Interfacial Phenomena and Specific Ion Effects via the Molecular Dynamics Framework

Jesse Prelesnik

A dissertation submitted in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

University of Washington
2021

Reading Committee:
Jim Pfaendtner, Chair
Chris Mundy
Anne McCoy

Program Authorized to Offer Degree:
Department of Chemistry

University of Washington

Abstract

Interfacial Phenomena via the Molecular Dynamics Framework

Jesse Prelesnik

Chair of the Supervisory Committee:
Jim Pfaendtner
Department of Chemistry

Intermolecular interactions are paramount to our understanding of collective physical properties, but present a scientific challenge in that the characteristic length scales of these interactions are below the diffraction limit for direct visualization, and depend on an ensemble of collective molecular motions that are intractable with contemporary quantum mechanics calculations. In such situations molecular dynamics (MD) is a versatile tool, employing empirical potentials to approximate molecular behavior and sample the underlying statistics rigorously. In recent years MD has experienced a renaissance, with sophisticated accelerated sampling techniques and a wide selection of parameterizations. This has allowed for complex environments such as solid-liquid interfaces to be studied in ways previously inaccessible. Interfacial phenomena are particularly important due to their relation to assembly processes, which are relevant to biological complexes and materials development, among other things. The present work explores aqueous response to heterogeneous interfaces for applications in hydrophobic aggregation, ordered self-assembly, and electronic device performance, in each case revealing distinct behaviors when aqueous ion species are exchanged. These specific ion effects have broad ramifications for the resulting physics, and are a crucial part of the sophisticated theoretical framework needed to predict outcomes *a priori*. Toward this goal, MD provides a route to explore underlying mechanisms, quantifying solution structure and applying macroscoping theory where relevant to connect microscopic phenomena with macroscopic outcomes.

Acknowledgements

I would like to express profound gratitude to Dr. Jim Pfaendtner for his support and guidance over the last two years of my PhD studies. I have learned a lot not only from his research expertise but from his attitudes about science and how to establish strong collaborations. It has been a rewarding experience to develop myself as a scientist under his guidance and I am honored to have worked with so many fantastic researchers after Jim took me under his wing. Among those closest to me is Dr. Chris Mundy, with whom I have spent uncountably many hours working directly and thinking carefully about challenging problems I could not have approached alone.

I would also like to thank my peers in the Pfaendtner group, both past and present, for mentorship and casual conversations that made my experiences seem relatable. Dr. Janani Sampath offered help frequently when I first joined the group, and detailed discussions with Sarah Alamdari, Luke Gibson, and Orion Dollar have been invaluable. There are many others of you who I had formative experiences with, too many to name — I am fortunate to be a part of a group that is so lively and friendly.

I am further grateful to Dr. Lutz Maibaum, who first encouraged my interests in chemical simulation and provided a thorough and robust background on which to develop my skills. His drive for detail and rigor will forever inspire my work. Furthermore, the Maibaum group was always there for me when I still had the most to learn. I would like mention James Ludwig by name, who came with me to the Pfaendtner group and has been a comrade at times I thought I was alone.

It has been a great pleasure to collaborate with Dr.s Rob Alberstein, Shuai Zhang, Harley Pyles, David Baker, Jim De Yoreo, Akif Tezcan, and Rick Remsing (see Chapter 3 of this Dissertation); and with Dr. David Ginger and his students Ramsess Quezada, Dr. Connor Bischak, and Dr. Raj Giridharagopal (see Chapter 4). Interdisciplinary collaborations have been precious and broadened my view of the scientific world in a way I could not have predicted.

This work was facilitated through the use of advanced computational, storage, and networking infrastructure provided by the Hyak supercomputer system at the University of Washington. Material from Chapter 3 is based on work supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, as part of the Energy Frontier Research Center program: CSSAS, the Center for the Science of Synthesis Across Scales under Award Number DE-SC0019288. Material from Chapter 4 is based on work that benefits and leverages DOE Basic Energy Sciences programs in materials science and UW's Washington State-funded Clean Energy Institute, as part of NW-IMPACT, the Northwest Institute for Materials Physics and Chemical Technology.

Declaration

The author declares no conflicts of interest in the work presented herein. Chapters of this thesis have been adapted from scientific manuscripts first-authored by Jesse Prelesnik, and coworkers are acknowledged in footnotes on the first page of each chapter. Permission is granted by the Proceedings of the National Academy of Sciences (PNAS) to include articles published therein by the lead author in their thesis or dissertation.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The last hundred years have seen a renaissance in our understanding of molecular structures, which provided a framework to build macroscopic phenomena from their interacting constituents. The characteristic time scales of collective motions are typically related to the size scales at hand, and are relatively well-understood at the two extremes: Systems of only a few atoms are understood with an exceptional level of detail, and at the other end the resulting macroscopic phenomena are directly measurable. Our understanding begins to break down at the meso-scale, in which the interactions of individual molecules are still paramount to an accurate description of the system but the number of atoms considered exceeds what is tractable with *ab initio* theory. When characteristic length scales do not exceed the wavelength of visible light ($<$ 300 nm), it becomes increasingly challenging to visualize the underlying interactions experimentally [1, 2, 3]. In such situations, it is valuable to characterize molecular motions via empirical models, which scale well enough to incorporate the many thousands of atoms required [4, 5].

1.1 Statistical Ensembles

The utility of molecular simulation is reliant on careful treatments of probability, which in the context of a large assembly of microscopic entities is known as statistical mechanics. In particular, statistical mechanics is built around the notion of *microstates*, which are states characterized by the values of all possible microscopic variables (e.g. positions) [6, 7]. A particular macroscopic observable, say the total energy of the system, corresponds to a large number of microscopic configurations with different internal coordinates but the same resulting energy. The number of microstates that point to this macrostate is known as the *density of states*, and for a system in equilibrium all of the corresponding same-energy microstates are taken to be equally probable [8]. Non-equilibrium statistical mechanics will be largely neglected in this work, although are also amenable to a rigorous statistical mechanical treatment [9, 10].

Even for a system in equilibrium, internal coordinates of a fluid are continually changing and macroscopic observables fluctuate about some average value. The nature of these fluctuations will vary slightly depending upon the statistical ensemble chosen – that is, some thermodynamic variables (such as particle number, N, volume V, or energy E) will be constant while others are permitted to fluctuate [6]. To replicate experimental conditions at room temperature and atmospheric

pressure, typically the canonical ensemble (NVT constant for temperature T) or the isothermal-isobaric ensemble (NPT constant for pressure P) are selected [11]. Regardless, in order to measure these quantities reliably, all of the relevant microstates must be sampled. Molecular dynamics, as a technique that propagates atoms through time, samples correlated microstates over short times but eventually will visit all parts of space that the system exists in – an idea known as *ergodicity* [12] that is valid for disordered systems that exhibit "randomness" (although pathological situations have been known to break the assumption of ergodicity [13, 14], including glassy substances [15]). The space of all internal coordinates that define the physical system, the *phase space*, is sampled with state probabilities corresponding to state energies E, following the Boltzmann distribution [11]:

$$P(E) \propto \exp(-E/k_B T) \tag{1.1}$$

where k_B is the Boltzmann constant and T is the temperature, giving a probability $P(E)$ of observing energy E. The proportionality constant is division by the *partition function* Z, which is the sum of all possible microstates. Doing so normalizes the probability distribution P [16].

$$Z \propto \int \exp(-E(\vec{r}^N, \vec{p}^N)/k_B T)\ \mathrm{d}\vec{r}^N \mathrm{d}\vec{p}^N \tag{1.2}$$

for positions $\vec{r}^N$ and momenta $\vec{p}^N$, written as an integral to illustrate that there are infinitely many microstates that differ infinitesimally (because both position and momentum are continuous quantities). In many situations, the observable of interest is only dependent upon position, in which case the partition function normalization reduces to the *configuration integral*:

$$Z \propto \int \exp(-E(\vec{r}^N)/k_B T)\ \mathrm{d}\vec{r}^N. \tag{1.3}$$

While this integral is not analytically tractable, being able to write down the probability of observing a macrostate allows for the distributions of closely-related systems to be related to one another, which is the basis of *enhanced sampling* techniques as described in Section 1.3.

1.2 Classical Hamiltonians

Classical simulations consider nuclei as point charges and build potential energy functions as a sum of empirically-parameterized terms that approximate quantum mechanical properties like equilibrium bond lengths and three-body angles [17, 18, 19]. While details vary substantively between implementations, all such empirical models suffer from a sort of over-fitting: Molecular parameterizations are based on specific experiments, such as infrared absorption, that are conducted with some set of state variables (temperature, pressure, etc.) and because of this the resulting force field will be most reliable in this vicinity [20]. For example, classical parameterizations of liquid water tend to fail at reproducing properties of ice, because the applied force field terms do not have the necessary temperature dependence [21, 22]. As a result, no model gets every physical feature correct, and the force field used for a given scientific question should be based on experiments under similar conditions.

For the simplest systems, like an ideal gas, the parameters in question are atomic charges and dispersion terms. Point charges are assigned to each nucleus, and the potential energy is often taken as a sum of pairwise terms for computational efficiency [20]. While three-body and higher-order interactions have been shown to improve simulation accuracy for highly-polarizable systems [23, 24], here we limit our scope to a two-body approximation. With this restriction, the Coulombic potential energy U_{Coulomb} of an N-atom system indexed on i and j having charges q_i and q_j may be computed as

$$U_{\text{Coulomb}}(r) \propto \sum_{i \neq j}^{N} \frac{q_i q_j}{r_{ij}^2}, \tag{1.4}$$

for separation r_{ij} between a pair of atoms and with a proportionality constant that depends on the system of units chosen. For a noble gas the atomic charges are identically zero, but in molecular systems these are challenging to define in a satisfactory way. The point charges q are quantum mechanically sensitive to molecular geometry [25], level of electronic structure theory [26], and choice of methodology to assign point charges from spatially heterogeneous orbitals [27].

Early attempts at assigning atomic partial charges, such as the Mulliken method, first compute the molecule's stationary state via a quantum mechanical electronic structure method and these orbitals are then partitioned between atoms [28]. That is, for each 3D grid point of the resulting orbitals, the closest nucleus is determined and a corresponding infinitessimal unit of charge is attributed to that nucleus. Doing so results in some finite fractional charge for the point nucleus, and this has a long history of use in molecular simulation through the 21st century [29]. In practice this method behaves poorly, however, and has been known for common molecules like water to be so incorrect that the sign of atomic partial charges disagrees with well-established chemical theory [27]. A more modern method is that of the Restrained Electrostatic Potential (RESP) [25], which seeks to determine the atomic partial charges that most accurately reproduce the electrostatic potential outside the molecules's own van Der Waals radius. This is done in an iterative least-squares fitting procedure, usually with penalty functions to ensure that no individual partial charge is too large in magnitude [30]. By considering the external field generated by molecular orbitals, the RESP method is more suitable for describing intermolecular interactions as needed for simulation.

Dispersive terms are captured via the Lennard-Jones (LJ) potential U_{LJ}, which is computationally inexpensive and resembles the Morse potential from quantum mechanics [31], which includes two parameters: an effective atomic radius, σ, and the depth of the attraction ϵ.

$$U_{\text{LJ}}(r) = \sum_{i \neq j}^{N} 4\epsilon \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^{6} \right] \tag{1.5}$$

Energetically optimum interactomic distances, set by the length scale σ (minimum energy separation $(2^{1/6})\sigma$), can be obtained via quantum mechanics calculations [32], while the attractive strength tends to be parameterized to emulate bulk properties such as heat of vaporization [33, 34].

Empirical simulations do not permit the formation or breakage of chemical bonds,

and so bonded atoms are never separated over long distances [20]. Therefore the bonded interaction is characterized by approximating the attractive well of the Morse potential via a harmonic restraint [35], under the premise that interatomic distances where the Morse potential and harmonic approximation substantively differ are improbable states. A harmonic potential, as written below, requires an equilibrium bond length r_0 and a characteristic vibrational "spring constant" k_{ij}.

$$U_{\text{bond}}(r) = \sum_{i \neq j}^{N} \frac{1}{2} k_{ij}(r_{ij} - r_0)^2 \tag{1.6}$$

Vibrational frequencies are typically calculated from quantum mechanical calculations [36] or spectroscopically [37], via techniques such as infrared absorption. Equilibrium bond lengths are determined via high-resolution beam scattering or crystallography [38]. The same can be said of angular equilibria and vibrational frequencies, which are also modeled with a harmonic potential. These parameters tend to be rather stiff (having a large k_{ij} compared to other energy magnitudes in the system) and are in general not sensitive to the choices of other parameters [39].

Finally, structural refinement is performed through the addition of torsional terms, which come in two kinds: so-called "proper dihedrals" span 4 atoms bonded in a linear chain, as in the carbon backbone of butane, and measures the angle between atoms 1 & 2 and 3 & 4. Proper dihedrals are commonly represented as Newman projections to illustrate staggered and eclipsed conformations. The other kind are "improper dihedrals" and are used to restrain sp^2-hybridized moieties as planar, preventing inversion about the center [40]. Proper dihedral potentials tend to be shallower than bonds and angles, and so are comparatively sensitive to the choices of other force field parameters. Torsions are typically defined last of all the force field parameters, because charge-charge interactions (among others) already have an influence on molecular structure. In order to correctly reproduce the torsional potential of some molecule, additional dihedral terms must be added to the force field *after* taking into account the potential energy of scanning that angular space due to pre-defined force field parameters [41].

In practice, this dihedral potential is typically obtained by comparing force field energies, with all parameters except torsions in place, against a quantum mechanical reference. This is performed across a scan of torsional angles, in multi-dimensional space (scanning multiple torsions as they may be coupled) and then those same configurations are evaluated according to the classical force field. The dihedral potential to be inserted into the classical force field is then the energetic difference, such that the empirical model matches the QM reference. This is employed as a fitted series of Fourier terms [39]. The mathematical description of the proper dihedral can take on several equivalent forms, and the one preferred here is the Ryckaert-Bellemans function:

$$U_{\text{torsion}}(\phi) = \sum_{n=0}^{5} C_n(\cos(\phi))^n \tag{1.7}$$

Here ϕ is the backbone dihedral angle, and C_n are fitted Fourier coefficients. Improper torsions are described harmonically, as in bonds and angles.

These combined atomic potentials (charges, dispersion, bonds, angles, and torsions, along with atomic mass m), defined for each sequence of "atom types" in

the system, constitute the empirical force field. As nuclei are propagated in time according to Newton's equations of motion, temperature is computed according to the velocity-mass kinetic energy equation and pressure according to the pressure tensor from kinetic energy and the virial equation. The kinetic energy E_{kin} can be determined from velocity vectors $\vec{v}$ as

$$E_{\text{kin}} = \frac{1}{2}\sum_{i}^{N} m_i |\vec{v}_i|^2 = \frac{1}{2} N_{\text{df}} k_{\text{B}} T, \tag{1.8}$$

where N_{df} is the number of degrees of freedom in the system, $3N - 3$ for an unconstrained system of N atoms [11]. While temperature is a scalar, the pressure tensor requires both the kinetic energy tensor $\mathbf{E}_{\text{kin}}$ and the virial tensor $\boldsymbol{\Xi}$ [42]. These are computed as

$$\mathbf{E}_{\text{kin}} = \frac{1}{2}\sum_{i}^{N} m_i \vec{v}_i \otimes \vec{v}_i \tag{1.9}$$

for atom i having velocity vector $\vec{v}_i$ and mass m_i as before, and

$$\boldsymbol{\Xi} = -\frac{1}{2}\sum_{i<j} \vec{r}_{ij} \otimes \vec{F}_{ij} \tag{1.10}$$

for position vector $\vec{r}_{ij}$ and force vector $\vec{F}_{ij}$ between atoms i and j. Together, the pressure $\mathbf{P}$ of a system with volume V is [43]

$$\mathbf{P} = \frac{2}{V}(\mathbf{E}_{\text{kin}} - \boldsymbol{\Xi}). \tag{1.11}$$

These relationships for temperature and pressure permit their instantaneous evaluation, and so can be tracked in time. Even in ensembles where P and T are not constant, for example the microcanonical NVE ensemble, these system observables tend to fluctuate about a well-defined average for a well-equilibrated system. It is also possible to employ a thermostat or barostat, a number of which algorithms have been employed [44, 45, 46], to access other thermodynamic ensembles in molecular dynamics simulations.

Finally, as a practical matter, one must consider the infinite extent of intermolecular interactions and their impacts on a system with a given set of boundaries. Dispersive contributions to the energy, the Lennard-Jones terms that decay as $1/r^6$, vanish in the limit of infinite r when integrated over spherical space, and so can be safely truncated at some large r without adversely impacting systemic measurements [47, 48]. Conversely, the Coulombic interaction that decays according to $1/r^2$ does not converge at any finite cut-off and can lead to severe artifacts if treated improperly [49].

In the condensed phase, as are the systems discussed herein, the infinite dimensions of space are taken into account via periodic boundary conditions (PBCs). The atoms constituting the simulated system reside in a central unit cell, and space beyond the unit cell is conceptually represented by identical images of the original unit cell, tiled in all space. Since each of these tiles is identical (repeated in space), as atoms are forced from the original simulation cell into an adjacent one, the original simulation cell experiences a reciprocal motion and an identical atom enters from

the opposite side. This always preserves the number of atoms in the simulation box while permitting atoms to diffuse freely, and in principle provides coordinates for all atoms in periodic space. Artifacts can arise if the simulation box if insufficiently small to avoid self-interaction with its own periodic image [50]. Treatment of electrostatics in this framework are achieved via the Particle Mesh Ewald (PME) method, that uses Fourier space to converge an infinite sum of electrostatic contributions [51, 52]. Close-range electrostatics are determined in real space, with some cut-off length at which to convert the calculation from real to reciprocal space, and a grid is employed for computational efficiency.

With this conceptual framework, a number of major force field families have emerged (CHARMM, OPLS-AA, GROMOS, etc.), as well as modifications to each to better suit individual applications. Molecular simulations have seen use in applications from theoretical toy problems to millions of atoms in the condensed phase, allowing researchers to computationally investigate microscopic phenomena with excellent resolution.

1.3 Sampling Phase Space with Molecular Dynamics

Molecular dynamics simulations propagate nuclei in time, which is achieved by computing forces and propagating nuclei according to Newton's equations of motion. In practice these propagations occur discretely, stepping forward in time by a finite amount. For infinitessimal time steps, integrating over time can be done with negligible error, but as the time step increases the trajectory through phase space begins to deviate from the ground truth [53, 54]. Having a large time step makes simulating long time scales more accessible because fewer computations must be done per unit time, and so simulation scientists are motivated to identify the maximum permissible step size. On a fundamental level, this time step must be small enough that the most rapid oscillatory behavior in the system can be well-approximated – that is, a sinusoidal motion cannot be captured by discretely sampling less than one point per period. The fastest motion in typical biochemical simulations are oscillations about hydrogen/heavy atom bonds, a frequency that is determined by the two masses and the empirical spring constant describing their bond [55]. Classically, heavier masses and weaker spring constants give lower-frequency motions, while atoms like hydrogen are the most problematic due to their light mass.

Integrating Newton's equations of motion is a task for which multiple algorithms have been proposed [56]. The most commonly relied upon is "velocity Verlet", which is advantageous over Verlet integration (without velocities) due to its time symmetry minimizing local discretization error in position, which scales with time t according to $\mathcal{O}(t^4)$ [57]. With a sufficiently small time step ($\lesssim$ 2 fs for typical biochemical simulations [58]) the discretization error is effectively nonexistent and the simulation can proceed for as many time steps as desired.

The principle of ergodicity, that a time-averaged quantity will approach the appropriate ensemble-average, applies in the limit of infinite time [12]. In practice, a finite-length simulation may not access all of the relevant microstates due to high energetic barriers separating basins, which result in improbable transitions from one collection of states to another (say, folded and unfolded states). In order to

obtain rigorous statistics about improbable events, they must be observed many times, compounding the problem of simulation length. Toward circumventing this practical conundrum, clever applications of statistical equalities permit simulations to be conducted with modified Hamiltonians and their results related back to the original ensemble. These methods are known as enhanced sampling techniques.

1.4 Enhanced Sampling Techniques

The enhanced sampling techniques utilized herein are of two kinds: The first of which are methods that include some additional potential energy term on top of the molecular Hamiltonians, which do not correspond to any underlying physics but rather serve to alter what regions of phase space are thermally accessible on the time scales simulated. As detailed below, it is possible to re-weight measured probabilities from one ensemble to another if the differences between their Hamiltonians are exactly known. The other kind of enhanced sampling method employed here hybridizes molecular dynamics simulations with Monte Carlo (MC) approaches, which accept or reject atom movements probabilistically. MD and MC are well-established as independent routes to obtaining probability distributions from a chemical force field.

1.4.1 Modified Hamiltonians

The most commonly applied enhanced sampling technique applied in MD simulations is to modify the Hamiltonian with a known perturbation, sample coordinates in phase space according to this modified ensemble, and then relate the resulting probability distribution to the un-modified ensemble [59]. The way this can be achieved is easily seen by writing the probability distribution of some variable $\hat{X}$, dependent on the positions of N atoms $\vec{r}^N$ in terms of the configuration integral:

$$P_{eq}(x) = \langle \delta(\hat{X} - x) \rangle_{eq} = \frac{\int \mathrm{d}\vec{r}^N \; \delta(\hat{X} - x) \exp(-\beta U)}{\int \mathrm{d}\vec{r}^N \exp(-\beta U)} \tag{1.12}$$

with x being some specific value of $\hat{X}$, U the total potential energy of the system for a set of coordinates $\vec{r}^N$, and $\beta = 1/k_\mathrm{B}T$. The variable $\hat{X}$ is taken to be dependent only on positions rather than momenta for notational ease, but this strategy can be generalized to other such collective variables that depend only on the state of a single simulation frame. In this notation, the Dirac delta function $\delta(\cdot)$ is a "picking function", zero everywhere except where the argument is zero. If U is an analytical function of $\hat{X}$ then the numerator reduces to

$$P_{eq}(x) = \frac{\exp(-\beta U(x))}{\int \mathrm{d}\vec{r}^N \exp(-\beta U(\hat{X}))}. \tag{1.13}$$

When some additional potential energy function $V(x)$, also explicitly a function of the observable $\hat{X}$, is added to the Hamiltonian, the resulting probability density $P_V(x)$ is clearly proportional to $\exp(-\beta U(x) - \beta V(x))$. If the configuration integral Z (or Z_b for the biased system) is taken to be a normalization constant for the resulting probability distribution, then a relationship between the two probabilities can be established:

$$P_{eq}(x) \propto P_V(x) \exp(\beta V(x)). \tag{1.14}$$

This simple relation, originally formulated by Torrie and Valleau in the '70s, is the basis for many enhanced sampling techniques [59]. A detailed formalism, in more general terms, is provided in Chapter 2 of this dissertation, applied in the context of umbrella sampling (with $V(x)$ a parabola). The important take-away from this relation is that, when a static bias is employed on a chemical simulation, it is possible in principle to deduce the probability distribution of a simulation that does not have the bias in place.

The functional form of $V(x)$ can be quite general, and one could argue that the "best" function $V(x)$, in terms of reducing the required computational runtime, is one that is exactly the inverse of the underlying free energy curve along x, because the end result is a free energy landscape that is perfectly flat and therefore sampling of the coordinate x occurs diffusively [60]. This is the objective of metadynamics, which spends some initial simulation period building the function $V(x)$ from a series of deposited Gaussian "hills". Initially, some Gaussian width σ^2 and height ω are set, and Gaussians are deposited periodically with characteristic time τ, centered at the present value of the collective variable x. As the time-dependent bias $V(x,t)$ accumulates at x, the hill height is typically tamped down exponentially according to some virtual temperature difference ΔT, a parameter, which allows the bias $V(x)$ to converge [61]. As an equation, one may write,

$$V(x,t) = \sum_{\tau=1}^{\tau_f} \omega\tau \exp\Big(-\frac{V(x,t)}{k_{\mathrm{B}}\Delta T}\Big) \exp\Big(-\sum_i \frac{(x_i - x_i(t))^2}{2\sigma_i^2}\Big). \tag{1.15}$$

This formulation allows for multiple coordinate-dependent variables x_i, and the functional dependence of $V(x_i,t)$ is on the full set of $\{x_i\}$. The metadynamics formulation is particularly valuable because the user needs only to specify a few characteristic input parameters, σ, τ, and ΔT, in order to achieve a well-converged free energy landscape that may encapsulate numerous energetic basins (assuming the variables x_i describe the transitions well). This is in contrast to manually-established parabolic energy biases that must be distributed along the coordinate x in a way that is dependent on the underlying free energy landscape in a way that is not known *a priori*. That is, if the underlying free energy landscape is very steep at some value of x, then harmonic energy restraints must be very stiff and closely-spaced to sample it thoroughly, and if parabolic minima are not well-placed then gaps are likely to appear in the resulting probability distribution. Metadynamics solves this problem by building the energy bias adaptively.

Since the well-tempered metadynamics bias $V(x)$ converges after some time, the bias potential may then be treated as constant and standard reweighting relations applied to obtain unbiased probability distributions along the chosen variables x [62]. Further modifications have been proposed to the well-tempered metadynamics formalism to extend its applicability to systems of many redundant collective variables, all of which contributing hills to the same underlying bias to expedite convergence [63]. These techniques have seen utility in sampling rare events such as peptide conformational changes.

Any technique that applies an energetic bias V that is a function of some collective variable x is dependent on the ability of the chosen collective variable to

accurately describe the process at hand. For example, if the rare event under study is a conformational change that is intimately connected with some torsional rotation, but the variable chosen to bias is some interatomic distance that describes the motion poorly, then the utility of applying a bias is diminished. The intent of applying such an energetic bias is expediency, to require fewer total simulation frames (and thus less computational expenditure) than would otherwise be required.

1.4.2 Monte Carlo Exchange

The convergence of simulated observables can also be expedited by running several simulation replicas in parallel and periodically exchanging their coordinates according to a Monte Carlo scheme that preserves microscopic reversibility. The most common version of Monte Carlo exchange in MD simulations is to set each simulation replica to a different temperature, increasing along an exponentially distributed ladder [64]. Those replicas that sample at high T have additional thermal energy to traverse energy barriers in configurational space, and can explore phase space rapidly compared to at low T. Once new regions of phase space have been accessed at high T, their coordinates may be swapped with other replicas at lower temperatures, allowing many energetic basins to be sampled at the temperature of interest (which is typically the lowest temperature setting of all the replicas).

Exchanging coordinates as the simulations propagate in time results in a trajectory that is discontinuous. That is, molecules do not proceed smoothly through space for a given replica but rather make large jumps at periodic intervals as configurations exchange and the replica takes on new coordinates. Therefore, Monte Carlo exchange techniques are valuable for determining thermodynamic properties but caution must be exercised when studying kinetic properties [65]. Here, an emphasis will be placed on extracting thermodynamic quantities.

A key feature of Monte Carlo algorithms in this context is the preservation of detailed balance, which is founded on principles of microscopic reversibility and states that at equilibrium the likelihood of some elementary process is balanced exactly by the reverse of such a process [66]. That is,

$$P(x'|x)P(x) = P(x|x')P(x') \tag{1.16}$$

for conditional probability $P(x|x')$ describing the likelihood of transitioning from any given state x' to another state x and the desired stationary distribution $P(x)$. Having established previously that $P_{eq}(x,\beta) = \exp(-\beta U(x))/Z(\beta)$, the probability of observing some set $\{x\} = \{x_1, x_2, ..., x_M\}$ for M replicas each with inverse temperature $\{\beta_m\}$ may be written [12],

$$P(\{x,\beta\}) = \prod_m^M P_{eq}(x_m,\beta_m). \tag{1.17}$$

This extended system comprised of independent simulation replicas also has an associated configuration integral,

$$\mathcal{Z} = \prod_m^M Z(\beta_m). \tag{1.18}$$

With this, the probabilities of exchanging two specific replicas a and b in the forward and reverse directions give rise to transition probability matrices that have a ratio equal to

$$\frac{P(x, \beta_b | x', \beta_a)}{P(x', \beta_b | x, \beta_a)} = \exp(-\Delta) \tag{1.19}$$

for Δ obtained from ratios of $P_{eq}(x, \beta)$ as

$$\Delta = (\beta_a - \beta_b)(U(x) - U(x')). \tag{1.20}$$

The relation given describing Δ pertains to the NVT ensemble, but has been generalized for the NPT ensemble as well [67]. The criterion for accepting the replica exchange between two systems can then be expressed,

$$P(x, \beta_b | x', \beta_a) = \begin{cases} 1 & \text{for } \Delta < 0 \\ \exp(-\Delta) & \text{for } \Delta > 0. \end{cases} \tag{1.21}$$

With a replica exchange acceptance criterion chosen, the replica exchange molecular dynamics (REMD) simulations may proceed with the guarantee that detailed balance is being obeyed, as it was included by construction [64]. REMD is applied in this work to determine the molar volume of a glassy material that at low temperatures does not sample phase space efficiently (see Chapter 4). By exchanging with replicas sampled at higher temperatures, quality statistics can be obtained at low temperatures by leveraging the ability of the other replicas to visit new states.

These tools are a common framework to many applications of molecular modeling in modern science. While specific theoretical applications vary from system to system (particularly when soft matter is being investigated and entropy is important), the statistical physics presented here are broadly applicable and have seen much utility in the study of interfacial phenomena, which are notoriously challenging to capture from a purely theoretical standpoint. These problems are exacerbated by the emergence of specific ion effects at interfaces, and thus MD simulations have become a staple in explaining experimental observations and predicting behaviors.

Chapter 2

Decoupling of the Water-Ion Density Fields Adjacent to Molecular-Sized Cavities

2.1 Abstract

[1]Bulk solution hydrophobic properties are well-studied between two separated phases. Interfacial effects, which dominate on the bulk length scale, are insufficient to describe the solvation of individual hydrophobic units too small to interrupt the solvent-solvent hydrogen bond network. Here we investigate the free energy dependence of evacuating a spherical cavity of varying size in aqueous solution (NaCl) on the cavity radius. Previous work in homogeneous medium indicates a transition from cubic to quadratic dependence on radius. We find that varying ion concentration does not shift this transition point substantially. By applying the 2D Weighted Histogram Analysis Method along one reaction coordinate and evaluating the free energy surface of another in classical molecular simulation, we acquired equilibrium statistics about cavities of intermediate sizes throughout this transition region and observe emerging order in water molecule orientations adjacent to the cavity. Notably, the density fields of water molecules and ions at the interface appear to be decoupled, and undergo depletion at different rates as the cavity size increases. These findings provide additional mechanistic insight to the liquid-vapor interface ion adsportion phenomenon as we observe the effect emerging prior to the formation of a water vapor layer.

2.2 Introduction

Understanding the thermodynamics of solvation and the solvent structure in the vicinity of a solute are fundamental goals of liquid state physical chemistry [68, 69, 70]. In this work, we study the solvation of purely hydrophobic solutes in aqueous electrolyte solutions. Using Molecular Dynamics (MD) computer simulations and free energy calculations, we quantify to what extent the presence of ions alters the well understood mechanism of solvation in neat water. Our work is motivated by

[1]Reproduced with permission from Jesse Prelesnik and Lutz Maibaum. Decoupling of the Water-Ion Density Fields Adjacent to Molecular-Sized Cavities. Pending submission.

the current understanding of two distinct phenomena that have been extensively studied over the past two decades. The first is the realization that when a large ($\gtrsim$ 1 nm) hydrophobic object is immersed in water, a solvent-depleted layer forms in its surroundings, which is separated from the bulk solution by an interface that resembles that between a liquid and its vapor in thermodynamic coexistence [71]. The propensity of water to form a vapor layer near such solutes explains their mutual effective attraction, the well-known hydrophobic effect [72]. The second phenomenon is the unexpected tendency of some ions to adsorb to water-air interfaces [73, 74, 75], in apparent contradiction to dielectric continuum theory [76, 77]. When combined, these two observations lead to the questions whether ions exhibit similar behavior near the liquid-vapor interface surrounding hydrophobic solutes, and to what extent this might change the thermodynamics of solvation.

Like many before us [78, 71, 79], we focus on purely hydrophobic solutes that do not have any attractive interactions with the surrounding water molecules. Inserting such a solute into a liquid is akin to forming a cavity, a region in space that is not accessible to solvent molecules. The free energy of inserting such a solute into the liquid is the same as that of growing a cavity of the same size [80]. While real solutes will generally exhibit some amount of attractive interactions with the solvent, those can often be treated perturbatively [81, 82, 83, 84]. The thermodynamic cost of inserting such a solute into liquid water depends on its size. If the solute is small, the water molecules in its vicinity can maintain much of their hydrogen bond network (Figure 2.1a), and the free energy scales with the volume of the solute [80]. If the solute is large, the solvent responds to its presence by the aforementioned creation of a vapor-liquid interface, which results in a free energy that is proportional to the surface area of the solute (Figure 2.1b). The transition between these two regimes occurs at solute sizes of around 1 nanometer for water at room temperature and pressure, and changes slightly depending on the thermodynamic conditions [71, 79].

