung, M.; Šamanić, I.; Puizina, J.; Pavela-Vrančić, M. Occurrence and Antibiotic Susceptibility Profiles of *Burkholderia Cepacia* Complex in Coastal Marine Environment. *Int. J. Environ. Health Res.* **2012**, *22* (6), 531–542. https://doi.org/10.1080/09603123.2012.667797.

(34) Malfatti, F.; Lee, C.; Tinta, T.; Pendergraft, M. A.; Celussi, M.; Zhou, Y.; Sultana, C. M.; Rotter, A.; Axson, J. L.; Collins, D. B.; Santander, M. V.; Anides Morales, A.; Aluwihare, L. I.; Riemer, N.; Grassian, V. H.; Azam, F.; Prather, K. A. Detection of Active Microbial Enzymes in Nascent Sea Spray Aerosol: Implications for Atmospheric Chemistry and Climate. *Environ. Sci. Technol. Lett.* **2019**. https://doi.org/10.1021/acs.estlett.8b00699.

(35) Aller, J. Y.; Kuznetsova, M. R.; Jahns, C. J.; Kemp, P. F. The Sea Surface Microlayer as a Source of Viral and Bacterial Enrichment in Marine Aerosols. *J. Aerosol Sci.* **2005**, *36* (5–6), 801–812. https://doi.org/10.1016/J.JAEROSCI.2004.10.012.

(36) Schiffer, J. M.; Luo, M.; Dommer, A. C.; Thoron, G.; Pendergraft, M.; Santander, M. V; Lucero, D.; Pecora De Barros, E.; Prather, K. A.; Grassian, V. H.; Amaro, R. E. Impacts of Lipase Enzyme on the Surface Properties of Marine Aerosols. *J Phys Chem Lett* **2018**, *9*, 22. https://doi.org/10.1021/acs.jpclett.8b01363.

(37) Pires De Oliveira, I.; Jara, G. E.; Martí Nez, L. Molecular Mechanism of Activation of Burkholderia Cepacia Lipase at Aqueous-Organic Interfaces. *Phys Chem Chem Phys* **2017**, *19*, 31499. https://doi.org/10.1039/c7cp04466f.

(38) Barbe, S.; Lafaquière, V.; Guieysse, D.; Monsan, P.; Remaud-Siméon, M.; André, I. Insights into Lid Movements of *Burkholderia Cepacia* Lipase Inferred from Molecular Dynamics Simulations. *Proteins Struct. Funct. Bioinforma.* **2009**, *77* (3), 509–523. https://doi.org/10.1002/prot.22462.

(39) Vial, L.; Chapalain, A.; Groleau, M.-C.; Déziel, E. The Various Lifestyles of the *Burkholderia Cepacia* Complex Species: A Tribute to Adaptation. *Environ. Microbiol.* **2011**, *13* (1), 1–12. https://doi.org/10.1111/j.1462-2920.2010.02343.x.

(40) Sánchez, D. A.; Tonetto, G. M.; Ferreira, M. L. Burkholderia Cepacia Lipase: A Versatile Catalyst in Synthesis Reactions. *Biotechnol. Bioeng.* **2018**, *115* (1), 6–24. https://doi.org/10.1002/bit.26458.

(41) Sasso, F.; Natalello, A.; Castoldi, S.; Lotti, M.; Santambrogio, C.; Grandori, R. Burkholderia Cepacia Lipase Is a Promising Biocatalyst for Biofuel Production. *Biotechnol. J.* **2016**, *11* (7), 954–960. https://doi.org/10.1002/biot.201500305.

(42) Eftaiha, A. F.; Wanasundara, S. N.; Paige, M. F.; Bowles, R. K. Exploring the Impact of Tail Polarity on the Phase Behavior of Single Component and Mixed Lipid Monolayers Using a MARTINI Coarse-Grained Force Field. *J. Phys. Chem. B* **2016**, *120* (31), 7641–7651. https://doi.org/10.1021/acs.jpcb.6b03970.

(43) Szőri, M.; Jedlovszky, P.; Roeselová, M. Water Adsorption on Hydrophilic and Hydrophobic Self-Assembled Monolayers as Proxies for Atmospheric Surfaces. A Grand Canonical Monte Carlo Simulation Study. *Phys. Chem. Chem. Phys.* **2010**, *12* (18), 4604. https://doi.org/10.1039/b923382b.

(44) Adams, E. M.; Wellen, B. A.; Thiraux, R.; Reddy, S. K.; Vidalis, A. S.; Paesani, F.; Allen, H. C. Sodium – Carboxylate Contact Ion Pair Formation Induces Stabilization of Palmitic Acid Monolayers at High PH. *Phys. Chem. Chem. Phys.* **2017**. https://doi.org/10.1039/C7CP00167C.

(45) Zhang, T.; Brantley, S. L.; Verreault, D.; Dhankani, R.; Corcelli, S. A.; Allen, H. C. Effect of PH and Salt on Surface PK a of Phosphatidic Acid Monolayers. *Langmuir* **2018**, *34* (1), 530–539. https://doi.org/10.1021/acs.langmuir.7b03579.

(46) Sandhiya, L.; Kolandaivel, P.; Senthilkumar, K. Oxidation and Nitration of Tyrosine by Ozone and Nitrogen Dioxide: Reaction Mechanisms and Biological and Atmospheric Implications. *J. Phys. Chem. B* **2014**. https://doi.org/10.1021/jp4106037.

(47) Liu, F.; Lakey, P. S. J.; Berkemeier, T.; Tong, H.; Kunert, A. T.; Meusel, H.; Cheng, Y.; Su, H.; Fröhlich-Nowoisky, J.; Lai, S.; Weller, M. G.; Shiraiwa, M.; Pöschl, U.; Kampf, C. J. Atmospheric Protein Chemistry Influenced by Anthropogenic Air Pollutants: Nitration and Oligomerization upon Exposure to Ozone and Nitrogen Dioxide. *Faraday Discuss.* **2017**, *200* (0), 413–427. https://doi.org/10.1039/C7FD00005G.

(48) Estillore, A. D.; Trueblood, J. V; Grassian, V. H. Atmospheric Chemistry of Bioaerosols: Heterogeneous and Multiphase Reactions with Atmospheric Oxidants and Other Trace Gases. **2016**. https://doi.org/10.1039/c6sc02353c.

(49) Schiffer, J. M.; Luo, M.; Dommer, A. C.; Thoron, G.; Pendergraft, M.; Santander, M. V; Lucero, D.; Pecora De Barros, E.; Prather, K. A.; Grassian, V. H.; Amaro, R. E. Impacts of Lipase Enzyme on the Surface Properties of Marine Aerosols. *J Phys Chem Lett* **2018**, *9*, 22. https://doi.org/10.1021/acs.jpclett.8b01363.

(50) Marsh, D. Protein Modulation of Lipids, and Vice-Versa, in Membranes. *Biochim. Biophys. Acta BBA - Biomembr.* **2008**, *1778* (7–8), 1545–1575. https://doi.org/10.1016/J.BBAMEM.2008.01.015.

(51) Mouchlis, V. D.; Chen, Y.; McCammon, J. A.; Dennis, E. A. Membrane Allostery and Unique Hydrophobic Sites Promote Enzyme Substrate Specificity. *J. Am. Chem. Soc.* **2018**, *140* (9), 3285–3291. https://doi.org/10.1021/jacs.7b12045.

(52) Mouchlis, V. D.; Bucher, D.; McCammon, J. A.; Dennis, E. A. Membranes Serve as Allosteric Activators of Phospholipase A2, Enabling It to Extract, Bind, and Hydrolyze Phospholipid Substrates. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (6), E516-25. https://doi.org/10.1073/pnas.1424651112.

(53) Adams, E.; Allen, H.; Adams, E. M.; Allen, H. C. Palmitic Acid on Salt Subphases and in Mixed Monolayers of Cerebrosides: Application to Atmospheric Aerosol Chemistry. *Atmosphere* **2013**, *4* (4), 315–336. https://doi.org/10.3390/atmos4040315.

(54) Fridlind, A. M.; Jacobson, M. Z. A Study of Gas-Aerosol Equilibrium and Aerosol PH in the Remote Marine Boundary Layer during the First Aerosol Characterization Experiment (ACE 1). **2000**, *105* (Ace 1).

(55) Dluhy, R. A. *Quantitative External Reflection Infrared Spectroscopic Analysis of Insoluble Monolayers Spread at the Air-Water Interface*; 1986; Vol. 90.

(56) Langmuir, /; Sierra-Hern Andez, M. R.; Allen, H. C. Incorporation and Exclusion of Long Chain Alkyl Halides in Fatty Acid Monolayers at the Air-Water Interface. *Langmuir* **2010**, *26* (24), 18806–18816. https://doi.org/10.1021/la1032674.

(57) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM. *J. Comput. Chem.* **2008**, *29*, 2967–2970. https://doi.org/10.1002/jcc.

(58) Huang, J.; Mackerell, A. D. CHARMM36 All-Atom Additive Protein Force Field: Validation Based on Comparison to NMR Data. *J. Comput. Chem.* **2013**, *34* (25), 2135–2145. https://doi.org/10.1002/jcc.23354.

(59) Marrink, S. J.; Tieleman, D. P. Structure and Dynamics of Lipid Monolayers: Theory and Applications. In *Biomembrane Frontiers: Nanostructures, Models and the Design of Life*; 2009; Vol. 2899311, pp 75–76. https://doi.org/10.1007/978-1-60761-314-5.

(60) Madhavi Sastry, G.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. Protein and Ligand Preparation: Parameters, Protocols, and Influence on Virtual Screening Enrichments. *J. Comput. Aided Mol. Des.* **2013**, *27* (3), 221–234. https://doi.org/10.1007/s10822-013-9644-8.

(61) Schrag, J. D.; Li, Y.; Cygler, M.; Lang, D.; Burgdorf, T.; Hecht, H. J.; Schmid, R.; Schomburg, D.; Rydel, T. J.; Oliver, J. D.; Strickland, L. C.; Dunaway, C. M.; Larson, S. B.; Day, J.; McPherson, A. The Open Conformation of a Pseudomonas Lipase. *Structure* **1997**, *5*, 187–202. https://doi.org/10.2210/PDB3LIP/PDB.

(62) Lomize, M. A.; Pogozheva, I. D.; Joo, H.; Mosberg, H. I.; Lomize, A. L. OPM Database and PPM Web Server: Resources for Positioning of Proteins in Membranes. *Nucleic Acids Res.* **2012**, *40* (D1), D370–D376. https://doi.org/10.1093/nar/gkr703.

(63) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. GROMACS: A Message-Passing Parallel Molecular Dynamics Implementation. *Comput. Phys. Commun.* **1995**, *91* (1–3), 43–56. https://doi.org/10.1016/0010-4655(95)00042-E.

(64) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25. https://doi.org/10.1016/J.SOFTX.2015.06.001.

(65) Towns, J.; Cockerill, T.; Dahan, M.; Foster, I.; Gaither, K.; Grimshaw, A.; Hazlewood, V.; Lathrop, S.; Lifka, D.; Peterson, G. D.; Roskies, R.; Scott, J. R.; Wilkens-Diehr, N. XSEDE: Accelerating Scientific Discovery. *Comput. Sci. Eng.* **2014**, *16* (5), 62–74. https://doi.org/10.1109/MCSE.2014.80.

(66) Nystrom, N. A.; Levine, M. J.; Roskies, R. Z.; Scott, J. R. Bridges: A Uniquely Flexible HPC Resource for New Communities and Data Analytics. In *Proceedings of the 2015 XSEDE Conference on Scientific Advancements Enabled by Enhanced Cyberinfrastructure - XSEDE '15*; 2015; pp 1–8. https://doi.org/10.1145/2792745.2792775.

(67) Pérez, F.; Granger, B. E. IPython: A System for Interactive Scientific Computing Python: An Open and General- Purpose Environment. *Comput. Sci. Eng.* **2007**, *9* (3), 21–29. https://doi.org/doi:10.1109/MCSE.2007.53.

(68) Phillips, M. C.; Chapman, D. Monolayer Characteristics of Saturated 1,2-Diacyl Phosphatidylcholines (Lecithins) and Phosphatidylethanolamines at the Air-Water Interface. *Biochim. Biophys. Acta BBA - Biomembr.* **1968**, *163* (3), 301–313. https://doi.org/10.1016/0005-2736(68)90115-6.

(69) Adams, E. M.; Verreault, D.; Jayarathne, T.; Cochran, R. E.; Stone, E. A.; Allen, H. C. Surface Organization of a DPPC Monolayer on Concentrated SrCl2 and ZnCl2 Solutions. *Phys. Chem. Chem. Phys. PCCP* **2016**, *18* (47), 32345–32357. https://doi.org/10.1039/c6cp06887a.

(70) Hunt, R. D.; Mitchell, M. L.; Dluhy, R. A. The Interfacial Structure of Phospholipid Monolayer Films: An Infrared Reflectance Study. *J. Mol. Struct.* **1989**, *214*, 93–109.

(71) Orendorff, C. J.; Ducey, M. W.; Pemberton, J. E. Quantitative Correlation of Raman Spectral Indicators in Determining Conformational Order in Alkyl Chains. *J Phys Chem A* **2002**, No. 520, 6991–6998. https://doi.org/10.1021/jp014311n.

(72) Snyder, R. G.; Hsu, S. L.; Krimm, S. Vibrational Spectra in the C-H Stretching Region and the Structure of the Polymetheylene Chain. *Spectrochim. Acta* **1978**, *34A*, 395–406.

(73) Wallach, D. F. H.; Verma, S. P.; Fookson, J. Application of Laser Raman and Infrared Spectroscopy to the Analysis of Membrane Structure. *Biochim. Biophys. Acta* **1979**, *559*, 153–208.

(74) Karvaly, B.; Loshchilova, E. Comments on the Quantitative Interpretation of Biomembrane Structure by Raman Spectroscopy. *BBA - Biomembr.* **1977**, *470* (3), 492–496. https://doi.org/10.1016/0005-2736(77)90140-7.

(75) Wei-Ning Huang; Shih-Che Sue; Da-Shin Wang; Po-Long Wu, and; Wu*, W. Peripheral Binding Mode and Penetration Depth of Cobra Cardiotoxin on Phospholipid Membranes as Studied by a Combined FTIR and Computer Simulation Approach†. **2003**. https://doi.org/10.1021/BI0344477.

(76) Meister, A.; Nicolini, C.; Waldmann, H.; Kuhlmann, J.; Kerth, A.; Winter, R.; Blume, A. Insertion of Lipidated Ras Proteins into Lipid Monolayers Studied by Infrared Reflection Absorption Spectroscopy (IRRAS). *Biophys. J.* **2006**, *91* (4), 1388–1401. https://doi.org/10.1529/biophysj.106.084624.

(77) Mendelsohn, R.; Brauner, J. W.; Gericke, A. External Infrared Reflection Absorption Spectrometry of Monolayer Films at the Air-Water Interface. *Annu Rev Phys Chern* **1995**, *46*, 305–339.

(78) Mendelsohn, R.; Mao, G.; Flach, C. R. Infrared Reflection-Absorption Spectroscopy: Principles and Applications to Lipid-Protein Interaction in Langmuir Films. *Biochim. Biophys. Acta* **2010**, *1798* (4), 788–800. https://doi.org/10.1016/j.bbamem.2009.11.024.

(79) Rusdi, M.; Moroi, Y. Study on Water Evaporation through 1-Alkanol Monolayers by the Thermogravimetry Method. *J. Colloid Interface Sci.* **2004**, *272* (2), 472–479. https://doi.org/10.1016/J.JCIS.2004.01.014.

(80) La Mer, V. K.; Healy, T. W.; Aylmore, L. A. G. The Transport of Water through Monolayers of Long-Chain n-Paraffinic Alcohols. *J. Colloid Sci.* **1964**, *19*, 673–684.

(81) McNamee, C. E.; Barnes, G. T.; Gentle, I. R.; Peng, J. B.; Steitz, R.; Probert, R. The Evaporation Resistance of Mixed Monolayers of Octadecanol and Cholesterol. *J. Colloid Interface Sci.* **1998**, *207* (2), 258–263. https://doi.org/10.1006/jcis.1998.5756.

(82) Barnes, G. T.; Bacon, K. J.; Ash, J. M. The Evaporation Resistances of Octadecanol-Cholesterol Mixed Monolayers. *J. Colloid Interface Sci.* **1980**, *76* (1), 263–264. https://doi.org/10.1016/0021-9797(80)90295-7.

(83) Costin, I. S.; Barnes, G. T. The Purification of Long-Chain Alcohols for Monolayer Studies. *J. Colloid Interface Sci.* **1967**, *25*, 584–585. https://doi.org/10.1016/0021-9797(67)90072-0.

(84) Brzozowska, A. M.; Duits, M. H. G.; Mugele, F. Stability of Stearic Acid Monolayers on Artificial Sea Water. *Colloids Surf. Physicochem. Eng. Asp.* **2012**, *407*, 38–48. https://doi.org/10.1016/j.colsurfa.2012.04.055.

(85) Patil, G. S.; Matthews, R. H.; Cornwell, D. G. Kinetics of the Processes of Desorption from Fatty Acid Monolayers. *J. Lipid Res.* **1973**, *14* (1), 26–31.

(86) Basu, S.; Biswas, P. Salt-Bridge Dynamics in Intrinsically Disordered Proteins: A Trade-off between Electrostatic Interactions and Structural Flexibility. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **2018**, *1866* (5–6), 624–641. https://doi.org/10.1016/J.BBAPAP.2018.03.002.

(87) Basu, S.; Mukharjee, D. Salt-Bridge Networks within Globular and Disordered Proteins: Characterizing Trends for Designable Interactions. *J. Mol. Model.* **2017**, *23* (7), 206. https://doi.org/10.1007/s00894-017-3376-y.

(88) Gruia, A. D.; Fischer, S.; Smith, J. C. *Molecular Dynamics Simulation Reveals a Surface Salt Bridge Forming a Kinetic Trap in Unfolding of Truncated Staphylococcal Nuclease*; 2003.

(89) Marcela P. Aliste; Justin L. MacCallum, and; Tieleman*, D. P. Molecular Dynamics Simulations of Pentapeptides at Interfaces: Salt Bridge and Cation−π Interactions. **2003**. https://doi.org/10.1021/BI027001J.

(90) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18* (12), 1463–1472. https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H.

(91) Páll, S.; Hess, B. A Flexible Algorithm for Calculating Pair Interactions on SIMD Architectures. *Comput. Phys. Commun.* **2013**, *184* (12), 2641–2650. https://doi.org/10.1016/j.cpc.2013.06.003.

(92) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81* (8), 3684–3690. https://doi.org/10.1063/1.448118.

Chapter 3

Calcium Bridging Drives Polysaccharide Co-Adsorption to a Proxy Sea Surface Microlayer

3.1 Abstract

Saccharides comprise a significant mass fraction of organic carbon in sea spray aerosol (SSA), but the mechanisms through which saccharides are transferred from seawater to the ocean surface and eventually into SSA are unclear. It is hypothesized that saccharides cooperatively adsorb to other insoluble organic matter at the air/sea interface, known as the sea surface microlayer (SSML). Using a combination of surface-sensitive infrared reflection-absorption spectroscopy and all-atom molecular dynamics simulations, we demonstrate that the marine-relevant, anionic polysaccharide alginate co-adsorbs to an insoluble palmitic acid monolayer via divalent cationic bridging interactions. Ca^{2+} induces the greatest extent of alginate co-adsorption to the monolayer, evidenced by the ~30% increase in surface coverage, whereas Mg^{2+} only facilitates one-third the extent of co-adsorption at seawater-relevant cation concentrations due to its strong hydration propensity. Na^{+} cations alone do not facilitate alginate co-adsorption, and palmitic acid protonation hinders the formation of divalent cationic bridges between the palmitate and alginate carboxylate moieties. Alginate co-adsorption is largely confined to the interfacial region beneath the monolayer headgroups, so surface pressure, and thus monolayer surface coverage, only changes the amount of alginate co-adsorption by less than 5%. Our results provide physical and molecular characterization of a potentially significant polysaccharide enrichment mechanism within the SSML.

3.2 Introduction

Saccharides constitute one of the most abundant classes of organic carbon in sea spray aerosol (SSA);[1–3] but the mechanisms through which these compounds are selectively transferred from bulk seawater to the ocean surface, known as the sea surface microlayer (SSML), and finally to SSA, are not fully understood.[4] Compared with seawater concentrations, saccharides are enriched 1.2–12.1-fold in the SSML,[5–7] 38–3700-fold in super-micron SSA,[8] and 100–930,000-fold in sub-micron SSA.[8,9] It is thought that soluble saccharides co-adsorb to insoluble organic films at the SSML and transfer into SSA via bubble bursting at the ocean surface.[10–14] Chemical

composition is a significant driver of SSA particle radiative properties, so climate models require predictive representations of marine aerosol composition to accurately model climate processes in the marine boundary layer.[10,15–18] SSA containing polysaccharides, especially polysaccharides within marine microgels, comprise a significant fraction of cloud condensation nuclei (CCN)[19–22] and ice nucleating particles (INPs),[23–30] thereby affecting cloud formation and albedo. Thus, characterization of polysaccharide enrichment mechanisms in SSA is imperative for improving aerosol representation in global climate models.

Recent work by Hasenecz e*t al.* has shown that the polysaccharide alginate is enriched in laboratory-generated marine aerosol, and alginate enrichment can be enhanced upon adding protein and additional $CaCl_2$ salt to the model seawater solution.[31] Alginate is a type of exopolymeric substance derived from marine brown algae and bacteria;[32,33] it is composed of (1→4)-linked α-L-guluronic (G) and β-D-mannuronic (M) monomers that form a block copolymer with random sequences of M-, G-, and MG-blocks.[34–39] Alginate polymers undergo ionic cross-linking to form hydrophilic gels via metal ion coordination primarily to the G residue carboxylic acid moieties.[36,39,40] The rigidity and stability of the hydrogels are largely driven by the M/G residue ratio and the molecular weight of the polymers,[39,41,42] but the ionic cross-linker identities and concentrations also play a crucial role.[43–48] Multivalent cations readily induce gelation, and the cation hydration free energies dictate the local interaction configurations with G residues.[48] Of the abundant seawater cations, Ca^{2+} is the most efficient gelling agent.[43,44,48] Mg^{2+} only induces gelation at high alginate concentrations (> 3 wt% alginate and $[Mg^{2+}] > 50$ mM) because of its high affinity for water such that the G residue carboxylate moieties cannot readily dehydrate Mg^{2+} for coordination.[45–48] Extensive experimental and computational studies[49–54] have investigated polysaccharide aggregation into surface-active marine gels via ionic coordination, but less is known about polysaccharide complexation to other surface-active organic molecules derived from marine biota. Enhanced saccharide enrichment in laboratory-generated SSA in the presence of divalent cations and other surface-active organic material strongly suggests a co-adsorption mechanism mediated by divalent cationic bridging.[12,31]

A divalent cation mediated co-adsorption mechanism was also postulated by Schill *et al.* to explain enrichment of the monosaccharide glucuronic acid in laboratory-generated SSA.[12] Glucuronic acid likely co-adsorbs to an insoluble palmitic acid (hexadecanoic acid, $CH_3(CH_2)_{14}COOH$) monolayer via seawater divalent cationic bridging interactions. Palmitic acid is one of the most abundant lipids in the SSML[55,56] and in nascent

SSA,[55,57,58] making it a good model for insoluble organic matter in an SSML proxy film.[59–62] Cooperative adsorption (co-adsorption) to insoluble lipid monolayers has been indirectly observed for other saccharides and polysaccharides as well.[10,11,13,63] Electrostatic interactions between charged saccharides and either charged or zwitterionic lipid headgroups have been the predominant mechanism of co-adsorption proposed. For example, the cationic polysaccharide chitosan primarily interacts with negatively charged and zwitterionic phospholipids through electrostatic interactions between the chitosan ammonium and phospholipid phosphate moieties.[64–69] Chitosan co-adsorption expands the monolayer,[70] and chitosan likely intercalates into the monolayer at low mean molecular area through dispersive interactions.[68,69] Divalent cationic bridging between the anionic polysaccharide dextran sulfate and the zwitterionic phospholipids 1,2-dipalmitoyl-phosphatidylethanolamine (DPPE) and 1,2-dimyristoyl-sn-glycero-3-phosphocholine (DMPC) has also been measured.[71–73] Ca^{2+} bridges the sulfate and phosphate moieties, whereas Na^{+} does not.[72] Strength of dextran sulfate co-adsorption increases with the number of calcium bridges formed, and unlike chitosan, dextran sulfate does not intercalate into the monolayer; instead, the authors argued that strong bridging interactions tether the polysaccharide to the underside of the phosphate headgroups.[73] Thus, divalent cationic bridging interactions could promote polysaccharide co-adsorption to the SSML and enrichment in SSA.

In this study, we examine the cationic bridging mechanism responsible for alginate co-adsorption to a deprotonated palmitic acid monolayer using surface-sensitive infrared reflection-absorption spectroscopy (IRRAS) and molecular dynamics simulations. In this work, we directly observe polyelectrolyte adsorption to an insoluble monolayer of the same charge state bridged by an ion of opposite charge. Ca^{2+} induces the greatest degree of alginate co-adsorption to the monolayer, but the divalent cationic bridges break apart upon palmitic acid protonation. Mg^{2+} also promotes co-adsorption to the deprotonated monolayer, but the interaction is much weaker due to the strong hydration of Mg^{2+}. Na^{+} alone does not facilitate alginate co-adsorption. Our detailed experimental and computational analyses of the alginate co-adsorption mechanism to an SSML proxy film suggest that the Ca^{2+} bridge to surface-active marine organic matter is an important driver of polysaccharide enrichment in the SSML, and thus, in SSA.

3.3 Experimental and Computational Methods

3.3.1 Materials

Calcium chloride dihydrate (Certified ACS, Fisher Chemical), magnesium chloride hexahydrate (Crystalline/Certified ACS, Fisher Chemical), d_{31}-palmitic acid (98%, Cambridge Isotope Laboratories, Inc.), and d_{33}-cetyl alcohol (d_{33}-hexadecanol, 98%, Cambridge Isotope Laboratories, Inc.) were used as received. Sodium chloride (99.5%, for biochemistry, ACROS Organics™) was baked for at least 10 hours in a furnace at 650° C to remove residual organic impurities.[74] Alginic acid (sodium salt, ACROS Organics™, Lot: A0406891) was also used as received from the same batch to maximize homogeneity in molecular weight and composition. All aqueous solutions were prepared with ultrapure water (18.2 MΩ·cm, Milli-Q Advantage A10, EMD Millipore) in glassware cleaned in a piranha acid bath. Acidification by atmospheric CO_2 causes the ultrapure water pH to be 5.8 ± 0.1; hence, the pH 5.8 aqueous solutions were not pH adjusted. Atmospheric CO_2 acidifies the solutions at seawater pH too, so the solutions were initially pH adjusted with sodium hydroxide pellets (98%, extra pure, ACROS Organics™) to 8.6 ± 0.1 to ensure that the pH would be 8.2 ± 0.1 throughout spectral acquisition. Lipids were dissolved in chloroform (Reagent ACS, 99.8+%, ACROS Organics™) to prepare ~1.25 mM solutions for spreading onto the aqueous surface.

3.3.2 Surface Pressure-Area Isotherms

Surface pressure – area isotherms were performed using the Wilhelmy plate method in a Teflon Langmuir trough with Delrin barriers (Biolin Scientific). The platinum Wilhelmy plate, trough, and barriers were cleaned thoroughly with ultrapure water and ethanol, and the Wilhelmy plate was fired with a Bunsen burner until red hot. Surface cleanliness of the aqueous solution was assessed by sweeping the barriers at maximum compression speed (270 mm/min/barrier) to ensure that the surface pressure did not rise above 0.20 mN/m. d_{31}-Palmitic acid dissolved in chloroform was spread dropwise onto the aqueous surface using a glass syringe (Hamilton), and the chloroform solvent was allowed to evaporate over 10 minutes. The barriers were symmetrically compressed at a rate of 5 mm/min/barrier, and constant surface pressure was maintained during spectroscopic measurements via slow barrier position fluctuations (1 mm/min/barrier).

3.3.3 Infrared Reflection-Absorption Spectroscopy

Infrared-reflection absorption spectroscopy was conducted using an in-lab-built optical setup. Two planar gold mirrors positioned at 48° relative to surface normal were placed inside a Fourier transform infrared spectrometer (Spectrum 100, PerkinElmer) to direct the incident beam towards the aqueous surface and to direct the reflected light towards the liquid nitrogen-cooled HgCdTe (MCT) detector. Spectra were collected with unpolarized light as an average of 400 scans in the single-beam mode, and the spectral resolution was set to 4 cm^{-1}. Energy values were recorded between 450 and 4000 cm^{-1} in 0.5 cm^{-1} increments. Experiments were repeated in at least triplicate, and each spectrum was reported as the average of all trials. Spectral background subtraction and peak integration were performed using OriginPro 9.0. The OH stretching region was analyzed by fitting a line between endpoints 2985 and 3000 cm^{-1} for baseline subtraction, and the area under the curve was integrated between these endpoints. The IRRAS spectra and integrated peak areas were numerically corrected to account for differences in monolayer mean molecular area (MMA) between experiments. The average MMA value of the solution containing alginate was divided by the corresponding average MMA value of the salt water solution. This ratio was then multiplied into the reflectance-absorbance values of the spectra containing alginate. Similar spectral analyses were performed in the COOH stretching (1150-1850 cm^{-1}) and CD_2 scissoring mode (1070-1110 cm^{-1}) regions, and descriptions of the peak fitting procedures can be found in Section 4.7. All spectra and data points represent averages of at least three measurements, and error bars represent one standard deviation from the mean.

3.3.4 Computational Methods

Explicit solvent all-atom molecular dynamics simulations were performed using NAMD[75] at a temperature of 298.15 K. The initial configurations of each system were constructed using Packmol.[76] To approximate a TC palmitic acid monolayer at pH 5.6, palmitic acid molecules were packed into two planar 25 Å x 25 Å monolayer slabs, each containing 96 molecules. A pH of 8.2 was approximated by packing 46 molecules each of palmitate and palmitic acid into monolayer slabs of the same dimensions. The monolayer slabs were then separated on one side by vacuum and the other side with a 60 Å-deep water box with 15 proxy alginate molecules (Figure 4.9). The initial structure for the alginate was constructed using the CHARMM-GUI glycan modeler[77–80] and consists of 6 (1→4)-β-linked M monomers. For each representative pH, the aqueous

phase either contained 0.47 M NaCl + 10 mM $CaCl_2$ or 0.47 M NaCl only. For the systems containing Ca^{2+}, all negative charges were balanced using random placement of calcium ions and otherwise balanced with sodium. Identical systems without alginate were also constructed as controls. TIP3P water[81] was used in conjunction with the CHARMM36 forcefield.[82] The system was simulated using constant number, volume and temperature (NVT) with temperature controlled via Langevin dynamics[83] with a coupling constant of 1 ps. The equations of motion were integrated with a timestep of 2 fs, with long-range electrostatics calculated every other timestep using the Particle Mesh Ewald method with a cutoff of 12 Å.[84] A correction was applied to prevent center of mass drift due to PME by enabling the zeroMomentum parameter. The systems were energy-minimized and equilibrated for 800 ps. Production simulations consisting of 200-ns trajectories each were then performed in triplicate for a total of 600 ns per system. The production run simulation length of 200 ns was chosen to allow for any long timescale diffusion and settling of ions and alginate. All simulations were performed on an AgilityValue Tower AMD ROME Workstation (Applied Data Systems Inc., Poway, CA) equipped with an NVIDIA GeForce RTX 3090 Graphics Card (Santa Clara, CA).

Analyses of the MD trajectories were performed in a Jupyter Notebook iPython environment. The density and radial distribution function (RDF) calculations were performed with pytraj.[85,86] Lipid order parameters were calculated using the GROMACS *gmx order* tool from GROMACS version 2018.3.[87]

3.4 Results and Discussion

The mechanism of alginate co-adsorption to a d_{31}-palmitic acid ($CD_3(CD_2)_{14}COOH$) monolayer was investigated experimentally as a function of solution ionic composition, d_{31}-palmitic acid protonation state, and surface pressure. We will use the terms “monolayer” to refer to the monomolecular layer of d_{31}-palmitic acid alone and “film” to describe alginate complexed to the d_{31}-palmitic acid monolayer. Deuterated palmitic acid was used to spectrally isolate the CD_2 scissoring mode from the carboxylate stretching region. The ionic composition was selected to model concentrations of the most abundant cations in seawater: 0.47 M NaCl, 10 mM $CaCl_2$, and 53 mM $MgCl_2$.[88,89] As a direct comparison to the 10 mM $CaCl_2$ solution, alginate co-adsorption in 10 mM $MgCl_2$ aqueous solution was also measured. All aqueous solutions have a background of 0.47 M NaCl to maintain high ionic strength as in seawater.

The carboxylic acid protonation state was varied through the solution pH values of 8.2 and 5.8. Palmitic acid has a reported surface pK_a between 8.34 and 8.7,[60,90–92] and the pK_a values of alginate G and M residues are 3.7 and 3.4, respectively.[93] Thus at pH 8.2, the pH of seawater,[94] palmitic acid is partially deprotonated and alginate is fully deprotonated. At pH 5.8, palmitic acid is mostly protonated, and alginate carboxylate groups remain deprotonated. Salts can deprotonate carboxylic acids at pH values significantly below the pK_a,[61,95,96] however, so a d_{33}-cetyl alcohol (d_{33}-hexadecanol, $CH_3(CH_2)_{15}OH$) monolayer was used as a control to study alginate co-adsorption to a fully protonated monolayer. Alginic acid is insoluble in water, so a lower pH value was not tested instead to avoid significant changes in solubility. Alginate hydrolysis kinetics are also enhanced in acidic solution, thereby changing the alginate molecular weight distribution in solution over time.[97,98] Deprotonated palmitate molecules are slightly more soluble in water,[60,61] so IRRAS spectra were maintained and collected at constant surface pressure to minimize any interpretation impacts from dissolution. The surface pressure values of 5 mN/m and 25 mN/m were chosen to represent the tilted condensed (TC) and untilted condensed (UC) two-dimensional monolayer phases, respectively.[99] Both phases exhibit long-range lateral order in the alkyl tails, but the d_{31}-palmitic acid molecules in the UC phase are closer together on average (more tightly packed) and less tilted relative to surface normal.[100]

To experimentally measure co-adsorption of alginate to the d_{31}-palmitic acid monolayer, IRRAS spectra were analyzed in the OH-stretching region (2985-3800 cm^{-1}), the COOH stretching region (1150-1850 cm^{-1}), and the CD_2 scissoring mode region (1070-1110 cm^{-1}). IRRAS spectra are plotted as reflectance-absorbance (RA) given by Eq. 3.1. Hence, the signal from the interfacial film is captured, whereas the signal from the bulk aqueous phase is subtracted. When the film reflectance is greater than the solution reflectance ($R/R_0 > 1$), the reflectance-absorbance values are negative. Conversely, if the reflected signal from the solution is greater than that of the film ($R/R_0 < 1$), then the reflectance-absorbance values are positive. To isolate the signal from alginate co-adsorption due to the presence of the d_{31}-palmitic acid monolayer, alginate was also included in the aqueous solution spectrum (R_0). Yet, the contribution of alginate adsorbed to the air-water interface due to surface activity alone is small and below our spectroscopic limit of detection.

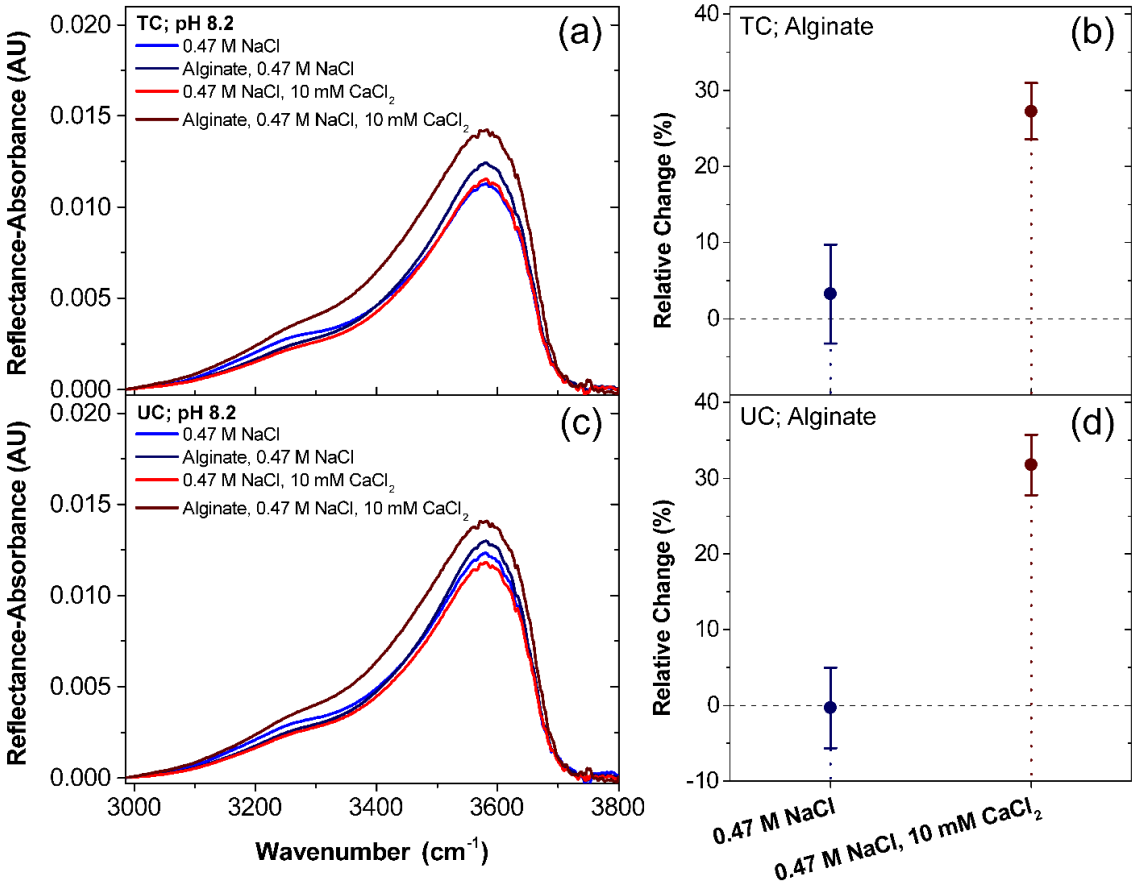


Figure 3.1: IRRAS spectra of the OH stretching region and the corresponding relative changes in integrated peak area of the film indicate that $CaCl_2$ induces significant alginate co-adsorption to the d_{31}-palmitic acid monolayer. Data points and error bars are color-coded to indicate differences in solution composition. Surface pressure was held constant in the (a) tilted condensed (5 mN/m) and (c) untilted condensed (25 mN/m) phases throughout spectral acquisition. The relative changes in integrated peak area between the solutions with and without alginate are quantified in the (b) tilted condensed and (d) untilted condensed phases. Positive relative change indicates alginate co-adsorption to the monolayer, and 0% relative change indicates no co-adsorption.

3.4.1 Experimental Evidence of Cation-Specific Alginate Co-Adsorption Mechanisms

The OH stretching region provides sensitive detection of the aqueous solution reflectance, making the spectra useful for quantifying changes in interfacial coverage as a function of alginate co-adsorption. As shown in **Figures 3.1a** and **3.1c**, the IRRAS OH-stretching modes are positive in sign, indicating that the OH region reflectance decreases upon spreading d_{31}-palmitic acid onto the aqueous solution. The magnitude of the reflectance-absorbance signal intensity also increases with compression of the monolayer to higher surface pressure. Thus, increasing organic surface coverage decreases reflectance from the underlying aqueous solution.

Consequently, alginate co-adsorption to d_{31}-palmitic acid molecules increases surface coverage, and the OH stretching region reflectance-absorbance values increase with alginate co-adsorption to the film. If there is no alginate co-adsorption, then the film spectrum with alginate in the solution directly overlaps the monolayer spectrum without alginate.

Figures 3.1a and **3.1c** illustrate changes in the OH stretching region spectra caused by increasing organic surface coverage due to alginate co-adsorption. The spectrum of d_{31}-palmitic acid spread onto an aqueous solution containing 50 ppm alginate, 0.47 M NaCl, and 10 mM $CaCl_2$ is significantly enhanced relative to d_{31}-palmitic acid spread onto the salt water solution. IRRAS signal corresponding to the NaCl and alginate solution is slightly greater relative to the NaCl solution alone, but the differences are much smaller in the absence of $CaCl_2$. The IRRAS film spectrum corresponding to alginate co-adsorption closely follows the shape of the monolayer spectrum on salt water solution, and the greatest OH stretching region signal enhancement occurs around ~3580 cm^{-1} which has been assigned to surface water molecules hydrogen bonded to a carboxylic acid group.[90,101,102] Therefore, alginate co-adsorption enhances the signal around 3580 cm^{-1} due to the alginate carboxylate hydration.

To better quantify organic surface coverage as a function of alginate co-adsorption, the MMA-corrected OH region was integrated between 2985 and 3800 cm^{-1}. The relative change in the integrated peak areas between the spectra with and without alginate was calculated using the following formula:

$$\text{Relative Change} = \frac{(A_{\text{Alginate}} - A_{\text{SW}})}{A_{\text{SW}}} \times 100\%. \qquad \text{Eq 3.1}$$

In this equation, A_{Alginate} represents the integrated peak area of the spectrum with alginate in the solution, and A_{SW} represents the integrated peak area of the spectrum containing only salt water (no alginate) in the solution. Alginate co-adsorption corresponds to a positive relative change in surface coverage, and no adsorption results in a 0% relative change. While the 0.47 M NaCl solution might induce some alginate co-adsorption to the TC monolayer, as shown by the small but positive relative change value in Figure 4.1b, the value is not statistically significant given that the error (represented as one standard deviation from the mean) is larger than the average relative change value. Addition of 10 mM $CaCl_2$ to the solution containing alginate induces a significant increase in the integrated peak area and consequently a 27% relative change in surface coverage. This increase suggests

an alginate co-adsorption mechanism of divalent cationic bridging between alginate and d_{31}-palmitic acid carboxylate moieties.

The differences between the TC and UC spectra and associated relative changes in surface coverage provide insights into the co-adsorption mechanism. The UC film spectrum corresponding to the 50 ppm alginate and 0.47 M NaCl solution nearly overlaps with the UC monolayer spectrum corresponding to the 0.47 M NaCl solution (**Figure 3.1c**), suggesting little to no alginate co-adsorption. The average relative difference in integrated peak area of the UC monolayer and film spectra is approximately 0% (**Figure 3.1d**), indicating a lack of alginate co-adsorption. Additionally, the slightly larger relative change in the TC spectrum surface coverage suggests that alginate expands the monolayer. It is possible that repulsive electrostatic interactions between the d_{31}-palmitic acid and alginate carboxylate moieties expand the monolayer, consequently increasing surface coverage in the TC phase. However, in the UC phase, increased dispersion interactions between the alkyl tails counteract the repulsive forces. This hypothesis is further supported by greater MMA expansion with alginate in the solution for the TC film relative to the UC film (see **Table 3.1**). Higher surface pressure increases alginate co-adsorption for the solution containing 10 mM $CaCl_2$ to a 32% relative change in surface coverage, and film expansion in the presence of alginate is not significantly different between the TC and UC phases. Thus, alginate likely does not intercalate between the d_{31}-palmitic acid molecules with $CaCl_2$ in the solution, and the UC film possibly provides more ideally spaced sites for the Ca^{2+} bridges.

4.4.2 Monolayer Protonation Impacts on Alginate Co-Adsorption

Palmitic acid protonation state significantly impacts the extent of alginate co-adsorption. For the d_{31}-palmitic acid monolayer at pH 5.8, there is a reduced degree of alginate co-adsorption to the film in comparison to the chemical system at pH 8.2. The d_{31}-palmitic acid TC and UC spectra corresponding to the solution containing alginate have higher reflectance-absorbance values across the entire OH stretching region (**Figures 3.2a** and **3.2c**), indicative of alginate co-adsorption. The relative changes in surface coverage presented in **Figures 3.2b** and **3.2d** quantify this observation, in which alginate co-adsorption at pH 5.8 results in a 14% and 9% increase in the integrated peak area for the TC and UC phase spectra, respectively. At pH 8.2, alginate co-adsorption results in a 27% and 32% integrated peak area increase for the TC and UC phase spectra, respectively.

Therefore, increased d_{31}-palmitic acid protonation decreases the extent of alginate co-adsorption. Unlike the interfacial film at pH 8.2, monolayer compression decreases the magnitude of alginate co-adsorption for the solution at pH 5.8. Alginate might be expanding the TC film similarly to the mechanism proposed for the solution containing only 0.47 M NaCl.

A fully protonated d_{33}-cetyl alcohol monolayer was examined as a control because it does not become partially deprotonated in the presence of salts at any of the pH values tested. Trends in the IRRAS spectra of d_{33}-cetyl alcohol alone are less clear (**Figures 3.2a** and **3.2c**). In both monolayer phases, the alginate solution enhances IRRAS signal at ~3580 cm^{-1} and decreases the signal between ~3050 and 3440 cm^{-1} relative to the salt water solution. The region around 3050 cm^{-1} has been attributed to cyclic carboxylic acid dimers observed in polyacrylic acid solutions,[102,103] and the region from ~3200 cm^{-1} to ~3400 cm^{-1} corresponds to a more ordered hydrogen bonding structure to a less-ordered water structure.[101,104,105] Further analysis reveals no significant differences between the OH region integrated peak areas for the solutions with and without alginate in both the TC and UC phases, resulting in a 0% relative change in surface coverage (**Figures 3.2b** and **3.2d**). It is likely that alginate perturbs the interfacial hydration structure but does not co-adsorb to the cetyl alcohol monolayer. Thus, a protonated monolayer blocks alginate co-adsorption, indicating the importance of electrostatic interactions in polysaccharide co-adsorption to an SSML proxy film.

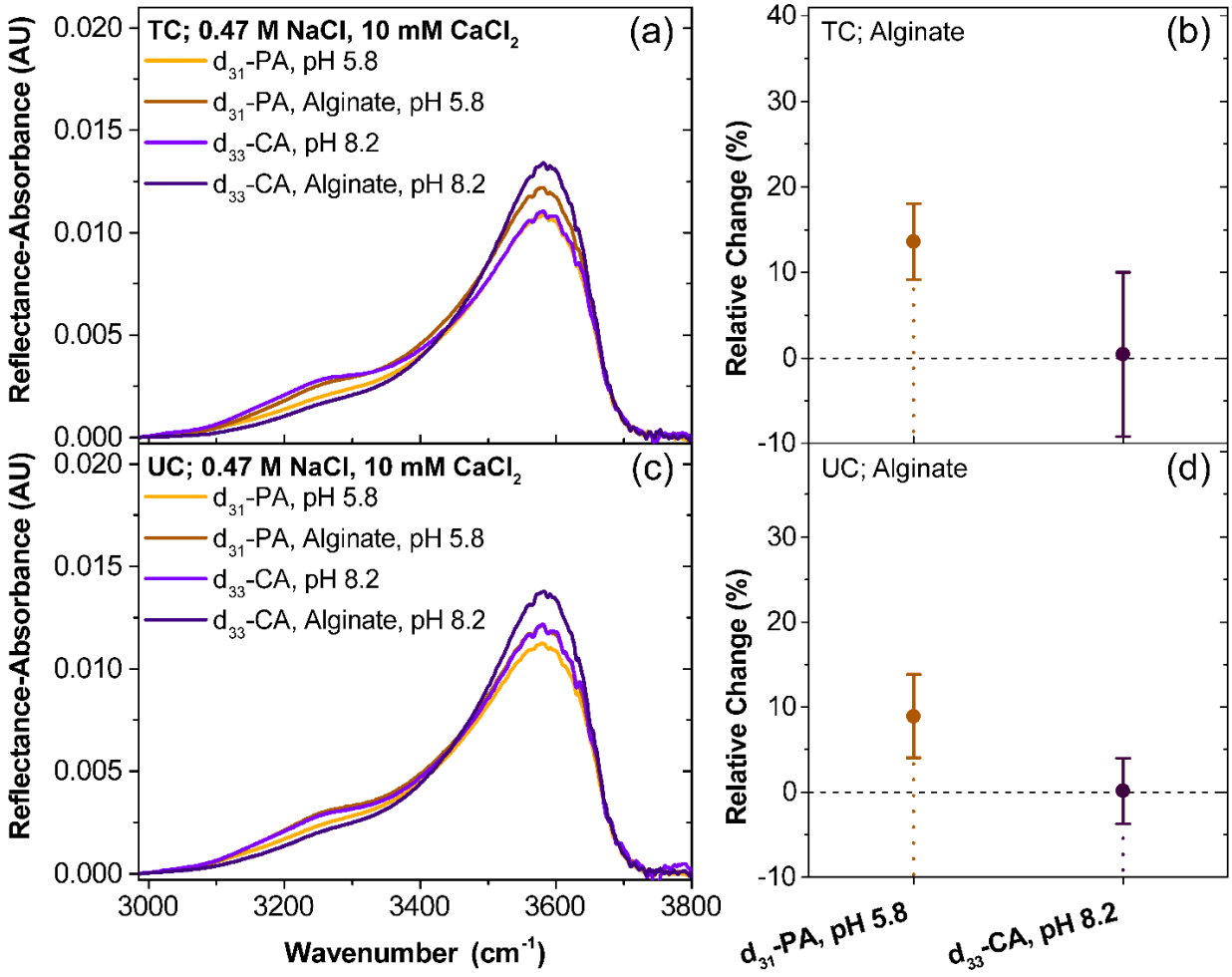


Figure 3.2: IRRAS spectra of the OH stretching region and the corresponding relative changes in integrated peak area of d_{31}-palmitic acid (d_{31}-PA) and d_{33}-cetyl alcohol (d_{33}-CA) monolayers show that headgroup protonation prevents alginate co-adsorption. Data points and error bars are color-coded to indicate differences in monolayer and solution composition. Surface pressure was held constant in the (a) tilted condensed (5 mN/m) and (c) untilted condensed (25 mN/m) phases throughout spectral acquisition. (Note that the light purple and dark brown data curves overlap in (c).) The relative changes in integrated peak area between the solutions with and without alginate are quantified in the (b) tilted condensed and (d) untilted condensed phases. Positive relative change indicates alginate co-adsorption to the monolayer, and 0% relative change indicates no co-adsorption.

The carboxylic acid spectral region (1150-1850 cm^{-1}) provides further insight into the d_{31}-palmitic acid protonation state and the extent of alginate co-adsorption. Harmonic vibrational frequency calculations were performed to predict the relative frequency shifts between the d_{31}-palmitic acid carboxylate and carboxylic acid modes and the alginate carboxylate modes (see **Section 3.7**). Additionally, peaks were fitted to Gaussian functions to resolve the center wavelengths and full width at half maximum (FWHM) values; the fitting procedure and summary of the Gaussian fits for each spectrum are summarized in **Section 3.7**. Spectra of a d_{31}-

palmitic acid TC (**Figure 3.3a**) and UC (**Figure 3.3c**) monolayer exhibit four negative peaks corresponding to the lipid carboxylic acid moieties and one positive band at ~1660 cm^{-1} corresponding to the water bending mode (δ H-O-H). The protonated carboxylic acid moiety is characterized by the C-OH stretching mode (ν C-OH, ~1270 cm^{-1}) and the C=O stretching mode (ν C=O, ~1720 cm^{-1}). Deprotonation of the carboxylic acid leads to the appearance of the COO^- symmetric (ν_S COO^-, ~1410 cm^{-1}) and asymmetric (ν_{AS} COO^-, ~1540 cm^{-1}) stretching modes (negative bands).

As anticipated, d_{31}-palmitic acid spread onto a 0.47 M NaCl solution at pH 8.2 is partially deprotonated, demonstrated by the presence of all four carboxylic acid and carboxylate stretching modes (**Figures 3.3a** and **3.3c**). Addition of 10 mM $CaCl_2$ further deprotonates the headgroups, as shown by the increase in COO^- stretching intensities and by the disappearance of the C=O stretching mode in both TC and UC phases (**Tables 3.6** and **3.8**). Decreasing the pH to 5.8 increases the extent of d_{31}-palmitic acid protonation; the C-OH and C=O stretching modes are most intense at this pH (**Table 3.10**). However, the presence of the COO^- stretching modes in the TC and UC phase spectra indicates that Ca^{2+} induces some deprotonation which has been shown previously.[61,106–108]

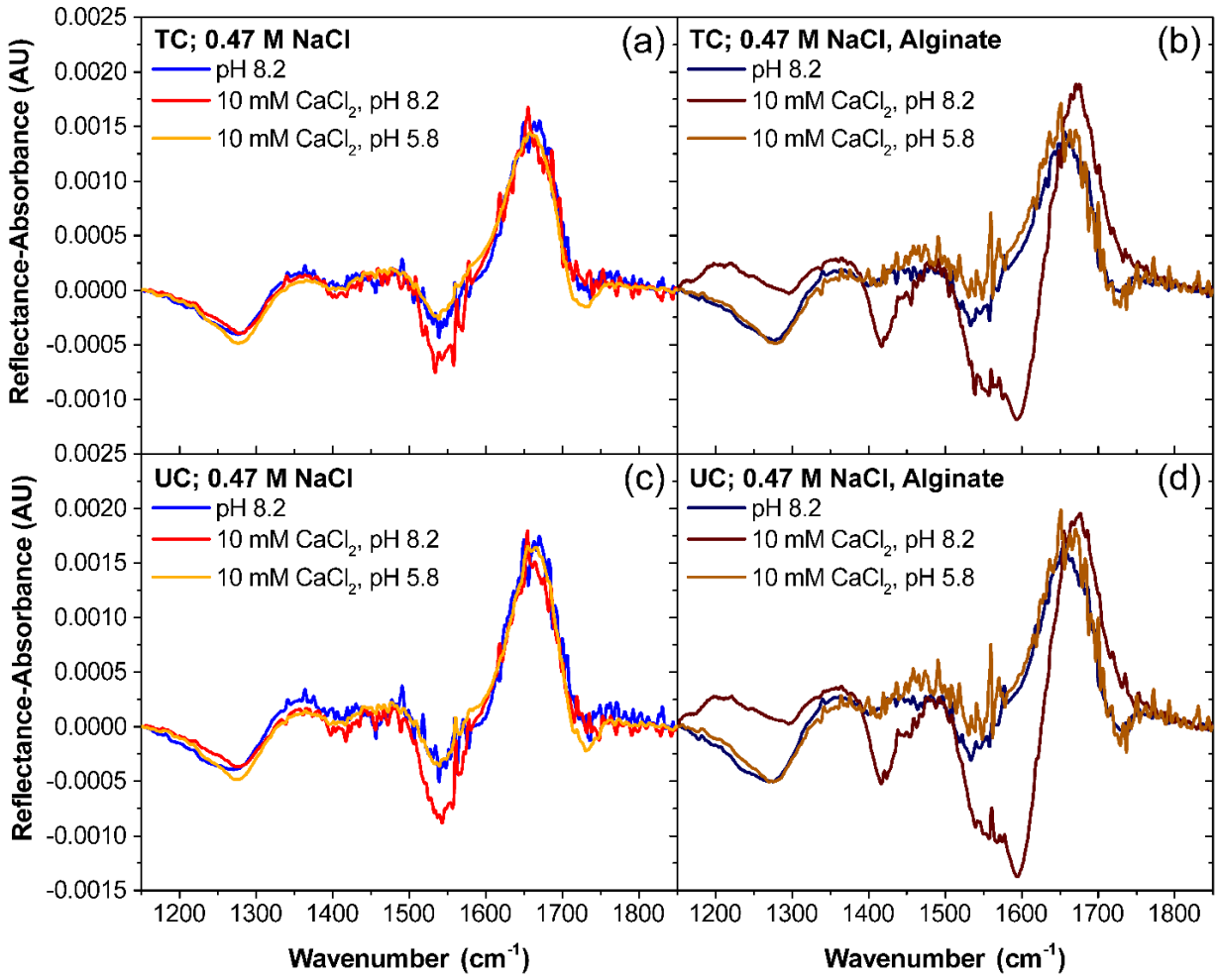


Figure 3.3: IRRAS spectra of the COOH stretching region provide direct evidence of alginate co-adsorption via Ca^{2+} bridging interactions to the d_{31}-palmitic acid monolayer at pH 8.2. Spectral lines are color-coded to indicate differences in solution composition. Surface pressure was held constant in the (a), (b) tilted condensed (5 mN/m) and (c), (d) untilted condensed (25 mN/m) phases throughout spectral acquisition. Spectra corresponding to the salt water solutions are shown in (a) and (c), and spectra corresponding to the salt water solutions containing 50 ppm alginate are shown in (b) and (d).

Direct measurement of alginate co-adsorption to the d_{31}-palmitic acid monolayer is observed in Figures 4.3b and 4.3d. There is a large increase in the negative COO^- stretching peak intensities and breadth for d_{31}-palmitic acid spread onto the 0.47 M NaCl, 10 mM $CaCl_2$, and 50 ppm alginate solution at pH 8.2. The asymmetric and symmetric stretches appear to split into higher and lower frequency bands, so an additional Gaussian function was used to fit both peaks (**Figure 3.13**). Vibrational frequency calculations predict a 47 cm^{-1} blue shift and a 20 cm^{-1} blue shift for the alginate asymmetric and symmetric COO^- stretching modes, respectively, relative to the corresponding d_{31}-palmitic acid modes (**Tables 3.3** and **3.5**). The theoretical

predictions quite closely match the experimental ~55 cm^{-1} blue shift for the second ν_{AS} COO^- peak and the ~37 cm^{-1} blue shift for the second ν_S COO^- peak (**Table 3.9**). Hence, the higher frequency COO^- stretching bands indicate alginate co-adsorption to the largely deprotonated monolayer in the presence of $CaCl_2$.

The carboxylate region also provides evidence for Ca^{2+} ionic bridges driving alginate co-adsorption to the d_{31}-palmitic acid monolayer. Alginate co-adsorption induces d_{31}-palmitic acid deprotonation, as shown by increased COO^- stretching intensities and decreased C-OH stretching intensity for the pH 8.2 solution containing 10 mM $CaCl_2$ (**Figure 3.3**). Similar peak broadening and intensity enhancement was observed for the phosphate headgroup vibrational modes of 1,2-dipalmitoyl-*sn*-glycero-3-phosphatidic acid (DPPA) upon arginine and guanidinium binding.[109] Palmitic acid deprotonation likely facilitates the formation of energetically favorable ionic complexes between alginate, Ca^{2+}, and palmitate. Furthermore, the ν_{AS} COO^-, ν_S COO^-, and ν C-OH modes blue shift upon addition of alginate to the solution (**Table 3.9**). The ν_{AS} COO^- mode blue shifts ~2.5 cm^{-1} in the TC phase but does not shift in the UC phase, the ν_S COO^- mode blue shifts 7 cm^{-1} in the TC phase and ~6 cm^{-1} in the UC phase, and the ν C-OH mode blue shifts ~13 cm^- in the TC phase and ~15 cm^{-1} in the UC phase. It is possible that the alginate carboxylate interacting with the Ca^{2+} ion complexed to the d_{31}-palmitate headgroup weakens ionic interactions between the monolayer carboxylate and the Ca^{2+} ion alone, thereby leading to an increased palmitic acid COO^- force constant and blue shifts in the palmitate carboxylate vibrational modes. However, it is more likely that the alginate carboxylate moieties further dehydrate the d_{31}-palmitate $\cdots$ Ca^{2+} complex upon alginate co-adsorption.

Spectra of d_{31}-palmitic acid spread onto the 0.47 M NaCl solution at pH 8.2 and the 0.47 M NaCl and 10 mM $CaCl_2$ solution at pH 5.8 do not exhibit any higher frequency COO^- stretching peaks upon alginate addition to the solution, suggesting minimal to no alginate co-adsorption (**Figure 3.3**). Secondly, the ν_{AS} COO^-, ν_S COO^-, and ν C-OH mode peak areas are insignificantly different between the solutions with and without alginate, meaning that alginate co-adsorption to the monolayer is unlikely (**Tables 3.6** and **3.7**; **Tables 3.10** and **3.11**). Smaller blue shifts in the ν_{AS} COO^-, ν_S COO^-, and ν C-OH modes are observed for the solutions containing 10 mM $CaCl_2$ and alginate at pH 5.8 as compared to the identical solution at pH 8.2 (**Table 3.11**). While these spectral shifts are not direct evidence of alginate co-adsorption to the monolayer, it is likely that

alginate perturbs the d_{31}-palmitic acid headgroup hydration structure via electrostatic interactions with the headgroups. The less numerous d_{31}-palmitate headgroups may facilitate Ca^{2+} bridging interactions with alginate carboxylate moieties at various points across the monolayer, as also indicated by the small relative change in surface coverage (**Figures 3.2b** and **3.2d**). However, having fewer Ca^{2+} bridging sites hinders concerted alginate co-adsorption. For the d_{33}-cetyl alcohol monolayer (**Figure 3.20**), there are no detectable peaks corresponding to alginate carboxylate modes, further supporting the lack of alginate co-adsorption to the fully protonated monolayer.

To determine if alginate co-adsorption perturbs d_{31}-palmitic acid interfacial organization, the CD_2 scissoring mode was analyzed as a function of solution composition, pH, and surface pressure (**Figure 3.4**). All conditions yield a CD_2 scissoring mode center wavelength of 1089 cm^{-1} (see **Table 3.16**), indicative of hexagonal lattice packing structure.[110] Furthermore, the absence of perturbations in the lattice packing structure, regardless of alginate presence in solution, indicates that alginate does not intercalate into the film. CD_2 scissoring mode spectra corresponding to the salt water and alginate solutions directly overlap for the 0.47 M NaCl solution at pH 8.2 (**Figures 3.4a** and **3.4d**) and for the 0.47 M NaCl and 10 mM $CaCl_2$ solution at pH 5.8 (**Figures 3.4c** and **3.4f**), further indicating no alginate co-adsorption. However, the CD_2 scissoring mode reflectance-absorbance signal magnitude in the TC and UC phase is significantly enhanced for d_{31}-palmitic acid spread onto the 0.47 M NaCl, 10 mM $CaCl_2$, and 50 ppm alginate solutions at pH 8.2 (**Figures 3.4b** and **3.4e**). Signal enhancement is most apparent in the high frequency regime, and a smaller extent of signal enhancement occurs in the lower frequency region of the spectra. From the harmonic frequency analysis results (**Tables 3.3** and **3.5**), the signal enhancement can be attributed to alginate C-OH stretching and bending and CH stretching. The CD_2 shoulder features overlap in frequency with the CD_2 scissoring modes, and particularly intense alginate vibrational transitions occur ~10 cm^{-1} lower than and ~20 cm^{-1} higher than the center frequency of the d_{31}-palmitate CD_2 scissoring mode. As a result, the signal enhancement surrounding the CD_2 scissoring peak can be confidently attributed to alginate co-adsorption to the monolayer via Ca^{2+} bridging interactions.

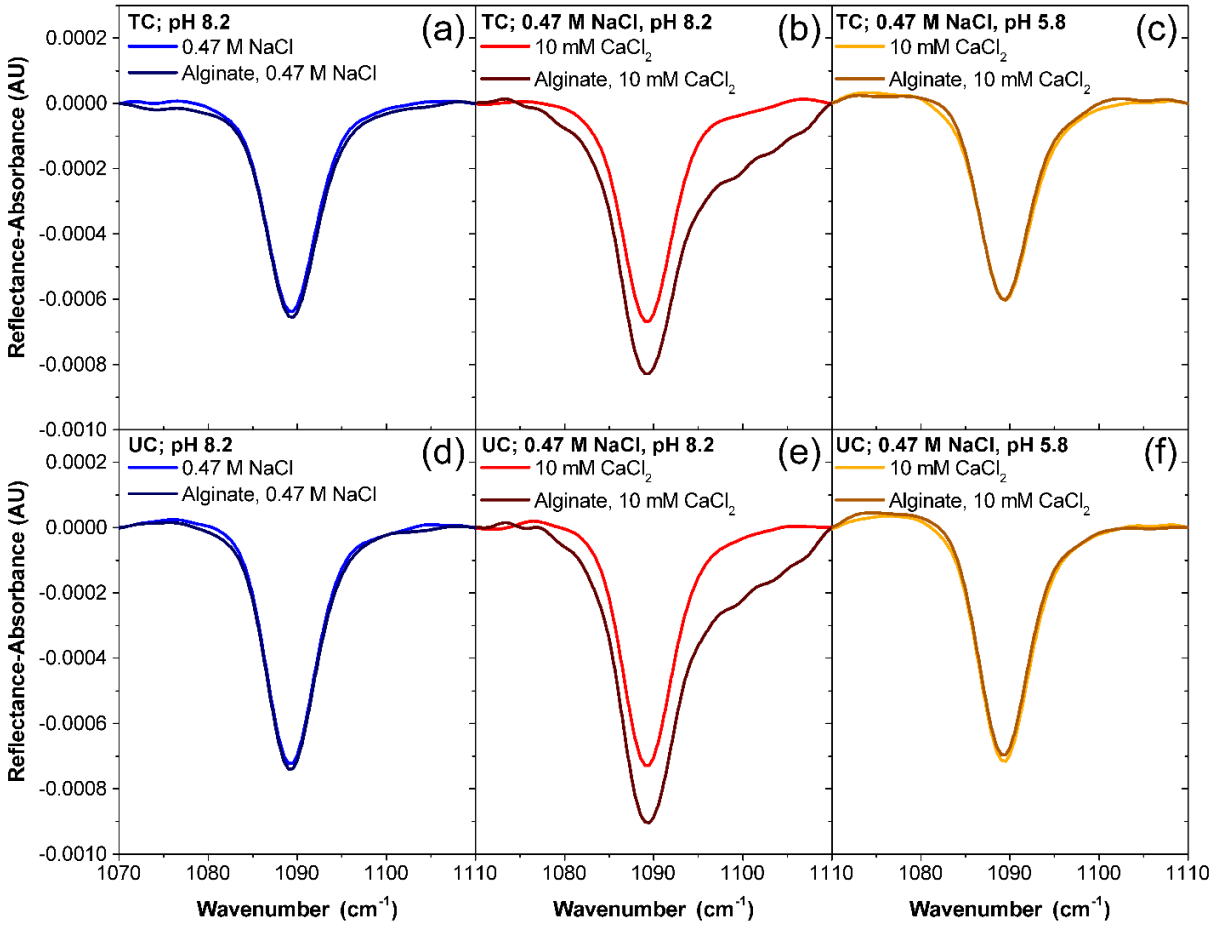


Figure 3.4: IRRAS spectra of the CD_2 scissoring mode region demonstrate alginate co-adsorption to the d_{31}-palmitic acid monolayer in the presence of 10 mM $CaCl_2$ at pH 8.2, and the spectra indicate no changes in d_{31}-palmitic acid lattice packing upon adsorption. Spectral lines are color-coded to indicate differences in solution composition. Surface pressure was held constant in the (a), (b), (c) tilted condensed (5 mN/m) and (d), (e), (f) untilted condensed (25 mN/m) phases throughout spectral acquisition.

3.4.3 Density and Radial Distribution Profiles from MD Simulations

Additional atomistic insights into the proposed divalent cationic bridging mechanism were obtained through explicit solvent all-atom molecular dynamics simulations of a TC palmitic acid monolayer at varying pH in the presence and absence of Ca^{2+} in the aqueous phase. **Figure 3.5** gives the number density and radial distribution profiles of selected species in each of the simulation conditions containing alginate at pH 8.2. The relative locations of the palmitic acid headgroups in the number density profiles (**Figures 3.5a** and **3.5b**) are in agreement with previous studies,[60,111] where the carboxylate headgroup of the palmitate sinks lower into the aqueous phase relative to the carboxylic acid. The peaks associated with the Ca^{2+} ion trace in **Figure 3.5a** are

indicative of selective calcium binding to the charged carboxylate headgroups. Similar coordination peaks associated with Na^+ are found in the sodium-only system shown in **Figure 3.5b**.

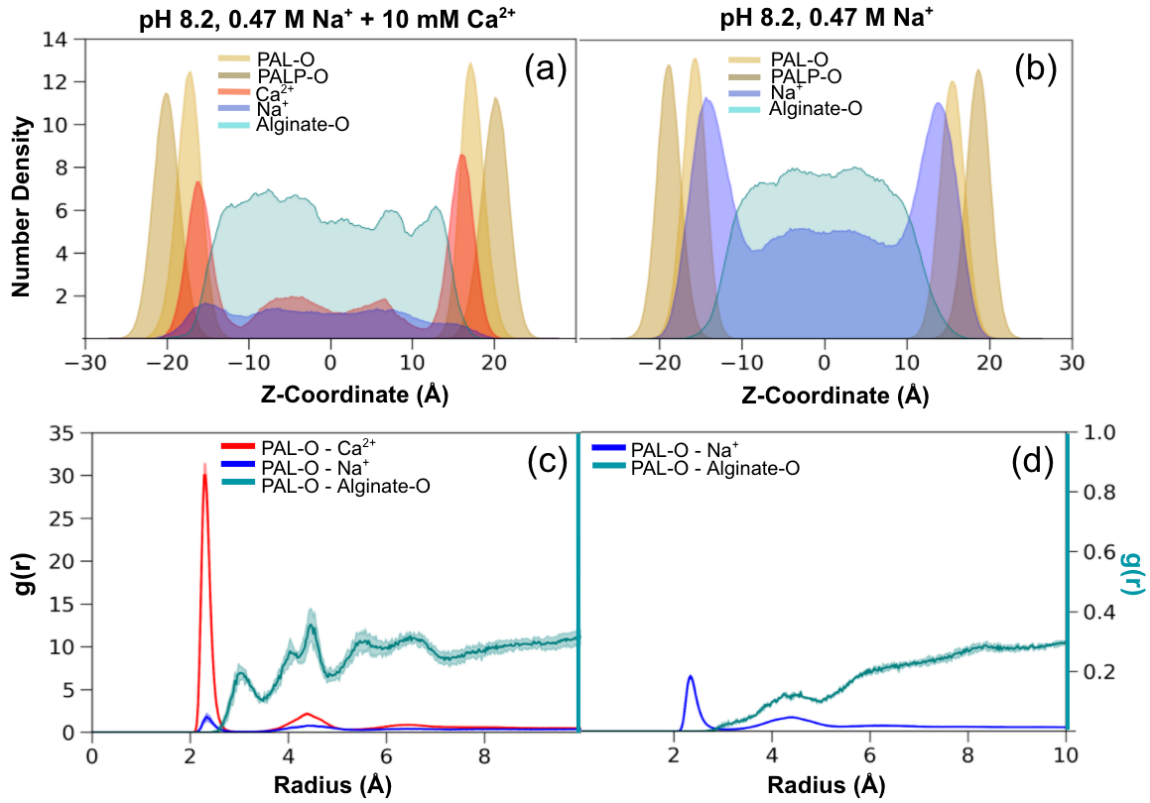


Figure 3.5: Number density profiles (a, b) and radial distribution functions (c-d) of selected species from molecular dynamics simulations at pH 8.2 with a 0.47 M NaCl subphase either in the presence (a) or absence (b) of Ca^{2+}. In the number density profiles, the oxygen atom of the palmitate headgroup carboxylate (PAL-O) is used to represent the location of palmitate (light tan). The OH-oxygen atom of the palmitic acid headgroup (PALP-O) is used to represent the location of palmitic acid (dark tan). Similarly, a carboxylate oxygen atom on each M monomer in alginate (Alginate-O) is used to quantify the distribution of alginate (teal). The radial distribution functions of Ca^{2+}, Na^+ and the alginate carboxylate with respect to the palmitate carboxylate are given for the systems in the presence (c) and absence (d) of Ca^{2+} at pH 8.2 over a 0.47 M NaCl subphase.