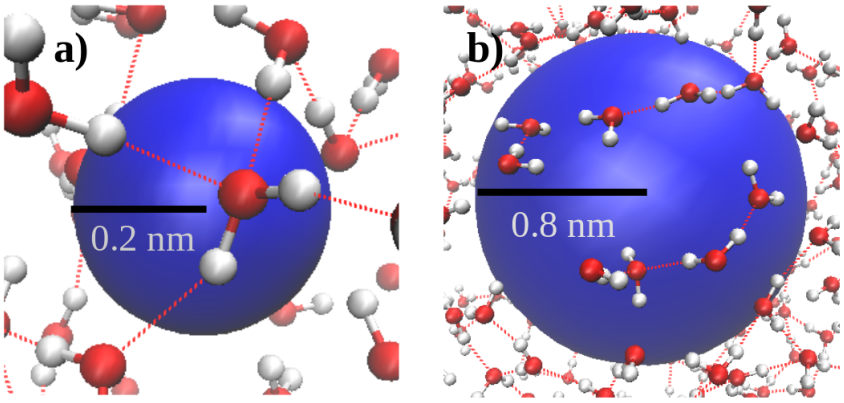

Figure 2.1: Representative simulation frames illustrating hydrogen bond networks between water molecules within 0.2 nm of the edge of the spherical cavity (blue) of radius (a) 0.2 nm and (b) 0.8 nm. Notice that in (a), the hydrogen bond network remains intact and spans the solute, while in (b) the cavity is geometrically incompatible with the network and water molecules close to the cavity are relatively sparse. Figure inspired by [71]

Our goal is to elucidate the effect of sodium and chloride ions on the thermodynamic cost of creating a cavity in water and on the molecular structure of the solvent in the vicinity of the cavity. As outlined in the following Methods section, we use

MD simulations in conjunction with Umbrella Sampling in our study. Because these simulations require the evaluation of continuous forces, we do not directly employ the hard sphere model that is typically used to study cavity formation in Monte Carlo approaches [80, 78, 85]. Instead we resort to a continuous analog of the cavity size, which is sufficiently similar to the actual cavity size that a probability distribution of the latter can be extracted from simulations that bias the former [86]. We can then structurally and thermodynamically characterize water molecules and ions near the cavity interface to identify couplings between species as the cavity grows.

2.3 Methods

All simulations were performed in GROMACS 5.1.4 [87] with the PLUMED 2.1 [88] suite. Each neutral system was composed of 2008 SPC/E water [89] molecules. Ionic systems included 500 mM and 2 M NaCl using OPLS-AA parameters, with each ion replacing a water molecule. We use periodic boundary conditions, the Berendsen barostat (1 bar), Nose-Hoover thermostat (298 K), fourth-order PME long-range interactions with Fourier spacing 0.12 nm, and real-space interaction cutoffs at 0.9 nm.

The simulated system is a homogeneous isothermal-isobaric fluid. Fluctuations cause local density to vary probabilistically, and small regions devoid of solvent transiently appear. Cavities of much greater radii than the Van der Waals radius of the solvent particles are exceedingly improbable, and may only be observed in a typical MD simulation under prohibitive simulation lengths [71, 90]. To characterize their statistics efficiently, specialized sampling algorithms are utilized. While enhanced sampling methods constitute a broad class, a common feature is the need to identify a generalized coordinate or "collective variable" (CV) that describes the rare event to be observed [91].

In the study of cavity formation, the relevant CV is one that measures the current size of the cavity, and is described as the magnitude of the vector connecting an arbitrary stationary origin to the nearest water molecule oxygen atom center, $\min(|\vec{r}_j|) \equiv R$. We calculate $P_R(r)$, the probability distribution of this CV. From this definition, the free energy of cavity formation may be expressed as [80]

$$\Delta G(R) = -k_{\mathrm{B}}T \ln\left\{\int_R^\infty \mathrm{d}r P_R(r)\right\}. \tag{2.1}$$

This integral is computed numerically, yielding the Gibbs free energy associated with cavity formation based upon the probability distribution obtained from simulation. The properties of $P_R(r)$ have been studied with prior Monte Carlo simulations in neat water across the full domain of cavity sizes [80, 71], while the behavior of $\Delta G(R)$ in the presence of salt has seen modest attention [79]. The presence of ions is thought to decrease the probability of a large cavity forming considering the increase in surface tension observed at the planar water-vapor interface [74]. Our interest lies in the structural characteristics of water and ions as this interface grows.

The PLUMED plugin was used to measure distances and apply bias energy to drive phase space exploration as detailed in the following section. While we are interested in the probability distribution of the cavity size R, this CV cannot be

biased in a MD simulation that requires smoothly varying forces. Instead, we define an analogous CV R^p,

$$R^p = \frac{\alpha}{\log \sum_j \exp(\frac{\alpha}{|\vec{r}_j|})}. \tag{2.2}$$

This form is implemented in the PLUMED plugin as a smoothly differentiable minimum distance convention [88]. The parameter α was selected to be 2.5 nm. This function well-approximates R, and R^p may then be biased to sample configurations with large cavities in MD simulations via Umbrella Sampling. We then reconstruct $P_R(r)$ using two-dimensional WHAM, a method detailed for our application below.

For analysis, hydrogen bonds were computed via a geometric definition which considers coordinate distances and angles, criteria which have been argued to provide similar results to experiment [92]. This was implemented by the GROMACS hbond utility with an angle parameter (Hydrogen-Donor-Acceptor) 30° and distance parameter (Donor-Acceptor) 0.36 nm.

2.4 Background: Two-Dimensional WHAM

The accelerated sampling technique employed here is the multivariate Weighted Histogram Analysis Method (WHAM), a Torrie-Valleau reweighting scheme applied to several simulations each on a restricted domain of observations, which are joined to one another in a statistically optimum fashion [59, 93]. This technique adds a (static) energetic bias to the system Hamiltonian. Sampling in the modified ensemble is proportional to the inverse of the Boltzmann factor of the energetic bias, and so un-biased information may be obtained from a biased simulation (see below). The benefit of introducing such a bias is to alter the region of phase space that the simulation visits, as measured by some CV. Since the underlying free energy landscape is not known *a priori*, in practice the energetic bias is often some quadratic function of the CV to ensure that a specific region of the reaction coordinate is sufficiently sampled [94]. This restricts the domain of observations, and so simulation "windows" are conducted in parallel with different energetic biases, on domains that have some overlap. Since each window may be unbiased into the same ensemble, probability distributions obtained from each window are combined into a single distribution spanning the entire domain.

The WHAM equations are presented here in their two-dimensional form to illustrate how a one-dimensional bias on R^p can be utilized to obtain probability distributions along R. This strategy is generally applicable when placing a bias on one CV expedites the phase space exploration of the other CV. For clarity of presentation and generality of application, the generic variables X and Y are used in place of R and R^p.

Consider the joint probability that a CV $\hat{X}(\vec{r}^N)$ takes on a specified value, x, while simultaneously some other CV $\hat{Y}(\vec{r}^N)$ takes on value y, which is written $P_{eq}(x, y)$. It may be expressed as an expectation value $\langle \delta(\hat{X}(\vec{r}^N) - x)\delta(\hat{Y}(\vec{r}^N) - y)\rangle$. In the language of statistical mechanics, this expected value may be computed from configuration integrals, assuming for simplicity that these variables are functions of system coordinates but not momenta. For notational brevity, these CVs will be

written $\hat{X}$ and $\hat{Y}$; while the function $U(\vec{r}^N)$, representing the potential energy as determined by the force field, may be abbreviated to U; and the external field applied by the user, indexed by simulation window on i and written $V_i(x, y)$, will be abbreviated as V_i.

$$\begin{aligned} P_{eq}(x,y) &= \langle \delta(\hat{X}-x)\delta(\hat{Y}-y)\rangle_{eq} \\ &= \frac{\int \mathrm{d}\vec{r}^N \ \delta(\hat{X}-x)\delta(\hat{Y}-y)\exp(-\beta U)}{\int \mathrm{d}\vec{r}^N \exp(-\beta U)}. \end{aligned} \tag{2.3}$$

In the context where the aforementioned bias energy is present, this expected value takes on a different form, denoted by the superscript b for the biased ensemble indexed on i.

$$\begin{aligned} P_i^b(x,y) &= \langle \delta(\hat{X}-x)\delta(\hat{Y}-y)\rangle_i \\ &= \frac{\int \mathrm{d}\vec{r}^N \ \delta(\hat{X}-x)\delta(\hat{Y}-y)\exp(-\beta U-\beta V_i)}{\int \mathrm{d}\vec{r}^N \exp(-\beta U-\beta V_i)}. \end{aligned} \tag{2.4}$$

The probability distribution obtained within the biased ensemble may be used to obtain the unbiased probability distribution via reweighting the observation histogram according to the bias factor $\exp(\beta V_i)$. A proportionality factor, not known *a priori*, accounts for the fact that the reweighted probability distribution $P_i^u(x, y)$ is not normalized.

$$\begin{aligned} P_i^u(x,y) &= \langle \delta(\hat{X}-x)\delta(\hat{Y}-y)\exp(\beta V_i)\rangle_i \\ &= \frac{\int \mathrm{d}\vec{r}^N \ \delta(\hat{X}-x)\delta(\hat{Y}-y)\exp(-\beta U)}{\int \mathrm{d}\vec{r}^N \exp(-\beta U-\beta V_i)} \\ &= \left[\frac{\int \mathrm{d}\vec{r}^N \ \exp(-\beta U)}{\int \mathrm{d}\vec{r}^N \exp(-\beta U-\beta V_i)}\right] P_{eq}(x,y). \end{aligned} \tag{2.5}$$

This proportionality constant may be interpreted in a number of ways. While it may be written as a ratio of configuration integrals as in equation (2.5), in umbrella sampling literature this value is concisely expressed in terms of the change in free energy F_i that results from imposing the external field V_i

$$\begin{aligned} \exp(-\beta F_i) &= \frac{\int \mathrm{d}\vec{r}^N \ \exp(-\beta U-\beta V_i)}{\int \mathrm{d}\vec{r}^N \exp(-\beta U)} \\ &= \langle \exp(-\beta V_i)\rangle_{eq}. \end{aligned} \tag{2.6}$$

With this definition, we may rewrite 2.5 as

$$P_{eq}(x,y) = \exp(-\beta F_i) P_i^u(x,y). \tag{2.7}$$

In practice, many independent simulation windows are run, and the probability

distribution $P_{eq}(x, y)$ is determined from a weighted sum of their results:

$$P_{eq}(x, y) = \sum_i w_i(x, y) \exp(-\beta F_i) P_i^u(x, y). \tag{2.8}$$

$$\sum_i w_i(x, y) = 1. \tag{2.9}$$

In WHAM, the weights are those that minimize the statistical error of $P_{eq}(x, y)$ subject to the above normalization constraint. Performing this optimization via the method of Lagrange multipliers and estimating the statistical error from within a simulation window from the properties of a Poisson distribution (as in [93]), these weights may be written as

$$w_i(x, y) = \frac{N_i \exp\left(-V_i(x, y) + F_i\right)}{\sum_j N_j \exp\left(-V_j(x, y) + F_j\right)}. \tag{2.10}$$

Here N_i is the "length" of simulation i, which is the number of coordinate sets used to compute $P_i^b(x, y)$. These weights may be evaluated for a given set of $\{F_i\}$ and an estimate of the unbiased probability distribution $P_{eq}(x, y)$ is obtained. In general the correct set of $\{F_i\}$ are not known *a priori*, and so an initial guess must be iteratively improved by self-consistently reevaluating $\{F_i\}$ from its definition, expressed in a way that is explicitly dependent on the distribution $P_{eq}(x, y)$:

$$\exp(-\beta F_i) = \int \mathrm{d}y \int \mathrm{d}x \exp(-\beta V_i(x, y)) P_{eq}(x, y). \tag{2.11}$$

Once $\{F_i\}$ has converged, the appropriate free energy constants are known to combine observations from numerous independent simulation trajectories with respect to the unbiased system, giving the appropriately-normalized $P_{eq}(x, y)$ [95]. To obtain one-dimensional landscapes along a single CV, the other degrees of freedom may be simply integrated out. This method may be trivially extended to arbitrarily many CVs, although with increasing computational expense [94], or used in conjunction with other biasing methods [96].

We applied a harmonic potential of the form $V(R, R^p) = \frac{k}{2}(R^p - R_0^p)^2$ in this work. Notice that V does not depend on R, which circumvents the problem of discontinuous forces. Writing the bias potential as $V(R, R^p)$ is useful in extracting thermodynamic quantities about R. A campaign of umbrella sampling simulations was launched with different biased values of R_0^p. Here, k descended from 1000 kJ/mol to 500 kJ/mol as R_0^p increased from 0.2 to 1.0 nm, with each simulation window having one k and one R_0^p parameter. Following the 2D WHAM procedure, we obtain $P_{eq}(R, R^p)$. We then integrate over R^p to obtain $P_{eq}(R) = P_R(r)$, and then integrate $P_R(r)$ according to Eq. 2.1 to obtain $\Delta G(R)$ as desired.

For properties such as the local density function relative to the cavity center for fixed cavity size, it is sufficient to gather statistics from a single simulation window that captures the relevant configurations and re-weight observations with the typical

Torrie-Valleau weights $\exp(\beta V_i)$ as in (2.7).

2.5 Results

We acquired equilibrium statistics about molecular sized spherical cavities using biased MD simulations to better understand the interplay between the tendency of ions to adsorb to liquid-vapor interfaces and the gradual formation of a liquid-vapor interface in water adjacent to a growing spherical cavity. The 2D WHAM algorithm was employed to calculate the free energy of cavity formation as a function of spherical cavity radius R. This computation uses Eq. (2.1) to manipulate the probability distribution acquired from simulation to obtain the free energy of generating a cavity of given size. For neat SPC/E water this result is known, and the computation here is in agreement with literature [80, 71].

We have further computed this quantity in 500 mM and 2 M aqueous NaCl solutions, in which the free energy increases more quickly with increasing salt concentration. In Figure 2.2 we see that the free energy scales according to R^3 for small cavities and as $\sim R^2$ for larger cavities, as expected [71]. The horizontal asymptotes represent the surface tension of a planar interface [97]. Computations of surface tension from simulations of the planar interface, using the mechanical definition of pressure and the pressure tensor [98], did not converge quickly enough to differentiate the surface tension of SPC/E water with and without NaCl to within statistical error. Instead, these amyptotes are taken from the well-known air-water interface surface tension of the SPC/E water model of 71 mJ/m^2 [99]. The surface tension of SPC/E water in the presence of ions is unknown, although it may be estimated based upon the behavior of real water. Empirical surface tension measurements with and without NaCl result in a ratio of surface tensions (with NaCl : without NaCl) slightly greater than unity. Here we assume that the surface tension of SPC/E water will increase at the same proportional rate with the addition of ions, and the well-known SPC/E surface tension is scaled up by this ratio to approximate its surface tension in the presence of NaCl. In the 500 mM and 2 M NaCl cases, these scaling factors are 1.005 and 1.012, respectively [100]. Note that the surface tension relevant to the study of cavities is slightly higher than that of the air-water interface because fluctuations present at the free interface are quenched by the presence of a wall [78].

The continuum theory of surface tension in Debye-Hückel electrolytes predicts, for low electrolyte concentrations, that surface tension should increase as a result of ion repulsion near the interface from their electrostatic image force [73]. Considering the already high surface tension of water, the addition of NaCl is expected to increase surface tension at most by ~ 2 %. This small deviation is observed in Figure 2.2, as evidenced by the free energy curves lying closely on top of one another.

These data suggest a weak trend emerges. A linear fit through the small size regime (considering radii 0-1.5 σ, for SPC/E Lennard-Jones parameter $\sigma = 0.3166$ nm) may be drawn where the free energy, normalized by R^2, scales linearly with R. It intersects the horizontal asymptote at somewhat smaller R values when NaCl is present, in a manner akin to increasing the temperature [71]. This is indicative that the free energy of cavity formation increases more quickly in the small size regime than the increase in planar interface surface tension would suggest. However, given that the horizontal asymptotes are estimates, and the fractional change in intersection R value is small, this apparent trend may be insignificant.

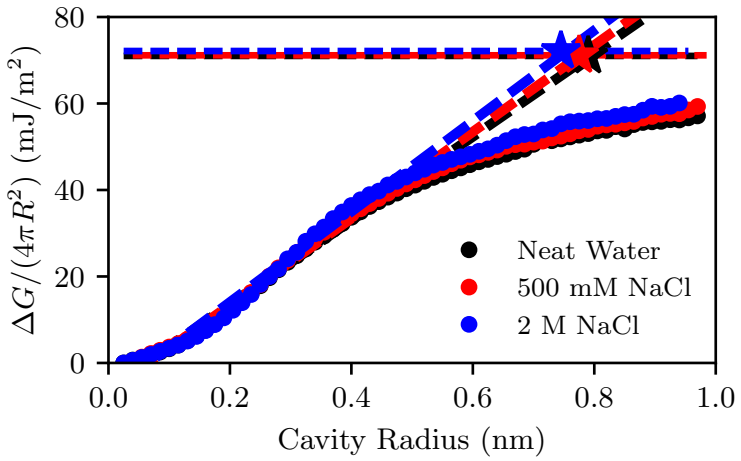

Figure 2.2: Gibbs free energy per unit surface area of spontaneously generating a spherical cavity of given radius, as measured by the distance from the nearest water oxygen atom to the center of the cavity (origin). The rate of free energy increase for small cavities (cavities smaller than ∼1.5 water solvation shells in radius, linear trend lines dashed) is greater in the presence of sodium chloride than would be anticipated from the increase in surface tension at the planar limit (horizontal dashed lines). This is indicated by the intersections of these lines, marked with stars, which tend toward smaller cavities with increasing NaCl concentration. Beyond ∼0.6 nm cavities, slow convergence produces less reliable results (see text). We estimate the statistical uncertainties using Monte Carlo bootstrapping, and we find the standard deviation to be less than the size of the points used to represent the data.

To understand the behavior of the system in the small cavity regime, the local density was computed as a function of distance from the origin [101]. Within the size regime of interest, approximately 0.15 nm to 0.45 nm, it is known that the solvent has not yet formed a vapor-water interface, and is still responding elastically to the cavity presence as indicated by an markedly greater density of water oxygen atoms immediately adjacent to the cavity (Figure 2.3a, H_2O) [80].

Despite the water density displaying little qualitative character change with increasing cavity size, the sodium and chloride ions displayed dramatically different distributions as the cavity size increases (Figure 2.3b Cl^-, 2.3c Na^+). By 0.45 nm these distributions qualitatively resemble the well-known result from NaCl at a planar interface, which exhibits both a depletion of ions near the boundary and a small peak in density [75]. We find that the ions undergo a transition in their local density profiles at much smaller cavity radii than the water molecules do.

On overlay of the local density functions for neat water, 500 mM NaCl, and 2 M NaCl further illustrate that the density of water molecules adjacent to the cavity is largely unimpacted by the depletion of ions from the same region (Figure 2.4).

The unchanging character of the water molecules in the presence of NaCl raises the mechanistic question of why the ions undergo density depletion. Toward this end, in Figure 2.5 we compute the orientations of water molecules adjacent to the cavity. Here the orientation of a water molecule relative to the cavity is defined as the cosine of the angle between two vectors: (i) The vector from the cavity center to the water oxygen, and (ii) the vector from the water oxygen to the average position of its two hydrogen atoms, the "dipole vector". The special case $\cos(\theta) = 1$ corresponds

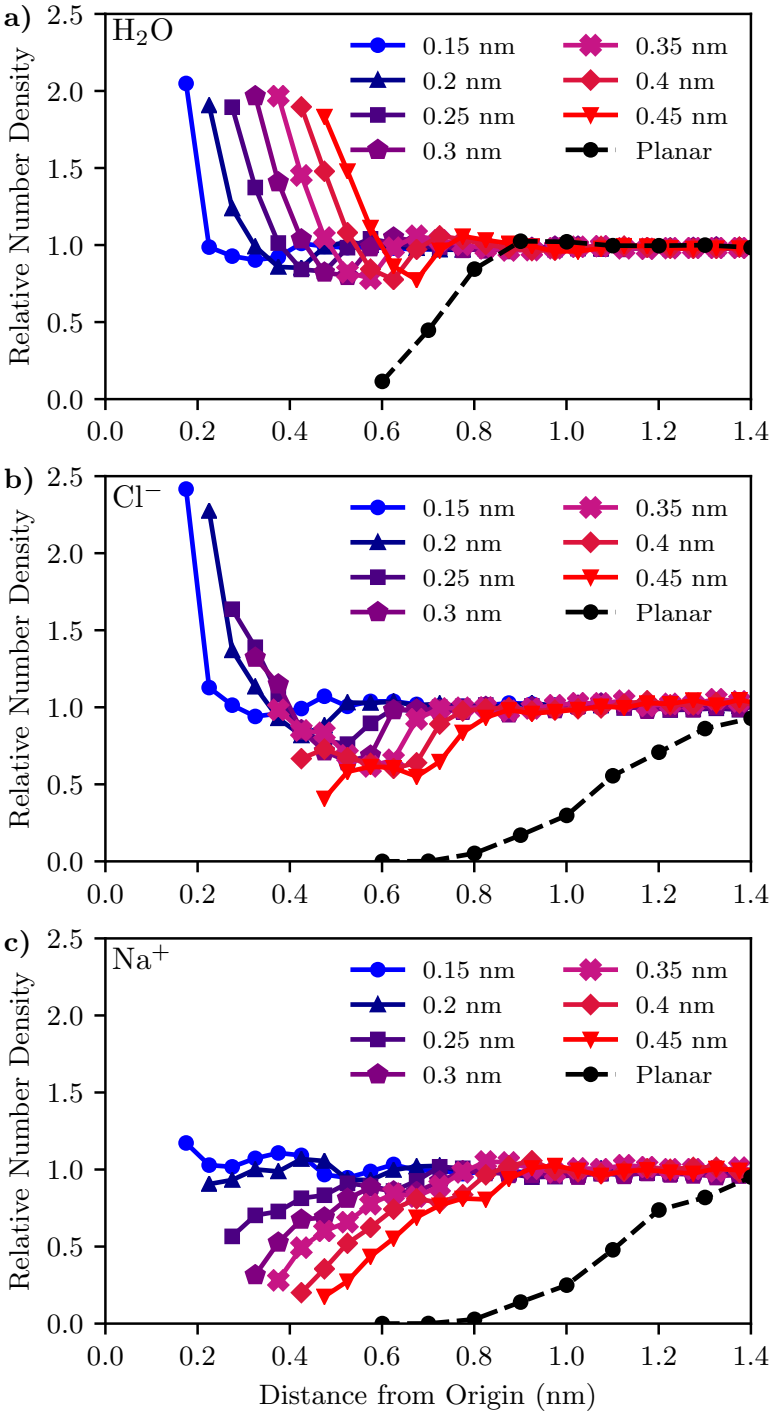

Figure 2.3: Ensemble average local number density of 2 M NaCl in water as a function of radial distance from the cavity center given that a cavity of specified radius is present, ± 0.025 nm. For Na^+ and Cl^-, density adjacent to the cavity is gradually depleted and ion adsoprtion peaks begin to occur over the cavity range indicated. Conversely, the density function of H_2O (measured using the oxygen atom) does not yet reflect the formation of a liquid-vapor interface as is known to occur in the large cavity limit. The dashed black lines represent the infinite cavity limit, a planar wall-water interface, shifted to superimpose on these axes.

to the water dipole vector pointing directly outward from the cavity center. The probability distribution of $\cos(\theta)$ is uniform for random orientation angle θ, which yields $\langle\cos(\theta)\rangle = 0.5$.

We observe a deviation from random orientation when a water molecule is adjacent to the cavity interface, a propensity that is magnified but qualitatively unchanged adjacent to larger cavities. This is indicated by probability distributions that deviate strongly from uniformity with greater contrast in probability density adjacent to the cavity. The presence of NaCl does not appear to alter these probability distributions significantly (Figure 2.5).

While the definition of the water molecule dipole vector averages over the precise positions of hydrogen atoms, complicating the interpretation of these orientation data, the region of high probability between $\cos(\theta) = 0.0$ and $\cos(\theta) = 0.5$ suggests that the dipole vector points nearly tangential to the cavity wall most frequently, with a tendency to orient with the positive end away from the cavity. This finding is consistent with the hydrogen bond network remaining intact in the presence of "small" solutes. Another probable region is found at almost exactly $\cos(\theta) = -1.0$, corresponding to the hydrogen atoms straddling the cavity. These configurations are consistent with previously determined distributions about small cavities in neat water [102].

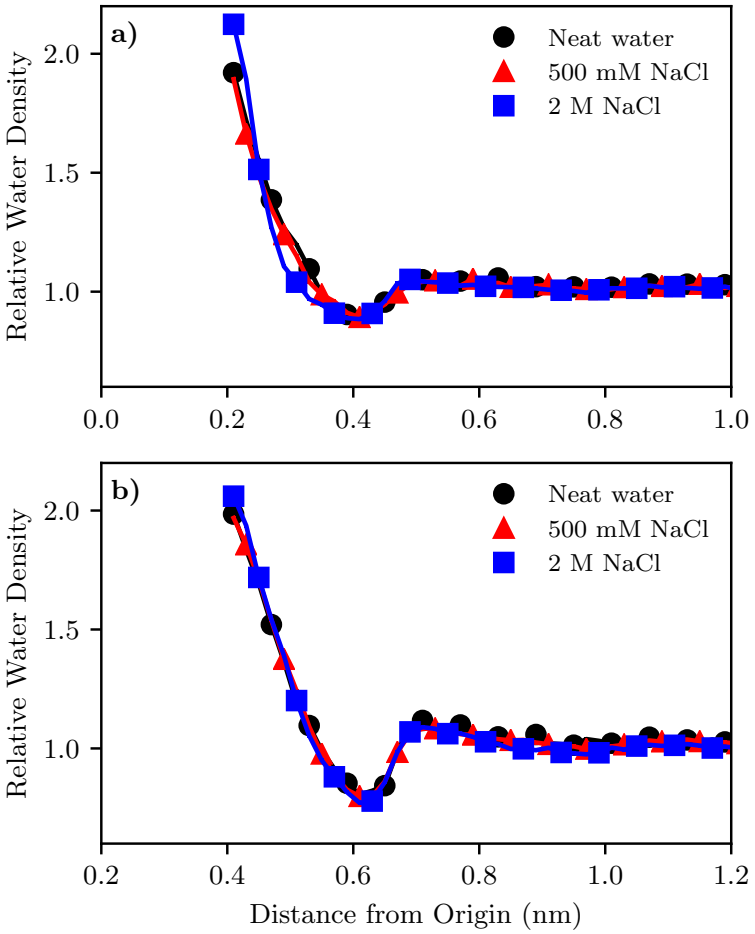

Figure 2.4: Spatial density function of water molecule oxygen atoms relative to the center of an (a) 0.2 nm, or (b) 0.4 nm cavity. The relative density of water molecules adjacent to the cavity is largely unaffected by the addition of NaCl.

To interpret this data in the context of the ion density field decoupling observed in Figure 2.3, we computed the average number of hydrogen bonds per water molecule as a function of distance from the cavity edge. We used a geometric criteria to determine hydrogen bonding (see Methods). For cavities between 0.15 nm and 0.45 nm, regardless of NaCl concentration, the average number of hydrogen bonds per water molecule was found to be within statistical error of the number associated with a water molecule in bulk solution, even if that water molecule was near the cavity interface. Minute changes in hydrogen bond patterns should be expected depending upon the water molecule's proximity to the cavity, considering the unique environment of the interface region compared to bulk. However, these differences could not be resolved with statistical significance here due to the large variance observed compared to the small magnitude of these hydrogen bonding differences. This speaks to the small cavity regime being consistent with an unbroken hydrogen bond network among water molecules, and that this network is not significantly disturbed by ions at the cavity interface.

2.6 Discussion

Adapted sampling methods of molecular simulation have been indispensible tools for mechanistically probing events beyond the resolution of experiment. In recent years, this class of methods has been instrumental in shaping the community's understanding of ion adsorption to interfaces: The phenomenon is driven by a repartitioning of the solvent energy, because the rotational anisotropy near the interface results in solvent polarizability differences [103]. This is corroborated by our data, even without explicit polarization in the force field, as seen in water molecules adjacent to larger cavities adopting stronger orientational preferences than near small cavities.

We find an orientational preference of water molecules adjacent to molecular sized cavities, having dipole moments with nonzero radial component (relative to the cavity). In particular, the positive pole tends to be oriented directly toward the cavity when not tangential to the surface, as seen in Figure 2.5. This water structuring may in part explain why anions tend to adsorb to interfaces more closely than cations immediately adjacent to the interface [75], and provide a Coulombic driving force for the adsorption peaks observed.

We observe an unexpected localization tendency of both ion species near a cavity of radius 0.15-0.45 nm. As cavity size increases, the ion density field responds more quickly than that of the water. This is likely related to the elastic nature of the water molecule hydrogen bond network about a cavity. While the water molecules adjacent to the cavity display qualitatively similar orientational profiles as the cavity grows, the elastic response of the water molecule hydrogen bond network to the cavity's increased volume implies decreased orientational freedom of participating water molecules. This entropy loss correlates with a lowered local dielectric [104], and could mechanistically explain why ions begin to deplete from the cavity interface prior to the hydrogen bond network breaking. The shoulder hills that develop for both Na^+ and Cl^- adjacent to cavities of $\sim$0.4 nm radius indicate an adsorption tendency that is not observed at the planar interface without any explicit three-body interactions to model polarizability, as has been demonstrated especially for polarizable ions like iodide [105, 106]. These shoulder features may have a Coulombic driving force related to the multiple prominent orientational basins of water

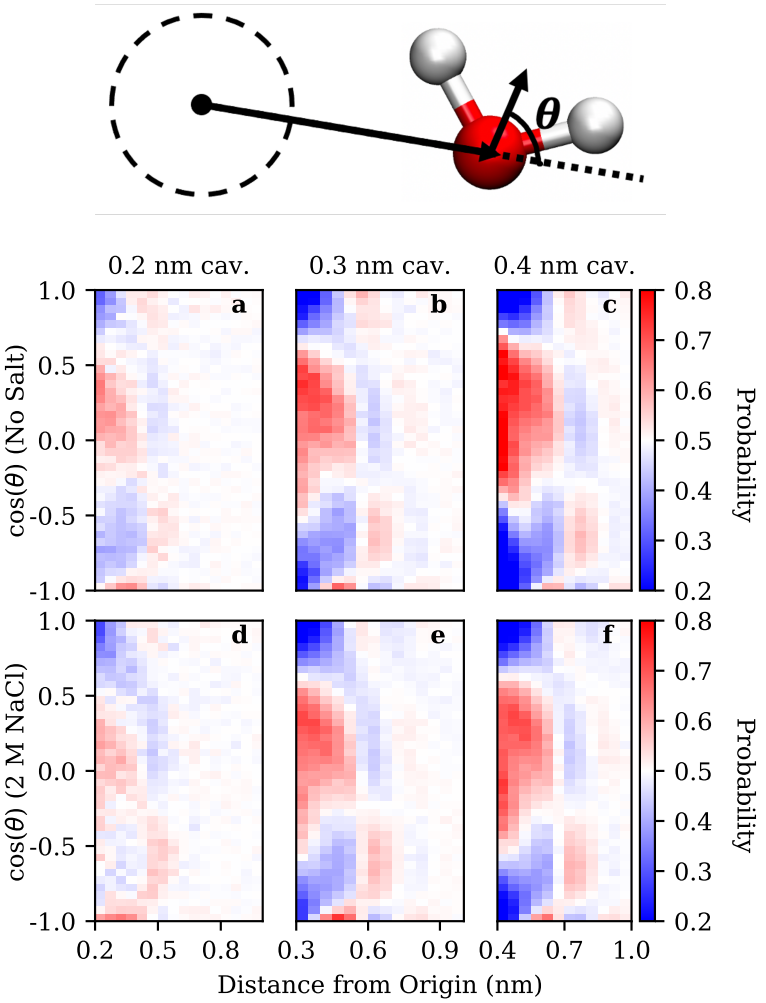

Figure 2.5: Orientation of water molecules as a function of radial distance from the origin, as measured by the cosine between the vectors (i) from the origin to the water oxygen and (ii) from the water oxygen to the average position of its two hydrogen atoms. Top graphic is an illustration of the angle θ. Subfigures are arranged in a grid for different experimental setups, with the top row indicating neat water, bottom row aqueous 2 M NaCl, and columns corresponding to 0.2 nm, 0.3 nm, and 0.4 nm cavities. Between cavities of radius 0.2 nm and 0.4 nm, the presence of NaCl does not influence the average water molecule orientation. Larger cavities are positively correlated with the intensity of water molecule orientational preference.

molecules within ∼0.5 nm from the interface. Since non-polarizable force fields display diminished ion adsorption effects compared to experiments and polarizable models [107], the intense localization seen here is likely a robust observation.