To determine whether alginate molecules co-adsorb to the monolayer by a cationic bridging mechanism as indicated by experiments, we can evaluate the alginate carboxylate traces. At pH 8.2, the alginate trace exhibits a broader, flatter profile in the presence of Ca^{2+} and a sharper rise in alginate coordination that occurs closer to the palmitate headgroups. We can also investigate this behavior using the radial distribution function (RDF) in which the probability density of the cation (Ca^{2+} or Na^+) or functional group (alginate COO^-) as a function of distance (r) from the palmitate headgroup is plotted (**Figures 3.5c-d**). There is a sharp cation peak as expected

at just over 2 Å in both systems. However, the alginate COO^- group exhibits distinctly different behavior between the two systems. The sharp peak at 3 Å in the presence of Ca^{2+} suggests that calcium likely serves as a bridge between the two COO^- moieties; conversely, in the absence of Ca^{2+}, the alginate peak occurs over much larger distances from the palmitate, has significantly lower probability, and is much broader. Furthermore, the RDFs of the water atoms (H_w and O_w) with respect to the palmitate COO^- are given in **Figure 3.21**. The sharp H_w and O_w peaks at approximately 2 and 3 Å, respectively, are associated with the water hydration shell around the palmitate COO^- headgroup.[111] With Ca^{2+} coordination, the COO^- headgroups are significantly less hydrated in comparison to Na^+-coordinated headgroups, where the solvation shell trace is much stronger. This observation agrees with the d_{31}-palmitic acid ν_{AS} COO^-, ν_S COO^-, and ν C-OH vibrational mode blue shifts upon alginate co-adsorption, indicative of monolayer headgroup dehydration. Taken together, these results strongly suggest that Ca^{2+} enables the adsorption of alginate to the palmitic acid monolayer at pH 8.2 through a divalent cation bridge.

In comparison, we also provide the number density profiles and RDFs for the monolayer systems at pH 5.8 in which the monolayer is fully protonated (**Figures 3.6a-d**). The number density plots exhibit distinctly different profiles from their higher pH counterparts; with Ca^{2+} present, rather than adsorbing to the surface, there is clear aggregation of the alginate in solution that is consistent with the gelation of alginate.[112–115] Without Ca^{2+}, the alginate shows a broader, flatter profile, indicating even distribution of alginate throughout solution and very little adsorption to the interface. The RDFs, in contrast with the higher pH system, show that alginates peak at just over 2 Å, which indicate that alginate weakly interacts with the monolayer via hydrogen bonding or dispersive forces between the alginate COO^- and palmitic acid -OH. The small cation hump at the same radius can be attributed to contact-ion pairing at the double-bonded oxygen of the carboxylic acid.[60] The subsequent increase in cation density after 4 Å may be attributed to either solvent-shared ion pairing between the cations and the first hydration shell of the headgroups, or, more likely, to contact-ion pairing between the cations and the alginate COO^-. These results are consistent with the experimental findings presented above, which suggest that palmitic acid protonation hinders alginate co-adsorption.

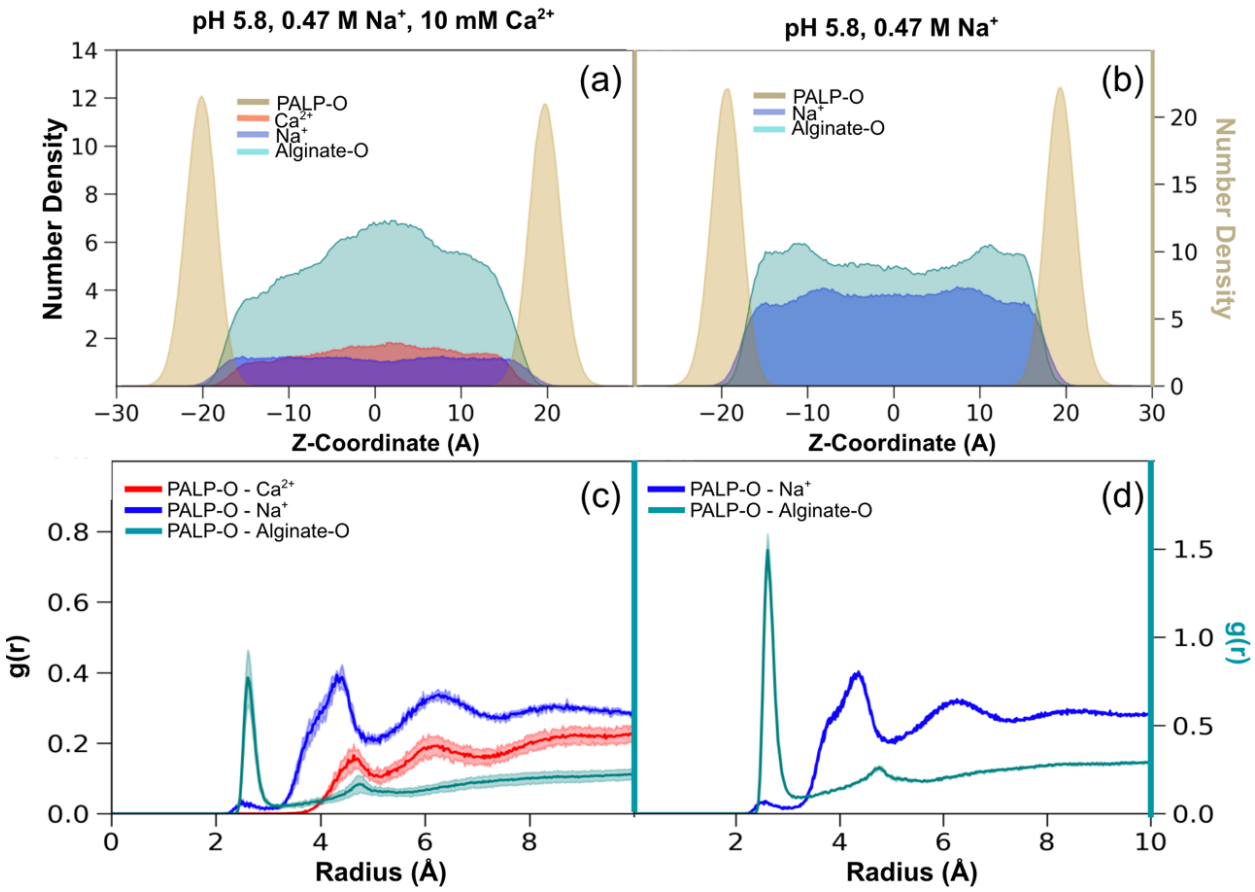


Figure 3.6: Number density (a-b) and radial distribution functions (c-d) from molecular dynamics simulations at pH 5.8 in which the palmitic acid monolayer is fully protonated.

3.4.4 Impacts of Alginate Adsorption on Palmitic Acid Interfacial Structure

The number density plots also show that, for all systems, the alginate trace goes to zero as the monolayer headgroup densities increase. Thus, there is little to no alginate intercalation into the monolayer. Alginate remains largely in the subsurface region, which is in good agreement with the experimental observation that hexagonal packing structure of the palmitic acid chains is preserved upon alginate co-adsorption.

To further investigate the impacts of alginate adsorption on the dynamics and organization of the palmitic acid monolayer, local order parameters were extracted from molecular dynamics simulations (**Figure 3.7** and **3.22**). These order parameters are calculated by

$$S_{CD} = \frac{3\times\langle cos^2\theta\rangle - 1}{2}, \tag{3.2}$$

where θ is the angle between the vector normal to the surface and adjacent carbon atoms, C_{n-1} and C_{n+1}. A high order parameter indicates high ordering among the alkyl chains with low motional anisotropy.

Figure 3.22 gives the order parameters for the palmitic acid and palmitate chains of the monolayers with and without alginate in the aqueous phase. With the exception of carbons 2 and 15, which tend to have low order parameters due to their geometric positions at the top and bottom of the monolayer, the order parameters remain relatively constant at low pH, indicating similar levels of ordering across the alkyl chains and thus no significant perturbations due to alginate interaction. However, at high pH, the order parameters vary across the chain. Additionally, the presence of alginate slightly decreases the order parameters, with the largest decrease seen in the absence of Ca^{2+}. Finally, compared to the low pH systems in which the order parameters steadily decrease from C4 to C2, the relative orders between C2 and C3 are reversed at high pH with C2 having a higher order parameter than C3. This reversal indicates that the headgroup carbons experience increased rotational rigidity at high pH.

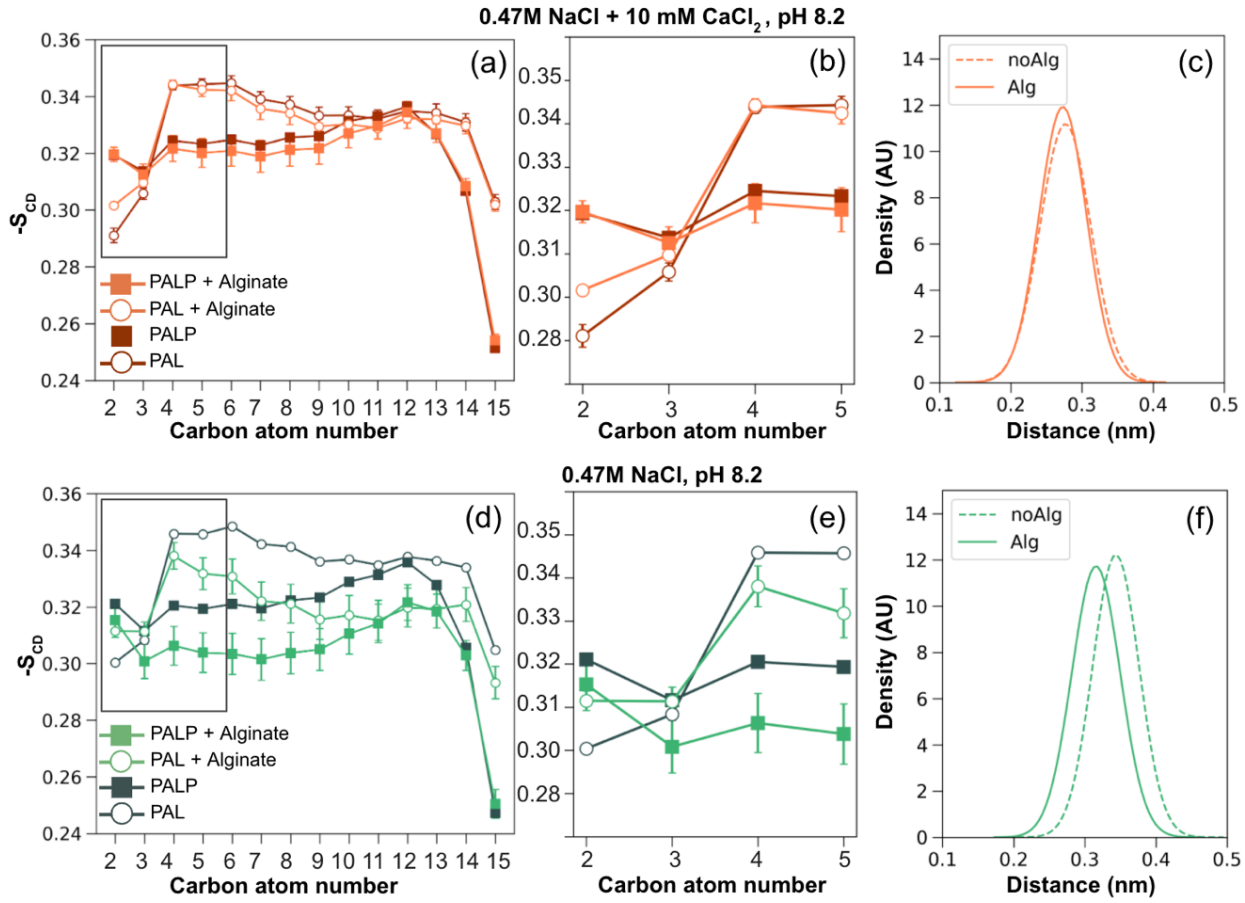


Figure 3.7: Local order parameters (a-b, d-e) and vertical headgroup separation (c, f) for palmitic acid monolayers at pH 8.2 in the presence (a-c) and absence (d-f) of calcium. Order parameters are calculated based on protonation state for all systems with and without alginate. Figures b and e are specific to the C2-C4 carbon range in the presence and absence of calcium, respectively.

To understand the physical basis of these structural differences, we recalculated order parameters based on residue protonation state (**Figure 3.7**). In all systems at pH 8.2, the deprotonated palmitic acid shows higher ordering than the protonated form, and the addition of alginate generally decreases ordering with the exception of C2 in both cases (**Figures 3.7a** and **3.7d**). In the presence of Ca^{2+}, alginate adsorption appears to increase ordering of the carboxylate groups but has no effect on the ordering of the protonated form (**Figure 3.7b**). In the absence of Ca^{2+}, the addition of alginate similarly increases ordering of the palmitate carboxylate headgroups but decreases ordering of the acid (**Figure 3.7e**).

To understand the differences in ordering based on headgroup protonation, we plotted the variations in headgroup position within the monolayer. Recalling that the palmitate headgroup tends to sink further into the

aqueous phase in comparison to the palmitic acid, in **Figures 3.7c** and **3.7f** we provide distributions of the height differences between carboxylate and carboxylic acid headgroups. The carboxylates are drawn into the aqueous phase, where they experience stabilizing ion-pairing with the cations, but the magnitude of the separation varies based on cation type. In the presence of Ca^{2+}, the headgroup separation decreases by nearly 1 Å to 2.8 Å, down from 3.8 Å without Ca^{2+}. The smaller difference in headgroup positions can be attributed to calcium-induced monolayer compression; Ca^{2+} binds neighboring carboxylate headgroups, thereby effectively shielding repulsive negative charges and bringing the headgroups closer together. Monovalent sodium does not participate in the ionic bridging and does less to screen the charges. Furthermore, whether alginate impacts the headgroup separation depends on the dominant adsorption or interaction mechanism. Because alginate adsorbs to the monolayer via calcium ion bridging, adsorption does not impact headgroup separation. However, without Ca^{2+} present, it is possible that the alginate COO^- slightly destabilizes the palmitate headgroups through repulsive interactions that are less mediated by Na^+, expanding the monolayer and allowing more space for the headgroups to align. Indeed, monolayer expansion is observed experimentally with increased MMA upon alginate adsorption (**Table 3.1**).

The differences in headgroup position between the protonated and deprotonated forms can explain the trends associated with the overall increase in ordering at the C2 carbon that are not seen in low pH systems. With the carboxylic acid headgroup more embedded into the monolayer, the rotational motion of that headgroup is significantly restricted, thus increasing the order parameter. Conversely, the carboxylate headgroups exposed to the aqueous phase experience a wider range of motion, and their order parameter decreases. In the case of alginate binding in the presence of Ca^{2+}, alginate has no impact on the carboxylic acids that are more deeply situated in the monolayer, but increases order associated with the carboxylates via the formation of calcium ion bridges that increase monolayer rigidity and decrease headgroup motion. In the case of alginate interacting with the monolayer without calcium bridges, we see a similar increase in order associated with the carboxylates and a decrease in order associated with the carboxylic acids. The mechanism for this behavior, although similar to that in calcium, is more likely explained by the decrease in headgroup separation associated with alginate interaction; the exposure of palmitic acid headgroups to the aqueous phase decreases order of the protonated form while at the same time increasing order of the carboxylate.

3.4.5 Investigation of magnesium bridging

With complementary insights gained into the divalent cationic bridging mechanism from both theory and experiment, the same alginate co-adsorption experiments were performed with $MgCl_2$, the most abundant divalent cation in seawater (~53 mM).[88,89] Mg^{2+} has such a strong hydration shell that the fully hydrated Mg^{2+} only weakly interacts with the alginate and palmitic acid carboxylate moieties.[45–48,101,108] Both 10 mM and 53 mM $MgCl_2$ solutions increase IRRAS reflectance-absorbance for the alginate-containing spectra at ~3580 cm^{-1}, and the magnitude of the signal increase is similar (**Figures 3.8a** and **3.8c**). Calculation of the relative change reveals that the higher $MgCl_2$ concentration causes a marginal increase in alginate co-adsorbed to the d_{31}-palmitic acid monolayer (**Figures 3.8b** and **3.8d**). With increasing $MgCl_2$ concentration, the relative change in integrated peak area increases from 8% to 10% and from 3% to 9% in the TC and UC phases, respectively. The reduction in OH region relative change with film compression (TC to UC phase) further suggests weak binding. Mg^{2+} is not as efficient in shielding the palmitic acid carboxylate moieties from negatively charged alginate carboxylate moieties, yielding monolayer expansion and increased surface coverage in the TC phase. Film compression to the UC phase reduces the extent of available space for monolayer expansion, and increased dispersion forces between the lipid tails counterbalance the repulsive electrostatic interactions. Despite Mg^{2+} being ~5 times more abundant than Ca^{2+} in seawater, alginate co-adsorption mediated by Mg^{2+} is ~3 times weaker in comparison to Ca^{2+} when comparing seawater relative concentrations of 10 mM Ca^{2+} and 53 mM Mg^{2+}. At 10 mM concentrations for both cations, Ca^{2+} outperforms Mg^{2+} by a factor of ~3 in the TC phase and by a factor of ~10 in the UC phase (**Figures 3.1** and **3.8**).

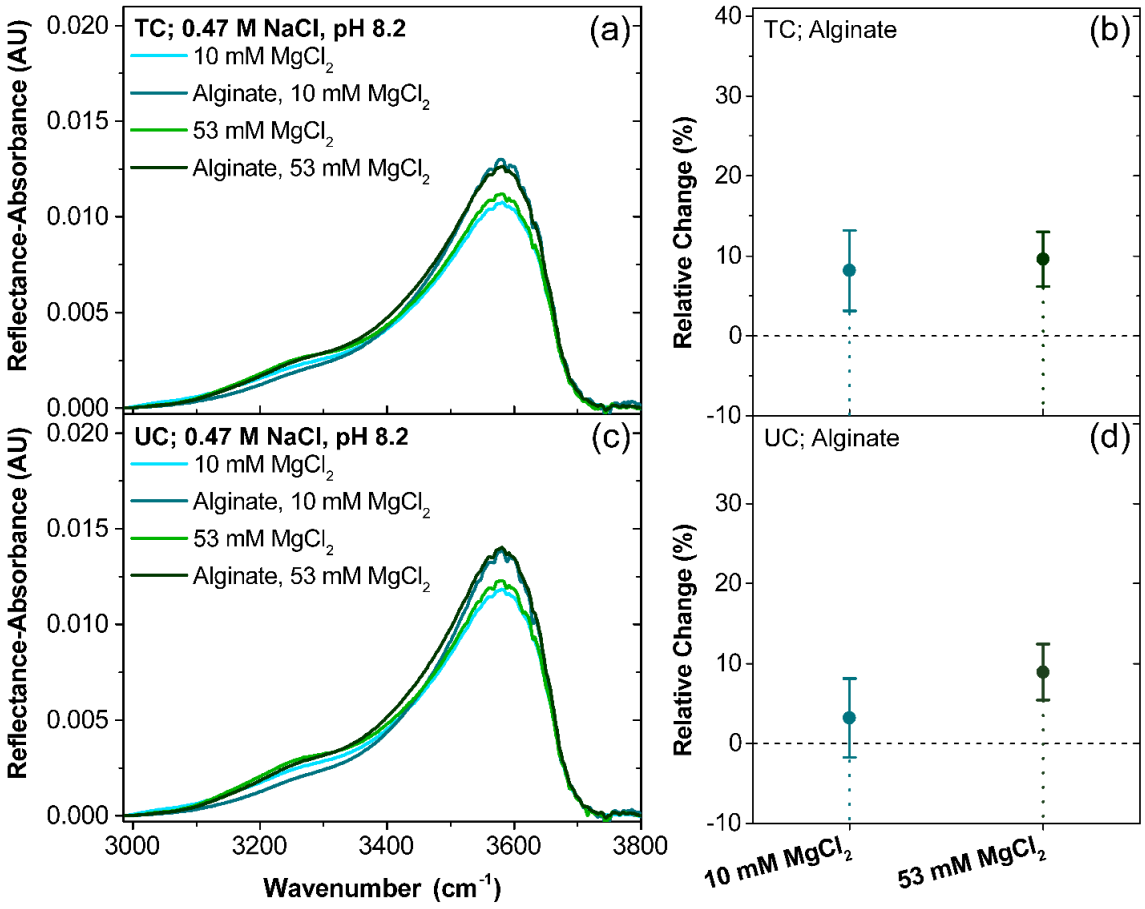


Figure 3.8: IRRAS spectra of the OH stretching region and the corresponding relative changes in integrated peak area indicate that Mg^{2+} induces weak alginate co-adsorption to the d_{31}-palmitic acid monolayer. Data points and error bars are color-coded to indicate differences in solution composition. Surface pressure was held constant in the (a) tilted condensed (5 mN/m) and (c) untilted condensed (25 mN/m) phases throughout spectral acquisition. The relative changes in integrated peak area between the solutions with and without alginate are quantified in the (b) tilted condensed and (d) untilted condensed phases. Positive relative change indicates alginate co-adsorption to the monolayer, and 0% relative change indicates no co-adsorption.

The carboxylate region provides additional support for this weak Mg^{2+} bridging co-adsorption mechanism (**Figure 3.23**). For d_{31}-palmitic acid spread onto the solution containing 10 mM $MgCl_2$, the ν_S COO^- mode is blue-shifted by 1 cm^{-1} in the TC phase and ~2.5 cm^{-1} in the UC phase, and the ν_{AS} COO^- mode is red-shifted by 1 cm^{-1} in the TC phase and ~3 cm^{-1} in the UC phase (**Tables 3.8** and **3.9**). These small spectral shifts suggest that the d_{31}-palmitate carboxylate headgroup becomes dehydrated with the addition of alginate to the solution,[95] perhaps via complexation of Mg^{2+} to the d_{31}-palmitate carboxylate moiety. The integrated carboxylate

peak areas do not change significantly between the salt water solution and alginate solution (**Tables 3.12** and **3.13**), suggesting no alginate co-adsorption.

Increasing the solution $MgCl_2$ concentration to 53 mM leads to some features of alginate co-adsorption to the monolayer (**Figure 3.23**). The d_{31}-palmitic acid ν_{AS} COO^-, ν_S COO^-, and ν C-OH modes blue shift upon alginate addition to the solution (**Tables 3.14** and **3.15**), similarly to the spectra corresponding to the solutions containing 10 mM $CaCl_2$. The ν_{AS} COO^- blue shifts ~0.5 cm^{-1} in the TC phase and ~3 cm^{-1} in the UC phase, the ν_S COO^- blue shifts ~4 cm^{-1} in the TC phase and ~5 cm^{-1} in the UC phase, and the ν C-OH blue shifts ~9 cm^{-1} in the TC phase and ~12 cm^{-1} in the UC phase. The blue shifts are smaller than those observed with the 10 mM $CaCl_2$ solution at pH 8.2, either an indication of fewer Mg^{2+} bridging interactions or weaker bridging interactions. Secondly, the ν C-OH peak area decreases while the ν_{AS} COO^- and ν_S COO^- peak areas increase with alginate present in the solution, further supporting the hypothesis of alginate co-adsorption.

The CD_2 scissoring modes of d_{31}-palmitic acid spread onto the $MgCl_2$ solutions corroborate the findings from the OH stretching and COOH stretching regions. At 10 mM $MgCl_2$, the salt water and alginate spectra nearly overlap (**Figures 3.24a** and **3.24c**). There is a small increase in peak intensity of the TC spectrum corresponding to the alginate solution that could be attributed to alginate weakly co-adsorbed to the TC monolayer. Then the alginate is squeezed out upon film compression, causing the peak intensity difference to disappear in the UC phase. The solution containing 53 mM $MgCl_2$ and alginate induces increased peak intensities in the d_{31}-palmitic acid CD_2 scissoring mode and the higher frequency regime relative to the spectra corresponding to the salt water solution (**Figures 3.24b** and **3.24d**). The signal enhancement is smaller in comparison to the system containing 10 mM $CaCl_2$ and alginate at pH 8.2, but the spectral trends match. Hence, the higher concentration of $MgCl_2$ facilitates some alginate co-adsorption to the monolayer through Mg^{2+} bridging interactions.

3.5 Conclusions

We directly observe alginate co-adsorption to an insoluble d_{31}-palmitic acid monolayer via divalent cationic bridging interactions using surface-sensitive infrared reflection-absorption spectroscopy (IRRAS) and molecular dynamics simulations. Ca^{2+} facilitates the greatest extent of alginate co-adsorption, as shown by the

appearance of alginate vibrational modes in the IRRAS spectra and by the ~27% and ~32% increase in surface coverage in the TC and UC phases, respectively. Alginate co-adsorption is dependent upon d_{31}-palmitic acid protonation state, however; d_{31}-palmitate promotes alginate co-adsorption in the presence of divalent cations, whereas protonation inhibits co-adsorption. Na^+ cations alone are insufficient in facilitating co-adsorption. Mg^{2+} induces ~3 times weaker alginate co-adsorption at a seawater concentration of 53 mM in comparison to 10 mM Ca^{2+}, and Mg^{2+} induces minimal co-adsorption when matching the Ca^{2+} seawater concentration (10 mM). The hydration free energy of Mg^{2+} is much higher than that of Ca^{2+}, meaning that Mg^{2+} cannot shed its hydration shell as readily to facilitate bridging interactions between the d_{31}-palmitate and alginate carboxylate moieties. The presence of alginate perturbs the hydration structure and dehydrates the d_{31}-palmitic acid carboxylic acid headgroups, but alginate co-adsorption does not change the d_{31}-palmitic acid lattice packing structure. Alginate co-adsorption is largely confined to the subsurface region of the film. Consequently, surface pressure plays a minimal role in the extent of alginate co-adsorption.

Our detailed experimental and computational characterization of the divalent cationic bridging interactions driving alginate co-adsorption to a sea surface microlayer (SSML) proxy film provides important physical and chemical insights into the potential mechanisms responsible for polysaccharide enrichment in sea spray aerosol (SSA). Ca^{2+} drives this bridging motif between the alginate and palmitic acid carboxylate moieties and outcompetes Mg^{2+} despite higher Mg^{2+} concentrations in seawater. Quantification of organic surface coverage via the OH stretching region integrated peak areas also provides a potentially useful parameter for aerosol representation in climate models. We demonstrate that soluble polysaccharides can interact electrostatically with other surface-active organic matter through seawater ionic bridging interactions, leading to polysaccharide surface enrichment in the SSML, and therefore, in SSA.

3.6 Acknowledgments

K.A.C.-F. designed the study, and H.C.A. and R.E.A. supervised the project and acquired funding. K.A.C.-F., M.E.F., and J.K. performed the experiments. A.C.D. performed the molecular dynamics simulations, and K.A.C.-F. performed the quantum chemistry calculations. K.A.C.-F. and A.C.D. analyzed the data and wrote the paper with input from all authors.

Chapter 3, in full, is a modified reprint of the material as it appears in "Carter-Fenk, K. A.; Dommer, A. C.; Fiamingo, M. E.; Kim, J.; Amaro, R. E.; Allen, H. C. *Calcium bridging drives polysaccharide co-adsorption to a proxy sea surface microlayer.* Phys. Chem. Chem. Phys. 2021, 23, 16401-16416." The dissertation author was a co-investigator and co-author of this work.

3.7 Supporting Information

3.7.1 Molecular Dynamics System Set-Up

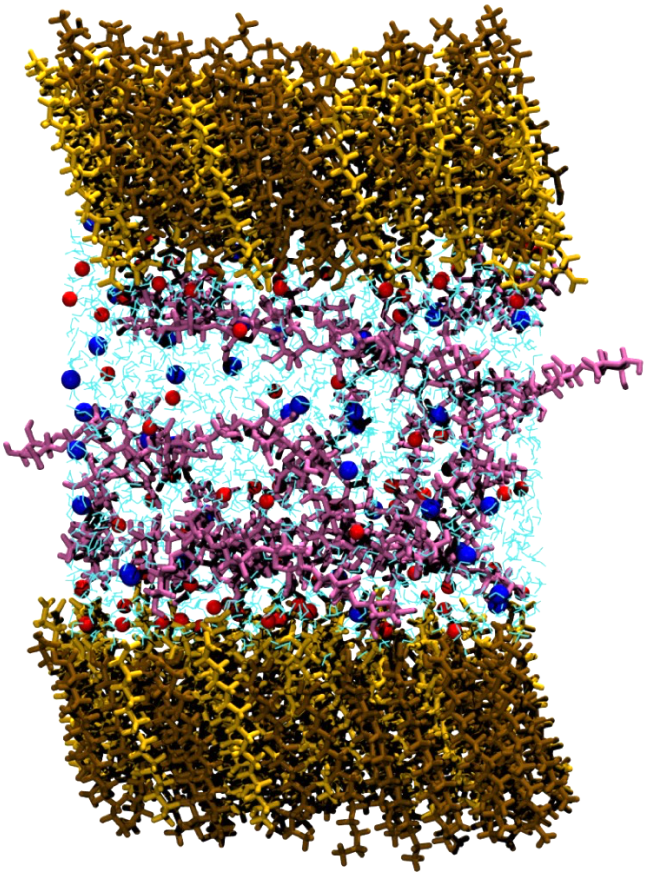


Figure 3.9: Example of a molecular dynamics monolayer system set up to include alginate. Pictured: palmitic acid (brown, licorice representation), palmitate (yellow, licorice), water (cyan, lines) alginate (mauve, licorice), calcium (red, Van der Waals) and sodium (blue, Van der Waals). Above and below the monolayers is a vacuum region where no molecules were placed to approximate atmospheric air pressure.

3.7.2 Mean Molecular Area Data

Mean molecular area (MMA) data of the d_{31}-palmitic acid and d_{33}-cetyl alcohol monolayers measured at constant surface pressure are tabulated below. Alginate consistently expands the monolayer relative to the salt

water subphase, and the magnitude of monolayer expansion is greater at 5 mN/m than at 25 mN/m. Enhanced dispersion interactions between the lipid alkyl chains and increased exchange interactions likely push alginate out of the plane of the compressed monolayer.

Table 3.1 Average mean molecular area values (Å^2/molecule) and one standard deviation from the mean (σ, Å^2/molecule) of d_{31}-palmitic acid (d_{31}-PA) and d_{33}-cetyl alcohol (d_{33}-CA) monolayers measured at constant surface pressures 5 mN/m and 25 mN/m.

Monolayer & Subphase	5 mN/m		25 mN/m	
	MMA	σ	MMA	σ
d_{31}-PA, 0.47 M NaCl, pH 8.2	22.66	0.29	20.57	0.33
d_{31}-PA, 50 ppm Alginate, 0.47 M NaCl, pH 8.2	25.48	0.44	21.50	0.44
d_{31}-PA, 0.47 M NaCl, 10 mM $CaCl_2$, pH 8.2	22.36	0.16	20.92	0.11
d_{31}-PA, 50 ppm Alginate, 0.47 M NaCl, 10 mM $CaCl_2$, pH 8.2	23.27	0.65	21.56	0.32
d_{31}-PA, 0.47 M NaCl, 10 mM $CaCl_2$, pH 5.8	25.21	0.24	22.67	0.20
d_{31}-PA, 50 ppm Alginate, 0.47 M NaCl, 10 mM $CaCl_2$, pH 5.8	27.29	0.48	22.99	0.92
d_{33}-CA, 0.47 M NaCl, 10 mM $CaCl_2$, pH 8.2	20.62	0.15	17.85	0.09
d_{33}-CA, 50 ppm Alginate, 0.47 M NaCl, 10 mM $CaCl_2$, pH 8.2	21.91	0.08	18.72	0.05
d_{31}-PA, 0.47 M NaCl, 10 mM $MgCl_2$, pH 8.2	24.27	0.04	21.40	0.14
d_{31}-PA, 50 ppm Alginate, 0.47 M NaCl, 10 mM $MgCl_2$, pH 8.2	26.51	0.13	22.08	0.17
d_{31}-PA, 0.47 M NaCl, 53 mM $MgCl_2$, pH 8.2	24.17	0.10	21.80	0.09
d_{31}-PA, 50 ppm Alginate, 0.47 M NaCl, 53 mM $MgCl_2$, pH 8.2	27.06	0.43	22.86	0.27

Table 3.2 The ratio of d_{31}-palmitic acid (d_{31}-PA) and d_{33}-cetyl alcohol (d_{33}-CA) MMA values (Alg/SW) and propagated error corresponding to the 50 ppm alginate subphase MMA divided by the seawater subphase MMA.

Monolayer & Subphase	5 mN/m		25 mN/m	
	Alg/SW	Error	Alg/SW	Error
d_{31}-PA, 0.47 M NaCl, pH 8.2	1.124	0.022	1.045	0.026
d_{31}-PA, 0.47 M NaCl, 10 mM $CaCl_2$, pH 8.2	1.041	0.029	1.030	0.016
d_{31}-PA, 0.47 M NaCl, 10 mM $CaCl_2$, pH 5.8	1.083	0.020	1.014	0.041
d_{33}-CA, 0.47 M NaCl, 10 mM $CaCl_2$, pH 8.2	1.062	0.008	1.049	0.005
d_{31}-PA, 0.47 M NaCl, 10 mM $MgCl_2$, pH 8.2	1.092	0.005	1.032	0.010
d_{31}-PA, 0.47 M NaCl, 53 mM $MgCl_2$, pH 8.2	1.120	0.017	1.048	0.012

3.7.3 D-Guluronate and d_{31}-Palmitic Acid Vibrational Frequency Calculations

To distinguish between peaks corresponding to the d_{31}-palmitic acid monolayer or the co-adsorbed alginate, harmonic vibrational frequency calculations were performed using Q-Chem v. 5.3.1.[116] D-guluronic acid, an alginate monomer primarily responsible for cation binding, was selected to model the alginate vibrational modes (**Figure 3.10a**).[36,39,40] Both protonated d_{31}-palmitic acid (**Figure 3.10b**) and deprotonated d_{31}-palmitate (**Figure 3.10c**) were modeled, and the atomic mass of 2.01410 was used for the deuterium atoms.[117] Geometry optimization and harmonic frequency analysis were performed at the EDF2/6-31+G* level of theory.[118] Frequencies, intensities, and vibrational mode assignments within the frequency region of the carboxylic acid headgroup and the C-D bending modes (1090-1850 cm^{-1}) are tabulated in Tables 4.3-4.5.

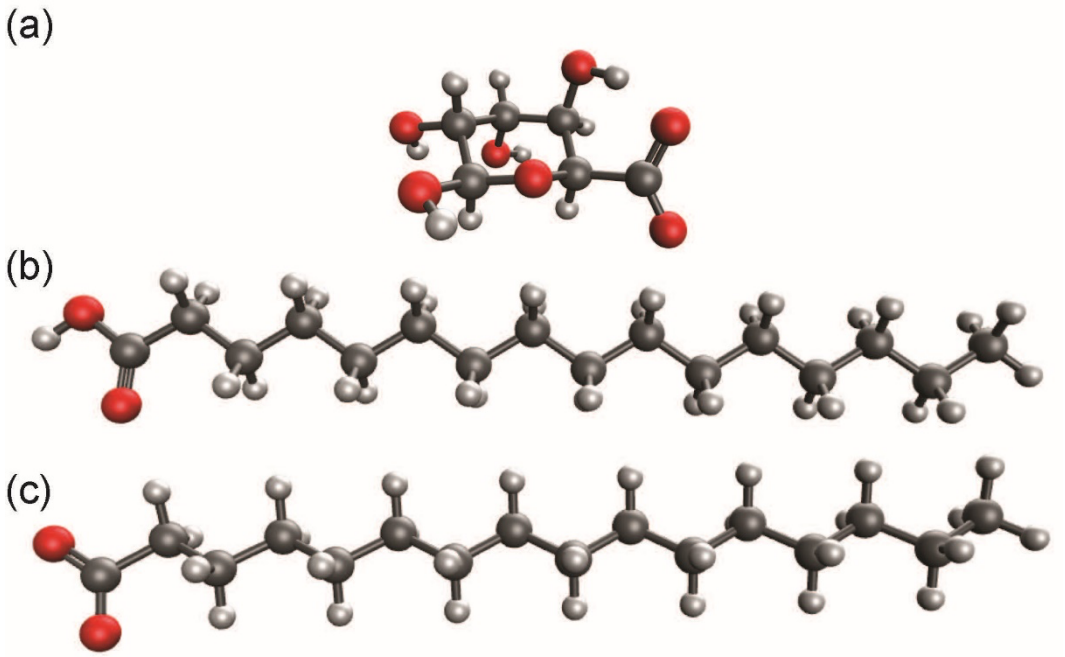


Figure 3.10: Geometry optimized structures of (a) D-guluronate, (b) d_{31}-palmitic acid, and (c) d_{31}-palmitate.

Table 3.3: Harmonic vibrational frequencies, intensities, and vibrational mode assignments of D-guluronate.

Wavenumber (cm^{-1})	Intensity (km/mol)	Mode Assignments
1094.24	111.741	C-O-H Bending, C-O-C Stretching
1100.50	145.998	C-OH Stretching
1116.31	80.623	C-O-C Stretching, C-O-H Bending
1148.48	211.874	C-O-H Stretching, C-H Wagging
1154.90	2.637	C-O-H Stretching and Bending
1206.32	34.863	C-O-H Bending, C-H Wagging
1242.62	32.793	C-H Wagging, C-O-H Bending
1264.93	32.675	C-H Wagging, C-O-H Bending
1297.55	18.810	C-H Wagging, C-O-H Bending
1317.06	32.035	C-H Wagging, C-O-H Bending
1325.07	15.798	C-H Wagging, C-O-H Bending
1340.30	18.708	C-H Wagging, C-O-H Bending
1352.16	121.662	C-H Wagging, C-O-H Bending, COO^- Symmetric Stretching
1356.34	27.790	C-H Wagging, C-O-H Bending, COO^- Symmetric Stretching
1391.73	127.213	COO^- Symmetric Stretching, C-H Wagging
1400.35	16.914	C-H Wagging, C-O-H Bending
1411.78	25.561	C-O-H Bending, C-H Wagging
1439.87	44.051	C-O-H Bending, C-H Wagging
1474.40	2.861	C-O-H Bending, C-H Wagging
1547.29	384.253	C-O-H Bending, COO^- Asymmetric Stretching
1725.18	423.674	COO^- Asymmetric Stretching, C-O-H Bending

Table 3.4: Harmonic vibrational frequencies, intensities, and vibrational mode assignments of d_{31}-palmitic acid.

Wavenumber (cm^{-1})	Intensity (km/mol)	Mode Assignments
1092.09	0.403	CD_3 Bending, CD_2 Bending & Wagging
1093.01	4.856	CD_3 Bending
1093.67	7.022	CD_2 Bending
1108.35	13.932	CD_2 Bending & Wagging, CD_3 Bending
1116.04	2.674	CD_2 Bending & Wagging
1117.22	0.663	CD_2 Bending & Wagging
1119.12	0.914	CD_2 Bending & Wagging
1121.34	0.634	CD_2 Bending
1123.82	1.570	CD_2 Bending
1125.90	0.640	CD_2 Bending
1127.50	11.375	CD_2 Bending
1128.73	8.575	CD_2 Bending, CD_3 Bending
1132.35	1.303	CD_2 Bending & Wagging, CD_3 Bending
1141.51	14.964	CD_2 Bending, C-O-H Bending & Stretching
1150.56	15.657	CD_2 Bending & Wagging, C-O-H Bending & Stretching
1159.16	46.521	C-O-H Bending & Stretching, CD_2 Bending & Wagging
1167.89	4.210	CD_2 Bending & Wagging, C-O-H Bending & Stretching
1177.22	4.233	CD_2 Bending, C-O-H Bending & Stretching
1189.07	65.964	C-O-H Bending & Stretching, CD_2 Wagging
1216.54	59.856	C-O-H Bending & Stretching, CD_2 Wagging
1241.05	15.368	CD_2 Wagging, C-O-H Bending
1260.91	6.855	CD_2 Wagging, C-O-H Bending
1274.81	0.569	CD_2 Wagging, C-O-H Bending
1282.93	0.528	CD_2 Wagging, C-O-H Bending
1286.54	0.032	CD_2 Wagging, C-O-H Bending
1288.46	0.160	CD_2 Wagging, C-O-H Bending
1369.64	120.711	C-O-H Bending & Stretching
1827.90	299.335	C=O Stretching, C-O-H Bending

Table 3.5: Harmonic vibrational frequencies, intensities, and vibrational mode assignments of d_{31}-palmitate.

Wavenumber (cm^{-1})	Intensity (km/mol)	Mode Assignments
1091.89	7.209	CD_3 Bending, CD_2 Bending & Wagging
1092.92	4.719	CD_3 Bending
1093.94	5.964	CD_2 Bending & Wagging, CD_3 Bending
1108.52	0.333	CD_2 Bending & Wagging, CD_3 Bending
1114.79	0.338	CD_2 Bending & Wagging
1116.53	0.014	CD_2 Bending & Wagging
1118.41	0.561	CD_2 Bending & Wagging, CD_3 Bending
1120.73	0.006	CD_2 Bending & Wagging, CD_3 Bending
1123.15	1.640	CD_2 Bending, CD_3 Bending
1125.37	0.002	CD_2 Bending, CD_3 Bending
1127.13	10.458	CD_2 Bending
1127.55	0.296	CD_2 Bending, CD_3 Bending
1131.37	0.628	CD_2 Bending & Wagging, CD_3 Bending
1140.96	0.701	CD_2 Bending & Wagging, CD_3 Bending
1149.66	0.109	CD_2 Bending & Wagging
1162.08	2.876	CD_2 Bending & Wagging, CD_3 Bending
1167.45	0.477	CD_2 Bending & Wagging, CD_3 Bending
1177.28	0.702	CD_2 Bending, CD_3 Bending
1198.48	0.953	CD_2 Bending & Wagging, CD_3 Bending
1228.77	0.582	CD_2 Bending & Wagging, CD_3 Bending
1250.96	1.753	CD_2 Wagging, CD_3 Bending
1265.36	0.348	CD_2 Wagging, CD_3 Bending
1275.19	0.701	CD_2 Wagging, CD_3 Bending
1281.76	0.054	CD_2 Wagging
1285.41	0.085	CD_2 Wagging
1371.58	328.038	COO^- Symmetric Stretching
1678.00	637.505	COO^- Asymmetric Stretching

3.7.4 Carboxylate Region Spectral Fitting

Peaks within the COOH stretching region (1150-1850 cm^{-1}) were fitted to Gaussian functions using OriginPro 9.0. The software uses the Levenberg-Marquardt algorithm to perform the nonlinear curve fitting routine. Eight points were placed along the baseline, and a 4^{th}-order polynomial function was fitted to those points and used to define the spectral baseline. Peak fitting parameters were not fixed.

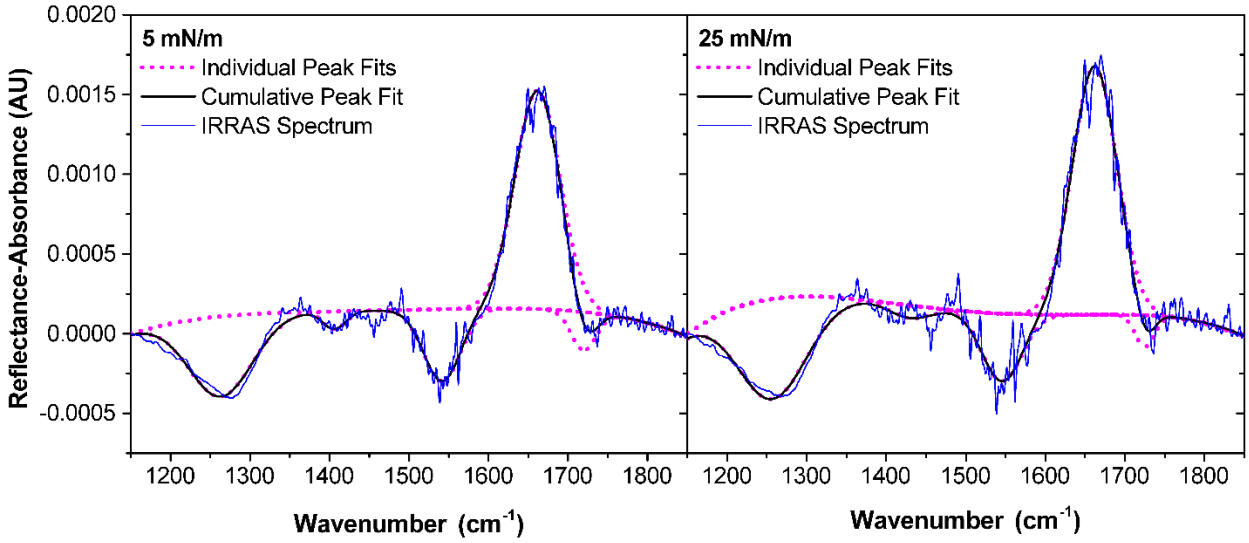


Figure 3.11: IRRAS spectra and corresponding peak fits of a d_{31}-palmitic acid monolayer at 5 mN/m (left) and 25 mN/m (right) spread onto an aqueous subphase of 0.47 M NaCl at pH 8.2.

Table 3.6: Center wavelengths (λ, cm^{-1}), reflectance-absorbance intensities (RA Int.), peak areas, and full width at half maximum (FWHM, cm^{-1}) values of Gaussian fits to IRRAS spectra in the COOH vibrational mode region of a d_{31}-palmitic acid monolayer (5 mN/m and 25 mN/m) spread onto an aqueous subphase of 0.47 M NaCl at pH 8.2.

Vibrational Mode	5 mN/m				25 mN/m			
	Center λ	RA Int.	Area	FWHM	Center λ	RA Int.	Area	FWHM
ν C-OH	1263.5	-5.02E-04	0.0501	93.7	1256.7	-6.31E-04	0.06895	102.7
ν_S COO^-	1406.3	-1.10E-04	0.00399	34.0	1429.7	-7.92E-05	0.00509	60.4
ν_{AS} COO^-	1541.6	-4.49E-04	0.0261	54.5	1545.6	-4.23E-04	0.0263	58.4
δ H-O-H	1661.5	0.00137	0.0975	66.9	1662.0	0.00156	0.109	66.0
ν C=O	1719.6	-2.40E-04	0.00942	36.8	1725.1	-1.96E-04	0.00586	28.1

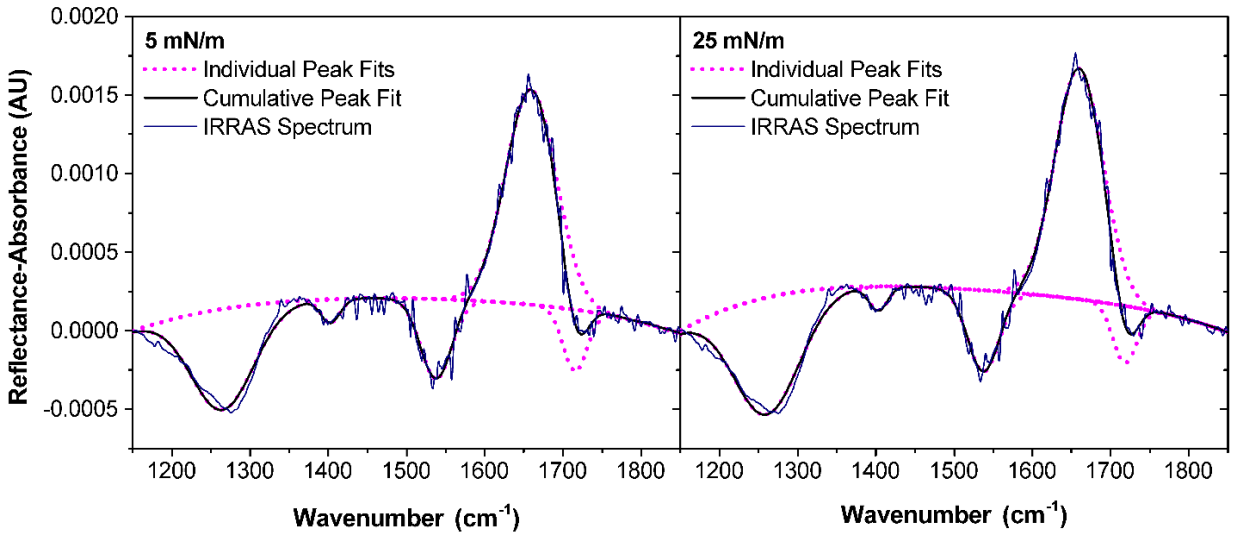


Figure 3.12: IRRAS spectra and corresponding peak fits of a d_{31}-palmitic acid monolayer at 5 mN/m (left) and 25 mN/m (right) spread onto an aqueous subphase of 0.47 M NaCl and 50 ppm alginate at pH 8.2.

Table 3.7: Center wavelengths (λ, cm^{-1}), reflectance-absorbance intensities (RA Int.), peak areas, and full width at half maximum (FWHM, cm^{-1}) values of Gaussian fits to IRRAS spectra in the COOH vibrational mode region of a d_{31}-palmitic acid monolayer (5 mN/m and 25 mN/m) spread onto an aqueous subphase of 0.47 M NaCl and 50 ppm alginate at pH 8.2.

Vibrational Mode	5 mN/m				25 mN/m			
	Center λ	RA Int.	Area	FWHM	Center λ	RA Int.	Area	FWHM
ν C-OH	1264.7	-6.42E-04	0.0643	94.1	1260.5	-7.46E-04	0.0789	99.3
ν_S COO^-	1402.6	-1.50E-04	0.00458	28.7	1402.9	-1.53E-04	0.00461	28.3
ν_{AS} COO^-	1537.8	-5.07E-04	0.0253	46.9	1537.3	-5.16E-04	0.0260	47.3
δ H-O-H	1658.9	0.00137	0.111	76.2	1659.6	0.00147	0.114	72.6
ν C=O	1715.4	-3.88E-04	0.0149	35.9	1718.3	-3.62E-04	0.0139	36.0

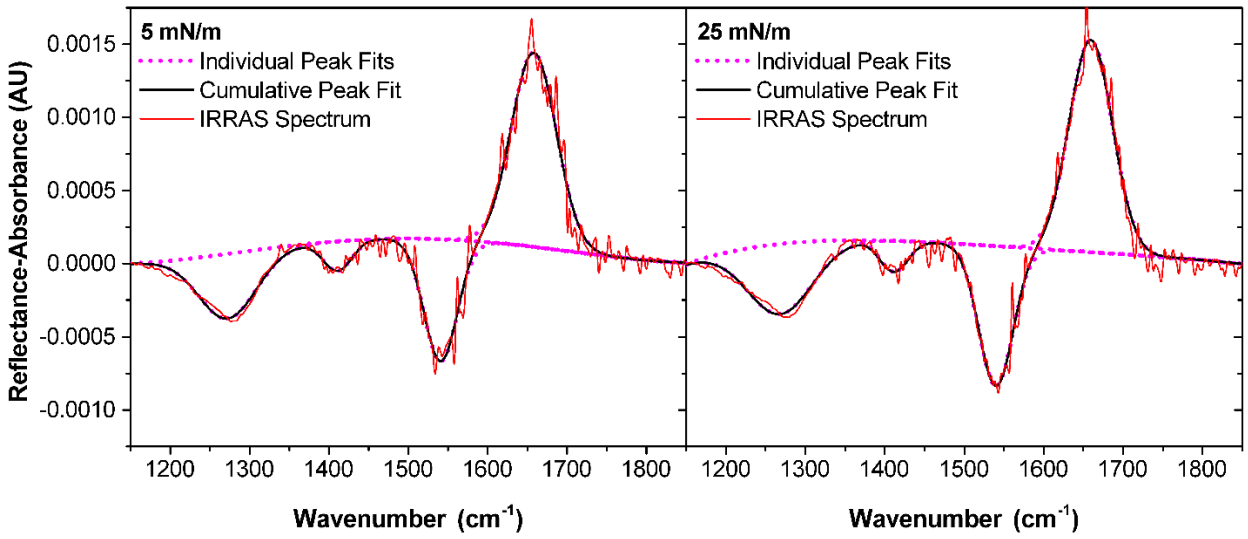


Figure 3.13: IRRAS spectra and corresponding peak fits of a d_{31}-palmitic acid monolayer at 5 mN/m (left) and 25 mN/m (right) spread onto an aqueous subphase of 0.47 M NaCl and 10 mM $CaCl_2$ at pH 8.2.

Table 3.8: Center wavelengths (λ, cm^{-1}), reflectance-absorbance intensities (RA Int.), peak areas, and full width at half maximum (FWHM, cm^{-1}) values of Gaussian fits to IRRAS spectra in the COOH vibrational mode region of a d_{31}-palmitic acid monolayer (5 mN/m and 25 mN/m) spread onto an aqueous subphase of 0.47 M NaCl and 10 mM $CaCl_2$ at pH 8.2.

Vibrational Mode	5 mN/m				25 mN/m			
	Center λ	RA Int.	Area	FWHM	Center λ	RA Int.	Area	FWHM
ν C-OH	1272.6	-4.42E-04	0.0401	85.0	1268.2	-4.81E-04	0.0497	97.1
ν_S COO^-	1410.8	-2.03E-04	0.00908	41.9	1411.5	-2.12E-04	0.00933	41.3
ν_{AS} COO^-	1540.6	-8.35E-04	0.0439	49.4	1539.9	-9.57E-04	0.0515	50.6
δ H-O-H	1658.0	0.00133	0.0946	67.0	1659.5	0.00145	0.0994	64.4

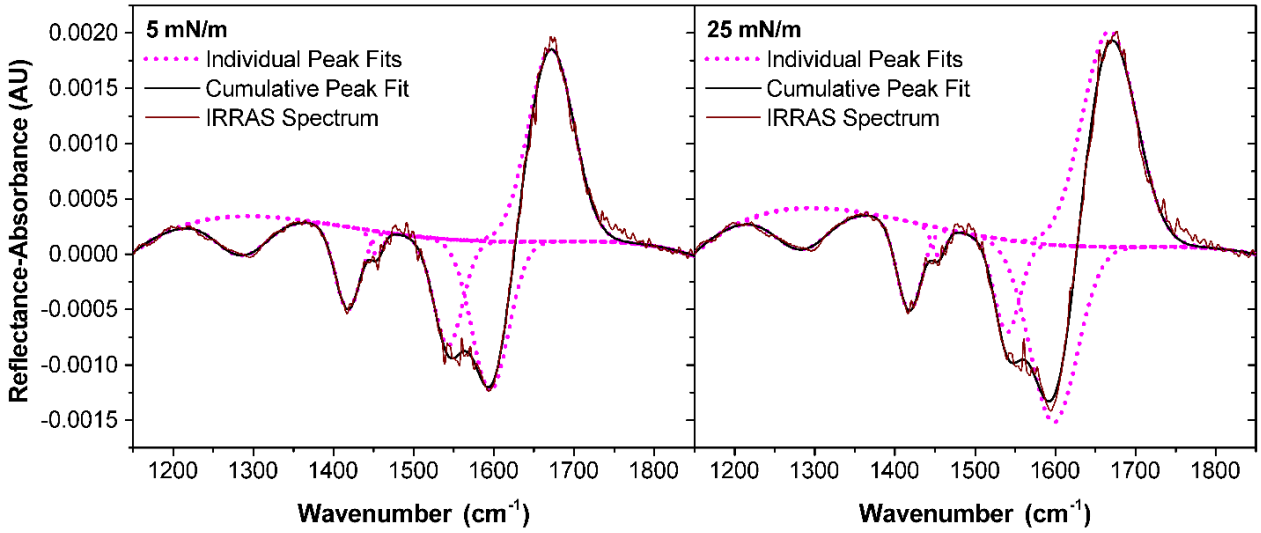


Figure 3.14: IRRAS spectra and corresponding peak fits of a d_{31}-palmitic acid monolayer at 5 mN/m (left) and 25 mN/m (right) spread onto an aqueous subphase of 0.47 M NaCl, 10 mM $CaCl_2$, and 50 ppm alginate at pH 8.2.

Table 3.9: Center wavelengths (λ, cm^{-1}), reflectance-absorbance intensities (RA Int.), peak areas, and full width at half maximum (FWHM, cm^{-1}) values of Gaussian fits to IRRAS spectra in the COOH vibrational mode region of a d_{31}-palmitic acid monolayer (5 mN/m and 25 mN/m) spread onto an aqueous subphase of 0.47 M NaCl, 10 mM $CaCl_2$, and 50 ppm alginate at pH 8.2.

Vibrational Mode	5 mN/m				25 mN/m			
	Center λ	RA Int.	Area	FWHM	Center λ	RA Int.	Area	FWHM
ν C-OH	1285.7	-3.58E-04	0.0291	76.3	1283.5	-3.77E-04	0.0318	79.2
ν_S COO^- (d_{31}-Palmitate)	1417.8	-7.45E-04	0.0293	36.9	1417.8	-8.07E-04	0.0314	36.5
ν_S COO^- (Alginate)	1454.3	-2.16E-4	0.00415	18.0	1454.7	-2.54E-4	0.00558	20.7
ν_{AS} COO^- (d_{31}-Palmitate)	1543.0	9.58E-04	0.0432	42.3	1539.8	-8.34E-04	0.0336	37.9
ν_{AS} COO^- (Alginate)	1595.4	0.00135	0.0768	53.6	1597.9	-0.00160	0.118	69.0
δ H-O-H	1671.3	0.00174	0.124	66.6	1667.3	0.00195	0.164	79.0

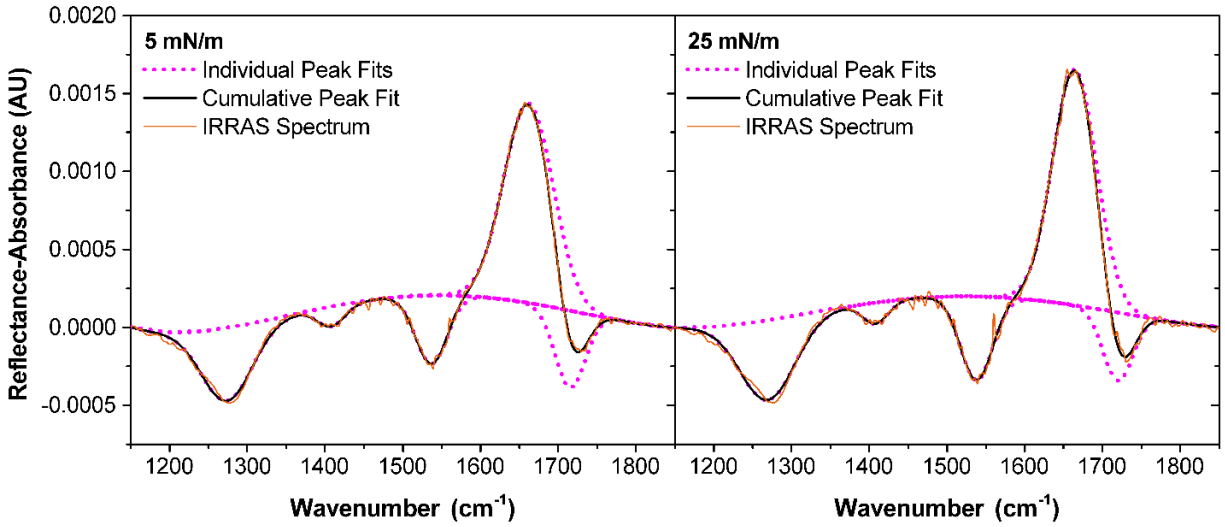


Figure 3.15: IRRAS spectra and corresponding peak fits of a d_{31}-palmitic acid monolayer at 5 mN/m (left) and 25 mN/m (right) spread onto an aqueous subphase of 0.47 M NaCl and 10 mM $CaCl_2$ at pH 5.8.

Table 3.10: Center wavelengths (λ, cm^{-1}), reflectance-absorbance intensities (RA Int.), peak areas, and full width at half maximum (FWHM, cm^{-1}) values of Gaussian fits to IRRAS spectra in the COOH vibrational mode region of a d_{31}-palmitic acid monolayer (5 mN/m and 25 mN/m) spread onto an aqueous subphase of 0.47 M NaCl and 10 mM $CaCl_2$ at pH 5.8.

Vibrational Mode	5 mN/m				25 mN/m			
	Center λ	RA Int.	Area	FWHM	Center λ	RA Int.	Area	FWHM
ν C-OH	1274.0	-4.67E-04	0.0377	75.8	1269.6	-5.11E-04	0.0448	82.4
ν_S COO^-	1408.2	-1.27E-04	0.00570	42.0	1407.9	-1.36E-04	0.00563	38.9
ν_{AS} COO^-	1536.9	-4.40E-04	0.0225	48.0	1538.0	-5.36E-04	0.0281	49.2
δ H-O-H	1661.6	0.00128	0.102	74.9	1664.1	0.00151	0.107	66.8
ν C=O	1715.2	-4.91E-04	0.0239	45.8	1719.7	-4.36E-04	0.0209	45.0

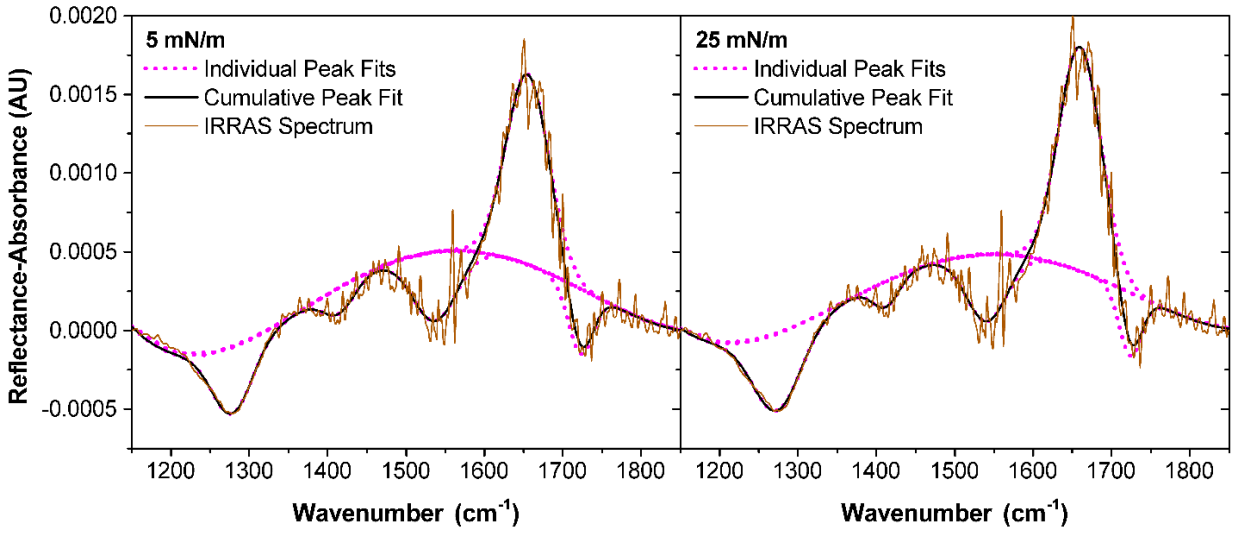


Figure 3.16: IRRAS spectra and corresponding peak fits of a d_{31}-palmitic acid monolayer at 5 mN/m (left) and 25 mN/m (right) spread onto an aqueous subphase of 0.47 M NaCl, 10 mM $CaCl_2$, and 50 ppm alginate at pH 5.8.

Table 3.11: Center wavelengths (λ, cm^{-1}), reflectance-absorbance intensities (RA Int.), peak areas, and full width at half maximum (FWHM, cm^{-1}) values of Gaussian fits to IRRAS spectra in the COOH vibrational mode region of a d_{31}-palmitic acid monolayer (5 mN/m and 25 mN/m) spread onto an aqueous subphase of 0.47 M NaCl, 10 mM $CaCl_2$, and 50 ppm alginate at pH 5.8.

Vibrational Mode	5 mN/m				25 mN/m			
	Center λ	RA Int.	Area	FWHM	Center λ	RA Int.	Area	FWHM
ν C-OH	1278.6	-4.18E-04	0.0264	59.2	1273.6	-4.86E-04	0.0363	70.1
ν_S COO^-	1414.3	-1.67E-04	0.00834	46.8	1412.1	-1.65E-04	0.00689	39.2
ν_{AS} COO^-	1536.8	-4.42E-04	0.0314	66.8	1540.6	-4.29E-04	0.0280	61.3
δ H-O-H	1654.9	0.00121	0.0829	64.6	1659.8	0.00142	0.0969	64.0
ν C=O	1722.6	-4.12E-04	0.0162	37.0	1724.1	-4.00E-04	0.0142	33.4

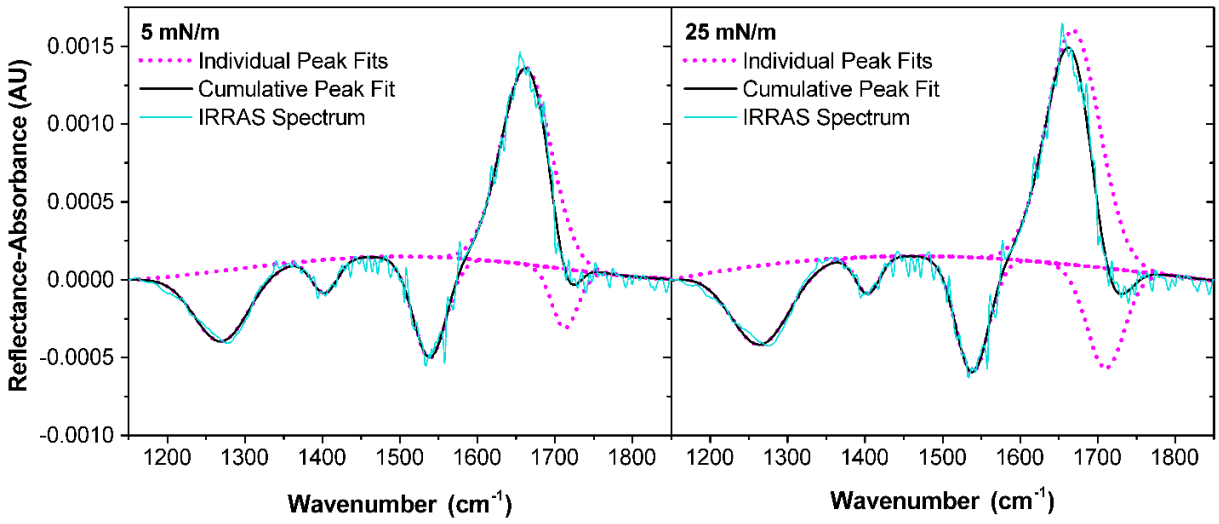


Figure 3.17: IRRAS spectra and corresponding peak fits of a d_{31}-palmitic acid monolayer at 5 mN/m (left) and 25 mN/m (right) spread onto an aqueous subphase of 0.47 M NaCl and 10 mM $MgCl_2$ at pH 8.2.

Table 3.12: Center wavelengths (λ, cm^{-1}), reflectance-absorbance intensities (RA Int.), peak areas, and full width at half maximum (FWHM, cm^{-1}) values of Gaussian fits to IRRAS spectra in the COOH vibrational mode region of a d_{31}-palmitic acid monolayer (5 mN/m and 25 mN/m) spread onto an aqueous subphase of 0.47 M NaCl and 10 mM $MgCl_2$ at pH 8.2.

Vibrational Mode	5 mN/m				25 mN/m			
	Center λ	RA Int.	Area	FWHM	Center λ	RA Int.	Area	FWHM
ν C-OH	1270.3	-4.53E-04	0.0398	82.6	1265.3	-5.08E-04	0.0484	89.4
ν_S COO^-	1403.0	-2.15E-04	0.00888	38.8	1403.3	-2.36E-04	0.00911	36.3
ν_{AS} COO^-	1538.4	-6.46E-04	0.0347	50.5	1538.2	-7.39E-04	0.0393	49.9
δ H-O-H	1662.6	0.00126	0.102	75.9	1668.1	0.00151	0.130	80.9
ν C=O	1712.6	-3.79E-04	0.0159	39.5	1710.1	-6.41E-04	0.0388	56.8

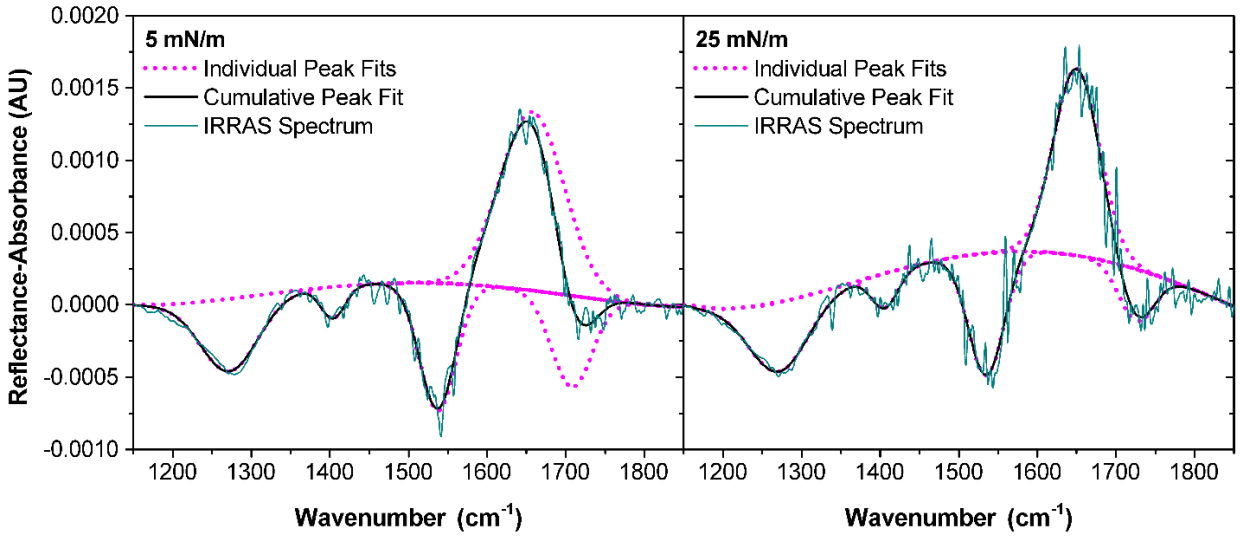


Figure 3.18: IRRAS spectra and corresponding peak fits of a d_{31}-palmitic acid monolayer at 5 mN/m (left) and 25 mN/m (right) spread onto an aqueous subphase of 0.47 M NaCl, 10 mM $MgCl_2$, and 50 ppm alginate at pH 8.2.

Table 3.13: Center wavelengths (λ, cm^{-1}), reflectance-absorbance intensities (RA Int.), peak areas, and full width at half maximum (FWHM, cm^{-1}) values of Gaussian fits to IRRAS spectra in the COOH vibrational mode region of a d_{31}-palmitic acid monolayer (5 mN/m and 25 mN/m) spread onto an aqueous subphase of 0.47 M NaCl, 10 mM $MgCl_2$, and 50 ppm alginate at pH 8.2.

Vibrational Mode	5 mN/m				25 mN/m			
	Center λ	RA Int.	Area	FWHM	Center λ	RA Int.	Area	FWHM
ν C-OH	1272.1	-4.99E-04	0.0455	85.5	1271.4	-4.78E-04	0.0439	86.2
ν_S COO^-	1404.0	-2.19E-04	0.00840	36.1	1405.7	-2.40E-04	0.0101	39.6
ν_{AS} COO^-	1537.3	-8.85E-04	0.0512	54.4	1534.9	-8.45E-04	0.0478	53.2
δ H-O-H	1657.2	0.00123	0.126	96.2	1649.7	0.00129	0.0944	68.7
ν C=O	1707.7	-6.31E-04	0.0405	60.3	1726.2	-3.58E-04	0.0200	52.4

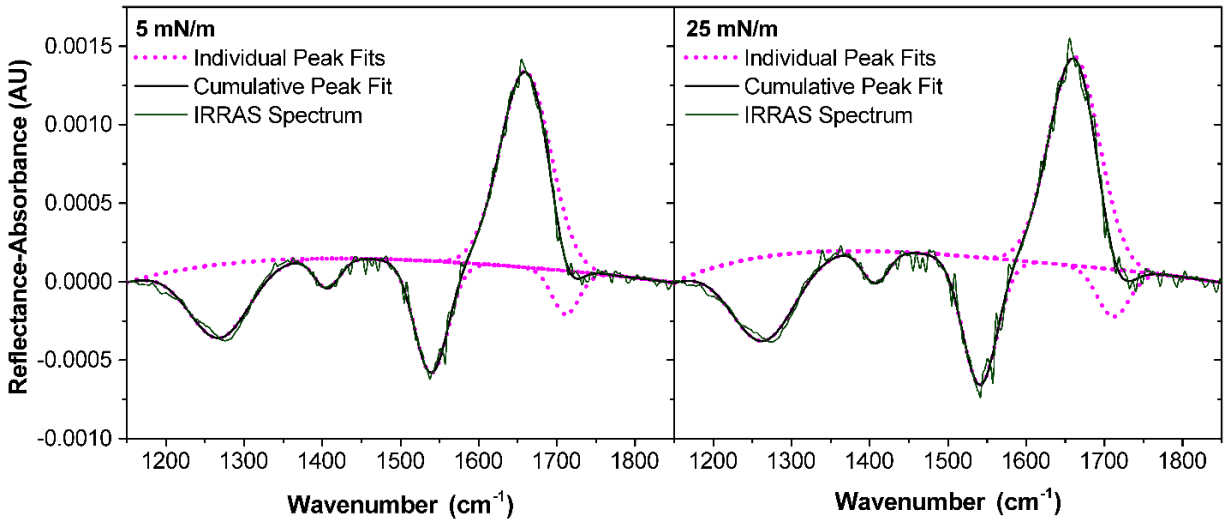


Figure 3.19: IRRAS spectra and corresponding peak fits of a d_{31}-palmitic acid monolayer at 5 mN/m (left) and 25 mN/m (right) spread onto an aqueous subphase of 0.47 M NaCl and 53 mM $MgCl_2$ at pH 8.2.