While it is striking that the structural characteristics of water undergo so little change while the ions respond dramatically, this too can be explained. In order to maintain favorable intermolecular attractions despite the presence of a hydrophobic barrier, the water molecules adopt a local density much greater than the bulk. In order for the water molecule density field to deplete from the interface as the ions do, these hydrogen bonds would be compromised, and so it does not occur (until those hydrogen bonds are geometrically incompatible with the cavity at radii ~ 1 nm). The free energy associated with maintaining these hydrogen bonds is evidently more impactful than the presence of ions, which are present at low concentrations relative to the water. The mole ratio of water to ions is also large enough to explain why the density of water molecules adjacent to the cavity relative to bulk concentration appears unperturbed by changing ion concentration.

2.7 Conclusion

Using 2D WHAM, biasing only one CV and recovering the probability distribution of another, we observed differences in the water and ion density fields in the region adjacent to a molecular-sized cavity. By considering the local orientation of water molecules, we identify dipole distributions as a contributing factor to this density decoupling, which is an observation likely to hold for other ion species in solution due to its Coulombic nature. Such ion adsorption problems have been of interest in the atmospheric sciences community for many years [108, 109]. These findings are complementary to the planar interface studies that have been conducted previously, and provide additional mechanistic insight toward understanding the hydrophobic effect in aqueous ionic media.

The effects of ion identity (size and charge effects) have not been examined here. Fluoride, for example, owing to its small size, displays adsorption tendencies more similar to a point charge than does chloride, and does not display surface adsorption at liquid/vapor interfaces as larger anions do [110]. It may be informative to perform a comparative study between different cations and anions to better understand the role of ion hydration at the small cavity interface.

The addition of a polarizable force field, while computationally expensive, would likely provide valuable insight as well: ion adsorption is more pronounced for highly polarizable species [107], and by neglecting explicit three-body interactions in the force field the ion adsorption behavior is affected. Ideally, both water molecules and ions should be treated this way. Using such a force field where adsorption to the planar interface is prominent, the nonmonatonic features of probability density traces shown here may be substantially enhanced.

The simplicity of the non-polarizable models employed here speak to the robustness of the observed density decoupling between water and ion species, which may have ramifications for a broad array of classical simulation studies that employ small hydrophobic solutes in aqueous salt solution.

2.8 Appendix: Additional Method Development

The work described in this chapter to characterize cavities in aqueous solution was achieved using 2D Umbrella Sampling as described in the main text, and is to be published as such. Prior to settling on that strategy, early attempts at generating cavities in MD simulation placed multiple degenerate biases on the distances of individual water molecules from the origin, resulting in a high-dimensional biased space that could not be efficiently treated by traditional WHAM. However, so long as states are being sampled in simulation that adequately capture to presence of a cavity, it should in principle be possible to determine $P_{eq}(R)$ from a set of biased simulations without relying on WHAM explicitly. Toward this goal, an algorithm was developed to align re-weighted segments of $P_{eq}(R)$ from each unbiased simulation $P_i^u(R)$ using geometric criteria, such that the residuals between $P_{eq}(R)$ and each window's estimate $P_i^u(R)$ at a given value of R is minimized. In what follows, the variable of interest in the study of cavities, R, is replaced with the generic variable x as the technique is generally applicable. This new method requires only the instantaneous values of the total applied bias energy to reweight probability distributions from individual simulations and then align simulation windows to provide a low-dimensional composite distribution along a desired CV.

Here we apply a generalization of the WHAM equations with a geometric emphasis, a numerical optimization technique that minimizes a cost function and is based on the multivariate Householder methods (which includes the multivariate Newton's method and Halley's method of root-finding) [111]. The probability distributions from windows indexed on i and coefficients F_i are the same as in WHAM, although the term that weights observations from different windows must be left in a more generic form (which comes from an intermediate step in the derivation of the weights Eq. 2.10 here, from [93]):

$$w_i(x) = \frac{\exp(2\beta F_i)\,[\sigma^2(\widehat{P_i^u}(x))]^{-1}}{\sum_j \exp(2\beta F_j)\,[\sigma^2(\widehat{P_j^u}(x))]^{-1}} \tag{2.12}$$

Here the symbol σ^2 denotes the statistical error of an associated quantity, and will be precisely defined below. One arrives at this expression following the method of Lagrange multipliers as discussed for equation (2.10), and then typically further substitutions are made based upon assumptions regarding the form of $\sigma^2(\widehat{P_i^u}(x))$. See the WHAM derivation from Roux 2001 for details on those steps [93]. In order to evaluate these weights, we must define the statistical error $\sigma^2(\widehat{P_i^u}(x))$.

For a stationary, stochastic, time-reversible process from simulation i with N_i independent observations indexed on n, one may define histogram bins m of width Δ, for which an "indicator function" ψ exists. Notice that this is a weighted counter, measuring whether an observable $\hat{X}$ falls within a given bin.

$$\psi_{m,i,n} = \begin{cases} \exp(\beta V_{i,n}) & \text{if } \hat{X}_{i,n} \in [m\Delta - \frac{\Delta}{2}, m\Delta + \frac{\Delta}{2}) \\ 0 & \text{else} \end{cases} \tag{2.13}$$

The resulting histogram, $h_{m,i}$ is their sum,

$$h_{m,i} = \sum_{n=1}^{N_i} \psi_{m,i,n}, \tag{2.14}$$

which, in the limit of infinite length and fine discretization, converges to the the unbiased probability distribution within that window [112]. It is implied that the CV value x corresponds with bin m,

$$\lim_{\Delta\to 0}\lim_{N_i\to\infty}\frac{h_{m,i}}{N_i\Delta}\to P_i^u(x). \tag{2.15}$$

If simulations are conducted in a way that samples phase space adequately, $(\frac{1}{N_i\Delta})*h_{m,i}\approx P_i^u(x)$, then the statistical error associated with a given histogram bin may be related to the uncertainty of the estimated probability distribution from simulation. To differentiate the actual probability from the estimate obtained in simulation, a hat will be used for the estimate, i.e. $\widehat{P_i^u}(x)$,

$$\left(\frac{1}{N_i\Delta}\right)^2 * \sigma^2(h_{m,i}) = \sigma^2(\widehat{P_i^u}(x)). \tag{2.16}$$

It is a well-known result ([112]) that the maximum likelihood estimate of the variance of a random variable whose average is $\langle\psi_{m,i}\rangle = N_i^{-1}\sum_{n=1}^{N_i}\psi_{m,i,n}$ may be expressed concisely as

$$\sigma^2(\psi_{m,i}) = \frac{g_{m,i}}{N_i}[\langle\psi_{m,i}^2\rangle - \langle\psi_{m,i}\rangle^2], \tag{2.17}$$

for statistical inefficiency g given as $(1+2\tau)$ for correlation time τ in units of the sampling interval. This result permits the variance within a histogram bin to be computed [112].

$$h_{m,i} = N_i\left[\frac{1}{N_i}\sum_{n=1}^{N_i}\psi_{m,i,n}\right] \tag{2.18}$$

$$\sigma^2(h_{m,i}) = (N_i)^2 * \sigma^2(\psi_{m,i}) \tag{2.19}$$

Putting the results of equations (2.16), (2.17), and (2.19) together, the statistical error associated with $\widehat{P_i^u}(x)$ may be written:

$$\sigma^2(\widehat{P_i^u}(x)) = \frac{g_{m,i}}{N_i\Delta^2}\left(\langle\psi_{m,i}^2\rangle - \langle\psi_{m,i}\rangle^2\right). \tag{2.20}$$

With $\sigma^2(\widehat{P_i^u}(x))$ known, the weights $w_i(x)$ can be obtained for a given set of $\{F_i\}$, which must be determined iteratively. An initial guess is made from the biased ensemble.

$$F_i = -\beta^{-1}\log\left(\frac{\int \mathrm{d}\vec{r}^N\exp(-\beta U)}{\int \mathrm{d}\vec{r}^N\exp(-\beta U)\exp(-\beta V_i)}\right) = \beta^{-1}\log\langle\exp(\beta V_i)\rangle_i.$$

While this equality is exact, the large range of values V_i may assume results in an expected value that converges slowly. Further iterative optimization of the set F_i is achieved via root-finding, as detailed in the next section.

Before departing, it is also useful to note that this formalism provides an avenue for estimating the uncertainty at a given point along $P_{eq}(x)$. One may write an expression for the variance between estimates of $P_{eq}(x)$ at a given x as taken from i simulation windows:

$$\sigma^2(\widehat{P_{eq}}(x)) = \sum_i w_i(x)^2 \, \exp(-2\beta F_i) \, \sigma^2(\widehat{P_i^u}(x)). \tag{2.21}$$

We observe that, when the statistically optimum weights (eq. 2.12) are substituted into (2.21), the variance of the optimally-weighted distribution is

$$\sigma^2(\widehat{P_{eq}}(x)) = \left[\sum_i \exp(2\beta F_i)[\sigma^2 \widehat{P_i^u}(x)]^{-1} \right]^{-1}. \tag{2.22}$$

This uncertainty estimate is easily accessible from simulation data and the determined set of $\{F_i\}$.

2.8.1 Optimization Via Root-Finding

The maximum likelihood estimation of F_i should minimize the residuals between $P_{eq}(x)$ and its constituent window estimates $\exp(\beta F_i)P_i^u(x)$, at each x for all windows. This is done variationally through the iteratively reweighted least squares method, updating subsequent estimates according to the derivatives of some cost function. Toward this end, we define the cost function f:

$$f \equiv \int \mathrm{d}x \, \sum_i w_i(x) \Big(P_{eq}(x) - \exp(\beta F_i) P_i^u(x)) \Big)^2 \tag{2.23}$$

The appropriate F coefficients are determined iteratively in a self-consistent manner via the multivariate Newton method. Alternative multivariate root-finding algorithms may be applied at the user's discretion. The quantities in column vector ∇f includes the partial derivatives of f that will decay to the zero vector as optimization proceeds.

$$\nabla f = \left[\frac{\partial f}{\partial F_1}, \frac{\partial f}{\partial F_2}, ..., \frac{\partial f}{\partial F_i} \right]^T \tag{2.24}$$

Also required is the matrix of mixed partial second derivatives, the Hessian matrix $\mathbf{H}$ of f with elements $H_{j,k}$:

$$H_{j,k} = \frac{\partial^2 f}{\partial F_j \partial F_k}. \tag{2.25}$$

Equations (2.24) and (2.25) together permit the set of coefficients F_i to be updated from iteration a to iteration $a+1$ [112].

$$\vec{F}_{a+1} = \vec{F}_a - \mathbf{H}^{-1} \nabla f \tag{2.26}$$

Recognizing that f is an explicit function of $\{F_i\}$, these derivatives may be taken analytically. Since the derivation variable F_j appears only within exponential functions, these derivatives take on relatively convenient forms:

$$\frac{\partial w_i(x)}{\partial F_j} = 2\beta w_i(x) \Big[\delta_{ij} - w_j(x) \Big] \tag{2.27}$$

$$\frac{\partial P_{eq}(x)}{\partial F_j} = \beta w_j(x) \Big\{ \exp(\beta F_j) P_j^u(x) - 2P_{eq}(x) \Big\} \tag{2.28}$$

Here δ_{ij} is the Kronecker delta. The derivatives of f are composed of these constituent terms. Combining the results of equations 2.27 and 2.28 with the observation that $\frac{\partial P_i^u(x)}{\partial F_j} = 0$, because $P_i^u(x)$ is constructed entirely from simulation data, there is now sufficient information to compute ∇f and $H_{j,k}$ for a given set of coefficients $\vec{F}_a$.

In practice, one F coefficient is fixed, invariant, to help prevent inconsistencies in normalization. The most physically reasonable solution is to include one unbiased simulation window in the analysis, which will have $F_0 = 0$ by definition. From equation 2.11, the F coefficient of the unbiased ensemble (where $V_i = 0$ always) is known with certainty *a priori*. However, provided normalization of the resulting probability distribution is performed afterward, for practical purposes any F coefficient may be fixed and the result unchanged. The multivariate Newton's method applied in this way is conceptually analogous to the traditional WHAM normalization self-consistency formula,

$$\exp(-\beta F_i) = \int \mathrm{d}x \exp(-\beta V_i(x)) P_{eq}(x). \tag{2.29}$$

Re-evaluating F_i from this equation enforces consistency between estimates for the coefficients F_i and its definition from configuration integrals in equation (2.6). Notice that if $V_i(x) = 0$ for all x, then F_i is identically zero. It is this fact that ensures the set F_i are self-consistent in the present derivation. The application here is more robust than traditional WHAM which assumes V is an explicit function of x, because in general $\langle \delta(\hat{X} - x) \exp(\beta V_i) \rangle_b \neq \exp(\beta V_i(x)) \langle \delta(\hat{X} - x) \rangle_b$. In the special case where V_i is a function exclusively of the CV x, where this equality holds, this derivation becomes conceptually synonymous with WHAM following the manipulation of $w_i(x)$ laid out in Roux 2001 [93].

By generalizing the form of the energetic bias utilized in Umbrella Sampling to be independent of the CV of interest, we have identified an additional avenue for analyzing cavity formation phenomena using the GROMACS MD framework. This approach is generally applicable to other systems for which this high-dimensional biasing strategy is pertinent. The results obtained are consistent with predictions obtained via 2D WHAM, and with high resolution (Figure 2.6).

This reweighting scheme suffers from statistical noise due to the difference between the ensemble being sampled and the ensemble of interest. As the two ensembles become less similar, the overlap between the ensemble being biased and the desired ensemble diminishes. Further, we do not make assumptions about $\langle \psi_{m,i}^2 \rangle - \langle \psi_{m,i} \rangle^2$. In situations where V_i varies greatly, which occurs in regions of phase space with a steep underlying free energy landscape, the average $\langle \psi_{m,i} \rangle$ converges slowly, and the problem is exacerbated when the processes of interest are not statistically well-represented in the data. Here, for example, queries regarding water molecules immediately adjacent to the cavity interface are poorly reported, by construction, due to the relatively small spherical shell volume that is relevant to the query.

A useful metric for assessing the impact of this inefficiency is to sort simulation configurations by weight and plot the proportion of the sum total weight that is accounted for by the top x configurations (Figure 2.7).

For an unweighted system, the fraction of sum total weight as a function of number of configurations considered is linear, because every configuration carries the

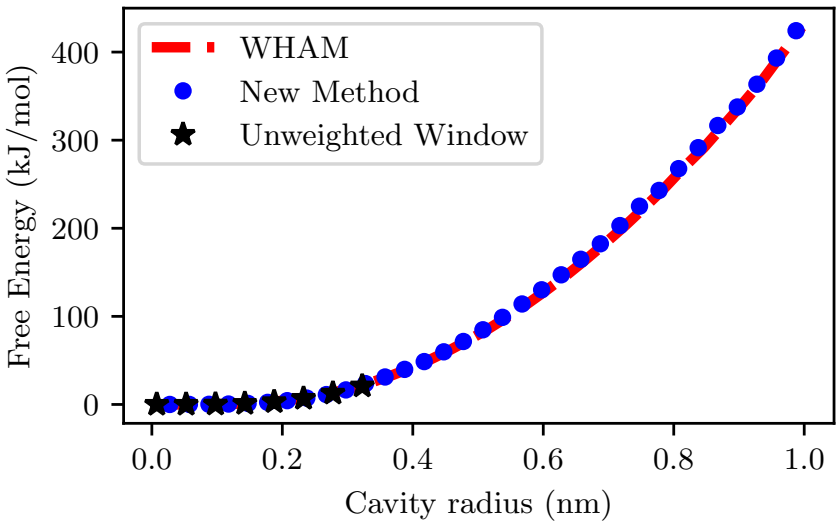

Figure 2.6: In neat water, free energy of evacuating a spherical cavity of radius R may be obtained without adapted sampling (black), with 2D WHAM (red), or the alternative method discussed here (blue). This case study illustrates the accuracy and effectiveness of the WHAM alternative relative the its parent method, despite the greater generality afforded here.

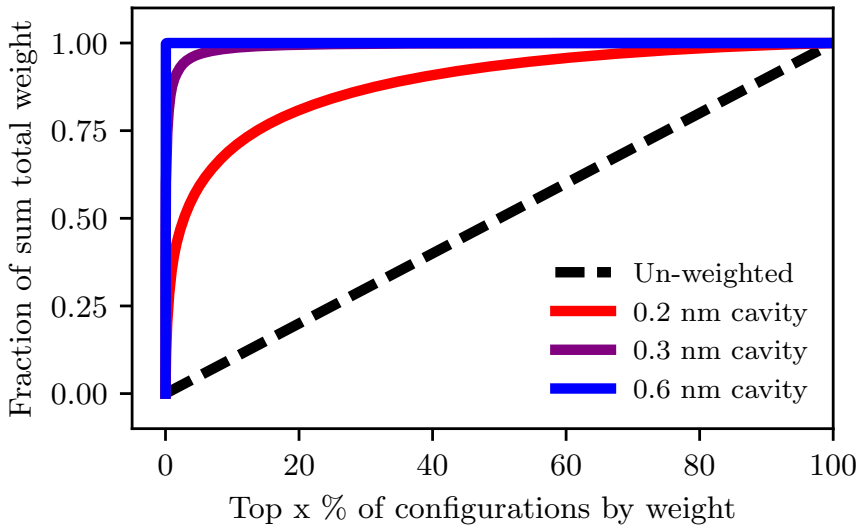

Figure 2.7: When simulation frames are arranged in order of descending weight magnitude, the sum of the most heavily weighted configurations over the sum of all frame weights does not approach unity linearly as would be the case if each frame were given the same weight. This effect is exacerbated in simulations of larger cavities, such that ensemble averages computed via re-weighting are dominated by very few configurations. This example was conducted under no salt conditions.

same weight. Simulations that contain some frames weighted very highly compared to others will deviate strongly from linearity, revealing that a few configurations (once reweighted) dominate the statistics acquired for the entire system. In such cases, prohibitive simulation lengths would be required for distributions to converge, and alternate strategies must be used.

We have presented here an application of the classic Torrie-Valleau reweighting scheme in a manner analogous to WHAM, with the advantage over traditional WHAM being the ability to compute the free energy profile along a collective variable axis that was not biased explicitly in the simulation. This method has wide-reaching applicability for collective variables defined from simple counting procedures that are discontinuous variables. Here this method was applied to cavity formation, defining the radius of a spherical cavity to be the vector length from an arbitrary center to the nearest heavy atom. Given the generality of this technique, for small cavities we have computed not only the free energy associated with spontaneously generating such a cavity, but have also acquired equilibrium statistics on the ensembles of interest, including local density, orientation, and hydrogen bonding accessed via the reweighting procedure discussed here. While only an efficient technique in regions of phase space where the variable being biased is sufficiently similar to the CV of interest, its flexibility is appreciated when concerned with the vast dimensionality reduction inherent to the CV of interest.

Chapter 3

Ion-Dependent Protein-Surface Interactions from Intrinsic Solvent Response

3.1 Abstract

[1]The phyllosilicate mineral muscovite mica is widely used as a surface template for the patterning of macromolecules, yet a molecular understanding of its surface chemistry under varying solution conditions, required to predict and control the self-assembly of adsorbed species, is lacking. We utilize all-atom molecular dynamics simulations in conjunction with an electrostatic analysis based in Local Molecular Field theory that affords a clean separation of long-range and short-range electrostatics. Using water polarization response as a measure of the electric fields that arise from patterned, surface-bound ions that direct the adsorption of charged macromolecules, we apply a Landau theory of forces induced by asymmetrically-polarized surfaces to compute protein-surface interactions for two muscovite-binding proteins (DHR10-mica6 and ^{C98}RhuA). Comparison of the pressure between surface and protein in high concentration KCl and NaCl aqueous solutions reveals ion-specific differences in far-field protein-surface interactions, neatly capturing the ability of ions to modulate the surface charge of muscovite that in turn selectively attracts one binding face of each protein over all others.

3.2 Background

Solid-liquid interfaces promote many phenomena that are of central importance in fields ranging from geochemistry to energy science [113, 114]. Moreover, the templating interactions between mineral surfaces and biomolecules are known to have far-reaching implications in our understanding of the origins of life and the synthesis of novel functional materials [115, 116]. Natural bioinorganic materials exhibit a diversity of physical properties adapted for organismal fitness [117,

[1]Reproduced with permission from Jesse Prelesnik, Robert Alberstein, Shuai Zhang, Harley Pyles, David Baker, Jim Pfaendtner, James De Yoreo, Akif Tezcan, Richard Remsing, and Christopher Mundy. Ion-Dependent Protein-Surface Interactions from Intrinsic Solvent Response. Proc. Natl. Acad. Sci. U. S. A., Accepted. (2021) © United States National Academy of Sciences.

118] (*e.g.*, nacre, composed of calcium carbonate platelets interconnected by proteins and other biopolymers to provide both strength and flexibility, and collagen, which forms bone, tendon, and cartilage as a function of mineral content) that are selectively produced under controlled conditions that modulate heterogeneous self-assembly. However, our fundamental understanding of such interfaces at the molecular scale has remained limited, precluding the development of synthetic biomaterials from first principles. Much progress has been made in experimentally probing solid-solution interfaces. Surface sensitive techniques such as X-ray reflectivity [119], second-harmonic spectroscopy [120], three-dimensional fast force mapping [121], dynamic force spectroscopy [122], and the surface force apparatus [123] continue to contribute to our understanding of how the properties of the solid surface in conjunction with the solution conditions (*e.g.*, ionic strength, electrolyte type) impact interfacial speciation and ultimately modulate the forces between surfaces. These concepts become particularly important at the nanoscale where the lengthscales of molecular fluctuations near an interface becomes comparable to that of the solution response (*e.g.*, screening) [124]. This convergence of scales, in turn, enables processes such as oriented attachment as a new pathway for synthesis of nanocrystals, which is known to be inextricably tied to the nature of both the surface and the solution conditions [125, 126, 127]. It is clear that our understanding of such interfacial phenomena requires a theoretical platform that effectively captures molecular scale details to understand physical outcomes at the macroscopic scale.

Herein, we utilize designed proteins as a sensitive probe of the underlying solution conditions and their effect on the synthesis of materials on a solid, templating surface. The construction of novel, synthetic biomaterials from designed, information-rich macromolecular building blocks (*e.g.*, proteins) promises precisely tailored material properties, but is confounded by the vast parameter space underpinning their self-assembly [128, 129]. Of these, materials that contain a high degree of ordering are particularly desirable targets, as they are amenable to structure determination methods, can serve as scaffolds with well-defined spacings for functionalization with chemical groups, and possess specific physicochemical properties based on the composition (and arrangement) of the individual building units [130]. Template materials with near-perfect basal cleavage are naturally well-suited to the production of low-dimensional ordered materials, as they provide atomically flat crystalline surfaces with well-defined surface properties. The phyllosilicate mineral mica, particularly the $C2/c$ symmetric pseudohexagonal muscovite, is an ideal example owing to its other desirable physical properties: it is chemically and thermally inert, insoluble, lightweight, insulating, flexible, inexpensive, and hydrophilic [131]. Consequently, various nanoscale objects such as DNA have been successfully adsorbed onto the atomically flat muscovite from aqueous solution in an orientationally specific manner, an important step toward building complex structures [132, 133].

Adsorbates of particular interest are proteins, which can be *de novo* designed and/or mutated to modulate mica binding affinity. Recently, a Designed Helical Repeat (DHR) protein featuring tandem repeats of helix-turn-helix subunits [134], designated DHR10-micaX (where X equals the number of repeats), was designed such that glutamate residues precisely matched the lattice spacing of the hexagonal array of potassium ions on the muscovite surface [135]. This allows the rod-shaped proteins to assemble in alignment with the three principal axes of the underlying

ion lattice as measured by atomic force microscopy (AFM) in a variety of solution conditions [135, 136]. Interestingly, its surface coverage and uniformity in orientation are sensitive to the ionic strength of the aqueous medium, among other variables such as protein concentration and ion species (Fig 3.1A/B).

At 3 M KCl, the 18-repeat DHR10-mica18 assembles into a 2D liquid-crystal-like phase similar to smectic or nematic phases with protein rods co-aligned and ordered into discrete rows that persist for many hundreds of nanometers, whereas at only 100 mM KCl, protein rods form clustered domains along all three available axes in a phase not theoretically predicted for spherocylinders using hard-core models [137]. This phase is not merely a kinetically trapped state, evidenced by its reversible transitions to and from this state upon adjusting ionic strength, suggesting that some specific protein-protein and/or protein-surface interaction must act to stabilize domains [135, 138]. Furthermore, exchanging potassium for sodium increases the critical concentration at which the 2D liquid crystal phase is formed, indicating a diminished binding affinity in sodium salt solution, and at low protein densities a dominant binding orientation is not obtained (see Fig 3.1B).

The principle of heteroepitaxial growth of proteins on muscovite appears to be quite general [139, 140, 141]. For example, the engineered protein ^{C98}RhuA also shows surface-induced crystallization into condition-dependent morphologies [142] despite an apparent disagreement between the inherent C_4 symmetry of the protein and the $C2/c$ symmetry of the pseudohexagonal muscovite surface. This occurs as the approximately square ($\sim 7 \times 7$ nm^2) proteins form corner-to-corner disulfide bonds to self-assemble into superstructures hundreds of nanometers across, locally oriented along either one or two directions depending upon solution concentrations of electrolytes as determined by AFM. Distinct crystalline morphologies can be selectively formed by tuning the concentrations and identity of the solution ions, despite using the same surface and building block (Fig 3.6) [142]. The examples of DHR and RhuA systems clearly illustrate that protein-muscovite binding behavior can be extensively tuned via ion effects to control the pathways and outcomes of heteroepitaxial protein self-assembly. See Fig 3.7 for amino acid sequences and protein schematics.

The inability of theory to accurately predict *a priori* each outcome hinders the development of more sophisticated, hierarchically structured, assemblies. Simple hard-wall packing models fail to capture all of the necessary physics, as specific protein-surface and protein-protein interactions are not considered [137]. Furthermore, high ionic strength conditions and specific ion effects are theoretically challenging even in the bulk solution phase, as Debye-Hückel coefficients cannot account for electrolyte non-ideality beyond dilute concentrations [143]. The present work approaches this knowledge gap with the intent of first understanding the solvent-structuring properties of each assembly component in isolation to compute far-field surface-protein interactions (Fig 3.1C).

We achieve this with three key modeling advancements:

1. We improve the accuracy of an empirical atomistic MD model of muscovite mica relative to experiment by treating aluminosilicate substitutions using a pseudo-atom with a mean-field philosophy, stoichiometrically averaging simulation parameters. Doing so enhances the chemical potential of muscovite for bound cations (a well-documented deficiency of empirical models at aqueous interfaces [144, 145]) and also eliminates measurement artifacts associated

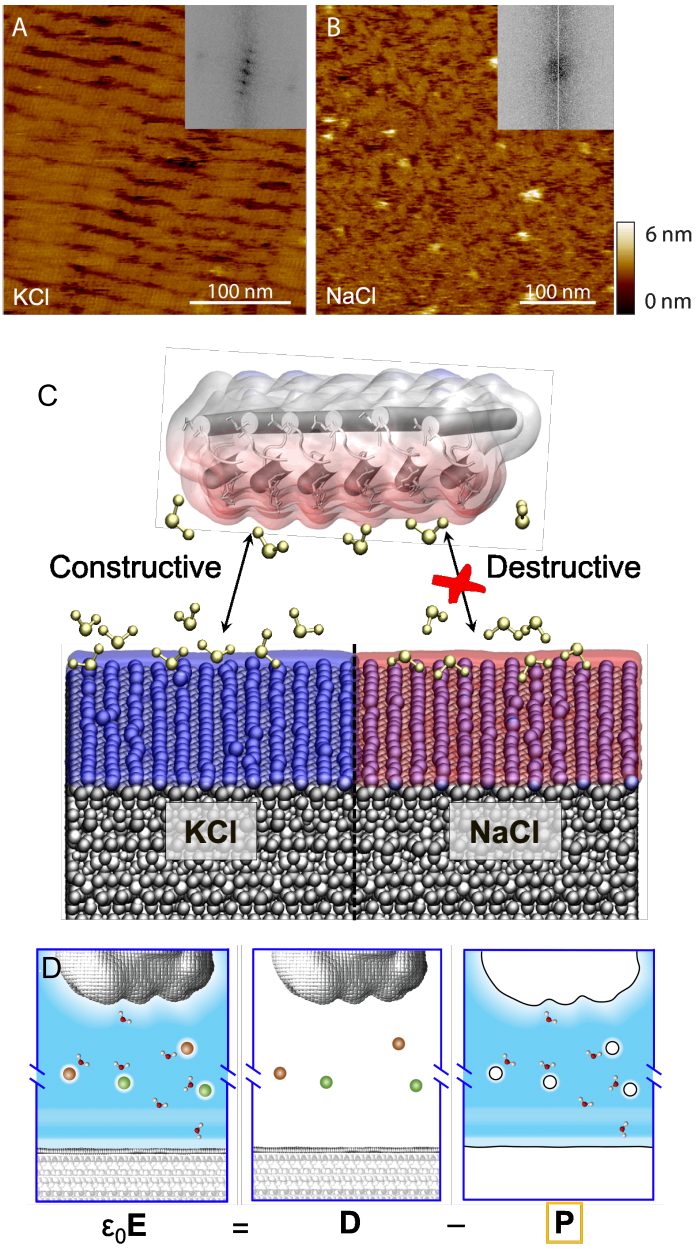

Figure 3.1: (A/B): AFM images of DHR10-Mica18 assembly. (A) 0.1 μM DHR10-Mica18 in 3 M KCl. (B) 0.1 μM DHR10-Mica18 in 3 M NaCl. The insets are Fast Fourier Transforms. The buffer is 20 mM Tris, pH 7. (C): Schematic representation of our key finding that macroscopic water polarization over muscovite is reversed in the presence of KCl vs. NaCl, which leads to either constructive or destructive interference with the polarization of a protein's surface depending on specific ion effects. For DHR10-mica6 (shown) the relative attraction strengths of the protein's front or back faces to muscovite correlate with reduced orientational specificity observed in AFM (panel B). (D): Illustration of decomposing the total electric field $\mathbf{E}$ into displacement field $\mathbf{D}$, consisting of background charges and the two surfaces (treated as fixed), and the electric polarization $\mathbf{P}$ due to water orientations, which is our focus here. This relation is written in SI units, where ϵ_0 is vacuum permittivity. Simulations of protein and muscovite are conducted separately.

with specific substitution patterns.

2. We quantify aqueous response through the lens of Local Molecular Field (LMF) theory, which isolates the low-intensity, long-range component of electrostatic potential [146, 147, 148, 149]. Doing so resolves a dipolar solvent structure that is otherwise hidden beneath higher-intensity, short range contributions to the total signal. We apply this approach to measure far-field electric polarization (Fig 3.1D).

3. We take the far-field polarization as measured from simulation as a scalar order parameter in the Landau free energy expansion between two planar surfaces that induce forces on one another via solvent polarization [150, 151]. Simulation data provide boundary values and a reference curve to fit an analytical solution of the order parameter and extract phenomenological parameters needed to compute the interaction pressure. By solving an analytical form of the Landau order parameter for asymmetric boundary conditions, we apply this approach to protein-surface interactions.

With this modeling approach, we quantify interaction strength between asymmetrically polarized surfaces via the long-range polarization-induced pressure, and relate intrinsic far-field specific ion effects from atomistic simulations to experimentally observed changes in assembly outcome.

3.3 Accurate Mean-Field Model of Muscovite

In an effort to improve performance of the MD model relative to experiment, we compared the behavior of a standard atomistic representation of muscovite to its mean-field variant described here (both a 5x3 unit cell, using the INTERFACE force field (IFF) [152]). We analyzed molecular densities normal to the surface (ZDFs) to determine the fraction of muscovite binding sites occupied by cations, which has been previously reported as underrepresenting the experimental limit of ∼100% [144, 153]. At 3 M KCl the explicit-substitution model retains less than 85% of its first-layer cations, which are present on the surface in a bimodal distribution (Fig 3.2A).

The mean-field model improves this occupancy to $97 \pm 1\%$ (Fig 3.2B) and provides a hexagonal lateral pattern that matches AFM measurements [135], which is obscured in the explicit substitution variant because solvent patterning over muscovite is imprinted by the lateral positions of Al sites (Fig 3.2C,D). The mean-field treatment provides much more realistic depictions of the bulk-scale surface than standard explicit-substitution representations in terms of modeling the perfect hexagonal array of ions that proteins are designed against.

Further simulations were conducted on large (20x12 unit cell) slabs of IFF or CLAYFF [154] mean-field muscovite, comparing solution conditions 3 M KCl to 3 M NaCl to mimic experiments. Measurements of molecular densities reveal subtle lateral inhomogeneity that persists despite the mean-field treatment, which is reflective of the pseudohexagonal symmetry of muscovite (Fig 3.8). A lateral probability density of water oxygen within a narrow volume slice above the muscovite surface shows characteristic striping (Fig 3.3). This finding is in alignment with prior simulation studies of parallel muscovite slabs at close separations, which show preference

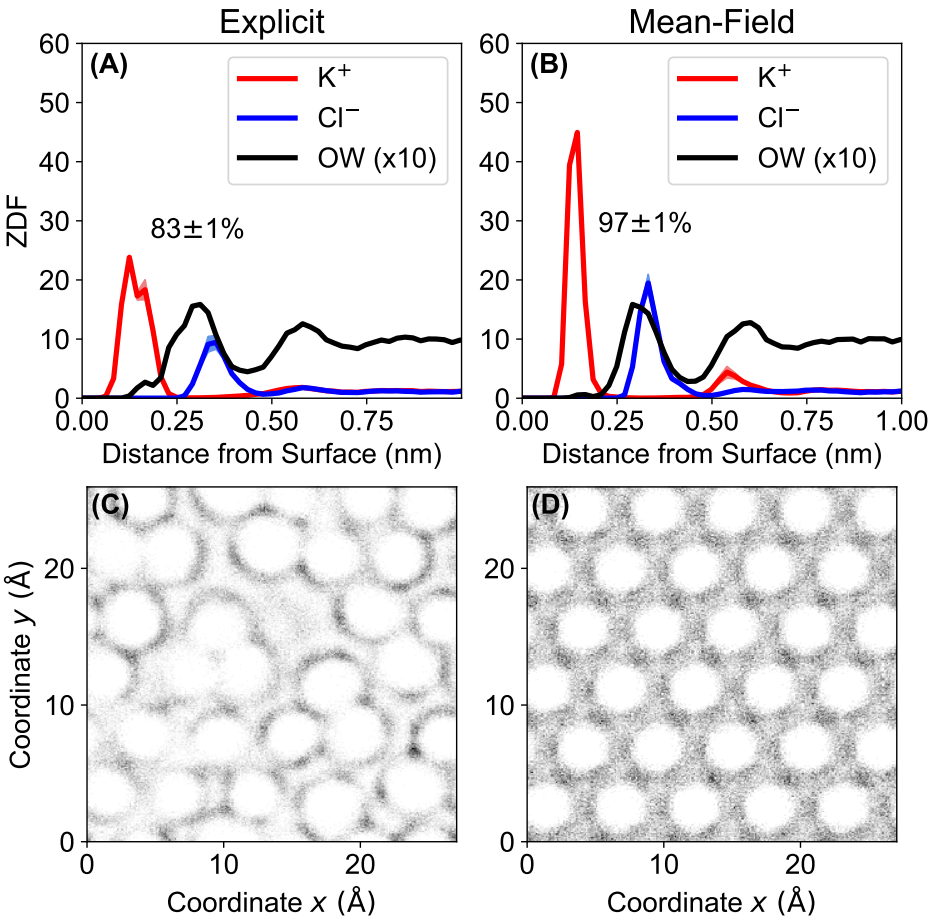

Figure 3.2: For 3 M KCl aqueous conditions, we have (A/B): Averaged probability density from atomic coordinates as a function of perpendicular distance from muscovite bedrock (ZDF), normalized to bulk, for (A) muscovite including explicit Al substitutions, and (B) muscovite in the mean-field substitution representation. Both are using the INTERFACE force field. Percentages are fraction of cation binding sites occupied, measured by integrating up to the first minimum in the potassium density. (C/D): Average probability density of lateral water oxygen positions a distance of 3.0 ± 0.3 Å from the muscovite surface as measured from the surface-exposed oxygen atoms. Patterning is uneven under explicit Al substitution (C) and organized under the mean-field treatment (D).

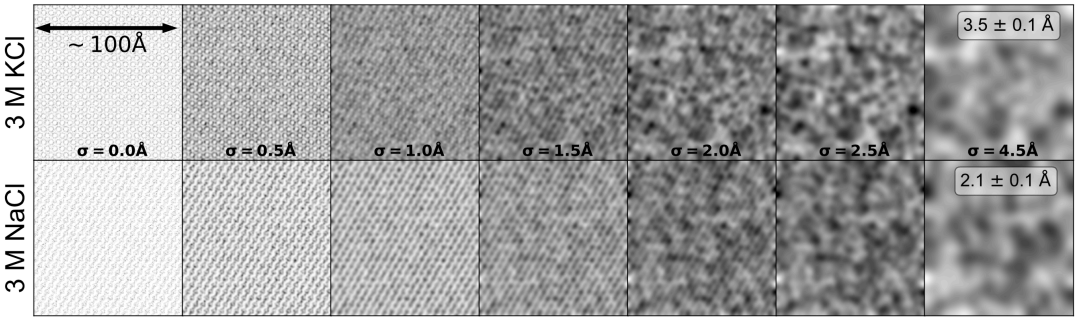

Figure 3.3: Lateral probability distribution of water molecules occupying the narrow region of 3.5 ± 0.1 Å (top row, 3 M KCl with IFF mean-field muscovite) or 2.1 ± 0.1 Å (bottom row, 3 M NaCl with IFF mean-field muscovite). This same probability density is smoothed by a Gaussian of width σ (a KDE of probability), in columns left-to-right, from σ=0.0 Å (no smoothing) to σ=4.5 Å in increments of 0.5 Å. Characteristic streaking can be seen from the lower left to upper right corners of each panel, up to a smoothing length of about 2.0 σ. This illustrates that some distinctive atomistic signatures are inaccessible to far-field smoothing procedures that use Gaussians of width $\sigma = 4.5$ Å.

for a single alignment orientation that is mediated by multiple water layers [155]. The anisotropy of muscovite has been experimentally shown to universally impose lateral asymmetry on the orientation of bound macromolecules, designed or natural [135, 142, 132, 133].

The attractions that give rise to surface assembly are complicated and numerous, but the electrostatic potential is particularly valuable here due to its slow decay. Solvent-structuring properties intrinsic to muscovite permeate far into the bulk and dominate the driving forces for macromolecules in solution, up to separations of ∼1 nm. Before short-range interactions (*i.e.*, confinement effects) manifest, long-range assembly forces are responsible for orienting the preferred face pointed toward the surface. As a starting point for understanding assembly, we have focused on the dilute limit in which protein-protein interactions are negligible. This dilute limit coincides with experiments, which typically use less than 1 μM of protein (with mean interparticle separations $\mathcal{O}$(100 nm)), and indications of solution-phase protein-protein aggregation are absent. We would also like to emphasize that polarization-induced interactions discussed here are long-ranged and are the driving force until about a nanometer from the surface, at which point the protein cannot easily reorient. The many-body effect of multiple proteins interacting on the surface is beyond the scope of the current work.

3.4 Quantifying Long-Range Solvent Response

We seek to isolate the far-field component of the full electrostatic potential computed from simulation, which is challenging to measure directly due to measurement noise over an intrinsically low-intensity interaction [156]. By separating the slowly-varying tail of the electric potential via the LMF formalism, we can accurately capture surface-induced electrostatic interactions that emanate far into the bulk [148]. In practice, this is accomplished by taking a convolution of atomic point charges with

a Gaussian kernel, as in the LMF smoothed charge density:

$$\rho^{\sigma}(\mathbf{r}) = \int \mathrm{d}\mathbf{r}'\rho(\mathbf{r}')\rho_G(\mathbf{r}-\mathbf{r}') \tag{3.1}$$

$$\rho_G(\mathbf{r}) = \frac{1}{\pi^{3/2}\sigma^3}\exp\big(-|\mathbf{r}|^2/\sigma^2\big). \tag{3.2}$$

The optimum value for the Gaussian width σ has been well-studied, and in aqueous solution is taken to be $\sigma = 0.45$ nm [147, 157]. Because the density is being smoothed or filtered over a length scale, σ, that is larger than molecular scales, atomic-scale features of the muscovite surface are washed out in the far-field representation. To demonstrate, we take the lateral probability density of the water oxygens and filter it through a Gaussian kernel density estimate (KDE) of increasing width, σ, following Eq. 3.1.

Lateral inhomogeneities observed in the oxygen density over muscovite persists when the signal is smoothed using a KDE Gaussian smoothing parameter from ∼0.5 Å up to about 2 Å, at which point the signal is laterally isotropic (Fig 3.3). When using a Gaussian width of 4.5 Å for the LMF far-field analysis lateral streaking is not observed, revealing that this striping tendency is a subtle near-field effect that cannot be captured using far-field dielectric theory. The same applies to explicitly-substituted muscovite models, in which lateral inhomogeneities that arise from specific aluminosilicate substitution patterns are washed out by LMF smoothing (see Fig 3.18).

Our present focus is in the direction of the surface normal, the vector along which a macromolecule approaches the surface. We use the macroscopic (or smoothed) polarization density as obtained through the LMF formalism as our order parameter. To access this quantity, we compute the atomistic electric polarization — the dipole moment per unit volume $\mathbf{P}(\mathbf{r}) = n(\mathbf{r})\langle\vec{\mu}(\mathbf{r})\rangle$ for an ensemble averaged water oxygen number density, $n(\mathbf{r})$, and an ensemble averaged water dipole vector as a function of spatial coordinates, $\langle\vec{\mu}(\mathbf{r})\rangle$ [158]. By the divergence theorem, Gauss's law for the polarization density field can be written in differential form, which we apply twice to convert the polarization $P_{\hat{z}}(z)$ into the LMF smoothed bound charge density $\rho_b^{\sigma}(z)$ and then back to the LMF smoothed polarization [159]:

$$\rho_b^{\sigma}(z) = -\frac{1}{4\pi}\int \mathrm{d}z'\partial_{\hat{z}}P_{\hat{z}}(z')\rho_G(z-z') \tag{3.3}$$

$$P_{\hat{z}}^{\sigma}(z) = -4\pi\int \mathrm{d}z\ \rho_b^{\sigma}(z). \tag{3.4}$$

The smoothed bound charge density $\rho_b^{\sigma}(z)$ is computed for these simulations over all space. For proteins, a rectangular region (visualized in Fig 3.9) was isolated as being confined between the protein's binding face and its periodic image along z. Signal from this rectangular subvolume was projected onto the z-axis using an arithmetic mean in x/y to isolate signal arising from the protein.

In comparing intrinsic properties of muscovite under KCl vs. NaCl in the LMF representation, we find qualitative contrast between water polarization that is subtle to detect at the atomistic scale. Strikingly, the LMF signal (far-field electric polarization due to water) changes sign in the two cases, which can be seen in both

$\rho_b^\sigma(z)$ and $P_{\hat{z}}(z)$ for muscovite (Fig 3.4A/B). This finding has ramifications for all surface-induced organizational processes on muscovite, such as heterogeneous ice nucleation, which occurs more efficiently in KCl than NaCl [160, 161]. We seek to determine the impact of this unexpected field reversal on energetics of protein attraction.

In alignment with the Landau description of two surfaces at $z = \pm D/2$, we change the coordinate frame of reference such that opposing faces of a macromolecule are at either end of the unit cell, and bulk solution is at $z = 0$. A visual example is shown in Fig 3.4A, where opposing surfaces at $z = \pm 30$ Å provide boundaries for the bulk. Simulations of proteins and muscovite individually were conducted in the same bulk environment, so charge density profiles can be concatenated together in bulk for a given salt treatment.

3.5 Theory for Polarization-Induced Interactions

Here we construct a theory for the polarization-induced forces felt by two planar surfaces with surface fields of arbitrary sign and intensity. For this we use a Landau theory framework described in detail by Netz and others [150, 151, 162]. The theory is built on a free energy density expansion about a general order parameter η, that describes the response of the solvent to two planar surfaces that are fixed at $z = \pm D/2$. Within Landau theory, the surface groups give rise to a surface field, $h_\pm$ for the surface at z = $\pm D/2$. To be consistent with the literature, we derive the Landau theory for a general order parameter $\eta(z)$, with the understanding that this refers to the far-field polarization for the problem of interest. In this case, we can write down the Landau free energy (per unit area) as

$$\frac{\mathcal{F}}{A} = \int_{-D/2}^{D/2} \mathrm{d}z \Big[a\eta^2(z) + c(\eta'(z))^2 \Big] + h_+\eta(D/2) + h_-\eta(-D/2), \tag{3.5}$$

where $\eta'(z)$ is the derivative of $\eta(z)$ with respect to z, and a and c are phenomenological parameters that give measure of the stiffness of the order parameter and the spatial range of interactions, respectively. The parameter a is a property of the bulk solvent, and when polarization in the relevant order parameter, one can show that $a = (1 - 1/\varepsilon)/2\epsilon_0$, where ε is the static dielectric constant of the solvent and ϵ_0 is the vacuum permittivity, as shown in the Supporting Information.

The functional form of $\eta(z)$ is determined by minimizing the free energy density with respect to the order parameter, giving a family of solutions for different boundary conditions (BCs). Prior work has used symmetric ($h_+ = h_- = h$) or antisymmetric ($h_+ = -h_- = h$) BCs. Here we solve for a more general solution of $\eta(z)$, making no assumptions about the signs or relative magnitudes of $h_\pm$, and use this solution to determine the polarization-induced pressure Π. We work at fixed surface field $h_\pm$ and define $\eta(D/2) = \eta_R$ and $\eta(-D/2) = \eta_L$ at the boundaries. We obtain the following expression for the order parameter profiles,

$$\eta(z) = \eta_R \frac{\sinh\left(\frac{D}{2\lambda} + \frac{z}{\lambda}\right)}{\sinh(D/\lambda)} + \eta_L \frac{\sinh\left(\frac{D}{2\lambda} - \frac{z}{\lambda}\right)}{\sinh(D/\lambda)}, \tag{3.6}$$

where $\lambda = (c/a)^{1/2}$. With $\eta(z)$ at the free energy minimum known, we are able to determine the surface fields h_+ and h_-,

$$h_\pm = -\frac{2c}{\lambda}[\eta_{R/L}\mathrm{coth}(D/\lambda) - \eta_{L/R}\mathrm{csch}(D/\lambda)]. \tag{3.7}$$

Substituting expressions 3.23 and 3.25 into Eq. 3.16 allows $\mathcal{F}/A$ to be evaluated in terms of surface fields $h_\pm$ and physical constants a and c for some separation D:

$$\frac{\mathcal{F}}{A} = \frac{-(h_+h_-)}{2(ac)^{1/2}}\mathrm{coth}\left(\frac{D}{2\lambda}\right)\left[1 + \frac{(h_+ - h_-)^2}{2h_+h_-}\frac{\mathrm{coth}(D/\lambda)}{\mathrm{coth}(D/2\lambda)}\right], \tag{3.8}$$

We can now evaluate the polarization-induced pressure between the surfaces as

$$\begin{aligned}\Pi(D) &= -\frac{d(\mathcal{F}/A)}{dD} \\ &= -\frac{h_+h_-}{4c}\mathrm{csch}^2\left(\frac{D}{2\lambda}\right)\left[1 + \frac{(h_+ - h_-)^2}{h{+}h_-}\frac{\sinh^2(D/2\lambda)}{\sinh^2(D/\lambda)}\right].\end{aligned} \tag{3.9}$$

In this notation a negative pressure is attractive and occurs when h_+ and h_- are of the same sign. For large D, the pressure scales as

$$\Pi \sim -\frac{h_+h_-}{c}e^{-D/\lambda}\left[1 + \frac{(h_+ - h_-)^2}{h_+h_-}e^{-D/\lambda}\right], \tag{3.10}$$

which indicates an exponentially decaying attraction at the largest D values.

When the surfaces are asymmetric, $|h_+| \neq |h_-|$, a second relevant lengthscale of $2D/\lambda$ emerges that can compete with the behavior of the pressure on a scale of D/λ, and this results in a nontrivial change in the sign of Π at

$$D/\lambda = 2\ \mathrm{cosh}^{-1}\left(\sqrt{\frac{-(h_+ - h_-)^2}{4h_+h_-}}\right) \sim \ln\left(\frac{-(h_+ - h_-)^2}{h_+h_-}\right). \tag{3.11}$$

The functional form of this intercept bounds the relative values of $h_\pm$ for a crossing to occur on $D > 0$. For example, Eq 3.11 is undefined for either symmetric or antisymmetric $h_\pm$, and so this was absent in previous work. An intercept is present for all $h_+h_- < 0$ if $h_+ \neq -h_-$, and occurs at greater D/λ for ratios of $h_\pm$ far from unity. In the present work, where λ is on the order of 2 Å, each instance that matched these criteria produced intercepts at $D/\lambda < 5$ Å and so may not be relevant to the far-field descriptors provided here. A detailed derivation of this theory and further discussion is provided in the Supplementary Information.

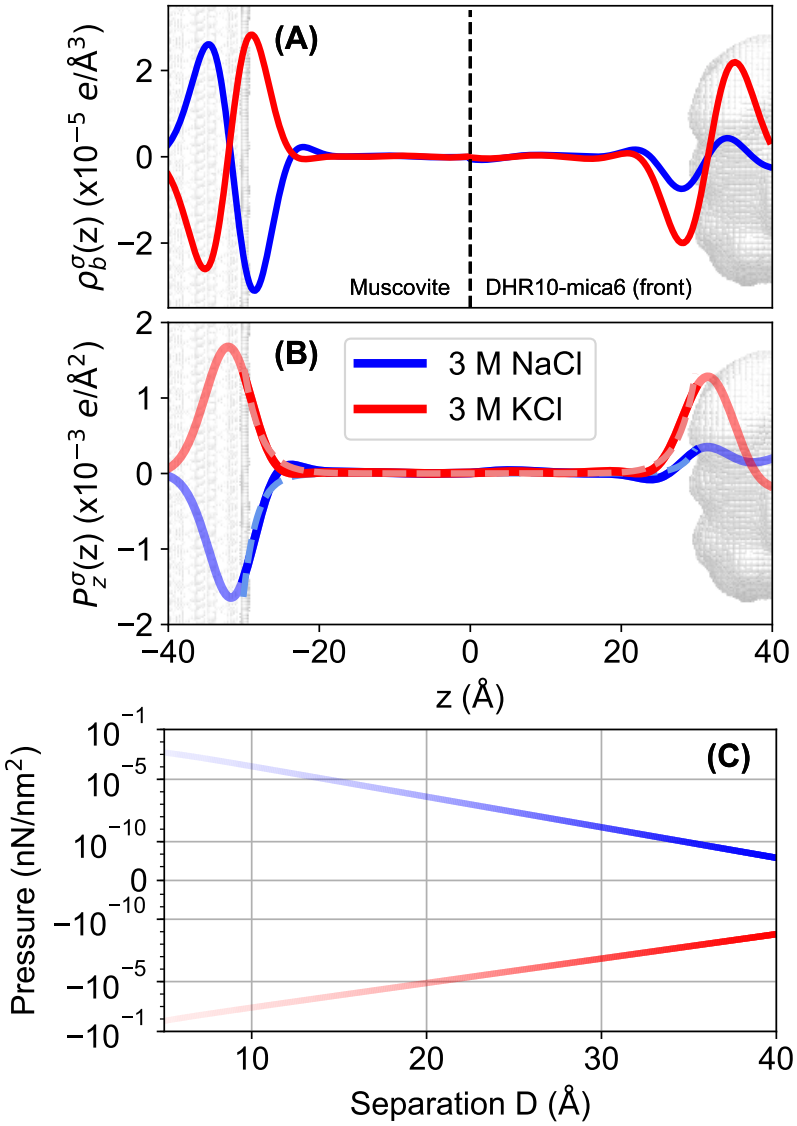

Figure 3.4: Demonstration of converting $\rho_b^\sigma(z)$ obtained from LMF-processed MD trajectories into polarization-induced pressure between muscovite and the front/back faces of DHR10-mica6. (A) A concatenated bound charge density profile is made by adjoining partial $\rho_b^\sigma(z)$ profiles from trajectories of protein or muscovite in isolation at the bulk, where solution conditions are identical. (B) We compute $P_z^\sigma(z)$ from $\rho_b^\sigma(z)$ via Eqs. 3.3 as in the main text, and then fit the functional form of order parameter $\eta(z)$ (Eq 3.6) to polarization on the region [-30 Å, +30 Å] that is solvent-accessible (fits are dashed lines). (C) Fitting $\eta(z)$ to $P_z^\sigma(z)$ determines the phenomenologica33l parameters needed to evaluate pressure Π (Eq 3.9) as a function of separation D, where a positive Π indicates repulsion. In all three panels, red coloration indicates 3 M KCl solution, while blue indicates 3 M NaCl.

There are two major results from this derivation: (1) By solving an analytical $\eta(z)$ (Eq 3.23) for asymmetric BCs we can fit the functional form of $\eta(z)$ to $P_{\hat{z}}(z)$ computed from simulation, and (2) The parameters determined from fitting to data (surface fields and λ) can be inserted into $\Pi(D)$ (Eq. 3.27) which provides an estimate of distance-dependent behavior.

It is apparent that, on the domain of z that is solvent accessible, the function $P_z^\sigma(z)$ has a sinh-like form that can be well-approximated by the analytical function $\eta(z)$. The fitted function $\eta(z)$ is overlaid in Fig 3.4B to illustrate. We take the parameters η_R and η_L from $P_z^\sigma(\pm D/2)$ and fit λ, which here ranges from 1.66 to 2.21 Å, larger than similar results for membranes [151]. The extracted parameter λ permits the pressure $\Pi(D)$ to be estimated for a given polarization profile.

For discrete simulations of "mean-field" muscovite or protein isolated in 3 M aqueous KCl or NaCl, we computed $P_{\hat{z}}^\sigma(z)$ for each surface. At large z these polarization densities are the same (for a given ion treatment) and so a composite $P_{\hat{z}}^\sigma(z)$ is generated by concatenating intrinsic polarization in the bulk region. The "composite" systems consist of the muscovite surface parallel to a flat protein face, either the notional binding face or the reverse side. Comparing results under KCl and NaCl conditions reveals qualitatively different far-field muscovite-protein interactions.

Figure 3.5 shows selected $\Pi(D)$ profiles for composite muscovite-protein systems. The notional binding face of ^{C98}RhuA is attracted to muscovite under both KCl and NaCl conditions, as indicated by the negative $\Pi(D)$. In contrast, the notional binding face of DHR10-mica6 is strongly attracted to muscovite under KCl conditions but is repelled in NaCl solution. The reverse is true of the backside of DHR10-mica6, which is not attracted to muscovite in the presence of KCl but is with NaCl. The qualities of these pressure profiles are determined by the relative signs and magnitudes of the surface fields generated by each body intrinsically, which can be explained with atomistic evidence from simulation. These results are consistent with experimental findings as detailed in the Discussion section.

3.6 Consistency with Atomistic Interpretations

One of the principal findings reported here is the change in sign of the far-field components of electric polarization due to water when K^+ is exchanged for Na^+ over muscovite. Considering that both systems consist of an ionic double layer with cations bound to mica, it is perhaps surprising that such a polarization reversal should occur. This observation can be reconciled with atomistic data: There are more surface vacancies for water molecules under NaCl conditions (as seen in Fig 3.10, close to the surface). Water molecules occupying cation binding pockets have a strong tendency to orient with the positive end of the dipole toward mica (Fig 3.11). Conversely, water molecules slightly further from the surface tend to interact strongly with the bound cations, and so orient their negative end toward the mica. By virtue of having more ion binding vacancies, the Na^+ condition sees a shift in relative intensity between these two populations of water molecules.

Furthermore, the presence of a secondary density peak in atomistic Na^+ position data reveals that the ionic double layer is less well-organized than with K^+ and fails to attract counterions as intensely (Fig 3.10). This gives rise to a broad orientational distribution of water at the interface, particularly ∼0.5 nm from the surface, and peaks in the water oxygen density function correspond to local water orientations

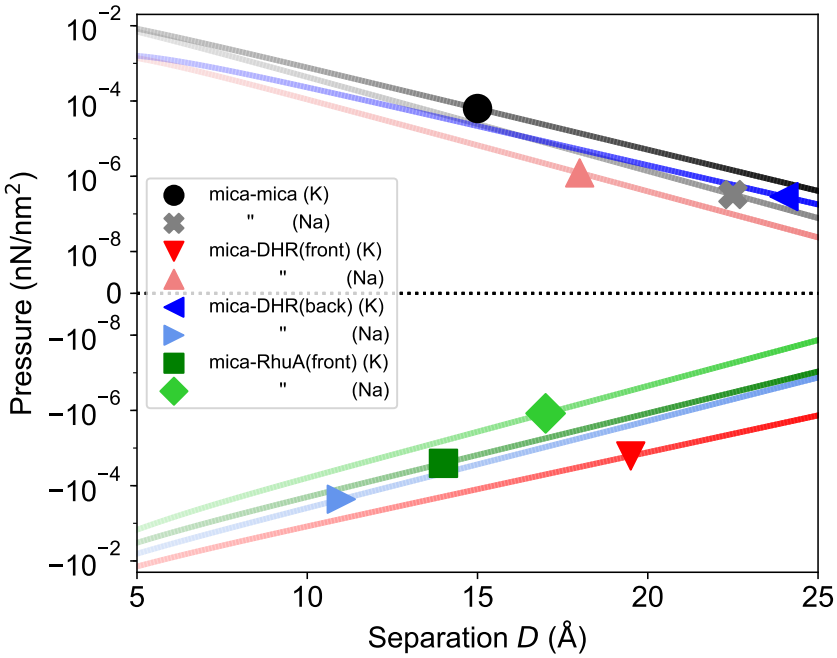

Figure 3.5: Selected far-field polarization-induced pressure curves demonstrating impacts of K^+/Na^+ exchange. The eight traces presented are to be thought of as four pairs, each representing a given muscovite-macromolecule pair under both 3 M KCl and 3 M NaCl conditions. These macromolecule pairings are: muscovite-muscovite (as reference, in black), muscovite-$DHR_{binding}$ (red), muscovite-DHR_{back} (blue), and muscovote-$^{C98}RhuA_{binding}$ (green). Pairs of traces are indicated by a shared color (dark vs. light shade for KCl vs. NaCl, respectively) and corresponding symbols: (●, ✖), (▼, ▲), (◄, ►), (■, ◆), listed in KCl/NaCl order. For example, the red pair of traces (▼, ▲) correspond to those in Fig 3.4C. With $\Pi > 0$ indicating repulsion we see muscovite-muscovite is repulsive under both KCl and NaCl, muscovite-$^{C98}RhuA_{binding}$ is attractive in both cases, and the muscovite-DHR interaction can be modulated to attractive or repulsive depending on which protein face is presented and the salt treatment. See Fig 3.7 for renderings of the two proteins to illustrate the notional "front" and "back" faces.

opposite that of K^+ (Fig 3.11). This is all to say that the LMF far-field polarity reversal on muscovite is consistent with the underlying atomistic structural changes.

The other intrinsic property of muscovite reported here, lateral anisotropy, does not contribute to far-field interactions as evidenced in Fig 3.3. The height(s) above muscovite at which lateral anisotropy can be detected are narrow, indicating that these streaking effects may only be felt by macromolecules when adsorbed. The finding of this lateral anisotropy does not contradict a previous AFM study of the azimuthal dependence interaction between two disjoining mica surfaces [163]. In Ref. [163] the forces between the two mica surfaces was measured at a predicted separation of roughly 1 nm. At this separation a weak azimuthal dependence of the force between the mica surfaces is observed reflecting the underlying basal symmetry of mica. However, no lateral anistropy was recorded that is consistent with Fig 3.3. The predicted separation of 1 nm between the two surfaces in Ref. [163] is likely too great a distance to sense the short-range lateral anistropy. Moreover, it was suggested that the weak azimuthal dependence of the force between mica surfaces at separations of ∼1 nm cannot be predicted with electrostatics alone [164, 163]. To the extent that the protein-mica surfaces remain close enough to sense the lateral anistropy in Fig 3.3 is possible due to the polarization induced forces predicted in Fig 3.5. This is distinct from Ref. [163] where two mica surfaces were found to feel a repulsive long-range interaction (see Fig 3.15) and therefore remain further apart. Additional study is needed to assess the interplay of lateral streaking and polarization along the normal, which here are presented as distinct observations. The ramifications of these striking features are currently being analyzed through the lens of close-range protein/surface interactions in a separate body of work.

Differences in LMF signal near the faces of proteins can be corroborated atomistically as well. For example, when K^+ is exchanged for Na^+ the binding face of DHR10-mica6 experiences a ∼2- to 3-fold deamplification in water polarization density $P_z^\sigma(z)$ (Fig 3.4A). This signal aligns with atomistic density distributions, which reveal a competition between cations and water molecules at the protein surface. While sodium ions show a marginally stronger affinity for the protein binding face than potassium, this is balanced with a depletion in the water density profile, and waters further from the surface showing a diminished angular preference. Density profiles on either side of DHR10-mica6 are included in Fig 3.12 and the consequent water orientation distributions are shown in Fig 3.13. Thus, atomistic simulations of a lone macromolecule in solution can provide sufficient information to predict surface adsorption behavior by evaluating the compatibility of its far-field response (due to atomic-scale structuring near the molecule) with those arising from the solid-liquid interface. Such relationships are challenging to disentangle without LMF analysis.

3.7 Discussion

Analyzing MD data through the lens of LMF theory has allowed us to connect intrinsic polarization to assembly processes via induced pressure from Landau theory. The unexpected activity of the DHR10-mica6 "back face" in NaCl results in an attractive $\Pi(D)$ as the protein approaches the surface with this orientation, in contrast to the binding face which interacts very weakly with muscovite in NaCl (Fig 3.5). The low-intensity nature of far-field interactions mean that these directional electrostatic biases on the protein are perturbations to its orientational distribution with respect

to muscovite. These predictions are valid at large D, as eventually short-range interactions become non-negligible and intrinsic far-field properties might interfere nonlinearly. These bound-state interactions will be characterized explicitly in a separate study.

The observation that the DHR10-micaX back face becomes well-ordered in NaCl conditions leads to the hypothesis that DHR10-micaX proteins may have substantive populations bound to muscovite through both of the protein faces in NaCl. This contrasts KCl conditions in which the binding face designed to match the muscovite potassium ion sublattice is expected to dominate. The back face of DHR10-micaX is not designed to match the muscovite lattice, which reduces the tendency for the protein to align along any of its pseudohexagonal axes. Indeed, experimental observations suggest diminished orientational specificity of protein binding in NaCl solution on average. AFM results show that the protein DHR10-mica18 displays weaker orientational specificity in aqueous NaCl than KCl (Fig 3.1). Furthermore, the most intense protein-muscovite interaction as predicted by $\Pi(D)$ is with the binding face of DHR10-micaX under KCl conditions, and our predictions are consistent with the experimentally-observed increase in critical concentration required to reach a 2D liquid crystal in NaCl solutions.

The other muscovite-binding protein ^{C98}RhuA has a relatively flat binding face and contacts muscovite under both KCl and NaCl conditions, as demonstrated by AFM experiments [142, 165, 166]. Our theory predicts that ^{C98}RhuA displays a weak attraction to muscovite from afar via $\Pi(D)$ in both cases (Fig 3.5). The lack of spatial ordering on the face of ^{C98}RhuA compared to muscovite, combined with the highly-screened environment studied here, results in weak polarization, $P_z^\sigma(z)$. This polarization in response to ^{C98}RhuA is dependent on the size of the subvolume used for analysis (see Fig 3.14), unlike DHR10-mica6. This is due to patterning of amino acids on and the curvature of the binding face of ^{C98}RhuA. These two effects wash out the signal from highly-charged residues at the center and our predictions for ^{C98}RhuA can be improved through three-dimensional order parameters. The large asymmetry between surface fields at muscovite and ^{C98}RhuA surfaces results in pressure profiles with a nontrivial zero-crossing at finite D that may be relevant for assembly. This non-monotonic form of $\Pi(D)$ occurs only for asymmetric surfaces; see the SI for further discussion on competing length scales that emerge when the polarization profile is asymmetric.

The magnitudes of the polarization-induced pressures and characteristic length scales reported here are similar to those reported in other computational, theoretical, and experimental studies [167, 168, 169, 170, 171, 172]. Measurements of pressure between two separated muscovite slabs vary depending on methodology and conditions, with estimates that are higher or lower than what is reported here at $\sim$2 nm separation. In particular the result depends sensitively on double layer's screening of the surface, which seems to be underestimated in empirical simulation models, as compared to experiment. Our predictions are consistent with pressure estimates available in literature, and a discussion on comparisons among pressure estimates is given in the SI subsection "Consistency with Simulation, Theory, and Experiment".