Table 3.14: Center wavelengths (λ, cm^{-1}), reflectance-absorbance intensities (RA Int.), peak areas, and full width at half maximum (FWHM, cm^{-1}) values of Gaussian fits to IRRAS spectra in the COOH vibrational mode region of a d_{31}-palmitic acid monolayer (5 mN/m and 25 mN/m) spread onto an aqueous subphase of 0.47 M NaCl and 53 mM $MgCl_2$ at pH 8.2.

Vibrational Mode	5 mN/m				25 mN/m			
	Center λ	RA Int.	Area	FWHM	Center λ	RA Int.	Area	FWHM
ν C-OH	1268.6	-4.69E-04	0.0447	89.6	1264.1	-5.37E-04	0.0543	95.1
ν_S COO^-	1406.0	-1.89E-04	0.00768	38.1	1406.8	-2.05E-04	0.00834	38.3
ν_{AS} COO^-	1539.3	-7.13E-04	0.0381	50.1	1540.6	-8.13E-04	0.0445	51.4
δ H-O-H	1659.5	0.00125	0.100	75.6	1661.1	0.00133	0.107	75.4
ν C=O	1711.0	-2.77E-04	0.0119	40.4	1711.1	-3.01E-04	0.0154	48.2

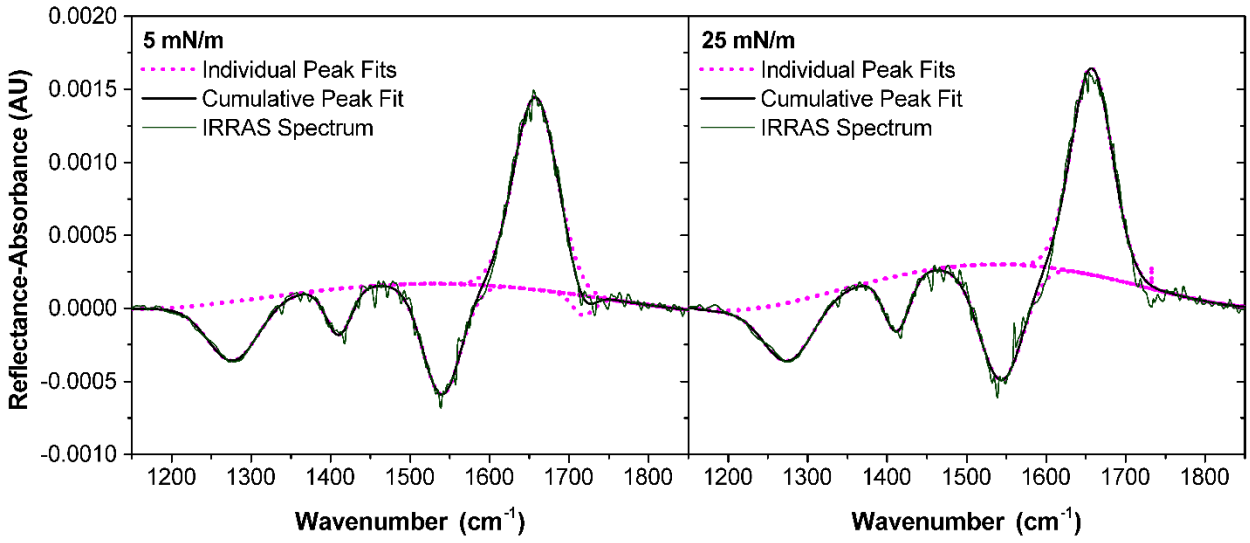


Figure 3.20: IRRAS spectra and corresponding peak fits of a d_{31}-palmitic acid monolayer at 5 mN/m (left) and 25 mN/m (right) spread onto an aqueous subphase of 0.47 M NaCl, 53 mM $MgCl_2$, and 50 ppm alginate at pH 8.2.

Table 3.15: Center wavelengths (λ, cm^{-1}), reflectance-absorbance intensities (RA Int.), peak areas, and full width at half maximum (FWHM, cm^{-1}) values of Gaussian fits to IRRAS spectra in the COOH vibrational mode region of a d_{31}-palmitic acid monolayer (5 mN/m and 25 mN/m) spread onto an aqueous subphase of 0.47 M NaCl, 53 mM $MgCl_2$, and 50 ppm alginate at pH 8.2.

Vibrational Mode	5 mN/m				25 mN/m			
	Center λ	RA Int.	Area	FWHM	Center λ	RA Int.	Area	FWHM
ν C-OH	1277.5	-4.05E-04	0.0322	74.7	1276.1	-3.97E-04	0.0308	72.9
ν_S COO^-	1410.3	-3.18E-04	0.0126	37.1	1412.1	-3.80E-04	0.0150	37.2
ν_{AS} COO^-	1539.9	-7.59E-04	0.0453	56.0	1543.5	-7.94E-04	0.0540	63.8
δ H-O-H	1657.6	0.00132	0.0910	64.9	1657.4	0.00141	0.0916	60.9
ν C=O	1717.7	-1.35E-04	0.00451	31.3	1732.6	1.60E-04	7.92E-07	0.00464

3.7.5 d_{33}-Cetyl Alcohol IRRAS Spectral Analysis

The carboxylic acid stretching region (1150-1850 cm^{-1}) was analyzed for the d_{33}-cetyl alcohol spectra to further examine the extent of alginate co-adsorption to the monolayer (**Figure 3.20**). Cetyl alcohol only exhibits the C-OH stretching mode, so any carboxylic acid peaks can be attributed to alginate co-adsorption. Neither the ν_S COO^-, ν_{AS} COO^-, nor the ν C=O stretches appear at either surface pressure, indicating no alginate co-adsorption. The δ H-O-H mode red shifts in response to alginate within the aqueous subphase, possibly caused by alginate solvation within the probing region of IRRAS.

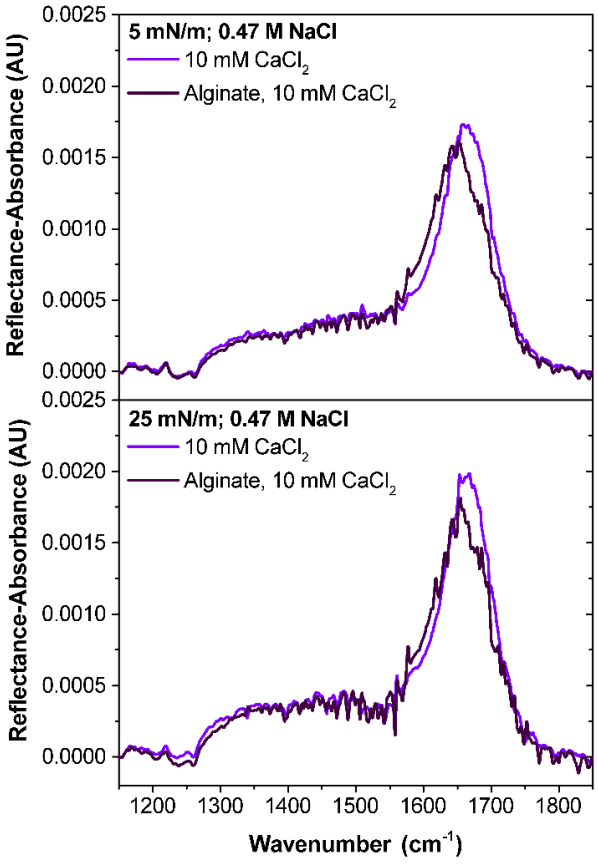


Figure 3.21: IRRAS spectra of d_{33}-cetyl alcohol collected and analyzed within the COOH stretching region at constant surface pressures of 5 mN/m (top) and 25 mN/m (bottom). The light purple curves correspond to the monolayer spread onto a 0.47 M NaCl and 10 mM $CaCl_2$ subphase at pH 8.2. The dark purple curves correspond to the monolayer spread onto a 0.47 M NaCl, 10 mM $CaCl_2$, and 50 ppm alginate subphase. All spectra represent averages of at least triplicate measurements.

3.7.6 CD_2 Scissoring Mode Analysis

The CD_2 scissoring mode region (1070-1110 cm^{-1}) was analyzed using OriginPro 9.0. A baseline was fitted to a line between the two endpoints of the region, and then the baseline was subtracted from the spectrum. Following baseline subtraction, the center frequency of the CD_2 scissoring mode was determined with the Peak Analyzer tool in Origin.

Table 3.16: Center wavelengths (λ, cm^{-1}) of the IRRAS CD_2 scissoring mode of d_{31}-palmitic acid (d_{31}-PA) and d_{33}-cetyl alcohol (d_{33}-CA) monolayers measured at constant surface pressures 5 mN/m and 25 mN/m.

Monolayer & Subphase	CD_2 Scissoring Mode λ	
	5 mN/m	25 mN/m
d_{31}-PA, 0.47 M NaCl, pH 8.2	1089.5	1089.0
d_{31}-PA, 50 ppm Alginate, 0.47 M NaCl, pH 8.2	1089.5	1089.0
d_{31}-PA, 0.47 M NaCl, 10 mM $CaCl_2$, pH 8.2	1089.4	1089.4
d_{31}-PA, 50 ppm Alginate, 0.47 M NaCl, 10 mM $CaCl_2$, pH 8.2	1089.4	1089.5
d_{31}-PA, 0.47 M NaCl, 10 mM $CaCl_2$, pH 5.8	1089.5	1089.5
d_{31}-PA, 50 ppm Alginate, 0.47 M NaCl, 10 mM $CaCl_2$, pH 5.8	1089.5	1089.5
d_{33}-CA, 0.47 M NaCl, 10 mM $CaCl_2$, pH 8.2	1089.0	1089.0
d_{33}-CA, 50 ppm Alginate, 0.47 M NaCl, 10 mM $CaCl_2$, pH 8.2	1089.0	1089.0
d_{31}-PA, 0.47 M NaCl, 10 mM $MgCl_2$, pH 8.2	1089.5	1089.0
d_{31}-PA, 50 ppm Alginate, 0.47 M NaCl, 10 mM $MgCl_2$, pH 8.2	1089.5	1089.0
d_{31}-PA, 0.47 M NaCl, 53 mM $MgCl_2$, pH 8.2	1089.5	1089.0
d_{31}-PA, 50 ppm Alginate, 0.47 M NaCl, 53 mM $MgCl_2$, pH 8.2	1089.5	1089.5

3.7.7 Radial Distribution and Number Density Profiles

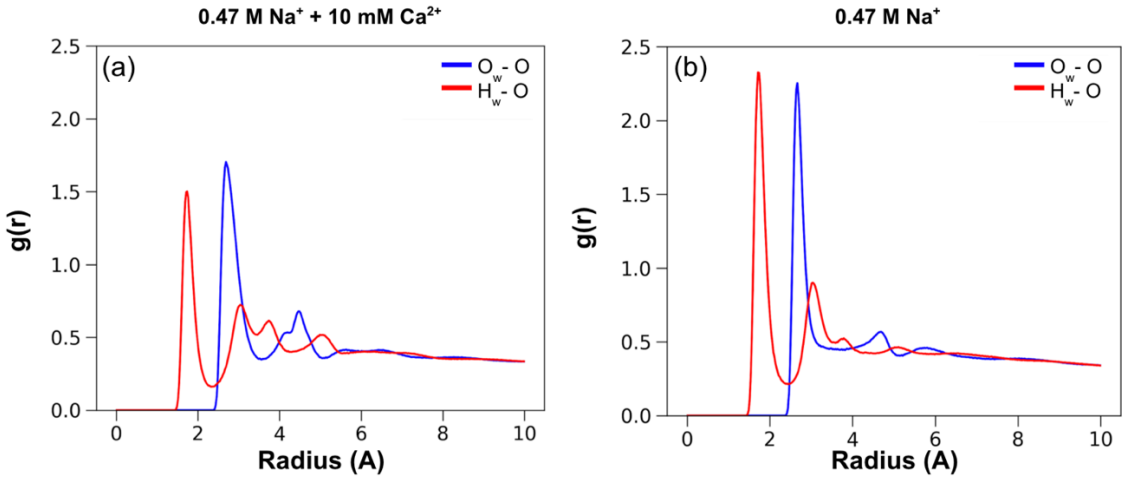


Figure 3.22: RDF of water O_w and H_w with respect to palmitate headgroup -OH oxygen in the presence (a) and absence (b) of calcium. Both systems contain alginate in the aqueous phase.

3.7.8 Local Order Parameters

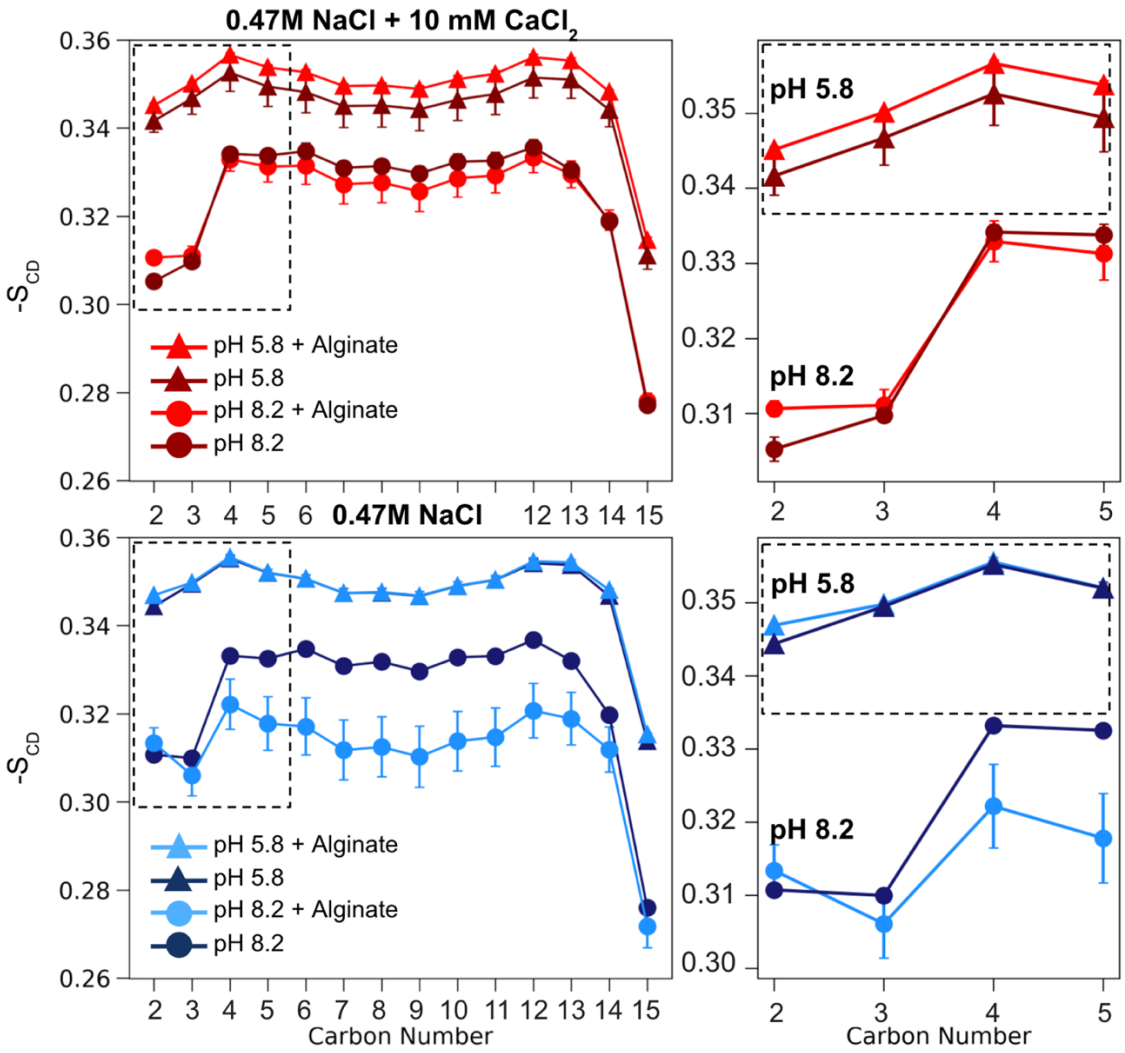


Figure 3.23: Local order parameters for all systems, not separated based on protonation state, with close-ups of the carbon C2-C4 regions.

3.7.9 IRRAS Spectra of Alginate Co-Adsorption via Magnesium Bridging

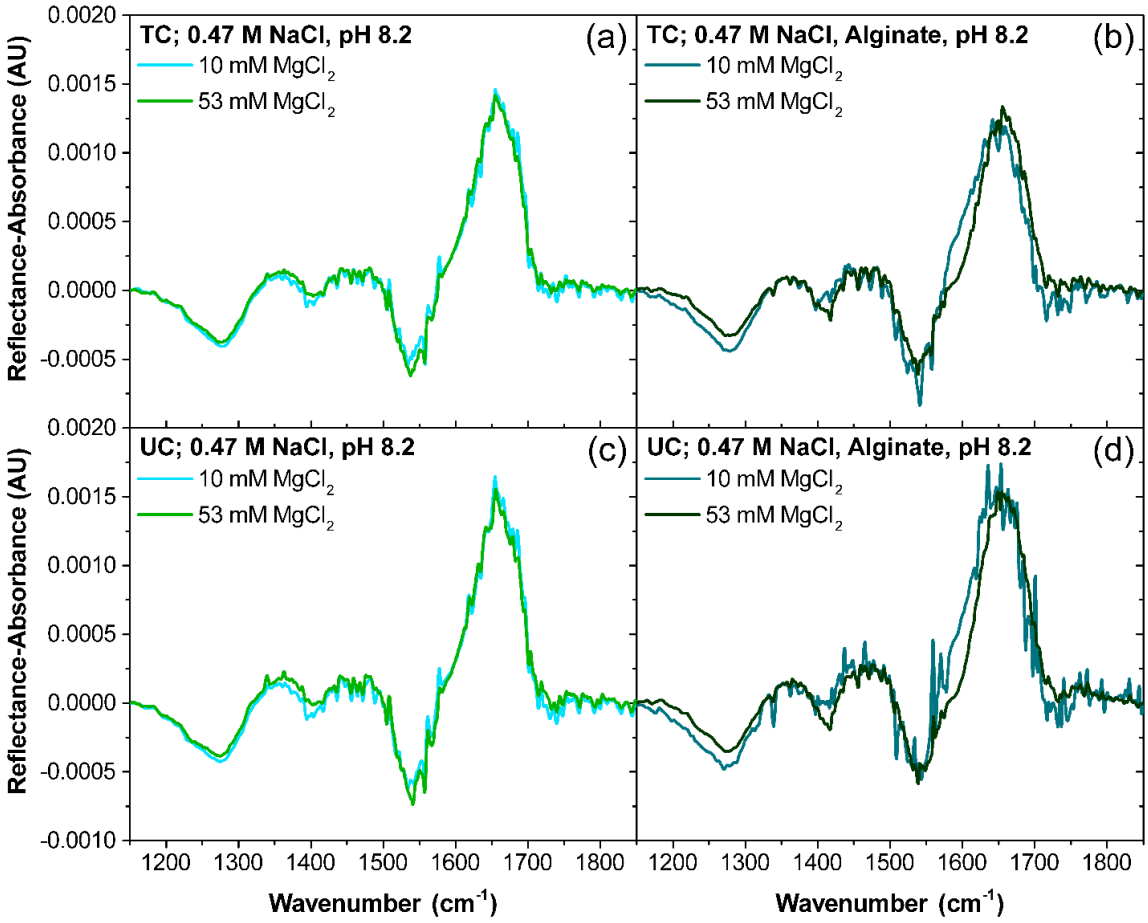


Figure 3.24: IRRAS spectra of the COOH stretching region indicate weak alginate co-adsorption to the d_{31}-palmitic acid monolayer via Mg^{2+} bridging interactions. Spectral lines are color-coded to indicate differences in solution composition. Surface pressure was held constant in the (a), (b) tilted condensed (5 mN/m) and (c), (d) untilted condensed (25 mN/m) phases throughout spectral acquisition. Spectra corresponding to the salt water solutions are shown in (a) and (c), and spectra corresponding to the salt water solutions containing 50 ppm alginate are shown in (b) and (d).

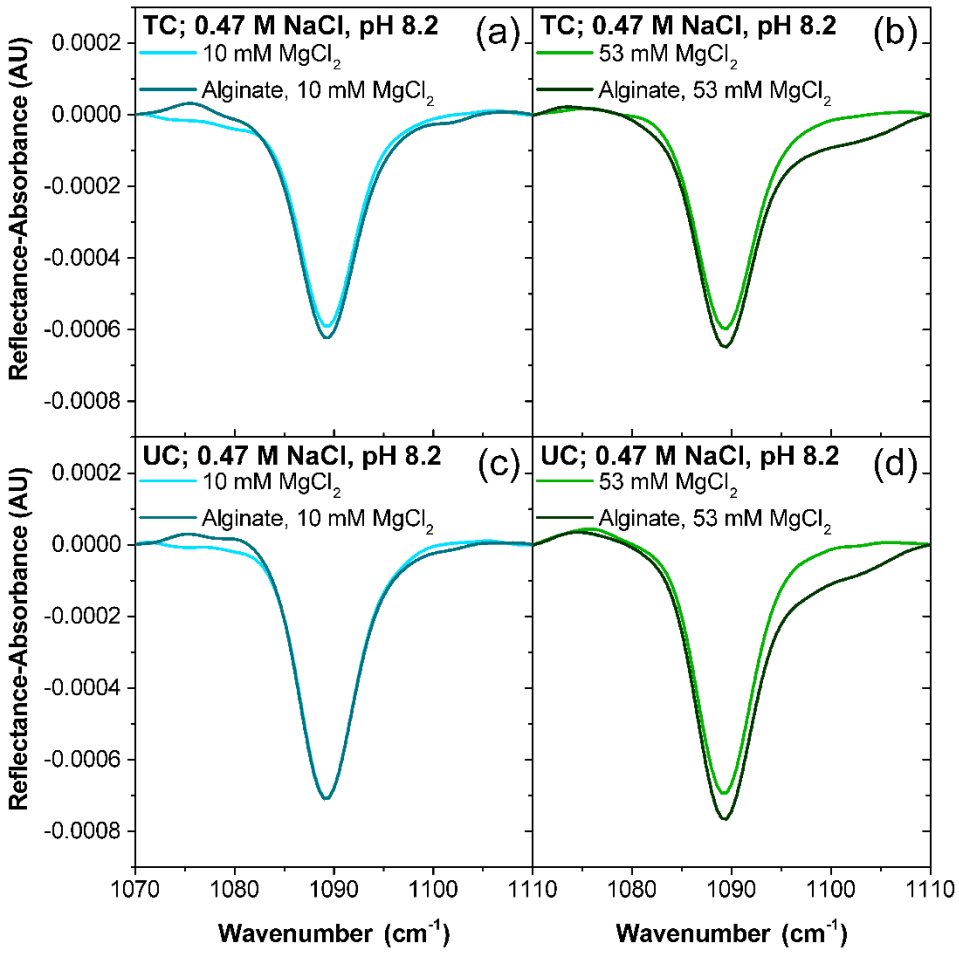


Figure 3.25: IRRAS spectra of the CD_2 scissoring mode region demonstrate weak alginate co-adsorption to the d_{31}-palmitic acid monolayer in the presence of 53 mM $MgCl_2$ at pH 8.2, and the spectra indicate no changes in the d_{31}-palmitic acid lattice packing upon adsorption. Spectral lines are color-coded to indicate differences in solution composition. Surface pressure was held constant in the (a), (b) tilted condensed (5 mN/m) and (c), (d) untilted condensed (25 mN/m) phases throughout spectral acquisition.

References

(1) Hawkins, L. N.; Russell, L. M. Polysaccharides, Proteins, and Phytoplankton Fragments: Four Chemically Distinct Types of Marine Primary Organic Aerosol Classified by Single Particle Spectromicroscopy. *Adv. Meteorol.* **2010**, *2010*, 1–14. https://doi.org/10.1155/2010/612132.

(2) Russell, L. M.; Hawkins, L. N.; Frossard, A. A.; Quinn, P. K.; Bates, T. S. Carbohydrate-like Composition of Submicron Atmospheric Particles and Their Production from Ocean Bubble Bursting. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107* (15), 6652–6657. https://doi.org/10.1073/pnas.0908905107.

(3) Aller, J. Y.; Radway, J. C.; Kilthau, W. P.; Bothe, D. W.; Wilson, T. W.; Vaillancourt, R. D.; Quinn, P. K.; Coffman, D. J.; Murray, B. J.; Knopf, D. A. Size-Resolved Characterization of the Polysaccharidic and Proteinaceous Components of Sea Spray Aerosol. *Atmos. Environ.* **2017**, *154*, 331–347. https://doi.org/10.1016/j.atmosenv.2017.01.053.

(4) Elliott, S.; Burrows, S. M.; Deal, C.; Liu, X.; Long, M.; Ogunro, O.; Russell, L. M.; Wingenter, O. Prospects for Simulating Macromolecular Surfactant Chemistry at the Ocean–Atmosphere Boundary. *Environ. Res. Lett.* **2014**, *9* (6), 064012. https://doi.org/10.1088/1748-9326/9/6/064012.

(5) Gao, Q.; Leck, C.; Rauschenberg, C.; Matrai, P. A. On the Chemical Dynamics of Extracellular Polysaccharides in the High Arctic Surface Microlayer. *Ocean Sci.* **2012**, *8* (4), 401–418. https://doi.org/10.5194/os-8-401-2012.

(6) Jayarathne, T.; Sultana, C. M.; Lee, C.; Malfatti, F.; Cox, J. L.; Pendergraft, M. A.; Moore, K. A.; Azam, F.; Tivanski, A. V.; Cappa, C. D.; Bertram, T. H.; Grassian, V. H.; Prather, K. A.; Stone, E. A. Enrichment of Saccharides and Divalent Cations in Sea Spray Aerosol During Two Phytoplankton Blooms. *Environ. Sci. Technol.* **2016**, *50* (21), 11511–11520. https://doi.org/10.1021/acs.est.6b02988.

(7) Thornton, D. C. O.; Brooks, S. D.; Chen, J. Protein and Carbohydrate Exopolymer Particles in the Sea Surface Microlayer (SML). *Front. Mar. Sci.* **2016**, *3*, 1–14. https://doi.org/10.3389/fmars.2016.00135.

(8) Hasenecz, E. S.; Jayarathne, T.; Pendergraft, M. A.; Santander, M. V.; Mayer, K. J.; Sauer, J.; Lee, C.; Gibson, W. S.; Kruse, S. M.; Malfatti, F.; Prather, K. A.; Stone, E. A. Marine Bacteria Affect Saccharide Enrichment in Sea Spray Aerosol during a Phytoplankton Bloom. *ACS Earth Space Chem.* **2020**, *4* (9), 1638–1649. https://doi.org/10.1021/acsearthspacechem.0c00167.

(9) Rastelli, E.; Corinaldesi, C.; Dell'Anno, A.; Lo Martire, M.; Greco, S.; Cristina Facchini, M.; Rinaldi, M.; O'Dowd, C.; Ceburnis, D.; Danovaro, R. Transfer of Labile Organic Matter and Microbes from the Ocean Surface to the Marine Aerosol: An Experimental Approach. *Sci. Rep.* **2017**, *7* (1), 11475. https://doi.org/10.1038/s41598-017-10563-z.

(10) Burrows, S. M.; Gobrogge, E.; Fu, L.; Link, K.; Elliott, S. M.; Wang, H.; Walker, R. OCEANFILMS-2: Representing Coadsorption of Saccharides in Marine Films and Potential Impacts on Modeled Marine Aerosol Chemistry. *Geophys. Res. Lett.* **2016**, *43* (15), 8306–8313. https://doi.org/10.1002/2016GL069070.

(11) Link, K. A.; Hsieh, C.-Y.; Tuladhar, A.; Chase, Z.; Wang, Z.; Wang, H.; Walker, R. A. Vibrational Studies of Saccharide-Induced Lipid Film Reorganization at Aqueous/Air Interfaces. *Chem. Phys.* **2018**, *512*, 104–110. https://doi.org/10.1016/j.chemphys.2018.02.011.

(12) Schill, S. R.; Burrows, S. M.; Hasenecz, E. S.; Stone, E. A.; Bertram, T. H. The Impact of Divalent Cations on the Enrichment of Soluble Saccharides in Primary Sea Spray Aerosol. *Atmosphere* **2018**, *9* (12), 476. https://doi.org/10.3390/atmos9120476.

(13) Link, K. A.; Spurzem, G. N.; Tuladhar, A.; Chase, Z.; Wang, Z.; Wang, H.; Walker, R. A. Organic Enrichment at Aqueous Interfaces: Cooperative Adsorption of Glucuronic Acid to DPPC Monolayers Studied with Vibrational Sum Frequency Generation. *J. Phys. Chem. A* **2019**, *123* (26), 5621–5632. https://doi.org/10.1021/acs.jpca.9b02255.

(14) Link, K. A.; Spurzem, G. N.; Tuladhar, A.; Chase, Z.; Wang, Z.; Wang, H.; Walker, R. A. Cooperative Adsorption of Trehalose to DPPC Monolayers at the Water–Air Interface Studied with Vibrational Sum Frequency Generation. *J. Phys. Chem. B* **2019**, *123* (42), 8931–8938. https://doi.org/10.1021/acs.jpcb.9b07770.

(15) Burrows, S. M.; Ogunro, O.; Frossard, A. A.; Russell, L. M.; Rasch, P. J.; Elliott, S. M. A Physically Based Framework for Modeling the Organic Fractionation of Sea Spray Aerosol from Bubble Film

Langmuir Equilibria. *Atmos. Chem. Phys.* **2014**, *14* (24), 13601–13629. https://doi.org/10.5194/acp-14-13601-2014.

(16) Quinn, P. K.; Collins, D. B.; Grassian, V. H.; Prather, K. A.; Bates, T. S. Chemistry and Related Properties of Freshly Emitted Sea Spray Aerosol. *Chem. Rev.* **2015**, *115* (10), 4383–4399. https://doi.org/10.1021/cr500713g.

(17) Cochran, R. E.; Laskina, O.; Trueblood, J. V.; Estillore, A. D.; Morris, H. S.; Jayarathne, T.; Sultana, C. M.; Lee, C.; Lin, P.; Laskin, J.; Laskin, A.; Dowling, J. A.; Qin, Z.; Cappa, C. D.; Bertram, T. H.; Tivanski, A. V.; Stone, E. A.; Prather, K. A.; Grassian, V. H. Molecular Diversity of Sea Spray Aerosol Particles: Impact of Ocean Biology on Particle Composition and Hygroscopicity. *Chem* **2017**, *2* (5), 655–667. https://doi.org/10.1016/j.chempr.2017.03.007.

(18) Bertram, T. H.; Cochran, R. E.; Grassian, V. H.; Stone, E. A. Sea Spray Aerosol Chemical Composition: Elemental and Molecular Mimics for Laboratory Studies of Heterogeneous and Multiphase Reactions. *Chem. Soc. Rev.* **2018**, *47* (7), 2374–2400. https://doi.org/10.1039/C7CS00008A.

(19) Leck, C.; Bigg, E. K. Biogenic Particles in the Surface Microlayer and Overlaying Atmosphere in the Central Arctic Ocean during Summer. *Tellus B* **2005**, *57* (4), 305–316. https://doi.org/10.3402/tellusb.v57i4.16546.

(20) Orellana, M. V.; Matrai, P. A.; Leck, C.; Rauschenberg, C. D.; Lee, A. M.; Coz, E. Marine Microgels as a Source of Cloud Condensation Nuclei in the High Arctic. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108* (33), 13612–13617. https://doi.org/10.1073/pnas.1102457108.

(21) Leck, C.; Gao, Q.; Mashayekhy Rad, F.; Nilsson, U. Size-Resolved Atmospheric Particulate Polysaccharides in the High Summer Arctic. *Atmos. Chem. Phys.* **2013**, *13* (24), 12573–12588. https://doi.org/10.5194/acp-13-12573-2013.

(22) Leck, C.; Svensson, E. Importance of Aerosol Composition and Mixing State for Cloud Droplet Activation over the Arctic Pack Ice in Summer. *Atmos. Chem. Phys.* **2015**, *15* (5), 2545–2568. https://doi.org/10.5194/acp-15-2545-2015.

(23) Pummer, B. G.; Bauer, H.; Bernardi, J.; Bleicher, S.; Grothe, H. Suspendable Macromolecules Are Responsible for Ice Nucleation Activity of Birch and Conifer Pollen. *Atmos. Chem. Phys.* **2012**, *12* (5), 2541–2550. https://doi.org/10.5194/acp-12-2541-2012.

(24) Hiranuma, N.; Möhler, O.; Yamashita, K.; Tajiri, T.; Saito, A.; Kiselev, A.; Hoffmann, N.; Hoose, C.; Jantsch, E.; Koop, T.; Murakami, M. Ice Nucleation by Cellulose and Its Potential Contribution to Ice Formation in Clouds. *Nat. Geosci.* **2015**, *8* (4), 273–277. https://doi.org/10.1038/ngeo2374.

(25) Pummer, B. G.; Budke, C.; Augustin-Bauditz, S.; Niedermeier, D.; Felgitsch, L.; Kampf, C. J.; Huber, R. G.; Liedl, K. R.; Loerting, T.; Moschen, T.; Schauperl, M.; Tollinger, M.; Morris, C. E.; Wex, H.; Grothe, H.; Pöschl, U.; Koop, T.; Fröhlich-Nowoisky, J. Ice Nucleation by Water-Soluble Macromolecules. *Atmos. Chem. Phys.* **2015**, *15* (8), 4077–4091. https://doi.org/10.5194/acp-15-4077-2015.

(26) Wilson, T. W.; Ladino, L. A.; Alpert, P. A.; Breckels, M. N.; Brooks, I. M.; Browse, J.; Burrows, S. M.; Carslaw, K. S.; Huffman, J. A.; Judd, C.; Kilthau, W. P.; Mason, R. H.; McFiggans, G.; Miller, L. A.; Nájera, J. J.; Polishchuk, E.; Rae, S.; Schiller, C. L.; Si, M.; Temprado, J. V.; Whale, T. F.; Wong, J. P. S.; Wurl, O.; Yakobi-Hancock, J. D.; Abbatt, J. P. D.; Aller, J. Y.; Bertram, A. K.; Knopf, D. A.; Murray,

B. J. A Marine Biogenic Source of Atmospheric Ice-Nucleating Particles. *Nature* **2015**, *525* (7568), 234–238. https://doi.org/10.1038/nature14986.

(27) Dreischmeier, K.; Budke, C.; Wiehemeier, L.; Kottke, T.; Koop, T. Boreal Pollen Contain Ice-Nucleating as Well as Ice-Binding 'Antifreeze' Polysaccharides. *Sci. Rep.* **2017**, *7*, 41890. https://doi.org/10.1038/srep41890.

(28) Hiranuma, N.; Adachi, K.; Bell, D. M.; Belosi, F.; Beydoun, H.; Bhaduri, B.; Bingemer, H.; Budke, C.; Clemen, H.-C.; Conen, F.; Cory, K. M.; Curtius, J.; DeMott, P. J.; Eppers, O.; Grawe, S.; Hartmann, S.; Hoffmann, N.; Höhler, K.; Jantsch, E.; Kiselev, A.; Koop, T.; Kulkarni, G.; Mayer, A.; Murakami, M.; Murray, B. J.; Nicosia, A.; Petters, M. D.; Piazza, M.; Polen, M.; Reicher, N.; Rudich, Y.; Saito, A.; Santachiara, G.; Schiebel, T.; Schill, G. P.; Schneider, J.; Segev, L.; Stopelli, E.; Sullivan, R. C.; Suski, K.; Szakáll, M.; Tajiri, T.; Taylor, H.; Tobo, Y.; Ullrich, R.; Weber, D.; Wex, H.; Whale, T. F.; Whiteside, C. L.; Yamashita, K.; Zelenyuk, A.; Möhler, O. A Comprehensive Characterization of Ice Nucleation by Three Different Types of Cellulose Particles Immersed in Water. *Atmos. Chem. Phys.* **2019**, *19* (7), 4823–4849. https://doi.org/10.5194/acp-19-4823-2019.

(29) Wolf, M. J.; Coe, A.; Dove, L. A.; Zawadowicz, M. A.; Dooley, K.; Biller, S. J.; Zhang, Y.; Chisholm, S. W.; Cziczo, D. J. Investigating the Heterogeneous Ice Nucleation of Sea Spray Aerosols Using Prochlorococcus as a Model Source of Marine Organic Matter. *Environ. Sci. Technol.* **2019**, *53* (3), 1139–1149. https://doi.org/10.1021/acs.est.8b05150.

(30) Zeppenfeld, S.; van Pinxteren, M.; Hartmann, M.; Bracher, A.; Stratmann, F.; Herrmann, H. Glucose as a Potential Chemical Marker for Ice Nucleating Activity in Arctic Seawater and Melt Pond Samples. *Environ. Sci. Technol.* **2019**, *53* (15), 8747–8756. https://doi.org/10.1021/acs.est.9b01469.

(31) Hasenecz, E. S.; Kaluarachchi, C. P.; Lee, H. D.; Tivanski, A. V.; Stone, E. A. Saccharide Transfer to Sea Spray Aerosol Enhanced by Surface Activity, Calcium, and Protein Interactions. *ACS Earth Space Chem.* **2019**, *3* (11), 2539–2548. https://doi.org/10.1021/acsearthspacechem.9b00197.

(32) Gorin, P. a. J.; Spencer, J. F. T. Exocellular Alginic Acid from Azotobacter Vinelandii. *Can. J. Chem.* **1966**, *44* (9), 993–998. https://doi.org/10.1139/v66-147.

(33) Linker, A.; Jones, R. S. A New Polysaccharide Resembling Alginic Acid Isolated from Pseudomonads. *J. Biol. Chem.* **1966**, *241* (16), 3845–3851.

(34) Haug, A.; Larsen, B.; Fykse, O.; Block-Bolten, A.; Toguri, J. M.; Flood, H. Quantitative Determination of the Uronic Acid Composition of Alginates. *Acta Chem. Scand.* **1962**, *16*, 1908–1918. https://doi.org/10.3891/acta.chem.scand.16-1908.

(35) Haug, A.; Larsen, B.; Smidsrød, O.; Smidsrød, O.; Eriksson, G.; Blinc, R.; Paušak, S.; Ehrenberg, L.; Dumanović, J. Studies on the Sequence of Uronic Acid Residues in Alginic Acid. *Acta Chem. Scand.* **1967**, *21*, 691–704. https://doi.org/10.3891/acta.chem.scand.21-0691.

(36) Haug, A.; Myklestad, S.; Larsen, B.; Smidsrød, O.; Eriksson, G.; Blinc, R.; Paušak, S.; Ehrenberg, L.; Dumanović, J. Correlation between Chemical Structure and Physical Properties of Alginates. *Acta Chem. Scand.* **1967**, *21*, 768–778. https://doi.org/10.3891/acta.chem.scand.21-0768.

(37) Grasdalen, H.; Larsen, B.; Smidsrød, O. A p.m.r. Study of the Composition and Sequence of Uronate Residues in Alginates. *Carbohyd. Res.* **1979**, *68* (1), 23–31. https://doi.org/10.1016/S0008-6215(00)84051-3.

(38) Johnson, F. A.; Craig, D. Q. M.; Mercer, A. D. Characterization of the Block Structure and Molecular Weight of Sodium Alginates. *J. Pharm. Pharmacol.* **1997**, *49* (7), 639–643. https://doi.org/10.1111/j.2042-7158.1997.tb06085.x.

(39) Lee, K. Y.; Mooney, D. J. Alginate: Properties and Biomedical Applications. *Prog. Polym. Sci.* **2012**, *37* (1), 106–126. https://doi.org/10.1016/j.progpolymsci.2011.06.003.

(40) Morris, E. R.; Rees, D. A.; Thom, D.; Boyd, J. Chiroptical and Stoichiometric Evidence of a Specific, Primary Dimerisation Process in Alginate Gelation. *Carbohyd. Res.* **1978**, *66* (1), 145–154. https://doi.org/10.1016/S0008-6215(00)83247-4.

(41) Fang, Y.; Al-Assaf, S.; Phillips, G. O.; Nishinari, K.; Funami, T.; Williams, P. A.; Li, L. Multiple Steps and Critical Behaviors of the Binding of Calcium to Alginate. *J. Phys. Chem. B* **2007**, *111* (10), 2456–2462. https://doi.org/10.1021/jp0689870.

(42) Hecht, H.; Srebnik, S. Structural Characterization of Sodium Alginate and Calcium Alginate. *Biomacromolecules* **2016**, *17* (6), 2160–2167. https://doi.org/10.1021/acs.biomac.6b00378.

(43) Kohn, R. Ion Binding on Polyuronates - Alginate and Pectin. *Pure Appl. Chem.* **1975**, *42* (3), 371–397. https://doi.org/10.1351/pac197542030371.

(44) Malovíková, A.; Rinaudo, M.; Milas, M. Comparative Interactions of Magnesium and Calcium Counterions with Polygalacturonic Acid. *Biopolymers* **1994**, *34* (8), 1059–1064. https://doi.org/10.1002/bip.360340809.

(45) Donati, I.; Cesàro, A.; Paoletti, S. Specific Interactions versus Counterion Condensation. 1. Nongelling Ions/Polyuronate Systems. *Biomacromolecules* **2006**, *7* (1), 281–287. https://doi.org/10.1021/bm050646p.

(46) Donati, I.; Asaro, F.; Paoletti, S. Experimental Evidence of Counterion Affinity in Alginates: The Case of Nongelling Ion Mg^{2+}. *J. Phys. Chem. B* **2009**, *113* (39), 12877–12886. https://doi.org/10.1021/jp902912m.

(47) Topuz, F.; Henke, A.; Richtering, W.; Groll, J. Magnesium Ions and Alginate Do Form Hydrogels: A Rheological Study. *Soft Matter* **2012**, *8* (18), 4877. https://doi.org/10.1039/c2sm07465f.

(48) Huynh, U. T. D.; Lerbret, A.; Neiers, F.; Chambin, O.; Assifaoui, A. Binding of Divalent Cations to Polygalacturonate: A Mechanism Driven by the Hydration Water. *J. Phys. Chem. B* **2016**, *120* (5), 1021–1032. https://doi.org/10.1021/acs.jpcb.5b11010.

(49) Chin, W.-C.; Orellana, M. V.; Verdugo, P. Spontaneous Assembly of Marine Dissolved Organic Matter into Polymer Gels. *Nature* **1998**, *391* (6667), 568–572. https://doi.org/10.1038/35345.

(50) Verdugo, P.; Alldredge, A. L.; Azam, F.; Kirchman, D. L.; Passow, U.; Santschi, P. H. The Oceanic Gel Phase: A Bridge in the DOM–POM Continuum. *Mar. Chem.* **2004**, *92* (1–4), 67–85. https://doi.org/10.1016/j.marchem.2004.06.017.

(51) Verdugo, P.; V. Orellana, M.; Chin, W.-C.; W. Petersen, T.; Eng, G. van den; Benner, R.; I. Hedges, J. Marine Biopolymer Self-Assembly : Implications for Carbon Cycling in the Ocean. *Faraday Discuss.* **2008**, *139* (0), 393–398. https://doi.org/10.1039/B800149A.

(52) Verdugo, P. Marine Microgels. *Annu. Rev. Mar. Sci.* **2012**, *4* (1), 375–400. https://doi.org/10.1146/annurev-marine-120709-142759.

(53) Li, X.; Leck, C.; Sun, L.; Hede, T.; Tu, Y.; Ågren, H. Cross-Linked Polysaccharide Assemblies in Marine Gels: An Atomistic Simulation. *J. Phys. Chem. Lett.* **2013**, *4* (16), 2637–2642. https://doi.org/10.1021/jz401276r.

(54) Sun, L.; Li, X.; Hede, T.; Tu, Y.; Leck, C.; Ågren, H. Molecular Dynamics Simulations Reveal the Assembly Mechanism of Polysaccharides in Marine Aerosols. *Phys. Chem. Chem. Phys.* **2014**, *16* (47), 25935–25941. https://doi.org/10.1039/C4CP03423F.

(55) Cochran, R. E.; Laskina, O.; Jayarathne, T.; Laskin, A.; Laskin, J.; Lin, P.; Sultana, C.; Lee, C.; Moore, K. A.; Cappa, C. D.; Bertram, T. H.; Prather, K. A.; Grassian, V. H.; Stone, E. A. Analysis of Organic Anionic Surfactants in Fine and Coarse Fractions of Freshly Emitted Sea Spray Aerosol. *Environ. Sci. Technol.* **2016**, *50* (5), 2477–2486. https://doi.org/10.1021/acs.est.5b04053.

(56) Rad, F. M.; Leck, C.; Ilag, L. L.; Nilsson, U. Investigation of Ultrahigh-Performance Liquid Chromatography/Travelling-Wave Ion Mobility/Time-of-Flight Mass Spectrometry for Fast Profiling of Fatty Acids in the High Arctic Sea Surface Microlayer. *Rapid Commun. Mass Sp.* **2018**, *32* (12), 942–950. https://doi.org/10.1002/rcm.8109.

(57) Tervahattu, H.; Juhanoja, J.; Kupiainen, K. Identification of an Organic Coating on Marine Aerosol Particles by TOF-SIMS. *J. Geophys. Res.* **2002**, *107* (D16), ACH 18-1. https://doi.org/10.1029/2001JD001403.

(58) Mochida, M.; Kitamori, Y.; Kawamura, K.; Nojiri, Y.; Suzuki, K. Fatty Acids in the Marine Atmosphere: Factors Governing Their Concentrations and Evaluation of Organic Films on Sea-Salt Particles. *J. Geophys. Res.* **2002**, *107* (D17), 4325. https://doi.org/10.1029/2001JD001278.

(59) Adams, E. M.; Allen, H. C. Palmitic Acid on Salt Subphases and in Mixed Monolayers of Cerebrosides: Application to Atmospheric Aerosol Chemistry. *Atmosphere* **2013**, *4* (4), 315–336. https://doi.org/10.3390/atmos4040315.

(60) Adams, E. M.; Wellen, B. A.; Thiraux, R.; Reddy, S. K.; Vidalis, A. S.; Paesani, F.; Allen, H. C. Sodium–Carboxylate Contact Ion Pair Formation Induces Stabilization of Palmitic Acid Monolayers at High PH. *Phys. Chem. Chem. Phys.* **2017**, *19* (16), 10481–10490. https://doi.org/10.1039/C7CP00167C.

(61) Rudd, B. A. W.; Vidalis, A. S.; Allen, H. C. Thermodynamic versus Non-Equilibrium Stability of Palmitic Acid Monolayers in Calcium-Enriched Sea Spray Aerosol Proxy Systems. *Phys. Chem. Chem. Phys.* **2018**, *20* (24), 16320–16332. https://doi.org/10.1039/C8CP01188E.

(62) Shrestha, M.; Luo, M.; Li, Y.; Xiang, B.; Xiong, W.; H. Grassian, V. Let There Be Light: Stability of Palmitic Acid Monolayers at the Air/Salt Water Interface in the Presence and Absence of Simulated Solar Light and a Photosensitizer. *Chem. Sci.* **2018**, *9* (26), 5716–5723. https://doi.org/10.1039/C8SC01957F.

(63) Degen, P.; Paulus, M.; Zwar, E.; Jakobi, V.; Dogan, S.; Tolan, M.; Rehage, H. Surfactant-Mediated Formation of Alginate Layers at the Water-Air Interface. *Surf. Interface Anal.* **2019**, *51* (11), 1051–1058. https://doi.org/10.1002/sia.6691.

(64) Pavinatto, F. J.; Pavinatto, A.; Caseli, L.; dos Santos, David S.; Nobre, T. M.; Zaniquelli, M. E. D.; Oliveira, O. N. Interaction of Chitosan with Cell Membrane Models at the Air−Water Interface. *Biomacromolecules* **2007**, *8* (5), 1633–1640. https://doi.org/10.1021/bm0701550.

(65) Wydro, P.; Krajewska, B.; Hąc-Wydro, K. Chitosan as a Lipid Binder: A Langmuir Monolayer Study of Chitosan−Lipid Interactions. *Biomacromolecules* **2007**, *8* (8), 2611–2617. https://doi.org/10.1021/bm700453x.

(66) Krajewska, B.; Wydro, P.; Jańczyk, A. Probing the Modes of Antibacterial Activity of Chitosan. Effects of PH and Molecular Weight on Chitosan Interactions with Membrane Lipids in Langmuir Films. *Biomacromolecules* **2011**, *12* (11), 4144–4152. https://doi.org/10.1021/bm2012295.

(67) Krajewska, B.; Wydro, P.; Kyzioł, A. Chitosan as a Subphase Disturbant of Membrane Lipid Monolayers. The Effect of Temperature at Varying PH: I. DPPG. *Colloid Surface A* **2013**, *434*, 349–358. https://doi.org/10.1016/j.colsurfa.2013.03.015.

(68) Pavinatto, A.; Souza, A. L.; Delezuk, J. A. M.; Pavinatto, F. J.; Campana-Filho, S. P.; Oliveira, O. N. Interaction of O-Acylated Chitosans with Biomembrane Models: Probing the Effects from Hydrophobic Interactions and Hydrogen Bonding. *Colloid Surface B* **2014**, *114*, 53–59. https://doi.org/10.1016/j.colsurfb.2013.09.034.

(69) Pavinatto, A.; Delezuk, J. A. M.; Souza, A. L.; Pavinatto, F. J.; Volpati, D.; Miranda, P. B.; Campana-Filho, S. P.; Oliveira Jr., O. N. Experimental Evidence for the Mode of Action Based on Electrostatic and Hydrophobic Forces to Explain Interaction between Chitosans and Phospholipid Langmuir Monolayers. *Colloid Surface B* **2016**, *145*, 201–207. https://doi.org/10.1016/j.colsurfb.2016.05.001.

(70) Parra-Barraza, H.; Burboa, M. G.; Sánchez-Vázquez, M.; Juárez, J.; Goycoolea, F. M.; Valdez, M. A. Chitosan−Cholesterol and Chitosan−Stearic Acid Interactions at the Air−Water Interface. *Biomacromolecules* **2005**, *6* (5), 2416–2426. https://doi.org/10.1021/bm050106s.

(71) de Meijere, K.; Brezesinski, G.; Zschörnig, O.; Arnold, K.; Möhwald, H. Structure Studies of a Phospholipid Monolayer Coupled to Dextran Sulfate. *Physica B* **1998**, *248* (1), 269–273. https://doi.org/10.1016/S0921-4526(98)00246-4.

(72) Huster, D.; Arnold, K. Ca^{2+}-Mediated Interaction between Dextran Sulfate and Dimyristoyl-*Sn*-Glycero-3-Phosphocholine Surfaces Studied by ^{2}H Nuclear Magnetic Resonance. *Biophys. J.* **1998**, *75* (2), 909–916. https://doi.org/10.1016/S0006-3495(98)77579-4.

(73) Huster, D.; Paasche, G.; Dietrich, U.; Zschörnig, O.; Gutberlet, T.; Gawrisch, K.; Arnold, K. Investigation of Phospholipid Area Compression Induced by Calcium-Mediated Dextran Sulfate Interaction. *Biophys. J.* **1999**, *77* (2), 879–887. https://doi.org/10.1016/S0006-3495(99)76939-0.

(74) Hua, W.; Verreault, D.; Adams, E. M.; Huang, Z.; Allen, H. C. Impact of Salt Purity on Interfacial Water Organization Revealed by Conventional and Heterodyne-Detected Vibrational Sum Frequency Generation Spectroscopy. *J. Phys. Chem. C* **2013**, *117* (38), 19577–19585. https://doi.org/10.1021/jp408146t.

(75) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14* (1), 33–38. https://doi.org/10.1016/0263-7855(96)00018-5.

(76) Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. PACKMOL: A package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.* **2009**, *30* (13), 2157–2164. https://doi.org/10.1002/jcc.21224.

(77) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM. *J. Comput. Chem.* **2008**, *29* (11), 1859–1865. https://doi.org/10.1002/jcc.20945.

(78) Jo, S.; Song, K. C.; Desaire, H.; MacKerell, A. D.; Im, W. Glycan Reader: Automated Sugar Identification and Simulation Preparation for Carbohydrates and Glycoproteins. *J. Comput. Chem.* **2011**, *32* (14), 3135–3141. https://doi.org/10.1002/jcc.21886.

(79) Park, S.-J.; Lee, J.; Patel, D. S.; Ma, H.; Lee, H. S.; Jo, S.; Im, W. Glycan Reader Is Improved to Recognize Most Sugar Types and Chemical Modifications in the Protein Data Bank. *Bioinformatics* **2017**, *33* (19), 3051–3057. https://doi.org/10.1093/bioinformatics/btx358.

(80) Park, S.-J.; Lee, J.; Qi, Y.; Kern, N. R.; Lee, H. S.; Jo, S.; Joung, I.; Joo, K.; Lee, J.; Im, W. CHARMM-GUI *Glycan Modeler* for Modeling and Simulation of Carbohydrates and Glycoconjugates. *Glycobiology* **2019**, *29* (4), 320–331. https://doi.org/10.1093/glycob/cwz003.

(81) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79* (2), 926–935. https://doi.org/10.1063/1.445869.

(82) Huang, J.; MacKerell, A. D. CHARMM36 All-Atom Additive Protein Force Field: Validation Based on Comparison to NMR Data. *J. Comput. Chem.* **2013**, *34* (25), 2135–2145. https://doi.org/10.1002/jcc.23354.

(83) Davidchack, R. L.; Handel, R.; Tretyakov, M. V. Langevin Thermostat for Rigid Body Dynamics. *J. Chem. Phys.* **2009**, *130* (23), 234101. https://doi.org/10.1063/1.3149788.

(84) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An $N \cdot \log(N)$ Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98* (12), 10089–10092. https://doi.org/10.1063/1.464397.

(85) Perez, F.; Granger, B. E. IPython: A System for Interactive Scientific Computing. *Comput. Sci. Eng.* **2007**, *9* (3), 21–29. https://doi.org/10.1109/MCSE.2007.53.

(86) Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9* (7), 3084–3095. https://doi.org/10.1021/ct400341p.

(87) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. GROMACS: A Message-Passing Parallel Molecular Dynamics Implementation. *Comput. Phys. Commun.* **1995**, *91* (1–3), 43–56. https://doi.org/10.1016/0010-4655(95)00042-E.

(88) Millero, F. J.; Feistel, R.; Wright, D. G.; McDougall, T. J. The Composition of Standard Seawater and the Definition of the Reference-Composition Salinity Scale. *Deep-Sea Res. Pt. I* **2008**, *55* (1), 50–72. https://doi.org/10.1016/j.dsr.2007.10.001.

(89) Millero, F. J. *Chemical Oceanography*, 4th ed.; Taylor & Francis: Boca Raton, 2013.

(90) Miranda, P. B.; Du, Q.; Shen, Y. R. Interaction of Water with a Fatty Acid Langmuir Film. *Chem. Phys. Lett.* **1998**, *286* (1–2), 1–8. https://doi.org/10.1016/S0009-2614(97)01476-0.

(91) Kanicky, J. R.; Shah, D. O. Effect of Degree, Type, and Position of Unsaturation on the PKa of Long-Chain Fatty Acids. *J. Colloid Interf. Sci.* **2002**, *256* (1), 201–207. https://doi.org/10.1006/jcis.2001.8009.

(92) McLean, D. S.; Vercoe, D.; Stack, K. R.; Richardson, D. The Colloidal PKa of Lipophilic Extractives Commonly Found in Pinus Radiata. *Appita J.* **2005**, *58* (5), 362.

(93) Bu, H.; Kjøniksen, A.-L.; Elgsaeter, A.; Nyström, B. Interaction of Unmodified and Hydrophobically Modified Alginate with Sodium Dodecyl Sulfate in Dilute Aqueous Solution: Calorimetric, Rheological, and Turbidity Studies. *Colloid Surface A* **2006**, *278* (1), 166–174. https://doi.org/10.1016/j.colsurfa.2005.12.016.

(94) Marion, G. M.; Millero, F. J.; Camões, M. F.; Spitzer, P.; Feistel, R.; Chen, C.-T. A. PH of Seawater. *Mar. Chem.* **2011**, *126* (1), 89–96. https://doi.org/10.1016/j.marchem.2011.04.002.

(95) Denton, J. K.; Kelleher, P. J.; Johnson, M. A.; Baer, M. D.; Kathmann, S. M.; Mundy, C. J.; Rudd, B. A. W.; Allen, H. C.; Choi, T. H.; Jordan, K. D. Molecular-Level Origin of the Carboxylate Head Group Response to Divalent Metal Ion Complexation at the Air–Water Interface. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116* (30), 14874–14880. https://doi.org/10.1073/pnas.1818600116.

(96) Luo, M.; Wauer, N. A.; Angle, K. J.; Dommer, A. C.; Song, M.; Nowak, C. M.; Amaro, R. E.; Grassian, V. H. Insights into the Behavior of Nonanoic Acid and Its Conjugate Base at the Air/Water Interface through a Combined Experimental and Theoretical Approach. *Chem. Sci.* **2020**, *11* (39), 10647–10656. https://doi.org/10.1039/D0SC02354J.

(97) Aida, T. M.; Yamagata, T.; Watanabe, M.; Smith, R. L. Depolymerization of Sodium Alginate under Hydrothermal Conditions. *Carbohyd. Polym.* **2010**, *80* (1), 296–302. https://doi.org/10.1016/j.carbpol.2009.11.032.

(98) Lu, J.; Yang, H.; Hao, J.; Wu, C.; Liu, L.; Xu, N.; Linhardt, R. J.; Zhang, Z. Impact of Hydrolysis Conditions on the Detection of Mannuronic to Guluronic Acid Ratio in Alginate and Its Derivatives. *Carbohyd. Polym.* **2015**, *122*, 180–188. https://doi.org/10.1016/j.carbpol.2015.01.008.

(99) Carter-Fenk, K. A.; Allen, H. C. Collapse Mechanisms of Nascent and Aged Sea Spray Aerosol Proxy Films. *Atmosphere* **2018**, *9* (12), 503. https://doi.org/10.3390/atmos9120503.

(100) Butt, H.-J.; Graf, K.; Kappl, M. *Physics and Chemistry of Interfaces*, Third, Revised and Enlarged Edition.; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2013.

(101) Tang, C. Y.; Huang, Z.; Allen, H. C. Interfacial Water Structure and Effects of Mg^{2+} and Ca^{2+} Binding to the COOH Headgroup of a Palmitic Acid Monolayer Studied by Sum Frequency Spectroscopy. *J. Phys. Chem. B* **2011**, *115* (1), 34–40. https://doi.org/10.1021/jp1062447.

(102) Balzerowski, P.; Meister, K.; Versluis, J.; Bakker, H. J. Heterodyne-Detected Sum Frequency Generation Spectroscopy of Polyacrylic Acid at the Air/Water-Interface. *Phys. Chem. Chem. Phys.* **2016**, *18* (4), 2481–2487. https://doi.org/10.1039/C5CP06177F.

(103) Tanaka, N.; Kitano, H.; Ise, N. Raman Spectroscopic Study of Hydrogen Bonding in Aqueous Carboxylic Acid Solutions. 3. Polyacrylic Acid. *Macromolecules* **1991**, *24* (10), 3017–3019. https://doi.org/10.1021/ma00010a060.

(104) Gragson, D. E.; Richmond, G. L. Investigations of the Structure and Hydrogen Bonding of Water Molecules at Liquid Surfaces by Vibrational Sum Frequency Spectroscopy. *J. Phys. Chem. B* **1998**, *102* (20), 3847–3861. https://doi.org/10.1021/jp9806011.

(105) Allen, H. C.; Casillas-Ituarte, N. N.; Sierra-Hernández, M. R.; Chen, X.; Tang, C. Y. Shedding Light on Water Structure at Air–Aqueous Interfaces: Ions, Lipids, and Hydration. *Phys. Chem. Chem. Phys.* **2009**, *11* (27), 5538–5549. https://doi.org/10.1039/B901209E.

(106) Le Calvez, E.; Blaudez, D.; Buffeteau, T.; Desbat, B. Effect of Cations on the Dissociation of Arachidic Acid Monolayers on Water Studied by Polarization-Modulated Infrared Reflection−Absorption Spectroscopy. *Langmuir* **2001**, *17* (3), 670–674. https://doi.org/10.1021/la000224v.

(107) Kundu, S.; Langevin, D. Fatty Acid Monolayer Dissociation and Collapse: Effect of PH and Cations. *Colloid Surface A* **2008**, *325* (1–2), 81–85. https://doi.org/10.1016/j.colsurfa.2008.04.037.

(108) Tang, C. Y.; Huang, Z.; Allen, H. C. Binding of Mg^{2+} and Ca^{2+} to Palmitic Acid and Deprotonation of the COOH Headgroup Studied by Vibrational Sum Frequency Generation Spectroscopy. *J. Phys. Chem. B* **2010**, *114* (51), 17068–17076. https://doi.org/10.1021/jp105472e.

(109) Neal, J. F.; Zhao, W.; Grooms, A. J.; Flood, A. H.; Allen, H. C. Arginine–Phosphate Recognition Enhanced in Phospholipid Monolayers at Aqueous Interfaces. *J. Phys. Chem. C* **2018**, *122* (46), 26362–26371. https://doi.org/10.1021/acs.jpcc.8b03531.

(110) Mendelsohn, R.; Brauner, J. W.; Gericke, A. External Infrared Reflection Absorption Spectrometry of Monolayer Films at the Air-Water Interface. *Annu. Rev. Phys. Chem.* **1995**, *46* (1), 305–334. https://doi.org/10.1146/annurev.pc.46.100195.001513.

(111) Lin, W.; Clark, A. J.; Paesani, F. Effects of Surface Pressure on the Properties of Langmuir Monolayers and Interfacial Water at the Air–Water Interface. *Langmuir* **2015**, *31* (7), 2147–2156. https://doi.org/10.1021/la504603s.

(112) Grant, G. T.; Morris, E. R.; Rees, D. A.; Smith, P. J. C.; Thom, D. Biological Interactions between Polysaccharides and Divalent Cations: The Egg-Box Model. *FEBS Lett.* **1973**, *32* (1), 195–198. https://doi.org/10.1016/0014-5793(73)80770-7.

(113) Stokke, B. T.; Smidsroed, O.; Bruheim, P.; Skjaak-Braek, G. Distribution of Uronate Residues in Alginate Chains in Relation to Alginate Gelling Properties. *Macromolecules* **1991**, *24* (16), 4637–4645. https://doi.org/10.1021/ma00016a026.

(114) Stokke, B. T.; Draget, K. I.; Smidsrød, O.; Yuguchi, Y.; Urakawa, H.; Kajiwara, K. Small-Angle X-Ray Scattering and Rheological Characterization of Alginate Gels. 1. Ca−Alginate Gels. *Macromolecules* **2000**, *33* (5), 1853–1863. https://doi.org/10.1021/ma991559q.

(115) Li, L.; Fang, Y.; Vreeker, R.; Appelqvist, I.; Mendes, E. Reexamining the Egg-Box Model in Calcium−Alginate Gels with X-Ray Diffraction. *Biomacromolecules* **2007**, *8* (2), 464–468. https://doi.org/10.1021/bm060550a.

(116) Shao, Y.; Gan, Z.; Epifanovsky, E.; Gilbert, A. T. B.; Wormit, M.; Kussmann, J.; Lange, A. W.; Behn, A.; Deng, J.; Feng, X.; Ghosh, D.; Goldey, M.; Horn, P. R.; Jacobson, L. D.; Kaliman, I.; Khaliullin, R. Z.; Kuś, T.; Landau, A.; Liu, J.; Proynov, E. I.; Rhee, Y. M.; Richard, R. M.; Rohrdanz, M. A.; Steele, R. P.; Sundstrom, E. J.; III, H. L. W.; Zimmerman, P. M.; Zuev, D.; Albrecht, B.; Alguire, E.; Austin, B.; Beran, G. J. O.; Bernard, Y. A.; Berquist, E.; Brandhorst, K.; Bravaya, K. B.; Brown, S. T.; Casanova, D.; Chang, C.-M.; Chen, Y.; Chien, S. H.; Closser, K. D.; Crittenden, D. L.; Diedenhofen, M.; Jr, R. A. D.; Do, H.; Dutoi, A. D.; Edgar, R. G.; Fatehi, S.; Fusti-Molnar, L.; Ghysels, A.; Golubeva-Zadorozhnaya, A.; Gomes, J.; Hanson-Heine, M. W. D.; Harbach, P. H. P.; Hauser, A. W.; Hohenstein, E. G.; Holden, Z. C.; Jagau, T.-C.; Ji, H.; Kaduk, B.; Khistyaev, K.; Kim, J.; Kim, J.; King, R. A.; Klunzinger, P.; Kosenkov, D.; Kowalczyk, T.; Krauter, C. M.; Lao, K. U.; Laurent, A. D.; Lawler, K. V.; Levchenko, S. V.; Lin, C. Y.; Liu, F.; Livshits, E.; Lochan, R. C.; Luenser, A.; Manohar, P.; Manzer, S. F.; Mao, S.-P.; Mardirossian, N.; Marenich, A. V.; Maurer, S. A.; Mayhall, N. J.; Neuscamman, E.; Oana, C. M.; Olivares-Amaya, R.; O'Neill, D. P.; Parkhill, J. A.; Perrine, T. M.; Peverati, R.; Prociuk, A.; Rehn, D. R.; Rosta, E.; Russ, N. J.; Sharada, S. M.; Sharma, S.; Small, D. W.; Sodt, A.; Stein, T.; Stück, D.; Su, Y.-C.; Thom, A. J. W.; Tsuchimochi, T.; Vanovschi, V.; Vogt, L.; Vydrov, O.; Wang, T.; Watson, M. A.; Wenzel, J.; White, A.; Williams, C. F.; Yang, J.; Yeganeh, S.; Yost, S. R.; You, Z.-Q.; Zhang, I. Y.; Zhang, X.; Zhao, Y.; Brooks, B. R.; Chan, G. K. L.; Chipman, D. M.; Cramer, C. J.; III, W. A. G.; Gordon, M. S.; Hehre, W. J.; Klamt, A.; III, H. F. S.; Schmidt, M. W.; Sherrill, C. D.; Truhlar, D. G.; Warshel, A.; Xu, X.; Aspuru-Guzik, A.; Baer, R.; Bell, A. T.; Besley, N. A.; Chai, J.-D.; Dreuw, A.; Dunietz, B. D.; Furlani, T. R.; Gwaltney, S. R.; Hsu, C.-P.; Jung, Y.; Kong, J.; Lambrecht, D. S.; Liang, W.; Ochsenfeld, C.; Rassolov, V. A.; Slipchenko, L. V.; Subotnik, J. E.; Voorhis, T. V.; Herbert, J. M.; Krylov, A. I.; Gill, P. M. W.; Head-Gordon, M. Advances in Molecular Quantum Chemistry Contained in the Q-Chem 4 Program Package. *Mol. Phys.* **2015**, *113* (2), 184–215. https://doi.org/10.1080/00268976.2014.952696.

(117) Atomic Masses and Isotopic Abundances. *CRC Handbook of Chemistry and Physics*; Rumble, J. R., Ed.; CRC Press/Taylor & Francis: Boca Raton, FL.

(118) Lin, C. Y.; George, M. W.; Gill, P. M. W. EDF2: A Density Functional for Predicting Molecular Vibrational Frequencies. *Aust. J. Chem.* **2004**, *57* (4), 365. https://doi.org/10.1071/CH03263.

Part II
Sea Spray Aerosol Morphology and Dynamics

Chapter 4

Cation-Driven Lipopolysaccharide Morphological Changes Impact Heterogeneous Reactions of Nitric Acid with Sea Spray Aerosol Particles

4.1 Abstract

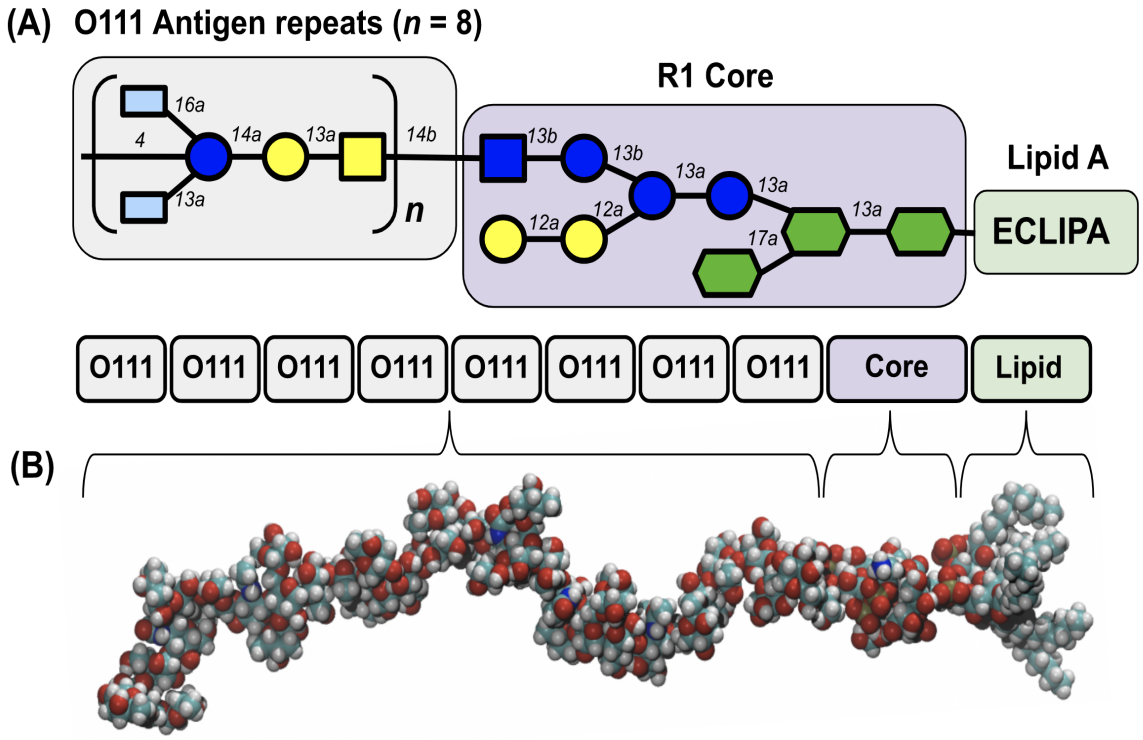
Lipopolysaccharides (LPS) in sea spray aerosol (SSA) particles have recently been shown to undergo heterogeneous reactions with HNO_3 in the atmosphere. Here, we integrate theory and experiment to further investigate how the most abundant sea salt cations, Na^+, Mg^{2+}, and Ca^{2+}, impact HNO_3 reactions with LPS-containing SSA particles. Aerosol reaction flow tube studies show that heterogeneous reactions of SSA particles with divalent cation (Mg^{2+} and Ca^{2+}) and LPS signatures were less reactive with HNO_3 than those dominated by monovalent cations (Na^+). All-atom molecular dynamics simulations of model LPS aggregates suggest that divalent cations cross-link the oligosaccharide chains to increase molecular aggregation and rigidity, which changes the particle phase and morphology, decreases water diffusion, and consequently decreases the reactive uptake of HNO_3. This study provides new insight into how complex chemical interactions between ocean-derived salts and biogenic organic species can impact the heterogeneous reactivity of SSA particles.