3.8 Conclusion

In this work we have shown how the intrinsic solvent-structuring properties of minerals and proteins with flat surfaces can be connected via a theoretical framework to categorize far-field interactions as attractive or repulsive for a given protein orientation. Relative interaction strengths in this picture correlate well with experimental observations of surface assembly. We construct an empirical all-atom muscovite model with improved cation occupancy (as a percentage of available sites) to compute solvent polarization in response to specific ion effects. The combined simulation/theoretic framework presented is broadly applicable to surface assembly problems and to predict a macromolecule's orientation upon approach.

The philosophy of this work takes water molecules as reporters for the underlying electric fields generated by ion organization in the vicinity of a liquid-solid interface. That is, by studying electric polarization rather than the entire electric field, we avoid boundary condition issues that arise from Poisson-Boltzmann analyses and omit the electric displacement field associated with ρ_{unbound}. The Landau theory approach taken here is advantageous because inputs are computed directly from simulation. While the present work demonstrates the efficacy of this strategy in high-concentration alkali salts in order to emulate experiments, we emphasize that this computational strategy is generally applicable to surfaces without well-ordered ion double layers at more typical ionic strengths.

One of the key findings from these simulations is that exchanging the identity of the aqueous salt from KCl to NaCl over muscovite results in a sign reversal in far-field electrostatic measurements ρ_b^σ and P_z^σ due to water molecules. This sign reversal can be understood as arising from the multi-binding-well character of Na^+ on muscovite, an observation that has been corroborated in XRR measurements [173]. Such a sign reversal is likely not exclusive to muscovite, and may occur on related clay surfaces (*i.e.*, montmorillonite, kaolinite, etc.) due to differences in surface organization for each ion.

This finding leads us to consider other polarization changes that occur on mineral-binding proteins that would impact the ability of the binding face to specifically orient toward the surface when approaching from solution. Qualitative changes to the long-range component of the electrostatic potential due to water is observed at some protein faces but not others, highlighting the importance of computing intrinsic properties from atomistic simulations. Furthermore, while atomistic measurements corroborate the LMF picture, this processing is essential to disambiguate the far-field character.

Ion-specific differences in surface polarization are potentially exploitable properties in engineering self-assembling materials: Taken in conjunction with high-intensity close-range interactions that do not flip sign with exchanging salt, it may be plausible to dynamically control heteroepitaxial growth by modulating salt conditions in order to amplify or deter far-field attraction while not disrupting the bound layer.

While the present work focuses on protein-surface interactions on nanometer length scales by extracting the far-field, macroscopic polarization, it is in principle possible to extend this framework to capture short-range oscillatory behavior in Landau theory as well. This combined simulation-theoretic approach will be valuable in coming to a predictive understanding of the conditions in which proteins assemble

on surfaces. The simulations required to extract intrinsic surface properties are simple to conduct and inexpensive, providing a potential high-throughput route to assessing the driving forces of assembly.

3.9 Materials and Methods

The geochemistry of muscovite is well-studied for the bulk material and reliable all-atom MD force fields have been published, particularly INTERFACE (IFF) [152] and CLAYFF (CFF) [154]. However, there are a number of practical issues with capturing the structural characteristics of the liquid-muscovite interface using either model: 1) the chemical potential of ion adsorption does not agree with experimental observation at low concentration and surface ion occupancy is underrepresented; 2) ion adsorption at the interfacial double layer depletes the aqueous environment due to finite size effects; and 3) lateral organization of water and ions is sensitive to aluminum substitution patterns within muscovite, which may be assigned in a variety of ways.

These are addressed as follows:

1. The fraction of surface cation binding sites that are occupied vs. vacant is underrepresented in MD simulation compared to experiment (both CFF and IFF), which is combated by utilizing a high chemical potential aqueous bath: Here we use 3 M monovalent salt, emulating experimental conditions in which DHR10-mica6 and ^{C98}RhuA bind muscovite. The high salinity simulated here is in emulation of experimental surface coverage near 100% occupancy, an "overcharged" surface with a strong ion double layer [135]. High occupancy is maintained down to 1 M salt, at which point competition from hydronium ions becomes an important factor, a situation that cannot be accurately captured in classical simulation.

2. Freshly cleaved, electrically neutral muscovite (as is prepared in default IFF or CFF repositories) only includes 50% of available cation sites occupied. When exposed to solution, vacancies are populated with ions from the bath. To prevent the finitely sized aqueous bath in MD from being depleted of ions due to surface sequestration, while also avoiding an enormous simulation box, the aqueous phase is prepared in MD with the anticipation that some number of ions will adsorb to the surface. That is, a simulation of aqueous muscovite includes sufficiently many ions to both 100% saturate the muscovite surface and maintain the desired aqueous concentration.

3. The lateral positioning of surface-exposed Al substitution sites (prevalent at a 1:3 stoichiometric ratio with Si) is adjustable and measured lateral inhomogeneities correlate with the substitution pattern chosen. To prevent measurement artifacts associated with specific Al distributions and to match ion distributions that mica-binding proteins are designed against, a novel mean-field approach is presented. Where explicit substitutions are used for comparison, the NMR-derived "random" pattern from IFF is preferred.

In this work we combine Si and Al into a statistically representative pseudo-atom, which removes lateral Al organization from consideration when conceptualizing far-field interactions. Briefly, this "mean-field" pseudo-atom approach was accomplished

by taking 1:3 weighted averages of the parameters for Al and Si. Conveniently, some parameters such as the Lennard-Jones well depth ϵ are identical for Al and Si both in CFF and IFF, and so the statistical average is trivial. In both CFF and IFF the radius σ was taken to be a weighted arithmetic mean of Al and Si, following the Lorentz-Berthelot combination rule [174]. Mass and charge were averaged similarly, for example yielding mass: $(3{*}m_{\mathrm{Si}} + 1{*}m_{\mathrm{Al}})/4 = 27.25764$ amu. The CFF topology does not include explicit bonds linking inorganic bodies, and so for CFF no further adjustments are required. Within IFF, such bond and angle definitions do exist, but parameters are identical for Si or Al in the same environment. Finally, oxygen atoms adjacent to substitution sites have their partial charges adjusted as well, in the same mean-field philosophy of lateral homogeneity.

Slabs of mica were prepared three double-layers thick ($\approx$ 6 nm total) in order to thoroughly insulate periodic images from one another. To further ensure that the muscovite slab did not bend due to thermal fluctuations, 1000 kJ/mol harmonic restraints were placed on a single layer of potassium ions embedded nearest the surface's center. This way, the surface as a whole retains its planar geometry while leaving surface layers flexible to respond to solvent. Models were constructed from force field to a 5x3 unit cell lattice ($\sim$ 2.6x2.7 nm^2) or a 20x12 unit cell lattice ($\sim$10.4x10.8 nm^2). The larger simulation cell is beneficial for identifying lateral inhomogeneities that persist despite usage of the mean-field Al substitution treatment, which would be reflective of the asymmetric hexagonal surface structure of muscovite.

Solvent-exposed alkali metal parameters were taken from the work of Joung and Cheatham (JC) [175], which have been specifically parameterized to prevent crystallization at high concentration (up to 3 M in this work). Protein simulations were conducted with the well-established CHARMM36 force field [18]. The CHARMM family of force fields was developed for use with the CHARMM-modified TIP3P water model (mTIP3P), which differs from the original TIP3P by the addition of Lennard-Jones sites to the hydrogen atoms. To apply the mTIP3P water model to CHARMM-based proteins while retaining the original TIP3P behavior with JC ions, NBFIX was used to set Lennard-Jones interactions to zero only for water hydrogen-ion pairs [176]. Doing so permits JC ions to interact with original TIP3P water while permitting the protein to interact with mTIP3P.

Proteins were restrained at four alpha carbons using harmonic potentials within GROMACS (1000 kJ/mol spring constants) to keep a fixed frame of reference. The restrained sites were selected to be at opposite corners of the protein to prevent tumbling motions, while permitting side chain and protein loop fluctuations to thoroughly sample interactions with the aqueous environment. For DHR10-mica6 these were alpha carbons 8A, 150A, 180A, and 258A, while for $^{\mathrm{C98}}$RhuA these were 79F on each tetrameric subunit.

Simulations were conducted using GROMACS 2018.3 [177, 178]. All simulations described utilized Particle Mesh Ewald electrostatics with 0.08 nm Fourier spacing, neighbor list and electrostatics cutoffs at 1.0 nm, and were conducted at 300 K. The Berendsen barostat was employed for NPT equilibration over 25 ns with bulk compressibility 0.000045 normal to the surface (z) and semiisotropic coupling preventing deviations in x/y. The full equilibration period is required to ensure that ions initialized in the bulk phase have sufficient time to adsorb onto the surface prior to 100 ns of production NVT. For the proteins, only 15 ns NPT equilibration is required

since these proteins do not sequester ions the same way muscovite does. However, a total of 150 ns of NVT simulation were conducted due the lower symmetry (and therefore reduced data quality) of proteins compared to muscovite. Error bars were computed by the method of bootstrapping, subdividing the simulated trajectory into ten non-overlapping subsections and computing standard deviations from the repeated results.

For slabs of muscovite only 5x3 wide, the system was sufficiently small to perform Parallel-Bias Well-Tempered Metadynamics with Partitioned Families (PBMetaD-PF) [179]. Briefly, this technique includes a time-dependent energetic bias on the z-coordinates of all K^+ and Cl^- ions using the redundancy of their coordinates to converge the bias potential in little time. Biasing ion z-coordinates in this way expedites the exploration of phase space and can reveal hidden energetic basins. PBMetaD-PF used Gaussian width sigma 0.01, with bias factor 5, initial hill height 1.25 kJ/mol, and pace 500 steps, at a target temperature of 300 K in accordance with previous ion-binding PMF studies [180]. Production runs were on the time scale of 200 ns in the NVT ensemble. Due to the fact that ∼150 ions of each kind contribute to each PMF, many may be surface-bound at once, and typical cation residence times on the order of nanoseconds, convergence was quickly achieved (Fig 3.15). We concluded that further PBMetaD-PF would not be required.

3.10 Derivation: Landau Theory of Induced Forces Between Asymmetrically Polarized Surfaces

3.10.1 The Landau Free Energy

We consider two planar surfaces, one fixed at $z = D/2$ and the other at $z = -D/2$. Each surface polarizes the solvent (water) in its vicinity, generally due to polar or charged surface functional groups. Within the Landau theory we presently derive, the surface groups give rise to a surface field, $h_\pm$ for the surface at $z = \pm D/2$. To be consistent with the literature [151, 150, 162], we will derive the Landau theory for a general order parameter $\eta(z)$, with the understanding that this refers to the polarization for the problem of interest. In this case, we can write down the Landau free energy (per unit area) as

$$\frac{\mathcal{F}}{A} = \int_{-D/2}^{D/2} \mathrm{d}z \Big[a\eta^2(z) + c(\eta'(z))^2 \Big] + h_+\eta(D/2) + h_-\eta(-D/2), \tag{3.12}$$

where a and c are expansion coefficients thought of as phenomenological parameters pertaining to bulk properties and decay rate of solvent ordering, respectively. In this notation $\eta'(z)$ is the derivative of $\eta(z)$ with respect to z.

3.10.2 Minimizing the Energy

We want to minimize the free energy to determine the order parameter $\eta(z)$ and surface fields $h_\pm$. To do this, we take the functional derivative of the free energy with respect to the order parameter, which will eventually be set to zero. This derivative is written

$$\frac{\delta(\mathcal{F}/A)}{\delta\eta(z')} = \int_{-D/2}^{D/2} dz\,[2a\eta(z)\delta(z-z') + 2c\eta'(z)\delta(z-z')] + h_+\delta\left(\frac{D}{2}-z'\right) + h_-\delta\left(\frac{D}{2}+z'\right) \tag{3.13}$$

$$= \int_{-D/2}^{D/2} dz\,[2a\eta(z)\delta(z-z') - 2c\eta''(z)\delta(z-z')] + 2c\eta'(z)\delta(z-z')\Big|_{-D/2}^{D/2} + h_+\delta\left(\frac{D}{2}-z'\right) + h_-\delta\left(\frac{D}{2}+z'\right) \tag{3.14}$$

$$= 2a\eta(z') - 2c\eta''(z') + \delta\left(\frac{D}{2}-z'\right)[h_+ + 2c\eta'(D/2)] + \delta\left(\frac{D}{2}+z'\right)[h_- - 2c\eta'(-D/2)]\,, \tag{3.15}$$

such that

$$\frac{\delta(\mathcal{F}/A)}{\delta\eta(z)} = 2a\eta(z) - 2c\eta''(z) + \delta\left(\frac{D}{2}-z\right)[h_+ + 2c\eta'(D/2)] + \delta\left(\frac{D}{2}+z\right)[h_- - 2c\eta'(-D/2)]\,. \tag{3.16}$$

At this point, we can set the bulk and surface terms to zero independently of one another,

$$\eta(z) - \left(\frac{c}{a}\right)\eta''(z) = 0, \tag{3.17}$$

$$h_+ + 2c\eta'(D/2) = 0, \tag{3.18}$$

$$h_- - 2c\eta'(-D/2) = 0. \tag{3.19}$$

We can evaluate the free energy at the minimum for any $\eta(z)$, without making any further specifications about the system.

$$\mathcal{F}/A = \int_{-D/2}^{D/2} dz\left[a\eta^2(z) - c\eta''(z)\eta(z)\right] + c\eta'(z)\eta(z)\Big|_{-D/2}^{D/2} + h_+\eta(D/2) + h_-\eta(-D/2) \tag{3.20}$$

$$= \int_{-D/2}^{D/2} dz\eta(z)\,[a\eta(z) - c\eta''(z)] + c\eta'(z)\eta(z)\Big|_{-D/2}^{D/2} + h_+\eta(D/2) + h_-\eta(-D/2) \tag{3.21}$$

$$= c\eta'(z)\eta(z)\Big|_{-D/2}^{D/2} + h_+\eta(D/2) + h_-\eta(-D/2), \tag{3.22}$$

where the term in brackets of the integrand is zero at the minimum via Eq. 3.17. We can now proceed by using the surface terms in Eqs. 3.18 and 3.19 to note that

$h_+ = -2c\eta'(D/2)$ and $h_- = 2c\eta'(-D/2)$. Now, the free energy is given by

$$\mathcal{F}/A = c\eta'(z)\eta(z)\Big|_{-D/2}^{D/2} - 2c\eta'(D/2)\eta(D/2) + 2c\eta'(-D/2)\eta(-D/2), \tag{3.23}$$

which is equal to

$$\mathcal{F}/A = -c\eta'(z)\eta(z)\Big|_{-D/2}^{D/2} \tag{3.24}$$

which is the general expression for the free energy at the minimum.

3.10.3 Order Parameter Profiles

We can determine $\eta(z)$ at the free energy minimum by solving Eq. 3.17, which has the general solution

$$\eta(z) = k_1 e^{z/\lambda} + k_2 e^{-z/\lambda}, \tag{3.25}$$

where $\lambda = \sqrt{c/a}$. To determine the coefficients k_1 and k_2, we specify the boundary conditions (BCs) for asymmetric order parameter profiles as,

$$\eta(D/2) = \eta_R \tag{3.26}$$

and

$$\eta(-D/2) = \eta_L. \tag{3.27}$$

This leads to two equations, $\eta_R = k_1 \exp(D/2\lambda) + k_2 \exp(-D/2\lambda)$ and

$\eta_L = k_1 \exp(-D/2\lambda) + k_2 \exp(D/2\lambda)$, which can be solved for the two coefficients. These coefficients are

$$k_1 = \frac{1}{2}\sinh^{-1}(D/\lambda)\left(\eta_R e^{D/2\lambda} - \eta_L e^{-D/2\lambda}\right), \tag{3.28}$$

and

$$k_2 = \frac{1}{2}\sinh^{-1}(D/\lambda)\left(\eta_L e^{D/2\lambda} - \eta_R e^{-D/2\lambda}\right). \tag{3.29}$$

We can plug these coefficients into Eq. 3.25, and, after some algebra, we obtain the following expression for the order parameter profiles,

$$\eta(z) = \eta_R \frac{\sinh\left(\frac{D}{2\lambda} + \frac{z}{\lambda}\right)}{\sinh(D/\lambda)} + \eta_L \frac{\sinh\left(\frac{D}{2\lambda} - \frac{z}{\lambda}\right)}{\sinh(D/\lambda)}. \tag{3.30}$$

3.10.4 Surface Fields

The surface fields h_+ and h_- can now be determined via Eqs. 3.18, 3.19, and 3.30. For this, we need the spatial derivative of the order parameter profile, which is

$$\eta'(z) = \frac{\eta_R}{\lambda}\frac{\cosh\left(\frac{D}{2\lambda}+\frac{z}{\lambda}\right)}{\sinh(D/\lambda)} - \frac{\eta_L}{\lambda}\frac{\cosh\left(\frac{D}{2\lambda}-\frac{z}{\lambda}\right)}{\sinh(D/\lambda)}. \tag{3.31}$$

We can now readily obtain the surface fields by evaluating $\eta'(z)$ at the boundaries, $z = \pm D/2$, to obtain

$$h_+ = -\frac{2c}{\lambda}\left[\eta_R \text{coth}(D/\lambda) - \eta_L \text{csch}(D/\lambda)\right] \tag{3.32}$$

and

$$h_- = -\frac{2c}{\lambda}\left[\eta_L \text{coth}(D/\lambda) - \eta_R \text{csch}(D/\lambda)\right]. \tag{3.33}$$

3.10.5 Working at Fixed Surface Field

We now have what we need to compute the free energy and the pressure between the two surfaces. However, we want to make sure that we are working at *fixed surface field* and not fixed order parameter, because fixed surface field corresponds to the experimental situation (the surface chemistry is not changing with D). In order to work at fixed surface field, we need to determine η_R and η_L in terms of h_+ and h_-, and we need to write $\eta(z)$ as a function of the surface fields. To do this, we can write $h_+ = A\eta_R + B\eta_L$ and $h_- = A\eta_L + B\eta_R$, where A and B can be determined by looking at Eqs. 3.32 and 3.33. We can solve this system of equations to get $\eta_L = (Ah_- - Bh_+)/(A^2 - B^2)$ and $\eta_R = (Ah_+ - Bh_-)/(A^2 - B^2)$. Using the values of A and B, one can show that $A^2 - B^2 = 4c^2/\lambda^2$. Then, we can write the order parameters at the boundaries as functions of the surface fields,

$$\eta_L = -\frac{\lambda}{2c}\text{csch}(D/\lambda)\left[h_- \cosh(D/\lambda) + h_+\right] \tag{3.34}$$

and

$$\eta_R = -\frac{\lambda}{2c}\text{csch}(D/\lambda)\left[h_+ \cosh(D/\lambda) + h_-\right]. \tag{3.35}$$

With these expressions for the order parameter at the boundaries, the order parameter profile is given by

$$\begin{aligned}\eta(z) = &-\frac{\lambda}{2c}\text{csch}^2(D/\lambda)\left[h_+ \cosh(D/\lambda) + h_-\right]\sinh(D/2\lambda + z/\lambda) \\ &-\frac{\lambda}{2c}\text{csch}^2(D/\lambda)\left[h_- \cosh(D/\lambda) + h_+\right]\sinh(D/2\lambda - z/\lambda).\end{aligned} \tag{3.36}$$

The spatial derivative of the order parameter can also be evaluated as

$$\eta'(z) = -\frac{1}{2c}\text{csch}^2(D/\lambda)\left[h_+\cosh(D/\lambda) + h_-\right]\cosh(D/2\lambda + z/\lambda) + \frac{1}{2c}\text{csch}^2(D/\lambda)\left[h_-\cosh(D/\lambda) + h_+\right]\cosh(D/2\lambda - z/\lambda). \tag{3.37}$$

While these are somewhat ugly expressions, we will see that they simplify when evaluated at the boundaries, as needed for the free energy.

3.10.6 Free Energy for Asymmetric Profiles

In this section, we finally evaluate the free energy, $\mathcal{F}/A$. To do so, we use Eq. 3.24,

$$\mathcal{F}/A = -c\eta'(D/2)\eta(D/2) + c\eta'(-D/2)\eta(D/2). \tag{3.38}$$

We proceed by evaluating each term separately, then plugging them into this expression for the free energy (rather than evaluating the product $\eta(z)\eta'(z)$ directly). We find

$$\eta(D/2) = -\frac{\lambda}{2c}\text{csch}(D/\lambda)\left[h_+\cosh(D/\lambda) + h_-\right] \tag{3.39}$$

$$\eta(-D/2) = -\frac{\lambda}{2c}\text{csch}(D/\lambda)\left[h_-\cosh(D/\lambda) + h_+\right] \tag{3.40}$$

$$\eta'(D/2) = -\frac{1}{2c}\left[h_+\text{coth}^2(D/\lambda) - h_+\text{csch}^2(D/\lambda)\right] \tag{3.41}$$

$$\eta'(-D/2) = -\frac{1}{2c}\left[h_-\text{csch}^2(D/\lambda) - h_-\text{coth}^2(D/\lambda)\right] \tag{3.42}$$

which we use to evaluate the two boundary terms in the free energy. After a lot of algebra, these two terms are

$$-c\eta(D/2)\eta'(D/2) = -\frac{\lambda}{4c}\text{csch}(D/\lambda)\left[h_+^2\cosh(D/\lambda) + h_+h_-\right] \tag{3.43}$$

and

$$c\eta(-D/2)\eta'(-D/2) = -\frac{\lambda}{4c}\text{csch}(D/\lambda)\left[h_-^2\cosh(D/\lambda) + h_+h_-\right]. \tag{3.44}$$

Putting everything together, and rearranging, we can write the free energy as

$$\mathcal{F}/A = -\frac{\lambda}{4c}\left[(h_+^2 + h_-^2)\text{coth}(D/\lambda) + 2h_+h_-\text{csch}(D/\lambda)\right]. \tag{3.45}$$

3.10.7 Comparison with Anti-/Symmetric Profiles

All of the results above are completely general. We did not make any assumptions about the sign of $\eta_{R/L}$ or $h_{+/-}$. In order to compare with previous derivations for symmetric or antisymmetric order parameter profiles, we can choose to re-write Eq. 3.45 in equivalent forms that are more comparable to previous results by Netz and coworkers [151]:

We can rearrange this expression for the free energy by noting that $(h_+^2 + h_-^2) =$

$(h_+ - h_-)^2 + 2h_+h_-$, and then using the trigonometric identity $\sinh(x)/[\cosh(x)+1] = \tanh(x/2)$. This gives a form that resembles that of a symmetric order parameter profile,

$$\frac{\mathcal{F}}{A} = -\frac{h_+h_-}{2(ac)^{1/2}}\coth\left(\frac{D}{2\lambda}\right)\left[1 + \frac{(h_+ - h_-)^2}{2h_+h_-}\frac{\coth(D/\lambda)}{\coth(D/2\lambda)}\right]. \tag{3.46}$$

In a completely equivalent way, we can alternatively rearrange Eq. 3.45 into a form that resembles the antisymmetric order parameter profile by making the substitution $(h_+^2 + h_-^2) = (h_+ + h_-)^2 - 2h_+h_-$, followed by the trigonometric relation $\coth(D/\lambda) - \mathrm{csch}(D/\lambda) = \tanh(D/2\lambda)$. Doing so results in the following free energy (completely equivalent to Eq. 3.46):

$$\frac{\mathcal{F}}{A} = \frac{h_+h_-}{2(ac)^{1/2}}\tanh\left(\frac{D}{2\lambda}\right)\left[1 - \frac{(h_+ + h_-)^2}{2h_+h_-}\frac{\tanh(D/2\lambda)}{\tanh(D/\lambda)}\right]. \tag{3.47}$$

Re-writing the free energy in the form of Eq. 3.46 reduces to the Netz *et al.* result when the polarization profile is symmetric, i.e. $h_+ = h_- = h$ [151]. Correspondingly, the free energy from Eq. 3.47 reduces to the antisymmetric result when $h_+ = -h_- = h$. Eqs. 3.45-3.47 are all equivalent and are re-written to highlight competing terms in different parameter limits.

3.10.8 Polarization-Induced Pressure Between Surfaces

We can now readily evaluate the pressure between surfaces as

$$\Pi = -\frac{d(\mathcal{F}/A)}{dD}. \tag{3.48}$$

While the pressure will be mathematically equivalent using any of Eqs. 3.45-3.47, we will evaluate this derivative twice using the free energy expressions that connect to the symmetric and antisymmetric order parameter profiles in order to highlight competing length scales that arise for asymmetric profiles.

For Comparison to Symmetric Order Parameter Profiles

Using Eq. 3.46 as the Landau free energy, the pressure between the surfaces may be evaluated as

$$\Pi = -\frac{h_+h_-}{4c}\mathrm{csch}^2\left(\frac{D}{2\lambda}\right)\left[1 + \frac{(h_+ - h_-)^2}{h_+h_-}\frac{\sinh^2(D/2\lambda)}{\sinh^2(D/\lambda)}\right] \tag{3.49}$$

which indicates an attractive interaction when $h_+ > 0$ and $h_- > 0$. For large D, the pressure scales as

$$\Pi \sim -\frac{h_+h_-}{c}e^{-D/\lambda}\left[1 + \frac{(h_+ - h_-)^2}{h_+h_-}e^{-D/\lambda}\right], \tag{3.50}$$

which indicates an exponentially decaying attraction at the largest D values. Another pressure component, that depends on the relative values of h_+ and h_- (and vanishes when they are identical) emerges at a length scale of $2D/\lambda$, and this can

be attractive or repulsive.

For Comparison to Antisymmetric Order Parameter Profiles

The pressure can be obtained by taking the negative derivative of Eq. 3.47 with respect to D, as above. This yields

$$\Pi = -\frac{h_+h_-}{4c}\text{sech}^2\left(\frac{D}{2\lambda}\right)\left\{1+\frac{(h_++h_-)^2}{h_+h_-}\left[\frac{\sinh^2(D/2\lambda)}{\sinh^2(D/\lambda)}-\frac{\tanh(D/2\lambda)}{\tanh(D/\lambda)}\right]\right\}. \tag{3.51}$$

At large D, the pressure scales the same as before, but we can rewrite it differently due to the substitution made for the surface fields:

$$\Pi \sim -\frac{h_+h_-}{c}e^{-D/\lambda}\left[1+\frac{(h_++h_-)^2}{h_+h_-}\left(e^{-D/\lambda}-1\right)\right], \tag{3.52}$$

or, to highlight the second, competing lengthscale that emerges for asymmetric profiles, we can write

$$\Pi \sim -\frac{h_+h_-}{c}e^{-D/\lambda}+\frac{(h_++h_-)^2}{c}\left[e^{-D/\lambda}-e^{-2D/\lambda}\right]. \tag{3.53}$$

We emphasize that the asymmetric order parameter profile approach taken here extends previous work and is capable of reproducing earlier results in the corresponding symmetric/antisymmetric parameter limits [151].

3.10.9 Competing Length Scales: Zero Crossing

The expressions for Π provided (Eqs. 3.49 and 3.51) have two characteristic length scales: (D/λ) and $(D/2\lambda)$. Having different prefactors, these terms compete with one another at small D for some combinations of $h_\pm$. We seek to identify conditions under which these terms contrast so strongly that a nontrivial zero in Π emerges.

We first re-write Eq. 3.49 multiplying through the square brackets and then set $\Pi = 0$.

$$\Pi = -\frac{h_+h_-}{4c}\text{csch}^2\left(\frac{D}{2\lambda}\right)+\frac{-(h_+-h_-)^2}{4c}\frac{\text{csch}^2(D/2\lambda)\,\sinh^2(D/2\lambda)}{\sinh^2(D/\lambda)} \tag{3.54}$$

$$= -\frac{h_+h_-}{4c}\text{csch}^2\left(\frac{D}{2\lambda}\right)+\frac{-(h_+-h_-)^2}{4c}\text{csch}^2\left(\frac{D}{\lambda}\right)=0. \tag{3.55}$$

We see that the factor of $4c$ can be cancelled. Then, re-arranging to separate D and $h_\pm$, we may write

$$\frac{-(h_+-h_-)^2}{h_+h_-}=\frac{\text{csch}^2(D/2\lambda)}{\text{csch}^2(D/\lambda)}=4\text{cosh}\left(\frac{D}{2\lambda}\right) \tag{3.56}$$

where the last equality utilizes the trigonometric identity $\sinh(2x) = 2\sinh(x)\cosh(x)$.

We may finally solve for the value D/λ where $\Pi = 0$ to obtain

$$D/\lambda = 2\ \cosh^{-1}\Big(\sqrt{\frac{-(h_+ - h_-)^2}{4h_+h_-}}\Big). \tag{3.57}$$

The inverse hyperbolic cosine function $\cosh^{-1}(x)$ is only defined for arguments $x > 1$. This places the following bounds on $h_\pm$ in order for a nontrivial zero to exist:

$$\frac{-(h_+ - h_-)^2}{4h_+h_-} > 1 \tag{3.58}$$

This indicates that in order for a sign reversal to occur for Π, h_+ must have a different sign than h_- and their magnitudes must satisfy

$$h_+h_-(h_+ + h_-)^2 < 0. \tag{3.59}$$

via Eq. 3.58. This inequality is not satisfied for either the symmetric nor antisymmetric boundary conditions (neither $h_+ = h_- = h$ nor $h_+ = -h_- = h$), illustrating how the second competing length scale only manifests for asymmetrical surface polarization as derived here. Any asymmetric combination of $h_\pm$ having opposite signs will result in characteristic nonmonotonic behavior.

3.10.10 Competing Length Scales: Pressure Turnaround

We further compute $\mathrm{d}\Pi/\mathrm{d}D$ to identify the abscissa of its stationary point. Using Eq. 3.55 as a starting place, we evaluate the derivative term-wise with the identity $\frac{\mathrm{d}}{\mathrm{d}x}\big(\mathrm{csch}^2(Ax)\big) = -2a\ \cosh(Ax)/\sinh^3(Ax)$ for arbitrary constant A. This gives

$$\frac{\mathrm{d}\Pi}{\mathrm{d}D} = \frac{h_+h_-}{4c\lambda}\frac{\cosh(D/2\lambda)}{\sinh^3(D/2\lambda)} + \frac{(h_+ - h_-)^2}{2c\lambda}\frac{\cosh(D/\lambda)}{\sinh^3(D/\lambda)}. \tag{3.60}$$

We now set the derivative equal to zero and rearrange to separate D and λ,

$$\frac{-2(h_+ - h_-)^2}{h_+h_-} = \frac{\cosh(D/2\lambda)\ \sinh^3(D/\lambda)}{\sinh^3(D/2\lambda)\ \cosh(D/\lambda)}. \tag{3.61}$$

We take advantage of the fact that $\sinh^3(D/\lambda)/\sinh^3(D/2\lambda) = 8\cosh^3(D/2\lambda)$ via the half-argument formula $\sinh(x) = 2\ \cosh(x/2)\ \sinh(x/2)$ to write

$$\frac{-2(h_+ - h_-)^2}{h_+h_-} = \frac{8\cosh^4(D/2\lambda)}{\cosh(D/\lambda)}, \tag{3.62}$$

followed by a second half-argument identity $\cosh^2(x/2) = \frac{1}{2}(\cosh(x) + 1)$ to finally express

$$\frac{-(h_+ - h_-)^2}{h_+h_-} = \cosh(D/\lambda) + \mathrm{sech}(D/\lambda) + 2. \tag{3.63}$$

The general relation $\cosh(x) + \mathrm{sech}(x) + A = 0$ with constant A has an accessible

solution $x = \pm\cosh^{-1}(\frac{1}{2}(\sqrt{A^2-4}-A))$ on the domain $A < 2$, which can easily be checked. In our case, $A = 2 + (h_+ - h_-)^2/h_+h_- = (h_+^2 + h_-^2)/h_+h_-$, and we take the positive solution:

$$D/\lambda = +\cosh^{-1}\left(\frac{1}{2}\left[\left|\frac{h_+^2 - h_-^2}{h_+h_-}\right| - \frac{h_+^2 + h_-^2}{h_+h_-}\right]\right) \tag{3.64}$$

which can be equivalently expressed with the conditional statement,

$$D/\lambda = \begin{cases} \cosh^{-1}(-h_+/h_-), & \text{if } h_+^3 h_- < h_-^3 h_+ \\ \cosh^{-1}(-h_-/h_+) & \text{otherwise.} \end{cases} \tag{3.65}$$

This illustrates that the polarization-induced pressure profile will have a stationary point at larger D/λ values for dissimilar h_+ and h_-. This expression is only valid when the inequality in Eq. 3.59 is satisfied.

As a visual demonstration, we plot Π as a function of D/λ and illustrate the stationary point and intercept for relevant parameter choices $h_\pm$ (Fig 3.17).

3.11 Polarization-Induced Interactions in the Presence of an Electric Field

In this section, we evaluate how the above results are changed by the presence of an external electric field $\mathcal{E}$. We will show that this enables us to relate the phenomenological parameter a in the Landau free energy to the static dielectric constant of the bulk solvent. In this system, the electric field couples to the total (macroscopic) polarization density of the system, $\mathbf{P}$. The electric field modifies the free energy density as

$$\frac{\mathcal{F}}{V} = \frac{\mathcal{F}_0}{V} - \mathcal{E}\cdot\mathbf{P}, \tag{3.66}$$

where

$$\mathbf{P} = \frac{1}{V}\int \mathrm{d}\mathbf{r}\eta(\mathbf{r}), \tag{3.67}$$

which becomes

$$P_z = \frac{1}{D}\int_{-D/2}^{D/2} \mathrm{d}z\eta(z) \tag{3.68}$$

for the planar system studied here, and

$$\frac{\mathcal{F}_0}{V} = \frac{1}{D}\int_{-D/2}^{D/2} \mathrm{d}z\left[a\eta^2(z) + c|\nabla\eta(z)|^2\right] + h_+\eta(D/2) + h_-\eta(-D/2) \tag{3.69}$$

is the free energy density in the absence of the electric field.

Now, we consider applying the field only in the z-direction. In this limit, the free

energy per unit area is

$$\frac{\mathcal{F}}{A} = \int_{-D/2}^{D/2} \mathrm{d}z \left[a\eta^2(z) + c|\nabla\eta(z)|^2 - \mathcal{E}_z\eta(z) \right] + h_+\eta(D/2) + h_-\eta(-D/2) \quad (3.70)$$

The functional derivative of this free energy allows us to minimize it, and this is given by

$$\begin{aligned} \frac{\delta(\mathcal{F}/A)}{\delta\eta(z)} &= 2a\eta(z) - 2c\eta''(z) - \mathcal{E}_z + \delta(D/2 - z)[h_+ + 2c\eta'(D/2)] \\ &\quad + \delta(D/2 + z)[h_- - 2c\eta'(-D/2)]. \end{aligned} \quad (3.71)$$

At this point, we again set the bulk and surface terms to zero independently of one another. This gives:

$$a\eta(z) - c\eta''(z) = \frac{\mathcal{E}_z}{2} \quad (3.72)$$

$$h_+ + 2c\eta'(D/2) = 0, \quad (3.73)$$

and

$$h_- - 2c\eta'(-D/2) = 0. \quad (3.74)$$

We can solve for the order parameter, as above, to obtain

$$\eta(z) = \frac{\mathcal{E}_z}{2a} + \tilde{\eta}_R \frac{\sinh(\frac{D}{2\lambda} + \frac{z}{\lambda})}{\sinh(D/\lambda)} + \tilde{\eta}_L \frac{\sinh(\frac{D}{2\lambda} - \frac{z}{\lambda})}{\sinh(D/\lambda)}, \quad (3.75)$$

where $\tilde{\eta}_{R/L} \equiv \eta_{R/L} - \mathcal{E}/2a$. To explicitly illustrate the effect of the electric field on the order parameter, we can rewrite it as

$$\eta(z) = \eta_0(z) + \frac{\mathcal{E}_z}{2a}\left[1 - \sinh^{-1}(D/\lambda)\left(\sinh\left(\frac{D}{2\lambda} + \frac{z}{\lambda}\right) + \sinh\left(\frac{D}{2\lambda} - \frac{z}{\lambda}\right)\right)\right]. \quad (3.76)$$

where $\eta_0(z)$ is the order parameter profile without the electric field. We can write the surface fields as

$$h_\pm = -\frac{2c}{\lambda}[\tilde{\eta}_{R/L}\mathrm{coth}(D/\lambda) - \tilde{\eta}_{L/R}\mathrm{csch}(D/\lambda)] \quad (3.77)$$

or, emphasizing the effect of the electric field,

$$h_\pm = h_\pm^{(0)} + \lambda\mathcal{E}_z[\mathrm{coth}(D/\lambda) - \mathrm{csch}(D/\lambda)]. \quad (3.78)$$

where $h_\pm^{(0)}$ are the surface fields in the absence of the electric field.

To determine the free energy at the minimum, we need to evaluate

$$\frac{\mathcal{F}}{A} = -\frac{\mathcal{E}_z}{2}\int_{-D/2}^{D/2} \eta(z)\mathrm{d}z - c\eta'(z)\eta(z)\Big|_{-D/2}^{D/2}. \quad (3.79)$$

Evaluating this expression, using the form of $\eta(z)$ given in Eq. 3.76, yields

$$\frac{\mathcal{F}}{A} = \frac{\tilde{\mathcal{F}}_0}{A} + \frac{\mathcal{E}_z}{4a}(h_+ + h_-) - D\left(\frac{\mathcal{E}_z}{2a}\right)^2 - \frac{\lambda\mathcal{E}_z}{2}(\tilde{\eta}_R + \tilde{\eta}_L)[\tanh^{-1}(D/\lambda) - \sinh^{-1}(D/\lambda)]. \tag{3.80}$$

where $\tilde{\mathcal{F}}_0/A$ has the form of the free energy in the absence of the field, but with the surface fields implicitly depending on the renormalized $\tilde{\eta}_{R/L}$, as denoted by the tilde. The pressure can also be obtained by differentiating with respect to D,

$$\Pi = \tilde{\Pi}_0 - \left(\frac{\mathcal{E}_z}{2a}\right)^2 + \frac{\mathcal{E}_z}{2}(\tilde{\eta}_R + \tilde{\eta}_L)[\mathrm{csch}^2(D/\lambda) - \coth(D/\lambda)\mathrm{csch}(D/\lambda)], \tag{3.81}$$

where $\tilde{\Pi}_0 = d(\tilde{\mathcal{F}}_0/A)/dD$.

At this point, we can determine a. To do so, we first note that the static dielectric constant of a bulk liquid, ε, can be obtained from a slab-like system in the presence of an electric field through the following relation:

$$\varepsilon = \frac{\mathcal{E}_z}{\mathcal{E}_{\mathrm{tot}}(z=0)}, \tag{3.82}$$

where $\mathcal{E}_{\mathrm{tot}}(z)$ is the total electric field in the system as a function of z, and we assume that the center of the slab ($z = 0$) is bulk. To ensure that this is correct, we will take $D \to \infty$ later below. From macroscopic electrostatics, we know that the total electric field in this case has two components, the static field $\mathcal{E}_z$ that physically arises from some fixed external charge distribution and the polarization field of the water in response to $\mathcal{E}_z$,

$$\mathcal{E}_{\mathrm{tot}}(z) = \mathcal{E}_z - 4\pi P_z(z), \tag{3.83}$$

where $P_z(z)$ is define as above and was evaluated in order to obtain the free energy,

$$P_z(z) = \frac{\mathcal{E}_z}{2a} + \frac{\lambda}{D}(\tilde{\eta}_R + \tilde{\eta}_L)[\tanh^{-1}(D/\lambda) - \sinh^{-1}(D/\lambda)]. \tag{3.84}$$

We can see that the static field generates a static polarization in response, as expected macroscopically. We can then evaluate the value of $\mathcal{E}_{\mathrm{tot}}$ at the center of slab,

$$\mathcal{E}_{\mathrm{tot}}(0) = \mathcal{E}_z\left(1 - \frac{2\pi}{a}\right) + \frac{4\pi\lambda}{D}(\tilde{\eta}_R + \tilde{\eta}_L)[\tanh^{-1}(D/\lambda) - \sinh^{-1}(D/\lambda)], \tag{3.85}$$

which upon taking the infinite separation limit becomes

$$\lim_{D\to\infty} \mathcal{E}_{\mathrm{tot}}(0) = \mathcal{E}_z\left(1 - \frac{2\pi}{a}\right). \tag{3.86}$$

The dielectric constant is then given by

$$\varepsilon = \frac{\mathcal{E}_z}{\mathcal{E}_z(1 - \frac{2\pi}{a})} = \left(1 - \frac{2\pi}{a}\right)^{-1}. \tag{3.87}$$

which can be rearranged to obtain an expression for a,

$$a = 2\pi\left(1 - \frac{1}{\varepsilon}\right)^{-1}. \tag{3.88}$$

In SI units, a is given by

$$a = \frac{1}{2\epsilon_0}\left(1 - \frac{1}{\varepsilon}\right)^{-1}, \tag{3.89}$$

which has units of energy times length divided by charge squared, and ϵ_0 is the vacuum permittivity. Variability in a can then be linked directly to bulk dielectric constants via this expression.

3.12 Additional Discussion

We have developed here a framework for computing polarization-induced pressure between stationary planar surfaces as a function of separation. Distance-dependent character of Π is to be treated cautiously at small D when surface fields may couple nonlinearly. The present work emphasizes intersurface distances, D, larger than molecular sizes, because the order parameter — the far-field electric polarization — is obtained by LMF smoothing over a scale of σ. We note that such a smoothing is consistent with that often performed in standard textbooks to move from microscopic to macroscopic electrostatic descriptions [158]. The Gaussian (LMF) smoothing used here is performed over a lengthscale of $\sigma = 0.45$ nm, and we expect the validity of the theory to be questionable as D approaches this scale.

Due to limits of model applicability at small D/λ, one must determine whether, if a pressure turnaround should occur, it is physically relevant to the system at hand. As indicated by Eq. 3.65, this determination can be made by comparing the abscissa of the stationary point (or of the root) to the lower bound of model applicability. While the characteristic length λ is expected to vary between systems, in the present work it is found to be on the order of $\sim$2 Å. For a liberal lower bound of model predictability, $D/\lambda \sim 5$, the difference in relative magnitude between h_+ and h_- must be greater than $\sim$75-fold in order for a stationary point in Π to occur on $D/\lambda > 5$. Such an extreme is not reported here, and all observed pressure turnarounds were found on $D/\lambda < 5$ where van der Waals and other close-range interactions are expected to dominate the pressure. The relevance of these attractive basins on protein-surface adsorption are yet to be determined.

This work does demonstrate that in principle $\Pi(D)$ can show highly non-monatonic behavior, and the emergence of an attractive well is an exploitable property for the purposes of controlling assembly. In the limit where one surface polarizes solvent very weakly compared to the other (giving rise to a stationary point at large D/λ and a low-intensity Π at that abscissa) it is plausible for the attractive Π to dominate far-field electrostatics at intermediate D even if the signs of $h_\pm$ suggest repulsion at $D \to \infty$. This complex and nuanced behavior provides new insight for the design of protein-surface interfaces.

3.13 Consistency among Simulation, Theory, and Experiment

Here we compare the distance-dependent magnitude of muscovite-to-muscovite pressure obtained herein to previous studies of related systems. Our pressures, like in many other simulation studies, overestimate disjoining pressures observed experimentally in a way consistent with poor electrostatic screening in non-polarizable classical empirical molecular models.

A relevant experimental paper on disjoining pressure measurements is the recent work by Smith *et al.* [172]. They measured a total disjoining pressure (converted from force) of $\sim 10^{-10}$ nN/nm^2 at a separation of 2 nm (between mica) at 2 M NaCl. This is a total disjoining pressure as obtained by Surface Force Apparatus (SFA). This value is independently corroborated by Pashley and Israelachvili, obtaining an almost identical result at 2 nm separation [168]. Our calculated polarization contribution to this pressure with appropriate units set by a (in 3 M NaCl/KCl) is $\sim 10^{-6}$ nN/nm^2. Interestingly, the discrepency in simulated versus measured pressures of Ref [172] are further corroborated by Li *et al* using AFM [169]. Here the geometry is that of a mica "chip" fastened to an AFM tip probing a macroscopic mica surface. The measurements of Ref. [169] were estimated to be at separations of $\mathcal{O}(1)$ nm and results in a force of 10^{-3} nN/nm^2. This study used continuum theory to estimate specific ion effects due to screening and charge regulation. Importantly, the estimate of pressure to *desolvate* an ion (while pushing hard) is roughly 10^{-1} to 10 nN/nm^2 [169]. We view this as an upper bound of disjoining pressures beyond which you are seeing other phenomena *e.g.* oriented attachment where desolvation is necessary. Theoretical estimates of disjoining pressures obtained by Sushko and Rosso [170] are much larger than the desolvation estimates even at mineral separation distaces $>$ 2 nm. Nevertheless, we believe that obtaining accurate disjoining pressures to compare to either simulation or experiment is beyond the scope of the classical density functional model (water is modelled as a neutral hard-sphere) and are designed to point to length-scales where important new phenomena may exist.

The agreement with the high-quality simulations of Shen and Bourg [167] is excellent. The agreement with Shen and Bourg also corroborates our LR smoothing procedure as the disjoining pressures obtained herein show agreement to $\sim$ 1 nm separation between mineral surfaces. Caveats are the different mineral surfaces (although related to muscovite) and ionic strength of the solution (1 M salt). Shen and Bourg subtract a DLVO-like term from the computed atomistic potential of mean force to isolate a remaining "hydration-like" component of the interaction. We believe that a similar analysis could be applied with the results and methods presented in this research. Specifically, our work provides an additional reduced description of the electrostatic interaction that is *consistent* with underlying simulation that can be subtracted from the raw (unsmoothed) atomistic force or pressure data. The remaining piece will contain SR electrostatics and additional LR forces not due to polarization. Because the SR electrostatics are known, the remaining LR scaling of the interactions can be isolated and compared to DLVO theory. Previous works have indicated that electrostatic hydration repulsion can exceed the magnitude of DLVO attraction at small separations [151], highlighting the importance of isolating competing terms to uncover driving forces.

We hypothesize that the (significant) overestimation of computed long-range

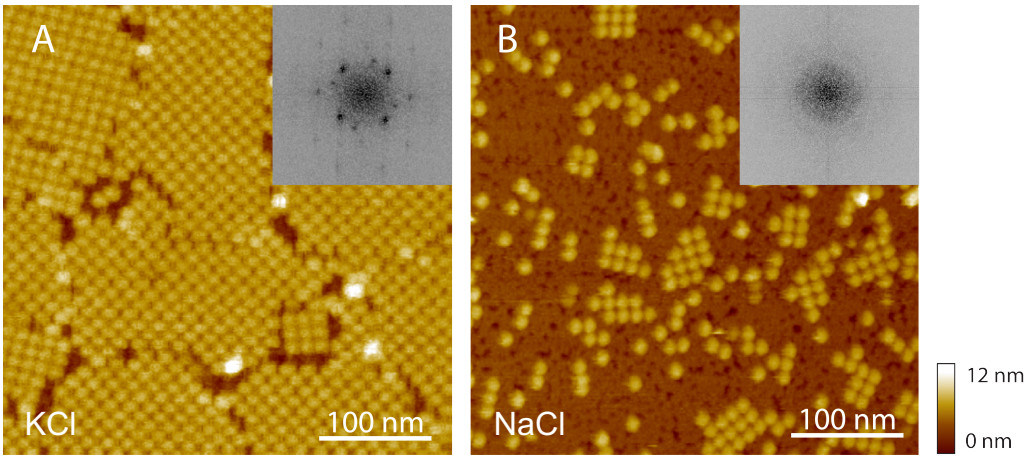

Figure 3.6: AFM images of ^{C98}RhuA assembly on muscovite. (A) 0.05 μM ^{C98}RhuA in 3 M KCl. (B) 0.05 μM ^{C98}RhuA in 3 M NaCl. The insets are Fast Fourier Transforms. The buffer is 20 mM Tris, pH 7. Exchanging KCl for NaCl is found to diminish protein coverage and reduces orientational specificity.

polarization pressure relative to measurement is due to the double layer's screening effect on the surface. As a simple exercise we can demonstrate the effect of screening length in our estimated disjoining pressures. If we review our muscovite-muscovite pressure data and change the screening length from 2 Å to 1 Å (otherwise staying consistent with the Landau theory framework) we recover values predicted by Perkin (see Figure 3.19). Importantly, the forces/pressures that are measured by Perkin *et al.* beyond 2 nm are due to LR electrostatic correlations (as opposed to dispersion) and thus, we believe, are a good comparison to our method. Our study suggests that empirical models under-screen the mineral surface. For ordered dipoles in a flat hexagonal array with lateral spacing on the order of $d = 5$Å, a decay length of $\lambda \approx \sqrt{3}d/4\pi$ has been suggested [181], which amounts to $\lambda \leq 1$ nm, again pointing toward an under-screening in our model. This is not surprising, in light of well-known issues that arise from neglecting dielectric contrast [182, 183] using the framework of non-polarizable empirical potentials. Overestimation of pressures in simulation have been reported previously, for similar and dissimilar systems [151]. It should be noted, however, that attempts to estimate λ from computer simulations have varied from 0.6 to 6 Å depending upon situation, illustrating the sensitivity of pressure outcomes to fitted boundary terms [181]. This suggests that faithful models of the double layer are important to be predictive – and that comparison to SFA force data and colloidal theory are relevant paths to connecting the simulation framework with experiment.

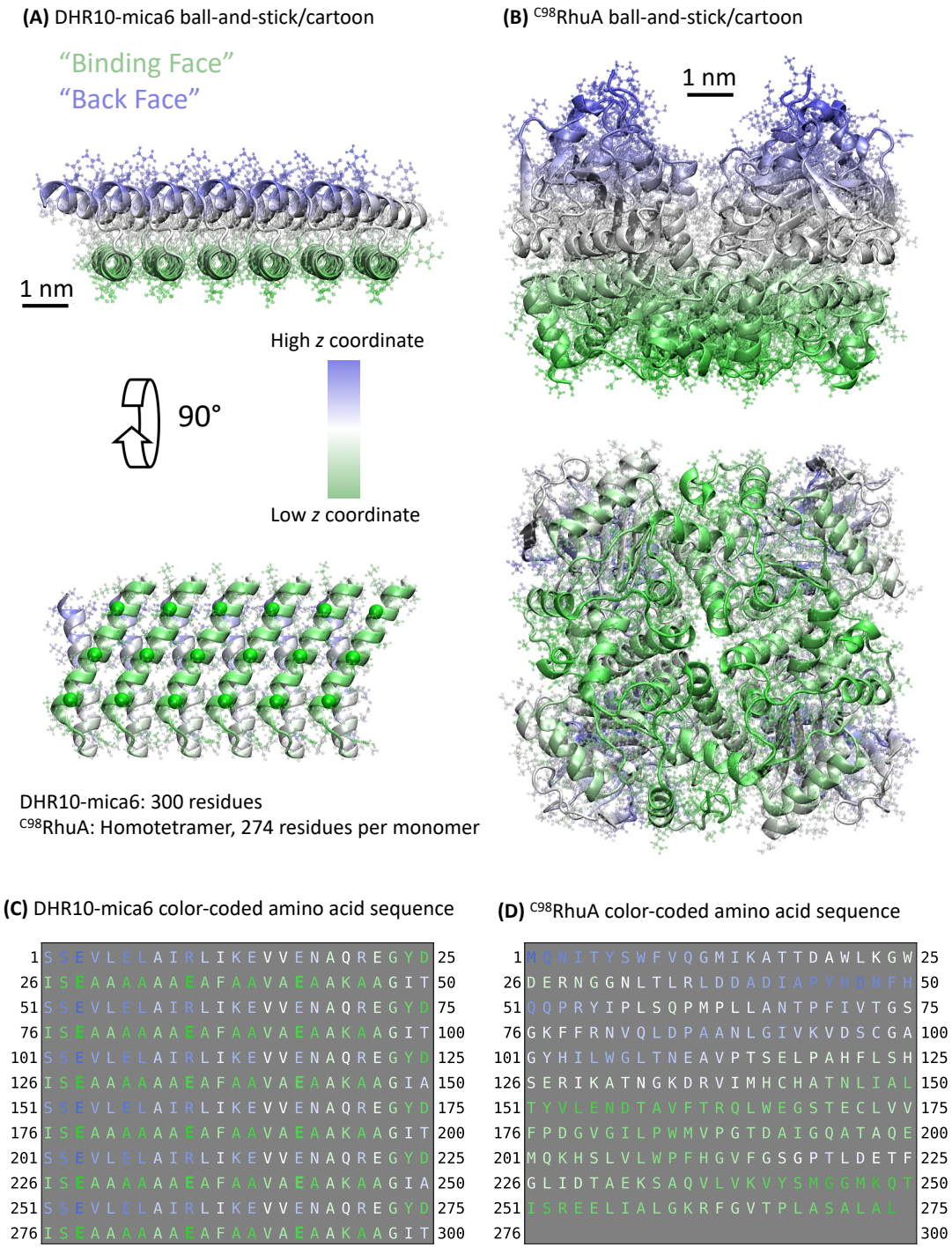

Figure 3.7: (A/B) Protein structures for (A) DHR10-mica6 and (B) ^{C98}RhuA as rendered using both a cartoon backbone and ball-and-stick side chain model, colored corresponding to depth in the z-dimension. The notional binding face appears in green (low z) while the opposite end, the "back" face, is blue (high z). For DHR10-mica6, the designed binding glutamate residue alpha carbons are marked by green vDW spheres to highlight the hexagonal patterning to match muscovite. (C/D) Amino acid sequences for the two proteins are color-coded according to the z-dimension in agreement with panels (A/B).

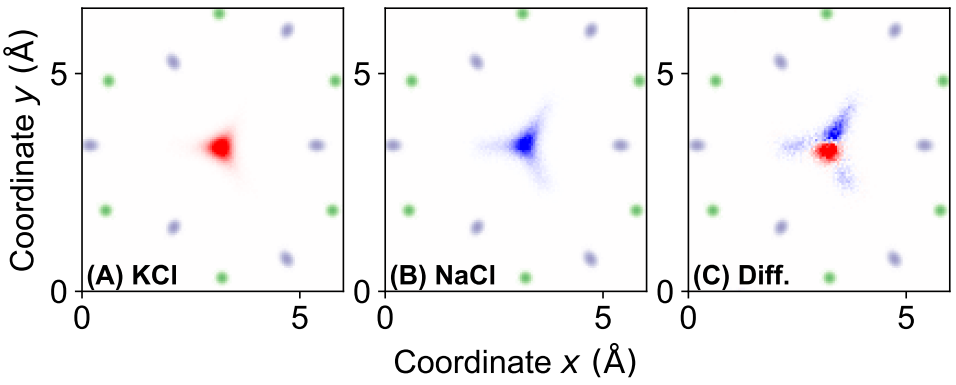

Figure 3.8: Lateral positioning of cations in their primary muscovite binding wells, measured from simulation for cations within the first density peak (within 0.2 to 0.3 nm from the surface bridging oxygen). The periphery of the binding pocket is outlined for reference with oxygen atoms (purple) and silicon (green). Panel (A) shows 3 M KCl conditions and potassium is in red. Panel (B) shows 3 M NaCl conditions and sodium is in blue. Panel (C) is a composite of (A) and (B), taking a difference between K^+ and Na^+ probability densities with regions of greater K^+ probability in red and greater Na^+ probability in blue, with white indicating equivalence. Panel (C) reveals that sodium is more likely than potassium to be displaced from the center of its cation binding well.

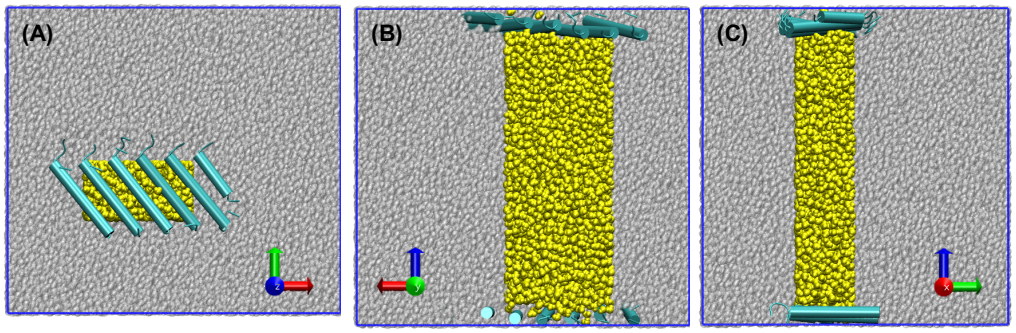

Figure 3.9: Visual representation of volume region attributed to the protein-confined space (yellow), with panels (A-C) corresponding to different viewing axes. Solution outside this volume was ignored when projecting $\rho(\mathbf{r})$ onto $\rho(z)$ (transparent gray). The three images are different angles of DHR10-mica6. Isolating a subvolume of space in this way amplifies the signal attributed to the protein compared to projecting averaging over all space, which includes excess solvent.

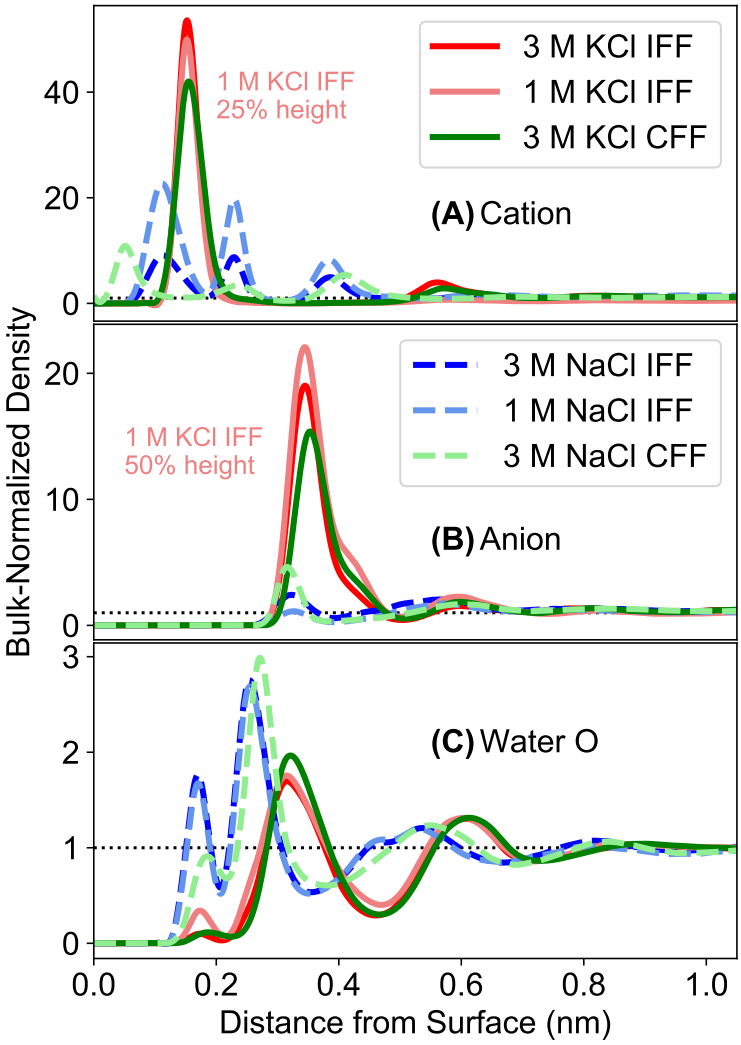

Figure 3.10: Probability densities from atomic coordinates of (A) cations, (B) anions, and (C) water oxygen, normalized by concentration in the bulk. Each panel compares six simulated salt treatment and mean-field substitution muscovite force field combinations: 3 M KCl, IFF (red); 1 M KCl, IFF (light red); 3 M KCl, CFF (green); 3 M NaCl, IFF (blue); 1 M NaCl, IFF (light blue); 3 M NaCl, CFF (light green). No qualitative change in density profile can be seen with changing concentration, only quantitative intensity relative to bulk. IFF and CFF disagree only in how closely a Na^+ ion can approach the surface.

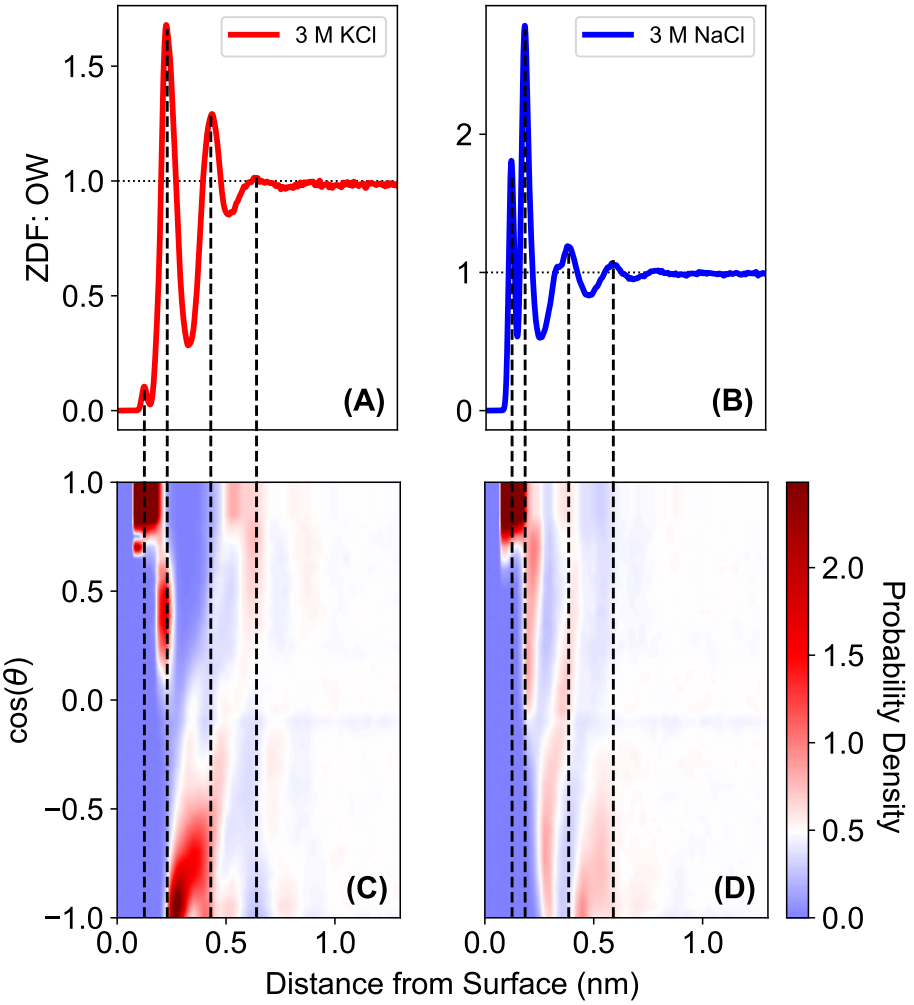

Figure 3.11: (A/B) Positional probabilty densities from atomic coordinates of water oxygen, normalized by concentration in the bulk, for (A) 3 M KCl, and (B) 3 M NaCl conditions. (C/D) Corresponding probability density of a water molecule dipole vector taking an angle θ with respect to muscovite surface normal given that the water oxygen is some distance from the surface. The value $\cos(\theta) = 1$ corresponds to the positive-to-negative vector of water being parallel to the surface normal into solution (positive end into the surface). Differences in water orientations from an atomistic perspective rationalize the surface polarization reversal observed in the LMF formalism.

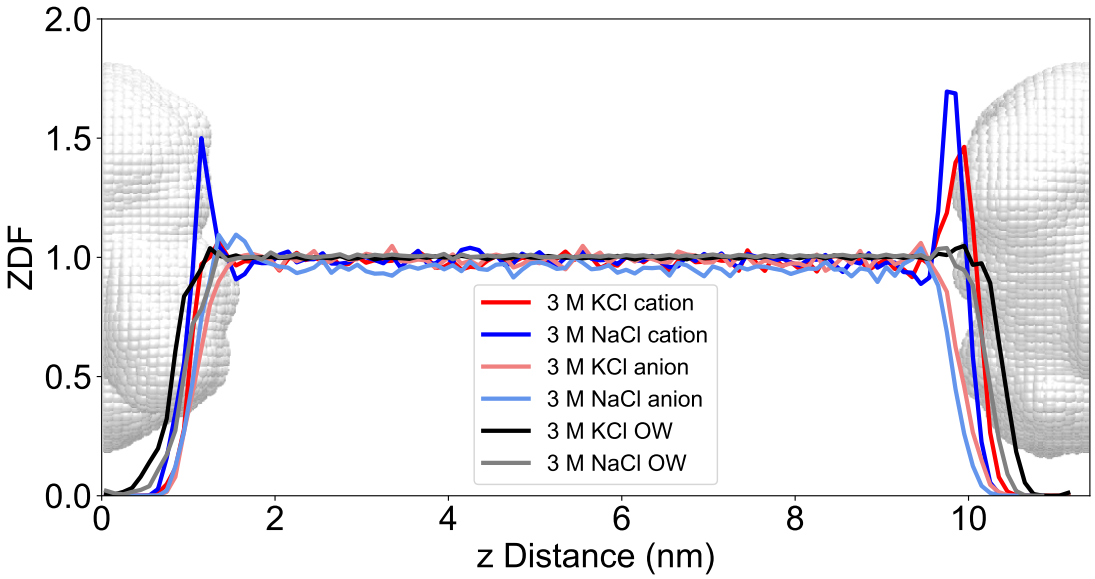

Figure 3.12: Probability density of cation and anion in the region between one face of DHR10-mica6 and the other (as in Fig S4). The binding face is at $\approx$ 10 nm and the opposing face is just above 0 nm. A small Cl^- adsorption peak can be seen in the NaCl case at $\sim$1.5 nm. Differences in ion aggregation tendences are an atomistic signature of solvent polarization and provides further evidence for the changing character of the back face of DHR10-mica6.