4.2 Introduction

Sea spray aerosol (SSA) particles are one of the most abundant naturally-produced aerosols in our atmosphere. Yet their heterogeneous reactivity in the atmosphere is not well-understood as chemically simple systems have only been investigated.[1–4] Understanding the heterogeneous reactivity of SSA particles is important for water uptake and cloud formation which is key to controlling our climate and reducing one of the greatest uncertainties in current climate models.[5] SSA particles are composed of a wide array of inorganic and organic components, ranging from simple mixtures of sea salt to entire microbial cells.[6–8] The composition of each SSA particle impacts its climate-relevant properties, including the interaction with gas phase species like HNO_3 that are important for regulating nitrogen levels in the global nitrogen cycle.[9] Up to 63.9% of SSA particles with aerodynamic sizes greater than 1 μm (supermicron SSA) contain polysaccharide signatures, including

lipopolysaccharides (LPS), as determined from Raman microspectroscopy.[8] Recently, LPS species have been shown to undergo acid-base reactions with HNO_3 in SSA particles, changing the paradigm of heterogeneous reaction between SSA and HNO_3 involving only inorganic species.[3,10]

LPS is a major component of outer-cell membranes of gram-negative bacteria, known constituents of microbial life within the sea surface microlayer, and a source of SSA components. The chemical structure of LPS, which primarily protects microbial cells, is expected to also impact SSA reaction pathways. Each LPS molecule is composed of three parts: (1) lipid A, (2) inner and outer core oligosaccharides, and (3) repeating O-antigens, a repetitive glycan polymer (Scheme 4.1).[11,12] Upon reaction with HNO_3, LPS can undergo acid-base chemistry by becoming protonated at the carboxylate and phosphate groups located within the inner and outer core oligosaccharides. This reaction process displaces coordinating cations (i.e., sodium, calcium, magnesium). The significance of this heterogeneous reaction pathway in nascent SSA particles was recently established through single particle studies by Raman microspectroscopy and aerosol time-of-flight mass spectrometry (ATOFMS).[10]



Scheme 4.1: LPS molecular structure and components. (A) Glycan symbol representation of model LPS used in this study. (B) Hard-sphere atomic model of LPS corresponding to the diagram in (A); white = hydrogen; teal = carbon; red = oxygen; blue = nitrogen; and tan = phosphorous. Detailed chemical structure can be found in Trueblood et al.[6]

Different cations are known to elicit unique morphological properties in LPS monolayers and bilayers.[13–16] Coughlin et al. reported that purified sodium LPS salt forms tube-like aggregates, while the presence of divalent cations calcium or magnesium induces the formation of LPS bilayer aggregates.[14,15] Many studies have also noted that the high activity of divalent cations leads to the displacement of monovalent cations weakly-bound to the saccharide headgroups, with calcium binding more tightly relative to magnesium.[13,15–18] The striking sensitivity of LPS aggregate morphology to cation type, and the high preference for calcium binding, has also been associated with the resistance of bacteria to commercial sanitizers; calcium is suggested to cross-link the saccharide chains together, preventing membrane penetration by the weaker cationic surfactants used in sanitizers.[13] A similar mechanism may play a role in the reactive uptake of trace atmospheric gases by marine aerosols, but the specific impacts of cation type on LPS-transformed SSA particle morphology and reactivity remain unclear.

Here, we sought to understand the specific impact of monovalent and divalent cations on heterogeneous reactions between LPS-transformed SSA particles and HNO_3. Both simplified salt-LPS model aerosols as well as nascent aerosols generated from a laboratory induced phytoplankton bloom were investigated. ATOFMS was used to reveal how cation valency drives LPS physicochemical changes in the aerosol phase, which in turn impacts the aerosol heterogeneous reactivity toward HNO_3.[19] Furthermore, explicit solvent all-atom molecular dynamics simulations of LPS bilayers under different cation conditions were performed to provide a molecular-level explanation of the reactivity behavior.[4] Taken together, these findings can provide substantial insight into the molecular mechanisms behind cation-driven LPS morphology, phase, and correlated reactivity.[20,21]

4.3 Results and Discussion

Chemical changes in LPS-containing SSA particles were investigated during the Investigation into Marine Particle Chemistry and Transfer Science (IMPACTS) campaign, where a phytoplankton bloom was induced using natural Pacific Ocean seawater.[7,8] The induced bloom in the ocean-atmosphere interaction facility (OAIF) produced SSA particles that are representative of those produced over the ocean at varying biological activities.[7,22] While two phytoplankton blooms were induced from the same seawater during the IMPACTS campaign (**Figure 4.1A**, green trace),[7] we focus on the second phytoplankton bloom in the study due to the concurrent bloom of heterotrophic bacteria (**Figure 4.1A**, orange marker) in the seawater as well as high enrichment of polysaccharides in the supermicron SSA population (**Figure 4.1A**, blue marker) as identified using Raman spectroscopic analysis by Cochran et al.[8] As each day contains data from tens to hundreds of thousands of particles, we further zoom into a single day during the 2nd bloom (**Figure 4.1A**, red box) where 63.9% of the Raman-active supermicron SSA particles contain polysaccharides.

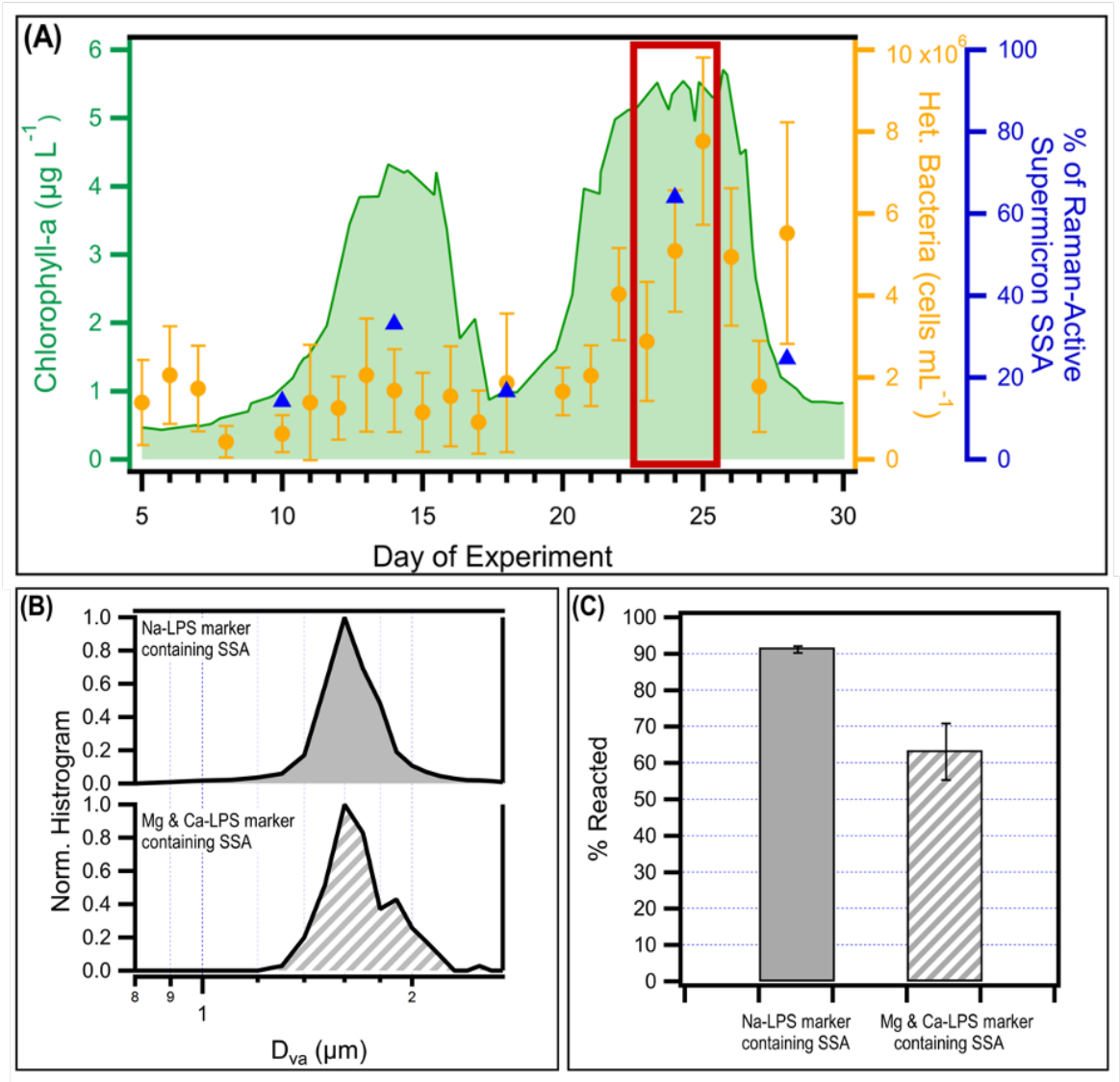


Figure 4.1: IMPACTS campaign data, including phytoplankton bloom time series and ATOFMS data corresponding to LPS-containing particles. (A) Seawater chlorophyll-a (green) and heterotrophic bacteria concentrations (orange) with the percentage of SSA containing polysaccharides, including LPS[6] (blue), over the course of the phytoplankton blooms from the IMPACTS campaign. Period of supermicron SSA polysaccharide-enrichment for further ATOFMS data analysis is noted by red box. (B) ATOFMS size histogram of Na-LPS (solid fill) and Mg/Ca-LPS SSA (dashed fill) from the IMPACTS campaign. (C) Percentage of reacted Na-LPS (solid fill) and Mg/Ca-LPS SSA (dashed fill) detected by ATOFMS with HNO_3 reaction flow tube from the IMPACTS 2014 campaign. Error bars in (C) represent 2σ for 95% confidence limit.

Using known LPS ion markers from reference ATOFMS mass spectra of LPS (**Figure 4.5**), we isolated 5,011 LPS-containing SSA particles out of a total of 52,918 particles (10.6 ± 0.3%) for one day of the study. Two types of particles were observed: 1) LPS-containing SSA with a dominant sodium signal (herein referred as Na-LPS SSA), and 2) LPS-containing SSA with dominant Mg and Ca signals (> 10% Rel. Area Intensity of

$^{24}Mg^{+}$ or $^{40}Ca^{+}$, herein referred as Mg/Ca-LPS SSA) where these particles made up of 98.0 ± 0.4% and 42.0 ± 2.2% of the total LPS-containing SSA particles, respectively. In these SSA particles, Mg/Ca-LPS SSA particles exhibited a more pronounced second peak in the larger aerodynamic sizes at 1.9 μm D_{va} (vacuum aerodynamic diameter, herein referred to as aerodynamic diameter) compared to Na-LPS SSA particles (**Figure 4.1B**). This can be attributed to either increase in particle size or change in shape as ATOFMS measures the particles' aerodynamic diameter and a change in shape leads to an apparent change in aerodynamic diameter.[23] Ultimately, when these particles were reacted with 350 ppb HNO_3 at 60 ± 2% relative humidity (RH), a lower percentage of Mg/Ca-LPS SSA particles reacted (63.6 ± 8.1%) than Na-LPS SSA particles (91.7 ± 0.8%) (**Figure 4.1C**). Given the level of chemical complexity of nascent SSA particles, we hypothesize that this shift in size and the associated differences in heterogeneous reactivity are due to aggregation of the LPS molecules with divalent cations that are enriched at the air-water interface of SSML and ejected as SSA particles.[24–27] We sought to investigate the role of cation valency using a more controlled model system consisting of specific salts and LPS using a synergistic experimental and computational approach.

Model systems of salt-LPS particles were produced by atomizing LPS solutions containing different counter cations: Na^{+}, Mg^{2+}, and Ca^{2+}. The ATOFMS, which provides size-resolved, dual polarity mass spectra of single particles,[19] can detect single soluble and insoluble particles, as well as agglomerates of different particles across a wide size range (**Figure 4.6**).[28] As done in the analysis of the IMPACTS campaign, select ion markers of LPS, obtained from a reference mass spectrum (**Figure 4.5**), were used to isolate particles containing LPS from the total population of atomized LPS particles. In the model systems, the generated aerosols are called "LPS particles" to easily distinguish this system with the complex system of the IMPACTS campaign.

In this study, we determined that morphology and particle size of LPS particles depends upon the counter cation present. Results from ATOFMS on these model LPS particles revealed a more pronounced shift from predominantly submicron to supermicron sizes depending upon the counter cation present (**Figure 4.2**). This is consistent with observations made during IMPACTS. Na-LPS particles were predominantly in the submicron size range, whereas Mg- and Ca-LPS particles were in the supermicron range. In order the visualize the physical changes, we took transmission electron microscopy (TEM, FEI Tecnai Spirit) images of the negative-stained[29,30] atomized LPS particles and revealed that Na-LPS particle exhibited long tubular structures

of single-layered structure, whereas Mg- and Ca-LPS particles exhibited bilayer sheets as previously observed in a study that characterized the LPS salts in bulk solution using TEM.[14]

It is unclear how this change in the morphology impacts the heterogeneous reactivity of SSA particles in the atmosphere. To investigate this, we reacted these model LPS particles in an aerosol flow tube using 10 ppb HNO_3 at 50 ± 2 % RH, a lower and more atmospherically relevant condition.[31] These simple model LPS particles showed very similar behavior to the complex SSA in **Figure 4.1C**. Na-LPS particles were more ~8 times more reactive with HNO_3 (24.1 ± 1.3%) compared to Mg- (2.5 ± 0.8%) or Ca-LPS (3.7 ± 1.2%) particles. These results all converge to support our original hypothesis that changes in LPS-cation structures formed with different counter cations present can change the atmospheric reactivity of SSA particles.

LPS molecules tend to form bilayer aggregates in the presence of different divalent counter cations as observed here and in previous studies (**Figure 4.2**).[14,32] To understand how these higher order LPS structures could alter heterogeneous pathways at the atomic level, we performed three all-atom explicit solvent molecular dynamics simulations of an E. Coli O111 LPS bilayer, the same LPS structure used in these model experiments. CHARMM-GUI was used to construct the model systems with coordinating cations.[33–35] Each LPS molecule contained eight O-antigen repeating units, an R1 core, and Lipid A, and was conjugated to Mg^{2+}, Ca^{2+} or Na^{+} counter cations to balance the charges.

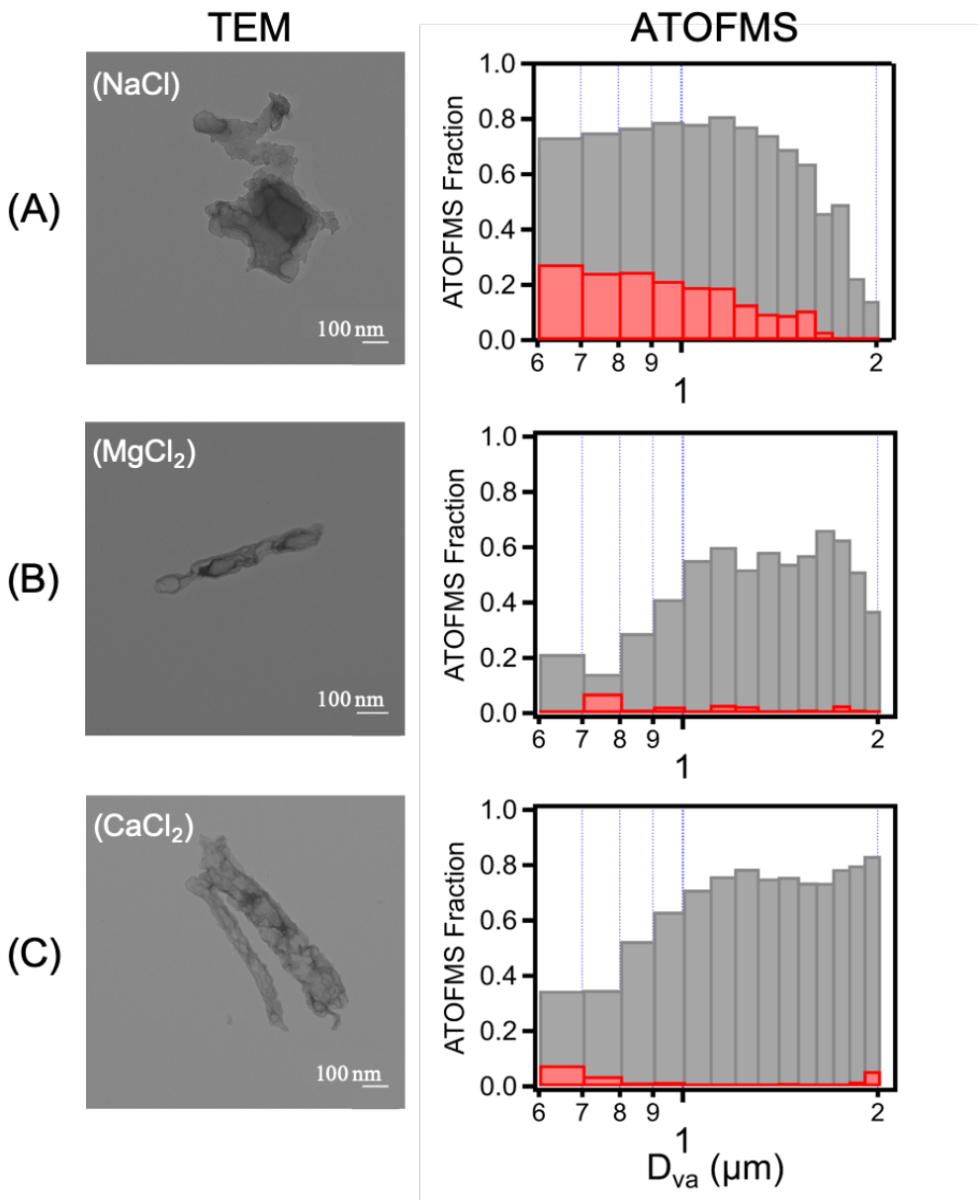


Figure 4.2: (LEFT) Transmission Electron Microscope images of negative-stained atomized particles, and (RIGHT) normalized size distribution of LPS particles detected by ATOFMS for (A) LPS spiked with 0.010 M NaCl, (B) LPS spiked with 0.010 M $MgCl_2$, (C) LPS spiked with 0.010 M $CaCl_2$. Scale bar on the image denotes 100 nm and the ATOFMS total LPS particles are shown in gray, whereas the reacted LPS particles are overlaid in red.

To elucidate the impact of the counter cation on LPS bilayer structure, we compared the equilibrium area per molecule and LPS chain length and orientations. The simulations show that, under constant pressure, the LPS bilayer in Ca^{2+} and Mg^{2+} matrices expanded laterally in area from their initial post-equilibration structures, while LPS in Na^{+} matrix contracted laterally. **Figure 4.3A** shows how this expansion is reflected in

the area per molecule of the LPS bilayers over the duration of each molecular dynamics trajectory. On average, the final area per molecule of Na-LPS molecules (1.91 $\pm$ 0.01 nm^2) was markedly lower than both Ca- and Mg-LPS (1.98 $\pm$ 0.02 and 2.01 $\pm$ 0.02 nm^2, respectively). Additionally, in both divalent cation matrices, the LPS chains underwent vertical compression compared to their Na-complexed LPS counterparts. In direct comparisons between monosaccharide units in Ca- or Mg-LPS chains with Na-LPS chains, both calcium and magnesium cations induced "crunching" in the oligosaccharides resulting in chain lengths consistently shorter than those of Na-LPS (**Figure 4.3B**). This observation is in line with previous work, which has documented the calcium-induced "collapse" of monolayers containing LPS isolated from *Pseudomona aeruginosa*.[36]

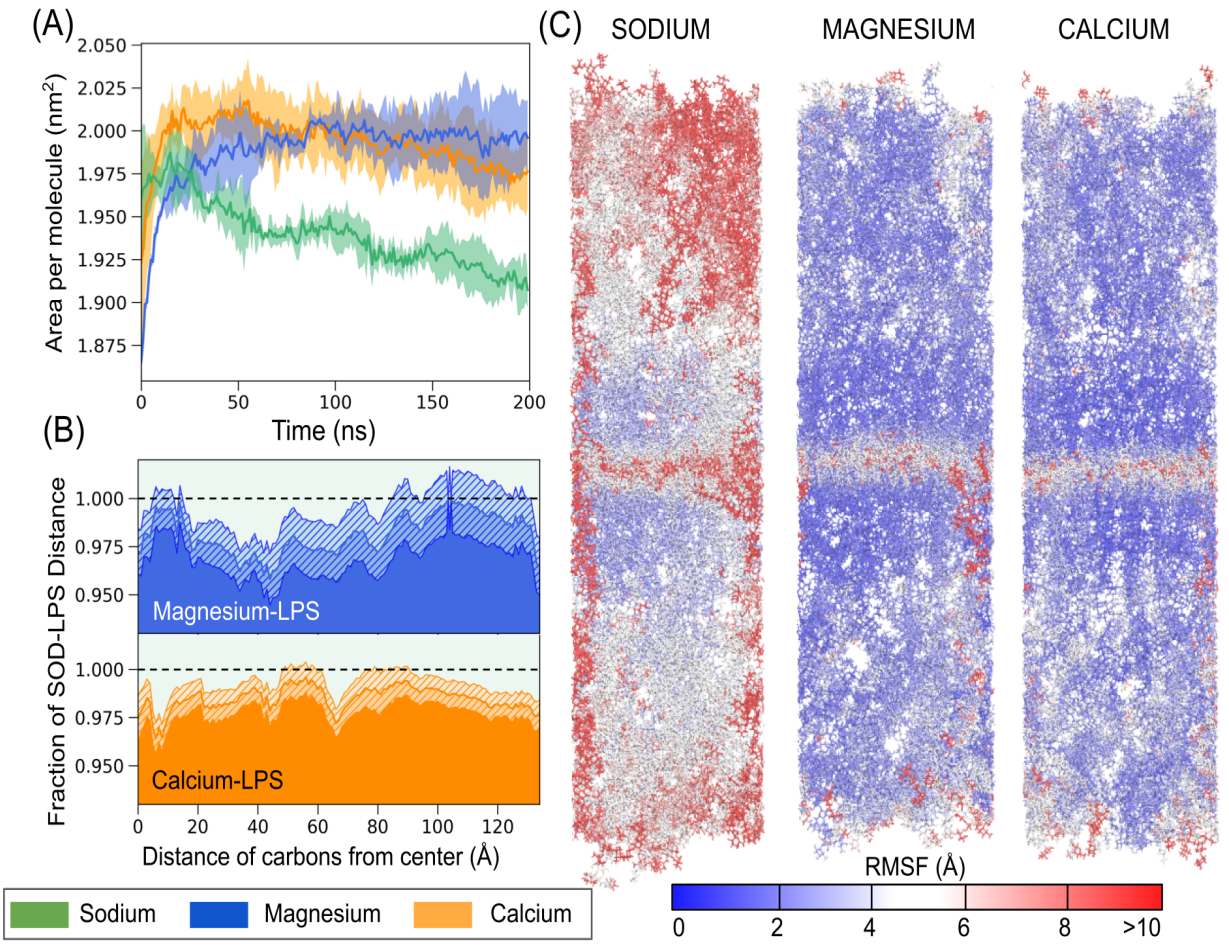


Figure 4.3: Characterization of morphology and dynamics from MD simulations. (A) Area per molecule (APM) measurements in nm^2 LPS-molecule^{-1} over each 200-ns trajectory. Replicates are indicated by average area per molecule over time with standard deviation shaded above and below, with Ca-LPS in orange, Na-LPS in green, and Mg-LPS in blue. (B) LPS chain length fractions of Mg-LPS (top) and Ca-LPS (bottom) as fractions of Na-LPS chain length. Error propagation bars are given by hatches. Dotted lines indicate unity. (C) Atomic-level root mean square fluctuations (RMSF) were calculated over the final 50 ns of simulation for replicates of Na-LPS (left), Mg-LPS (center), and Ca-LPS (right) bilayers. Atoms are colored by RMSF value, with red indicating more fluctuations and blue indicating less fluctuations. Note that red atoms located at the edges of the simulation box may be due to artifacts of molecules moving across periodic boundaries and not corrected by the gmx rmsf module.

To understand the impacts of cation valency on the LPS aggregate dynamics, we compared the atomic-level root mean square fluctuations (RMSF). RMSF calculations indicate that the O-antigen chains on Na-LPS exhibited a much higher degree of flexibility than those complexed with magnesium or calcium, which adopted a more rigid structure (**Figure 4.3C**). These calculations are in line with experimentally-measured

compressibilities for salt-LPS monolayers, where Na-LPS monolayers demonstrate high compressibility compared to Ca-LPS.[37] However, the impacts of cation type on the core region flexibility of the LPS molecules is much less pronounced, indicating that specific cation interactions within the saccharide chains are more impactful to the overall dynamics of LPS bilayers than interactions with the R1 core or Lipid A.

To understand how the morphological changes in the LPS bilayer impact the reactivity of the LPS aggregate, we calculated water diffusion coefficients and water density for each system. The diffusivity of water throughout the LPS when complexed to sodium was significantly higher than when complexed to magnesium or calcium (**Figure 4.4A**). Additionally, measurements of water density throughout the bilayers indicate that water molecules in the Na LPS bilayers were able to occupy space between the O-antigen regions of the LPS molecules. Conversely, the water in Ca/Mg LPS bilayers is mainly concentrated in the core region of the LPS (**Figure 4.4B**). The dehydration of the saccharide regions as indicated by low water occupation and diffusion constants in Ca/Mg LPS bilayers likely accounts for the reduced HNO_3 reactivity of these LPS particle types, as the morphology of these Ca/Mg LPS particles prevents diffusion of dissolved HNO_3 into the LPS core where it can undergo protonation reactions at the phosphate and carboxylate sites.

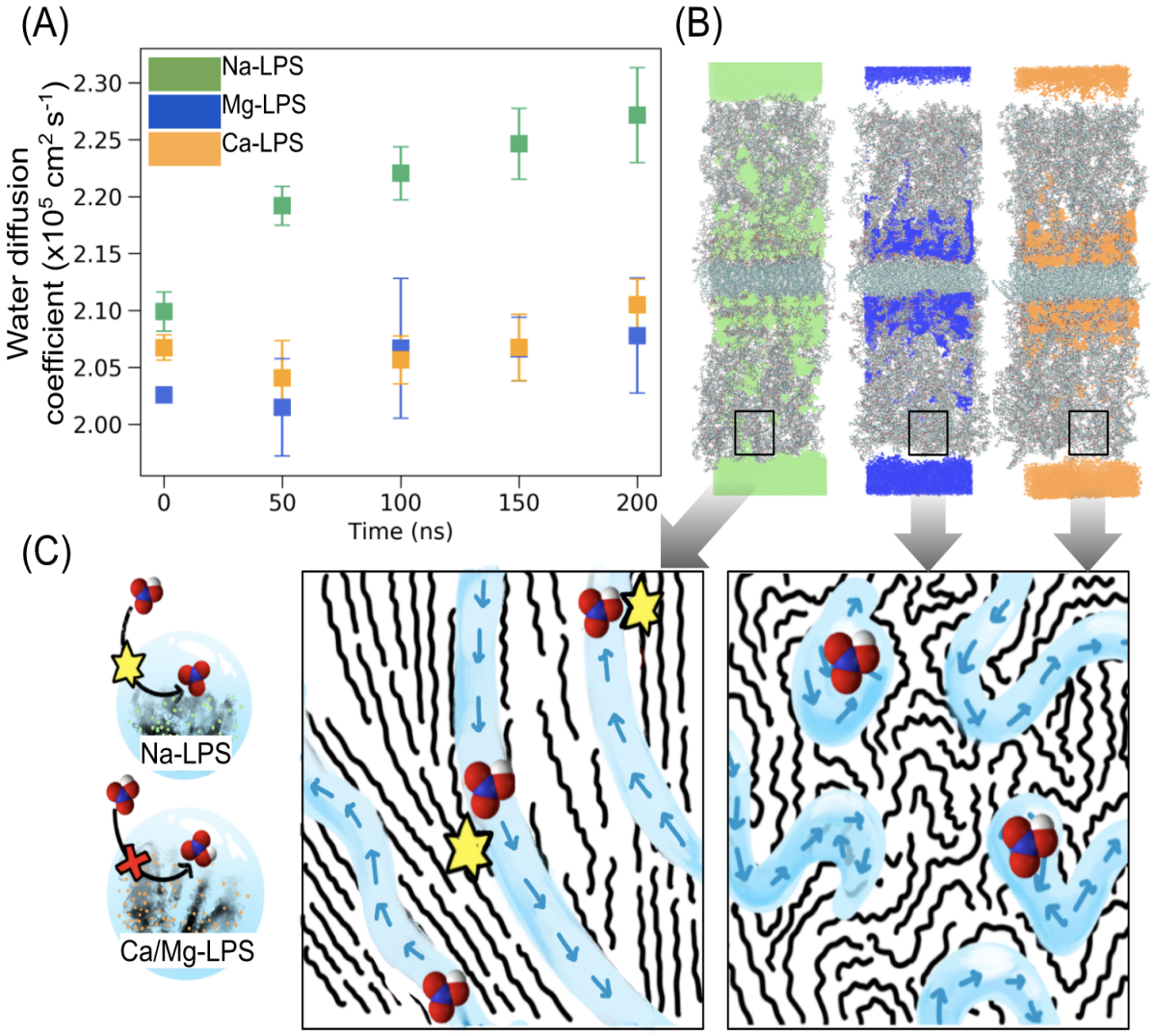


Figure 4.4: Water diffusion and density calculations from MD simulations. (A) Water diffusion coefficients of LPS bilayers complexed to either sodium (green), magnesium (blue) or calcium (orange). (B) Water density measurements for replicate 1 of each counter-cation averaged over the entire 200-ns trajectory. Water is visualized by color, where Na-LPS is given by green, Mg-LPS is given by blue, and Ca-LPS is given by orange volume maps, overlaid by corresponding LPS molecules. (C) Schematic representing how water diffusion impacts HNO_3 reactivity based on structural changes in the LPS molecules.

One possible explanation for the cation-dependent differences in morphology and reactivity of LPS bilayers is that divalent cations, by chelating two singly-charged reactive sites each, deform the structures of the oligosaccharides while positioning them into ideal cross-linking conformations, which simultaneously dehydrates the O-antigen regions and blocks the movement of water.[13,15,16,36,37] Monovalent sodium cations do not appear to significantly disrupt the alignment of the LPS molecules; however, divalent cations induce

structural deformation away from the bilayer normal in the oligosaccharide chains, which increases the molecular footprint of each LPS molecule and causes the bilayer to expand. Notably, the increase in area per molecule does not correspond to an increase in water diffusion throughout aggregate; rather, the LPS conformations hinder water diffusion by forming a relatively dehydrated and intertwined network of saccharide chains, also known as microgels, which have been observed in marine environments.[38] Although the Na-LPS bilayer adopts a lower area per molecule, the lower specificity of the saccharide chains to sodium and the alignment of the molecules allows for greater molecular flexibility. This more flexible structure promotes the diffusion of water and small molecules such as nitric acid, which increases the overall heterogeneous reactivity of Na-LPS aggregates compared to the divalent analogs.

4.4 Conclusion

SSA particles have been shown to exhibit a wide range of heterogeneous reactivities at the single particle level.[1,39] Results from this study provide a detailed atomic-level examination of the factors that helps explain the observed variation in reactivity in LPS-containing SSA particles. Specifically, cation type elicits differences in physicochemical properties of LPS-containing marine aerosols, highlighting the role that counter cations play on aerosol structure and atmospheric reactivity. This discovery not only has important implications for other atmospheric heterogeneous reactions including oxidation reactions, but also in key surface processes such as water uptake, cloud droplet and ice nucleation.[40–44] This study demonstrates how the addition of biogenically-derived LPS molecules can impact SSA physicochemical properties and reactivity, warranting further investigations into how these morphological changes impact relevant properties including cloud and ice forming abilities.

4.5 Acknowledgments

This work is funded by NSF Center for Aerosol Impacts on Chemistry of the Environment (CAICE), National Science Foundation Center for Chemical Innovation (NSF CHE-1801971). Computing support through XSEDE was provided by NSF CHE-060073N to REA. The authors would like to thank all collaborators involved with the NSF CAICE IMPACTS intensive campaign and the Scripps Institution of Oceanography Hydraulic Laboratory staff. The authors thank O. Ryder, M. Santander, and J. Dowling for their valuable contributions and

G. Castillon, M. Ellisman, and the National Center of Microscopy and Imaging Research for the TEM component of this study.

Chapter 4, in full, is a modified reprint of the material as it appears in "Lee, C.† and Dommer, A. C.†; Schiffer, J. M.; Amaro, R. E.; Grassian, V. H.; Prather, K. A. *Cation-driven lipopolysaccharide morphological changes impact nitric acid heterogeneous reactivity in sea spray aerosols*. J. Phys. Chem. Lett. 2021, 12 (20), 5023-5029." The dissertation author was a co-investigator and co-author of this work.

4.6 Supporting Information

Further information on the experiment methods and setup, size distribution of atomized LPS particles, and ATOFMS data showing the averaged mass spectra of LPS particles and molecular dynamics simulations are given in the Supporting Information. This material is available free of charge via the Internet at http://pubs.acs.org. All data and IPython notebooks for this work is available as part of the UCSD Library Digital Collections (https://doi.org/10.6075/J0862F04 & https://library.ucsd.edu/dc/collection/bb96275693).

4.6.1 Lipopolysaccharide (LPS) Aggregation and Data Analysis

Lipopolysaccharide (LPS) powder (Sigma-Aldrich, L4130, lot #: 064M4125V, 085M4107V), $NaCl_{(s)}$ (Acros Organics, 42429-5000, lot #: B014114413), $MgCl_2 \cdot 6H_2O$ (Sigma Aldrich, M33-500), and $CaCl_2 \cdot 2H_2O$ (Alfa Aesar, 33296, lot #: U13B016), were purchased and used without further purification. Method of LPS aggregation was adapted from previous studies.[45–48] 0.1 g of LPS was dissolved in 200 mL of ultrapure water (Millipore) to be analyzed as a reference standard. Three additional solutions of same LPS concentration were prepared with NaCl added to the 1st solution (0.1169g), $MgCl_2 \cdot 6H_2O$ added to the 2nd solution (0.4066g), and $CaCl_2 \cdot 2H_2O$ added to the 3rd solution (0.2904 g), all ending with salt concentration of 0.01 M. All solutions with LPS were sonicated for minimum of 30 min to ensure complete dissolution of reagents. Aliquots (100 mL) were taken out of each solution immediately after preparation to be atomized for initial analysis. Rest of the solutions were incubated for 72 hours at room temperature (24 ± 1 °C), constantly stirred using an orbital shaker. The solutions were then atomized for analysis.

Collison atomizer[49] was used to atomize the solutions (1.5 SLPM). Atomized particles were dried using two diffusion driers (RH < 10%) and sent to aerosol time-of-flight mass spectrometer (ATOFMS). ATOFMS

provided single particle size and chemical composition between 0.3-3.0 μm D_{va} (vacuum aerodynamic diameter).[50] Further details on the instrument can be found elsewhere.[50] In brief, particles are drawn into the ATOFMS nozzle inlet and are accelerated through two stages of differential pumping. The particles then pass through two orthogonally placed continuous wave lasers (Coherent, OBIS 1236438, 405 nm/50 mW and OBIS 1236443 488 nm/ 30 mW, respectively). Particle velocity is calculated based on the particle transit time between the two lasers. External calibration curve generated using polystyrene latex spheres of known diameter and density is used to convert the velocity to D_{va}. Particle velocity also triggers pulsed Q-switched Nd:YAG laser (Quantel, 266 nm, 8 ns width, 700 μm spot size, 3×10^7 W cm^{-2}) that desorbs and ionizes each particle. Positive and negative ions formed from each particles are detected simultaneously using a dual-polarity reflectron time-of-flight mass spectrometer, equipped with microchannel plate detectors (MCP, Photonis, 931377). Acquired data were then imported into MATLAB (The Math-Works, Inc.) with software toolkit FATES[51] for further data analysis. Equations (4.1) and (4.2) were used to calculate the uncertainty in the reported fractions of particles, where F is fraction, x is number of particles containing select ion marker, N is total number of particles, and SE stands for standard error of 1σ. Error reported in this study is 2σ for 95% confidence limit.

$$F = \frac{x}{N} \tag{4.1}$$

$$SE = \sqrt{\frac{F(1-F)}{N}} \tag{4.2}$$

After sampling with ATOFMS, the particle flow was sent to aerosol particle sizer (APS, TSI Model 3321, 1 SLPM sample flow), and scanning mobility particle sizer (SMPS Model 3936 and 3938, operating at 0.3/3.0 SLPM sample/sheathe flow). APS and SMPS measures particles with aerodynamic diameter (D_a) between 0.6 and 20 μm, and mobility diameters (D_m) between 0.013 and 0.7 μm, respectively. Averaged number size distributions from the two instruments were merged by converting both Da and D_m to the physical diameter (D_p),[52] assuming all particles had a density of 1.8 g cm^{-3} and were spherical.[53] It is important to note that as the sampled particles were dried, the spherical particle assumption may not be accurate at all times.

Aerosol reaction flow tube (0.048 m i.d. x 1.5 m length) was used to probe the reactivity of the particles. HNO_3 permeation tube (Kin-Tek, HRT-010.00-2022/60) in a temperature-controlled chamber (30 ºC) was

utilized as a constant source of reactant gas and all reaction studies were performed after 30 min of system equilibration and atomizing NaCl particles through the flow tube to confirm the reaction as a daily standard. The concentration of the HNO_3 for the LPS aggregate reactivity study was approximately 10 ppb, with residence time and relative humidity (RH) of the flow tube of 1.8 min and 50 ± 2 %, respectively. The exit flow of the flow tube was split between a RH and temperature probe (Vaisala, HMP110) and ATOFMS which sampled the particles after passing through two diffusion driers.

Atomized LPS particles were also collected for transmission electron microscopy (TEM) imaging using Aerosol Devices Inc. Liquid Spot Sampler for 60 min at 1 SLPM into 0.75 mL of liquid volume. Bulk sample prior to atomization were also taken. Bulk (5 µL) and spot sampler collected liquid (10 µL) were then applied to the TEM grids (Fisher Scientific, 50-930-847) and were prepared and stained using the methods outlined in Booth et al.[54] and Passmore and Russo[55] except for using uranyl acetate and not uranyl formate. These samples were then imaged using FEI Tecnai Spirit TEM with the settings of 120 kV, 1.5 s exposure time, and spot size of 3. Overview images were taken with ~20,000x magnification, where the detailed images reported in the manuscript are taken with ~50,000x magnification. Images and its metadata were acquired using SerialEM (https://bio3d.colorado.edu/SerialEM/, v.3.1.1a) and were analyzed using ImageJ (https://imagej.nih.gov/ij/).

4.6.2 Lipopolysaccharide (LPS) System Preparation for Molecular Dynamics Simulations

CHARMMGUI[56] was used to build the LPS bilayer starting structure. The *E.Coli* O111 antigen and R1 core structure were chosen, with 8 repeating units of the O111 antigen. The starting box structure was determined as 70 Å for the plane of the bilayer, with a 30 Å water buffer on both sides. Each molecule of LPS was placed in the bilayer with a surface area of 190 $Å^2$ per LPS, for a total of 52 LPS molecules. The phosphates and carboxyl groups of the O111 repeats and Core domain were balanced with either Ca^{2+}, Mg^{2+}, or Na^+. Each phosphate maintained a net charge of -1 in the simulation to replicate a ~ pH of less than 6 environment in the context of a sea spray aerosol. Approximately 77,000 TIP3P waters[57–59] were placed in each system in the water buffer region as well as between the O antigens. The CHARMM36 force field[60] was used to parameterize the molecules and the GROMACS 2018 MD engine[61,62] was used to perform simulations.

4.6.3 Lipopolysaccharide (LPS) All-Atom Molecular Dynamics Simulations

Simulations were minimized and equilibrated using a NVIDIA GK110 GPU (GeForce GTX Titan, NVIDIA, Santa Clara, CA). All production runs utilized the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562.[63] Specifically, simulations were performed on the Bridges supercomputer, which is supported by NSF award number ACI-1445606 at the Pittsburgh Supercomputing Center (PSC).[64] The structure went through one step of minimization using the steepest descent gradient[65] algorithm and six steps of equilibration were performed. Each step of equilibration was performed for 250 ps of simulation, for a total of 1.5 ns of equilibration. GPU-enabled GROMACS 2018.3 production runs were carried out as an NTP ensemble at 298.15 K, with a 2-fs time step, and particle mesh Ewald[66] electrostatic approximation with a nonbonding cutoff of 12 Å. Langevin dynamics[67] with a collision frequency of 1.0 ps^{-1}, a pressure relaxation time of 2.0 ps, and with SHAKE bond constraints on hydrogen bonds.[68]

4.6.4 Lipopolysaccharide Simulation Analyses

Analyses were performed using built-in GROMACS modules and MDTraj[69] in the IPython Jupyter Notebook environment.[70] All IPython notebooks for this work will be made available as part of the UCSD Library Digital Collections: Center for Aerosol Impacts on Chemistry of the Environment (CAICE) (https://library.ucsd.edu/dc/collection/bb96275693). Images were taken with VMD.[71]

4.6.5 Investigation into Marine PArticle Chemistry and Transfer Science (IMPACTS) Campaign Experiment Methods

Detailed experimental methods, including the phytoplankton bloom methods, on-line and off-line instrument sampling, and other sampling protocols can be found elsewhere.[72–75] Simplified schematic drawing of the ocean-atmosphere interaction facility (OAIF), detailing the experiment setup involving the ATOFMS and the aerosol flow tube, is depicted in **Figure 4.7**. Filtered air using HEPA, activated carbon, and potassium permanganate filters were supplied to the headspace of the wave flume, with air flow speed of approximately 5 cm s^{-1}. Hydraulic paddle generated waves that broke off the angled beach, produced realistic sea spray aerosols (SSA) over the course of the induced phytoplankton bloom experiment. Port placed downwind of the breaking

waves sampled the aerosols to the connected instruments in the study through a laminar flow manifold. Flow tube parameters for the IMPACTS campaign were: HNO_3 concentration of approximately 350 ppb, with residence time and relative humidity (RH) of the flow tube of 2 min and 60 ± 2 %, respectively.

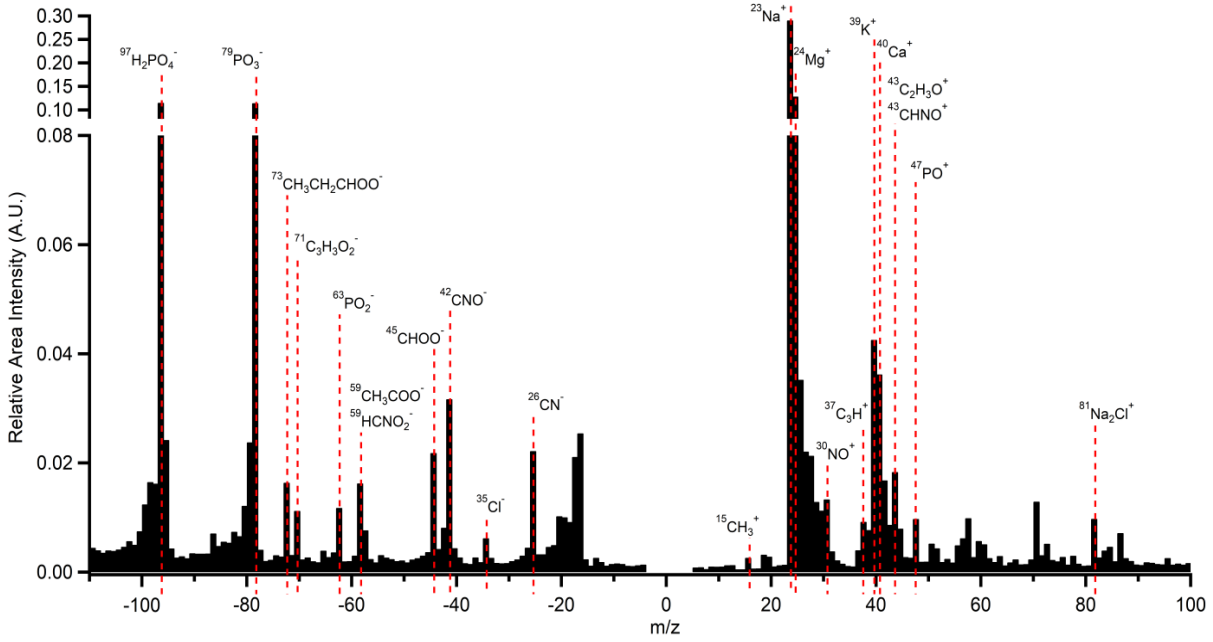


Figure 4.5: Averaged mass spectra of atomized LPS particles, with select ion markers used for identification of LPS containing particles in atomized samples of LPS salt solutions.

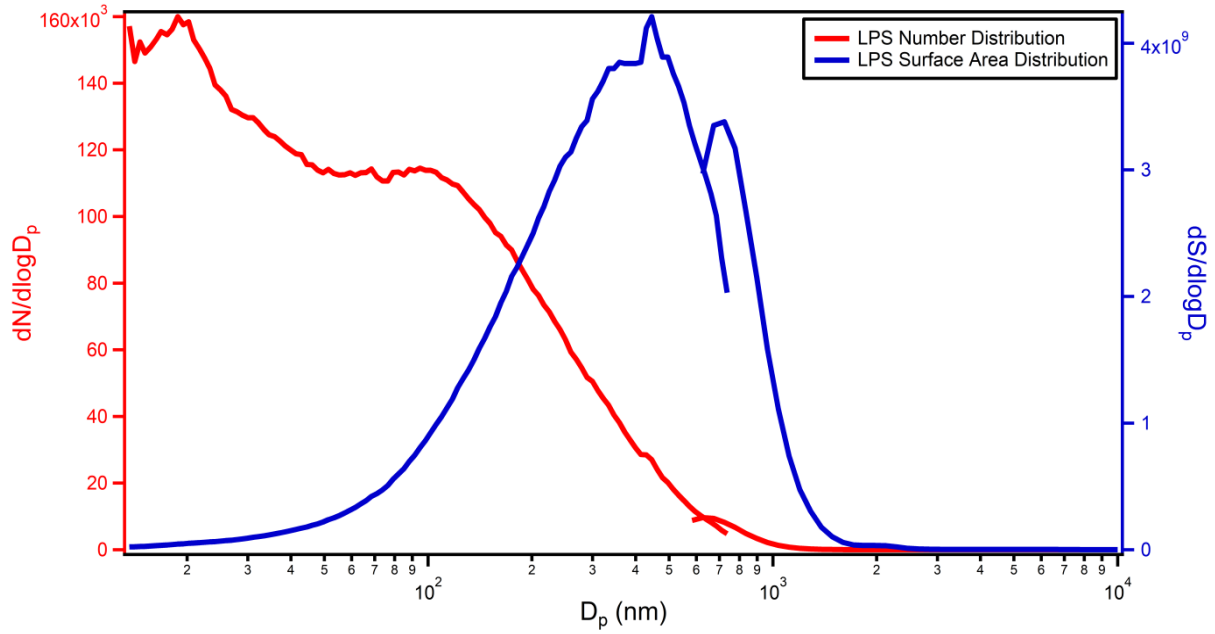


Figure 4.6: Merged APS and SMPS Number (red) and calculated surface area (blue) size distributions of atomized LPS particles.

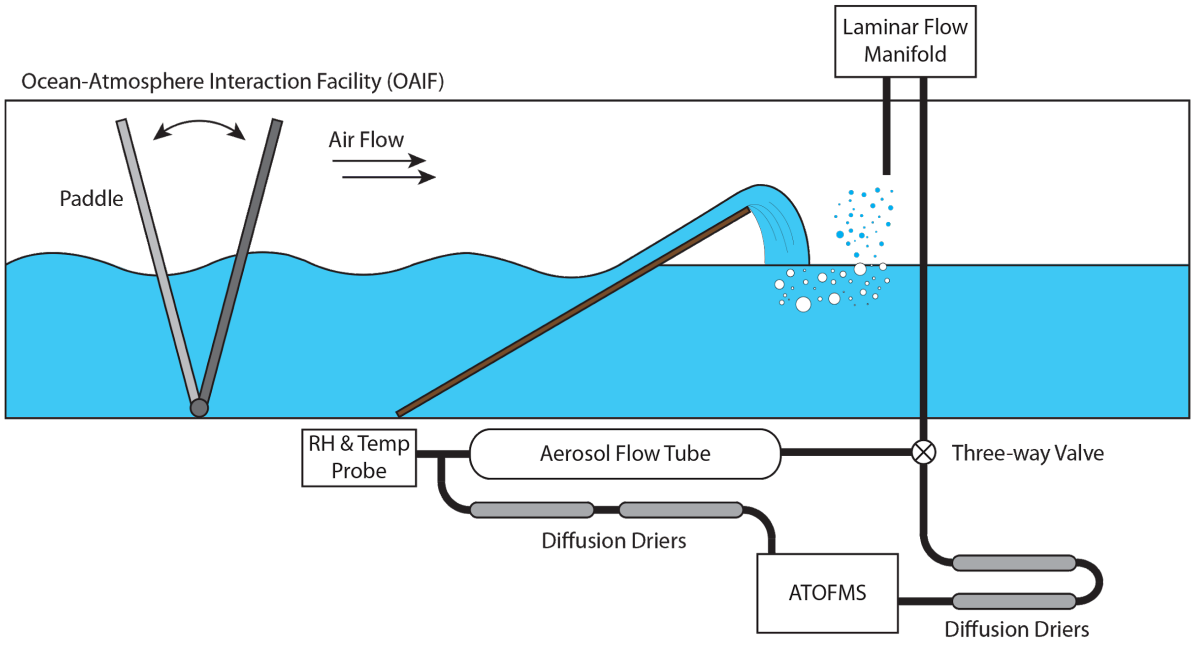


Figure 4.7: Schematic drawing of the simplified IMPACTS campaign experiment showing the OAIF and the ATOFMS with aerosol flow tube. Angled beach for breaking waves is colored brown.

References

(1) Abbatt, J. P. D.; Lee, A. K. Y.; Thornton, J. A. Quantifying Trace Gas Uptake to Tropospheric Aerosol: Recent Advances and Remaining Challenges. *Chemical Society Reviews* **2012**, *41* (19), 6555. https://doi.org/10.1039/c2cs35052a.

(2) Bertram, T. H.; Cochran, R. E.; Grassian, V. H.; Stone, E. A. Sea Spray Aerosol Chemical Composition: Elemental and Molecular Mimics for Laboratory Studies of Heterogeneous and Multiphase Reactions. *Chemical Society Reviews* **2018**, *47* (7), 2374–2400. https://doi.org/10.1039/C7CS00008A.

(3) Estillore, A. D.; Trueblood, J. V.; Grassian, V. H. Atmospheric Chemistry of Bioaerosols: Heterogeneous and Multiphase Reactions with Atmospheric Oxidants and Other Trace Gases. *Chem. Sci.* **2016**, *7* (11), 6604–6616. https://doi.org/10.1039/C6SC02353C.

(4) Schiffer, J. M.; Mael, L. E.; Prather, K. A.; Amaro, R. E.; Grassian, V. H. Sea Spray Aerosol: Where Marine Biology Meets Atmospheric Chemistry. **2019**, *19*, 46. https://doi.org/10.1021/acscentsci.8b00674.

(5) Lewis, E. R.; Schwartz, S. E. Sea Salt Aerosol Production: Mechanisms, Methods, Measurements and Models. *Geophysical Monograph Series* **2004**, *152*, 413. https://doi.org/10.1029/GM152.

(6) Patterson, J. P.; Collins, D. B.; Michaud, J. M.; Axson, J. L.; Sultana, C. M.; Moser, T.; Dommer, A. C.; Conner, J.; Grassian, V. H.; Stokes, M. D.; Deane, G. B.; Evans, J. E.; Burkart, M. D.; Prather, K. A.; Gianneschi, N. C. Sea Spray Aerosol Structure and Composition Using Cryogenic Transmission

Electron Microscopy. *ACS Central Science* **2016**, *2* (1), 40–47. https://doi.org/10.1021/acscentsci.5b00344.

(7) Wang, X.; Sultana, C. M.; Trueblood, J.; Hill, T. C. J.; Malfatti, F.; Lee, C.; Laskina, O.; Moore, K. A.; Beall, C. M.; McCluskey, C. S.; Cornwell, G. C.; Zhou, Y.; Cox, J. L.; Pendergraft, M. A.; Santander, M. V.; Bertram, T. H.; Cappa, C. D.; Azam, F.; DeMott, P. J.; Grassian, V. H.; Prather, K. A. Microbial Control of Sea Spray Aerosol Composition: A Tale of Two Blooms. *ACS Central Science* **2015**, *1* (3), 124–131. https://doi.org/10.1021/acscentsci.5b00148.

(8) Cochran, R. E.; Laskina, O.; Trueblood, J. V.; Estillore, A. D.; Morris, H. S.; Jayarathne, T.; Sultana, C. M.; Lee, C.; Lin, P.; Laskin, J.; Laskin, A.; Dowling, J. A.; Qin, Z.; Cappa, C. D.; Bertram, T. H.; Tivanski, A. V; Stone, E. A.; Prather, K. A.; Grassian, V. H. Molecular Diversity of Sea Spray Aerosol Particles: Impact of Ocean Biology on Particle Composition and Hygroscopicity. *Chem* **2017**, *2* (5), 655–667. https://doi.org/10.1016/j.chempr.2017.03.007.

(9) Pryor, S. C.; Sørensen, L. L. Nitric Acid–Sea Salt Reactions: Implications for Nitrogen Deposition to Water Surfaces. *Journal of Applied Meteorology* **2000**, *39* (5), 725–731. https://doi.org/10.1175/1520-0450-39.5.725.

(10) Trueblood, J. V.; Estillore, A. D.; Lee, C.; Dowling, J. A.; Prather, K. A.; Grassian, V. H. Heterogeneous Chemistry of Lipopolysaccharides with Gas-Phase Nitric Acid: Reactive Sites and Reaction Pathways. *Journal of Physical Chemistry A* **2016**, *120* (32), 6444–6450. https://doi.org/10.1021/acs.jpca.6b07023.

(11) Aurell, C. A.; Wistrom, A. O. Critical Aggregation Concentrations of Gram-Negative Bacterial Lipopolysaccharides (LPS). *Biochemical and biophysical research communications* **1998**, *253* (1), 119–123. https://doi.org/10.1006/bbrc.1998.9773.

(12) Wu, E. L.; Engström, O.; Jo, S.; Stuhlsatz, D.; Yeom, M. S.; Klauda, J. B.; Widmalm, G.; Im, W. Molecular Dynamics and NMR Spectroscopy Studies of E. Coli Lipopolysaccharide Structure and Dynamics. *Biophysical Journal* **2013**, *105* (6), 1444–1455. https://doi.org/10.1016/j.bpj.2013.08.002.

(13) Thoma, J.; Abuillan, W.; Furikado, I.; Habe, T.; Yamamoto, A.; Gierlich, S.; Kaufmann, S.; Brandenburg, K.; Gutsmann, T.; Konovalov, O.; Inoue, S.; Tanaka, M. Specific Localisation of Ions in Bacterial Membranes Unravels Physical Mechanism of Effective Bacteria Killing by Sanitiser. *Scientific Reports* **2020**, *10* (1), 1–12. https://doi.org/10.1038/s41598-020-69064-1.

(14) Coughlin, R. T.; Haug, A.; McGroarty, E. J. Physical Properties of Defined Lipopolysaccharide Salts. *Biochemistry* **1983**, *22* (1978), 2007–2013.

(15) Garidel, P.; Rappolt, M.; Schromm, A. B.; Howe, J.; Lohner, K.; Andrä, J.; Koch, M. H. J.; Brandenburg, K. Divalent Cations Affect Chain Mobility and Aggregate Structure of Lipopolysaccharide from Salmonella Minnesota Reflected in a Decrease of Its Biological Activity. *Biochimica et Biophysica Acta - Biomembranes* **2005**, *1715* (2), 122–131. https://doi.org/10.1016/j.bbamem.2005.07.013.

(16) Schneck, E.; Schubert, T.; Konovalov, O. V.; Quinn, B. E.; Gutsmann, T.; Brandenburg, K.; Oliveira, R. G.; Pink, D. A.; Tanaka, M. Quantitative Determination of Ion Distributions in Bacterial Lipopolysaccharide Membranes by Grazing-Incidence X-Ray Fluorescence. *Proceedings of the National Academy of Sciences of the United States of America* **2010**, *107* (20), 9147–9151. https://doi.org/10.1073/pnas.0913737107.

(17) Coughlin, R. T.; Tonsager, S.; McGroarty, E. J. Quantitation of Metal Cations Bound to Membranes and Extracted Lipopolysaccharide of Escherichia Coli. *Biochemistry* **1983**, *22* (8), 2002–2007. https://doi.org/10.1021/bi00277a041.

(18) Parikh, S. J.; Chorover, J. Infrared Spectroscopy Studies of Cation Effects on Lipopolysaccharides in Aqueous Solution. *Colloids and Surfaces B: Biointerfaces* **2007**, *55* (2), 241–250. https://doi.org/10.1016/j.colsurfb.2006.12.014.

(19) Gard, E. E.; Mayer, J. E.; Morrical, B. D.; Dienes, T.; Fergenson, D. P.; Prather, K. A. Real-Time Analysis of Individual Atmospheric Aerosol Particles: Design and Performance of a Portable ATOFMS. *Analytical Chemistry* **1997**, *69* (20), 4083–4091. https://doi.org/10.1021/ac970540n.

(20) Garner, J.; Park, K. Chemically Modified Natural Polysaccharides to Form Gels. In *Polysaccharides*; Springer International Publishing: Cham, 2015; pp 1555–1582. https://doi.org/10.1007/978-3-319-16298-0_31.

(21) Herrmann, M.; Schneck, E.; Gutsmann, T.; Brandenburg, K.; Tanaka, M. Bacterial Lipopolysaccharides Form Physically Cross-Linked, Two-Dimensional Gels in the Presence of Divalent Cations. *Soft Matter* **2015**. https://doi.org/10.1039/c5sm01002k.

(22) Prather, K. A.; Bertram, T. H.; Grassian, V. H.; Deane, G. B.; Stokes, M. D.; Demott, P. J.; Aluwihare, L. I.; Palenik, B. P.; Azam, F.; Seinfeld, J. H.; Moffet, R. C.; Molina, M. J.; Cappa, C. D.; Geiger, F. M.; Roberts, G. C.; Russell, L. M.; Ault, A. P.; Baltrusaitis, J.; Collins, D. B.; Corrigan, C. E.; Cuadra-Rodriguez, L. a; Ebben, C. J.; Forestieri, S. D.; Guasco, T. L.; Hersey, S. P.; Kim, M. J.; Lambert, W. F.; Modini, R. L.; Mui, W.; Pedler, B. E.; Ruppel, M. J.; Ryder, O. S.; Schoepp, N. G.; Sullivan, R. C.; Zhao, D. Bringing the Ocean into the Laboratory to Probe the Chemical Complexity of Sea Spray Aerosol. *Proceedings of the National Academy of Sciences of the United States of America* **2013**, *110* (19), 7550–7555. https://doi.org/10.1073/pnas.1300262110.

(23) McMurry, P. A Review of Atmospheric Aerosol Measurements. *Atmospheric Environment* **2000**, *34* (12–14), 1959–1999. https://doi.org/10.1016/S1352-2310(99)00455-0.

(24) Denton, J. K.; Kelleher, P. J.; Johnson, M. A.; Baer, M. D.; Kathmann, S. M.; Mundy, C. J.; Wellen Rudd, B. A.; Allen, H. C.; Choi, T. H.; Jordan, K. D. Molecular-Level Origin of the Carboxylate Head Group Response to Divalent Metal Ion Complexation at the Air–Water Interface. *Proceedings of the National Academy of Sciences* **2019**, *116* (30), 14874–14880. https://doi.org/10.1073/pnas.1818600116.

(25) Jayarathne, T.; Sultana, C. M.; Lee, C.; Malfatti, F.; Cox, J. L.; Pendergraft, M. A.; Moore, K. A.; Azam, F.; Tivanski, A. V.; Cappa, C. D.; Bertram, T. H.; Grassian, V. H.; Prather, K. A.; Stone, E. A. Enrichment of Saccharides and Divalent Cations in Sea Spray Aerosol during Two Phytoplankton Blooms. *Environmental Science and Technology* **2016**, *50* (21), 11511–11520. https://doi.org/10.1021/acs.est.6b02988.

(26) Duce, R. A.; Hoffman, E. J. Chemical Fractionation at the Air/Sea Interface. *Annual Review of Earth and Planetary Sciences* **1976**, *4* (1), 187–228. https://doi.org/10.1146/annurev.ea.04.050176.001155.

(27) Gaston, C. J.; Furutani, H.; Guazzotti, S. a.; Coffee, K. R.; Bates, T. S.; Quinn, P. K.; Aluwihare, L. I.; Mitchell, B. G.; Prather, K. a. Unique Ocean-Derived Particles Serve as a Proxy for Changes in Ocean Chemistry. *Journal of Geophysical Research: Atmospheres* **2011**, *116* (18), 1–13. https://doi.org/10.1029/2010JD015289.

(28) Creamean, J. M.; Lee, C.; Hill, T. C.; Ault, A. P.; DeMott, P. J.; White, A. B.; Ralph, F. M.; Prather, K. A. Chemical Properties of Insoluble Precipitation Residue Particles. *Journal of Aerosol Science* **2014**, *76*, 13–27. https://doi.org/10.1016/j.jaerosci.2014.05.005.

(29) Passmore, L. A.; Russo, C. J. *Specimen Preparation for High-Resolution Cryo-EM*, 1st ed.; Elsevier Inc., 2016; Vol. 579. https://doi.org/10.1016/bs.mie.2016.04.011.

(30) Booth, D. S.; Avila-Sakar, A.; Cheng, Y. Visualizing Proteins and Macromolecular Complexes by Negative Stain EM: From Grid Preparation to Image Acquisition. *Journal of visualized experiments : JoVE* **2011**, No. 58, 1–7. https://doi.org/10.3791/3227.

(31) Santos, N. C.; Silva, A. C.; Castanho, M. a R. B.; Martins-Silva, J.; Saldanha, C. Evaluation of Lipopolysaccharide Aggregation by Light Scattering Spectroscopy. *ChemBioChem* **2003**, *4* (1), 96–100. https://doi.org/10.1002/cbic.200390020.

(32) Snyder, S.; Kim, D.; McIntosh, T. J. Lipopolysaccharide Bilayer Structure: Effect of Chemotype, Core Mutations, Divalent Cations, and Temperature. *Biochemistry* **1999**, *38* (33), 10758–10767. https://doi.org/10.1021/bi990867d.

(33) Jo, S.; Kim, T.; Iyer, V. G. V. G.; Im, W. CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM. *Journal of Computational Chemistry* **2008**, *29* (11), 1859–1865. https://doi.org/10.1002/jcc.20945.

(34) Huang, J.; Mackerell, A. D. CHARMM36 All-Atom Additive Protein Force Field: Validation Based on Comparison to NMR Data. *Journal of Computational Chemistry* **2013**, *34* (25), 2135–2145. https://doi.org/10.1002/jcc.23354.

(35) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D. CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nature Methods* **2016**, *14* (1), 71–73. https://doi.org/10.1038/nmeth.4067.

(36) Schneck, E.; Papp-Szabo, E.; Quinn, B. E.; Konovalov, O. V; Beveridge, T. J.; Pink, D. A.; Tanaka, M. Calcium Ions Induce Collapse of Charged O-Side Chains of Lipopolysaccharides from Pseudomonas Aeruginosa. *Journal of The Royal Society Interface* **2009**, *6*, S671-8. https://doi.org/10.1098/rsif.2009.0190.focus.

(37) Jeworrek, C.; Evers, F.; Howe, J.; Brandenburg, K.; Tolan, M.; Winter, R. Effects of Specific versus Nonspecific Ionic Interactions on the Structure and Lateral Organization of Lipopolysaccharides. *Biophysical Journal* **2011**, *100* (9), 2169–2177. https://doi.org/10.1016/j.bpj.2011.03.019.

(38) Orellana, M. V; Matrai, P. A.; Leck, C.; Rauschenberg, C. D.; Lee, A. M.; Coz, E. Marine Microgels as a Source of Cloud Condensation Nuclei in the High Arctic. *Proceedings of the National Academy of Sciences of the United States of America* **2011**, *108* (33), 13612–13617. https://doi.org/10.1073/pnas.1102457108.

(39) Ault, A. P.; Guasco, T. L.; Baltrusaitis, J.; Ryder, O. S.; Trueblood, J. V.; Collins, D. B.; Ruppel, M. J.; Cuadra-Rodriguez, L. a.; Prather, K. A.; Grassian, V. H. Heterogeneous Reactivity of Nitric Acid with Nascent Sea Spray Aerosol: Large Differences Observed between and within Individual Particles. *Journal of Physical Chemistry Letters* **2014**, *5* (15), 2493–2500. https://doi.org/10.1021/jz5008802.

(40) Andreae, M. O.; Rosenfeld, D. Aerosol-Cloud-Precipitation Interactions. Part 1. The Nature and Sources of Cloud-Active Aerosols. *Earth-Science Reviews* **2008**, *89* (1–2), 13–41. https://doi.org/10.1016/j.earscirev.2008.03.001.

(41) Collins, D. B.; Ault, A. P.; Moffet, R. C.; Ruppel, M. J.; Cuadra-Rodriguez, L. A.; Guasco, T. L.; Corrigan, C. E.; Pedler, B. E.; Azam, F.; Aluwihare, L. I.; Bertram, T. H.; Roberts, G. C.; Grassian, V. H.; Prather, K. A. Impact of Marine Biogeochemistry on the Chemical Mixing State and Cloud Forming Ability of Nascent Sea Spray Aerosol. *Journal of Geophysical Research: Atmospheres* **2013**, *118* (15), 8553–8565. https://doi.org/10.1002/jgrd.50598.

(42) McCluskey, C. S.; Hill, T. C. J.; Sultana, C. M.; Laskina, O.; Trueblood, J. V.; Santander, M. V.; Beall, C. M.; Michaud, J. M.; Kreidenweis, S. M.; Prather, K. A.; Grassian, V. H.; DeMott, P. J. A Mesocosm Double Feature: Insights into the Chemical Make-Up of Marine Ice Nucleating Particles. *Journal of the Atmospheric Sciences* **2018**, *75* (7), 2405–2423. https://doi.org/10.1175/jas-d-17-0155.1.

(43) Sandhiya, L.; Kolandaivel, P.; Senthilkumar, K. Oxidation and Nitration of Tyrosine by Ozone and Nitrogen Dioxide: Reaction Mechanisms and Biological and Atmospheric Implications. *The Journal of Physical Chemistry B* **2014**, *118* (13), 3479–3490. https://doi.org/10.1021/jp4106037.

(44) Estillore, A.; Morris, H.; Or, V. W.; Lee, H. D.; Alves, M. R.; Marciano, M. A.; Laskina, O.; Qin, Z.; Tivanski, A. V; Grassian, V. Linking Hygroscopicity and Surface Microstructure of Model Inorganic Salts, Simple and Complex Carbohydrates, and Authentic Sea Spray Aerosol Particles. *Physical Chemistry Chemical Physics* **2017**, *19* (31), 21101–21111. https://doi.org/10.1039/C7CP04051B.

(45) Aurell, C. A.; Wistrom, A. O. Critical Aggregation Concentrations of Gram-Negative Bacterial Lipopolysaccharides (LPS). *Biochemical and biophysical research communications* **1998**, *253* (1), 119–123. https://doi.org/10.1006/bbrc.1998.9773.

(46) Coughlin, R. T.; Haug, A.; McGroarty, E. J. Physical Properties of Defined Lipopolysaccharide Salts. *Biochemistry* **1983**, *22* (1978), 2007–2013.

(47) Li, L.; Luo, R. G. Use of Ca+2 to Re-Aggregate Lipopolysaccharide (LPS) in Hemoglobin Solutions and the Subsequent Removal of Endotoxin by Ultrafiltration. *Biotechnology Techniques* **1998**, *12* (2), 119–122. https://doi.org/10.1023/A:1008884332586.

(48) Parikh, S. J.; Chorover, J. Infrared Spectroscopy Studies of Cation Effects on Lipopolysaccharides in Aqueous Solution. *Colloids and Surfaces B: Biointerfaces* **2007**, *55* (2), 241–250. https://doi.org/10.1016/j.colsurfb.2006.12.014.

(49) May, K. R. The Collison Nebulizer: Description, Performance and Application. *Journal of Aerosol Science* **1973**, *4* (3), 235–243. https://doi.org/10.1016/0021-8502(73)90006-2.

(50) Gard, E. E.; Mayer, J. E.; Morrical, B. D.; Dienes, T.; Fergenson, D. P.; Prather, K. A. Real-Time Analysis of Individual Atmospheric Aerosol Particles: Design and Performance of a Portable ATOFMS. *Analytical Chemistry* **1997**, *69* (20), 4083–4091. https://doi.org/10.1021/ac970540n.

(51) Sultana, C. M.; Cornwell, G. C.; Rodriguez, P.; Prather, K. A. FATES: A Flexible Analysis Toolkit for the Exploration of Single-Particle Mass Spectrometer Data. *Atmospheric Measurement Techniques* **2017**, *10* (4), 1323–1334. https://doi.org/10.5194/amt-10-1323-2017.

(52) DeCarlo, P. F.; Slowik, J. G. Particle Morphology and Density Characterization by Combined Mobility and Aerodynamic Diameter Measurements. Part 1: Theory. *Aerosol Science and Technology* **2004**, *38* (12), 1185–1205. https://doi.org/10.1080/027868290903907.

(53) Zelenyuk, A.; Imre, D.; Cuadra-Rodriguez, L. A.; Ellison, B. Measurements and Interpretation of the Effect of a Soluble Organic Surfactant on the Density, Shape and Water Uptake of Hygroscopic Particles. *Journal of Aerosol Science* **2007**, *38* (9), 903–923. https://doi.org/10.1016/j.jaerosci.2007.06.006.

(54) Booth, D. S.; Avila-Sakar, A.; Cheng, Y. Visualizing Proteins and Macromolecular Complexes by Negative Stain EM: From Grid Preparation to Image Acquisition. *J. Vis. Exp* **2011**, No. 58, 3227–3227. https://doi.org/10.3791/3227.

(55) Passmore, L. A.; Russo, C. J. *Specimen Preparation for High-Resolution Cryo-EM*, 1st ed.; Elsevier Inc., 2016; Vol. 579, p 86. https://doi.org/10.1016/bs.mie.2016.04.011.

(56) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM. *Journal of computational chemistry* **2008**, *29*, 2967–2970. https://doi.org/10.1002/jcc.

(57) Lu, J.; Qiu, Y.; Baron, R.; Molinero, V. Coarse-Graining of TIP4P/2005, TIP4P-Ew, SPC/E, and TIP3P to Monatomic Anisotropic Water Models Using Relative Entropy Minimization. *Journal of Chemical Theory and Computation* **2014**, *10* (9), 4104–4120. https://doi.org/10.1021/ct500487h.

(58) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *The Journal of Chemical Physics* **1983**, *79* (2), 926–935. https://doi.org/10.1063/1.445869.

(59) Vega, C.; Abascal, J. L. F. Simulating Water with Rigid Non-Polarizable Models: A General Perspective. *Physical Chemistry Chemical Physics* **2011**, *13* (44), 19663–19663. https://doi.org/10.1039/c1cp22168j.

(60) Huang, J.; Mackerell, A. D. CHARMM36 All-Atom Additive Protein Force Field: Validation Based on Comparison to NMR Data. *Journal of Computational Chemistry* **2013**, *34* (25), 2135–2145. https://doi.org/10.1002/jcc.23354.

(61) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. GROMACS: A Message-Passing Parallel Molecular Dynamics Implementation. *Computer Physics Communications* **1995**, *91* (1–3), 43–56. https://doi.org/10.1016/0010-4655(95)00042-E.

(62) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25. https://doi.org/10.1016/J.SOFTX.2015.06.001.

(63) Towns, J.; Cockerill, T.; Dahan, M.; Foster, I.; Gaither, K.; Grimshaw, A.; Hazlewood, V.; Lathrop, S.; Lifka, D.; Peterson, G. D.; Roskies, R.; Scott, J. R.; Wilkens-Diehr, N. XSEDE: Accelerating Scientific Discovery. *Computing in Science & Engineering* **2014**, *16* (5), 62–74. https://doi.org/10.1109/MCSE.2014.80.

(64) Nystrom, N. A.; Levine, M. J.; Roskies, R. Z.; Scott, J. R. Bridges: A Uniquely Flexible HPC Resource for New Communities and Data Analytics. In *Proceedings of the 2015 XSEDE Conference on Scientific*

Advancements Enabled by Enhanced Cyberinfrastructure - XSEDE '15; 2015; pp 1–8. https://doi.org/10.1145/2792745.2792775.

(65) Ziegel, E.; Press, W.; Flannery, B.; Teukolsky, S.; Vetterling, W. *Numerical Recipes: The Art of Scientific Computing*; 1987; Vol. 29, p 501. https://doi.org/10.2307/1269484.