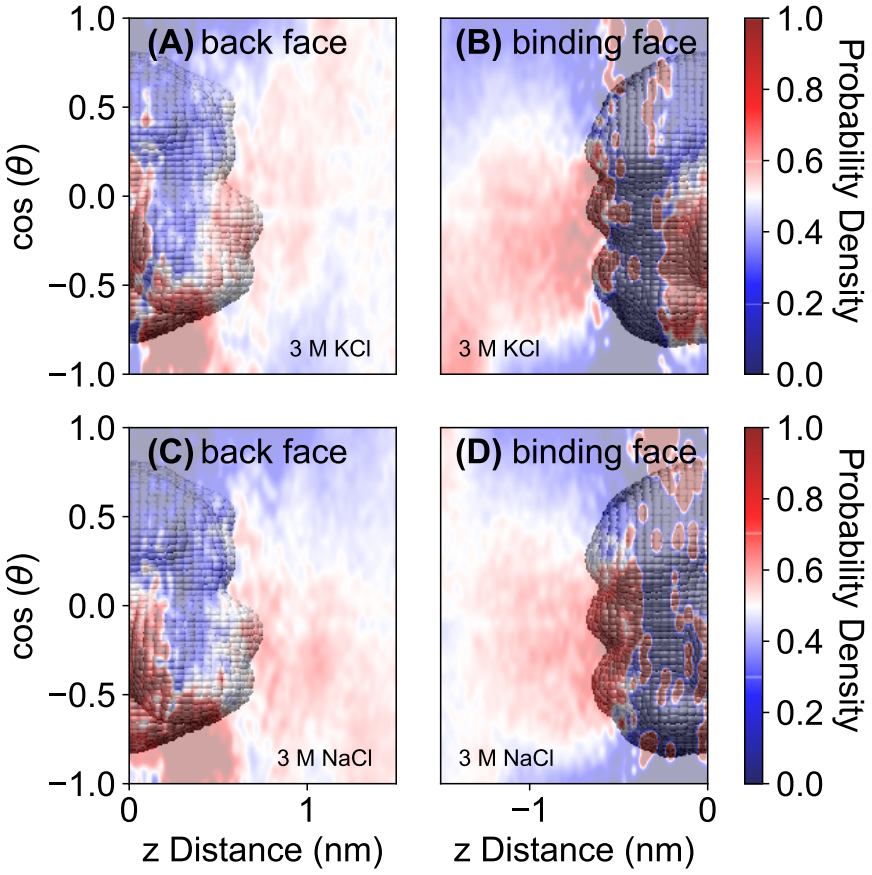

Figure 3.13: Probability distribution of water molecule orientation as a function of distance from either the back face (A/C) or the binding face (B/D) of DHR10-mica6, under 3 M KCl (A/B) or 3 M NaCl (C/D) conditions. The angle θ is between a water molecule dipole, negative to positive, and the vector {0,0,-1} (here running right to left). The left two panels represent water near the back face of the protein (overlaid as a ghost for visual reference) while the right two panels show the protein binding face, under KCl conditions (top) or NaCl conditions (bottom). Atomistic differences between aqueous K^+ and Na^+ corroborate the LMF result in terms of polarization intensity. The non-specificity of the side chain-cation interaction results in a similar cation density distribution adjacent to the protein binding face for both K^+ and Na^+.

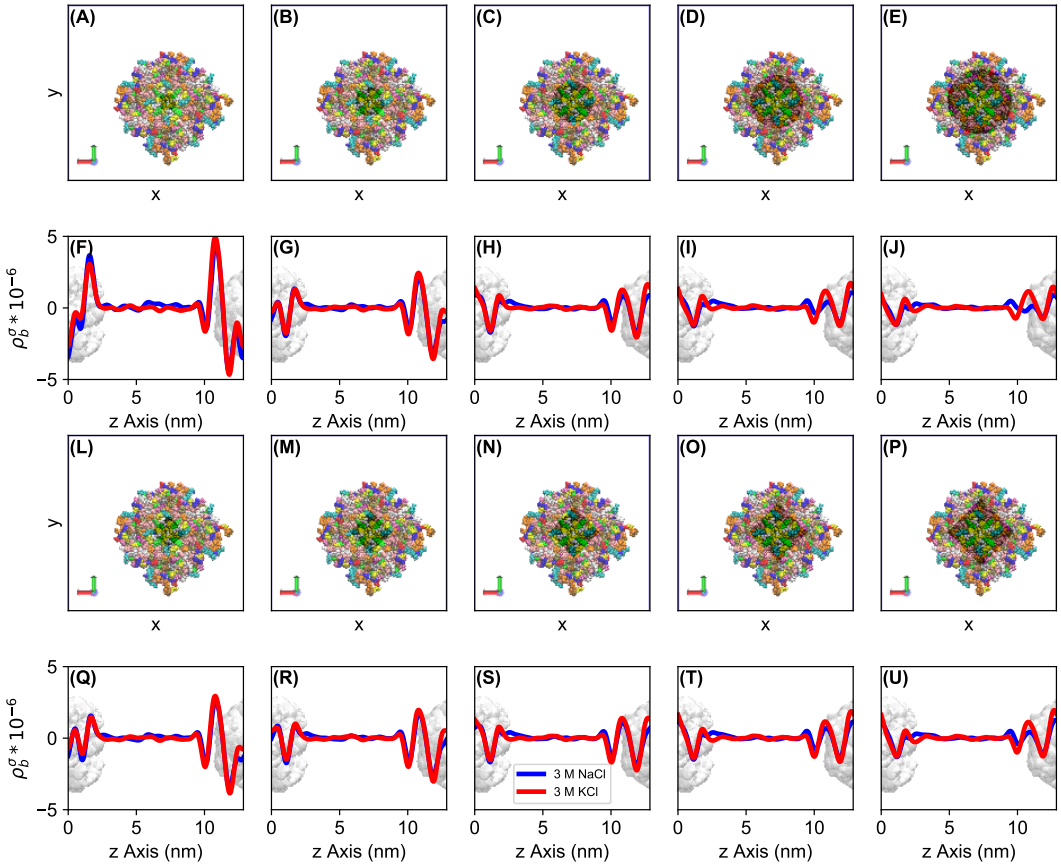

Figure 3.14: Ramifications of taking different x/y cuts on the projection $\rho_b^\sigma(z)$. Panels (A-E) show the lateral positioning of this cylindrical subvolume relative to the protein, with radii increasing from 7 Å to 24 Å. Panels (F-J) show corresponding projections $\rho_b^\sigma(z)$. At smaller subvolume radii, $\rho_b^\sigma(z)$ is similar for KCl and NaCl, while at larger subvolume radii the two begin to differ. The comparison is then replicated for a rectangular subvolume, represented in panels (L-P) and (Q-U), respectively.

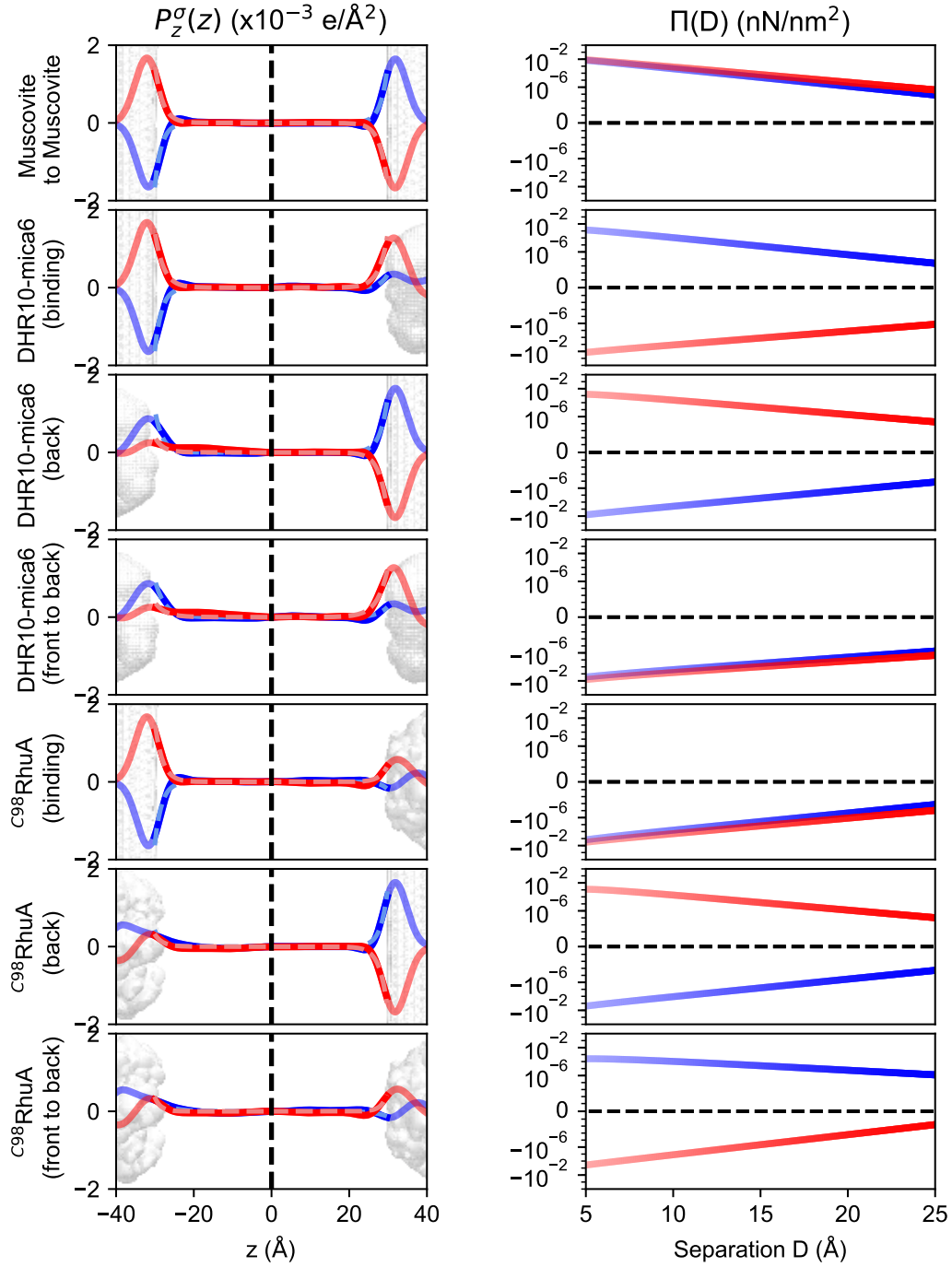

Figure 3.15: Assorted osmotic pressure profiles, complementary to Fig 4 in the main text in format. The functional form of $\eta(z)$ (as in Eq. 3.30 above) is fitted to electric polarization $P_z^\sigma(z)$. In each panel on the left side, the function $\eta(z)$ is shown as a dashed line over the solid trace $P_z^\sigma(z)$ from simulation. In each case, red represents 3 M KCl aqueous conditions while blue represents 3 M NaCl. Data from discrete simulations are concatenated at $z = 0$ in bulk, as marked by a dashed black line. Extracted parameters are used to evaluate Eq. 3.49 as a function of D, which are represented in the right column with any given row of panels corresponding to the same pair of macromolecules. Predictions about protein binding to muscovite are consistent with experiment.

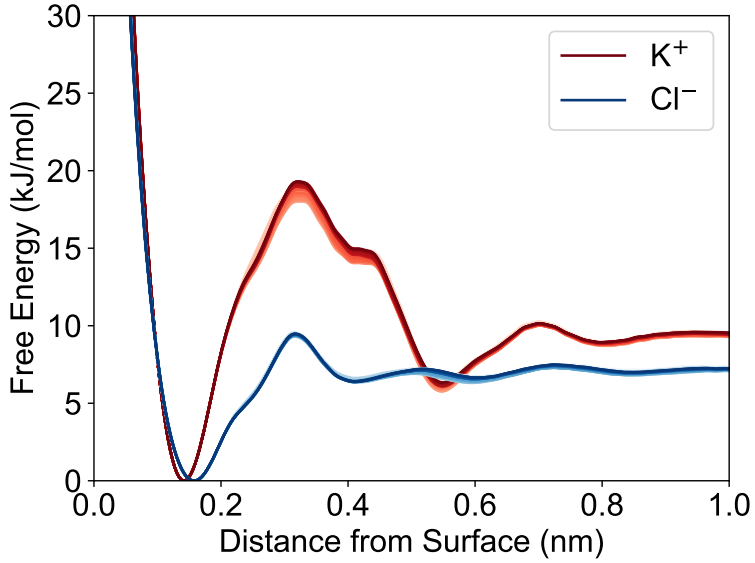

Figure 3.16: Progression of metadynamics PMF of ions over INTERFACE force field mean-field treatment muscovite in contact with 3 M KCl as a function of simulation time, from faint shades to darker in 10 evenly-spaced increments 10 ns apart from time 100 to time 200 ns out of 200 ns total simulation, which was the segment used for analysis. The lines all lie on top of one another, indicating a well-converged free energy landscape.

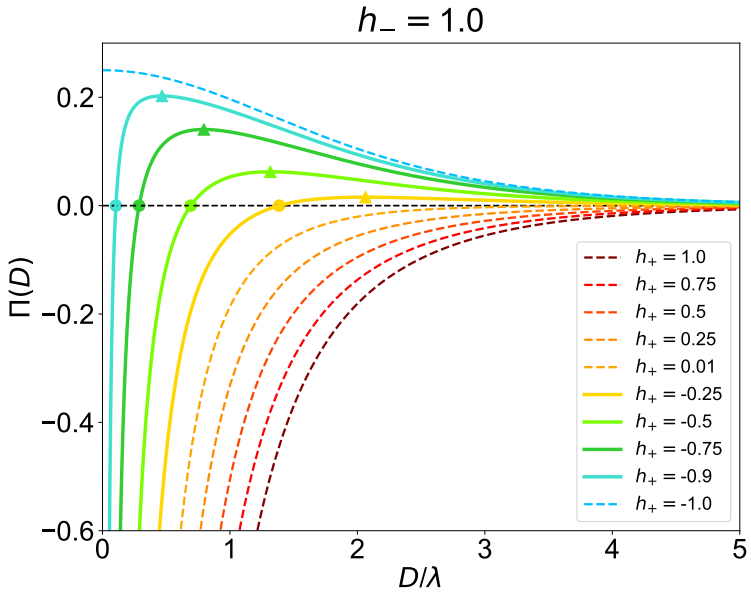

Figure 3.17: Osmotic pressure Π is plotted as a function of D/λ for $c = 1$ and $a = 1$. With $h_- = 1.0$ for all traces, the value of h_+ varies from -1 to 1. Dashed traces do not satisfy the inequality from Eq. 3.59 and so have neither intercepts nor stationary points for finite $D > 0$. Solid traces do satisfy Eq. 3.59 and so have intercepts marked closed circles and stationary points marked by triangles.

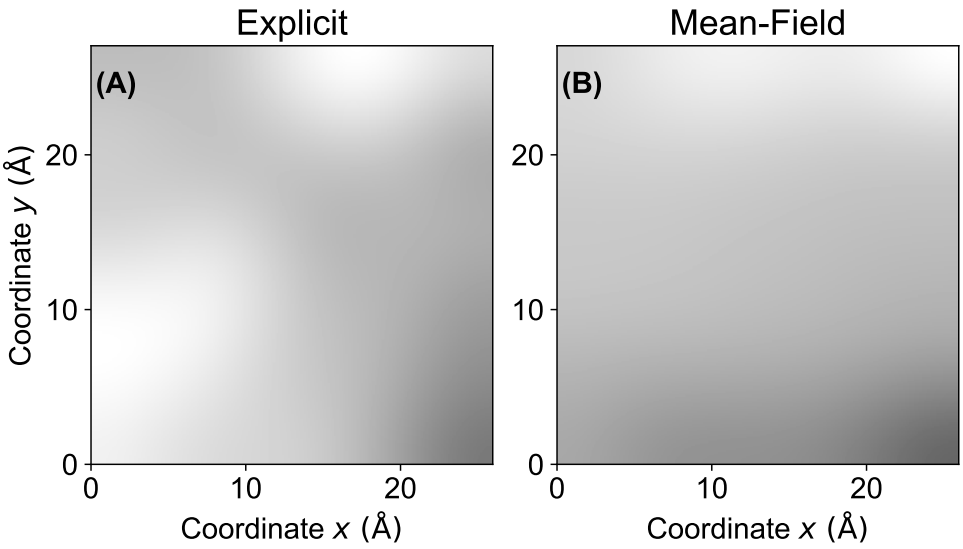

Figure 3.18: Average probability density of lateral water oxygen positions a distance of 3.0 ± 0.3 nm from the muscovite surface as measured from the surface-exposed oxygen atoms. After smoothing on the length scale of σ=0.45 nm, lateral inhomogeneities are washed out under both explicit Al substitution (C) and mean-field treatment (D).

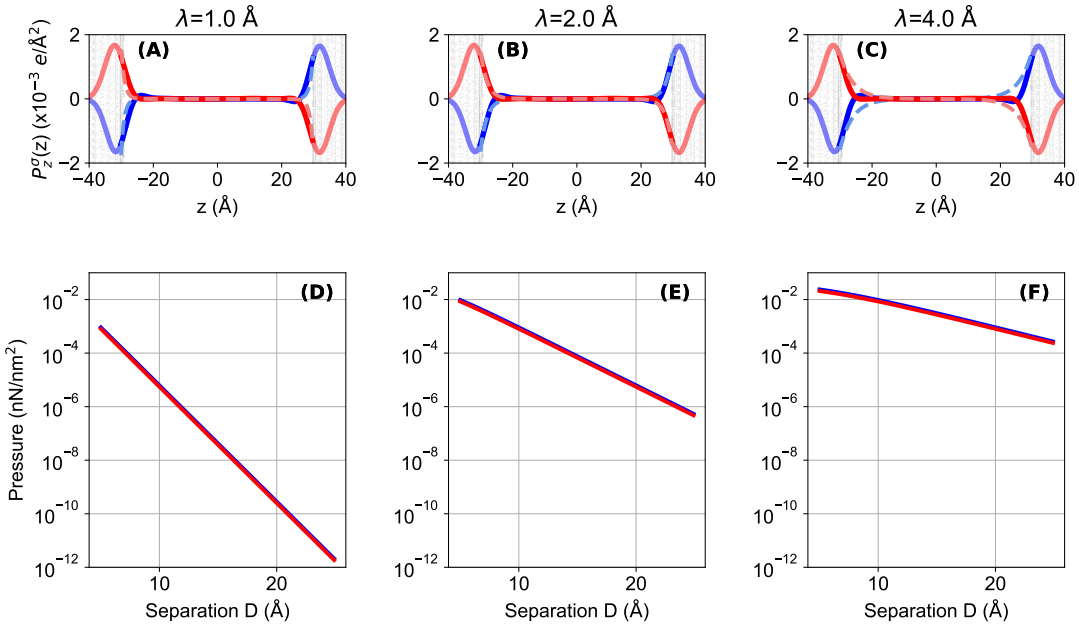

Figure 3.19: Demonstration of the effect of manipulating Landau decay length λ for muscovite-to-muscovite interactions. Panels (A)-(C) show how choosing a λ value other than $\sim$2 Å results in a poor fit relative to $P_z^\sigma(z)$ from simulation, while panels (D)-(F) reveal the substantive ramifications of this change on estimated pressure. Importantly, if $\lambda = 1$Å, coinciding with a more heavily screened surface, then pressure decays quickly and a value of 10^{-10} nN/nm^2 is obtained at 20 Å separation as in the experiments of Perkin *et al.*. If λ=4Å, coinciding with a weakly-screened environment, then larger pressures are maintained and have magnitudes comparable to other simulation works [171]

Chapter 4

Coupled Charge Carrier / Counterion Mobility in Organic Electrochemical Cells

4.1 Abstract

[1]Development of continually improved electronic devices from flexible and biocompatible organic semiconductors is hindered by a limited understanding of the molecular mechanisms by which electron carriers along the polymer backbones are coupled to the diffusion of aqueous counterions that are present for charge compensation. To shed light on the slow pair diffusion of electron hole and molecular counterions in the model material polythiophene, a two-sided computational study is presented that probes the ability of disordered polymer strands to confine intercalated counterions and of those counterions' abilities to restrict those electron transfer (ET) reactions that are favorable. A key modeling advancement, a fixed point charge model of the delocalized electron hole in a classical molecular dynamics force field, provides a sufficient driving force for counterions to permeate the polymer phase at low doping levels, with molecular stoichiometries that correlate with their relative hydrophobicities. Interestingly, the identity of an intercalated counterion has little impact on the distribution of ET reactions observed, pointing toward the importance of the number of counterions present and overall solution intercalation rather than specific polymer-ion interactions.

4.2 Introduction

Conjugated organic polymers that have been electrochemically doped are semiconducting materials with applications in device manufacture (actuators, capacitors, etc.) that are preferable to traditional inorganic materials in medical applications due to an improved bio-compatibility [184, 185, 186]. They benefit from a number of technical advantages as well, including flexibility of the polymer that permits spin coating or ink jet printing deposition techniques onto various substrates, and a

[1]Reproduced with permission from Jesse Prelesnik, Ramsess Quezada, David Ginger, and Christopher Mundy. Coupled Charge Carrier / Counterion Mobility in Organic Electrochemical Cells. Work in progress.

diversity of conjugated backbone moieties and side chain functionalities [187]. This presents an opportunity to develop neuromorphic transistors and other such devices. The flexible nature of organic polymers permits aqueous solution to permeate the semiconductor, which couples intermolecular electron transfer (ET) between polymers with the motions of diffusive counterion species [188, 189]. The resulting device performance is exquisitely sensitive to design choices such as counterion identity and polymer side chain functionalization, among other things, which are relationships that must be fully understood to develop better-performing materials.

Of particular interest is the reversible doping/de-doping process, by which the polymer phase swells with aqueous solution in response to an applied potential and a steady state is reached [190, 191]. This occurs as the conjugated polymer backbone accepts electron carriers, and the oxidized or reduced polymer is charge-balanced by counterions from solution. The rate at which this occurs is dependent on specific polymer-ion interactions, as electronic and ionic transport rates are coupled [192]. Poly(3-hexylthiophene-2,5-diyl) (P3HT) is a benchmark conjugated polymer, in part due to its facile synthesis, regioregularity, and efficient π-π stacking, which gives strong optical and electronic responses [193]. While many of its properties have been well-studied, such as threshold voltage or the thickness of the polymer layer, the role of electrolyte identity is seldom emphasized [194]. The electronic properties of P3HT are altered in the presence of dopant, altering π-π stacking distances, which contributes to a delicate interplay between polymer chain charge transfer and compensatory ion transport.

Interestingly, polyatomic ion dopants such as hexafluorophosphate (PF_6^-) and bis(trifluoromethane)sulfonimide ($TFSI^-$) produce greater currents than halides, and the degree of ion mobility within the polymer phase is not simply a function of ion volume [194]. In fact, the doping process has been found to occur more quickly for bulkier anions. Further development of the devices that rely on these physics would be expedited with a more complete mechanistic understanding of the ion/polymer pair properties that result in enhanced ion mobility.

Particularly vexing is the coupling between counterion diffusivity and electron hole inter-chain reaction rate, which requires co-localization as they propagate across the device for neutralization. This coupling means that the overall pair mobility is rate-limited by the inability of either species to diffuse freely. The pair mobility can be assessed via "moving front" experiments in which half of a polymer film is doped and the gradual oxidation of the un-doped section of the polymer film can be monitored via optical response [195, 196]. It is this pair mobility that we ultimately seek to maximize, and to do so it is vital to understand the motions of the counterion and electron hole independently of one another in order to disentangle their coupling, to understand what components are rate-limiting for charge transfer in devices.

Molecular models have the potential to shine light on the mechanisms underlying these complex interactions, with precise control over polymer morphology and the ability to identify interactions between functional groups atomistically. P3HT has been well-studied at various levels of theory, from comparisons of quantum mechanical methods [197, 198, 199, 200, 201] to coarse-grained beads [202, 203, 204, 205, 206], with a delicate interplay between molecular detail and scalability. On one hand, the extensively delocalized conjugated π-system along the thiophene backbone requires a high level of theory to accurately capture its propensity for po-

larization, while on the other hand the solvent-intercalated polymer system is highly heterogeneous and involves length scales well beyond what quantum mechanics can practically investigate. The middle ground chosen here is with empirically parameterized atomistic molecular dynamics (MD) simulations, which allow for systems spanning nanometers and many thousands of atoms while maintaining control over molecular charge distributions.

This model is used to approach the problem of counterion/polaron co-diffusion from two sides: (1) A biphasic polymer/aqueous system with a bath of counterions is allowed to intermix as a result of polymer doping, revealing molecular stoichiometries in the swollen polymer phase and providing counterion diffusivity when in the presence of a doped polymer to assess the polymers' ability to confine the ion, and (2) When in the presence of a counterion, the rate of polymer-polymer electron hopping is estimated from classical simulation to assess the ability of a counterion to direct the electron transfers that are viable. These two components feed into a complexation mechanism by which charge hopping and counterion diffusion operate in succession along favorable "percolarion" pathways [207].

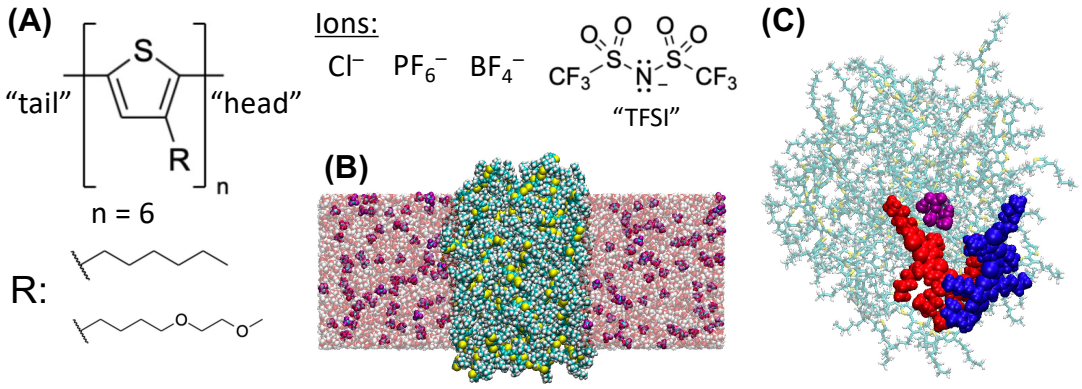

Figure 4.1: Schematic of the systems used herein. (A) illustrates the structure of the thiophene backbone, with "head" and "tail" directionalities labeled relative to the side chain attachment point to illustrate the difference between regioregular (head-to-tail connectivities only) and regiorandom (where head-to-head or tail-to-tail connections are also permitted, with random distribution). The R-group two side chains considered here are also illustrated, alongside the counterions employed to neutralize the electron hole on thiophene. Aqueous counterions are always K^+ in this study. Panel (B) illustrates the biphasic simulation setup used here, in which 128 hexamers (in teal) constitute the organic phase and an excess of water solvent (transparent red) at 0.5 M ionic strength (ions in purple) is driven to intercalate into the polymer phase due after doping has occurred. Panel (C) illustrates the electron transfer systems used, with 32 hexamers (transparent teal) and a single counterion/water cluster with some number of water molecules determined by simulations from panel (B). In this illustration, an electron hole is transferring from the blue polymer to the red to better neutralize with the purple cluster of counterion and waters.

A key modeling advancement presented here is the reparameterization of a popular force field for P3HT to accommodate a delocalized electron hole along the backbone, which is used for both parts (1) and (2) of the theoretical strategy. In

part (1), the presence of oxidized polymers provides a sufficiently hydrophilic environment to drive solvent influx from an adjacent aqueous bath, drawing counterions into polymer-confined spaces along with some water. Such penetrance is not observed in un-doped polymer systems [208], which are quite hydrophobic with relative permittivity typically less than 5 [209] and remain phase-segregated with aqueous solution. In part (2), a single counterion and its accompanying water molecules (as determined from part (1)) are placed within a polymer melt containing a single oxidized polymer. By applying an MD-based strategy to estimate the rate of polymer-polymer ET via Marcus theory [210], as has been applied in the biochemical community previously [211, 212], the impact of the counterion on ET rate can be directly estimated.

The remainder of this article describes the molecular model applied and then presents results and discussion on the ability of counterions and electron holes to influence one another in this framework. Polymer side chain identity, counterion identity, and polymer side chain regioregularity (uniformity of head-to-tail connectivity) are varied to assess their ramifications for bulk transport in operable devices. This information can then be extrapolated to other polymers and ion types to predict conditions that would promote efficient charge transport in the absence of experimental data.

4.3 Methods

4.3.1 Validation of Oligomeric Thiophenes to Represent Polymers

In choosing the theoretical model to apply, three specific prior findings were accounted for:

1. Experiments on the same patch of spin-coated P3HT indicate an anti-correlation between swelling and stiffness, where swelling is greatest in the softest areas of the film [213]. These correspond to the least-ordered regions of the P3HT film, in contrast to the well-aligned crystalline regions that have been well-studied. That is, amorphous regions of the P3HT melt swell the most with solvent, and are therefore most relevant to studying intercalated aqueous solution in the polymer phase.

2. A disordered polymer melt contains a disjointed conjugated π system, disrupting the extensive electronic delocalization that is typically seen [214]. The persistence length of the amorphous polymer has been debated [215, 216, 217], but is widely acknowledged to be on the order of 10 repeat units before kinks in the backbone effectively randomize the relative orientations of disparate polymer segments in the disordered phase. Moreover, electron carriers in the polymer backbone can only be delocalized across a conjugated system, and are therefore confined to linear regions of the polymer [218]. This permits molecular models of P3HT to effectively capture the behavior of disordered polymers using shorter oligomers on the order of the persistence length. Here we choose to model hexamers that have been thermally annealed, which we show reproduce structural and dynamical features of the disordered polymer

phase modeled with longer chains. The hexameric alkylthiophene modeled here will be abbreviated as P3HT, despite being insufficiently long to be a polymer by conventional standards.

3. Prior simulation works have suggested a number of classical empirical force fields for P3HT, with different parameterization philosophies. A recent comparative paper [219] established the force field of Moreno *et al.* [39] as most accurately reproducing the disordered phase as determined by X-ray scattering distances. The P3HT force field in question is based on the OPLS-AA family, maintaining features of the original wherever applicable – for example, alkyl groups share parameters with small organic molecules like propane [17]. This permits facile side chain functionalization to adjust counterion coordination abilities.

Disordered oligomers of P3HT have been previously reported in literature, providing a protocol for orientational randomization and thermal annealing [220, 208, 219]. These same works provide torsional distributions, backbone planarity metrics, and radial distribution functions for comparative purposes and to validate that our hexamers preserve the desired properties. All classical simulations reported herein were conducted using GROMACS 2020.4, with ensembles and simulation durations reported where relevant.

An important feature of the P3HT film is its glassy character, displaying long-timescale self-diffusivity and rotational autocorrelation [214, 221]. To ensure that the hexameric P3HT modeled here is in the appropriate phase at 300 K, we determine the glass transition temperature T_g via a change in the thermal expansion coefficient, $\alpha = \frac{1}{V}(\partial V/\partial T)_p$ for volume V, temperature T, and pressure p. The derivative of system volume with temperature is determined via Replica Exchange Molecular Dynamics [222]. This strategy was employed on a slab of polymers that was annealed first at 550 K (12 ns NPT), then 435 K (10 ns NPT), and finally 300 K (10 ns NPT, 20 ns NVT), both above, below, and near the expected glass transition temperature. The high temperatures (and excessive initial volume giving a density of $<$0.1 g/cm^3) ensure that molecular configurations are well-randomized to mimic the disordered phase before being completely condensed at the target temperature.