(66) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N·log(N) Method for Ewald Sums in Large Systems. *The Journal of Chemical Physics* **1993**, *98* (12), 10089–10092. https://doi.org/10.1063/1.464397.

(67) Lamm, G.; Szabo, A. Langevin Modes of Macromolecules. *The Journal of Chemical Physics* **1986**, *85* (12), 7334–7334. https://doi.org/10.1063/1.451373.

(68) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *Journal of Computational Physics* **1977**, *23* (3), 327–341. https://doi.org/10.1016/0021-9991(77)90098-5.

(69) McGibbon, R. T.; Beauchamp, K. A.; Harrigan, M. P.; Klein, C.; Swails, J. M.; Hernández, C. X.; Schwantes, C. R.; Wang, L.-P.; Lane, T. J.; Pande, V. S. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophysical Journal* **2015**, *109* (8), 1528–1532. https://doi.org/10.1016/j.bpj.2015.08.015.

(70) Pérez, F.; Granger, B. E. *IPython: A System for Interactive Scientific Computing*.

(71) Humphery Dalke, A. S., K. ,. W. VMD - Visual Molecular Dynamics. *J. Molec. Graph* **1996**, *14*, 33–38.

(72) Wang, X.; Sultana, C. M.; Trueblood, J.; Hill, T. C. J.; Malfatti, F.; Lee, C.; Laskina, O.; Moore, K. A.; Beall, C. M.; McCluskey, C. S.; Cornwell, G. C.; Zhou, Y.; Cox, J. L.; Pendergraft, M. A.; Santander, M. V.; Bertram, T. H.; Cappa, C. D.; Azam, F.; DeMott, P. J.; Grassian, V. H.; Prather, K. A. Microbial Control of Sea Spray Aerosol Composition: A Tale of Two Blooms. *ACS Central Science* **2015**, *1* (3), 124–131. https://doi.org/10.1021/acscentsci.5b00148.

(73) Cochran, R. E.; Laskina, O.; Jayarathne, T.; Laskin, A.; Laskin, J.; Lin, P.; Sultana, C.; Lee, C.; Moore, K. A.; Cappa, C. D.; Bertram, T. H.; Prather, K. A.; Grassian, V. H.; Stone, E. A. Analysis of Organic Anionic Surfactants in Fine and Coarse Fractions of Freshly Emitted Sea Spray Aerosol. *Environmental Science and Technology* **2016**, *50* (5), 2477–2486. https://doi.org/10.1021/acs.est.5b04053.

(74) Cochran, R. E.; Laskina, O.; Trueblood, J.; Estillore, A. D.; Morris, H. S.; Jayarathne, T.; Sultana, C. M.; Lee, C.; Lin, P.; Laskin, J.; Laskin, A.; Dowling, J.; Qin, Z.; Cappa, C. D.; Bertram, T. H.; Tivanski, A. V.; Stone, E. A.; Prather, K. A.; Grassian, V. H. Molecular Characterization of Sea Spray Particles : Influence of Ocean Biology on Particle Composition and Interaction with Water. *Chem* **2017**, *2*, 655–667. https://doi.org/10.1016/j.chempr.2017.03.007.

(75) Jayarathne, T.; Sultana, C. M.; Lee, C.; Malfatti, F.; Cox, J. L.; Pendergraft, M. A.; Moore, K. A.; Azam, F.; Tivanski, A. V.; Cappa, C. D.; Bertram, T. H.; Grassian, V. H.; Prather, K. A.; Stone, E. A. Enrichment of Saccharides and Divalent Cations in Sea Spray Aerosol during Two Phytoplankton Blooms. *Environmental Science and Technology* **2016**. https://doi.org/10.1021/acs.est.6b02988.

Chapter 5

All-Atom Molecular Dynamics Simulations Reveal Microemulsion-Like Phase Behavior in Model Submicron Marine Aerosols

5.1 Abstract

Submicron sea spray aerosols ejected through bursting film caps at the ocean surface contain nearly 50% organic material by mass. Their morphology impacts their climate-relevant properties such as cloud condensation, uptake of atmospheric gases, and interaction with solar radiation. The study of submicron aerosols is hindered by limitations in single-particle analysis techniques, necessitating the use of computational methods to interrogate these systems at the molecular level. Here, large-scale all-atom molecular dynamics simulations are performed on 40-nm model aerosols to investigate how chemical complexity impacts the partitioning of organic material throughout the particle. We consider a range of organic constituents with varying chemical properties while also probing the impact of divalent cations. Our results indicate that despite variations in chemical complexity, complete phase separation between organic and aqueous components is unlikely. Surfactants readily partition between both the surface and interior and may play a role in stabilizing less-soluble organics in dispersed droplets. Our work suggests that submicron organic marine aerosols exhibit a microemulsion-like morphology in contrast to widely-accepted phase separation models.

5.2 Introduction

Atmospheric aerosols can impact climate by nucleating cloud droplets and ice crystals, scattering, absorbing, and reflecting solar radiation, and reacting with or sequestering atmospheric gases (e.g., pollutants). The representation of these aerosols in climate models contributes significant uncertainty to our predictions of global warming or cooling.[1,2] Sea spray aerosols (SSA) released into the atmosphere through bubble bursting at the ocean surface contain a wide range of biological and organic molecules, including protein, saccharides, alkanes, and even whole or fragmented bacteria and viruses.[3–7] The immense chemical diversity of SSA,

combined with significant limitations in single particle analysis techniques, makes the estimation of their climate impacts challenging.

Aerosol morphology is a critical driver of atmospheric properties; particle surfaces are responsible for interacting with gases and sunlight,[8–11] while the organization of their interior regulates the absorption and diffusion of water which can control reactivity and phase state.[12,13] The current understanding of SSA morphology is derived from experimental observation. It has been widely accepted that SSA adopt a so-called "core-shell" morphology,[13–18] in which organic materials phase separate to the aerosol surface to form a thick organic coating around the aqueous core (**Figure 5.5A**). Other biphasic morphologies have been observed, as well as an evenly-mixed homogeneous morphology (**Figure 5.5**).[19–26] Microscopy techniques such as atomic force microscopy (AFM), scanning electron microscopy (SEM), and cryogenic transmission electron microscopy (cryo-TEM), while extraordinarily valuable, require the deposition of aerosols onto a substrate which alters particle shape and surface area,[27] necessarily compromising the structural integrity of the particle. Other techniques such as microfluidics and aerosol optical tweezers (AOT) are similarly limited in their size resolution, aerosol generation methods, and for AOT, ability to investigate non-spherical particles.

Molecular dynamics (MD) simulations have become a more popular method in recent years as an alternative to conventional wet lab techniques to interrogate a variety of properties of nanoscale aerosols, including water condensation,[28–34] atmospheric gas uptake,[35] coalescence,[36] and phase state.[17,30,32,33,37–41] A majority of studies investigate small aerosol clusters containing low molecular-weight mono- and dicarboxylic acids related to anthropogenic or natural emission of volatile organic carbon. Other studies have approximated sea spray by combining medium-chain fatty acids or free amino acids with saltwater clusters. A central limitation to all MD studies is balancing system size and complexity with simulation length, making the simulation of large, complex aerosols for statistically significant timescales particularly challenging. Thus, the aerosol models studied to date have largely contained simple binary mixtures in the size range of ~2-12 nm with simulation lengths of ~2-10 ns. However, despite the short timescales and small system sizes, these simulations have overwhelmingly supported experimental evidence that insoluble organics fully phase separate, typically resulting in an organic coating at the particle surface.

In the present work, we aim to bridge the gap between experimental resolution and previous small-scale studies by 1) extending the aerosol size up to four-fold in comparison to previous all-atom MD simulations; and 2) increasing the chemical complexity of both the organic and inorganic phases to be more representative of nascent sea spray. Mimicking realistic SSA complexity, we simulate three different 40-nm diameter aerosol systems for 500-1000 ns each in triplicate. For each system, we increase the chemical complexity of the organic phase starting with a simple fatty acid and protein mixture, then adding perturbations of lipopolysaccharides, free glucose, and neutral glucose oligosaccharides. We evaluate particle shape and size, partitioning of the organic material, and extent of surface coverage by organic surfactants. Our simulations give us new insights into how particle phase and morphology are impacted by chemical diversity, and how that in turn may impact climate-relevant properties of submicron sea spray.

5.3 Experimental

5.3.1 Experimental Design.

The driving science question was to investigate the impacts of chemical complexity on the phase, morphology, and dynamics of nanoscale marine aerosols. We had several limitations to keep in mind while designing the experiments. First, the CHARMM36 forcefield[42–44] was used for all simulations, and care was taken to ensure that only molecules that had been sufficiently parameterized within the forcefield were investigated. Additionally, ion concentrations were kept within the forcefield accuracy limitations. Secondly, 30-50 nm diameter aerosols are the most difficult to study experimentally due to their small size yet correspond to computationally large systems containing multiple millions of atoms and require leadership class computing facilities. We thus needed to carefully plan our simulations so that they would be computationally feasible and would still give statistically rigorous results. Finally, the largest limitation is the vast chemical complexity of nascent marine aerosols. There is consensus on, broadly, the major classes of organic compounds, but narrowing down our selection of molecular components to include the most impactful species was non-trivial. We ultimately chose to build three 40-nm diameter aerosol systems to be run in triplicate (9 total simulations), with each system representing a step up in chemical complexity, i.e., adding a new molecular component to the organic mass fraction.

The chemical components of the SSA models were selected based on the most up-to-date molecular analyses of nascent submicron SSA; although, it is worth noting that at the time of writing, only an estimated 25% of the total chemical species found in SSA has been fully characterized.[45] There is, however, a consensus on the classes of molecules observed in SSA. Fatty acids make up an estimated 2-20% of the observed organic species by mass, with chain lengths varying from 12-18 carbons for the most abundant subset.[7] Thus, our most basic system ("B") is composed of an organic fraction containing only saturated fatty acids in a 1:2:4:3 ratio of chain lengths 12, 14, 16, and 18 carbons, respectively. The mass fraction of proteinaceous material in SSA is unknown, but most likely highly variable. We chose to initialize our simple system with a number of *Burkholderia cepacia* lipases (BCL). BCL is a robust enzyme found in SSA that has, notably, been observed to retain some activity after aerosolization.[4,46,47] This enzyme was also selected because of the authors' familiarity with its structural, biological, and enzymatic properties, as well as the ease with which collaborators could incorporate it into subsequent laboratory studies.[48,49] The number of lipases remains constant throughout our study at a mass fraction of approximately 3%. Finally, we initialized our basic system with 0.4 M NaCl to reflect the salinity of sea water.

To understand the specific impacts of chemical complexity on submicron SSA, we chose to both increase the cation complexity as well as the complexity of the organic species. In the subsequent two steps, the aqueous phase contained approximately 0.11 M Mg^{2+}, 25 mM Ca^{2+}, and 20 mM K^{+}, in addition to the existing NaCl concentration. When considering the components of the organic fraction, we focused on systems that had been studied before both computationally and experimentally, so that our simulations could ultimately be validated by laboratory experiments. For the first step, system "L", we added lipopolysaccharides, which are components of bacterial cell membranes that are often found in SSA due to cellular breakdown. We expected that this addition would be a valuable perturbation to the simple system in two major ways: first, it had been discovered that LPS-containing particles readily uptake and react with nitric acid in the atmosphere and are thus chemically significant.[12,50] Second, based on previous studies, we expected that the interaction between LPS and divalent cations such as Ca^{2+} would cause a phase change[12,51–56] within the particle that could be explored computationally.

Finally, it is estimated that at least 2-10% of the organic carbon mass in submicron SSA is made up of poly- and oligosaccharides, as well as free saccharides, the most abundant monomer of which is glucose.[45,57,58] Based on this data, for the final system (labeled "A"), we added a combination of free glucose and laminarin, a neutral glucose oligosaccharide. Recently, Richards et al. demonstrated that free glucose and other monosaccharides in sufficiently high concentrations and in the presence of divalent cations—typical conditions within atmospheric aerosols—form supramolecular ion-bridging interactions that can induce phase transitions to a rigid gel.[59] This unique phase behavior had been observed for macromolecular systems (such as those containing LPS described above) but not yet with small molecules. This final step up in complexity allows us to explore the impacts of neutral saccharides within the overall mixture and computationally explore any ionic bridging interactions that could lead to phase changes.

There are many caveats to our analysis given that the chemical compositions we explored are simplified approximations of realistic complexity. Given the enormous amount of chemical complexity within SSA, we maintain that the best approach is to start with a simple baseline and test small perturbations to understand the unique impacts of each contribution, which also gives us the opportunity to explore and extract previously unknown and untested intermolecular interactions in atomistic detail. We note that all fatty acids here are fully protonated to reflect the low pH environment of SSA.[60–62] The fatty acid protonation state, as well as the lack of unsaturated fatty acids, gives the monolayers and aggregates unique properties that may not necessarily reflect the true properties found in natural SSA. For example, fully protonated fatty acids are largely insoluble, and their monolayers are highly rigid and gel-like, especially without double bonds which would increase its fluidity.[63] Additionally, protonated fatty acids preferentially aggregate into oil droplets in the aqueous phase, whereas anionic forms tend to assemble into micellar and lamellar structures.[63,64] Finally, the work described in this paper represents primary, rather than aged, SSA; additional studies containing perturbations with more oxidized species are underway.

5.3.2 Constructing Sea Spray Aerosol Models

All fatty acid structures and glucose are readily available and parametrized in the CHARMM36m[42–44] forcefield. Our laminarin oligosaccharide model was constructed using the CHARMM-GUI glycan modeler.[65–

[67] The oligosaccharide contained a total of 15 glucose monomers linked together with $\beta(1\rightarrow3)$ linkages. Branches were incorporated at monomers 3 and 9 with $\beta(1\rightarrow6)$ linkages (see **Figure 5.6**). The *E. coli* lipopolysaccharide structure included here was generated using CHARMM-GUI's LPS Modeler[68] with Type 1 Lipid A, an R1 core, and 8 O11-antigen repeats. Details for this structure can be found in Lee and Dommer et al.[12] The structure of BCL (PDB ID: 3LIP)[69] was prepared for a pH 5 environment using NAMD *Psfgen*.[70] The number of water molecules was calculated based on an estimated[45] relative humidity for the particles of 70%, amounting to ~50% water by mass.[45]

The ingredients were distributed randomly during packing so as to allow the components to self-assemble without introducing any bias into the system. All molecular structures were randomly packed into 16 geometrically-distinct segments (to later combine into a sphere) using PACKMOL[71] in 3 successive rounds of packing. The most difficult structures to place (macromolecular structures BCL, laminarin, and LPS) were packed first, followed by small molecules including fatty acids and glucose, and then ions and TIP3P[72,73] water. Packing the ingredients into small segments, which could then be patched together into one 40-nm diameter sphere, rather than packing all ingredients into a single large sphere at once, significantly reduced the packing time needed. The authors note that for future system-building, packing time can be further reduced by packing the ingredients at their respective concentrations into one small box and simply replicating and translating the box as needed. We would also like to note that it is possible to easily avoid catastrophic calculation errors when determining the ion counts and neutralizing the total system charge by using VMD and NAMD's *Psfgen, Autoionize,* and *Solvate* plugins,[70,74] which are invaluable tools for building large and complex systems.

Due to the difficulty of packing all of the ingredients to our target mass percentages (and significant computing resource restrictions), each replicate was constructed independently and immediately launched into simulation. Subsequent replicates were generated by modifying the ingredient recipe slightly to improve the chances of hitting our target mass percentage values. Additionally, fluctuations in the random seed in the PACKMOL packing algorithm introduced subtle differences in the number of each ingredient placed for each of the packed segments. Thus, the initial molecular configurations between replicates of the same system contain small random, as well as guided, variations. **Table 5.1** outlines the specific molecular components and their mass percentages within each model SSA construct.

Table 5.1: Molecular components of model SSAs and their mass percentages. Because of the variable nature of various carbohydrates and protein, the molecular weights of selected structures used in this work are provided. The systems are labeled B for the simplest system, "L" for the medium complexity system, and "A" for the highest complexity system. Numbers 1-3 represent each individual replicate.

Chemical Component, MW (g/mol)	B1	B2	B3	L1	L2	L3	A1	A2	A3
Organics									
Dodecanoic acid	3.40	3.67	3.63	2.59	2.79	3.81	2.06	2.00	2.10
Tetradecanoic acid	7.75	8.38	8.31	5.91	6.37	7.97	4.64	4.58	4.80
Hexadecanoic acid	17.41	18.85	18.68	13.26	14.3	14.65	10.44	10.31	10.81
Octadecanoic acid	14.48	15.67	15.52	11.04	11.9	13.54	8.68	8.58	9.00
BCL (34,100)	2.31	2.47	2.44	2.6	2.85	2.60	2.56	2.53	2.66
LPS (10,798)	--	--	--	6.70	7.23	6.58	5.41	6.42	6.73
Glucose	--	--	--	--	--	--	3.39	3.90	4.09
Laminarin (506)	--	--	--	--	--	--	0.46	3.64	3.82
Inorganics									
Water	51.92	48.07	48.55	56.60	53.22	49.64	58.30	55.25	53.10
Na^+	1.01	1.13	1.12	0.16	0.17	0.16	0.4	0.37	0.39
Ca^{2+}	--	--	--	0.18	0.20	0.18	0.18	0.24	0.24
Mg^{2+}	--	--	--	0.09	0.10	0.09	0.17	0.23	0.24
K^+	--	--	--	0.07	0.08	0.07	0.16	0.23	0.24

5.3.3 All-Atom Molecular Dynamics Simulation

The authors received a generous allocation through NSF-CAICE to run dynamics on the UIUC Blue Waters and TACC Frontera supercomputing facilities. Memory-Optimized NAMD 2.14, which has excellent scaling for very large systems across thousands of nodes,[70,75–77] was used to run all all-atom explicit-solvent molecular dynamics simulations with periodic boundary conditions. Each 40-nm aerosol was placed in an empty box approximately 55 nm to a side. The authors note that, due to the large size of the SSA model, small deformations in the aerosol as it evolves could cause it to interact with itself across a PBC boundary if the box is not sufficiently large. We found that leaving a 75 Å buffer on all sides of the aerosol particle was sufficient to prevent self-interaction.

Each system was energy minimized, heated, and equilibrated in multiple steps. Conjugate-gradient energy minimizations were performed with fatty acid headgroups harmonically restrained first at 100 kcal/mol Å^2 and then 10 kcal/mol Å^2 for 15,000 cycles each. The systems were heated from 23 K to 298.15 K in increments of 25 K for 100 ps each at a 2-fs timestep. Finally, the systems were equilibrated at 298.15 K by slowly releasing harmonic restraints on the lipid headgroups over 500 ps. Production runs were carried out for 500 ns for each system and extended to 1 μs in total for the B1 and A1 replicates. For all procedures, particle-mesh Ewald[78] electrostatics were employed for long-range electrostatic interactions; non-bonded van der Waals interactions and short-range electrostatics were calculated with a 12 Å cut-off. The SHAKE[79] algorithm was used to fix hydrogen bond lengths and a Langevin thermostat with a damping coefficient of 5/ps was applied to maintain temperature control at 298.15 K. **Table 5.2** details the atom counts for each system, as well as the total time simulated for each replicate.

Table 5.2: Atom counts and total simulation time for each system. The systems are labeled B for the simplest system, "L" for the medium complexity system, and "A" for the highest complexity system. Numbers 1-3 represent each individual replicate.

	B1	**B2**	**B3**	**L1**	**L2**	**L3**	**A1**	**A2**	**A3**
Atom count ($\times 10^6$)	3.083	2.907	2.933	2.649	2.462	2.731	2.102	2.670	2.546
Simulation time (ns)	1,000	515	502	533	574	530	1000	597	593

5.4 Results and Discussion

To track the evolution of aerosol shape, we quantify the degree of asphericity, ϕ, and the relative shape anisotropy parameter, κ^2, which can be derived from the radius of gyration tensors. A detailed description of the calculations and derivations used here is given in the Supporting Information. As the nanoaerosols evolved, their initial spherical shape often deformed into an ellipsoid, which can be characterized using the asphericity term. This term is zero when the particle is a perfect sphere and increases as the shape evolves away from sphericity. In a complementary analysis, κ^2 was used to describe the symmetry of the particle. For rod-like symmetry, κ^2 approaches 1, where all atoms lie along a line, while a value of 0 indicates that the particle has a higher degree of symmetry, such as that of a perfect tetrahedron or sphere. The evolution of ϕ and κ^2 for each

replicate of each aerosol system is plotted over time in **Figure 5.1**. Our analysis of aerosol shape was based on the one developed by Karadima et al.[17,18,80]

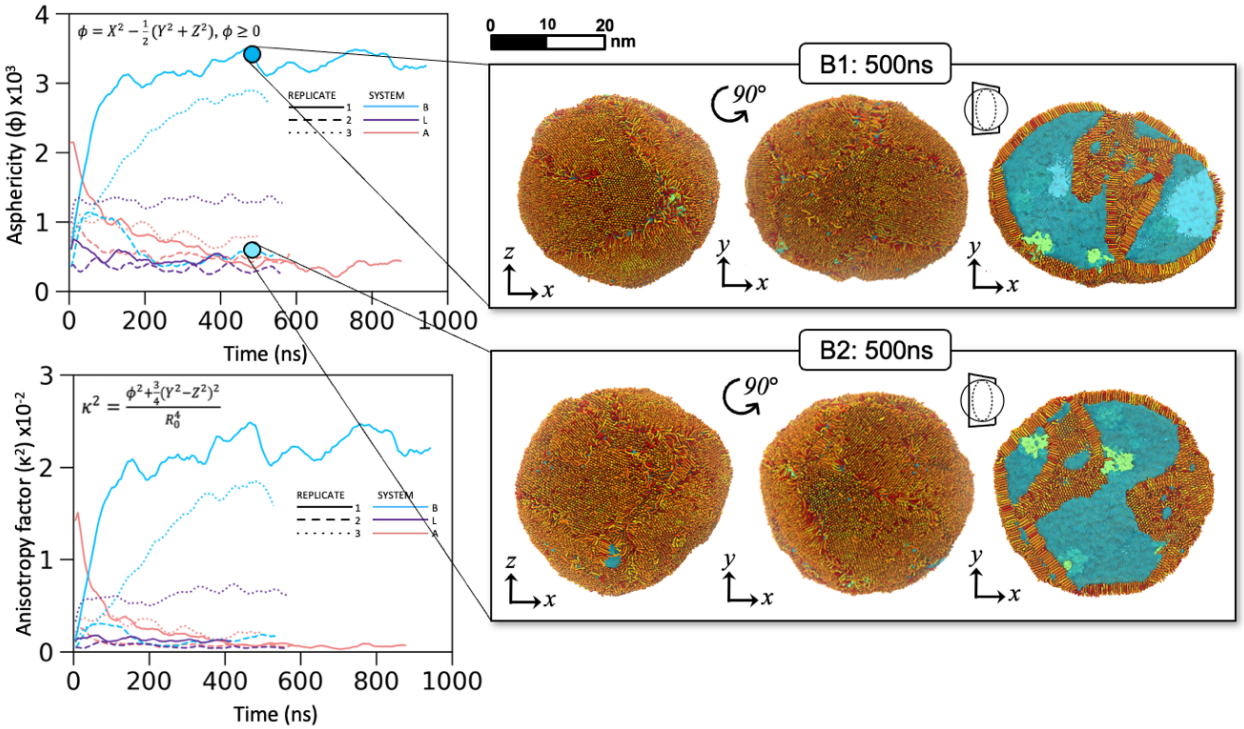


Figure 5.1: Evolution of aerosol shape over time, designated by the asphericity factor (top left) and relative shape anisotropy factor (bottom left). MD snapshots of B1 and B2, corresponding to the *xy* and *xz* planes at 500 ns, are provided on the right for visual shape comparison. Lipids are colored orange, red, and yellow, BCL appears in green, and water in blue.

Figure 5.1 indicates that the largest deviations from sphericity occur with the evolution of the B systems, which contain organic fractions composed of only fatty acids. The first and third B replicates also show the largest changes in symmetry, corresponding to the most dramatic changes in the κ^2 parameter. To illustrate the variations in aerosol shape, we provide MD snapshots from the 500 ns timepoint for replicates B1 and B2. The B1 replicate (solid line) evolves into an ellipsoidal shape within the first 100 ns of simulation time. B2 (dashed line) initially approaches and hovers around ϕ = 1000 before falling to ϕ = 500 at 200 ns, which corresponds to a more spherical and symmetric structure than B1 and more closely resembles the shape of the other systems.

These dramatic evolutions in particle shape appear to correspond to the higher fatty acid content relative to the other systems. Fatty acids, particularly protonated fatty acids, have a high hydrophobicity and favor

formation of highly ordered structures on a rapid timescale. This self-assembly has been explored experimentally on bulk phase fatty acids in aqueous solutions[63] and our work highlights a similar behavior. Once the surface saturates with a monolayer, fatty acids begin to aggregate in bulk, forming micelles and lamellar vesicles of tunable size based on fatty acid chain length, temperature, pH, and salt concentration.[63,81–90] In fact, vesicle formation within intact supermicron sea spray has been observed previously by Patterson et al. *via* cryogenic transmission electron microscopy.[3] However, nanoaerosols do not have sufficient volume for micelle or vesicle formation. Rather, the excess lipids—those not occupying space at the surface—will aggregate into amorphous oil droplets that either adsorb to the monolayer or are freely suspended in the aqueous phase. The aggregates that form in the present simulations often take the form of a bilayer and/or lipid agglomerate. For replicates B1 and B3, the aggregate is a bilayer that bisects the aerosol nearly completely, leading to the evolution of an ellipsoidal shape; in the more spherical B2, the aggregate consists of double bilayers that adhere to one side of the particle (**Figure 5.1**).

Of the 9 systems, 7 evolve rapidly and then reach a relatively constant value for both parameters within 200 ns. B3 and A1, however, are much slower to converge. It is difficult to say without simulating to longer timescales whether any of the systems have reached a true equilibrium. The asphericity plot of replicate B2 shows an extended plateau between t = 30 and t = 200 ns; if simulation time had stopped at 150 ns, the data would have appeared to be converged at that point. B1 seems to show a general plateau, however, there is still a significant variability to these parameters at this plateau. We believe that, while the simulations may not have reached a true equilibrium, they have reached a local energy minimum that is likely an important structure when considering the highly dynamic atmospheric environment of nascent sea spray.

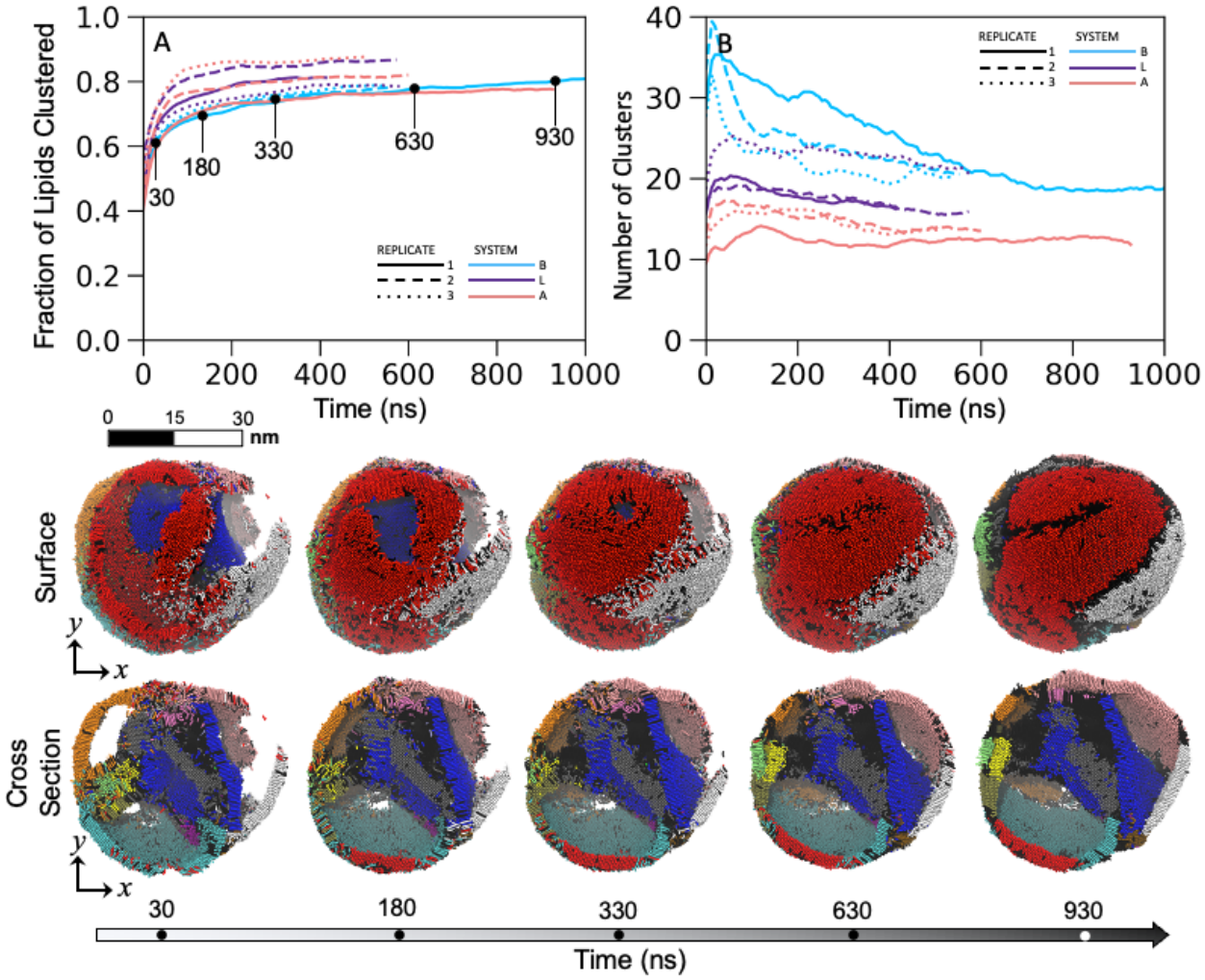


Figure 5.2: Lipid clustering analysis. DBSCAN clustering reveals the kinetics of lipid aggregation over the course of the simulation. A) Fraction of lipids clusters over time, and B) Total number of clusters evolved over time. In the lower panel, we provide snapshots from MD simulations of the B1 replicate, with unique clusters colored by ID. Only the lipids are visualized.

To understand the surfactant aggregation patterns, we clustered the fatty acids by position and the vector formed between the head and tail carbon atoms using the DBSCAN clustering method. This method doesn't require a predetermined number of clusters and allows for unclustered groups (**Figure 5.2**). The carbon vector was selected as a clustering parameter to characterize the formation of monolayers in which all lipids obtain a similar directional orientation and tilt; that is, the hydrophobic tails align and the carboxylic headgroups facing the aqueous phase. The results of this analysis show that the surfactants aggregate rapidly into long-lived clusters in under 200 ns, with over 60% of the total lipids clustered in the first 30 ns. **Figure 5.2** shows the fraction of

total lipids clustered as well as the total number of clusters calculated over the duration of the simulations. Here, we also provide snapshots from the MD simulations for replicate B1, with each cluster colored individually. Upon visual inspection, the first timepoint (30ns) already shows lipids organizing into their final clusters, with latter timesteps illustrating increased molecular alignment, greater surface coverage, and fewer unclustered patches.

One of the goals of this study was also to understand how all of the organic material is distributed between the aerosol surface and interior. This partitioning of organic material is of particular importance to atmospheric chemistry. Aerosols provide surfaces for water evaporation and condensation (influencing hygroscopicity[91] and cloud condensation)[92], reactive uptake of atmospheric gases,[10,50,93] and ice nucleation.[94,95] At planar interfaces, long chain saturated fatty acid surfactants form a tightly-packed monolayer at the air/aqueous interface that is known to have high rigidity, low fluidity, and impermeability to water—these are the aggregates that were identified above with our clustering method.[63,96–100] However, these monolayer properties can be disrupted or otherwise influenced by additional chemical components. For example, adhesion to the aqueous side of the monolayer by polyelectrolytes, such as LPS or alginate,[12,101] will slow molecular diffusion in all directions. Insertion of cosurfactants such as alcohols, cholesterols, or even shorter chain fatty acids may result in a more porous, higher-fluidity membrane.[96,100,102–106] It has also been shown that BCL can insert into surfactant monolayers and significantly disrupt the lipid ordering,[48,49] making the monolayers less stable and more permeable.

To quantitatively evaluate the distribution of organic material throughout each aerosol, we subdivided each particle into concentric ellipsoids of equal volume, following the general methods described by Karadima et al.[18] For the last snapshot of each simulation, we applied a convex hull method to approximate the bounding ellipsoid, then found a best-fit ellipsoidal mesh to describe the surface of the aerosol (See Section 6.7 for details). Using atom selections in VMD, we extracted the atoms located in each region defined as the surface, the bulk, or the core, and calculated their mass percentages (**Figure 5.3**, Illustration). We plotted the distribution of material by mass throughout the three regions (**Figure 5.3**, top panel).

This analysis shows that the organic material largely distributes to either the surface or the core. Additionally, the distribution is common throughout all 9 systems, which suggests that the chemical complexity

of the organic fraction does not impact the distribution of material. This is a critical finding given that the so-called "core-shell" morphology is the widely accepted paradigm for organic aerosol morphology. In other words, it is traditionally thought that the organic material should completely phase separate from the aqueous core to coat the particle surface as one thick layer.

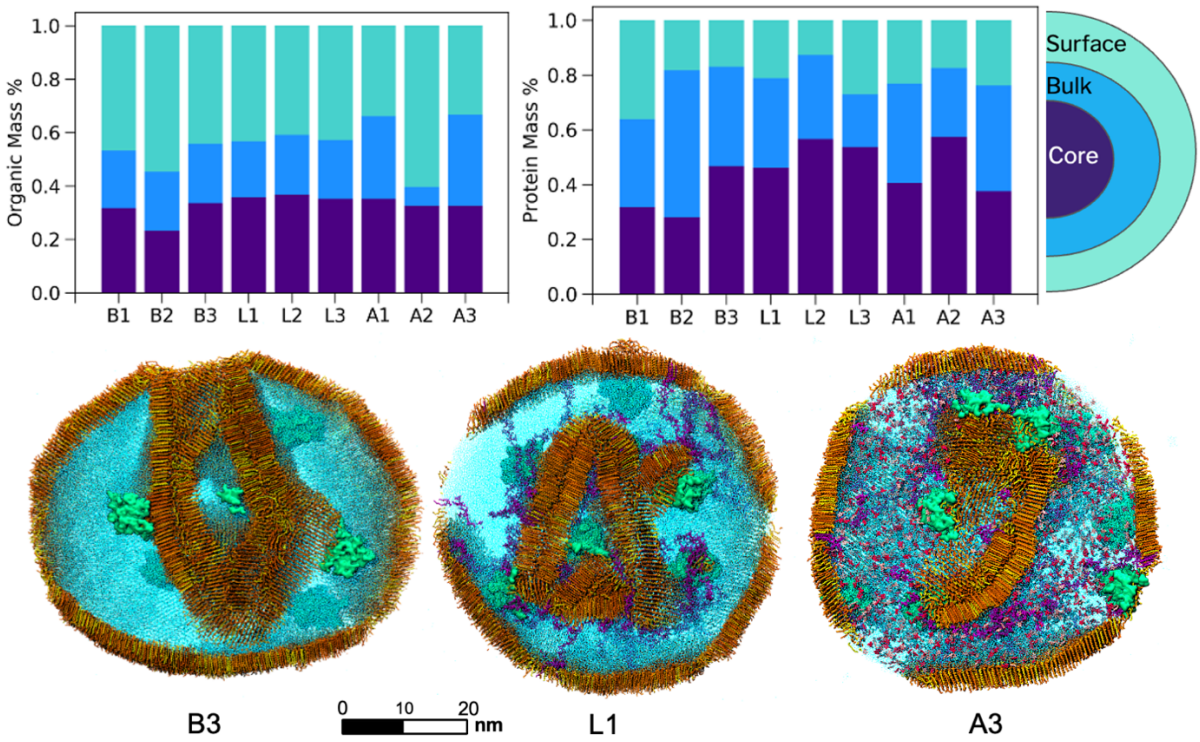


Figure 5.3: Distribution of organic material throughout the particle. Top panel: distribution of organic material (left) and protein (right) by mass percentage throughout the surface, bulk, and core regions. (See illustration to right for region labels). Bottom panel: Selected snapshots representing cross-sections from B1, L1, and A3 replicates at their final timesteps.

However, when considering the true complexity of the organic phase, our results should not be surprising. Rather than consisting of oily alkanes which will completely phase separate, the majority of the organic material in our simulations is amphiphilic or water-soluble in character. Both fatty acids and LPS molecules contain hydrophobic tails and polar headgroups, and can thus participate in monolayer, micellar, and bilayer aggregation. Glucose, due to its small size and hydroxy groups, is highly soluble and will remain in solution. Bulky, branched laminarin, although less soluble than glucose, still contains polar hydroxyl groups that give it partial solubility. Most notably, the hydrophobic material aggregates into inclusions in the aerosol center.

Mael et al. recently observed this phenomenon during water uptake experiments: neither complete phase separation nor dissolution occurs in the case of some organics, leading to the formation of inclusions.[91]

Finally, we wanted to understand how organic material aggregates at the surface. We expected that the most hydrophobic components, the fatty acid surfactants, would saturate the surface. Using the clustering data extracted from the simulations and shown above, we identified the fatty acid clusters that formed at the surfaces and estimated their total surface coverage (See **Section 5.7** for calculation details). **Figure 5.4** (lower panel) shows the fraction of surface area covered by monolayers compared to the total surface area of the final ellipsoid. As expected, the fraction of surface covered by lipids decreases with the total surfactant load: the percentage of fatty acids comprising the organic mass fraction decreases with increasing chemical complexity, from 100% (B) to 80% (L) to 70% (A).

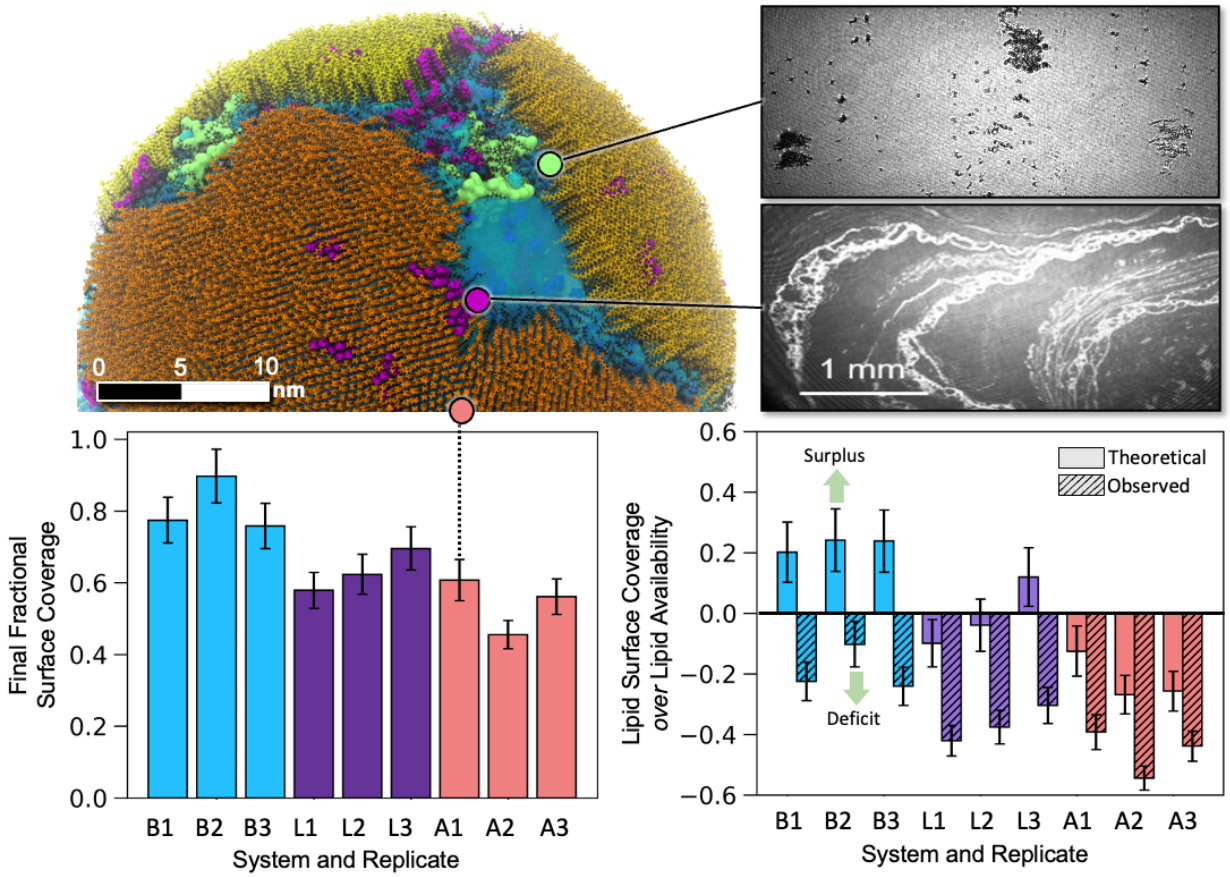


Figure 5.4: Fatty acid patches coat the particle surface, with other surface-active material disrupting the surface coverage. In A) a representative snapshot of A1 is shown highlight fatty acid clusters in yellow and orange, LPS in purple, and BCL in green. B) BAM imagery highlights the surface disruption of fatty acid monolayers either with BCL (top) or LPS (bottom). C) Surface coverage by fatty acids calculations shows a decrease in surface coverage as fatty acid content decreases and as the organic complexity increases. D) Theoretical calculations of fatty acid surface coverage overestimate the coverage as compared to the observed surface coverage.

To understand to what extent chemical complexity impacts how the surface saturates with fatty acids, we need to understand how many surfactants actually cover the surface with respect to how many surfactant molecules are available and compare this to the total surface area available. Thus, based on the number of lipids available and the final geometric surface area of the aerosol, we estimated how much of the surface should have theoretically been covered by lipids. In **Figure 5.4** we compare the observed surface saturation with the theoretical saturation should every "site" at the surface have been occupied by a surfactant molecule. A value of zero indicates that there are exactly the same number of lipids as there are available sites at the surface, where every surface site is occupied by a lipid. A positive value indicates that there are more lipids than surface sites

available (surplus), and *vice versa* for negative values, *i.e.,* a deficit means that there are less lipids than sites available for occupation at the surface. In the basic system (B), there are more lipids available than can fit at the surface, and we should expect 100% surface saturation. However, we observe that in each replicate, there are available sites at the surface that are unoccupied by lipids. For the most complex system (A), there is a lipid deficit for all replicates, but significantly less surfactants are present at the surface than would be expected. For all systems, there is significantly less surface covered by lipids than would be expected.

There are many explanations for these results. The calculation assumes that lipids will continue to accumulate at the air/water interface until the surface is completely saturated. This assumption is based on the large free energy difference, ΔG, between a surfactant molecule at the surface *vs* the monomer in bulk. The partitioning between the two states at equilibrium can be described by a Boltzmann distribution,

$$\frac{p_i}{p_j} = e^{(\varepsilon_j - \varepsilon_i)/kT} = e^{\Delta G/kT}, \tag{5.1}$$

where p is probability, ε is energy, and the subscripts i and j correspond to the surface or dissolved/monomeric state. At the air-sea interface, fatty acid surfactants are stabilized at the surface by the hydrophobic effect and have a probability of occurring at the interface roughly proportional to the number of carbon atoms in their tails.[63,107,108] Determining the free energies corresponding to the current fatty acid mixture is out of the scope of this work, but we provide a qualitative discussion of the impacts of chemical complexity on the distribution of organic material.

Notably, as we pointed out earlier, surface active molecular species other than the fatty acids in these models are also competing for spots at the interface. BCL and LPS, which were not included in the lipid clustering analysis, are non-conventional surfactants. BCL contains a hydrophobic region that preferentially embeds into nonpolar/aqueous interfaces,[48,49,109,110] and LPS contains Lipid A, which consists generally of two β(1→6)-linked glucosamine units attached to 6 acyl chains with structural similarity to fatty acids C12-C14.[56,111,112] Accounting for the regions where Lipid A or BCL accumulates at the surface may explain a small, but not insignificant, lipid deficit at the surface.

An important observation from the lipid clustering analysis provided above is that the fatty acids that accumulate at the surface do not do so uniformly across the entire surface, but rather, group into tessellated

clusters with unique tilt angles, illustrated by the red and white patches visible in **Figure 5.2.** The clustered lipid patches, which are highly ordered, adopt a lower curvature than would otherwise be expected (**Figure 5.8**) and are separated by "rivers" in which the lipids relax into a more disordered phase. This behavior can be attributed to the balance between the high bending modulus of the fatty acid monolayer (i.e., energy required to perturb the equilibrium curvature) and the line tension of the curved particle surface. Upon visual inspection, we found that both LPS and BCL accumulate at aerosol surfaces at these rivers between lipid patches. **Figure 5.4** shows a snapshot of the A1 particle surface with BCL (green) and LPS (purple) inserting between two fatty acid patches. This motif is consistent across all simulations containing BCL or LPS. BCL is more likely to be incorporated at the disordered river-like regions at the surface, rather than be integrated into a lipid patch. On the other hand, we found that the Lipid A region of LPS adsorbs to the surface in the rivers, but it also inserts readily into the fatty acid patches (likely due to the ability of fatty acid-like alkyl chains on Lipid A to align with the monolayer). Interestingly, we also found instances where the saccharide headgroup of the LPS also adsorbed to the surface in the river regions.

To corroborate our findings that BCL and LPS are likely to be found at the interface and that they preferentially insert between fatty acid patches, we conducted Brewster angle microscopy (BAM) experiments in which we added BCL or LPS into monolayers of our fatty acid mixture. Resulting BAM images are provided in **Figure 5.4.** Bright regions of the images indicate high organic content while dark regions indicate interfacial water. BCL (**Figure 5.4,** top image) appears to significantly disrupt the fatty acid monolayers, as expected and observed previously.[48,49] However, it also appears to aggregate, forming holes in the monolayer with little to no organized fatty acid density. The LPS, on the other hand (**Figure 5.4**, bottom image), aggregates at the surface to form rivers, regions of high and low organic content. The behavior seen here, although describing a phenomenon at a larger scale than our simulations, is consistent with computational observations. It is unclear, however, what the orientation of the LPS is at the interface, but our simulations suggest that the entire LPS molecule is surface active. The Lipid A region can insert directly into and contribute to the bright regions with high fatty acid density, while the saccharide headgroup can adsorb to the surface and appear in the darker, river-like regions. Further studies can help elucidate the mechanisms reported here.

We posit that in organic SSA, the organic fraction is less likely to completely phase separate (as many models suggest), and more likely to exhibit microemulsion-like characteristics. Although the aerosol models simulated here are not large enough in volume to form a true emulsion, insights from this work suggest that the organics found in SSA have a range of chemical properties conducive to the formation and stabilization of less-soluble domains. Proteins and saccharides, for example, have widely been used as emulsification agents in industrial applications.[113–120] Polyelectrolyte saccharides in particular, such as the saccharides in LPS or alginate, have been found to form stabilized oil-in-water emulsions with cationic surfactants and surfactant-like proteins.[117,120,121] Combined with the recent work suggesting that low-solubility biological macromolecules likely form inclusions in the particle center as they uptake water,[91] we suggest that the surfactant-like properties of marine organic material can further enable the stabilization of such partially-soluble inclusions.

5.5 Conclusion

We use ultra-large all-atom molecular dynamics simulations of submicron marine aerosols to understand the link between chemical complexity and aerosol morphology. We show that fatty acid surfactants readily aggregate and distribute to the surface and into oily aggregates within the aerosol. Simulations combined with BAM microscopy imaging suggest that rigid lipid monolayers at curved aerosol surfaces form discontinuous patches, separated by disordered regions in which amphiphilic species such as BCL and LPS can aggregate. The overall distribution of organic material throughout the particle is consistent across variations in the organic phase complexity, and in more complex particles, less of the aerosol surface is covered. We argue that organic SSA may not always phase separate to adopt a "core-shell" morphology where the aerosol surface is coated with a thick organic layer. Rather, we propose that organic SSA containing the full chemical complexity of biogenic marine material more likely adopts a microemulsion-like morphology, where amphiphilic surfactant-like species stabilize domains of less-soluble organic inclusions in the aerosol center.

5.6 Acknowledgments

We are grateful to the NSF for funding for this work through the grant CHE-1801971. Simulations were run on UIUC Blue Waters and TACC Frontera. ACD thanks Jim Phillips for invaluable help optimizing the NAMD code, which was crucial to getting these systems running. ACD also thanks Lorenzo Casalino for help

with big system building and rendering. ACD, REA, and NAW conceptualized the experiments presented in this work with supporting discussions from VHG and KAP. We are also grateful to Tim Bertram, Chris Cappa, Gil Nathanson and many other NSF-CAICE investigators for helpful discussions. ACD built the models, ran the simulations, wrote the manuscript and performed all analyses. NAW wrote and perfected the absolutely beautiful lipid clustering function and contributed to thorough editing of the manuscript. KJA and ML performed BAM microscopy experiments.

Chapter 5, in full, is in preparation for submission for publication and is presented in draft form containing mainly computational findings. It is presented as it may appear in "Dommer, A. C.; Wauer, N. A.; Angle, K. J.; Luo, M.; Grassian, V. H.; Amaro, R. E. *All-Atom Molecular Dynamics Simulations Reveal Microemulsion-Like Phase Behavior in Model Submicron Marine Aerosols*. 2022." The dissertation author was the primary investigator and co-author of this work.

5.7 Supporting Information

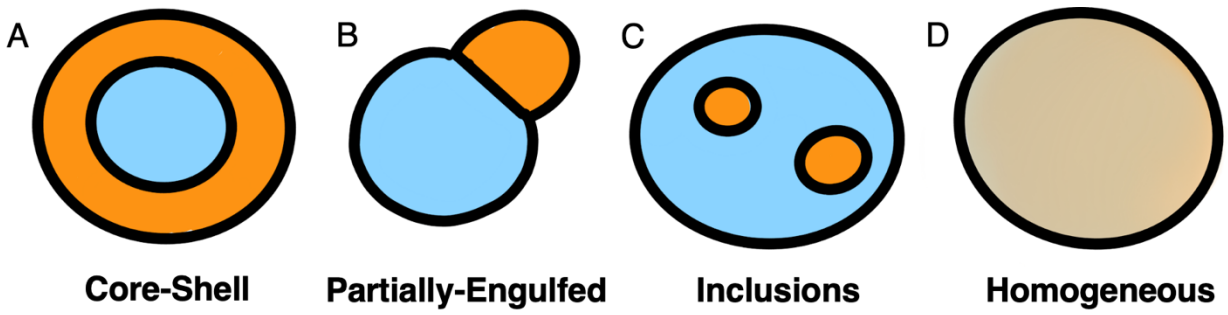


Figure 5.5: Possible SSA morphologies. Orange and blue colors represent organics and water, respectively.

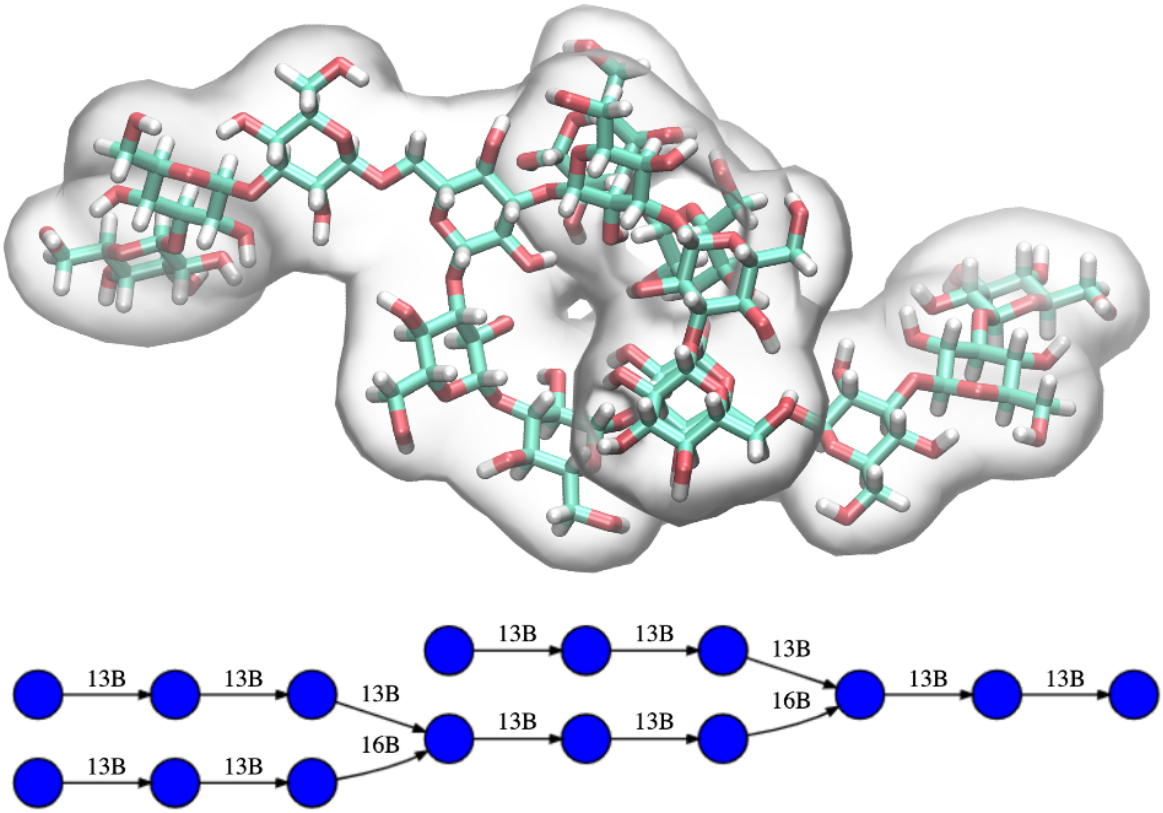


Figure 5.6: Rendering of the laminarin molecular structure with two branch points located at monomers 6 and 9. Below, a schematic diagram illustrating the linkages applied in the CHARMM-GUI Glycan Modeler.

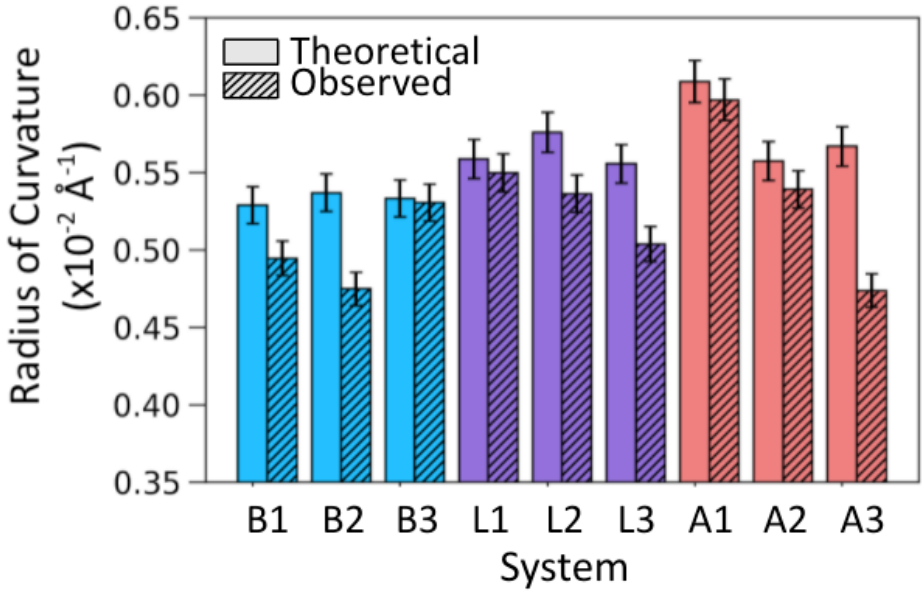


Figure 5.7: Estimates of surface lipid curvature compared to expected value of the curvature at the surface at that point given the best fit ellipsoid.

5.7.1 Calculation of Asphericity and Relative Shape Anisotropy

Here, we use the definition of asphericity derived by Theodorou and Suter:[80]

$$\phi = X^2 - \frac{1}{2}(Y^2 + Z^2),\ \phi \geq 0. \tag{5.2}$$

The quantities X^2, Y^2, and Z^2, are given by the symmetric radius of gyration tensor,

$$S = \begin{bmatrix} X^2 & XY & XZ \\ XY & Y^2 & YZ \\ XZ & YZ & Z^2 \end{bmatrix}, \tag{5.3}$$

such that the eigenvalues are $X^2 \geq Y^2 \geq Z^2$. S, derived from the definition

$$S_{m,n} = \frac{1}{N}\sum_{k=1}^{N} r_m^k r_n^k \tag{5.4}$$

where N represents the number of atoms in the system, $m \in \{1,2,3\}$, and $n \in \{1,2,3\}$, is computed directly from atomic coordinates. The squared radius of gyration, R_0^2, is given by the first invariant of S,

$$R_0^2 = X^2 + Y^2 + Z^2, \tag{5.5}$$

and is used to calculate the relative shape anisotropy, κ^2. This term is a shape descriptor of the symmetry of the particle; that is, a value of 1 indicates a rod-like symmetry, where all atoms lie along a line, while a value of 0 indicates the particle has a higher degree of symmetry, such as that of a perfect tetrahedron or sphere.[1,3] Here, we use the definition of κ^2 given by Theodorou and Suter:[1]

$$\kappa^2 = \frac{\phi^2 + \frac{3}{4}(Y^2 - Z^2)^2}{R_0^4}. \tag{5.6}$$

A close inspection of this equation reveals that the quantity in the numerator is a function of the asphericity and the parameter *acylindricity*, given by $Y^2 - Z^2$, where a value of zero indicates perfect cylindrical symmetry. A more detailed description of these quantities is out of the scope of this work but can be found in the provided references.

5.7.2 Organic Distribution by Volume

To quantify distribution of organic material throughout the particle, we calculate the best-fit ellipsoid to the dataset. Let *n*, M be the total particle mass, and *a, b,* and *c* correspond to the dimensions of each semi-axis, as

illustrated by **Figure 5.8**. *I* can be calculated using the internal *measure inertia* command in VMD,[74] which returns the eigenvalues and the principal axes.

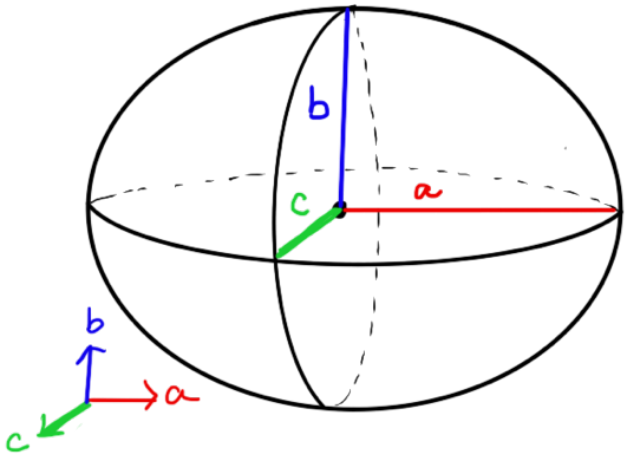


Figure 5.8: Sketch of an ellipsoid with axes of symmetry labeled.

We then subdivide the ellipsoid into three regions by volume, termed "core," "bulk," and "shell," using the approach outlined by Karadima et. al.[18] Drawing concentric ellipsoids along the principle axis each enveloping approximately a third of the volume, we can describe the distribution of mass by type into each region of the particle.

5.7.3 Estimation of Surface Curvature

Here, we use the inverse radius of the best fit sphere as our definition of curvature. That is, for the given set of datapoints, we find the best fit sphere using a NumPy linear algebra least squares function. In this case, the data points are the points defined by the fatty acid headgroups. **Figure 5.9** shows some example sphere fits to lipid selected lipid clusters.

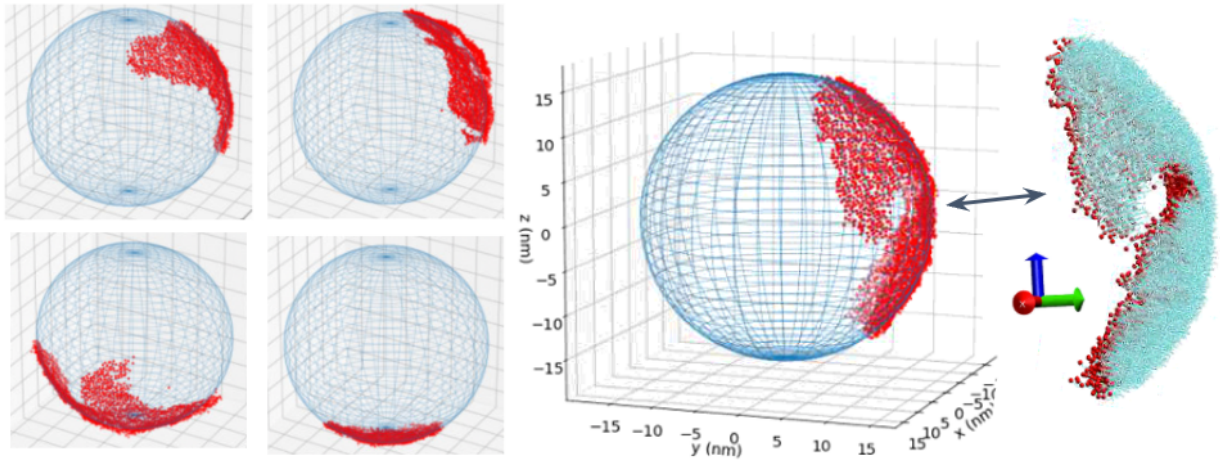


Figure 5.9: Spherical fits to selected lipid clusters to demonstrate the sphere fit function.

5.7.4 Estimation of Surface Area Coverage

To estimate the total surface area, we must first know the area per lipid of the particular surface cluster. Surface clusters were identified manually by visual inspection. The estimation of area per lipid is non-trivial. There are many methods one could use for this calculation, but the procedure we followed is enumerated below.

1. The radial distribution function (RDF) was calculated between the headgroups (the first carbon atoms) using MDTraj.[122] This allows us to extract the average distance between headgroups.
2. Using the RDF plot, we used NumPy[123] to fit a gaussian curve of the function

$$g(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right) \tag{5.7}$$

to the first peak, which corresponds to the distribution of nearest headgroups. We then extract the mean value (μ) as the radius and the standard deviation (σ) as the error associated with the calculation.

3. Using the radius *r* of the best fit sphere to the lipid cluster calculated *via* **Section 5.7.3**, we can estimate the average area covered by a lipid on a curved surface using the geometric principles of a spherical cap, illustrated in **Figure 5.10**, where the distance between two lipid headgroups is defined as 2*a*.

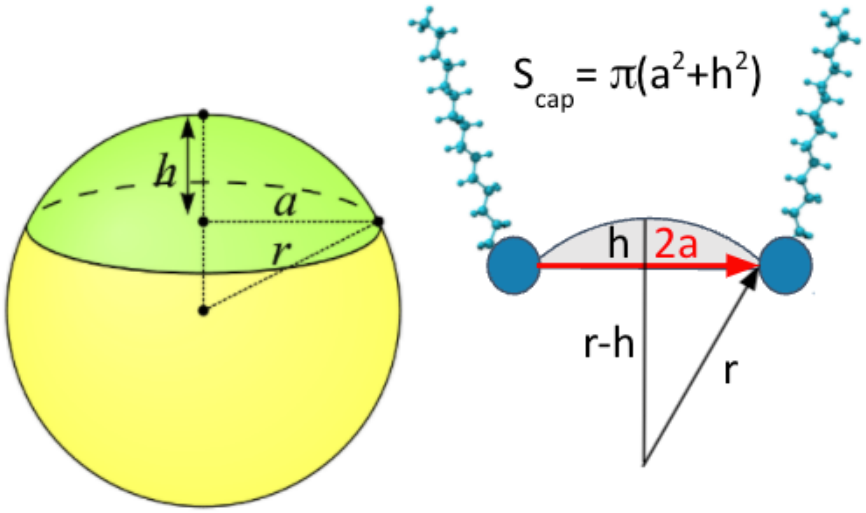


Figure 5.10: Diagram of spherical cap calculation.

4. Multiplying the resulting area by the number of lipids in the cluster and propagating error appropriately gives the estimate of the total area covered by the lipid cluster.

5.7.5 Estimation of Ellipsoidal Surface Area

One equation estimating the surface area of an ellipsoid is given by Knud Thomsen's Formula:

$$S \approx 4\pi \left(\frac{(ab)^p + (ac)^p + (bc)^p}{3} \right)^{1/p} \tag{6.8}$$

where the constant $p = 1.6075$ gives an approximate error ≤1.061%.[124–126] The estimated surface area can thus be calculated using the semi-axis values extracted from the best fit bounding ellipsoid calculation from **Section 5.7.1**.

5.7.6 Error Propagation

Error was propagated using the equations in **Table 5.3** for the analytical propagation of uncertainty.[127]

Type of Analysis	Example	Equation	Eq
Addition/Subtraction	$x = a + b - c$	$s_x = \sqrt{s_a^2 + s_b^2 + s_c^2}$	6.9
Multiplication/Division	$x = a \times b \div c$	$\frac{s_x}{x} = \sqrt{(\frac{s_a}{a})^2 + (\frac{s_b}{b})^2 + (\frac{s_c}{c})^2}$	6.10

References

(1) Haywood, J.; Boucher, O. Estimates of the Direct and Indirect Radiative Forcing Due to Tropospheric Aerosols: A Review. *Reviews of Geophysics* **2000**, *38* (4), 513–543. https://doi.org/10.1029/1999RG000078.

(2) Cochran, R. E.; Ryder, O. S.; Grassian, V. H.; Prather, K. A. Sea Spray Aerosol: The Chemical Link between the Oceans, Atmosphere, and Climate. *Accounts of Chemical Research* **2017**, *50* (3), 599–604. https://doi.org/10.1021/acs.accounts.6b00603.

(3) Patterson, J. P.; Collins, D. B.; Michaud, J. M.; Axson, J. L.; Sultana, C. M.; Moser, T.; Dommer, A. C.; Conner, J.; Grassian, V. H.; Stokes, M. D.; Deane, G. B.; Evans, J. E.; Burkart, M. D.; Prather, K. A.; Gianneschi, N. C. Sea Spray Aerosol Structure and Composition Using Cryogenic Transmission Electron Microscopy. *ACS Central Science* **2016**, acscentsci.5b00344-acscentsci.5b00344. https://doi.org/10.1021/acscentsci.5b00344.

(4) Malfatti, F.; Lee, C.; Tinta, T.; Pendergraft, M. A.; Celussi, M.; Zhou, Y.; Sultana, C. M.; Rotter, A.; Axson, J. L.; Collins, D. B.; Santander, M. V.; Anides Morales, A.; Aluwihare, L. I.; Riemer, N.; Grassian, V. H.; Azam, F.; Prather, K. A. Detection of Active Microbial Enzymes in Nascent Sea Spray Aerosol: Implications for Atmospheric Chemistry and Climate. *Environmental Science & Technology Letters* **2019**. https://doi.org/10.1021/acs.estlett.8b00699.

(5) Jayarathne, T.; Sultana, C. M.; Lee, C.; Malfatti, F.; Cox, J. L.; Pendergraft, M. A.; Moore, K. A.; Azam, F.; Tivanski, A. V.; Cappa, C. D.; Bertram, T. H.; Grassian, V. H.; Prather, K. A.; Stone, E. A. Enrichment of Saccharides and Divalent Cations in Sea Spray Aerosol during Two Phytoplankton Blooms. *Environmental Science and Technology* **2016**. https://doi.org/10.1021/acs.est.6b02988.

(6) Cochran, R. E.; Laskina, O.; Trueblood, J.; Estillore, A. D.; Morris, H. S.; Jayarathne, T.; Sultana, C. M.; Lee, C.; Lin, P.; Laskin, J.; Laskin, A.; Dowling, J.; Qin, Z.; Cappa, C. D.; Bertram, T. H.; Tivanski, A. V.; Stone, E. A.; Prather, K. A.; Grassian, V. H. Molecular Characterization of Sea Spray Particles : Influence of Ocean Biology on Particle Composition and Interaction with Water. *Chem* **2017**, *2*, 655–667. https://doi.org/10.1016/j.chempr.2017.03.007.

(7) Cochran, R. E.; Laskina, O.; Jayarathne, T.; Laskin, A.; Laskin, J.; Lin, P.; Sultana, C.; Lee, C.; Moore, K. A.; Cappa, C. D.; Bertram, T. H.; Prather, K. A.; Grassian, V. H.; Stone, E. A. Analysis of Organic Anionic Surfactants in Fine and Coarse Fractions of Freshly Emitted Sea Spray Aerosol. *Environmental Science and Technology* **2016**, *50* (5), 2477–2486. https://doi.org/10.1021/acs.est.5b04053.

(8) Ryder, O. S.; Campbell, N. R.; Morris, H.; Forestieri, S.; Ruppel, M. J.; Cappa, C.; Tivanski, A.; Prather, K.; Bertram, T. H. Role of Organic Coatings in Regulating N2O5 Reactive Uptake to Sea Spray Aerosol. *Journal of Physical Chemistry A* **2015**, *119* (48), 11683–11692. https://doi.org/10.1021/acs.jpca.5b08892.

(9) Zhao, Z.; Husainy, S.; Stoudemayer, C. T.; Smith, G. D. Reactive Uptake of NO3 Radicals by Unsaturated Fatty Acid Particles. *Physical Chemistry Chemical Physics* **2011**, *13* (39), 17809–17817. https://doi.org/10.1039/c1cp21790a.

(10) Moise, T.; Rudich, Y. Reactive Uptake of Ozone by Aerosol-Associated Unsaturated Fatty Acids: Kinetics, Mechanism, and Products. *Journal of Physical Chemistry A* **2002**, *106* (27), 6469–6476. https://doi.org/10.1021/jp025597e.

(11) Shrestha, M.; Luo, M.; Li, Y.; Xiang, B.; Xiong, W.; Grassian, V. H. Let There Be Light: Stability of Palmitic Acid Monolayers at the Air/Salt Water Interface in the Presence and Absence of Simulated Solar Light and a Photosensitizer. *Chemical Science* **2018**, *9* (26), 5716–5723. https://doi.org/10.1039/C8SC01957F.

(12) Lee, C.; Dommer, A. C.; Schiffer, J. M.; Amaro, R. E.; Grassian, V. H.; Prather, K. A. Cation-Driven Lipopolysaccharide Morphological Changes Impact Heterogeneous Reactions of Nitric Acid with Sea Spray Aerosol Particles. *The Journal of Physical Chemistry Letters* **2021**, *12* (20), 5023–5029. https://doi.org/10.1021/acs.jpclett.1c00810.

(13) Lee, H. D.; Ray, K. K.; Tivanski, A. V. Directly Probing the Phase States and Surface Tension of Individual Submicrometer Particles Using Atomic Force Microscopy. *ACS Symposium Series* **2018**, *1299*, 245–259. https://doi.org/10.1021/bk-2018-1299.ch012.

(14) Qiu, Y.; Molinero, V. Morphology of Liquid-Liquid Phase Separated Aerosols. *Journal of the American Chemical Society* **2015**, *137* (33), 10642–10651. https://doi.org/10.1021/jacs.5b05579.

(15) Huang, Y.; Mahrt, F.; Xu, S.; Shiraiwa, M.; Zuend, A.; Bertram, A. K. Coexistence of Three Liquid Phases in Individual Atmospheric Aerosol Particles. *PNAS* **2021**, *118* (16). https://doi.org/10.1073/pnas.2102512118.

(16) O'Brien, R. E.; Wang, B.; Kelly, S. T.; Lundt, N.; You, Y.; Bertram, A. K.; Leone, S. R.; Laskin, A.; Gilles, M. K. Liquid–Liquid Phase Separation in Aerosol Particles: Imaging at the Nanometer Scale. *Environ. Sci. Technol.* **2015**, *49* (8), 4995–5002. https://doi.org/10.1021/acs.est.5b00062.

(17) Karadima, K. S.; Mavrantzas, V. G.; Pandis, S. N. Molecular Dynamics Simulation of the Local Concentration and Structure in Multicomponent Aerosol Nanoparticles under Atmospheric Conditions. *Phys. Chem. Chem. Phys.* **2017**, *19* (25), 16681–16692. https://doi.org/10.1039/C7CP02036H.

(18) Karadima, K. S.; Mavrantzas, V. G.; Pandis, S. N. Insights into the Morphology of Multicomponent Organic and Inorganic Aerosols from Molecular Dynamics Simulations. *Atmospheric Chemistry and Physics* **2019**, *19* (8), 5571–5587. https://doi.org/10.5194/acp-19-5571-2019.

(19) Reid, J. P.; Dennis-Smither, B. J.; Kwamena, N.-O. A.; Miles, R. E. H.; Hanford, K. L.; Homer, C. J. The Morphology of Aerosol Particles Consisting of Hydrophobic and Hydrophilic Phases: Hydrocarbons, Alcohols and Fatty Acids as the Hydrophobic Component. *Phys. Chem. Chem. Phys.* **2011**, *13* (34), 15559–15572. https://doi.org/10.1039/C1CP21510H.

(20) Gorkowski, K.; Beydoun, H.; Aboff, M.; Walker, J. S.; Reid, J. P.; Sullivan, R. C. Advanced Aerosol Optical Tweezers Chamber Design to Facilitate Phase-Separation and Equilibration Timescale Experiments on Complex Droplets. *Aerosol Science and Technology* **2016**, *50* (12), 1327–1341. https://doi.org/10.1080/02786826.2016.1224317.

(21) Bertram, A. K.; Martin, S. T.; Hanna, S. J.; Smith, M. L.; Bodsworth, A.; Chen, Q.; Kuwata, M.; Liu, A.; You, Y.; Zorn, S. R. Predicting the Relative Humidities of Liquid-Liquid Phase Separation, Efflorescence, and Deliquescence of Mixed Particles of Ammonium Sulfate, Organic Material, and Water Using the Organic-to-Sulfate Mass Ratio of the Particle and the Oxygen-to-Carbon Elemental Ratio of the Organic Component. *Atmospheric Chemistry and Physics* **2011**, *11* (21), 10995–11006. https://doi.org/10.5194/acp-11-10995-2011.

(22) You, Y.; Renbaum-Wolff, L.; Carreras-Sospedra, M.; Hanna, S. J.; Hiranuma, N.; Kamal, S.; Smith, M. L.; Zhang, X.; Weber, R. J.; Shilling, J. E.; Dabdub, D.; Martin, S. T.; Bertram, A. K. Images Reveal That Atmospheric Particles Can Undergo Liquid-Liquid Phase Separations. *Proceedings of the National Academy of Sciences* **2012**, *109* (33), 13188–13193. https://doi.org/10.1073/pnas.1206414109.

(23) Renbaum-Wolff, L.; Song, M.; Marcolli, C.; Zhang, Y.; Liu, P. F.; Grayson, J. W.; Geiger, F. M.; Martin, S. T.; Bertram, A. K. Observations and Implications of Liquid–Liquid Phase Separation at Highrelative Humidities in Secondary Organic Material Produced by <I>A</I>-Pinene Ozonolysis without Inorganic Salts. *Atmos. Chem. Phys.* **2016**, *16* (12), 7969–7979. https://doi.org/10.5194/acp-16-7969-2016.

(24) Altaf, M. B.; Zuend, A.; Freedman, M. A. Role of Nucleation Mechanism on the Size Dependent Morphology of Organic Aerosol. *Chem. Commun.* **2016**, *52* (59), 9220–9223. https://doi.org/10.1039/C6CC03826C.

(25) Song, M.; Marcolli, C.; Krieger, U. K.; Lienhard, D. M.; Peter, T. Morphologies of Mixed Organic/Inorganic/Aqueous Aerosol Droplets. *Faraday Discuss.* **2013**, *165* (0), 289–316. https://doi.org/10.1039/C3FD00049D.

(26) Gorkowski, K.; Donahue, N. M.; Sullivan, R. C. Emulsified and Liquid–Liquid Phase-Separated States of α-Pinene Secondary Organic Aerosol Determined Using Aerosol Optical Tweezers. *Environ. Sci. Technol.* **2017**, *51* (21), 12154–12163. https://doi.org/10.1021/acs.est.7b03250.

(27) Lee, H. D.; Tivanski, A. V. Atomic Force Microscopy: An Emerging Tool in Measuring the Phase State and Surface Tension of Individual Aerosol Particles. *Annual Review of Physical Chemistry* **2021**, *72* (1), 235–252. https://doi.org/10.1146/annurev-physchem-090419-110133.

(28) Ergin, G.; Takahama, S. Carbon Density Is an Indicator of Mass Accommodation Coefficient of Water on Organic-Coated Water Surface. *J. Phys. Chem. A* **2016**, *120* (18), 2885–2893. https://doi.org/10.1021/acs.jpca.6b01748.

(29) Chakraborty, P.; Zachariah, M. R. Sticking Coefficient and Processing of Water Vapor on Organic-Coated Nanoaerosols. *J. Phys. Chem. A* **2008**, *112* (5), 966–972. https://doi.org/10.1021/jp076442f.

(30) Ma, X.; Chakraborty, P.; Henz, B. J.; Zachariah, M. R. Molecular Dynamic Simulation of Dicarboxylic Acid Coated Aqueous Aerosol: Structure and Processing of Water Vapor. **2011**, 11.

(31) Zhao, Z.; Kong, K.; Wang, S.; Zhou, Y.; Cheng, D.; Wang, W.; Zeng, X. C.; Li, H. Understanding Hygroscopic Nucleation of Sulfate Aerosols: Combination of Molecular Dynamics Simulation with Classical Nucleation Theory. *J. Phys. Chem. Lett.* **2019**, 7.

(32) Darvas, M.; Picaud, S. Water Adsorption around Oxalic Acid Aggregates: A Molecular Dynamics Simulation of Water Nucleation on Organic Aerosols. **2011**, 10.

(33) Li, X.; Hede, T.; Tu, Y.; Leck, C.; Ågren, H. Cloud Droplet Activation Mechanisms of Amino Acid Aerosol Particles: Insight from Molecular Dynamics Simulations. *Tellus B: Chemical and Physical Meteorology* **2013**, *65* (1), 20476. https://doi.org/10.3402/tellusb.v65i0.20476.

(34) Sun, L.; Hede, T.; Tu, Y.; Leck, C.; Ågren, H. Combined Effect of Glycine and Sea Salt on Aerosol Cloud Droplet Activation Predicted by Molecular Dynamics Simulations. *J. Phys. Chem. A* **2013**, *117* (41), 10746–10752. https://doi.org/10.1021/jp407538x.

(35) Karadima, K. S.; Mavrantzas, V. G.; Pandis, S. N.; Vega, C.; De Miguel, E.; Li, W.; Pak, C. Y.; Wang, X.; Tse, Y. L. S.; Ilčin, M.; Michalík, M.; Kováčiková, K.; Káziková, L.; Lukeš, V.; Phys, J. C.; Chakraborty, P.; Zachariah, M. R.; Takahama, S.; Russell, L. M. Uptake of Common Atmospheric Gases by Organic-Coated Water Droplets. *Journal of Physical Chemistry C* **2019**, *123* (31), 18924–18931. https://doi.org/10.1021/acs.jpcc.9b03252.

(36) Pak, C. Y.; Li, W.; Tse, Y.-L. S. Free Energy and Dynamics of Organic-Coated Water Droplet Coalescence. *J. Phys. Chem. C* **2020**, 9.

(37) Radola, B.; Picaud, S.; Vardanega, D. Molecular Dynamics Simulations of the Interaction between Water Molecules and Aggregates of Acetic or Propionic Acid Molecules. *J. Phys. Chem. B* **2015**, 13.

(38) Li, X.; Hede, T.; Tu, Y.; Leck, C.; Ågren, H. Surface-Active Cis-Pinonic Acid in Atmospheric Droplets: A Molecular Dynamics Study. *J. Phys. Chem. Lett.* **2010**, 5.

(39) Feng, Y.-J.; Huang, T.; Wang, C.; Liu, Y.-R.; Jiang, S.; Miao, S.-K.; Chen, J.; Huang, W. Π-Hydrogen Bonding of Aromatics on the Surface of Aerosols: Insights from Ab Initio and Molecular Dynamics Simulation. *J. Phys. Chem. B* **2016**, 7.

(40) Li, X.; Hede, T.; Tu, Y.; Leck, C. Glycine in Aerosol Water Droplets: A Critical Assessment of Ko¨hler Theory by Predicting Surface Tension from Molecular Dynamics Simulations. *Atmos. Chem. Phys.* **2011**, 9.

(41) Roose, A.; Toubin, C.; Dusanter, S.; Riffault, V.; Duflot, D. Classical Molecular Dynamics Study of Small-Chain Carboxylic Acid Aerosol Particles. *ACS Earth and Space Chemistry* **2019**, 10.

(42) Klauda, J. B.; Venable, R. M.; Freites, J. A.; O'Connor, J. W.; Tobias, D. J.; Mondragon-Ramirez, C.; Vorobyov, I.; MacKerell, A. D.; Pastor, R. W. Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. *The Journal of Physical Chemistry B* **2010**, *114* (23), 7830–7843. https://doi.org/10.1021/jp101759q.

(43) Huang, J.; Mackerell, A. D. CHARMM36 All-Atom Additive Protein Force Field: Validation Based on Comparison to NMR Data. *Journal of Computational Chemistry* **2013**, *34* (25), 2135–2145. https://doi.org/10.1002/jcc.23354.

(44) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D. CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nature methods* **2017**, *14* (1), 71–73. https://doi.org/10.1038/nmeth.4067.

(45) Bertram, T. H.; Cochran, R. E.; Grassian, V. H.; Stone, E. A. Sea Spray Aerosol Chemical Composition: Elemental and Molecular Mimics for Laboratory Studies of Heterogeneous and Multiphase Reactions. *Chemical Society Reviews* **2018**, *47* (7), 2374–2400. https://doi.org/10.1039/C7CS00008A.

(46) Maravić, A.; Skočibušić, M.; Šprung, M.; Šamanić, I.; Puizina, J.; Pavela-Vrančić, M. Occurrence and Antibiotic Susceptibility Profiles of Burkholderia Cepacia Complex in Coastal Marine Environment. *International Journal of Environmental Health Research* **2012**, *22* (6), 531–542. https://doi.org/10.1080/09603123.2012.667797.

(47) Sánchez, D. A.; Tonetto, G. M.; Ferreira, M. L. Burkholderia Cepacia Lipase: A Versatile Catalyst in Synthesis Reactions. *Biotechnology and Bioengineering* **2018**, *115* (1), 6–24. https://doi.org/10.1002/bit.26458.

(48) Schiffer, J. M.; Luo, M.; Dommer, A. C.; Thoron, G.; Pendergraft, M.; Santander, M. V.; Lucero, D.; Pecora De Barros, E.; Prather, K. A.; Grassian, V. H.; Amaro, R. E. Impacts of Lipase Enzyme on the Surface Properties of Marine Aerosols. *J. Phys. Chem. Lett* **2018**, *9*, 22–22. https://doi.org/10.1021/acs.jpclett.8b01363.

(49) Luo, M.; Dommer, A. C.; Schiffer, J. M.; Rez, D. J.; Mitchell, A. R.; Amaro, R. E.; Grassian, V. H. Surfactant Charge Modulates Structure and Stability of Lipase-Embedded Monolayers at Marine-Relevant Aerosol Surfaces. *Langmuir* **2019**, *35* (27), 9050–9060. https://doi.org/10.1021/acs.langmuir.9b00689.

(50) Trueblood, J. V.; Estillore, A. D.; Lee, C.; Dowling, J. A.; Prather, K. A.; Grassian, V. H. Heterogeneous Chemistry of Lipopolysaccharides with Gas-Phase Nitric Acid: Reactive Sites and Reaction Pathways. *Journal of Physical Chemistry A* **2016**, *120* (32), 6444–6450. https://doi.org/10.1021/acs.jpca.6b07023.

(51) Coughlin, R. T.; Tonsager, S.; McGroarty, E. J. Quantitation of Metal Cations Bound to Membranes and Extracted Lipopolysaccharide of Escherichia Coli. *Biochemistry* **1983**, *22* (8), 2002–2007. https://doi.org/10.1021/bi00277a041.

(52) Garidel, P.; Rappolt, M.; Schromm, A. B.; Howe, J.; Lohner, K.; Andrä, J.; Koch, M. H. J.; Brandenburg, K. Divalent Cations Affect Chain Mobility and Aggregate Structure of Lipopolysaccharide from Salmonella Minnesota Reflected in a Decrease of Its Biological Activity. *Biochimica et Biophysica Acta - Biomembranes* **2005**, *1715* (2), 122–131. https://doi.org/10.1016/j.bbamem.2005.07.013.

(53) Jeworrek, C.; Evers, F.; Howe, J.; Brandenburg, K.; Tolan, M.; Winter, R. Effects of Specific versus Nonspecific Ionic Interactions on the Structure and Lateral Organization of Lipopolysaccharides. *Biophysical Journal* **2011**, *100* (9), 2169–2177. https://doi.org/10.1016/j.bpj.2011.03.019.

(54) Li, L.; Luo, R. G. Use of Ca+2 to Re-Aggregate Lipopolysaccharide (LPS) in Hemoglobin Solutions and the Subsequent Removal of Endotoxin by Ultrafiltration. *Biotechnology Techniques* **1998**, *12* (2), 119–122. https://doi.org/10.1023/A:1008884332586.

(55) Schneck, E.; Schubert, T.; Konovalov, O. V.; Quinn, B. E.; Gutsmann, T.; Brandenburg, K.; Oliveira, R. G.; Pink, D. A.; Tanaka, M. Quantitative Determination of Ion Distributions in Bacterial Lipopolysaccharide Membranes by Grazing-Incidence X-Ray Fluorescence. *Proceedings of the National Academy of Sciences of the United States of America* **2010**, *107* (20), 9147–9151. https://doi.org/10.1073/pnas.0913737107.

(56) Snyder, S.; Kim, D.; McIntosh, T. J. Lipopolysaccharide Bilayer Structure: Effect of Chemotype, Core Mutations, Divalent Cations, and Temperature. *Biochemistry* **1999**, *38* (33), 10758–10767. https://doi.org/10.1021/bi990867d.

(57) Wang, X.; Deane, G. B.; Moore, K. A.; Ryder, O. S.; Stokes, M. D.; Beall, C. M.; Collins, D. B.; Santander, M. V.; Burrows, S. M.; Sultana, C. M.; Prather, K. A. The Role of Jet and Film Drops in Controlling the Mixing State of Submicron Sea Spray Aerosol Particles. *Proceedings of the National Academy of Sciences of the United States of America* **2017**, *114* (27), 6978–6983. https://doi.org/10.1073/pnas.1702420114.

(58) Russell, L. M.; Hawkins, L. N.; Frossard, A. A.; Quinn, P. K.; Bates, T. S. Carbohydrate-like Composition of Submicron Atmospheric Particles and Their Production from Ocean Bubble Bursting. *PNAS* **2010**, *107* (15), 6652–6657. https://doi.org/10.1073/pnas.0908905107.

(59) Richards, D. S.; Trobaugh, K. L.; Hajek-Herrera, J.; Price, C. L.; Sheldon, C. S.; Davies, J. F.; Davis, R. D. Ion-Molecule Interactions Enable Unexpected Phase Transitions in Organic-Inorganic Aerosol. *Science Advances* **2020**, *6* (47), eabb5643. https://doi.org/10.1126/sciadv.abb5643.

(60) Angle, K. J.; Crocker, D. R.; Simpson, R. M. C.; Mayer, K. J.; Garofalo, L. A.; Moore, A. N.; Garcia, S. L. M.; Or, V. W.; Srinivasan, S.; Farhan, M.; Sauer, J. S.; Lee, C.; Pothier, M. A.; Farmer, D. K.; Martz, T. R.; Bertram, T. H.; Cappa, C. D.; Prather, K. A.; Grassian, V. H. Acidity across the Interface from the Ocean Surface to Sea Spray Aerosol. *PNAS* **2021**, *118* (2). https://doi.org/10.1073/pnas.2018397118.

(61) Luo, M.; Wauer, N. A.; Angle, K. J.; Dommer, A. C.; Song, M.; Nowak, C. M.; Amaro, R. E.; Grassian, V. H. Insights into the Behavior of Nonanoic Acid and Its Conjugate Base at the Air/Water Interface through a Combined Experimental and Theoretical Approach. *Chemical Science* **2020**, *11* (39). https://doi.org/10.1039/d0sc02354j.

(62) Wellen, B. A.; Lach, E. A.; Allen, H. C. Surface PKa of Octanoic, Nonanoic, and Decanoic Fatty Acids at the Air-Water Interface: Applications to Atmospheric Aerosol Chemistry. *Physical Chemistry Chemical Physics* **2017**, *19* (39), 26551–26558. https://doi.org/10.1039/c7cp04527a.

(63) Budin, I.; Prwyes, N.; Zhang, N.; Szostak, J. W. Chain-Length Heterogeneity Allows for the Assembly of Fatty Acid Vesicles in Dilute Solutions. *Biophysical Journal* **2014**, *107* (7), 1582–1590. https://doi.org/10.1016/j.bpj.2014.07.067.

(64) Budin, I.; Debnath, A.; Szostak, J. W. Concentration-Driven Growth of Model Protocell Membranes. *Journal of the American Chemical Society* **2012**, *134* (51), 20812–20819. https://doi.org/10.1021/ja310382d.

(65) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM. *Journal of computational chemistry* **2008**, *29*, 2967–2970. https://doi.org/10.1002/jcc.

(66) Lee, J.; Cheng, X.; Swails, J. M.; Yeom, M. S.; Eastman, P. K.; Lemkul, J. A.; Wei, S.; Buckner, J.; Jeong, J. C.; Qi, Y.; Jo, S.; Pande, V. S.; Case, D. A.; Brooks, C. L.; MacKerell, A. D.; Klauda, J. B.; Im, W. CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *Journal of Chemical Theory and Computation* **2016**, *12* (1), 405–413. https://doi.org/10.1021/acs.jctc.5b00935.

(67) Park, S. J.; Lee, J.; Qi, Y.; Kern, N. R.; Lee, H. S.; Jo, S.; Joung, I.; Joo, K.; Lee, J.; Im, W. CHARMM-GUI Glycan Modeler for Modeling and Simulation of Carbohydrates and Glycoconjugates. *Glycobiology* **2019**. https://doi.org/10.1093/glycob/cwz003.

(68) Lee, J.; Patel, D. S.; Ståhle, J.; Park, S. J.; Kern, N. R.; Kim, S.; Lee, J.; Cheng, X.; Valvano, M. A.; Holst, O.; Knirel, Y. A.; Qi, Y.; Jo, S.; Klauda, J. B.; Widmalm, G.; Im, W. CHARMM-GUI Membrane Builder for Complex Biological Membrane Simulations with Glycolipids and Lipoglycans. *Journal of Chemical Theory and Computation* **2019**. https://doi.org/10.1021/acs.jctc.8b01066.

(69) Schrag, J. D.; Li, Y.; Cygler, M.; Lang, D.; Burgdorf, T.; Hecht, H.-J.; Schmid, R.; Schomburg, D.; Rydel, T. J.; Oliver, J. D.; Strickland, L. C.; Dunaway, C. M.; Larson, S. B.; Day, J.; McPherson, A. The Open Conformation of a Pseudomonas Lipase. *Structure* **1997**, *5* (2), 187–202. https://doi.org/10.1016/S0969-2126(97)00178-0.

(70) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *Journal of Computational Chemistry*. 2005. https://doi.org/10.1002/jcc.20289.

(71) Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. PACKMOL: A package for building initial configurations for molecular dynamics simulations. *Journal of Computational Chemistry* **2009**, *30* (13), 2157–2164. https://doi.org/10.1002/jcc.21224.

(72) Lu, J.; Qiu, Y.; Baron, R.; Molinero, V. Coarse-Graining of TIP4P/2005, TIP4P-Ew, SPC/E, and TIP3P to Monatomic Anisotropic Water Models Using Relative Entropy Minimization. *Journal of Chemical Theory and Computation* **2014**, *10* (9), 4104–4120. https://doi.org/10.1021/ct500487h.

(73) Vega, C.; Abascal, J. L. F. Simulating Water with Rigid Non-Polarizable Models: A General Perspective. *Physical Chemistry Chemical Physics* **2011**, *13* (44), 19663–19663. https://doi.org/10.1039/c1cp22168j.

(74) Humphery Dalke, A. S., K. ,. W. VMD - Visual Molecular Dynamics. *J. Molec. Graph* **1996**, *14*, 33–38.

(75) Phillips, J.; Zheng, G.; Kumar, S.; Kale, L. NAMD: Biomolecular Simulation on Thousands of Processors. In *Proceedings of the IEEE/ACM SC2002 Conference, Technical Paper 277*; IEEE Press: Baltimore, Maryland, 2002; pp 1–18.

(76) Phillips, J. C.; Hardy, D. J.; Maia, J. D. C.; Stone, J. E.; Ribeiro, J. V.; Bernardi, R. C.; Buch, R.; Fiorin, G.; Hénin, J.; Jiang, W.; McGreevy, R.; Melo, M. C. R.; Radak, B.; Skeel, R. D.; Singharoy, A.; Wang, Y.; Roux, B.; Aksimentiev, A.; Luthey-Schulten, Z.; Kalé, L. V.; Schulten, K.; Chipot, C.; Tajkhorshid, E. Scalable Molecular Dynamics on CPU and GPU Architectures with NAMD. *J. Chem.\ Phys.* **2020**, *153*, 044130. https://doi.org/10.1063/5.0014475.

(77) Stone, J. E.; Hynninen, A.-P.; Phillips, J. C.; Schulten, K. Early Experiences Porting the NAMD and VMD Molecular Simulation and Analysis Software to GPU-Accelerated OpenPOWER Platforms. *International Workshop on OpenPOWER for HPC (IWOPH'16)* **2016**, 188–206.

(78) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N·log(N) Method for Ewald Sums in Large Systems. *The Journal of Chemical Physics* **1993**, *98* (12), 10089–10092. https://doi.org/10.1063/1.464397.

(79) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *Journal of Computational Physics* **1977**, *23* (3), 327–341. https://doi.org/10.1016/0021-9991(77)90098-5.

(80) Theodorou, D. N.; Suter, U. W. Shape of Unperturbed Linear Polymers: Polypropylene. *Macromolecules* **1985**, *18* (6), 1206–1214. https://doi.org/10.1021/ma00148a028.

(81) Chen, I. A.; Walde, P. From Self-Assembled Vesicles to Protocells. *Cold Spring Harbor perspectives in biology* **2010**, *2* (7), 1–13. https://doi.org/10.1101/cshperspect.a002170.

(82) Douliez, J.-P.; Houssou, B. H.; Fameau, A.-L.; Navailles, L.; Nallet, F.; Grélard, A.; Dufourc, E. J.; Gaillard, C. Self-Assembly of Bilayer Vesicles Made of Saturated Long Chain Fatty Acids. *Langmuir* **2016**, *32* (2), 401–410. https://doi.org/10.1021/acs.langmuir.5b03627.

(83) Misuraca, L.; Caliò, A.; Grillo, I.; Grélard, A.; Oger, P. M.; Peters, J.; Demé, B. High-Temperature Behavior of Early Life Membrane Models. *Langmuir* **2020**, *36* (45), 13516–13526. https://doi.org/10.1021/acs.langmuir.0c02258.

(84) Gebicki, J. M.; Hicks, M. Ufasomes Are Stable Particles Surrounded by Unsaturated Fatty Acid Membranes. *Nature* **1973**, *243* (5404), 232–234. https://doi.org/10.1038/243232a0.

(85) Hargreaves, W. R.; Deamer, D. W. Liposomes from Ionic, Single-Chain Amphiphiles. *Biochemistry* **1978**, *17* (18), 3759–3768. https://doi.org/10.1021/bi00611a014.

(86) Cistola, D. P.; Hamilton, J. A.; Jackson, D.; Small, D. M. Ionization and Phase Behavior of Fatty Acids in Water: Application of the Gibbs Phase Rule. *Biochemistry* **1988**, *27* (6), 1881–1888. https://doi.org/10.1021/bi00406a013.

(87) Douliez, J.-P.; Gaillard, C. Self-Assembly of Fatty Acids: From Foams to Protocell Vesicles. *New J. Chem.* **2014**, *38* (11), 5142–5148. https://doi.org/10.1039/C4NJ00914B.

(88) Segré, D.; Ben-Eli, D.; Deamer, D. W.; Lancet, D. The Lipid World. *Orig Life Evol Biosph* **2001**, *31* (1), 119–145. https://doi.org/10.1023/A:1006746807104.

(89) Motomura, K.; Yamanaka, M.; Aratono, M. Thermodynamic Consideration of the Mixed Micelle of Surfactants. *Colloid & Polymer Sci* **1984**, *262* (12), 948–955. https://doi.org/10.1007/BF01490027.

(90) Aratono, M.; Villeneuve, M.; Takiue, T.; Ikeda, N.; Iyota, H. Thermodynamic Consideration of Mixtures of Surfactants in Adsorbed Films and Micelles. *Journal of Colloid and Interface Science* **1998**, *200* (1), 161–171. https://doi.org/10.1006/jcis.1997.5369.

(91) Mael, L. E.; Busse, H. L.; Peiker, G.; Grassian, V. H. Low-Temperature Water Uptake of Individual Marine and Biologically Relevant Atmospheric Particles Using Micro-Raman Spectroscopy. *J Phys Chem A* **2021**, *125* (44), 9691–9699. https://doi.org/10.1021/acs.jpca.1c08037.

(92) Farmer, D. K.; Cappa, C. D.; Kreidenweis, S. M. Atmospheric Processes and Their Controlling Influence on Cloud Condensation Nuclei Activity. *Chemical Reviews* **2015**, *115* (10), 4199–4217. https://doi.org/10.1021/cr5006292.

(93) Brink, H. M. T. Reactive Uptake of HNO3 and H2SO4 in Sea-Salt (NaCl) Particles. *Journal of Aerosol Science* **1998**, *29* (1–2), 57–64. https://doi.org/10.1016/S0021-8502(97)00460-6.

(94) DeMott, P. J.; Mason, R. H.; McCluskey, C. S.; Hill, T. C. J.; Perkins, R. J.; Desyaterik, Y.; Bertram, A. K.; Trueblood, J. V.; Grassian, V. H.; Qiu, Y.; Molinero, V.; Tobo, Y.; Sultana, C. M.; Lee, C.; Prather, K. A. Ice Nucleation by Particles Containing Long-Chain Fatty Acids of Relevance to Freezing by Sea Spray Aerosols. *Environmental Science: Processes and Impacts* **2018**, *20* (11), 1559–1569. https://doi.org/10.1039/c8em00386f.

(95) Zhang, X.-X.; Chen, M.; Fu, M. Impact of Surface Nanostructure on Ice Nucleation. *The Journal of Chemical Physics* **2014**, *141* (12), 124709–124709. https://doi.org/10.1063/1.4896149.

(96) Archer, R. J.; Mer, V. K. L. The Rate of Evaporation of Water through Fatty Acid Monolayers. **1955**, *59*, 9.

(97) Heikkila, R. E.; Kwong, C. N.; Cornwell, D. G. Stability of Fatty Acid Monolayers and the Relationship between Equilibrium Spreading Pressure, Phase Transformations, and Polymorphic Crystal Forms. *Journal of Lipid Research* **1970**, *11* (3), 190–194. https://doi.org/10.1016/s0022-2275(20)42982-7.

(98) Li, S.; Du, L.; Wei, Z.; Wang, W. Aqueous-Phase Aerosols on the Air-Water Interface: Response of Fatty Acid Langmuir Monolayers to Atmospheric Inorganic Ions. *Science of the Total Environment* **2017**, *580*, 1155–1161. https://doi.org/10.1016/j.scitotenv.2016.12.072.

(99) Takahama, S.; Russell, L. M. A Molecular Dynamics Study of Water Mass Accommodation on Condensed Phase Water Coated by Fatty Acid Monolayers. *Journal of Geophysical Research: Atmospheres* **2011**, *116* (D2). https://doi.org/10.1029/2010JD014842.

(100) Henry, D. J.; Dewan, V. I.; Prime, E. L.; Qiao, G. G.; Solomon, D. H.; Yarovsky, I. Monolayer Structure and Evaporation Resistance: A Molecular Dynamics Study of Octadecanol on Water. *Journal of Physical Chemistry B* **2010**, *114* (11), 3869–3878. https://doi.org/10.1021/jp909544a.

(101) Carter-Fenk, K. A.; Dommer, A. C.; Fiamingo, M. E.; Kim, J.; Amaro, R. E.; Allen, H. C. Calcium Bridging Drives Polysaccharide Co-Adsorption to a Proxy Sea Surface Microlayer. *Phys. Chem. Chem. Phys.* **2021**, *23* (30), 16401–16416. https://doi.org/10.1039/D1CP01407B.

(102) Brzozowska, A. M.; Duits, M. H. G.; Mugele, F. Stability of Stearic Acid Monolayers on Artificial Sea Water. *Colloids and Surfaces A: Physicochemical and Engineering Aspects* **2012**, *407*, 38–48. https://doi.org/10.1016/j.colsurfa.2012.04.055.

(103) La Mer, V. K.; Healy, T. W.; Aylmore, L. A. G. The Transport of Water through Monolayers of Long-Chain n-Paraffinic Alcohols. *Journal of Colloid Science* **1964**, *19*, 673–684.

(104) Fomina, E. S.; Vysotsky, Y. B.; Belyaeva, E. A.; Vollhardt, D.; Fainerman, V. B.; Miller, R. On Hexagonal Orientation of Fatty Alcohols in Monolayers at the Air/Water Interface: Quantum-Chemical Approach. *Journal of Physical Chemistry C* **2014**, *118* (8), 4122–4130. https://doi.org/10.1021/jp409911a.

(105) Vysotsky, Y. B.; Fomina, E. S.; Belyaeva, E. A.; Fainerman, V. B.; Vollhardt, D. On the Inclusion of Alkanes into the Monolayer of Aliphatic Alcohols at the Water/Alkane Vapor Interface: A Quantum Chemical Approach. *Physical Chemistry Chemical Physics* **2013**, *15* (6), 2159–2176. https://doi.org/10.1039/c2cp43713a.

(106) Kartashynska, E. S.; Vysotsky, Y. B.; Vollhardt, D.; Fainerman, V. B.; Zakharov, A. Y. Theoretical Description of Mixed Film Formation at the Air/Water Interface: Carboxylic Acids-Alcohols. *ACS Omega* **2018**, *3* (12), 16693–16705. https://doi.org/10.1021/acsomega.8b02583.

(107) Stephenson, B. C.; Goldsipe, A.; Beers, K. J.; Blankschtein, D. Quantifying the Hydrophobic Effect. 1. A Computer Simulation–Molecular-Thermodynamic Model for the Self-Assembly of Hydrophobic and Amphiphilic Solutes in Aqueous Solution. *J. Phys. Chem. B* **2007**, *111* (5), 1025–1044. https://doi.org/10.1021/jp065696i.

(108) Stephenson, B. C.; Goldsipe, A.; Beers, K. J.; Blankschtein, D. Quantifying the Hydrophobic Effect. 2. A Computer Simulation–Molecular-Thermodynamic Model for the Micellization of Nonionic Surfactants in Aqueous Solution. *J. Phys. Chem. B* **2007**, *111* (5), 1045–1062. https://doi.org/10.1021/jp065697a.

(109) Barbe, S.; Lafaquière, V.; Guieysse, D.; Monsan, P.; Remaud-Siméon, M.; André, I. Insights into Lid Movements of Burkholderia Cepacia Lipase Inferred from Molecular Dynamics Simulations. *Proteins: Structure, Function, and Bioinformatics* **2009**, *77* (3), 509–523. https://doi.org/10.1002/prot.22462.

(110) Pires De Oliveira, I.; Jara, G. E. Molecular Mechanism of Activation of Burkholderia Cepacia Lipase at Aqueous-Organic Interfaces. *Phys. Chem. Chem. Phys* **2017**, *19*, 31499–31499. https://doi.org/10.1039/c7cp04466f.

(111) Coughlin, R. T.; Haug, A.; McGroarty, E. J. Physical Properties of Defined Lipopolysaccharide Salts. *Biochemistry* **1983**, *22* (1978), 2007–2013.

(112) Wu, E. L.; Engström, O.; Jo, S.; Stuhlsatz, D.; Yeom, M. S.; Klauda, J. B.; Widmalm, G.; Im, W. Molecular Dynamics and NMR Spectroscopy Studies of E. Coli Lipopolysaccharide Structure and Dynamics. *Biophysical Journal* **2013**, *105* (6), 1444–1455. https://doi.org/10.1016/j.bpj.2013.08.002.

(113) Evans, M.; Ratcliffe, I.; Williams, P. A. Emulsion Stabilisation Using Polysaccharide–Protein Complexes. *Current Opinion in Colloid & Interface Science* **2013**, *18* (4), 272–282. https://doi.org/10.1016/j.cocis.2013.04.004.

(114) Dickinson, E. Hydrocolloids as Emulsifiers and Emulsion Stabilizers. *Food Hydrocolloids* **2009**, *23* (6), 1473–1482. https://doi.org/10.1016/j.foodhyd.2008.08.005.

(115) Siew, C. K.; Williams, P. A.; Cui, S. W.; Wang, Q. Characterization of the Surface-Active Components of Sugar Beet Pectin and the Hydrodynamic Thickness of the Adsorbed Pectin Layer. *J. Agric. Food Chem.* **2008**, *56* (17), 8111–8120. https://doi.org/10.1021/jf801588a.

(116) Schmitt, C.; Aberkane, L.; Sanchez, C. 16 - Protein–Polysaccharide Complexes and Coacervates. In *Handbook of Hydrocolloids (Second Edition)*; Phillips, G. O., Williams, P. A., Eds.; Woodhead Publishing Series in Food Science, Technology and Nutrition; Woodhead Publishing, 2009; pp 420–476. https://doi.org/10.1533/9781845695873.420.

(117) Cooper, C. L.; Dubin, P. L.; Kayitmazer, A. B.; Turksen, S. Polyelectrolyte–Protein Complexes. *Current Opinion in Colloid & Interface Science* **2005**, *10* (1), 52–78. https://doi.org/10.1016/j.cocis.2005.05.007.

(118) Matalanis, A.; Jones, O. G.; McClements, D. J. Structured Biopolymer-Based Delivery Systems for Encapsulation, Protection, and Release of Lipophilic Compounds. *Food Hydrocolloids* **2011**, *25* (8), 1865–1880. https://doi.org/10.1016/j.foodhyd.2011.04.014.

(119) Vinayahan, T.; Williams, P. A.; Phillips, G. O. Electrostatic Interaction and Complex Formation between Gum Arabic and Bovine Serum Albumin. *Biomacromolecules* **2010**, *11* (12), 3367–3374. https://doi.org/10.1021/bm100486p.

(120) Costa, C.; Medronho, B.; Filipe, A.; Mira, I.; Lindman, B.; Edlund, H.; Norgren, M. Emulsion Formation and Stabilization by Biomolecules: The Leading Role of Cellulose. *Polymers (Basel)* **2019**, *11* (10), 1570. https://doi.org/10.3390/polym11101570.

(121) Piacentini, E.; Mazzei, R.; Giorno, L. Comparison between Lipase Performance Distributed at the O/W Interface by Membrane Emulsification and by Mechanical Stirring. *Membranes* **2021**, *11* (2), 137. https://doi.org/10.3390/membranes11020137.

(122) McGibbon, R. T.; Beauchamp, K. A.; Harrigan, M. P.; Klein, C.; Swails, J. M.; Hernández, C. X.; Schwantes, C. R.; Wang, L.-P.; Lane, T. J.; Pande, V. S. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophysical Journal* **2015**, *109* (8), 1528–1532. https://doi.org/10.1016/j.bpj.2015.08.015.

(123) Harris, C. R.; Millman, K. J.; van der Walt, S. J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N. J.; Kern, R.; Picus, M.; Hoyer, S.; van Kerkwijk, M. H.; Brett, M.; Haldane, A.; del Río, J. F.; Wiebe, M.; Peterson, P.; Gérard-Marchant, P.; Sheppard, K.; Reddy, T.; Weckesser, W.; Abbasi, H.; Gohlke, C.; Oliphant, T. E. Array Programming with NumPy. *Nature* **2020**, *585* (7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2.

(124) Klamkin, M. S. Elementary Approximations to the Area of N-Dimensional Ellipsoids. *The American Mathematical Monthly* **1971**, *78* (3), 280–283. https://doi.org/10.2307/2317530.

(125) Dunkl, C. F.; Ramirez, D. E. Computing Hyperelliptic Integrals for Surface Measure of Ellipsoids. *ACM Trans. Math. Softw.* **1994**, *20* (4), 413–426. https://doi.org/10.1145/198429.198430.

(126) Dunkl, C. F.; Ramirez, D. E. Algorithm 736: Hyperelliptic Integrals and the Surface Measure of Ellipsoids. *ACM Trans. Math. Softw.* **1994**, *20* (4), 427–435. https://doi.org/10.1145/198429.198431.

(127) Douglas A. Skoog; F. James Holler; Stanley R. Crouch. *Principles of Instrumental Analysis*, 6th ed.; Thomson Brooks/Cole, 2007.

Part III
Special Applications to COVID-19 Research

Chapter 6

AI-Driven Multiscale Simulations Illuminate Mechanisms of SARS-CoV-2 Spike Dynamics

6.1 Abstract

We develop a generalizable AI-driven workflow that leverages heterogeneous HPC resources to explore the time-dependent dynamics of molecular systems. We use this workflow to investigate the mechanisms of infectivity of the SARS-CoV-2 spike protein, the main viral infection machinery. Our workflow enables more efficient investigation of spike dynamics in a variety of complex environments, including within a complete SARS-CoV-2 viral envelope simulation, which contains 305 million atoms and shows strong scaling on ORNL Summit using NAMD. We present several novel scientific discoveries, including the elucidation of the spike's full glycan shield, the role of spike glycans in modulating the infectivity of the virus, and the characterization of the flexible interactions between the spike and the human ACE2 receptor. We also demonstrate how AI can accelerate conformational sampling across different systems and pave the way for the future application of such methods to additional studies in SARS-CoV-2 and other molecular systems.

6.2 Justification

We:

- develop an AI-driven multiscale simulation framework to interrogate SARS-CoV-2 spike dynamics,
- reveal the spike's full glycan shield and discover that glycans play an active role in infection, and
- achieve new high watermarks for classical MD simulation of viruses (305 million atoms) and the weighted ensemble method (600,000 atoms).

6.3 Performance Attributes

Table 6.1: Performance attributes.

Performance Attribute	Our Submission
Category of achievement	Scalability, Time-to-solution
Type of method used	Explicit, Deep Learning
Results reported on the basis of	Whole application including I/O
Precision reported	Mixed Precision
System scale	Measured on full system
Measurement Mechanism	Hardware performance counters, Application timers, Performance Modeling

6.4 Overview of the Problem

The SARS-CoV-2 virus is the causative agent of COVID19, a world- wide pandemic that has infected over 35 million people and killed over one million. As such it is the subject of intense scientific investigations. Researchers are interested in understanding the structure and function of the proteins that constitute the virus, as this knowledge aids in the understanding of transmission, infectivity, and potential therapeutics.

A number of experimental methods, including x-ray crystallography, cryoelectron (cryo-EM) microscopy, and cryo-EM tomography are able to inform on the structure of viral proteins and the other (e.g., host cell) proteins with which the virus interacts. Such structural information is vital to our understanding of these molecular machines, however, there are limits to what experiments can tell us. For example, achieving high resolution structures typically comes at the expense of dynamics: flexible parts of the proteins (e.g., loops) are often not resolved, or frequently not even included in the experimental construct. Glycans, the sugar-like structures that decorate viral surface proteins, are particularly flexible and thus experimental techniques are currently unable to provide detailed views into their structure and function beyond a few basic units. Additionally, these experiments can resolve static snapshots, perhaps catching different states of the protein, but they are unable to elucidate the thermodynamic and kinetic relationships between such states.

In addition to the rich structural datasets, researchers have used a variety of proteomic, glycomic, and other methods to determine detailed information about particular aspects of the virus. In one example, deep sequencing methods have informed on the functional implications of mutations in a key part of the viral spike

protein.[1] In others, mass spectrometry approaches have provided information about the particular composition of the glycans at particular sites on the viral protein.[2,3] These data are each valuable in their own right but exist as disparate islands of knowledge. Thus there is a need to integrate these datasets into cohesive models, such that the fluctuations of the viral particle and its components that cause its infectivity can be understood.

In this work, we used all-atom molecular dynamics (MD) simulations to combine, augment, and extend available experimental datasets in order to interrogate the structure, dynamics, and function of the SARS-CoV-2 spike protein (**Figure 6.1**). The spike protein is considered the main infection machinery of the virus because it is the only glycoprotein on the surface of the virus and it is the molecular machine that interacts with the human host cell receptor, ACE2, at the initial step of infection. We have developed MD simulations of the spike protein at three distinct scales, where each system (and scale) is informative, extensive, and scientifically valuable in its own right (as will be discussed). This includes the construction and simulation of the SARS-CoV-2 viral envelope that contains 305 million atoms and is thus among one of the largest and most complex biological systems ever simulated (**Figure 6.1A**). We employ both conventional MD as well as the weighted ensemble enhanced sampling approach (which again breaks new ground in terms of applicable system size). We then collectively couple these break-through simulations with artificial intelligence (AI) based methods as part of an integrated workflow that transfers knowledge gained at one scale to 'drive' (enhance) sampling at another.

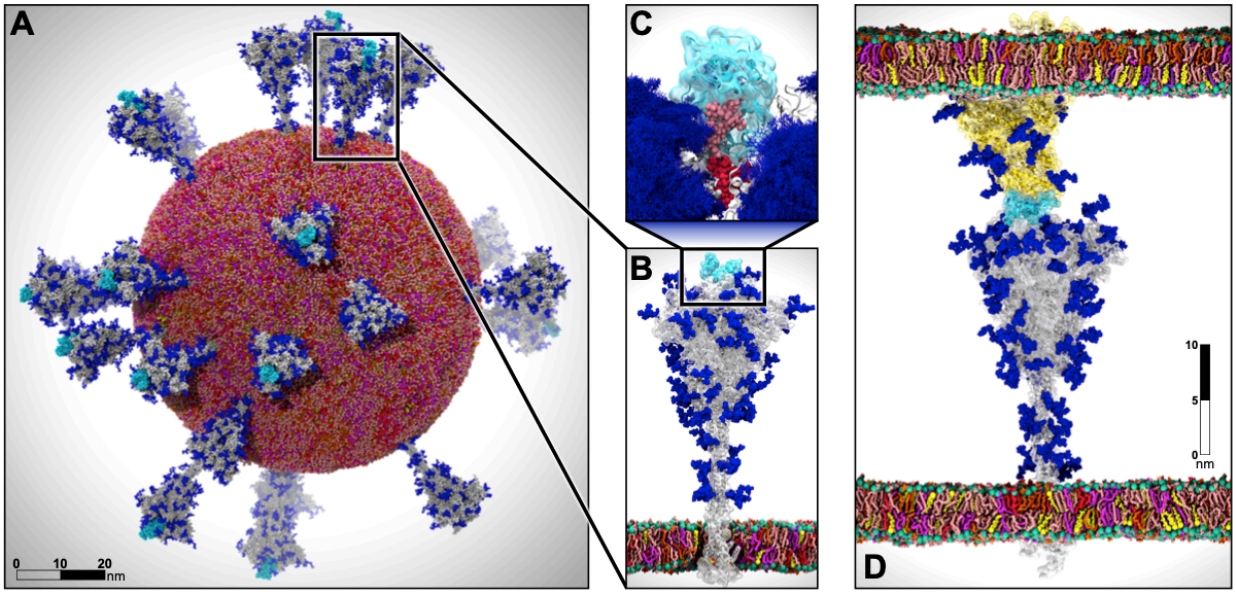


Figure 6.1: Multiscale modeling of SARS-CoV-2. A) All-atom model of the SARS-CoV-2 viral envelope (305 M atoms), including 24 spike proteins (colored in gray) in both the open (16) and closed states (8). The RBDs in the "up" state are highlighted in cyan) N-/O-Glycans are shown in blue. Water molecules and ions have been omitted for clarity. B) Full-length model of the glycosylated SARS-CoV-2 spike protein (gray surface) embedded into an ERGIC-like lipid bilayer (1.7 M atoms). RBD in the "up" state is highlighted in cyan. C) The glycan shield is shown by overlaying multiple conformations for each glycan collected at subsequent timesteps along the dynamics (blue bush-like representation). Highlighted in pink and red are two N-glycans (linked to N165 and N234, respectively) responsible for the modulation of the RBD dynamics, thus priming the virus for infection. The RBD "up" is depicted with a cyan surface. D) Two-parallel-membrane system of the spike-ACE2 complex (8.5 M atoms). The spike protein, embedded into an ERGIC-like membrane, is depicted with a gray transparent surface, whereas ACE2 is shown with a yellow transparent surface and it is embedded into a lipid bilayer mimicking the composition of mammalian cell membranes. Glycans are shown in blue, whereas water has been omitted for clarity. Visualizations were created in VMD using its custom GPU-accelerated ray tracing engine.

An additional significant challenge faced in bringing this work to fruition is that it pushes the boundaries of several fields simultaneously, including biology, physics, chemistry, mathematics, and computer science. It is intersectional in nature and requires the collective work of and effective communication among experts in each of these fields to construct, simulate, and analyze such systems - all while optimizing code performance to accelerate scientific discovery against SARS-CoV-2.

Our work has brought HPC to bear to provide unprecedented detail and atomic-level understanding of virus particles and how they infect human cells. Our efforts shed light on many aspects of the spike dynamics and function that are currently inaccessible with experiment and have provided a number of experimentally testable hypotheses - some of which have already been experimentally validated. By doing so, we provide new

understandings for vaccine and therapeutic development, inform on basic mechanisms of viral infection, push technological and methodological limits for molecular simulation, and bring supercomputing to the forefront in the fight against COVID19.

6.4.1 Methods

6.4.1.1 Full-length, fully glycosylated spike protein

In this work, we built two full-length glycosylated all-atom models of the SARS-CoV-2 S protein in both closed and open states, fully detailed in Casalino et al.[4] The two all-atom models were built starting from the cryo-EM structures of the spike in the open state (PDB ID: 6VSB[5]) where one receptor binding domain (RBD) is in the "up" conformation, and in the closed state, bearing instead three RBDs in the "down" conformation (PDB ID: 6VXX[6]). Given that the experimental cryo-EM structures were incomplete, the remaining parts, namely (i) the missing loops within the head (residues 16–1141), (ii) the stalk (residues 1141–1234) and (iii) the cytosolic tail (residues 1235–1273), were modelled using MODELLER[7] and I- TASSER.[8] The resulting full-length all-atom constructs were subsequently N-/O-glycosylated using the Glycan Reader & Modeler tool[9] integrated into Glycan Reader[10] in CHARMM-GUI.[11] Importantly, an asymmetric glycoprofile was generated (e.g., not specular across monomers) taking into account the N-/O-glycans heterogeneity as described in the available glycoanalytic data.[2,3] The two glycosylated systems were embedded into their physiological environment composed of an ERGIC-like lipid bilayer[12,13] built using CHARMM-GUI[14], explicit TIP3P water molecules[15], and neutralizing chloride and sodium ions at 150 mM concentration, generating two final systems each tallying 1.7 million atoms. Using CHARMM36 all-atom additive force fields[16,17] and NAMD 2.14[18], the systems were initially relaxed through a series of minimization, melting (for the membrane), and equilibration cycles. The equilibrated systems were then subjected to multiple replicas of all-atom MD simulation production runs of the open (6x) and closed (3x) systems on the NSF Frontera computing system at the Texas Advanced Computing Center (TACC). A cumulative extensive sampling of 4.2 and 1.7 μs was attained for the open and closed systems, respectively. Additionally, a third, mutant system bearing N165A and N234A mutations was built from the open system in order to delete the N-linked glycans and delineate their structural role in the RBD dynamics. This system was also simulated for 4.2 μs in 6 replicas.[4]

6.4.1.2 ACE2-RBD complex MD simulations

The model of the ACE2- RBD complex was based on cryo-EM structure trapping ACE2 as a homo-dimer co-complexed with two RBDs and B0AT1 trans- porter (PDB ID 6M17[19]). Upon removal of B0AT1, ACE2 missing residues at the C terminal end were modeled using I-TASSER, whereas those missing at the N terminal end were taken from 6M0J and properly positioned upon alignment of the N terminal helix. Zinc sites including the ions and the coordinating residues were copied from 1R42. The construct was fully N-/O-glycosylated using CHARMM-GUI tools for glycan modeling, reproducing the glycan heterogeneity for ACE2 and RBD reported in the available glycoanalytic data.[2,20,21] Similarly, the apo ACE2 homo-dimer was also built upon removal of the RBDs from the holo construct. The glycosylated models were embedded into separate lipid patches with a composition mimicking that of mammalian cellular membranes and simulated in explicit water molecules at 150 mM ion concentration, affording two final systems of 800,000 atoms each. MD simulations were performed using CHARMM36 all-atom additive force along with NAMD 2.14. The MD protocol was identical to that adopted for the simulation of the full-length spike and it is fully described in Casalino et al.[4] This work is fully detailed in Barros et al.[22]

6.4.1.3 Weighted ensemble simulations of spike opening

The spike must undergo a large conformational change for activation and binding to ACE2 receptors, where the receptor binding domain transitions from the “down’,’ or closed state to the “up,” or open state. [5] Such conformational changes occur on biological timescales generally not accessible by classical molecular dynamics simulations.[23] To simulate the full unbiased path at atomic resolution, we used the weighted ensemble (WE) enhanced sampling method.[24,25] Instead of running one single long simulation, the WE method runs many short simulations in parallel along the chosen reaction coordinates. The trajectories that rarely sample high energy regions are replicated, while the trajectories that frequently sample low energy regions are merged, which makes sampling rare events computationally tractable and gives enhanced sampling. The trajectories also carry probabilities or weights, which are continuously updated, and there is no statistical bias added to the system. Hence, we are able to directly obtain both thermodynamic and kinetic properties from the WE simulations.[26]

For this study, the closed model of the glycosylated spike from Casalino et al. was used as the initial structure by only keeping the head domain. The WE simulations were run using the highly scalable WESTPA software[27], with the Amber GPU accelerated molecular dynamics engine[28,29] version 18. Chamber[30] was used to convert CHARMM36 force fields and parameters from the system developed by Casalino et al. into an Amber readable format. A TIP3P[15] water box with at least 10 Å between protein and box edges was used with 150 mM NaCl, leading the total number of atoms to 548,881. Amber minimization was carried out in two stages. First the solvent was minimized for 10,000 cycles with sugars and proteins restrained with a weight of 100 kcal/mol $Å^2$, followed by unrestrained minimization for 100,000 cycles. Next the system was incrementally heated to 300 K over 300 ps. Equilibration and production were carried out in 2 fs timesteps with SHAKE[31] constraints on non-polar hydrogens and NPT ensemble. Pressure and temperature were controlled with Monte Carlo barostat and Langevin thermostat with 1 ps-1 collision frequency. The particle-mesh Ewald (PME) method was used with 10 Å cutoff for non-bonded interactions. The system was first equilibrated for 21 ns of conventional MD. The RMSD of the alpha carbons began to level off around 16 ns, and 24 structures were taken at regular intervals between 16 and 21 ns to use as equally weighted basis states for the WE simulation.

For each WE, tau was set to 100 ps of MD production followed by progress coordinate evaluation, and splitting / merging of walkers and updating weights, with a maximum of 8 walkers per bin. A two dimensional progress coordinate was defined by (i): the distance between the center of mass (COM) of the alpha carbons in the structured region of the spike helical core, and the alpha carbons in the four main beta sheets of the RBD (refers to RBD from chain A unless otherwise specified) and (ii): the RMSD of the alpha carbons in the four main beta sheets of the RBD to the initial structure (obtained from 1 ns equilibration). This simulation was run for 8.77 days on 80 P100 GPUs on Comet at SDSC collecting a comprehensive sampling of 7.5 μs, with bin spacing continuously monitored and adjusted to maximize sampling.

After extensive sampling of the RBD closed state, the second progress coordinate was changed to the RMSD of the alpha carbons in the four main beta sheets of the RBD compared to the final open structure, obtained from system 1, after 1 ns of equilibration carried out with identical methods as the closed structure described above, which was initially calculated as 11.5 Å. This allowed more efficient sampling of the transition to the open state by focusing sampling on states which are closer in rotational or translational space to the final

state, rather than sampling all conformations that are distinctly different from the closed state. Bin spacing was continuously monitored and adjusted to maximize traversing the RMSD coordinate. The full transition was confirmed when the RMSD coordinate reached below 6 Å and the RBD COM coordinate reached above 8.5 Å (**Figure 6.2**). The simulation was stopped for analysis after 1099 iterations, upon running for 26.74 days on 100 V100 GPUs on Longhorn at TACC and harvesting 70.0 μs.

A second, independent WE simulation was conducted to determine if the findings of the initial simulation were reproducible, and to use the information on the free energy landscape of the successful transition in the first WE to inform bin spacing and target state definition to run an unsupervised simulation. After 19.64 days on 100 V100 GPUs on TACC Longhorn and 51.5 μs of comprehensive sampling, successful transitions to the open state were observed, as well as further open states, in which the RBD was observed to be peeling off of the spike core.

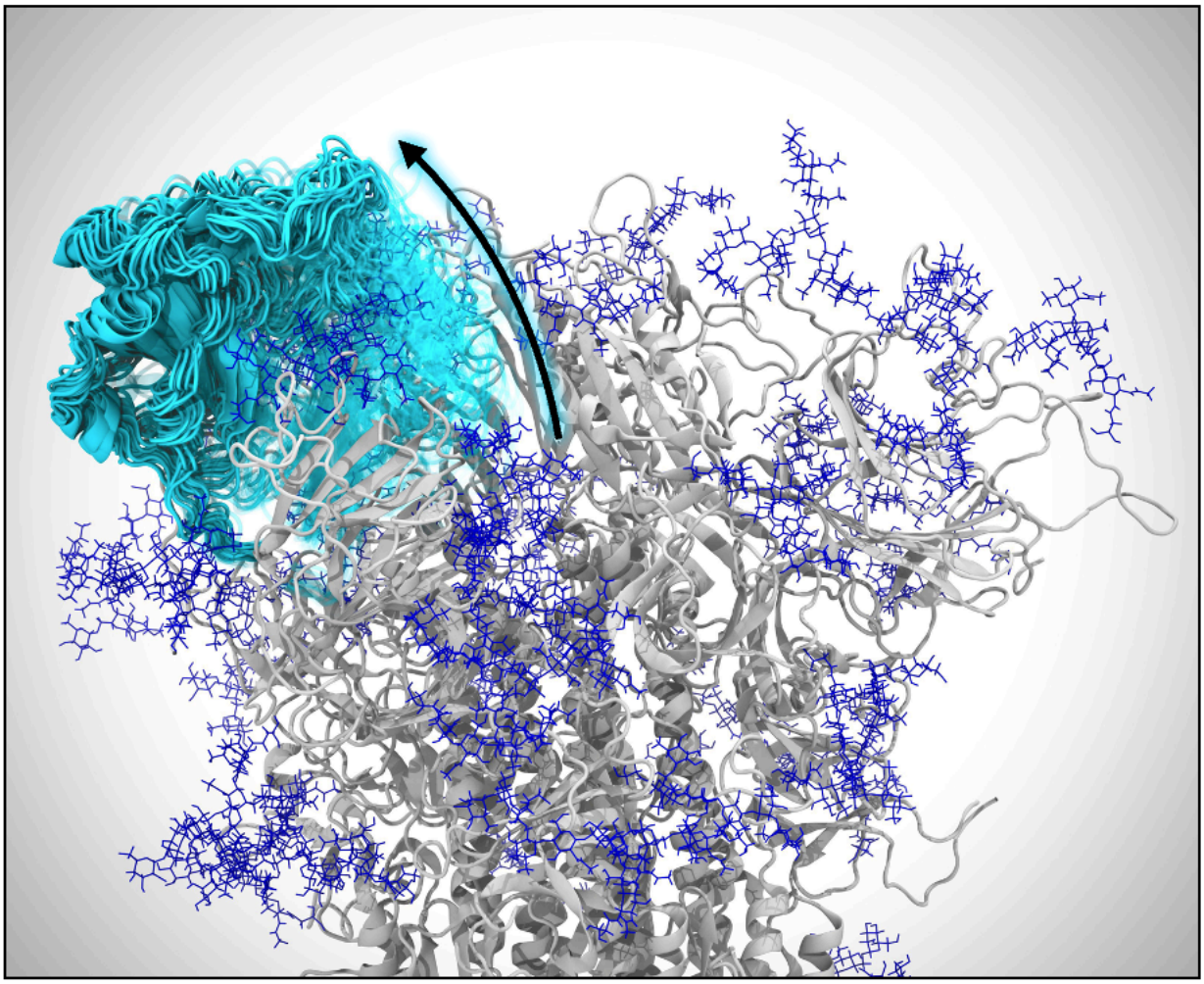


Figure 6.2: Opening of the spike protein. VMD visualization of weighted ensemble simulations shows the transition of the spike's RBD from the closed state to the open state. Many conformations of the RBD along its opening pathway are represented at the same time using cyan cartoons and a transparency gradient. Glycans appear as dark blue.

6.4.1.4 Two-parallel-membrane system of the spike-ACE2 complex

The SARS-CoV-2 virus gains entry into the host cell through a membrane fusion process taking place upon the recognition of the ACE2 receptors exposed on the host cell. This binding event triggers several, dramatic conformational changes within the spike protein, which becomes primed to pull the two membranes together for fusion, allowing the virus to pour the viral RNA into the host cell. In order to disentangle the mechanistic intricacies underlying this key process, we exploited the wealth of information obtained from the individual simulations described above to assemble an all-atom complex between the full-length spike and the ACE2 dimer. As a first step, equilibrated structures of the spike in the open state and of the ACE2-RBD complex

were extracted from their respective individual simulations. Subsequently, the spike protein was superimposed onto the ACE2- RBD complex by aligning the spikes' RBD "up" with the RBD of the ACE2-RBD complex, allowing for a fairly vertical arrangement of the new construct. In order to preserve the best possible binding interface, the RBD of the spike was discarded, whereas the RBD from the ACE2-RBD complex was retained and linked to the rest of the spike. The spike-ACE2 complex was embedded into a double membrane system: the spike's transmembrane do- main was inserted into a 330 Å 330 Å ERGIC-like lipid bilayer, whereas for ACE2 a mammalian cellular membrane of the same dimension was used. The two membranes were kept parallel to each other, allowing the use of an orthorhombic box. In order to facilitate the water and ion exchange between the internal and external compartment, an outer-membrane-protein-G (OmpG) porin folded into a beta barrel was embedded into each membrane. The OmpG equilibrated model was obtained from Chen et al.[32] The generated two-membrane construct was solvated with explicit TIP3P water molecules, with the total height of the external water compartment matching the internal one exhibiting a value of 380 Å. Sodium and chloride ions were added at a concentration of 150 mM to neutralize the charge and reshuffled to balance the charge between the two compartments.

The composite system, counting 8,562,698 atoms with an orthorhombic box of 330 Å 330 Å 850 Å, was subjected to all-atom MD simulation on the Summit computing system at ORNL using NAMD 2.14 and CHARMM36 all-atom additive force fields. Two cycles of conjugate gradient energy minimization and NPT pre-equilibration were conducted using a 2 fs timestep for a total of 3 ns. During this phase, the ACE2 and spike proteins and the glycans were harmonically restrained at 5 kcal/mol, allowing for the relaxation of the two lipid bilayers, the OmpG porins, water molecules and ions within the context of the double membrane system. We remark that the two lipid patches were previously equilibrated, therefore not requiring a melting phase at this stage. The dimension of the cell in the xy plane was maintained constant while allowing fluctuation along the z axis. Upon this initial pre- equilibration phase, a 17 ns NPT equilibration was performed by releasing all the restraints, preparing the system for production run. From this point, three replicas were run or a total of 522 ns comprehensive simulation time. By using the trained AI learning model, three conformations were extracted from this set of simulations, each of them representing a starting point of a new replica with re-initialized

velocities. A total of three additional simulations were therefore performed, collecting ~180 ns and bringing the total simulation time to ~702 ns.

6.4.1.5 SARS-CoV-2 viral envelope.

The full-scale viral envelope was constructed using the LipidWrapper program (v1.2) previously developed and described by Durrant et al.[33] A 350 Å x 350 Å lipid bilayer patch used as the pdb input was generated using CHARMM-GUI with an ERGIC-like lipid composition and an estimated area per lipid of 63 Å. An icospherical mesh with a 42.5 nm radius, in accordance with experimentally-observed SARS-CoV-2 radii, was exported as a collada file from Blender (v2.79b) and used as the surface file.[34] LipidWrapper was run in a Python 2.7 conda environment with lipid headgroup parameters "_P,CHL1_O3", a lipid clash cut-off of 1.0 Å, and filling holes enabled. The final bilayer pdb was solvated in a 110 nm cubic box using explicit TIP3P water molecules and neutralized with sodium and chloride ions to a concentration of 150 mM. The final system contained 76,134,149 atoms.

Since the LipidWrapper program operates via tessellation, lipid clash removal, and a subsequent lipid patching algorithm, the bilayer output attains a lower surface pressure than that of a bilayer of the same lipid composition at equilibrium. Due to this artifact, as the bilayer equilibrates, the lipids undergo lateral compression resulting in the unwanted formation of pores. Thus, the envelope was subjected to multiple rounds of minimization, heating, equilibration, and patching until the appropriate equilibrium surface pressure was reached.

All-atom MD simulations were performed using NAMD 2.14 and CHARMM36 all-atom additive force fields. The conjugate-gradient energy minimization procedure included two phases in which the lipid headgroups were restrained with 100 and 10 kcal/mol weights, respectively, at 310K for 15,000 cycles each. The membrane was then melted by incremental heating from 25 K to 310 K over 300 ps prior to NPT equilibration. The equilibration sequentially released the harmonic restraints on the lipid headgroups from 100 to 0 kcal/mol over 0.5 ns. Following this sequence, the structure was visually evaluated to determine whether to continue equilibration or to proceed with pore patching. Most structures continued with unrestrained equilibration for 4–26 ns prior to patching, with longer unrestrained equilibrations attributed to later, more stable envelopes.

Patching of the envelope was done by overlapping the initial LipidWrapper bilayer output with the newly-equilibrated envelope. All superimposed lipids within 2.0 Å of the equilibrated lipids were removed to eliminate clashes. Superimposed lipids within 4.0 Å of an equilibrated cholesterol molecule were also removed to eliminate ring penetrations. The patched system, with new lipids occupying the pores, was then re-solvated, neutralized, and subjected to the next round of minimization, heating, and equilibration.

After ten rounds of equilibration and patching, 24 spike proteins with glycans, 8 in the closed and 16 in the open state, were inserted randomly on the envelope using a house tcl script. A random placement algorithm was used in accordance with experimental microscopy imaging which has suggested that there is no obvious clustering of the spikes and no correlation between RBD state and location on the spike surface[34] The number of spikes was selected based on experimental evidence reporting a concentration of 1 spike per 1000 nm^2 on the envelope.[34] The new structure containing spikes was re-solvated, neutralized, and processed to remove clashing lipids prior to further simulation. The resulting cubic solvent box was 146 nm per side and contained 304,780,149 atoms. The spike-inclusive envelope was then subjected to three more equilibration and patching sequences. The final virion used for all-atom MD production runs had a lipid envelope of 75 nm in diameter with a full virion diameter of 120 nm. The complete equilibration of the viral envelope totaled 41 ns on the TACC Frontera system and 75 ns on ORNL Summit. Full-scale viral envelope production simulations were performed on Summit for a total of 84 ns in an NPT ensemble at 310 K, with a PME cutoff of 12 Å for non-bonded interactions.

6.5 Current State of the Art

6.5.1 Parallel molecular dynamics

NAMD[35] has been developed over more than two decades, with the goal of harnessing parallel computing to create a computational microscope[36,37] enabling scientists to study the structure and function of large biomolecular complexes relevant to human health. NAMD uses adaptive, asynchronous, message-driven execution based on Charm++.[38,39] It was one of the first scientific applications to make use of heterogeneous computing with GPUs[40], and it implements a wide variety of advanced features supporting state-of-the-art simulation methodologies. Continuing NAMD and Charm++ developments have brought improved work decomposition and distribution approaches and support for low overhead hardware-specific messaging layers,

enabling NAMD to achieve greater scalability on larger parallel systems.[41,42] NAMD incorporates a collective variables module supporting advanced biasing methods and a variety of in-situ analytical operations.[43] Simulation preparation, visualization, and post-hoc analysis are performed using both interactive and offline parallel VMD jobs.[44–47] NAMD has previously been used to study viruses and large photosynthetic complexes on large capability-oriented and leadership class supercomputing platforms, enabling the high-fidelity determination of the HIV-1 capsid structure[48], the characterization of substrate binding in influenza[49], and the structure and kinetics of light harvesting bacterial organelles.[50]

6.5.2 Weighted Ensemble MD simulations

The weighted ensemble (WE) method is an enhanced sampling method for MD simulations that can be orders of magnitude more efficient than standard simulations in generating pathways and rate constants for rare-event processes. WE runs many short simulations in parallel, instead of one long simulation, and directly gives both thermodynamic and kinetic properties, which most enhanced sampling methods cannot do. The simulations go through “resampling" where simulations are merged for over-sampled regions and replicated for rare regions so that regions are continuously sampled regardless of energy barriers. The simulations also carry probabilities or “weights" that are continuously updated and no statistical bias is added to the system, so we are able to directly obtain both thermodynamic (e.g., free energy landscape) and kinetic (e.g., rates and pathways) properties from the simulation. In addition, the WE method is one of the few methods that can obtain continuous unbiased pathways between states, so this was the most suitable method for us to obtain and observe the closed to open transition for the spike system. Before the WE method was applied to the spike system under investigation here (about 600,000 atoms), the largest system used for the WE method was the barnase-barnstar complex (100,000 atoms).[51]

6.5.3 AI-driven multiscale MD simulations

A number of approaches, including deep learning methods, have been developed for analysis of long timescale MD simulations.[52] These linear, non-linear, and hybrid ML approaches cluster the simulation data along a small number of latent dimensions to identify conformational transitions between states.[53,54] Our group developed a deep learning approach, namely the variational autoencoder that uses convolutional filters on contact

maps (from MD simulations) to analyze long time-scale simulation datasets and organize them into a small number of conformational states along biophysically relevant reaction coordinates.[55] We have used this approach to characterize protein conformational landscapes.[56] However, with the spike protein, the intrinsic size of the simulation posed a tremendous challenge in scaling our deep learning approaches to elucidate conformational states relevant to its function.

Recently, we extended our approach to adaptively run MD simulation ensembles to fold small proteins. This approach, called DeepDriveMD[57], successively learns which parts of the conformational landscape have been sampled sufficiently and initiates simulations from undersampled regions of the conformational landscape (that also constitute "interesting" features from a structural perspective of the protein). While a number of adaptive sampling techniques exist[58–64] , including based on reinforcement learning methods[65], these techniques have been demonstrated on prototypical systems. In this paper, we utilize the deep learning framework to suggest additional points for sampling and do not necessarily use it in an adaptive manner to run MD simulations (mainly due to the limitations posed by the size of the system). However, extensions to our framework for enabling support of such large-scale systems are straightforward and further work will examine such large-scale simulations.

6.6 Innovations Realized

6.6.1 Parallel molecular dynamics

Significant algorithmic improvements and performance optimizations have been required for NAMD to achieve high performance on the GPU-dense Summit architecture.[18,39] New CUDA kernels for computing the short-range non-bonded forces were developed that implement a "tile list" algorithm for decomposing the workload into lists of finer grained tiles that more fully and equitably distribute work across the larger SM (streaming multiprocessor) counts in modern NVIDIA GPUs. This new decomposition uses the symmetry in Newton's Third Law to eliminate redundant calculation without incurring additional warp-level synchronization. CUDA kernels also were added to offload the calculation of the bonded force terms and non-bonded exclusions. Although these terms account for a much smaller percentage of the work per step than that of the short-range non-bonded forces, NAMD performance on Summit benefits from further reduction of CPU workload. NAMD also benefits from the portable high-performance communication layer in Charm++ that communicates using

the IBM PAMI (Parallel Active Messaging Interface) library, which improves performance by up to 20% over an MPI-based implementation.[39,41]

Additional improvements have benefited NAMD performance on Frontera. Recent developments in Charm++ now include support for the UCX (Unified Communication X) library which improves performance and scaling for Infiniband-based networks. Following the release of NAMD 2.14, a port of the CUDA tile list algorithm to Intel AVX-512 intrinsics was introduced, providing a 1.8 performance gain over the "Sky Lake" (SKX) builds of NAMD.

A significant innovation in NAMD and VMD has been the development of support for simulation of much larger system sizes, up to two billion atoms. Support for larger systems was developed and tested through all-atom modeling and simulation of the protocell as part of the ORNL CAAR (Center for Accelerated Application Readiness) program that provided early science access to the Summit system. This work has greatly improved the performance and scalability of internal algorithms and data structures of NAMD and VMD to allow modeling of biomolecular systems beyond the previous practical limitation on the order of 250 million atoms. This work has redefined the practical simulation size limits in both NAMD and VMD and their associated file formats, added new analysis methods specifically oriented toward virology[66], and facilitates modeling of cell-scaled billion-atom assemblies, while making smaller modeling projects significantly more performant and streamlined than before.

6.6.2 Multiscale molecular dynamics simulations

Often referred to as "computational microscopy," MD simulations are a powerful class of methods that enable the exploration of complex biological systems, and their time-dependent dynamics, at the atomic level. The systems studied here push state of the art in both their size and complexity. The system containing a full-length, fully- glycosylated spike protein, embedded in a realistic viral membrane (with composition that mimics the endoplasmic reticulum) contains essentially all of the biological complexity known about the SARS-CoV-2 spike protein. The composite system contains 1.7 million atoms and combines data from multiple cryo-EM, glycomics, and lipidomics datasets. The system was simulated with conventional MD out to microseconds in length, and several mutant systems were simulated and validated with independent experiments.

A related set of experiments utilizing the weighted ensemble method, an enhanced sampling technique, explored a truncated version of the spike protein (600,000 atoms with explicit solvent) in order to simulate an unbiased spike protein conformational transition from the closed to open state. This is the largest system, by an order of magnitude, that has been simulated using the WE method (biggest system until now was 60,000 atoms). Using calculations optimized to efficiently make use of extensive GPU resources, we obtained several full, unbiased paths of the glycosylated spike receptor binding domain activation mechanism.

The second system increases the complexity by an order of magnitude by combining the spike system described above with a full- length, fully-glycosylated model of the ACE2 receptor bound into a host cell plasma membrane. This system represents the encounter complex between the spike and the ACE2 receptor, contains two parallel membranes of differing composition, has both the spike and ACE2 fully glycosylated, and forming a productive binding event at their interface. The composite system contains 8.5 million atoms with explicit water molecules and provides unseen views into the critical handshake that must occur between the spike protein and the ACE2 receptor to begin the infection cascade.

Our final system is of the SARS-CoV-2 viral envelope. This system incorporates 24 full-length, fully-glycosylated spike proteins into a viral membrane envelope of realistic (ER-like) composition, where the diameter of the viral membrane is 80nm and the diameter of the virion, inclusive of spikes, is 146 nm. Until now, the largest system disclosed in a scientific publication was the influenza virus, which contained 160 million atoms. The SARS-CoV-2 viral envelope simulation developed here contains a composite 305 million atoms, and thus breaks new ground for MD simulations of viruses in terms of particle count, size, and complexity.

Moreover, typical state of the art simulations are run in isolation, presenting each as a self-contained story. While we also do that for each of the systems presented here, we advance on state of the art by using an AI-driven workflow that drives simulation at one scale, with knowledge gained from a disparate scale. In this way, we are able to explore relevant phase space of the spike protein more efficiently and in environments of increasing complexity.

6.6.3 Using AI for driving multiscale simulations

MD simulations such as the ones described above generate tremendous amounts of data. For e.g., the simulations of the WE sampling of the spike protein's closed- to-open state generated over 100 terabytes of data. This imposes a heavy burden in terms of understanding the intrinsic latent dimensions along which large-scale conformational transitions can be characterized. A key challenge then is to use the raw simulation datasets (either coordinates, contact matrices, or other data collected as part of a standard MD runs) to cluster conformational states that have been currently sampled, to identify biologically relevant transitions between such states (e.g., open/closed states of spike), and suggest conformational states that may not be fully sampled to characterize these transitions.[54] To deal with the size and complexity of these simulation datasets, approaches that analyze 3D point clouds are more appropriate. Indeed, such approaches are becoming more commonly utilized for characterizing protein binding pockets and protein-ligand interactions. We posited that such representations based on the Cα representation of protein structures could be viable to characterize large-scale conformational changes within MD simulation trajectories. We leverage the 3D PointNet based[67] adversarial autoencoder (3D-AAE) developed by Zamorski and colleagues[68] to analyze the spike protein trajectories. In this work, we employ the chamfer distance based reconstruction loss and a Wasserstein[69] adversarial loss with gradient penalty[70] to stabilize training. The original PointNet backbone treats the point cloud as unordered, which is true for general point clouds. In our case however, the protein is essentially a 1D embedding into a 3D space. This allows us to define a canonical order of points, i.e., the order in which they appear in the chain of atoms. For that reason, we increase the size-1 1D convolutional encoder kernels from the original PointNet approach to larger kernels up to size 5. This allows the network to not only learn features solely based on distance, but also based on neighborhood in terms of position of each atom in the chain. We found that a 4-layer encoder network with kernel sizes [5, 3, 3, 1, 1] and filter sizes [64, 128, 256, 256, 512] performs well for most tasks. A final dense layer maps the vectors into latent space with dimensionality 64. For the generator, we only use unit size kernels with filter dimensions [64, 128, 512, 1024, 3] respectively (the output filter size is always the dimensionality of the problem). The discriminator is a 5 layer fully connected network with layer widths [512, 512, 128, 64, 1].

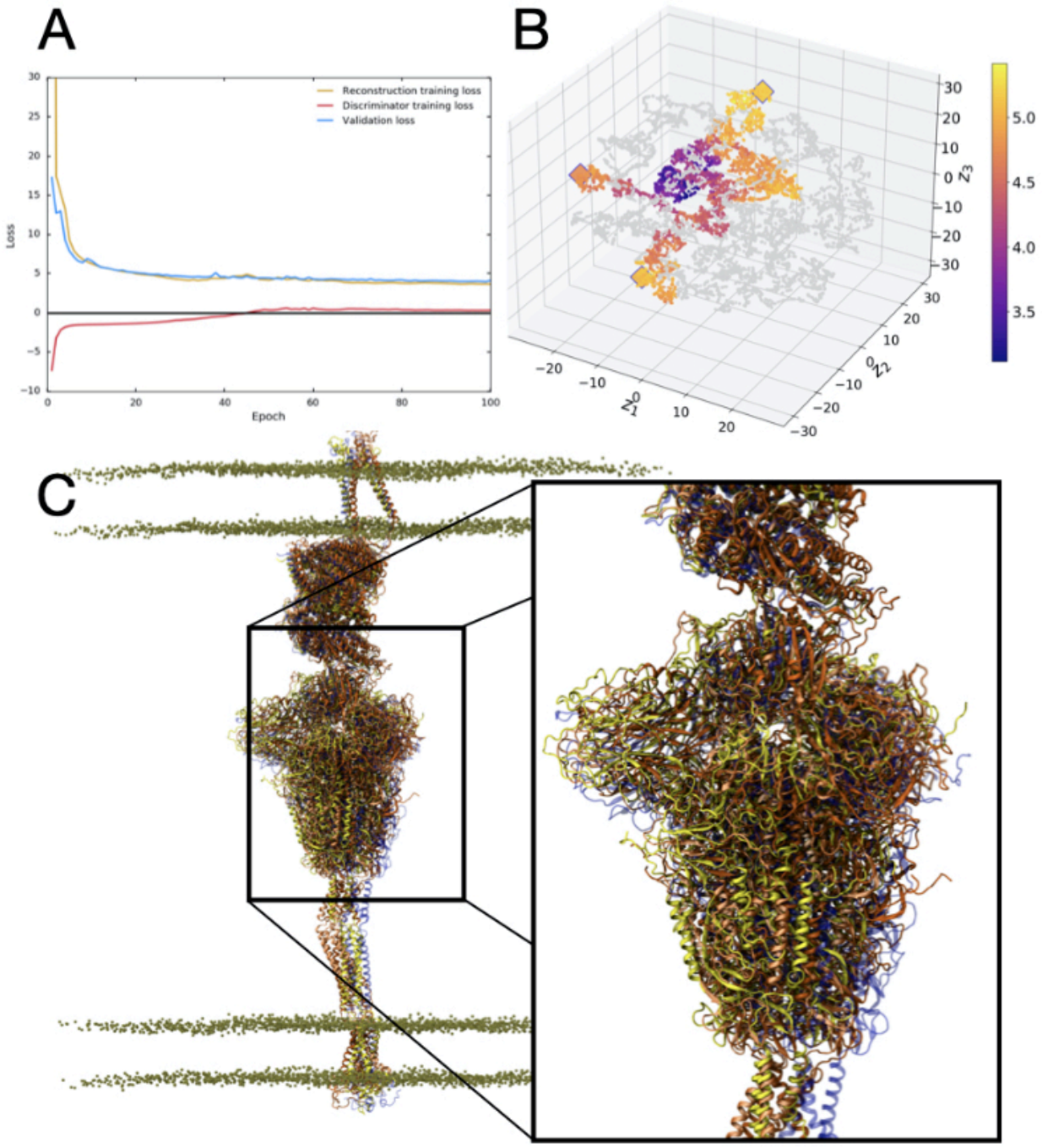


Figure 6.3: 3D-AAE training and test results. A) The loss progression for reconstruction, discriminator and validation loss over 100 epochs. B) The t-SNE plot visualization of the reduced latent space, with training embeddings represented in grey and test examples represented in color over the range of RMSD values. Outliers identified in the outlier detection stage are represented with an outlined diamond. C) VMD visualization of outlier structures (yellow, orange, dark orange) aligned and compared to the starting structure (blue).

The trajectories from the WE simulations were used to build a combined data set consisting of 130,880 examples. The point cloud data, representing the coordinates of the 3,375 backbone Cα atoms of the protein,

was randomly split into training (80%) and validation input (20%) and was used to train the 3D-AAE model for 100 epochs using a batch size of 32. The data was projected onto a latent space of 64 dimensions constrained by a gaussian prior distribution with a standard deviation of 0.2. The loss optimization was performed with the Adam optimizer, a variant of stochastic gradient descent, using a learning rate of 0.0001. We also added hyperparameters to scale individual components of the loss. The reconstruction loss was scaled by 0.5 and the gradient penalty by a factor of 10.

The embedding learned from the 3D-AAE model summarizes a latent space that is similar to variational autoencoders, except that 3D-AAEs tend to be more robust to outliers within the simulation data. The embeddings learned from the simulations allow us to cluster the conformations (in an unsupervised manner) based on their similarity in overall structure, which can be typically measured using quantities such as root-mean squared deviations (RMSD).

We trained the model using several combinations of hyperparameters, mainly varying learning rate, batch size and latent dimension. For visualizing and assessing the quality of the model in terms latent space structure, we computed t-SNE[71] dimensionality reductions on the high-dimensional embeddings from the validation set. A good model should generate clusters with respect to relevant biophysical observables not used in the training process. Therefore, we painted the t-SNE plot with the root mean squared deviation (RMSD) of each structure to the starting conformation and observed intelligible clustering of RMSD values. We tested this model on a set of trajectories from the full-scale spike-ACE2 system, using the same atom selection (3,375 Cα atoms) as the corresponding WE spike protein. We subsequently performed outlier detection using the local outlier factor (LOF) algorithm, which uses distance from neighboring points to identify anomalous data. The goal of the outlier detection step is to identify conformations of the protein that are most distinct from the starting structure, in order to story board important events in the transition of the protein from an open to closed conformation. Although the number of outlier conformations detected can be a parameter that the end-user can specify, we selected 20 outlier conformations, based on the extreme LOF scores. These conformations were visualized in VMD and further analyzed using tilt angles of the stalk and the RBD. The final selection included 3 structures which were used as the starting conformations for the next set of simulations. These ‘outlier’ conformers are cycled through additional MD simulations that are driven by the ML-methods.

6.7 How Performance Was Measured

6.7.1 3D-AAE

Since this application dominantly utilizes the GPU, we do not need to profile CPU FLOPs. Instead, we measure FLOPs for all precisions using the methodology explained in Yang et al.[72] with the NVIDIA NSight Compute 2020 GPU profiling tool. We collect floating point instructions of relevant flavors (i.e. adds, mults, fmas (fused multiply adds) and tensor core operations for FP16, FP32 and FP64) and multiply those with weighting factors of {1, 1, 2, 512} respectively in order to transform those into FLOP counts. The sum of all these values for all precisions will yield our overall mixed precision FLOP count. To exclude FLOPs occurring during initialization and shut- down, we wrap the training iteration loop into start/stop profiler hooks provided by the NVIDIA CuPy Python package.

Table 6.2: NAMD AVX-512 FP operation breakdown.

FP Instr.	Ops	% total	FP Instr.	Ops	% total
DblScalar	4.99e16	26.9%	SglScalar	2.09e15	1.1%
Dbl128b	6.86e15	3.7%	Sgl128b	3.61e15	1.9%
Dbl256b	1.06e17	57.1%	Sgl256b	1.18e16	6.3%
Dbl512b	4.96e15	2.7%	Sgl512b	3.43e14	0.2%

6.7.2 NAMD

NAMD performance metrics were collected on TACC Frontera, using the Intel msr-tools utilities, with NAMD 2.14 with added Intel AVX-512 support. FLOP counts were measured for each NAMD simulation with runs of two different step counts. The results of the two simulation lengths were subtracted to eliminate NAMD startup operations, yielding an accurate estimate of the marginal FLOPs per step for a continuing simulation.[73]

FLOP counts were obtained by reading the hardware performance counters on all CPU cores on all nodes, using the rdmsr utility from msr-tools.4 At the beginning of each job, the “TACC stats” system programs the core performance counters to count the 8 sub-events of the Intel FP_ARITH_INST_RETIRED. Counter values are summed among the 56 cores in each node, and ultimately among each node. Each node-summed counter value is scaled by the nominal SIMD-width of the floating-point instruction being counted and the 8 classes are added together to provide the total FLOP count per node. The hardware counters do not take masked

SIMD instructions into account. SIMD lanes that are masked-out still contribute to the total FLOPs, however static analysis of the AVX-512-enabled NAMD executable showed that only 3.7% of FMA instructions were masked.

A breakdown of floating-point instruction execution frequency for the AVX-512 build of NAMD across 2048 nodes is shown in **Table 6.1**. For CPU versions of NAMD, arithmetic is performed in double precision, except for single-precision PME long-range electrostatics calculations and associated FFTs. In the GPU-accelerated NAMD on Summit, single-precision arithmetic is used for both PME and also for short-range non-bonded force calculations, significantly increasing the fraction of single-precision instructions, at the cost of requiring a mixed-precision patch-center-based atomic coordinate representation to maintain full force calculation precision.

6.8 Performance Results

6.8.1 3D-AAE training performance

We used the aforementioned recipe for GPU profiling to determine the performance for the 3D-AAE training. We measure the FLOP counts individually for 2 training and 1 validation steps for a batch size of 32. The latent dimension of the model is a free hyperparameter and affects the FLOP count. We trained three models with latent dimensions 32, 64, 128 in order to determine an optimal model for the task and thus we profile and report numbers for all of those. All models were trained for 100 epochs with batch size 32 on a single V100 GPU each. As mentioned above, the train/validation dataset split is 80%/20% and we do one validation pass after each training epoch. Thus, we can assume that this fraction translates directly into the FLOP counts for these alternating two stages. Our sustained performance numbers are computed using this weighted FLOP count average and the total run time. In order to determine peak performance, we compute the instantaneous FLOP rate for the fastest batch during training. Note that the 3D-AAE does exclusively use float (FP32) precision. The performance results are summarized in **Table 6.2.** Although the model is dense linear algebra heavy, it is also rather lightweight so it cannot utilize the full GPU and thus only delivering 20% of theoretical peak performance.

Table 6.3: 3D-AAE training performance on one V100 GPU.

Latent Dimensions	Peak TFLOP/s	Sustained TFLOP/s
32	2.96	0.97
64	3.16	2.28
128	3.13	0.91

Table 6.4: NAMD simulation floating point ops per timestep.

NAMD Simulation	Atoms	FLOPS/step
ACE2-RBD complex	800k	21.57 GFLOPS/step
Single Spike	1.7M	47.96 GFLOPS/step
Spike-ACE2 complex	8.5M	243.7 GFLOPS/step
SARS-CoV-2 virion	305M	8.3511 TFLOPS/step

As expected, the peak performance is very consistent between the runs. The big difference in sustained performance between latent dim 64 and the other two models is that the frequency for computing the t-SNE was significantly reduced, i.e., from every epoch to every 5th. The t-SNE computation and plotting happens after each validation in a background thread on the CPU, but the training epochs can be much shorter than the t-SNE time. In that case, the training will stall until the previous t-SNE has completed. Evidently, decreasing the t-SNE frequency reduces that overhead significantly. We expect that the other models would perform similarly if we would have enabled this optimization for those runs as well. The remaining difference in peak vs. sustained performance can be explained by other overhead, e.g., storing embedding vectors, model checkpoints and the initial scaffolding phase. Furthermore, it includes the less FLOP-intensive validation phase whereas the peak estimate is obtained from the FLOP-heavy training phase.

6.8.2 NAMD simulation performance

Low-level NAMD performance measurements were made on the TACC Frontera system, to establish baseline counts of FLOPs per timestep for the four different biomolecular systems simulated as part of this work, summarized in **Table 6.3** with the breakdown of CPU FLOPs described in **Table 6.1**. Sustained NAMD performance measurements were obtained using built-in application timers over long production science runs of several hours, including all I/O, and reported in units of nanoseconds per day of simulation time. NAMD sustained simulation performance for the spike-ACE2 complex is summarized for the TACC Frontera and ORNL Summit systems in **Table 6.4** and **Figure 6.4**. NAMD sustained simulation performance, parallel speedup, and

scaling efficiency are reported for the full SARS-CoV-2 virion in **Table 6.5.** Peak NAMD mixed-precision FLOP rates on ORNL Summit are estimated in **Table 6.6** by combining sustained performance measurements with FLOPs/timestep measurements.

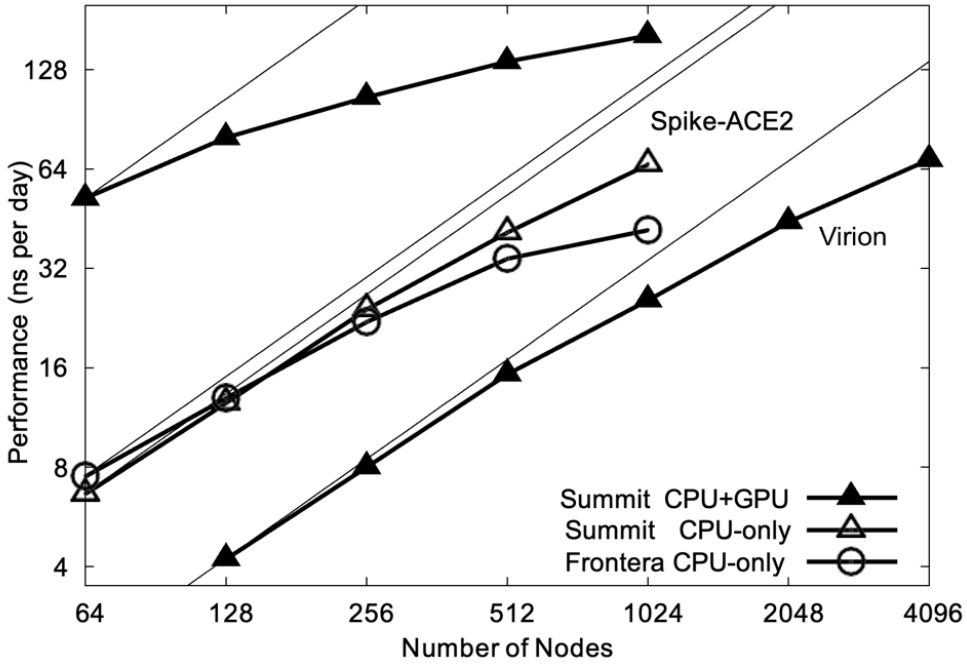


Figure 6.4: NAMD scaling on Summit and Frontera for 8.5M- atom spike-ACE2 complex (upper lines) and 305M-atom virion (lower line). Thin lines indicate linear scaling.

6.9 Implications

Our major scientific achievements are:

(1) **We characterize for the first time the glycan shield of the full-length SARS-CoV-2 spike protein (including the stalk) and find that two N-glycans linked to N165 and N234 have a functional role in modulating the dynamics of the spike's RBD.** This unprecedented finding establishes a major new role of glycans in this system as playing an active role in infection, beyond shielding (Fig. 1C) (Casalino et al., 2020b).

(2) **We discover that the human ACE2 receptor has a flexible hinge in the linker region near the membrane that enables it to undergo exceptionally large angular motions relative to the plane of the membrane.** We predict this flexibility will aid forming productive complexes with the spike protein and may serve as a mechanical energy source during the cell fusion process (Barros et al., 2020).

(3) **We openly share our models, methods, and data, making them freely available to the scientific community.** We are committed to the shared set of principles outlined in Ref. (Amaro and Mulholland, 2020): depositing findings as preprints in advance of formal peer review, making available our models at the time of deposition into a preprint server (Barros et al., 2020), and releasing the full datasets upon peer review (Casalino et al., 2020b). By doing so, the reproducibility and robustness of our findings and methods are enhanced, and the scientific findings from our simulations are amplified through reuse by others.

(4) **We describe for the first time unbiased pathways for the full closed-to-open transition of the spike's RBD (Fig. 2), where knowledge of this pathway has the potential to inform on mechanisms of viral infection as well as potentially aid in the discovery of novel druggable pockets within the spike.** Our work set a new milestone for the use of the weighted ensemble method in biomolecular simulation, increasing applicable system size by an order or magnitude over current state of the art.

(5) **We characterize the spike's flexibility in the context of ACE2 binding.** One of the most important properties of the spike protein is its intrinsic flexibility, a key feature that facilitates the interaction with the ACE2 receptors exposed on the host cell. Cryo-EM and cryo-ET structural data revealing the architecture of the SARS-CoV-2 viral particle showed that the spike can tilt up to 60º with respect to the perpendicular to the membrane.[34,74] Behind this flexibility is the structural organization of the extra-virion portion of the spike, composed of two major domains, the stalk and the head, that are connected through a flexible junction that has been referred to as "hip" (Figure 7.5A). Moreover, the stalk can be further divided into an upper and a lower leg, which correspond to the extra-virion alpha-helices of the coil-coiled trimeric bundle, and the transmembrane domain, which can be intended as the foot of this organizational scaffold. The stalk's upper leg, lower leg and the foot are interspersed by highly flexible loops defined as "knee" and "ankle" junctions (**Figure 6.5A**). We then harnessed DeepDriveMD to perform adaptive MD on the Spike-ACE2 8.5 million atoms system. Following this workflow, we extracted three conformations from the first set of Spike-ACE2 MD simulations (replicas 1-3) and subsequently used them as starting points for a new round of MD (replicas 4-6). We then calculated the distribution of the overall spike tilting with respect to the perpendicular to the membrane (**Figure 6.5E**)

and of other three angles involving the stalk, namely the "hip" angle between the stalk's up- per leg and the head (**Figure 6.5B**), the "knee" angle between the stalk's lower and upper legs (**Figure 6.5C**), and the "ankle" angle between the perpendicular to the membrane and the stalk's lower leg (**Figure 6.5D**). The AI-driven adaptive MD approach expanded the conformational space explored, especially for the knee and hip angles, showing average values of 18.5º ± 7.7º and 13.8º ± 7.6º for replicas 1-3, shifted to 30.4º ± 5.1º and 18.8º ± 6.0º for the subsequent set of MD (replicas 4-6), respectively. The population shift is less pronounced for the ankle, exhibiting an average angle of 21.8º ± 2.7º. These results, in agreement with the data from Turonova et al.[75] that however did not consider the spike in complex with ACE2, reveal large hinge motions throughout the stalk and between the stalk and the head that accommodate the interaction between the spike's RBD and the ACE2 receptor, preventing the disruption of the binding interface. This is further highlighted by the overall tilting of the spike that remains well defined around 7.3º ± 2.0º (**Figure 6.5E**) showing that the stalk's inner hinge motions prevent a larger scale bending that could potentially disrupt the RBD-ACE2 interaction.

(6) **Our approach points to the very near-term ability to accelerate the sampling of dynamical configurations of the complicated viral infection machinery within in the context of its full biological complexity using AI.** The enormous amount of data arising from MD and WE simulations of the single spike served to build and train an AI model using the variational autoencoder deep learning approach, which we demonstrate to accelerate dynamical sampling of the spike in a larger, more complex system (i.e., the two parallel membrane spike-ACE2 complex). Thus, the combination of the AI-driven workflows together with the ground- breaking simulations opens the possibility to overcome a current major bottleneck in the development and use of such ultra-large scale MD simulations, which relates to the efficient and effective sampling of the conformational dynamics of a system with so many degrees of freedom. The scientific implications of such a techno- logical advance, in terms of understanding of the basic science of molecular mechanisms of infection as well as the development of novel therapeutics, are vast.

(7) **We establish a new high watermark for the atomic-level simulation of viruses with the simulation of the SARS-CoV-2 viral envelope, tallying 305 million atoms including explicit water molecules,**

and exhibiting a strong scaling on Summit. The virion has a realistic ERGIC-like membrane, contains 24 fully glycosylated full-length spikes (in both the open and closed states) and replicates the spatial patterning and density of viral proteins as determined from cryoelectron tomography experiments. These groundbreaking simulations, just now in the process of being fully analyzed, set the stage for future work on SARS-CoV-2 that will be unprecedented in terms of their ability to more closely mimic realistic biological conditions. This includes, for example, the ability to explore the interactions of the virus with multiple receptors on the host cell, or multiple antibodies. It will allow researchers to explore the correlated dynamics of the molecular pieceparts on the surface of the virus and the host cell, and the effects of curvature on such behavior. It will be used as the ground-truth in the development of other simulation approaches, including coarse grained simulation methods, which are under development.[76] It will aid in the development of methods related to the construction of complicated biological membranes. And the list goes on.

(8) **We developed an AI-driven workflow as a generalizable framework for multiscale simulation.** Though we focus here on advances made relevant to COVID19, the methods and work- flow established here will be broadly applicable to the multiscale simulation of molecular systems.

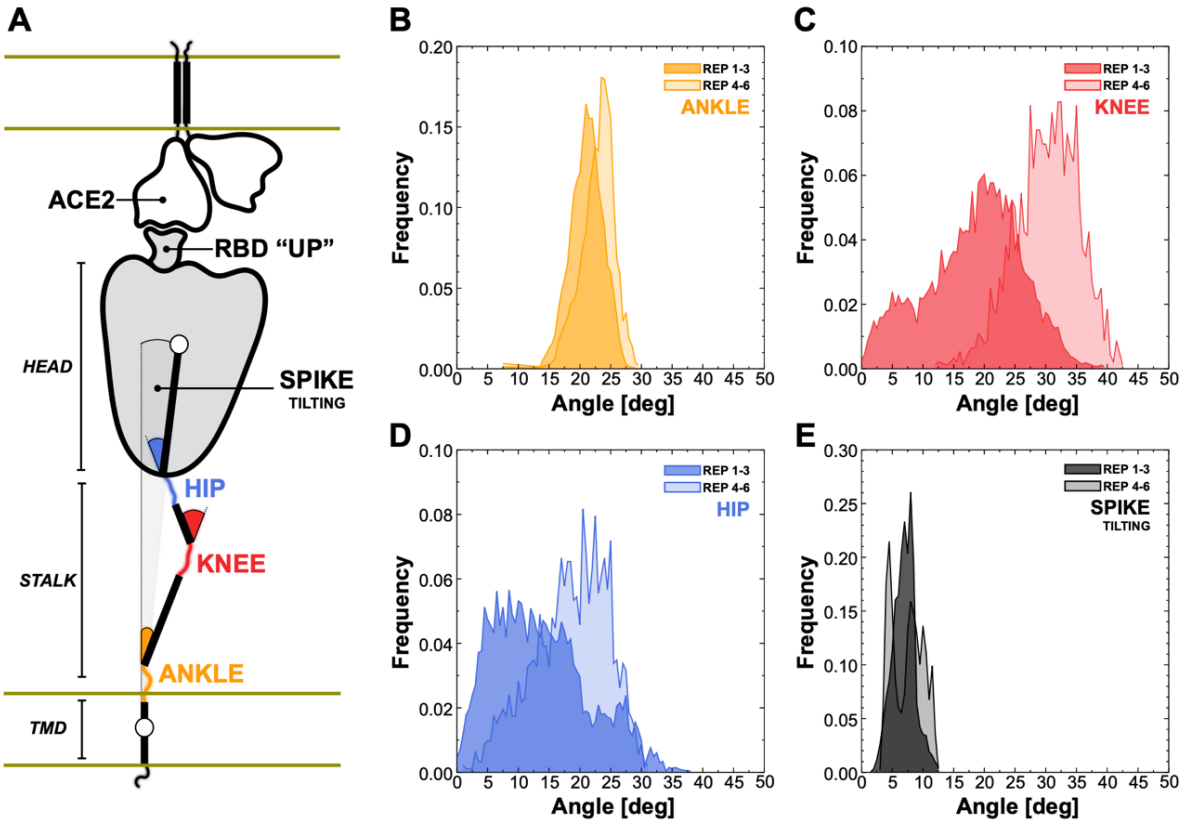


Figure 6.5: Flexibility of the spike bound to the ACE2 receptor. A) Schematic representation of the two-parallel-membrane system of the spike-ACE2 complex. (B-E) Distributions of the ankle, knee, hip and spike-tilting angles resulting from MD replicas 1-3 (darker color) and 4-6 (lighter color). Starting points for replicas 4-6 have been selected using DeepDriveMD.

6.10 Acknowledgments

The authors thank D. Maxwell, B. Messer, J. Vermaas, and the Oak Ridge Leadership Computing Facility at Oak Ridge National Laboratory supported by the DOE under Contract DE-AC05-00OR22725. We also thank the Texas Advanced Computing Center Frontera team, especially D. Stanzione and T. Cockerill, and for compute time made available through a Director's Discretionary Allocation (NSF OAC-1818253). We thank the Argonne Leadership Computing Facility supported by the DOE under DE-AC02-06CH11357. NAMD and VMD are funded by NIH P41-GM104601. The NAMD team thanks Intel and M. Brown for contributing the AVX-512 tile list kernels. Anda Trifan acknowledges support from a DOE CSGF (DE-SC0019323). This work was supported by NIH GM132826, NSF RAPID MCB-2032054, an award from the RCSA Research Corp., a UC San Diego Moore's Cancer Center 2020 SARS-COV-2 seed grant, to R.E.A. This research was supported

by the Exascale Computing Project (17-SC-20-SC), a collaborative effort of the US DOE Office of Science and the National Nuclear Security Administration. Research was supported by the DOE through the National Virtual Biotechnology Laboratory, a consortium of DOE national laboratories focused on response to COVID-19, with funding from the Coronavirus CARES Act. This work used resources, services, and support from the COVID-19 HPC Consortium (https://covid19- hpc-consortium.org/), a private-public effort uniting government, industry, and academic leaders who are volunteering free compute time and resources in support of COVID-19 research. We dedicate this contribution to the memory of Klaus Schulten.

Chapter 6, in full, is a modified reprint of the material as it appears in "Casalino, L. †, Dommer, A. C. †, and Gaieb, Z. †; Barros, E. P.; Sztain, T.; Ahn, S.; Trifan, A.; Brace, A.; Ma, H.; Lee, H.; Bogetti, A.; Khalid, S.; Chong, L.; Simmerling, C.; Hardy, D. J.; Maia, J. D.; Phillips, J. C.; Kurth, T.; Stern, A.; Huang, L.; McCalpin, J.; Tatineni, M.; Gibbs, T.; Stone, J. E.; Jha, S.; Ramanathan, A.; Amaro, R. E. *AI-Driven Multiscale Simulations Illuminate Mechanisms of SARS-CoV-2 Spike Dynamics.* Int. J. High Perform. Comput. Appl. 2021, 35 (5), 432-451." The dissertation author was a primary co-investigator and co-author of this work.

References

(1) Starr, T. N.; Greaney, A. J.; Hilton, S. K.; Ellis, D.; Crawford, K. H. D.; Dingens, A. S.; Navarro, M. J.; Bowen, J. E.; Tortorici, M. A.; Walls, A. C.; King, N. P.; Veesler, D.; Bloom, J. D. Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. *Cell* **2020**, *182* (5), 1295-1310.e20. https://doi.org/10.1016/j.cell.2020.08.012.

(2) Shajahan, A.; Supekar, N. T.; Gleinich, A. S.; Azadi, P. Deducing the N-and O-Glycosylation Profile of the Spike Protein of Novel Coronavirus SARS-CoV-2. *Glycobiology 2020*, 1–8. https://doi.org/10.1093/glycob/cwaa042.

(3) Watanabe, Y.; Allen, J. D.; Wrapp, D.; McLellan, J. S.; Crispin, M. Site-Specific Glycan Analysis of the SARS-CoV-2 Spike. *Science (New York, N.Y.)* **2020**. https://doi.org/10.1126/science.abb9983.

(4) Casalino, L.; Gaieb, Z.; Goldsmith, J. A.; Hjorth, C. K.; Dommer, A. C.; Harbison, A. M.; Fogarty, C. A.; Barros, E. P.; Taylor, B. C.; Mclellan, J. S.; Fadda, E.; Amaro, R. E. Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. *ACS Central Science* **2020**. https://doi.org/10.1021/acscentsci.0c01056.

(5) Wrapp, D.; Wang, N.; Corbett, K. S.; Goldsmith, J. A.; Hsieh, C.-L.; Abiona, O.; Graham, B. S.; McLellan, J. S. Cryo-EM Structure of the 2019-NCoV Spike in the Prefusion Conformation. *Science (New York, N.Y.)* **2020**, *1263* (March), 1260–1263. https://doi.org/10.1126/science.abb2507.

(6) Walls, A. C.; Park, Y. J.; Tortorici, M. A.; Wall, A.; McGuire, A. T.; Veesler, D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **2020**, *181* (2), 281-292.e6. https://doi.org/10.1016/j.cell.2020.02.058.

(7) Šali, A.; Blundell, T. L. Comparative Protein Modelling by Satisfaction of Spatial Restraints. *Journal of Molecular Biology* **1993**, *234* (3), 779–815. https://doi.org/10.1006/jmbi.1993.1626.

(8) Zhang, Y. I-TASSER Server for Protein 3D Structure Prediction. *BMC Bioinformatics* **2008**, *9* (1), 1–8. https://doi.org/10.1186/1471-2105-9-40.

(9) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM. *Journal of computational chemistry* **2008**, *29*, 2967–2970. https://doi.org/10.1002/jcc.

(10) Jo, S.; Song, K. C.; Desaire, H.; MacKerell, A. D.; Im, W. Glycan Reader: Automated Sugar Identification and Simulation Preparation for Carbohydrates and Glycoproteins. *Journal of Computational Chemistry* **2011**. https://doi.org/10.1002/jcc.21886.

(11) Park, S. J.; Lee, J.; Qi, Y.; Kern, N. R.; Lee, H. S.; Jo, S.; Joung, I.; Joo, K.; Lee, J.; Im, W. CHARMM-GUI Glycan Modeler for Modeling and Simulation of Carbohydrates and Glycoconjugates. *Glycobiology* **2019**. https://doi.org/10.1093/glycob/cwz003.

(12) Casares, D.; Escribá, P. V.; Rosselló, C. A. Membrane Lipid Composition: Effect on Membrane and Organelle Structure, Function and Compartmentalization and Therapeutic Avenues. *International Journal of Molecular Sciences* **2019**, *20* (9). https://doi.org/10.3390/ijms20092167.

(13) Van Meer, G.; Voelker, D. R.; Feigenson, G. W. Membrane Lipids: Where They Are. *Nature Reviews Molecular Cell Biology* **2009**, *101* (1), 1–4. https://doi.org/10.1038/nrm2330.Membrane.

(14) Wu, E. L.; Cheng, X.; Jo, S.; Rui, H.; Song, K. C.; Dávila-Contreras, E. M.; Qi, Y.; Lee, J.; Monje-Galvan, V.; Venable, R. M.; Klauda, J. B.; Im, W. CHARMM-GUI Membrane Builder toward Realistic Biological Membrane Simulations. *Journal of Computational Chemistry*. 2014. https://doi.org/10.1002/jcc.23702.

(15) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *The Journal of Chemical Physics* **1983**, *79* (2), 926–935. https://doi.org/10.1063/1.445869.

(16) Guvench, O.; Hatcher, E.; Venable, R. M.; Pastor, R. W.; MacKerell, A. D. CHARMM Additive All-Atom Force Field for Glycosidic Linkages between Hexopyranoses. *Journal of Chemical Theory and Computation* **2009**, *5* (9), 2353–2370. https://doi.org/10.1021/ct900242e.

(17) Huang, J.; Mackerell, A. D. CHARMM36 All-Atom Additive Protein Force Field: Validation Based on Comparison to NMR Data. *Journal of Computational Chemistry* **2013**, *34* (25), 2135–2145. https://doi.org/10.1002/jcc.23354.

(18) Phillips, J. C.; Hardy, D. J.; Maia, J. D. C.; Stone, J. E.; Ribeiro, J. V.; Bernardi, R. C.; Buch, R.; Fiorin, G.; Hénin, J.; Jiang, W.; McGreevy, R.; Melo, M. C. R.; Radak, B.; Skeel, R. D.; Singharoy, A.; Wang, Y.; Roux, B.; Aksimentiev, A.; Luthey-Schulten, Z.; Kalé, L. V.; Schulten, K.; Chipot, C.; Tajkhorshid, E. Scalable Molecular Dynamics on CPU and GPU Architectures with NAMD. *J. Chem.\ Phys.* **2020**, *153*, 044130. https://doi.org/10.1063/5.0014475.

(19) Yan, R.; Zhang, Y.; Li, Y.; Xia, L.; Guo, Y.; Zhou, Q. Structural Basis for the Recognition of SARS-CoV-2 by Full-Length Human ACE2. *Science* **2020**, *367* (6485), 1444–1448. https://doi.org/10.1126/science.abb2762.

(20) Sun, Z.; Ren, K.; Zhang, X.; Chen, J.; Jiang, Z.; Jiang, J.; Ji, F.; Ouyang, X.; Li, L. Mass Spectrometry Analysis of Newly Emerging Coronavirus HCoV-19 Spike Protein and Human ACE2 Reveals Camouflaging Glycans and Unique Post-Translational Modifications. *Engineering* **2020**. https://doi.org/10.1016/j.eng.2020.07.014.

(21) Zhao, P.; Praissman, J. L.; Grant, O. C.; Cai, Y.; Xiao, T.; Rosenbalm, K. E.; Aoki, K.; Kellman, B. P.; Bridger, R.; Barouch, D. H.; Brindley, M. A.; Lewis, N. E.; Tiemeyer, M.; Chen, B.; Woods, R. J.; Wells, L. Virus-Receptor Interactions of Glycosylated SARS-CoV-2 Spike and Human ACE2 Receptor. *Cell Host and Microbe* **2020**, *28* (4), 586-601.e6. https://doi.org/10.1016/j.chom.2020.08.004.

(22) Barros, E. P.; Casalino, L.; Gaieb, Z.; Dommer, A. C.; Wang, Y.; Fallon, L.; Raguette, L.; Belfon, K.; Simmerling, C.; Amaro, R. E. The Flexibility of ACE2 in the Context of SARS-CoV-2 Infection. *bioRxiv* **2020**. https://doi.org/10.1101/2020.09.16.300459.

(23) Onuchic, J. N.; Luthey-Schulten, Z.; Wolynes, P. G. THEORY OF PROTEIN FOLDING: The Energy Landscape Perspective. *Annual Review of Physical Chemistry* **1997**, *48* (1), 545–600. https://doi.org/10.1146/annurev.physchem.48.1.545.

(24) Huber, G. A.; Kim, S. Weighted-Ensemble Brownian Dynamics Simulations for Protein Association Reactions. *Biophysical Journal* **1996**, *70* (1), 97–110.

(25) Zuckerman, D. M.; Chong, L. T. Weighted Ensemble Simulation: Review of Methodology, Applications, and Software. *Annual Review of Biophysics* **2017**, *46*, 43–57. https://doi.org/10.1146/annurev-biophys-070816-033834.

(26) Zhang, B. W.; Jasnow, D.; Zuckerman, D. M. The "Weighted Ensemble" Path Sampling Method Is Statistically Exact for a Broad Class of Stochastic Processes and Binning Procedures. *The Journal of Chemical Physics* **2010**, *132* (5), 054107. https://doi.org/10.1063/1.3306345.

(27) Zwier, M. C.; Adelman, J. L.; Kaus, J. W.; Pratt, A. J.; Wong, K. F.; Rego, N. B.; Suárez, E.; Lettieri, S.; Wang, D. W.; Grabe, M.; Zuckerman, D. M.; Chong, L. T. WESTPA: An Interoperable, Highly Scalable Software Package for Weighted Ensemble Simulation and Analysis. *Journal of Chemical Theory and Computation* **2015**, *11* (2), 800–809. https://doi.org/10.1021/ct5010615.

(28) Götz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *Journal of Chemical Theory and Computation* **2012**, *8* (5), 1542–1555. https://doi.org/10.1021/ct200909j.

(29) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *Journal of Chemical Theory and Computation* **2013**, *9* (9), 3878–3888. https://doi.org/10.1021/ct400314y.

(30) Crowley, M. F.; Williamson, M. J.; Walker, R. C. CHAMBER: Comprehensive Support for CHARMM Force Fields within the AMBER Software. *International Journal of Quantum Chemistry* **2009**, *109* (15), 3767–3772. https://doi.org/10.1002/qua.22372.

(31) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *Journal of Computational Physics* **1977**, *23* (3), 327–341. https://doi.org/10.1016/0021-9991(77)90098-5.

(32) Chen, M.; Khalid, S.; Sansom, M. S. P.; Bayley, H. Outer Membrane Protein G: Engineering a Quiet Pore for Biosensing. *Proceedings of the National Academy of Sciences* **2008**, *105* (17), 6272–6277. https://doi.org/10.1073/pnas.0711561105.

(33) Durrant, J. D.; Amaro, R. E. LipidWrapper: An Algorithm for Generating Large-Scale Membrane Models of Arbitrary Geometry. *PLoS Computational Biology* **2014**, *10* (7). https://doi.org/10.1371/journal.pcbi.1003720.

(34) Ke, Z.; Oton, J.; Qu, K.; Cortese, M.; Zila, V.; McKeane, L.; Nakane, T.; Zivanov, J.; Neufeldt, C. J.; Cerikan, B.; Lu, J. M.; Peukes, J.; Xiong, X.; Kräusslich, H. G.; Scheres, S. H. W.; Bartenschlager, R.; Briggs, J. A. G. Structures and Distributions of SARS-CoV-2 Spike Proteins on Intact Virions. *Nature* **2020**, 1–7. https://doi.org/10.1038/s41586-020-2665-2.

(35) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *Journal of Computational Chemistry*. 2005. https://doi.org/10.1002/jcc.20289.

(36) Lee, E. H.; Hsin, J.; Sotomayor, M.; Comellas, G.; Schulten, K. Discovery through the Computational Microscope. *Structure* **2009**, *17*, 1295–1306.

(37) Shaw, D. E.; Deneroff, M. M.; Dror, R. O.; Kuskin, J. S.; Larson, R. H.; Salmon, J. K.; Young, C.; Batson, B.; Bowers, K. J.; Chao, J. C.; Eastwood, M. P.; Gagliardo, J.; Grossman, J. P.; Ho, C. R.; Ierardi, D. J.; Kolossváry, I.; Klepeis, J. L.; Layman, T.; McLeavey, C.; Moraes, M. A.; Mueller, R.; Priest, E. C.; Shan, Y.; Spengler, J.; Theobald, M.; Towles, B.; Wang, S. C. Anton, a Special-Purpose Machine for Molecular Dynamics Simulation. *SIGARCH Comput. Archit. News* **2007**, *35*, 1–12.

(38) Kalé, L. V.; Zheng, G. Chapter 1: The Charm++ Programming Model. In *Parallel Science and Engineering Applications: The Charm++ Approach*; Kale, L. V., Bhatele, A., Eds.; CRC Press, Inc.: Boca Raton, FL, USA, 2013; pp 1–16. https://doi.org/10.1201/b16251.

(39) Acun, B.; Hardy, D. J.; Kale, L.; Li, K.; Phillips, J. C.; Stone, J. E. Scalable Molecular Dynamics with NAMD on the Summit System. *IBM\ J.\ Res.\ Dev.* **2019**, *62*, 4:1-4:9.

(40) Phillips, J. C.; Stone, J. E.; Schulten, K. Adapting a Message-Driven Parallel Application to GPU-Accelerated Clusters. In *SC '08: Proceedings of the 2008 ACM/IEEE Conference on Supercomputing*; IEEE Press: Piscataway, NJ, USA, 2008; pp 1–9.

(41) Kumar, S.; Mamidala, A. R.; Faraj, D. A.; Smith, B.; Blocksome, M.; Cernohous, B.; Miller, D.; Parker, J.; Ratterman, J.; Heidelberger, P.; Chen, D.; Steinmacher-Burrow, B. PAMI: A Parallel Active Message Interface for the Blue Gene/Q Supercomputer. In *2012 IEEE 26th International Parallel and Distributed Processing Symposium*; 2012; pp 763–773. https://doi.org/10.1109/IPDPS.2012.73.

(42) Phillips, J. C.; Sun, Y.; Jain, N.; Bohm, E. J.; Kalé, L. V. Mapping to Irregular Torus Topologies and Other Techniques for Petascale Biomolecular Simulation. In *Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis*; SC '14; IEEE Press, 2014; pp 81–91. https://doi.org/10.1109/SC.2014.12.

(43) Fiorin, G.; Klein, M. L.; Hénin, J. Using Collective Variables to Drive Molecular Dynamics Simulations. *Mol. Phys.* **2013**, *111* (22–23), 3345–3362.

(44) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *Journal of Molecular Graphics* **1996**. https://doi.org/10.1016/0263-7855(96)00018-5.

(45) Stone, J. E.; Isralewitz, B.; Schulten, K. Early Experiences Scaling VMD Molecular Visualization and Analysis Jobs on Blue Waters. In *Extreme Scaling Workshop (XSW), 2013*; 2013; pp 43–50. https://doi.org/10.1109/XSW.2013.10.

(46) Stone, J. E.; Vandivort, K. L.; Schulten, K. GPU-Accelerated Molecular Visualization on Petascale Supercomputing Platforms. In *Proceedings of the 8th International Workshop on Ultrascale Visualization*; UltraVis '13; ACM: New York, NY, USA, 2013; p 6:1-6:8.

(47) Stone, J. E.; Sener, M.; Vandivort, K. L.; Barragan, A.; Singharoy, A.; Teo, I.; Ribeiro, J. V.; Isralewitz, B.; Liu, B.; Goh, B. C.; Phillips, J. C.; MacGregor-Chatwin, C.; Johnson, M. P.; Kourkoutis, L. F.; Hunter, C. N.; Schulten, K. Atomic Detail Visualization of Photosynthetic Membranes with GPU-Accelerated Ray Tracing. *Parallel Computing* **2016**, *55*, 17–27. https://doi.org/10.1016/j.parco.2015.10.015.

(48) Zhao, G.; Perilla, J. R.; Yufenyuy, E. L.; Meng, X.; Chen, B.; Ning, J.; Ahn, J.; Gronenborn, A. M.; Schulten, K.; Aiken, C.; Zhang, P. Mature HIV-1 Capsid Structure by Cryo-Electron Microscopy and All-Atom Molecular Dynamics. *Nature* **2013**, *497*, 643–646. https://doi.org/10.1038/nature12162.

(49) Durrant, J. D.; Kochanek, S. E.; Casalino, L.; Ieong, P. U.; Dommer, A. C.; Amaro, R. E. Mesoscale All-Atom Influenza Virus Simulations Suggest New Substrate Binding Mechanism. *ACS Central Science* **2020**, *6* (2), 189–196. https://doi.org/10.1021/acscentsci.9b01071.

(50) Singharoy, A.; Maffeo, C.; Delgardo, K.; Swainsbury, D. J. K.; Sener, M.; Kleinekathöfer, U.; Isralewitz, B.; Teo, I.; Chandler, D.; Stone, J.; Phillips, J.; Pogorelov, T.; Mallus, M. I.; Chipot, C.; Luthey-Schulten, Z.; Tieleman, P.; Hunter, C. N.; Tajkhorshid, E.; Aksimentiev, A.; Schulten, K. Atoms to Phenotypes: Molecular Design Principles of Cellular Energy Metabolism. *Cell* **2019**, *179*, 1098–1111.

(51) Saglam, A. S.; Chong, L. T. Protein–Protein Binding Pathways and Calculations of Rate Constants Using Fully-Continuous, Explicit-Solvent Simulations. *Chemical science* **2019**, *10* (8), 2360–2372.

(52) Noe, F.; Tkatchenko, A.; Müller, K.-R.; Clementi, C. Machine Learning for Molecular Simulation. *Annual Review of Physical Chemistry* **2020**, *71* (1), 361–390. https://doi.org/10.1146/annurev-physchem-042018-052331.

(53) Bernetti, M.; Bertazzo, M.; Masetti, M. Data-Driven Molecular Dynamics: A Multifaceted Challenge. *Pharmaceuticals* **2020**, *13* (9). https://doi.org/10.3390/ph13090253.

(54) Ramanathan, A.; Savol, A. J.; Burger, V. M.; Quinn, S.; Agarwal, P.; Chennubhotla, C. Statistical Inference for Big Data Problems in Molecular Biophysics; 2012.

(55) Bhowmik, D.; Gao, S.; Young, M. T.; Ramanathan, A. Deep Clustering of Protein Folding Simulations. *BMC Bioinformatics* **2018**, *19* (18), 484. https://doi.org/10.1186/s12859-018-2507-5.

(56) Romero, R.; Ramanathan, A.; Yuen, T.; Bhowmik, D.; Mathew, M.; Munshi, L. B.; Javaid, S.; Bloch, M.; Lizneva, D.; Rahimova, A.; Khan, A.; Taneja, C.; Kim, S.-M.; Sun, L.; New, M. I.; Haider, S.; Zaidi,

M. Mechanism of Glucocerebrosidase Activation and Dysfunction in Gaucher Disease Unraveled by Molecular Dynamics and Deep Learning. *Proceedings of the National Academy of Sciences* **2019**, *116* (11), 5086–5095. https://doi.org/10.1073/pnas.1818411116.

(57) Lee, H.; Turilli, M.; Jha, S.; Bhowmik, D.; Ma, H.; Ramanathan, A. DeepDriveMD: Deep-Learning Driven Adaptive Molecular Simulations for Protein Folding. In *2019 IEEE/ACM Third Workshop on Deep Learning on Supercomputers (DLS)*; 2019; pp 12–19.

(58) Allison, J. R. Computational Methods for Exploring Protein Conformations. *Biochemical Society Transactions* **2020**, *48* (4), 1707–1724. https://doi.org/10.1042/BST20200193.

(59) Bonati, L.; Zhang, Y.-Y.; Parrinello, M. Neural Networks-Based Variationally Enhanced Sampling. *Proceedings of the National Academy of Sciences* **2019**, *116* (36), 17641–17647. https://doi.org/10.1073/pnas.1907975116.

(60) Kasson, P. M.; Jha, S. Adaptive Ensemble Simulations of Biomolecules. *Current Opinion in Structural Biology* **2018**, *52*, 87–94. https://doi.org/10.1016/j.sbi.2018.09.005.

(61) Lamim Ribeiro, J. M.; Tiwary, P. Toward Achieving Efficient and Accurate Ligand-Protein Unbinding with Deep Learning and Molecular Dynamics through RAVE. *Journal of Chemical Theory and Computation* **2019**, *15* (1), 708–719. https://doi.org/10.1021/acs.jctc.8b00869.

(62) Ribeiro, J. M. L.; Bravo, P.; Wang, Y.; Tiwary, P. Reweighted Autoencoded Variational Bayes for Enhanced Sampling (RAVE). *The Journal of Chemical Physics* **2018**, *149* (7), 072301. https://doi.org/10.1063/1.5025487.

(63) Wang, Y.; Ribeiro, J. M. L.; Tiwary, P. Machine Learning Approaches for Analyzing and Enhancing Molecular Dynamics Simulations. *Current Opinion in Structural Biology* **2020**, *61*, 139–145.

(64) Wang, Y.; Ribeiro, J. M. L.; Tiwary, P. Past–Future Information Bottleneck for Sampling Molecular Reaction Coordinate Simultaneously with Thermodynamics and Kinetics. *Nature Communications* **2019**, *10* (1), 3573. https://doi.org/10.1038/s41467-019-11405-4.

(65) Pérez, A.; Herrera-Nieto, P.; Doerr, S.; De Fabritiis, G. AdaptiveBandit: A Multi-Armed Bandit Framework for Adaptive Sampling in Molecular Simulations. *Journal of Chemical Theory and Computation* **2020**, *16* (7), 4685–4693. https://doi.org/10.1021/acs.jctc.0c00205.

(66) Gonzalez-Arias, F.; Reddy, T.; Stone, J.; Hadden-Perilla, J.; Perilla, J. Scalable Analysis of Authentic Viral Envelopes on FRONTERA. *Computing in Science and Engineering* **2020**. https://doi.org/10.1109/MCSE.2020.3020508.

(67) Qi, C. R.; Su, H.; Mo, K.; Guibas, L. J. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. 2017.

(68) Zamorski, M.; Zięba, M.; Klukowski, P.; Nowak, R.; Kurach, K.; Stokowiec, W.; Trzciński, T. Adversarial Autoencoders for Compact Representations of 3D Point Clouds. *Computer Vision and Image Understanding* **2020**, *193*, 102921. https://doi.org/10.1016/j.cviu.2020.102921.

(69) Arjovsky, M.; Bottou, L. Towards Principled Methods for Training Generative Adversarial Networks. *arXiv preprint arXiv:1701.04862* **2017**.

(70) Gulrajani, I.; Ahmed, F.; Arjovsky, M.; Dumoulin, V.; Courville, A. Improved Training of Wasserstein GANs. 2017.

(71) van der Maaten, L. J. P.; Hinton, G. E. Visualizing High-Dimensional Data Using t-SNE. *Journal of Machine Learning Research* **2008**, *9* (nov), 2579–2605.

(72) Yang, C. Hierarchical Roofline Analysis: How to Collect Data Using Performance Tools on Intel CPUs and NVIDIA GPUs. 2020.

(73) Phillips, J.; Zheng, G.; Kumar, S.; Kale, L. NAMD: Biomolecular Simulation on Thousands of Processors. In *Proceedings of the IEEE/ACM SC2002 Conference, Technical Paper 277*; IEEE Press: Baltimore, Maryland, 2002; pp 1–18.

(74) Yao, H.; Song, Y.; Chen, Y.; Shi, Y.; Li, L.; Correspondence, S. L.; Wu, N.; Xu, J.; Sun, C.; Zhang, J.; Weng, T.; Zhang, Z.; Wu, Z.; Cheng, L.; Shi, D.; Lu, X.; Lei, J.; Crispin, M.; Li, S. Molecular Architecture of the SARS-CoV-2 Virus. *Cell* **2020**, *183*. https://doi.org/10.1016/j.cell.2020.09.018.

(75) Turoňová, B.; Sikora, M.; Schürmann, C.; Hagen, W. J. H.; Welsch, S.; Blanc, F. E. C.; von Bülow, S.; Gecht, M.; Bagola, K.; Hörner, C.; van Zandbergen, G.; Landry, J.; de Azevedo, N. T. D.; Mosalaganti, S.; Schwarz, A.; Covino, R.; Mühlebach, M. D.; Hummer, G.; Krijnse Locker, J.; Beck, M. In Situ Structural Analysis of SARS-CoV-2 Spike Reveals Flexibility Mediated by Three Hinges. *Science* **2020**, eabd5223. https://doi.org/10.1126/science.abd5223.

(76) Yu, A.; Pak, A. J.; He, P.; Monje-Galvan, V.; Casalino, L.; Gaieb, Z.; Dommer, A. C.; Amaro, R. E.; Voth, G. A. A Multiscale Coarse-Grained Model of the SARS-CoV-2 Virion. *bioRxiv* **2020**, 2020.10.02.323915. https://doi.org/10.1101/2020.10.02.323915.

Chapter 7

#COVIDisAirborne: AI-Enabled Multiscale Computational Microscopy of Delta SARS-CoV-2 in a Respiratory Aerosol

7.1 Abstract

We seek to completely revise current models of airborne transmission of respiratory viruses by providing never-before-seen atomic- level views of the SARS-CoV-2 virus within a respiratory aerosol. Our work dramatically extends the capabilities of multiscale computational microscopy to address the significant gaps that exist in current experimental methods, which are limited in their ability to interrogate aerosols at the atomic/molecular level and thus obscure our understanding of airborne transmission. We demonstrate how our integrated data-driven platform provides a new way of exploring the composition, structure, and dynamics of aerosols and aerosolized viruses, while driving simulation method development along several important axes. We present a series of initial scientific discoveries for the SARS-CoV-2 Delta variant, noting that the full scientific impact of this work has yet to be realized.

7.2 Justification

We develop a novel HPC-enabled multiscale research framework to study aerosolized viruses and the full complexity of species that comprise them. We present technological and methodological advances that bridge time and length scales from electronic structure through whole aerosol particle morphology and dynamics.

7.3 Performance Attributes

Table 7.1 Performance Attributes.

Performance Attribute	Our Submission
Category of achievement	Scalability, Time-to-solution
Type of method used	Explicit, Deep Learning
Results reported on the basis of	Whole application including I/O
Precision reported	Mixed Precision
System scale	Measured on full system
Measurement Mechanism	Hardware performance counters, Application timers, Performance Modeling

7.4 Overview of the Problem

Respiratory pathogens, such as SARS-CoV-2 and influenza, are the cause of significant morbidity and mortality worldwide. These respiratory pathogens are spread by virus-laden aerosols and droplets that are produced in an infected person, exhaled, and transported through the environment[1] (**Figure 7.1**). Medical dogma has long focused on droplets as the main transmission route for respiratory viruses, where either a person has contact with an infected surface (fomites) or direct droplet transmission by close contact with an infected individual. However, as we continue to observe with SARS-CoV-2, airborne transmission also plays a significant role in spreading disease. We know this from various super spreader events, e.g., during a choir rehearsal.[2] Intervention and mitigation decisions, such as the relative importance of surface cleaning or whether and when to wear a mask, have unfortunately hinged on a weak understanding of aerosol transmission, to the detriment of public health.

A central challenge to understanding airborne transmission has been the inability of experimental science to reliably probe the structure and dynamics of viruses once they are inside respiratory aerosol particles. Single particle experimental methods have poor resolution for smaller particles (<1 micron) and are prone to sample destruction during collection. Airborne viruses are present in low concentrations in the air and are similarly prone to viral inactivation during sampling. In addition, studies of the initial infection event, for example in the deep lung, are limited in their ability to provide a detailed understanding of the myriad of molecular interactions and dynamics taking place in situ. Altogether, these knowledge gaps hamper our collective ability to understand mechanisms of infection and develop novel effective antivirals, as well as prevent us from developing concrete, science-driven mitigation measures (e.g., masking and ventilation protocols).

Here, we aim to reconceptualize current models of airborne transmission of respiratory viruses by providing never-before-seen views of viruses within aerosols. Our approach relies on the use of all-atom molecular dynamics (MD) simulations as a multiscale 'computational microscope.' MD simulations can synthesize multiple types of biological data (e.g., multiresolution structural datasets, glycomics, lipidomics, etc.) into cohesive, biologically 'accurate' structural models. Once created, we then approximate the model down to its many atoms, creating trajectories of its time dependent dynamics under cell-like (or in this case, aerosol-like)

conditions. Critically, MD simulations are more than just 'pretty movies.' MD equations are solved in a theoretically rigorous manner, allowing us to compute experimentally testable macroscopic observables from time-averaged microscopic properties. What this means is that we can directly connect MD simulations with experiments, each vali- dating and providing testable hypotheses to the other, which is the real power of the approach. An ongoing challenge to the successful application of such methods, however, is the need for technological and methodological advances that make it possible to access length scales relevant to the study of large, biologically complex systems (spanning nanometers to one micron in size) and, correspondingly, longer timescales (microseconds to seconds).

Such challenges and opportunities manifest in the study of aerosolized viruses. Aerosols are generally defined as being less than 5 microns in diameter, able to float in the air for hours, travel significant distances (e.g., can fill a room, like cigarette smoke), and be inhaled. Fine aerosols < 1 micron in size can stay in the air for over 12 hours and are enriched with viral particles.[3,4] Our work focuses on these finer aerosols that travel deeper into the respiratory tract. Several studies provide the molecular recipes necessary to reconstitute respiratory aerosols according to their actual biologically-relevant composition.[5,6] These aerosols can contain lipids, cholesterol, albumin (protein), various mono- and di-valent salts, mucins, other surfactants, and water (**Figure 7.1**). Simulations of aerosolized viruses embody a novel framework for the study of aerosols: they will allow us and others to tune different species, relative humidity, ion concentrations, etc. to match experiments that can directly and indirectly connect to and inform our simulations, as well as test hypotheses. Some of the species under study here, e.g., mucins, have not yet been structurally characterized or explored with simulations and thus the models we generate are expected to have impact beyond their roles in aerosols.

In addition to varying aerosol composition and size, the viruses themselves can be modified to reflect new variants of concern, where such mutations may affect interactions with particular species in the aerosol that might affect its structural dynamics and/or vi- ability. The virion developed here is the Delta variant (B.1.617.2 lineage) of SARS-CoV-2 (**Figure 7.2**) which presents a careful integration of multiple biological datasets: (1) a complete viral envelope with realistic membrane composition, (2) fully glycosylated full-length spike proteins integrating 3D structural coordinates from multiple cryoelectron microscopy (cryo-EM) studies[7–10] (3) all biologically known features (post-translational modifications, palmitoylation, etc.), (4) any other known

membrane proteins (e.g., the envelope (E) and membrane (M) proteins), and (5) virion size and patterning taken directly from cryoelectron tomography (cryo-ET). Each of the individual components of the virus are built up before being integrated into the composite virion, and thus represent useful molecular-scale scientific contributions in their own right.[11,12]

Altogether in this work, we dramatically extend the capabilities of data-driven, multiscale computational microscopy to provide a new way of exploring the composition, structure, and dynamics of respiratory aerosols. While a seemingly limitless number of putative hypotheses could result from these investigations, the first set of questions we expect to answer are: How does the virus exist within a droplet of the same order of magnitude in size, without being affected by the air-water interface, which is known to destroy molecular structure[13] ? How does the biochemical composition of the droplet, including pH, affect the structural dynamics of the virus? Are there species within the aerosols that "buffer" the viral structure from damage, and are there particular conditions under which the impact of those species changes? Our simulations can also provide specific parameters that can be included in physical models of aerosols, which still assume a simple water or water-salt composition even though it is well known that such models, e.g., using kappa-Kohler theory, break down significantly as the molecular species diversify.[14]

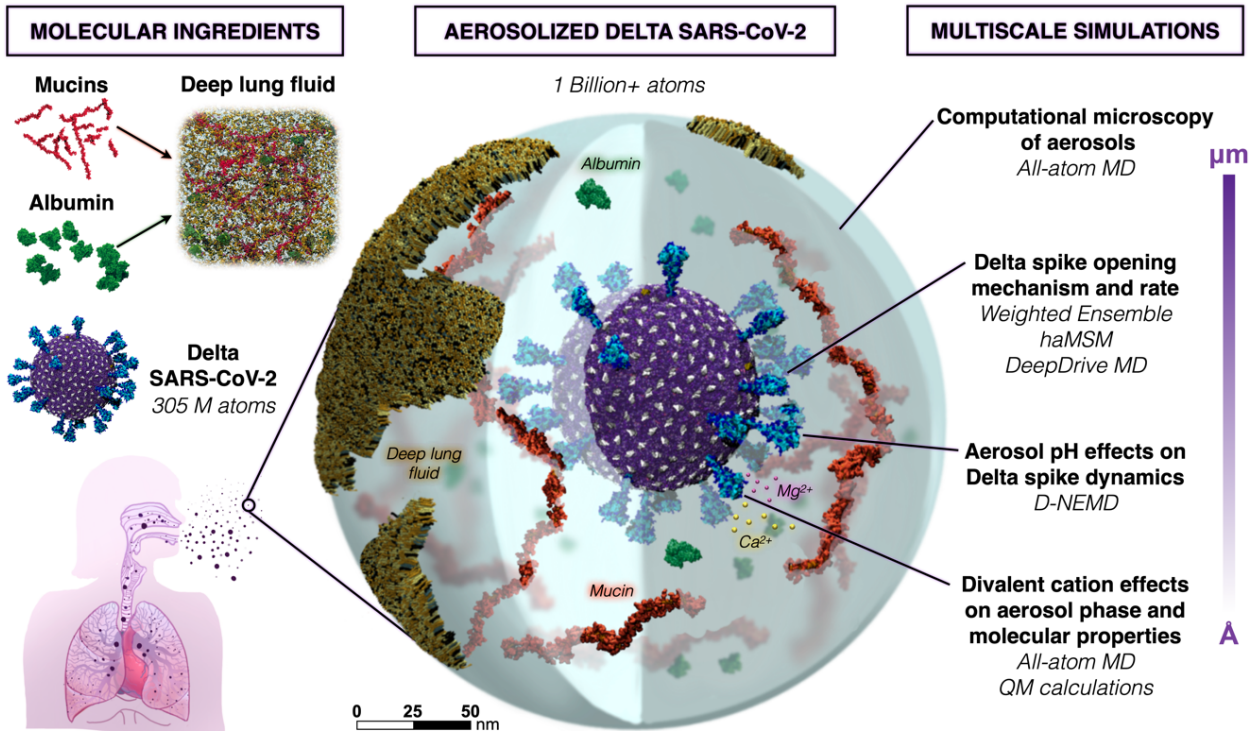


Figure 7.1: Overall schematic depicting the construction and multiscale simulations of Delta SARS-CoV-2 in a respiratory aerosol. (N.B.: The size of divalent cations has been increased for visibility.)

7.5 Current State of the Art

Current experimental methods are unable to directly interrogate the atomic-level structure and dynamics of viruses and other molecules within aerosols. Here we showcase computational microscopy as a powerful tool capable to overcome these significant experimental limitations. We present the major elements of our multiscale computational microscope and how they come together in an integrated manner to enable the study of aerosols across multiple scales of resolution. We demonstrate the impact such methods can bring to bear on scientific challenges that until now have been intractable and present a series of new scientific discoveries for SARS-CoV-2.

7.5.1 Parallel molecular dynamics

All-atom molecular dynamics simulation has emerged as an increasingly powerful tool for understanding the molecular mechanisms underlying biophysical behaviors in complex systems. Leading simulation engines, NAMD[15], AMBER[16], and GROMACS[17], are broadly useful, with each providing unique strengths in terms of

specific methods or capabilities as required to address a particular biological question, and in terms of their support for particular HPC hardware platforms. Within the multiscale computational microscopy platform developed here, we show how each of these different codes contributes different elements to the overall framework, oftentimes utilizing different computing modalities/architectures, while simultaneously extending on state-of-the-art for each. Structure building, simulation preparation, visualization, and post-hoc trajectory analysis are performed using VMD on both local workstations and remote HPC resources, enabling modeling of the molecular systems studied herein.[18–22] We show how further development of each of these codes, considered together within the larger-scale collective framework, enables the study of SARS-CoV-2 in a wholly novel manner, with extension to numerous other complex systems and diseases.

7.5.2 AI-enhanced WE simulations

Because the virulence of the Delta variant of SARS-CoV-2 may be partly attributable to spike protein (S) opening, it is of pressing interest to characterize the mechanism and kinetics of the process. Although S-opening in principle can be studied via conventional MD simulations, in practice the system complexity and timescales make this wholly intractable. Splitting strategies that periodically replicate promising MD trajectories, among them the weighted ensemble (WE) method[23,24], have enabled simulations of the spike opening of WT SARS-CoV-2.[12,25] WE simulations can be orders of magnitude more efficient than conventional MD in generating pathways and rate constants for rare events (e.g., protein folding[26] and binding[27]). The WESTPA software for running WE[28] is well-suited for high-performance computing with nearly perfect CPU/GPU scaling. The software is interoperable with any dynamics engine, including the GPU-accelerated AMBER dynamics engine[29] that is used here. As shown below, major upgrades to WESTPA (v. 2.0) have enabled a dramatic demonstration of spike opening in the Delta variant (Figures 7.5 and 7.6) and exponentially improved analysis of spike-opening kinetics.

The integration of AI techniques with WE can further enhance the efficiency of sampling rare events.[11,30,31] One frontier area couples unsupervised linear and non-linear dimensionality reduction methods to identify collective variables/progress coordinates in high-dimensional molecular systems.[32,33] Such methods may be well suited for analyzing the aerosolized virus. Integrating these approaches with WE simulations is

advantageous in sampling the closed open transitions in the Delta S landscape (Figure 7.5) as these unsupervised AI approaches automatically stratify progress coordinates (Figure 7.5D).

7.5.3 Dynamical Non-Equilibrium MD

Aerosols rapidly acidify during flight via reactive uptake of atmospheric gases, which is likely to impact the opening/closing of the S protein.[5,34] Here, we describe the extension of dynamical non-equilibrium MD (D-NEMD)[35] to investigate pH effects on the Delta S. D-NEMD simulations[35] are emerging as a useful technique for identifying allosteric effects and communication pathways in proteins[36,37], including recently identifying effects of linoleic acid in the WT spike.[38] This approach complements equilibrium MD simulations, which provide a distribution of configurations as starting points for an ensemble of short non-equilibrium trajectories under the effect of the external perturbation. The response of the protein to the perturbation introduced can then be determined using the Kubo-Onsager by directly tracking the change in atomic positions between the equilibrium and non-equilibrium simulations at equivalent points in time.

7.5.4 OrbNet

Ca^{2+} ions are known to play a key role in mucin aggregation in epithelial tissues.[39] Our RAV simulations would be an ideal case-study to probe such complex interactions between Ca^{2+}, mucins, and the SARS-CoV-2 virion in aerosols. However, Ca^{2+} binding energies can be difficult to capture accurately due to electronic dispersion and polarization, terms which are not typically modeled in classical mechanical force fields. Quantum mechanical (QM) methods are uniquely suited to capture these subtle interactions. Thus, we set out to estimate the correlation in Ca^{2+} binding energies between CHARMM36m and quantum mechanical estimates enabled via AI with OrbNet. Calculation of energies with sufficient accuracy in biological systems can, in many cases, be adequately described with density functional theory (DFT). However, its high cost limits the applicability of DFT in comparison to fixed charge force-fields. To capture quantum quality energetics at a fraction of the computational expense, we employ a novel approach (OrbNet) based on the featurization of molecules in terms of symmetry-adapted atomic orbitals and the use of graph neural network methods for deep-learning quantum-mechanical properties.[40] Our method outperforms existing methods in terms of its training efficiency and

transferable accuracy across diverse molecular systems, opening a new pathway for replacing DFT in large-scale scientific applications such as those explored here.[41]

7.6 Innovations Realized

7.6.1 Construction and simulation of SARS-CoV-2 in a respiratory aerosol

Our approach to simulating the entire aerosol follows a composite framework wherein each of the individual molecular pieces is refined and simulated on its own before it is incorporated into the composite model. Simulations of each of the components are useful in their own right, and often serve as the basis for biochemical and biophysical validation and experiments.[11]

Throughout, we refer to the original circulating SARS-CoV-2 strain as "WT", whereas all SARS-CoV-2 proteins constructed in this work represent the Delta variant (**Figure 7.2**) All simulated membranes reflect mammalian ER-Golgi intermediate compartment (ERGIC) mimetic lipid compositions. VMD[18,22], psfgen[42] and CHARMM- GUI[43] were used for construction and parameterization. Topologies and parameters for simulations were taken from CHARMM36m all-atom additive force fields.[44–50] NAMD was used to perform MD simulations,[15] adopting similar settings and protocols as in Casalino et al.[11] All systems underwent solvation, charge neutralization, minimization, heating, and equilibration prior to production runs. Refer to **Table 7.2** for abbreviations, PBC dimensions, total number of atoms, and total equilibration times for each system of interest.

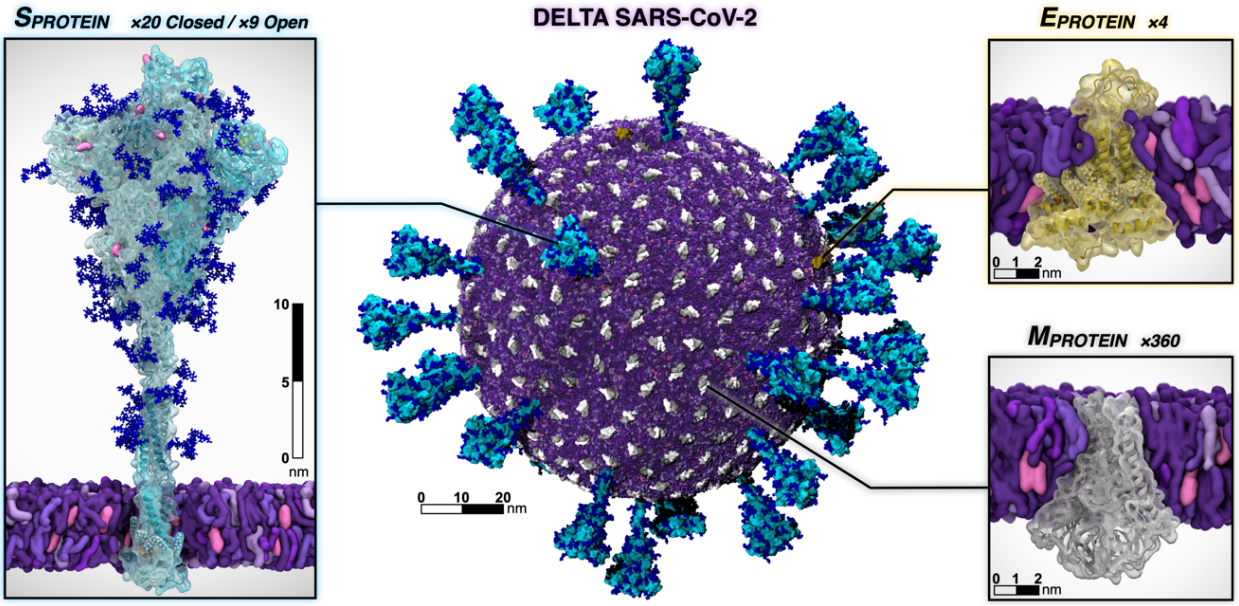


Figure 7.2: Individual protein components of the SARS-CoV-2 Delta virion. The spike is shown with the surface in cyan and with Delta's mutated residues and deletion sites highlighted in pink and yellow, respectively. Glycans attached to the spike are shown in blue. The E protein is shown in yellow and the M protein is shown in silver and white. Visualized with VMD.

7.6.1.1 Simulating the SARS-CoV-2 structural proteins.

Fully glycosylated Delta spike (S) structures in open and closed conformations were built based on WT constructs from Casalino et al.[11] with the following mutations: T19R, T95I, G142D, E156G, Δ157-158, L452R, T478K, D614G, P681R, and D950N.[8,51] Higher resolved regions were grafted from PDB 7JJI.[7] Additionally, coordinates of residues 128-167 accounting for a drastic conformational change seen in the Delta variant S (graciously made available to us by the Veesler Lab) were similarly grafted onto our constructs.[8] Finally, the S proteins were glycosylated following work by Casalino et al.[11] By incorporating the Veesler Lab's bleeding-edge structure and highly resolved regions from 7JJI, our models represent the most complete and accurate structures of the Delta S to date. The S proteins were inserted into membrane patches and equilibrated for 3x110 ns. For nonequilibrium and weighted ensemble simulations, a closed S head (SH, residues 13- 1140) was constructed by removing the stalk from the full-length closed S structure, then re-solvated, neutralized, minimized, and subsequently passed to WE and D-NEMD teams. The M protein was built from a structure graciously provided by the Feig Lab (paper in prep). The model was inserted into a membrane patch and equilibrated for 700 ns. RMSD-based clustering was used to select a stable starting M protein conformation.

From the equilibrated and clustered M structure, VMD's Mutator plugin[18] was used to incorporate the I82T mutation onto each M monomer to arrive at the Delta variant M. To construct the most complete E protein model to-date, the structure was patched together by re- solving incomplete PDBs 5X29,[52] 7K3G,[53] and 7M4R.[54] To do so, the transmembrane domain (residues 8-38) from 7K3G was aligned to the N-terminal domain (residues 1-7) and residues 39 to 68 of 5X29 and residues 69 to 75 of 7M4R by their Cα atoms. E was then inserted into a membrane patch and equilibrated for 40 ns.

7.6.1.2 Constructing the SARS-CoV-2 Delta virion.

The SARS-CoV-2 Delta virion (V) model was constructed following Casalino et al.[55] using CHARMM-GUI,[56] LipidWrapper,[57] and Blender,[58] using a 350 Å lipid bilayer with an equilibrium area per lipid of 63 Å^2 and a 100 nm diameter Blender icospherical surface mesh.[59] The resulting lipid membrane was solvated in a 1100 Å^3 waterbox and subjected to 4 rounds of equilibration and patching.[55] 360 M dimers and 4 E pentamers were then tiled onto the surface, followed by random placement of 29 full-length S proteins (9 open, 20 closed) according to experimentally observed S protein density.[60] M and E proteins were oriented with intravirion C-termini. After solvation in a 1460 Å waterbox, the complete V model tallied >305 million atoms (Table 8.2). V was equilibrated for 41 ns prior to placement in the respiratory aerosol (RA) model. The equilibrated membrane was 90 nm in diameter and remains in close structural agreement with the experimental studies.[60]

Table 7.2: Summary of all systems constructed in this work. See Figure 8.3 for illustration of aerosol construction.

[a]systems	[b]Abb.	[c](Å × Å × Å)	[d]N_a	[e](ns)
[f]M dimers	M	125 × 125 × 124	164,741	700
[f]E pentamers	E	123 × 125 × 102	136,775	41
Spikes:				
[f](open)	S	206 × 200 × 410	1,692,444	330
[f](closed)	S	204 × 202 × 400	1,658,224	330
[g](closed, head)	SH	172 × 184 × 206	615,593	73μs
Mucins:				
[f]short mucin 1	m_1	123 × 104 × 72	87,076	25
[f]short mucin 2	m_2	120 × 101 × 72	82,155	25
[f]long mucin 1	m_3	810 × 104 × 115	931,778	23
[f]long mucin 2	m_4	904 × 106 × 109	997,029	15
[f]long mucin 3	m_5	860 × 111 × 113	1,040,215	18
[f]S+m1/m2+ALB	SMA	227 × 229 × 433	2,156,689	840
[f]Virion	V	1460 × 1460 × 1460	305,326,834	41
[f]Resp.Aero.+Vir.	RAV	2834 × 2820 × 2828	1,016,813,441	2.42
TOTAL FLOPS			2.4 ZFLOPS	

[a]M, E, S, SH, and V models represent SARS-CoV-2 Delta strain. [b]Abbreviations used throughout document. [c]Periodic boundary dimensions. [d]Total number of atoms. [e]Total aggregate simulation time, including heating and equilibration runs. [f]Simulated with NAMD. [g]Simulated with NAMD, AMBER, and GROMACS.

7.6.1.3 Building and simulating the respiratory aerosol.

Respiratory aerosols contain a complex mixture of chemical and biological species. We constructed a respiratory aerosol (RA) fluid based on a composition from artificial saliva and surrogate deep lung fluid recipes.[6] This recipe includes 0.7 mM DPPG, 6.5 mM DPPC, 0.3 mM cholesterol, 1.4 mM Ca^{2+}, 0.8 mM Mg^{2+}, and 142 mM Na^{+}, human serum albumin (ALB) protein, and a composition of mucins (**Figure 7.3**) Mucins are long polymer-like structures that are decorated by dense, heterogeneous, and complex regions of O-glycans. This work represents the first of its kind as, due to their complexity, the O-glycosylated regions of mucins have never before been constructed for molecular simulations. Two short (m1, m2, 5 nm) and three long (m3, m4, m5 55 nm) mucin models were constructed following known experimental compositions of protein and glycosylation sequences[39,61–64] with ROSETTA[65] and CHARMM GUI Glycan Modeler.[43] Mucin models (short and long) were solvated, neutralized by charge matching with Ca^{2+} ions, minimized, and equilibrated for 15-25 ns each (**Table 7.2**). Human serum albumin (ALB), which is also found in respiratory aerosols, was constructed from PDB 1AO6.[66] ALB was solvated, neutralized, minimized, and equilibrated for 7ns. Equilibrated structures

of ALB and the three long mucins were used in construction of the RAV with m3+m4+m5 added at 6 g/mol and ALB at 4.4 g/mol.

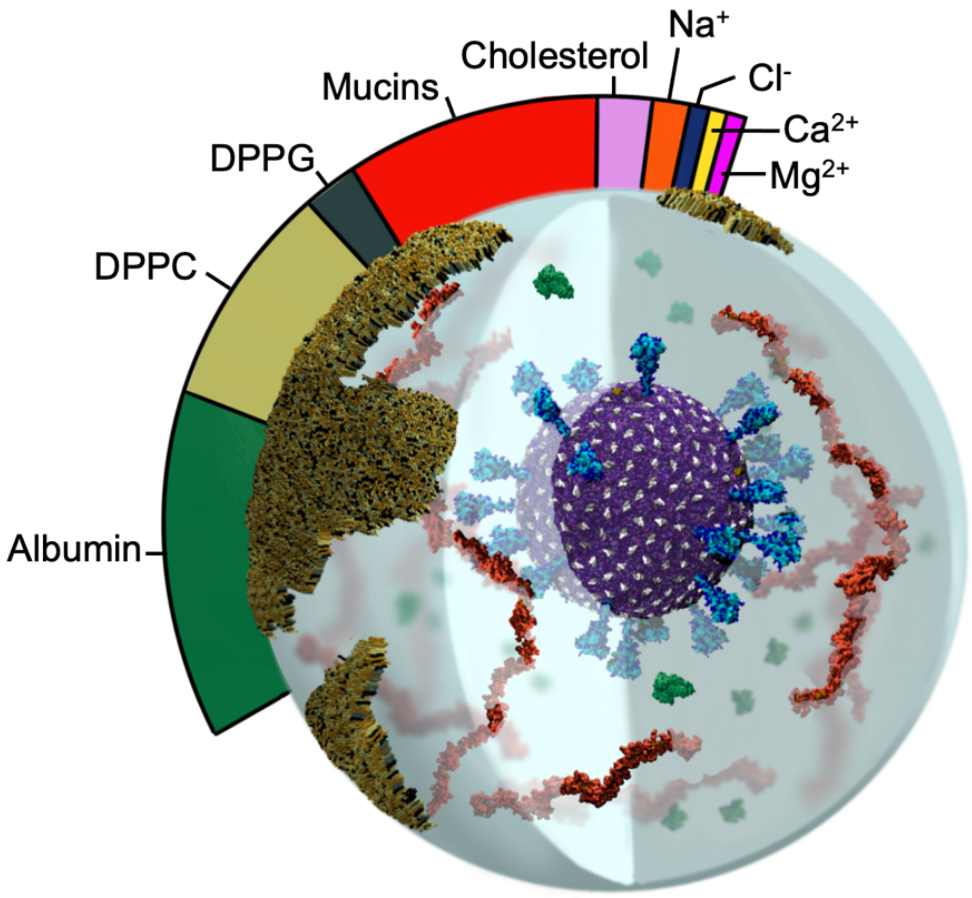


Figure 7.3: Image of RAV with relative mass ratios of RA molecular components represented in the color bar. Water content is dependent on the relative humidity of the environment and is thus omitted from the molecular ratios.

7.6.1.4 Constructing the respiratory aerosolized virion model.

A 100 nm cubic box with the RA fluid recipe specified above was built with PACKMOL,[67] minimized, equilibrated briefly on TACC Frontera, then replicated to form a 300 nm cube. The RA box was then carved into a 270 nm diameter sphere. To make space for the placement of V within the RA, a spherical selection with volume corresponding to that of the V membrane + S crown (radius 734 Å) was deleted from the center of the RA. The final equilibrated V model, including surrounding equilibrated waters and ions (733 Å radius), was translated into the RA. Atom clashes were resolved using a 1.2 Å cutoff. Hydrogen mass repartitioning[68] was applied to the structure to improve performance. The simulation box was increased to 2800 Å per side to provide

a 100 Å vacuum atmospheric buffer. The RAV simulation was conducted in an NVT ensemble with a 4 fs timestep. After minimizing, the RAV was heated to 298 K with 0.1 kcal/mol Å2 restraints on the viral lipid headgroups, then equilibrated for 1.5 ns. Finally, a cross-section of the RAV model – including and open S, m1/m2, and ALB (called the SMA system) – was constructed with PACKMOL to closely observe atomic scale interactions within the RAV model (**Figure 7.4**).

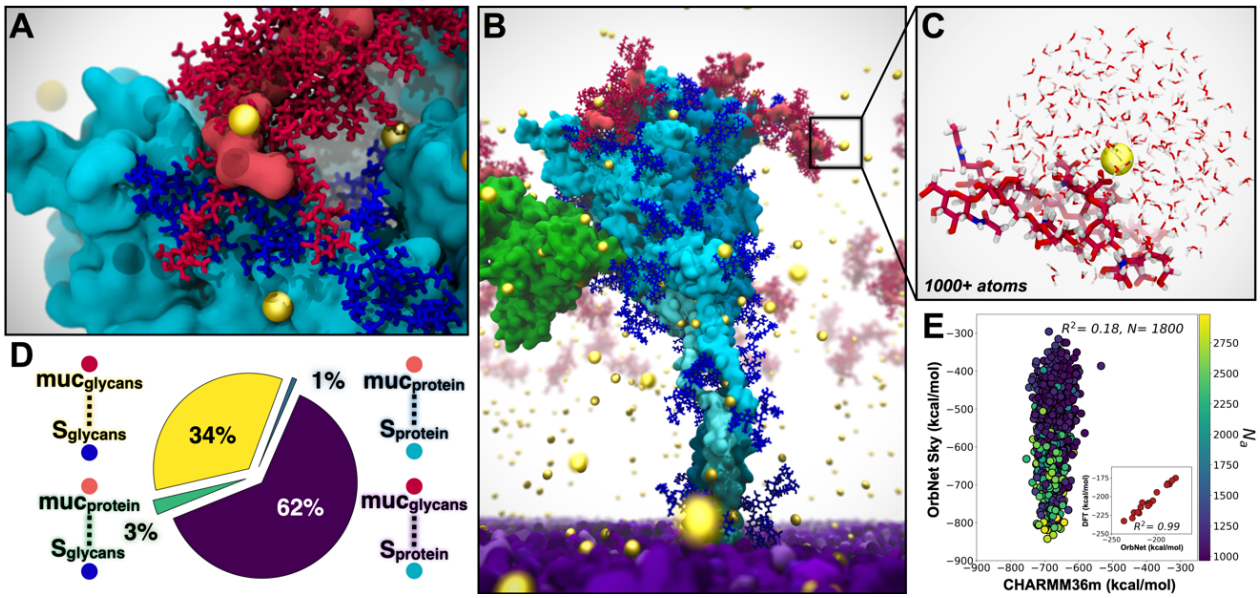


Figure 7.4: SMA system captured with multiscale modeling from classical MD to AI-enabled quantum mechanics. For all panels: S protein shown in cyan, S glycans in blue, m1/m2 shown in red, ALB in orange, Ca2+ in yellow spheres, viral membrane in purple. A) Interactions between mucins and S facilitated by glycans and Ca2+. B) Snapshot from SMA simulations. C) Ex- ample Ca2+ binding site from SMA simulations (1800 sites, each 1000+ atoms) used for AI-enabled quantum mechanical estimates from OrbNet Sky. D) Quantification of contacts between S and mucin from SMA simulations. E) OrbNet Sky energies vs CHARMM36m energies for each sub-selected system, colored by total number of atoms. Performance of OrbNet Sky vs. DFT in subplot (ωB97x-D3/def-TZVP, R2=0.99, for 17 systems of peptides chelating Ca2+. Visualized with VMD.

7.6.2 Parameter evaluation with OrbNet

Comparison to quantum methods reveals significant polarization effects and shows that there is opportunity to improve the accuracy of fixed charge force fields. For the large system sizes associated with solvated Ca^{2+}-protein interaction motifs (over 1000 atoms, even in aggressively truncated systems) conventional quantum mechanics methods like density functional theory (DFT) are im- practical for analyzing a statistically significant ensemble of distinct configurations (see discussion in Performance Results). In contrast, OrbNet allows for DFT accuracy with over 1000-fold speed-up, providing a useful method for benchmarking and

refining the force- field simulation parameters with quantum accuracy.[41] To confirm the accuracy of OrbNet versus DFT (ωB97X- D/def2-TZVP), the inset of **Figure 7.4E** correlates the two methods for the Ca2+-binding energy in a benchmark dataset of small Ca^{2+}- peptide complexes.[69] The excellent correlation of OrbNet and DFT for the present use case is clear from the inset figure; six datapoints were removed from this plot on the basis of a diagnostic applied to the semi-empirical GFN-xTB solution used for feature generation of OrbNet.

Figure 8.4E presents a comparison of the validated OrbNet method with the CHARMM36m force field for 1800 snapshots taken from the SMA MD simulations. At each snapshot, a subsystem containing a solvated Ca^{2+}-protein complex was extracted (**Figure 7.4E**) with protein bonds capped by hydrogens. For both OrbNet and the force field, the Ca^{2+}-binding energy was computed and shown in the correlation plot. Lack of correlation between OrbNet and the force field identifies important polarization effects, absent in a fixed charge description. Similarly, the steep slope of the best-fit line in **Figure 7.4E** reflects the fact that some of the configurations sampled using MD with the CHARMM36m force field are relatively high in energy according to the more accurate OrbNet potential. This approach allows us to test and quantify limitations of empirical force fields, such as lack of electronic polarization.

The practicality of OrbNet for these simulation snapshots with 1000+ atoms offers a straightforward multiscale strategy for refining the accuracy of the CHARMM36m force field. By optimizing the partial charges and other force field parameters, improved correlation with OrbNet for the subtle Ca^{2+}-protein interactions could be achieved, leading to near-quantum accuracy simulations with improved configurational sampling. The calculations presented here present a proof-of-concept of this iterative strategy.

7.6.3 AI-WE simulations of Delta spike opening

While our previous WE simulations of the WT SARS-CoV-2 S- opening[12] were notable in generating pathways for a seconds-timescale process of a massive system, we have made two critical technological advancements in the WESTPA software that greatly enhance the efficiency and analysis of WE simulations. These advances enabled striking observations of Delta-variant S opening (**Figures 7.5** and **7.6**). First, in contrast to prior manual bins for controlling trajectory replication, we have developed automated and adaptive binning that enables more efficient surmounting of large barriers via early identification of “bottleneck” regions.[70]

Second, we have parallelized, memory-optimized, and implemented data streaming for the history-augmented Markov state model (haMSM) analysis scheme[71] to enable application to the TB-scale S-opening datasets. The haMSM approach estimates rate constants from simulations that have not yet reached a steady state.[72]

Our WE simulations generated >800 atomically detailed, Delta- variant S-opening pathways (**Figures 7.5B** and **7.6**) of the receptor binding domain (RBD) switching from a glycan-shielded 'down' to an ex- posed 'up' state using 72 μs of total simulation time within 14 days using 192 NVIDIA V100 GPUs at a time on TACC's Longhorn supercomputer. Among these pathways, 83 reach an 'open' state that aligns with the structure of the human ACE2-bound WT S protein[73] and 18 reach a dramatically open state (**Figure 7.6**). Our haMSM analysis of WT WE simulations successfully provided long-timescale (steady-state) rate constants for S-opening based on highly transient information (**Figure 7.5C**).

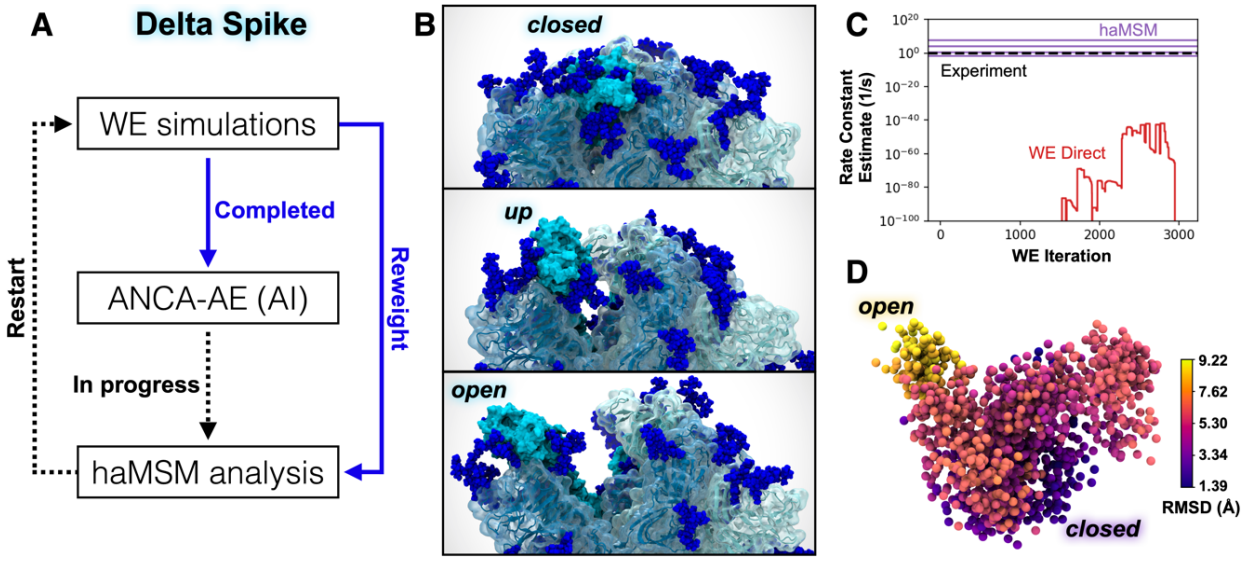


Figure 7.5: Delta-variant spike opening from WE simulations, and AI/haMSM analysis. A) The integrated workflow. B) Snapshots of the 'down', 'up', and 'open' states for Delta S-opening from a representative pathway generated by WE simulation, which represents 105 speedup compared to conventional MD. C) Rate-constant estimation with haMSM analysis of WE data (pur- ple lines) significantly improves direct WE computation (red), by comparison to experimental measurement (black dashed). Varying haMSM estimates result from different featurizations which will be individually cross-validated. D) The first three dimensions of the ANCA-AE embeddings depict a clear separation between the closed (darker purple) and open (yellow) con- formations of the Delta spike. A sub-sampled landscape is shown here where each sphere represents a conformation from the WE simulations and colored with the root-mean squared deviations (Å) with respect to the closed state. Visualized with VMD.

We also leveraged a simple, yet powerful unsupervised deep learning method called Anharmonic Conformational Analysis enabled Autoencoders (ANCA-AE)[33] to extract conformational states from our long-timescale WE simulations of Delta spike opening (**Figure 7.5A,D**). ANCA-AE first minimizes the fourth order correlations in atomistic fluctuations from MD simulation datasets and projects the data onto a low dimensional space where one can visualize the anharmonic conformational fluctuations. These projections are then input to an autoencoder that further minimizes non-linear correlations in the atomistic fluctuations to learn an embedding where conformations are automatically clustered based on their structural and energetic similarity. A visualization of the first three dimensions from the latent space articulates the RBD opening motion from its closed state (**Figure 7.5D**). It is notable that while other deep learning techniques need special purpose hardware (such as GPUs), the ANCA-AE approach can be run with relatively modest CPU resources and can therefore scale to much larger systems (e.g., the virion within aerosol) when optimized.

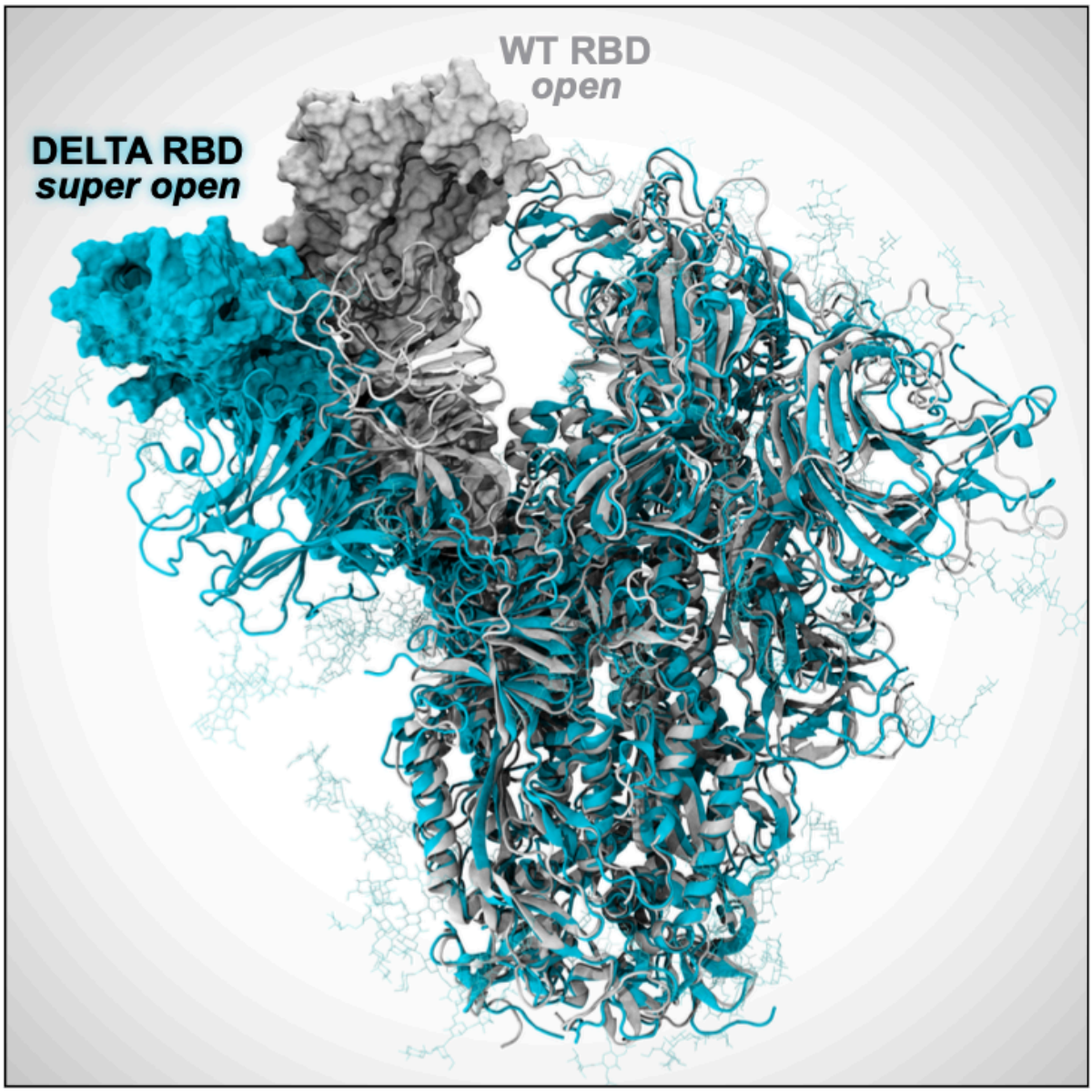


Figure 7.6: WE simulations reveal a dramatic opening of the Delta S (cyan), compared to WT S (white). While further investigation is needed, this super open state seen in the Delta S may indicate increased capacity for binding to human host-cell receptors.

7.6.4 D-NEMD explores pH effects on Delta spike

We performed D-NEMD simulations of the SH system with GROMACS[74] using a ΔpH=2.0 (from 7.0 to 5.0) as the external perturbation. We ran 3 200-ns equilibrium MD simulations of SH to generate 87 configurations (29 configurations per replicate) that were used as the starting points for multiple short (10 ns) D-NEMD trajectories under the effect of the external perturbation (ΔpH=2.0). The effect of a ΔpH was modelled

by changing the protonation state of histidines 66, 69, 146, 245, 625, 655, 1064, 1083, 1088, and 1101 (we note that other residues may also become protonated[75] ; the D-NEMD approach can also be applied to examine those). The structural response of the S to the pH decrease was investigated by measuring the difference in the position for each Cα atom between the equilibrium and corresponding D-NEMD simulation at equivalent points in time[38], namely after 0, 0.1, 1, 5 and 10 ns of simulation. The D-NEMD simulations reveal that pH changes, of the type expected in aerosols, affect the dynamics of functionally important regions of the spike, with potential implications for viral behavior (**Figure 7.7**). As this approach involves multiple short independent non-equilibrium trajectories, it is well suited for cloud computing. All D-NEMD simulations were performed using Oracle Cloud.

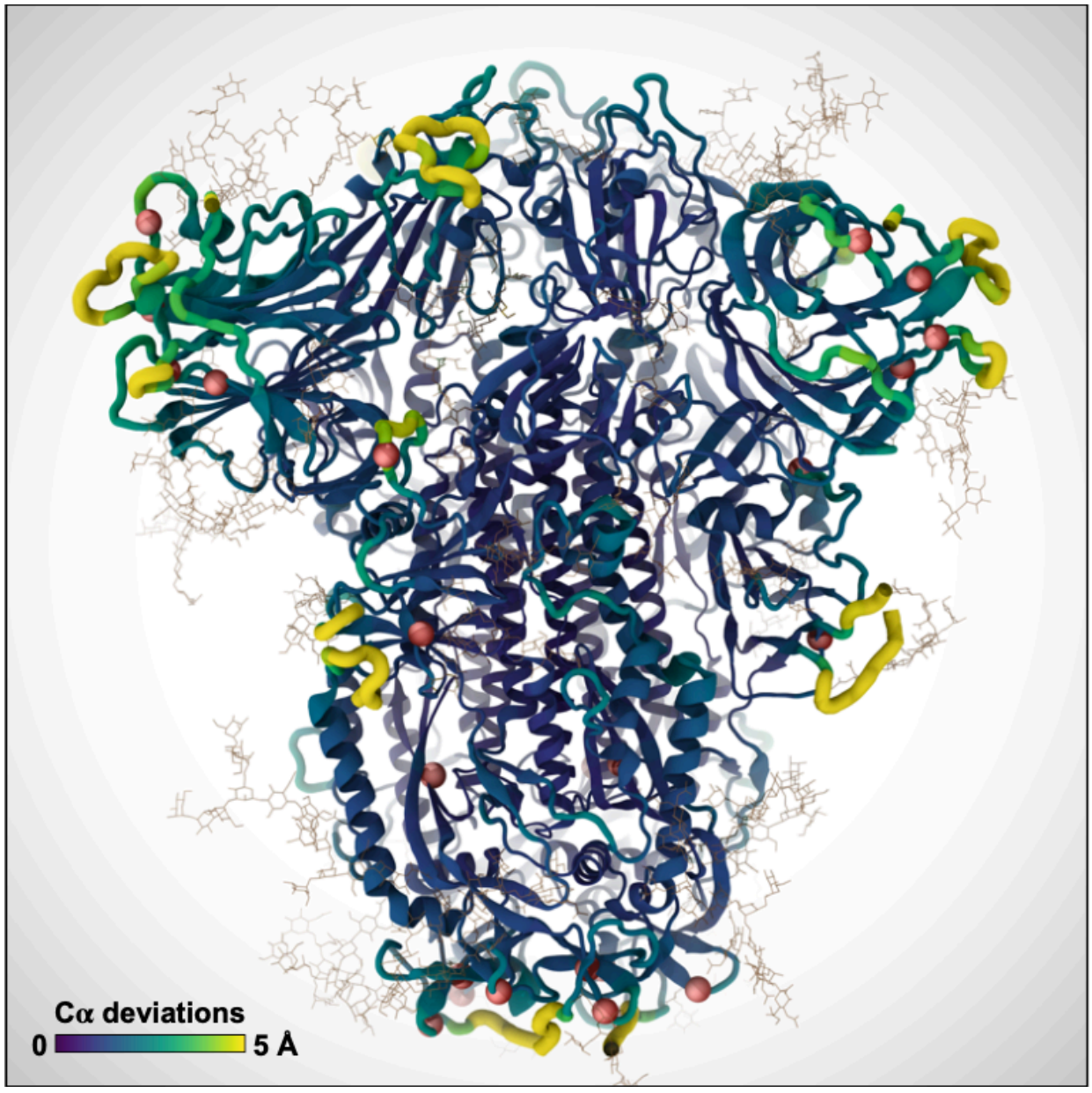


Figure 7.7: D-NEMD simulations reveal changes in key functional regions of the S protein, including the receptor binding domain, as the result of a pH decrease. Color scale and ribbon thickness indicate the degree of deviation of C*α* atoms from their equilibrium position. Red spheres indicate the location of positively charged histidines.

7.7 How performance was measured

7.7.1 WESTPA

For the WE simulations of spike opening using WESTPA, we defined the time to solution as the total simulation time required to generate the first spike opening event. Spike opening is essentially impossible to observe via conventional MD. WESTPA simulations were run using the AMBER20 dynamics engine and 192 NVIDIA V100 GPUs at a time on TACC's Longhorn supercomputer.

7.7.2 NAMD

NAMD performance metrics were collected using hardware performance counters for FLOPs/step measurements, and application- internal timers for overall simulation rates achieved by production runs including all I/O for simulation trajectory and checkpoint output. NAMD FLOPs/step measurements were conducted on TACC Frontera, by querying hardware performance counters with the rdmsr utility from Intel msr-tools and the "TACC stats" system programs. For each simulation, FLOP counts were measured for NAMD simulation runs of two different step counts. The results of the two simulation lengths were subtracted to eliminate NAMD startup operations, yielding an accurate estimate of the marginal FLOPs per step for a continuing simulation.[76] Using the FLOPs/step values computed for each simulation, overall FLOP rates were computed by dividing the FLOPs/step value by seconds/step performance data reported by NAMD internal application timers during production runs.

Table 7.3: MD simulation floating point ops per timestep.

MD Simulation	Code	Atoms	[a]FLOPs/step
Spike, head	AMBER, GROMACS	0.6M	62.14 GFLOPs/step
Spike	NAMD	1.7M	43.05 GFLOPs/step
S+m_1/m_2+ALB	NAMD	2.1M	54.86 GFLOPs/step
Resp. Aero.+Vir.	NAMD	1B	25.81 TFLOPs/step

[a] FLOPs/step data were computed by direct FLOP measurements from hardware performance counters for NAMD simulations, or by using the application-reported FLOP rates and ns/day simulation performance in the case of GROMACS.

7.7.3 GROMACS

GROMACS 2020.4 benchmarking was performed on Oracle Cloud Infrastructure (OCI) compute shape BM.GPU4.8 consisting of 8 NVIDIA A100 tensor core GPUs, and 64 AMD Rome CPU cores. The simulation

used for benchmarking contained 615,563 atoms and was run for 500,000 steps with 2 fs time steps. The simulations were run on increasing numbers of GPUs, from 1 to 8, using 8 CPU cores per GPU, running for both the production (Nose-Hoover) and GPU- accelerated (velocity rescaling) thermostats. Particle–mesh Ewald (PME) calculations were pinned to a single GPU, with additional GPUs for multi-GPU jobs used for particle–particle calculations. Performance data (ns/day and average single-precision TFLOPS, calculated as total number of TFLOPs divided by total job walltime) were reported by GROMACS itself. Each simulation was repeated four times and average performance figures reported.

7.8 Performance

7.8.1 NAMD performance

NAMD was used to perform all of the simulations listed in **Table 7.2**, except for the closed spike (SH) simulations described further below. With the exception of the aerosol and virion simulation, the other NAMD simulations used conventional protocols and have performance and parallel scaling characteristics that closely match the results reported in our previous SARS-CoV-2 research.[55] NAMD 2.14 scaling performance for the one billion- atom respiratory aerosol and virion simulation run on ORNL Summit is summarized in **Tables 7.4** and **8.5**. A significant performance challenge associated with the aerosol virion simulation relates to the roughly 50% reduction in particle density as compared with a more conventional simulation with a fully populated periodic cell. The reduced particle density results in large regions of empty space that nevertheless incur additional overheads associated with both force calculations and integration, and creates problems for the standard NAMD load balancing scheme that estimates the work associated with the cubic “patches” used for parallel domain decomposition. The PME electrostatics algorithm and associated 3-D FFT and transpose operations encompass the entire simulation unit cell and associated patches, requiring involvement in communication and reduction operations despite the inclusion of empty space. Enabling NAMD diagnostic output on a 512-node 1B-atom aerosol and virion simulation revealed that ranks assigned empty regions of the periodic cell had 66 times the number of fixed-size patches as ranks assigned dense regions. The initial load estimate for an empty patch was changed from a fixed 10 atoms to a runtime parameter with a default of 40 atoms, which reduced the patch ratio from 66 to 19 and doubled performance on 512 nodes.

Table 7.4: NAMD performance: Respiratory Aerosol + Virion, 1B atoms, 4 fs timestep w/ HMR, and PME every 3 steps.

Nodes	Summit CPU + GPU	Speedup	Efficiency
256	4.18 ns/day	~1.0×	~100%
512	7.68 ns/day	1.84×	92%
1024	13.64 ns/day	3.27×	81%
2048	23.10 ns/day	5.53×	69%
4096	34.21 ns/day	8.19×	51%

Table 7.5: Peak NAMD FLOP rates, ORNL Summit.

NAMD Simulation	Atoms	Nodes	Sim rate	Performance
Resp. Aero.+Vir.	1B	4096	34.21 ns/day	2.55 PFLOPS

7.8.2 WESTPA performance

Our time to solution for WE simulations of spike opening (to the "up" state) (**Figure 7.5**) using the WESTPA software and AMBER20 was 14 μs of total simulation time, which was completed in 4 days using 192 NVIDIA V100 GPUs at a time on TACC's Longhorn supercomputer. For reference, conventional MD would require an expected 5 orders of magnitude more computing. The WESTPA software is highly scalable, with nearly perfect scaling out to >1000 NVIDIA V100 GPUs and this scaling is expected to continue until the filesystem is saturated. Thus, WESTPA makes optimal use of large supercomputers and is limited by filesystem I/O due to the periodic restarting of trajectories after short time intervals.

7.8.3 AI-enhanced WE simulations

DeepDriveMD is a framework to coordinate the concurrent execution of ensemble simulations and drive them using AI models.[77,78] DeepDriveMD has been shown to improve the scientific performance of diverse problems: from protein folding to conformation of protein-ligand complexes. We coupled WESTPA to DeepDriveMD, which is responsible for resource dynamism and concurrent heterogeneous task execution (ML and AMBER). The coupled workflow was executed on 1024 nodes on Summit (OLCF), and, in spite of the spatio-temporal heterogeneity of tasks involved, the resource utilization was in the high 90%. Consistent with earlier studies, the coupling of WESTPA to DeepDriveMD results in a 100x improvement in the exploration of phase space.

7.8.4 GROMACS performance

Figure 8.8 shows GROMACS parallelizes well across the 8 NVIDIA A100 GPUs available on each BM.GPU4.8 instance used in the *Cluster in the Cloud* running on OCI. There is a performance drop for two GPUs due to inefficient division of the PME and particle-particle tasks. Methods to address this exist for the two GPU case[17], but were not adopted as we were targeting maximum raw performance across all 8 GPUs. Production simulations achieved 27% of the peak TFLOPS available from the GPUs. Multiple simulations were run across 10 such compute nodes, enabling the ensemble to run at an average combined speed of 425 TFLOPS and sampling up to 1μs/day. We note that the calculations will be able to run 20%–40% faster once the Nose-Hoover thermostat that is required for the simulation is ported to run on the GPU. Benchmarking using a velocity rescaling thermostat that has been ported to GPU shows that this would enable the simulation to extract 34% of the peak TFLOPS from the cards, enabling each node to achieve an average speed of 53.4 TFLOPS, and 125 ns/day. A cluster of 10 nodes would enable GROMACS to run at an average combined speed of over 0.5 PFLOPs, simulating over 1.2 μs/day.

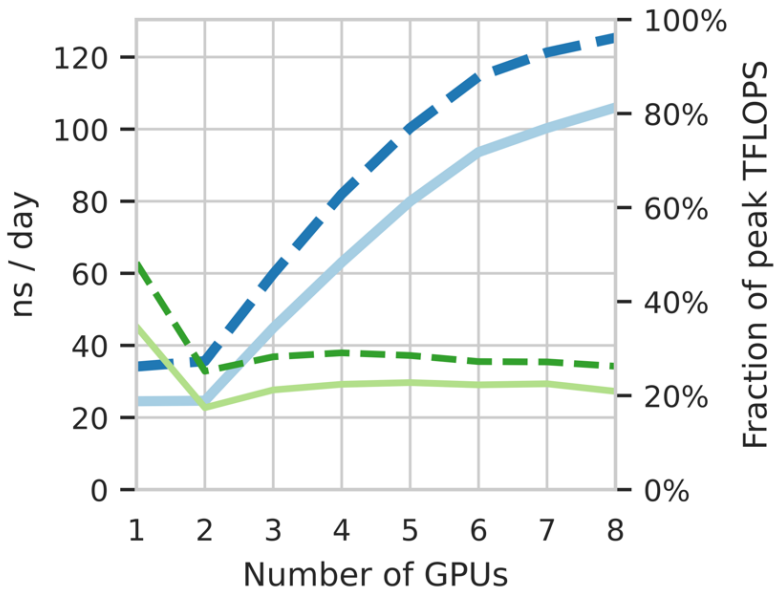


Figure 7.8: GROMACS performance across 1–8 A100 GPUs in ns/day (thicker, blue lines) and the fraction of maximum theoretical TFLOPS (thinner, green lines); production setup shown with solid line and runs with the GPU-accelerated thermostat in dashed.

A significant innovation is that this power is available on demand: Cluster in the Cloud with GPU-optimized GROMACS was provisioned and benchmarked within one day of inception of the project. This was handed to the researcher, who submitted the simulations. Automatically, up to ten BM.GPU4.8 compute nodes were provisioned on-demand based on requests from the Slurm scheduler. These simulations were performed on OCI, using *Cluster in the Cloud* to manage automatic scaling.

Cluster in the Cloud was configured to dynamically provision and terminate computing nodes based on the workload. Simulations were conducted using GROMACS 2020.4 compiled with CUDA sup- port. Multiple simultaneous simulations were conducted, with each simulation utilizing a single BM.GPU4.8 node without multi-node parallelism.

This allowed all production simulations to be completed within 2 days. The actual compute cost of the project was less than $6125 USD (on-demand OCI list price). The huge reduction in “time to science” that low-cost cloud enables changes the way that researchers can access and use HPC facilities. In our opinion, such a setup enables “exclusive on-demand” HPC capabilities for the scientific community for rapid advancement in science.

7.8.5 OrbNet performance

Prior benchmarking reveals that OrbNet provides over 1000-fold speedup compared to DFT. For the calculations presented here, the cost of corresponding high-quality range-separated DFT calculations (wB97X-D/def2-TZVP) can be estimated. In **Figure 7.4E**, we consider system sizes which would require 14,000–47,000 atomic orbitals for wB97X-D/def2-TZVP, exceeding the range of typical DFT evaluations. Estimation of the DFT computational cost of the 1811 configurations studied in **Figure 7.4E** suggests a total of 115M core-hours on NERSC Cori Haswell nodes; in contrast, the OrbNet calculations for the current study require only 100k core-hours on the same nodes. DFT cost estimates were based on extrapolation from a dataset of over 1M ChEMBL molecules ranging in size from 40 to 107 atom systems considering only the cubic cost component of DFT.[41]

7.9 Implications

Our major scientific achievements are:

(1) We showcase an extensible AI-enabled multiscale computational framework that bridges time and length scales from electronic structure through whole aerosol particle morphology and dynamics.

(2) We develop all-atom simulations of respiratory mucins and use these to understand the structural basis of interaction with the SARS-CoV-2 spike protein. This has implications for viral binding in the deep lung, which is coated with mucins. We expect the impact of our mucin simulations to be far reaching, as malfunctions in mucin secretion and folding have been implicated in progression of severe diseases such as cancer and cystic fibrosis.

(3) We present a significantly enhanced all-atom model and simulation of the SARS-CoV-2 Delta virion, which includes the hundreds of tiled M-protein dimers and the E-protein ion channels. This model can be used as a basis to understand why the Delta virus is so much more infectious than the WT or alpha variants.

(4) We develop an ultra-large (1 billion+) all-atom simulation capturing massive chemical and biological complexity within a respiratory aerosol. This simulation provides the first atomic level views of virus-laden aerosols and is already serving as a basis to develop an untold number of experimentally testable

hypotheses. An immediate example suggests a mechanism through which mucins and other species, e.g., lipids, which are present in the aerosol, arrange to protect the molecular structure of the virus, which otherwise would be exposed to the air-water interface. This work also opens the door for developing simulations of other aerosols, e.g., sea spray aerosols, that are involved in regulating climate.

(5) We evidence how changes in pH, which are expected in the aerosol environment, may alter dynamics and allosteric communication pathways in key functional regions of the Delta spike protein.

(6) We characterize atomically detailed pathways for the spike- opening process of the Delta variant using WE simulations, revealing a dramatically open state that may facilitate binding to human host cells.

(7) We demonstrate how parallelized haMSM analysis of WE data can provide physical rate estimates of spike opening, improving prior estimates by many orders of magnitude. The pipeline can readily be applied to the any variant spike protein or other complex systems of interest.

(8) We show how HPC and cloud resources can be used to significantly drive down time-to-solution for major scientific efforts as well as connect researchers and greatly enable complex collaborative interactions.

(9) We demonstrate how AI coupled to HPC at multiple levels can result in significantly improved effective performance, e.g., with AI-driven WESTPA, and extend the reach and do- main of applicability of tools ordinarily restricted to smaller, less complex systems, e.g., with OrbNet.

(10) While our work provides a successful use case, it also exposes weaknesses in the HPC ecosystem in terms of support for key steps in large/complex computational science campaigns. We find lack of widespread support for high performance remote visualization and interactive graphical sessions for system preparation, debugging, and analysis with diverse science tools to be a limiting factor in such efforts.

7.10 Acknowledgments

We thank Prof. Kim Prather for inspiring and informative discussions about aerosols and for her commitment to convey the air- borne nature of SARS-CoV-2. We thank D. Veesler for sharing the Delta spike NTD coordinates in advance of publication. We thank B. Messer, D. Maxwell, and the Oak Ridge Leadership

Computing Facility at Oak Ridge National Laboratory supported by the DOE under Contract DE-AC05-00OR22725. We thank the Texas Advanced Computing Center Frontera team, especially D. Stanzione and T. Cockerill, and for compute time made available through a Director's Discretionary Allocation (NSF OAC-1818253). We thank the Argonne Leadership Computing Facility supported by the DOE under DE-AC02-06CH11357. We thank the Pittsburgh Supercomputer Center for providing priority queues on Bridges-2 through the XSEDE allocation NSF TG-CHE060063. We thank N. Kern and J. Lee of the CHARMM-GUI support team for help converting topologies between NAMD and GROMACS. We thank J. Copperman, G. Simpson, D. Aristoff, and J. Leung for valuable discussions and sup- port from NIH grant GM115805. NAMD and VMD are funded by NIH P41-GM104601. This work was supported by the NSF Center for Aerosol Impacts on Chemistry of the Environment (CAICE), National Science Foundation Center for Chemical Innovation (NSF CHE-1801971), as well as NIH GM132826, NSF RAPID MCB-2032054, an award from the RCSA Research Corp., a UC San Diego Moore's Cancer Center 2020 SARS-CoV-2 seed grant, to R.E.A. This work was also supported by Oracle Cloud credits and related resources provided by the Oracle for Research program.

Chapter 7, in full, has been accepted for publication and is presented as it will appear in "Dommer, A., [†] Casalino, L.[†], and Kearns, F.[†]; Rosenfeld, M.; Wauer, N.; Ahn, S-H.; Russo, J.; Oliveira, S.; Morris, C.; Bogetti, A.; Trifan, A.; Brace, A.; Sztain, T.; Clyde, A.; Ma, H.; Chennubhotla, C.; Lee, H.; Turilli, M.; Khalid, S.; Tamayo-Mendoza, T.; Welborn, M.; Christensen, A.; Smith, D.; Qiao, Z.; Sirumalla, S.; O'Connor, M.; Manby, F.; Anandkumar, A.; Hardy, D.; Phillips, J.; Stern, A.; Romero, J.; Clark, D.; Dorrell, M.; Maiden, T.; Huang, L.; McCalpin, J.; Woods, C.; Williams, M.; Barker, B.; Rajapaksha, H.; Pitts, R.; Gibbs, T.; Stone, J.; Zuckerman, D.; Mulholland, A.; Miller, T.; Jha, S.; Ramanathan, A.; Chong, L.; Amaro, R. *#COVIDisAirborne: AI-Enabled Multiscale Computational Microscopy of Delta SARS-CoV-2 in a Respiratory Aerosol.* Int. J. High Perform. Comput. Appl. 2022." The dissertation author was a primary co-investigator and co-author of this work.

References

(1) Wang, C. C.; Prather, K. A.; Sznitman, J.; Jimenez, J. L.; Lakdawala, S. S.; Tufekci, Z.; Marr, L. C. Airborne Transmission of Respiratory Viruses. *Science* **2021**, *373* (6558), eabd9149. https://doi.org/10.1126/science.abd9149.

(2) Miller, S. L.; Nazaroff, W. W.; Jimenez, J. L.; Boerstra, A.; Buonanno, G.; Dancer, S. J.; Kurnitski, J.; Marr, L. C.; Morawska, L.; Noakes, C. Transmission of SARS-CoV-2 by Inhalation of Respiratory Aerosol in the Skagit Valley Chorale Superspreading Event. *Indoor Air* **2021**, *31* (2), 314–323. https://doi.org/10.1111/ina.12751.

(3) Coleman, K. K.; Tay, D. J. W.; Sen Tan, K.; Ong, S. W. X.; Son, T. T.; Koh, M. H.; Chin, Y. Q.; Nasir, H.; Mak, T. M.; Chu, J. J. H.; Milton, D. K.; Chow, V. T. K.; Tambyah, P. A.; Chen, M.; Wai, T. K. Viral Load of SARS-CoV-2 in Respiratory Aerosols Emitted by COVID-19 Patients While Breathing, Talking, and Singing. *Clinical Infectious Diseases* **2021**. https://doi.org/10.1093/cid/ciab691.

(4) Fennelly, K. P. Particle Sizes of Infectious Aerosols: Implications for Infection Control. *The Lancet Respiratory Medicine* **2020**, *8* (9), 914–924. https://doi.org/10.1016/S2213-2600(20)30323-4.

(5) Vejerano, E. P.; Marr, L. C. Physico-Chemical Characteristics of Evaporating Respiratory Fluid Droplets. *Journal of the Royal Society Interface* **2018**, *15* (139), 1–10. https://doi.org/10.1098/rsif.2017.0939.

(6) Walker, J. S.; Archer, J.; Gregson, F. K. A.; Michel, S. E. S.; Bzdek, B. R.; Reid, J. P. Accurate Representations of the Microphysical Processes Occurring during the Transport of Exhaled Aerosols and Droplets. *ACS Central Science* **2021**. https://doi.org/10.1021/acscentsci.0c01522.

(7) Bangaru, S.; Ozorowski, G.; Turner, H. L.; Antanasijevic, A.; Huang, D.; Wang, X.; Torres, J. L.; Diedrich, J. K.; Tian, J.-H.; Portnoff, A. D.; Patel, N.; Massare, M. J.; Yates, J. R.; Nemazee, D.; Paulson, J. C.; Glenn, G.; Smith, G.; Ward, A. B. Structural Analysis of Full-Length SARS-CoV-2 Spike Protein from an Advanced Vaccine Candidate. *Science* **2020**, *370* (6520), 1089–1094. https://doi.org/10.1126/science.abe1502.

(8) McCallum, M.; Walls, A. C.; Sprouse, K. R.; Bowen, J. E.; Rosen, L.; Dang, H. V.; deMarco, A.; Franko, N.; Tilles, S. W.; Logue, J.; Miranda, M. C.; Ahlrichs, M.; Carter, L.; Snell, G.; Pizzuto, M. S.; Chu, H. Y.; Van Voorhis, W. C.; Corti, D.; Veesler, D. Molecular Basis of Immune Evasion by the Delta and Kappa SARS-CoV-2 Variants. *bioRxiv* **2021**. https://doi.org/10.1101/2021.08.11.455956.

(9) Walls, A. C.; Park, Y. J.; Tortorici, M. A.; Wall, A.; McGuire, A. T.; Veesler, D. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **2020**, *181* (2), 281-292.e6. https://doi.org/10.1016/j.cell.2020.02.058.

(10) Wrapp, D.; Wang, N.; Corbett, K. S.; Goldsmith, J. A.; Hsieh, C.-L.; Abiona, O.; Graham, B. S.; McLellan, J. S. Cryo-EM Structure of the 2019-NCoV Spike in the Prefusion Conformation. *Science* **2020**, *367* (6483), 1260–1263. https://doi.org/10.1126/science.abb2507.

(11) Casalino, L.; Gaieb, Z.; Goldsmith, J. A.; Hjorth, C. K.; Dommer, A. C.; Harbison, A. M.; Fogarty, C. A.; Barros, E. P.; Taylor, B. C.; Mclellan, J. S.; Fadda, E.; Amaro, R. E. Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. *ACS Central Science* **2020**. https://doi.org/10.1021/acscentsci.0c01056.

(12) Sztain, T.; Ahn, S.-H.; Bogetti, A. T.; Casalino, L.; Goldsmith, J. A.; Seitz, E.; McCool, R. S.; Kearns, F. L.; Acosta-Reyes, F.; Maji, S.; Mashayekhi, G.; McCammon, J. A.; Ourmazd, A.; Frank, J.; McLellan, J. S.; Chong, L. T.; Amaro, R. E. A Glycan Gate Controls Opening of the SARS-CoV-2 Spike Protein. *Nature Chemistry 2021 13:10* **2021**, *13* (10), 963–968. https://doi.org/10.1038/s41557-021-00758-3.

(13) D'Imprima, E.; Floris, D.; Joppe, M.; Sánchez, R.; Grininger, M.; Kühlbrandt, W. Protein Denaturation at the Air-Water Interface and How to Prevent It. *eLife* **2019**, *8*, e42747. https://doi.org/10.7554/eLife.42747.

(14) Petters, M. D.; Kreidenweis, S. M. A Single Parameter Representation of Hygroscopic Growth and Cloud Condensation Nucleus Activity. *Atmospheric Chemistry and Physics* **2007**, *7* (8), 1961–1971. https://doi.org/10.5194/acp-7-1961-2007.

(15) Phillips, J. C.; Hardy, D. J.; Maia, J. D. C.; Stone, J. E.; Ribeiro, J. V.; Bernardi, R. C.; Buch, R.; Fiorin, G.; Hénin, J.; Jiang, W.; McGreevy, R.; Melo, M. C. R.; Radak, B.; Skeel, R. D.; Singharoy, A.; Wang, Y.; Roux, B.; Aksimentiev, A.; Luthey-Schulten, Z.; Kalé, L. V.; Schulten, K.; Chipot, C.; Tajkhorshid, E. Scalable Molecular Dynamics on CPU and GPU Architectures with NAMD. *J. Chem.\ Phys.* **2020**, *153*, 044130. https://doi.org/10.1063/5.0014475.

(16) Case, D.; Ben-Shalom, I.; Brozell, S.; article is an open \ldots, D. C.-\ldots Switzerland. T. AMBER 2018; University of California: San Francisco, CA, USA, 2018.© 2019 by the Authors.

(17) Páll, S.; Zhmurov, A.; Bauer, P.; Abraham, M.; Lundborg, M.; Gray, A.; Hess, B.; Lindahl, E. Heterogeneous Parallelization and Acceleration of Molecular Dynamics Simulations in GROMACS. *The Journal of Chemical Physics* **2020**, *153* (13), 134110. https://doi.org/10.1063/5.0018516.

(18) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *Journal of Molecular Graphics* **1996**. https://doi.org/10.1016/0263-7855(96)00018-5.

(19) Sener, M.; Levy, S.; Stone, J. E.; Christensen, A. J.; Isralewitz, B.; Patterson, R.; Borkiewicz, K.; Carpenter, J.; Hunter, C. N.; Luthey-Schulten, Z.; Cox, D. Multiscale Modeling and Cinematic Visualization of Photosynthetic Energy Conversion Processes from Electronic to Cell Scales. *Parallel Computing* **2021**, 102698.

(20) Stone, J. E.; Isralewitz, B.; Schulten, K. Early Experiences Scaling VMD Molecular Visualization and Analysis Jobs on Blue Waters. In *Extreme Scaling Workshop (XSW), 2013*; 2013; pp 43–50. https://doi.org/10.1109/XSW.2013.10.

(21) Stone, J. E.; Vandivort, K. L.; Schulten, K. GPU-Accelerated Molecular Visualization on Petascale Supercomputing Platforms. In *Proceedings of the 8th International Workshop on Ultrascale Visualization*; UltraVis '13; ACM: New York, NY, USA, 2013; p 6:1-6:8.

(22) Stone, J. E.; Sener, M.; Vandivort, K. L.; Barragan, A.; Singharoy, A.; Teo, I.; Ribeiro, J. V.; Isralewitz, B.; Liu, B.; Goh, B. C.; Phillips, J. C.; MacGregor-Chatwin, C.; Johnson, M. P.; Kourkoutis, L. F.; Hunter, C. N.; Schulten, K. Atomic Detail Visualization of Photosynthetic Membranes with GPU-Accelerated Ray Tracing. *Parallel Computing* **2016**, *55*, 17–27. https://doi.org/10.1016/j.parco.2015.10.015.

(23) Huber, G. A.; Kim, S. Weighted-Ensemble Brownian Dynamics Simulations for Protein Association Reactions. *Biophysical Journal* **1996**, *70* (1), 97–110.

(24) Zuckerman, D. M.; Chong, L. T. Weighted Ensemble Simulation: Review of Methodology, Applications, and Software. *Annual Review of Biophysics* **2017**, *46*, 43–57. https://doi.org/10.1146/annurev-biophys-070816-033834.

(25) Zimmerman, M. I.; Porter, J. R.; Ward, M. D.; Singh, S.; Vithani, N.; Meller, A.; Mallimadugula, U. L.; Kuhn, C. E.; Borowsky, J. H.; Wiewiora, R. P.; al, et. SARS-CoV-2 Simulations Go Exascale to Predict Dramatic Spike Opening and Cryptic Pockets across the Proteome. *Nature Chemistry* **2021**, *13* (7), 651–659. https://doi.org/10.1038/s41557-021-00707-0.

(26) Adhikari, U.; Mostofian, B.; Copperman, J.; Petersen, A.; Zuckerman, D. M. Computational Estimation of Ms-Sec Atomistic Folding Times. *Journal of the American Chemical Society* **2019**, 6519–6526. https://doi.org/10.1101/427393.

(27) Saglam, A. S.; Chong, L. T. Protein–Protein Binding Pathways and Calculations of Rate Constants Using Fully-Continuous, Explicit-Solvent Simulations. *Chemical science* **2019**, *10* (8), 2360–2372.

(28) Zwier, M. C.; Adelman, J. L.; Kaus, J. W.; Pratt, A. J.; Wong, K. F.; Rego, N. B.; Suárez, E.; Lettieri, S.; Wang, D. W.; Grabe, M.; Zuckerman, D. M.; Chong, L. T. WESTPA: An Interoperable, Highly Scalable Software Package for Weighted Ensemble Simulation and Analysis. *Journal of Chemical Theory and Computation* **2015**, *11* (2), 800–809. https://doi.org/10.1021/ct5010615.

(29) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *Journal of Chemical Theory and Computation* **2013**, *9* (9), 3878–3888. https://doi.org/10.1021/ct400314y.

(30) Brace, A.; Lee, H.; Ma, H.; Trifan, A.; Turilli, M.; Yakushin, I.; Munson, T.; Foster, I.; Jha, S.; Ramanathan, A. Achieving 100X Faster Simulations of Complex Biological Phenomena by Coupling ML to HPC Ensembles. 2021.

(31) Noe, F.; Tkatchenko, A.; Müller, K.-R.; Clementi, C. Machine Learning for Molecular Simulation. *Annual Review of Physical Chemistry* **2020**, *71* (1), 361–390. https://doi.org/10.1146/annurev-physchem-042018-052331.

(32) Bhowmik, D.; Gao, S.; Young, M. T.; Ramanathan, A. Deep Clustering of Protein Folding Simulations. *BMC Bioinformatics* **2018**, *19* (18), 484. https://doi.org/10.1186/s12859-018-2507-5.

(33) Clyde, A.; Galanie, S.; Kneller, D. W.; Ma, H.; Babuji, Y.; Blaiszik, B.; Brace, A.; Brettin, T.; Chard, K.; Chard, R.; Coates, L.; Foster, I.; Hauner, D.; Kertesz, V.; Kumar, N.; Lee, H.; Li, Z.; Merzky, A.; Schmidt, J. G.; Tan, L.; Titov, M.; Trifan, A.; Turilli, M.; Van Dam, H.; Chennubhotla, S. C.; Jha, S.; Kovalevsky, A.; Ramanathan, A.; Head, M. S.; Stevens, R. High Throughput Virtual Screening and Validation of a SARS-CoV-2 Main Protease Non-Covalent Inhibitor. *bioRxiv* **2021**. https://doi.org/10.1101/2021.03.27.437323.

(34) Warwicker, J. A Model for PH Coupling of the SARS-CoV-2 Spike Protein Open/Closed Equilibrium. *Briefings in Bioinformatics* **2021**, *22* (2), 1499–1507. https://doi.org/10.1093/bib/bbab056.

(35) Ciccotti, G.; Ferrario, M. Non-Equilibrium by Molecular Dynamics: A Dynamical Approach. *Molecular Simulation* **2016**, *42* (16), 1385–1400. https://doi.org/10.1080/08927022.2015.1121543.

(36) Galdadas, I.; Qu, S.; Oliveira, A. S. F.; Olehnovics, E.; Mack, A. R.; Mojica, M. F.; Agarwal, P. K.; Tooke, C. L.; Gervasio, F. L.; Spencer, J.; Bonomo, R. A.; Mulholland, A. J.; Haider, S. Allosteric Communication in Class A β-Lactamases Occurs via Cooperative Coupling of Loop Dynamics. *eLife* **2021**, *10*, e66567. https://doi.org/10.7554/eLife.66567.

(37) Oliveira, A. S. F.; Shoemark, D. K.; Campello, H. R.; Wonnacott, S.; Gallagher, T.; Sessions, R. B.; Mulholland, A. J. Identification of the Initial Steps in Signal Transduction in the A4β2 Nicotinic Receptor: Insights from Equilibrium and Nonequilibrium Simulations. *Structure* **2019**, *27* (7), 1171-1183.e3. https://doi.org/10.1016/j.str.2019.04.008.

(38) Oliveira, A. S. F.; Ciccotti, G.; Haider, S.; Mulholland, A. J. Dynamical Nonequilibrium Molecular Dynamics Reveals the Structural Basis for Allostery and Signal Propagation in Biomolecular Systems. *The European Physical Journal B* **2021**, *94* (7), 144. https://doi.org/10.1140/epjb/s10051-021-00157-0.

(39) Hughes, G. W.; Ridley, C.; Collins, R.; Roseman, A.; Ford, R.; Thornton, D. J. The MUC5B Mucin Polymer Is Dominated by Repeating Structural Motifs and Its Topology Is Regulated by Calcium and PH. *Scientific Reports* **2019**, *9* (1), 17350. https://doi.org/10.1038/s41598-019-53768-0.

(40) Qiao, Z.; Welborn, M.; Anandkumar, A.; Manby, F. R.; Miller, T. F. OrbNet: Deep Learning for Quantum Chemistry Using Symmetry-Adapted Atomic-Orbital Features. *The Journal of Chemical Physics* **2020**, *153* (12), 124111. https://doi.org/10.1063/5.0021955.

(41) Christensen, A. S.; Sirumalla, S. K.; Qiao, Z.; O'Connor, M. B.; Smith, D. G. A.; Ding, F.; Bygrave, P. J.; Anandkumar, A.; Welborn, M.; Manby, F. R.; au2, T. F. M. I. OrbNet Denali: A Machine Learning Potential for Biological and Organic Chemistry with Semi-Empirical Cost and DFT Accuracy. 2021.

(42) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *Journal of Computational Chemistry*. 2005. https://doi.org/10.1002/jcc.20289.

(43) Park, S. J.; Lee, J.; Qi, Y.; Kern, N. R.; Lee, H. S.; Jo, S.; Joung, I.; Joo, K.; Lee, J.; Im, W. CHARMM-GUI Glycan Modeler for Modeling and Simulation of Carbohydrates and Glycoconjugates. *Glycobiology* **2019**. https://doi.org/10.1093/glycob/cwz003.

(44) Beglov, D.; Roux, B. Finite Representation of an Infinite Bulk System: Solvent Boundary Potential for Computer Simulations. *The Journal of Chemical Physics* **1994**. https://doi.org/10.1063/1.466711.

(45) Guvench, O.; Hatcher, E.; Venable, R. M.; Pastor, R. W.; MacKerell, A. D. CHARMM Additive All-Atom Force Field for Glycosidic Linkages between Hexopyranoses. *Journal of Chemical Theory and Computation* **2009**, *5* (9), 2353–2370. https://doi.org/10.1021/ct900242e.

(46) Han, K.; Venable, R. M.; Bryant, A.-M.; Legacy, C. J.; Shen, R.; Li, H.; Roux, B.; Gericke, A.; Pastor, R. W. Graph–Theoretic Analysis of Monomethyl Phosphate Clustering in Ionic Solutions. *The Journal of Physical Chemistry B* **2018**, *122* (4), 1484–1494. https://doi.org/10.1021/acs.jpcb.7b10730.

(47) Huang, J.; Mackerell, A. D. CHARMM36 All-Atom Additive Protein Force Field: Validation Based on Comparison to NMR Data. *Journal of Computational Chemistry* **2013**, *34* (25), 2135–2145. https://doi.org/10.1002/jcc.23354.

(48) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D. CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nature methods* **2017**, *14* (1), 71–73. https://doi.org/10.1038/nmeth.4067.

(49) Klauda, J. B.; Venable, R. M.; Freites, J. A.; O'Connor, J. W.; Tobias, D. J.; Mondragon-Ramirez, C.; Vorobyov, I.; MacKerell, A. D.; Pastor, R. W. Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. *The Journal of Physical Chemistry B* **2010**, *114* (23), 7830–7843. https://doi.org/10.1021/jp101759q.

(50) Venable, R. M.; Luo, Y.; Gawrisch, K.; Roux, B.; Pastor, R. W. Simulations of Anionic Lipid Membranes: Development of Interaction-Specific Ion Parameters and Validation Using NMR Data. *The Journal of Physical Chemistry B* **2013**, *117* (35), 10183–10192. https://doi.org/10.1021/jp401512z.

(51) Kannan, S. R.; Spratt, A. N.; Cohen, A. R.; Naqvi, S. H.; Chand, H. S.; Quinn, T. P.; Lorson, C. L.; Byrareddy, S. N.; Singh, K. Evolutionary Analysis of the Delta and Delta Plus Variants of the SARS-CoV-2 Viruses. *Journal of Autoimmunity* **2021**, *124*, 102715. https://doi.org/10.1016/j.jaut.2021.102715.

(52) Surya, W.; Li, Y.; Torres, J. Structural Model of the SARS Coronavirus E Channel in LMPG Micelles. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2018**, *1860* (6), 1309–1317. https://doi.org/10.1016/j.bbamem.2018.02.017.

(53) Mandala, V. S.; McKay, M. J.; Shcherbakov, A. A.; Dregni, A. J.; Kolocouris, A.; Hong, M. Structure and Drug Binding of the SARS-CoV-2 Envelope Protein Transmembrane Domain in Lipid Bilayers. *Nature Structural and Molecular Biology* **2020**, *27* (12), 1202–1208.

(54) Chai, J.; Cai, Y.; Pang, C.; Wang, L.; McSweeney, S.; Shanklin, J.; Liu, Q. Structural Basis for SARS-CoV-2 Envelope Protein Recognition of Human Cell Junction Protein PALS1. *Nature Communications* **2021**, *12* (1), 3433. https://doi.org/10.1038/s41467-021-23533-x.

(55) Casalino, L.; Dommer, A. C.; Gaieb, Z.; Barros, E. P.; Sztain, T.; Ahn, S.-H.; Trifan, A.; Brace, A.; Bogetti, A. T.; Clyde, A.; Ma, H.; Lee, H.; Turilli, M.; Khalid, S.; Chong, L. T.; Simmerling, C.; Hardy, D. J.; Maia, J. D.; Phillips, J. C.; Kurth, T.; Stern, A. C.; Huang, L.; McCalpin, J. D.; Tatineni, M.; Gibbs, T.; Stone, J. E.; Jha, S.; Ramanathan, A.; Amaro, R. E. AI-Driven Multiscale Simulations Illuminate Mechanisms of SARS-CoV-2 Spike Dynamics. *The International ournal of High Performance Computing Applications* **2021**, *35* (5), 432–451. https://doi.org/10.1177/10943420211006452.

(56) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM. *Journal of computational chemistry* **2008**, *29*, 2967–2970. https://doi.org/10.1002/jcc.

(57) Durrant, J. D.; Amaro, R. E. LipidWrapper: An Algorithm for Generating Large-Scale Membrane Models of Arbitrary Geometry. *PLoS Computational Biology* **2014**, *10* (7). https://doi.org/10.1371/journal.pcbi.1003720.

(58) Blender Online Community. Blender - a 3D Modelling and Rendering Package. Blender Foundation: Blender Institute, Amsterdam 2020.

(59) Turoňová, B.; Sikora, M.; Schürmann, C.; Hagen, W. J. H.; Welsch, S.; Blanc, F. E. C.; von Bülow, S.; Gecht, M.; Bagola, K.; Hörner, C.; van Zandbergen, G.; Landry, J.; de Azevedo, N. T. D.; Mosalaganti, S.; Schwarz, A.; Covino, R.; Mühlebach, M. D.; Hummer, G.; Krijnse Locker, J.; Beck, M. In Situ

Structural Analysis of SARS-CoV-2 Spike Reveals Flexibility Mediated by Three Hinges. *Science* **2020**, eabd5223. https://doi.org/10.1126/science.abd5223.

(60) Ke, Z.; Oton, J.; Qu, K.; Cortese, M.; Zila, V.; McKeane, L.; Nakane, T.; Zivanov, J.; Neufeldt, C. J.; Cerikan, B.; Lu, J. M.; Peukes, J.; Xiong, X.; Kräusslich, H. G.; Scheres, S. H. W.; Bartenschlager, R.; Briggs, J. A. G. Structures and Distributions of SARS-CoV-2 Spike Proteins on Intact Virions. *Nature* **2020**, 1–7. https://doi.org/10.1038/s41586-020-2665-2.

(61) Mariethoz, J.; Alocci, D.; Gastaldello, A.; Horlacher, O.; Gasteiger, E.; Rojas-Macias, M.; Karlsson, N. G.; Packer, N. H.; Lisacek, F. Glycomics@ExPASy: Bridging the Gap*. *Molecular and Cellular Proteomics* **2018**, *17* (11), 2164–2176. https://doi.org/10.1074/mcp.RA118.000799.

(62) Symmes, B. A.; Stefanski, A. L.; Magin, C. M.; Evans, C. M. Role of Mucins in Lung Homeostasis: Regulated Expression and Biosynthesis in Health and Disease. *Biochemical Society Transactions* **2018**, *46* (3), 707–719. https://doi.org/10.1042/BST20170455.

(63) Thomsson, K. A.; Schulz, B. L.; Packer, N. H.; Karlsson, N. G. MUC5B Glycosylation in Human Saliva Reflects Blood Group and Secretor Status. *Glycobiology* **2005**, *15* (8), 791–804. https://doi.org/10.1093/glycob/cwi059.

(64) Markovetz, M. R.; Subramani, D. B.; Kissner, W. J.; Morrison, C. B.; Garbarine, I. C.; Ghio, A.; Ramsey, K. A.; Arora, H.; Kumar, P.; Nix, D. B.; Kumagai, T.; Krunkosky, T. M.; Krause, D. C.; Radicioni, G.; Alexis, N. E.; Kesimer, M.; Tiemeyer, M.; Boucher, R. C.; Ehre, C.; Hill, D. B. Endotracheal Tube Mucus as a Source of Airway Mucus for Rheological Study. *American Journal of Physiology-Lung Cellular and Molecular Physiology* **2019**, *317* (4), L498–L509. https://doi.org/10.1152/ajplung.00238.2019.

(65) Raveh, B.; London, N.; Schueler-Furman, O. Sub-Angstrom Modeling of Complexes between Flexible Peptides and Globular Proteins. *Proteins: Structure, Function, and Bioinformatics* **2010**, *78* (9), 2029–2040. https://doi.org/10.1002/prot.22716.

(66) Sugio, S.; Kashima, A.; Mochizuki, S.; Noda, M.; Kobayashi, K. Crystal Structure of Human Serum Albumin at 2.5 Å Resolution. *Protein Engineering, Design and Selection* **1999**, *12* (6), 439–446. https://doi.org/10.1093/protein/12.6.439.

(67) Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. PACKMOL: A package for building initial configurations for molecular dynamics simulations. *Journal of Computational Chemistry* **2009**, *30* (13), 2157–2164. https://doi.org/10.1002/jcc.21224.

(68) Hopkins, C. W.; Le Grand, S.; Walker, R. C.; Roitberg, A. E. Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *Journal of Chemical Theory and Computation* **2015**, *11* (4), 1864–1874. https://doi.org/10.1021/ct5010406.

(69) Hu, X.; Lenz-Himmer, M.-O.; Baldauf, C. Better Force Fields Start with Better Data – A Data Set of Cation Dipeptide Interactions. 2021.

(70) Torrillo, P. A.; Bogetti, A. T.; Chong, L. T. A Minimal, Adaptive Binning Scheme for Weighted Ensemble Simulations. *The Journal of Physical Chemistry A* **2021**, *125* (7), 1642–1649. https://doi.org/10.1021/acs.jpca.0c10724.

(71) Copperman, J.; Zuckerman, D. M. Accelerated Estimation of Long-Timescale Kinetics from Weighted Ensemble Simulation via Non-Markovian "Microbin" Analysis. *Journal of Chemical Theory and Computation* **2020**, *16* (11), 6763–6775.

(72) Suarez, E.; Lettieri, S.; Zwier, M. C.; Stringer, C. A.; Subramanian, S. R.; Chong, L. T.; Zuckerman, D. M. Simultaneous Computation of Dynamical and Equilibrium Information Using a Weighted Ensemble of Trajectories. *Journal of chemical theory and computation* **2014**, *10* (7), 2658–2667.

(73) Benton, D. J.; Wrobel, A. G.; Xu, P.; Roustan, C.; Martin, S. R.; Rosenthal, P. B.; Skehel, J. J.; Gamblin, S. J. Receptor Binding and Priming of the Spike Protein of SARS-CoV-2 for Membrane Fusion. *Nature* **2020**, *588* (7837), 327–330. https://doi.org/10.1038/s41586-020-2772-0.

(74) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25. https://doi.org/10.1016/J.SOFTX.2015.06.001.

(75) Lobo, V. R.; Warwicker, J. Predicted PH-Dependent Stability of SARS-CoV-2 Spike Protein Trimer from Interfacial Acidic Groups. *Computational and Structural Biotechnology Journal* **2021**, *19*, 5140–5148. https://doi.org/10.1016/j.csbj.2021.08.049.

(76) Phillips, J.; Zheng, G.; Kumar, S.; Kale, L. NAMD: Biomolecular Simulation on Thousands of Processors. In *Proceedings of the IEEE/ACM SC2002 Conference, Technical Paper 277*; IEEE Press: Baltimore, Maryland, 2002; pp 1–18.

(77) Brace, A.; Salim, M.; Subbiah, V.; Ma, H.; Emani, M.; Trifa, A.; Clyde, A. R.; Adams, C.; Uram, T.; Yoo, H.; Hock, A.; Liu, J.; Vishwanath, V.; Ramanathan, A. Stream-AI-MD: Streaming AI-Driven Adaptive Molecular Simulations for Heterogeneous Computing Platforms. In *Proceedings of the Platform for Advanced Scientific Computing Conference*; Association for Computing Machinery: New York, NY, USA, 2021.

(78) Lee, H.; Turilli, M.; Jha, S.; Bhowmik, D.; Ma, H.; Ramanathan, A. DeepDriveMD: Deep-Learning Driven Adaptive Molecular Simulations for Protein Folding. In *2019 IEEE/ACM Third Workshop on Deep Learning on Supercomputers (DLS)*; 2019; pp 12–19.