Once all replicas have thoroughly exchanged and visited the other temperature settings, an average volume is determined at each temperature and a plot of volume as a function of temperature is generated. On either side of the glass transition temperature the temperature dependence is linear, as schematized in Figure 4.10. The intersection point of the two linear trend lines indicates the glass transition temperature, which here is approximately 435 K. Knowing the glass transition temperature, the polymers can be annealed according to previously established protocols involving both high and low temperatures. Indeed, this phase transition can be confirmed by determining the end-to-end molecular autocorrelation time, which displays qualitatively different behavior for temperatures above and below the glass transition temperature T_g (Figure 4.2). Importantly, below T_g the polymer end-to-end vector does not substantively decorrelate over the simulation length, illustrating the slow characteristic time scales of the system at hand.

The homogeneous polymer phase was characterized structurally and dynamically for regioregular and regiorandom side chain connectivities, for hexyl and ether side chain functionalities (polymers named P3HT and P3APPT, respectively). Relevant

parameters include torsional distributions, the radial distribution functions $g(r)$ between bodies, bulk diffusivity, and ring planarity as a function of distance (Figure 4.3).

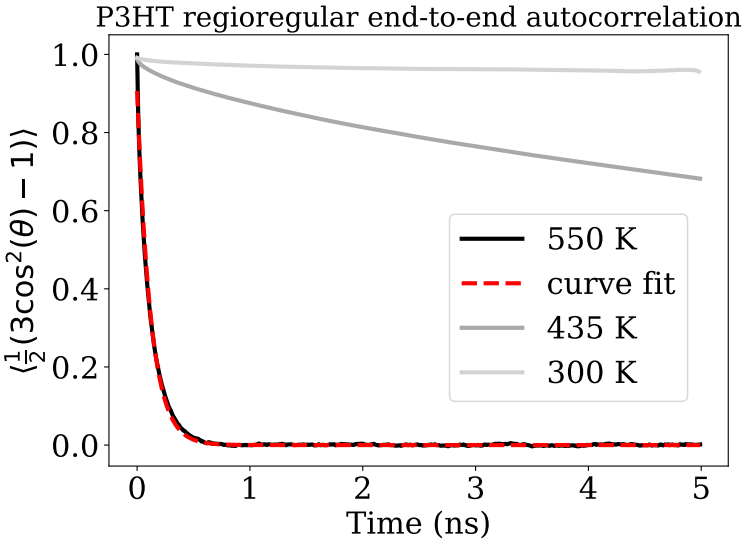

Figure 4.2: End-to-end autocorrelation function of regioregular P3HT with hexyl side chains. Above T_g = 435 K the end-to-end autocorrelation (black) obeys a standard exponential decay and can be fitted (dashed red) with a characteristic time constant on the length scale of nanoseconds. At $T > T_g$, however (light gray trace), the autocorrelation obeys roughly exponential behavior but on such a long time scale that a function cannot be appropriately fit on time scales relevant to chemical simulation. This resistance to molecular reorientation is what characterizes the glassy polymer state and is why simulations are to be conducted in replicates rather than extended to great lengths.

The radial distribution functions $g(r)$ between inter-chain thiophene ring centers of mass show little deviation as a result of side chain or regioregularity (Fig 4.3A). The $g(r)$ functions between thiophene ring centers of mass and intermolecular side chain terminal CH_3 groups similarly show small differences when comparing regioregular and regiorandom chemistries (Fig 4.3B). The alkyl side chain variant P3HT displays a diminished side chain-backbone interaction compared to the the hexyl functionalized P3APPT. This may be due to enhanced side chain-side chain interactions, since the the dipoles introduced by ether linkages allow those side chains to engage in stronger Coulombic attractions. While consistent with prior simulation works that report on the disordered P3HT melt via chemical simulation [220, 219], it is worth noting that experimental characterizations report more substantive differences between regioregular and regiorandom polymer films on the whole, which include some degree of crystalline and amorphous regions both [213, 223].

Torsional distributions show no statistically significant differences between side chain identities, which is to be expected since the most important dihedrals being considered (backbone planarity and side chain-backbone orientation) are defined via atoms that are unchanged from one model to another (Fig 4.3B, 4.3E). However, differences can be observed when regioregular chemistry is exchanged for regiorandom, particularly in the backbone S-C-C-S dihedral (Fig 4.3C). The regiorandom chemistries tend to prefer the *trans* backbone configuration (π rad) so as to avoid

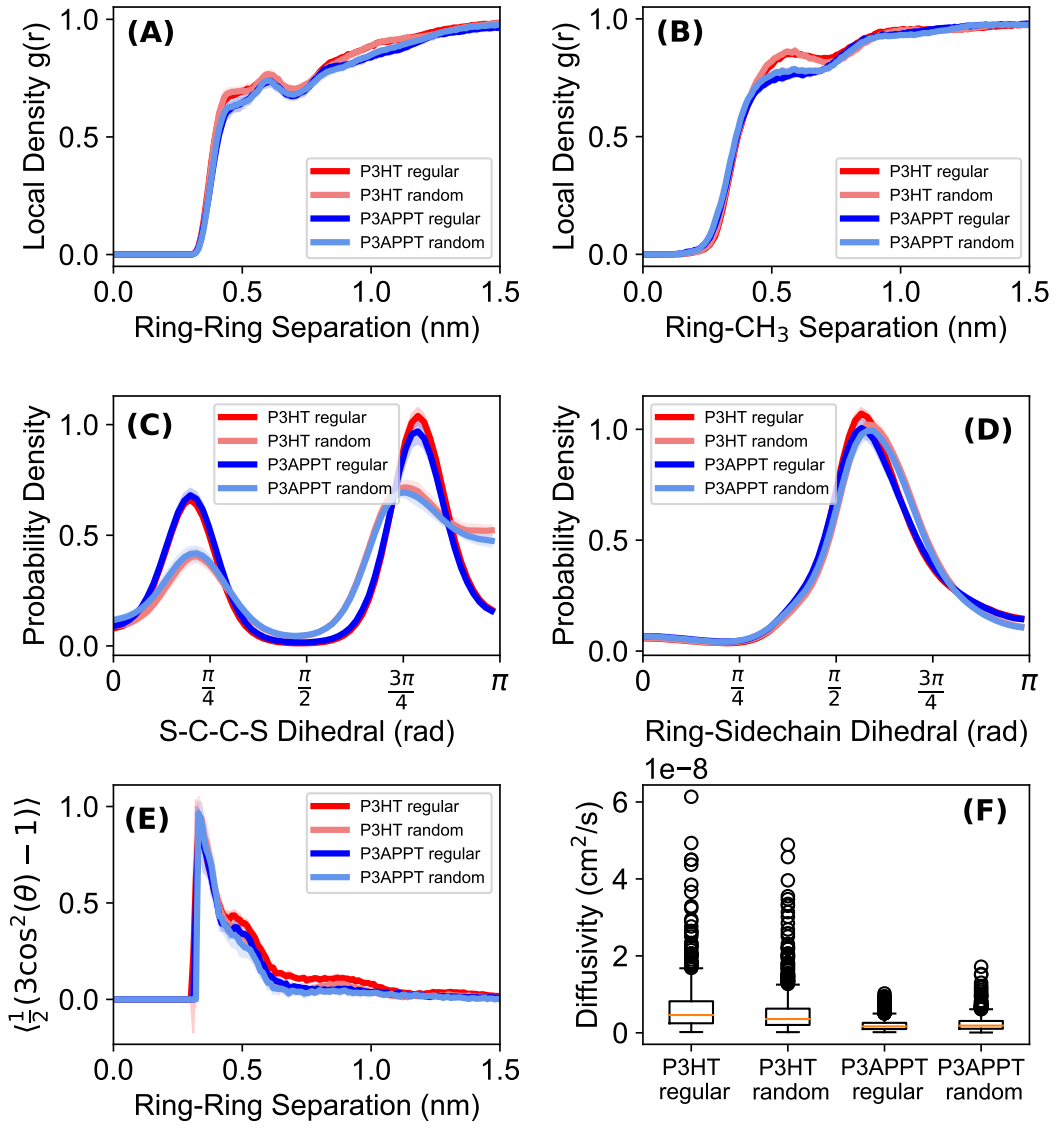

Figure 4.3: Characterization of homogeneous oligothiophenes functionalized with hexyl or ether side chains, and either regioregular or regiorandom connectivity. (A) The radial distribution function $g(r)$ between intermolecular backbone thiophene ring centers of mass, and its counterpart (B) the function $g(r)$ between thiophene ring center of mass and the center of mass of side chain terminal CH_3 groups, intermolecularly. (C) Probability distribution of thiophene-thiophene dihedrals in a homogeneous melt, as measured from S-C-C-S, with an angle of π corresponding to a *trans* configuration, and its counterpart (D), the dihedral distribution of the side side chain with respect to the backbone, as measured between the first two carbons of the side chain and the nearest two C_β and C_α carbons of the thiophene ring, such that $\pi/2$ corresponds to a perpendicular orientation. (E) A thiophene-thiophene co-planarity metric obtained from the second Legendre polynomial of the angle formed between the normal vectors of two intermolecular thiophene rings, in which a value of unity corresponds to perfectly co-planar rings and a value of zero indicates no correlation. (F) Molecular diffusion constants for individual oligothiophene molecules, obtained from mean-square-displacement, as a box and whisker plot. All values are less than 10^{-7}.

side chain steric clashes. In all cases, the side chain tends to be nearly perpendicular to the plane defined by the thiophene ring (Fig 4.3D), with only slight differences between regioregular and regiorandom chemistries. Again, the regiorandom functionalization tends toward a direction that reduces steric clash when adjacent side chains are put in close proximity.

The extent of backbone-backbone co-planarity may be assessed on a thiophene-by-thiophene basis, analyzing the distribution of angles θ formed between the normal vectors of each ring. By taking the second Legendre polynomial of this value, $P_2(\theta) = \langle(1/2)(3\cos^2(\theta) - 1)\rangle$, three distinct states may be produced: $P_2(\theta) = 1$ corresponds to perfectly co-planar rings, in which at all times $\cos(\theta) = \pm 1$; $P_2(\theta) = -0.5$ corresponds to perpendicular rings with $\cos(\theta) = 0$; and $P_2(\theta) = 0$ indicates no correlation between ring-ring angles. With this metric evaluated as a function of distance, one may observe how ring-ring co-planarity decays with separation. Indeed, Fig 4.3E shows that for all side chain functionalizations and regioregularities, at the closest separations thiophene rings must be perfectly co-planar for steric reasons. At larger distances, regioregular P3HT retains co-planarity better than does its regiorandom counterpart to a minor extent, while for P3APPT the distinction between regioregularities cannot be made.

Finally, molecular diffusivity of each polymer molecule (Fig 4.3F) is on the order of $<10^{-8}$ cm^2/s. While side-chain dependent differences are observed, these magnitudes are all ∼1,000-fold slower than the self-diffusivity of water, once again confirming that the polymers are in an extremely glassy state. When mobility is this low, differences between polymers are near-negligible. These characteristics demonstrate that the hexameric functionalized thiophenes presented here display the desired properties of a polythiophene.

4.3.2 Force Field Modification: Oxidized Polymer

Prior atomistic simulations confirm the observation reported here that P3HT in its neutral state is highly hydrophobic and does not swell without prior oxidation [208, 224]. Useful information about the nature of polymer-counterion interactions may be obtained from such simulations, but without solvent influx the doping process is not well-described. To extract quantities like counterion diffusivity in the polymer phase (a proxy for its doping and de-doping rate), it is necessary to model the polymer in its oxidized form to drive the swelling process. This presents a challenge because no existing classical force field attempts to model the delocalized electron hole along the polymer backbone.

Since no chemical bonds are made nor broken when the polymer is oxidized, the protocol to re-parameterize the P3HT hexamer to its charged form can be taken directly from the parent publication of the neutral form [39]. Specifically, bond, angle, and Lennard-Jones parameters are maintained as originally specified in Moreno *et al.*, and the parameters of side chains (those taken directly from OPLS-AA) are identical. The quantities of principal interest are atomic partial charges, namely those along the polymer backbone and extending to the first carbon atom of the side chain directly bound to the oxidized backbone. Partial charges are determined via the Restrained Electrostatic Potential (RESP) method, which seeks to reproduce the electrostatic potential felt at grid points outside the molecule's own van Der Waals radius [225, 30]. Atomic partial charges are refined according to least-squares

fitting to most accurately reproduce the field produced from QM calculations. The level of electronic structure theory chosen is ground-state unrestricted density functional theory (DFT), using the Becke 3-parameter Lee-Yang-Parr (B3LYP) hybrid exchange-correlation functional and 6-311G** basis set which includes polarizing d- and p-functions for second-row elements and hydrogens, respectively [226]. RESP calculations were conducted using the Gaussian 16 electronic structure package [227].

The target molecule for RESP is a hexamer of thiophenes with regioregular ethyl side chains. Side chains were truncated from the hexyl groups on P3HT to accelerate calculation convergence, since alkyl groups are not being reparameterized. First, a geometry optimization is performed. Notably, the ground state geometry of P3HT differs in its neutral and cationic states, with the latter having an enhanced propensity for planarity. This phenomenon is well-known for radical cations in conjugated π-systems [228, 229]. Since the partial charges that emerge from RESP are dependent on single-point geometry, RESP was performed for both ground state geometries in both of the two possible charge states for a total of four QM calculations. This produces two sets of point charges for both the neutral and cationic states, and atom-by-atom charge differences δq were taken at each geometry between the two charge states.

In empirical classical force fields, different atoms that have identical bonded environments should have the same underlying parameters. Since the P3HT model here is subtly asymmetrical due to the regioregular side chain configuration, atoms that should have identical charges in the final force field differ slightly at the QM level. To reconcile this, differences δq of atoms belonging to the same "atom type" are averaged. The atom types defined by Moreno were then adjusted by this δq, ensuring that the sum of atomic charges per molecule remained an integer. It is worth noting that while the Moreno force field was parameterized using multiple oligothiophene molecules for a transferable force field that permits building arbitrary chains, the reparameterization presented here applies only to the hexameric oligothiophene, as charge delocalization along the backbone is sensitive to the unbroken conjugation length.

While bond and angle terms are rather stiff and not expected to change substantively upon oxidation, the dihedral terms in a classical force field tend to be much smaller in magnitude and the resulting torsional distribution is sensitive to changes in atomic partial charges. Therefore, the dihedrals must be reparameterized following charge adjustment. This is accomplished by comparing energies from the classical force field against a QM "ground truth", scanning along the dihedral coordinate. The relevant torsions here are the thiophene-thiophene backbone twist and side chain torsion relative to the backbone ring. Because these two degrees of freedom are coupled, QM calculations (at the same level of theory as before) were conducted as 2D scans spanning 180° each and including 7 state points along each axis for a total of 49 conformations. Each degree of freedom scanned exhibits even two-fold symmetry. Since the molecule is composed of repeat units, only one torsion (the inner-most) for each of two dihedral types needed to be considered. Geometries were optimized with an applied restraint to prevent deviation from the dihedral state point specified, and all other degrees of freedom were allowed to relax. Then, the 49 resulting geometries from QM calculations were exported and relaxed according to the MD force field with dihedral coefficients set to zero (while freezing the atoms that constitute the relevant dihedrals). Energy differences between QM and MD

Table 4.1: Summary of atomic partial charges adjusted following RESP employed in DFT minimizations as detailed in the main text. Charges are given in units of e^-. Atom descriptions are taken from the Moreno *et al.*.

Atom description	Neutral Polymer	Positive Polymer
R**C**H2-C_β centrale	-0.045	-0.075
R**C**H2-C_β terminale	-0.060	-0.080
S centrale HR	-0.051	-0.031
S terminale HR	-0.006	0.029
$\mathbf{C}_\alpha$ centrale	-0.0045	0.0355
$\mathbf{C}_\alpha$ terminale	-0.182	-0.127
$\mathbf{C}_\beta$ centrale to R	-0.015	0.005
$\mathbf{C}_\beta$ centrale to H	-0.1475	-0.0975
$\mathbf{C}_\beta$ terminale to R	-0.015	0.025
$\mathbf{C}_\beta$ terminale to H	-0.1475	-0.1175
H thiophene	0.1475	0.1575

Table 4.2: Fitted torsion coefficients (as in Fig 4.4) in kJ/mol as written in the Ryckaert-Bellemans functional form, $V(\phi) = \sum_{n=0}^{5} C_n(\cos(\phi))^n$.

(kJ/mol)	C0	C1	C2	C3	C4	C5
Backbone	37.450	4.099	-85.547	-5.104	41.505	3.008
(kJ/mol)	C0	C1	C2	C3	C4	C5
Side chain	-0.212	-9.936	19.057	37.218	-8.858	-15.880

across all 49 state points were then projected onto a single axis for each dihedral via the formula $F(x_1) = -k_\mathrm{B}T \ln \int \exp(-F(x_1, x_2)/k_\mathrm{B}T)\mathrm{d}x_2$ [230], generating a potential energy curve to be added into the MD model to match the QM energies. Finally, the functional form of this potential in terms of Fourier coefficients (in the Ryckaert-Bellemans representation) was determined by least-squares fitting.

The torsional coefficients are included in Table 2. Compared to the neutral torsions, the backbone penalty for rotation becomes much steeper, and the *trans* basin remains preferable to the *cis*. The side chain torsion differs from that published in Moreno *et al.* in two ways: Firstly, Moreno fitted the side chain dihedral based upon 3-ethylthiophene, a monomer. The parameterization here, on the central unit of an oligomer, introduces substantive steric clash when the side chain is oriented toward its nearest neighboring ring (at the 0° mark). Furthermore, the 180° state is stabilized and 0° destabilized as a result of backbone oxidation, which is verified on dihedral scans of 3-ethylthiophene in the neutral and charged states (Figure 4.11).

With both atomic partial charges and torsions successfully reparameterized, the classical P3HT model has been re-cast in its oxidized state. Characterizing the polymer film in its oxidized state as a homogeneous condensed phase cannot be done without some counter-charge to ensure net neutrality, and so one cannot directly

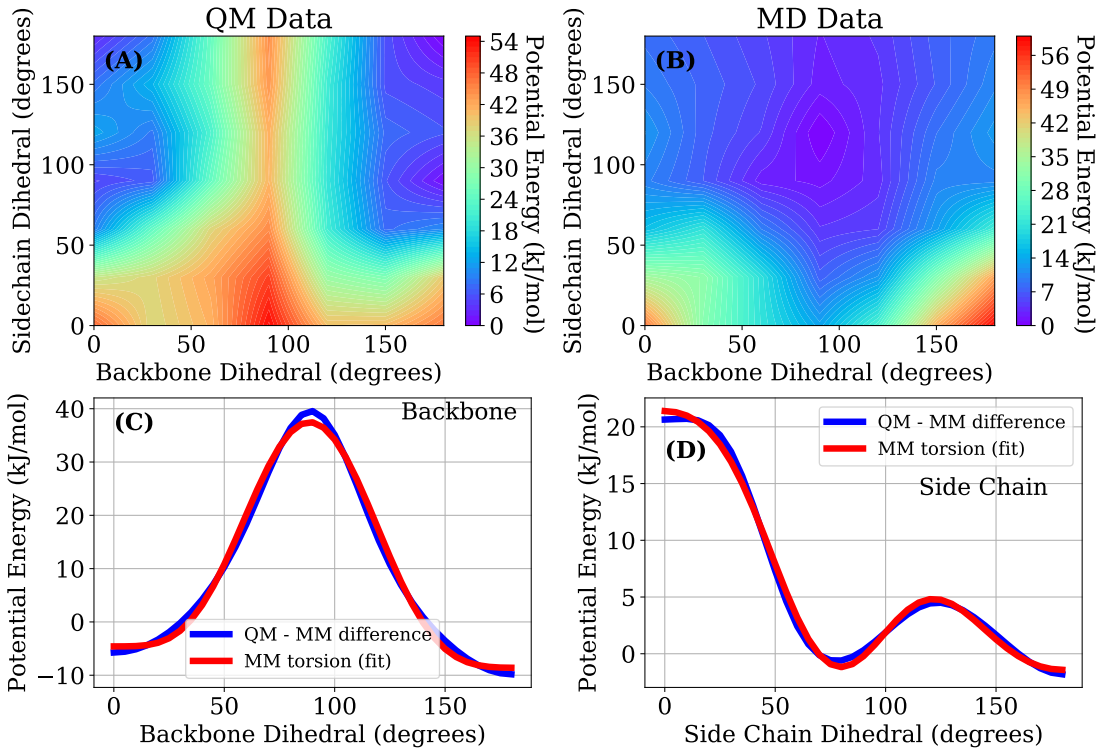

Figure 4.4: Graphical illustration of the fitting procedure. Panels (A) and (B) show 2D total energy plots about 49 state points in the torsion space as computed first from DFT and then those same configurations minimized in the MM force field with the relevant torsions set to zero, as detailed in the main text. In both cases the torsions that define the state point were constrained. Panels (C) and (D) illustrate the projections of the QM - MD energy difference onto a single torsional axis (blue), and the final fitted function (red) that is inserted into the MD engine.

compare distribution functions of the polymer slab in its neutral and charged states. Moreover, a polymer slab with every molecule containing a positive charge corresponds with some of the highest applied potentials seen in literature [194], and is a regime that we avoid. Rather, some molar ratio of charged and neutral polymers are inter-mixed to generate some *average* polymer charge density that is finely tunable to modulate the extent of solvent intercalation. When dihedral distributions are obtained from oxidized oligomers in production runs with neutral polymers and counterions present (as in the next section), differences between the neutral and cationic distributions are in line with expectations, showing an enhanced preference for a planar backbone and the side chains remain in a predominantly 90°conformation (Figure 4.12). Due to the glassy nature of the neutral polymer, large conformational changes are not expected on the time scales simulated, and ensemble sampling is ensured by simulating eight independent replicas of the same systems.

By utilizing a classical force field based upon the well-developed OPLS-AA model, it is also possible to modify the side chain identities to a wide range of chemical functionalizations. One important caveat to doing so is that, if the side chain contains an electronically donating/withdrawing group immediately adjacent to the thiophene backbone, the electronic character of the π-system will be altered and force field modifications will be required even in the case of the neutral polymer. In the present study, this was avoided as a form of control, such that differences between systems with different side chains differ *only* in their side chains, rather than also modifying the backbone parameters. Furthermore, side chain functionalizations that contain labile protons were also avoided, as the chemical environment within the polymer phase that determines the resulting protonation state of each group is unclear. The present work emphasizes aliphatic and ether side chains, due to their prevalence in experimental works and well-behaved nature from a simulation perspective [208, 231].

4.3.3 Measuring Solvent Intercalation

With a model for charged and neutral P3HT on-hand, the first question to be answered is to what extent the oxidized polymers will drive solvent influx. Equilibrated polymer slabs were put into contact with an aqueous bath at 0.5 M K^+/counterion salt concentration and run for 10 ns at 300 K with 1 fs time steps as a pre-run before 45 ns at 300 K with 3 fs time steps as a production run. The ensemble employed was semiisotropic NP_zT, with pressure coupling only in the z dimension, and atmospheric pressure as set by the Berendsen barostat. To achieve a 3 fs time step without destabilizing the system, the Hydrogen Mass Repartitioning (HMR) scheme was employed [58]. In brief, HMR artificially inflates the mass of hydrogen atoms by distributing some mass from adjacent heavy atoms onto the hydrogen, preserving the total mass of the system. By making hydrogen atoms slightly heavier and heavy atoms slightly lighter, the fastest bond oscillations in the system are slowed, allowing for longer time steps and more efficient simulations. While the dynamical properties of molecules like water will differ with and without HMR, trends are still clearly detectable: the bulk diffusivity of ions in solution will be slowed across the board relative to their non-HMR counterparts. For the purposes herein, it is pertinent only to rank the diffusivity of different ions, rather than extract exact values.

Each production run simulation was evaluated thermodynamically using probability density functions of each species along the z-coordinate, normal to the polymer-aqueous interface. These probability density functions serve to provide average molecular occupancy within the polymer phase. Since the particle number N is constant, the number of molecules within the polymer phase may be obtained by integrating the probability density function, accounting for the total number of a given species. The polymer "interior" is defined as the inner-most 3 nm ($\sim$50% polymer peak FWHM) of the polymer phase so as to avoid edge effects. Statistical uncertainty was estimated from the eight independent simulation replicas. From these measurements, both the number of counterions that penetrate the polymer phase and the number of accompanying water molecules per counterion may be evaluated (Fig 4.5). At the low doping levels chosen, volume swelling of the polymer is on the order of 1% and not statistically significant.

The other quantity of interest is determining some characteristic time scale of counterion-polymer interactions, which can be captured with diffusion rates. This observable is complicated by the interfacial nature of the polymer/aqueous composite system, with at least three distinct populations of counterions depending upon their location (bulk aqueous solution, at the interface, and in bulk polymer). Molecular diffusivity from mean-squared-displacement (MSD) is defined in the limit of infinite time, and presents a challenge to measure accurately because a counterion is liable to transition from one region of the simulation to another in finite time, and within the two regions the ion will in general have a different diffusion constant D [232]. Here, it is observed that ions that intercalate into the polymer phase tend to stay there for the remainder of the simulation (>20 ns), which in practice is sufficient to compute D from MSD [233]. Those ions that come and go from the interfacial region to aqueous bulk are not of principal interest, and so an exact report of their diffusion constant is unimportant. Each molecular diffusion constant obtained is assessed in conjunction with the counterion center of mass z-coordinate at the time origin to verify its membership to a group identified as bulk polymer phase. The slope of MSD with time is determined via linear fit from 10% to 90% of the domain along time, as is performed in the GROMACS MSD implementation.

4.3.4 Computing Electron Transfer Rates Classically

The performance of a device that utilizes organic semiconductors is dependent on both the counterion diffusivity in the polymer phase and the electron transfer rate from one π backbone to another. Of particular interest is the way in which a counterion influences electron transfer rate, by proximity energetically favoring some transfers over others.

The polymer phase presents many candidate donors and acceptors at any given time point, with the number of combinations rising like the number of candidate thiophene rings squared, considering all intermolecular ring-ring combinations both forward and reverse. To avoid self-interaction finite size effects, these condensed phase simulations must contain dozens of thiophene oligomers, and so the expense of analyzing ET rates for a single counterion species/side chain combination can be exorbitant at a sophisticated level of theory. Therefore, liberal assumptions must be made in their computations.

The electron transfer rate in question occurs without changes in chemical bonds,

with the two entities remaining intact before, after, and during the transfer. Such a redox reaction, called an "outer sphere" electron transfer (ET) [234], is well-described by the Marcus rate equation [210, 235] so long as electronic coupling is weak (compared to $k_{\mathrm{B}}T$). The resulting equation takes its form from transition state theory [236]:

$$k_{ET} = \frac{H_{\mathrm{DA}}^2}{\hbar}\sqrt{\frac{\pi}{\lambda k_B T}}\exp\left[-\frac{(\Delta A+\lambda)^2}{4\lambda k_B T}\right] \tag{4.1}$$

where $\hbar$ is the reduced Planck constant, H_{DA} is the electronic coupling between donor D and acceptor A (one of which will carry a positive charge at any instant), ΔA is the Helmholtz free energy of reaction, and λ is the so-called reorganization energy which comes from solvent realignment that must occur for the charge transfer to occur. The form of the Marcus rate equation follows from describing the donor and acceptor energetic basins as being parabolic along the reaction coordinate, molecular charging in this case, with the point of parabolic intersection (the transition state energy) being

$$\Delta A^{\ddagger} = (\Delta A+\lambda)^2/4\lambda. \tag{4.2}$$

The exponential prefactor contains the probabilty of an electron jump via Landau-Zener theory [237, 235]. The relevant quantities to be computed are then H_{DA}, λ, and ΔA, each of which are to be computed as an average.

A fundamental quantity in the Marcus formulation which is used to compute λ and ΔA is the energy gap coordinate ΔE, the order parameter defined as the vertical ionization energy between the initial (0) and final (1) states of the redox reaction evaluated for a set of molecular coordinates $\{\mathbf{R}_I\}$ [238]:

$$\Delta E(\{\mathbf{R}_I\}) = E_1(\{\mathbf{R}_I\}) - E_0(\{\mathbf{R}_I\}). \tag{4.3}$$

A given set of molecular coordinates are obtained from a simulation trajectory that samples either the 0 or 1 states, which is indicated on the energy gap coordinate with a subscript, i.e. ΔE_1 indicating an ensemble of states propagated in the product state.

For some coordinate-dependent observable $\mathcal{O}(\{\mathbf{R}_I\})$, its canonical average in the ensemble $M = 0$ or 1 may be defined as [239],

$$\langle\mathcal{O}\rangle_M = \frac{\int \mathcal{O}(\{\mathbf{R}_I\})\exp(-E_M(\{\mathbf{R}_I\})/k_{\mathrm{B}}T)\prod_I \mathrm{d}\mathbf{R}_I}{\int \exp(-E_M(\{\mathbf{R}_I\})/k_{\mathrm{B}}T)\prod_I \mathrm{d}\mathbf{R}_I}. \tag{4.4}$$

Using this notation, the probability distribution of the energy gap taking on a given value X may be expressed in terms of the Dirac delta function as,

$$p_M(X) = \langle\delta(\Delta E(\{\mathbf{R}_I\}) - X)\rangle_M. \tag{4.5}$$

A key assumption in Marcus theory is that this probability distribution is Gaussian along X with an effective spring constant k_M, taking on the form [238],

$$p_M(X) = \sqrt{\frac{k_M}{2\pi k_{\mathrm{B}}T}}\exp[-k_M(X-\Delta E_M)^2/2k_{\mathrm{B}}T], \tag{4.6}$$

where the exponential argument is a harmonic oscillator and the prefactor normalizes

the probability distribution. The oxidation free energy ΔA is defined through a thermodynamic perturbation formula known as the Zwanzig relation [240],

$$\Delta A = -k_{\mathrm{B}}T\ln\langle\exp(-E(\{\mathbf{R}_I\})/k_{\mathrm{B}}T)\rangle_0. \tag{4.7}$$

This relation may be re-written as a cumulant expansion, which for Gaussian probability distributions may be rigorously truncated at the second moment [239], giving

$$\Delta A = \Delta E_0 - \frac{\sigma_0^2}{2k_{\mathrm{B}}T} = \Delta E_1 + \frac{\sigma_1^2}{2k_{\mathrm{B}}T} \tag{4.8}$$

for variance of energy fluctuations in each oxidation state given by

$$\sigma_M^2 = \langle(\Delta E(\{\mathbf{R}_I\}) - \Delta E_M)^2\rangle_M. \tag{4.9}$$

A second key assumption in Marcus theory is that the curvatures of the two parabolic basins are the same as one another, an approximation that is particularly well-grounded in situations where the donor and acceptor species are chemically identical. This approximation means that $k_1 = k_0 = k$ and $\sigma_0^2 = \sigma_1^2 = \sigma^2$. As a result, the expression for ΔA in Eq. 4.8 may be added to itself for $M = 0$ and $M = 1$ and divided by two, removing the dependence on σ^2 by cancellation and leaving

$$\Delta A = \frac{1}{2}(\Delta E_1 + \Delta E_0). \tag{4.10}$$

This is the expression provided most frequently in applications such as constrained density functional theory (CDFT). To understand the effort involved in such a simulation, it is informative to expand ΔE_M in terms of its constituent ensemble-averaged energies, written in a notation where the subscript on E indicates whether the potential energy function describes the electron hole as being on the donor or acceptor molecule, and the brackets indicate the ensemble of states from which the average is taken:

$$\Delta A = (\langle E_1\rangle_1 - \langle E_0\rangle_0) + \frac{1}{2}[(\langle E_1\rangle_0 - \langle E_1\rangle_1) - (\langle E_0\rangle_1 - \langle E_0\rangle_0)]. \tag{4.11}$$

The first parenthetical term can be identified as the total internal energy of the reaction, ΔU, while the latter terms constitute the entropic contribution. It is noteworthy that an ET reaction which incurs little change in system entropy will exhibit a symmetry between $\eta_1 = (\langle E_1\rangle_0 - \langle E_1\rangle_1)$ and $\eta_0 = (\langle E_0\rangle_1 - \langle E_0\rangle_0)$, which cancel, $\eta_1 - \eta_0 = 0$, if $T\Delta S = 0$.

When this quantity is being computed classically, an additional term $\Delta\mathrm{IP}_{\mathrm{DA}}$ also appears as an additive constant, the difference in vacuum ionization potential of the bodies carrying the positive charge [211]. This quantity is computed quantum mechanically and inserted as a correction term for the deficiencies of the classical force field. In the present situation, since the donor and acceptor molecules have identical thiophene backbones, $\Delta\mathrm{IP}_{\mathrm{DA}} \approx 0$. In the case of regiorandom side chain appendages this expression is not exact due to subtle differences in electron localization due to the locations of side chain appendages between molecules, but the difference is expected to be small.

The reorganization energy arises from the separation of time scales between

electrons and nuclei, by which the electron transfer is considered near-instantaneous compared to the time scale of nuclear relaxation [241]. This is combined with the notion that in order for the electron transfer to occur, the nuclei must be positioned in a configuration compatible with the successor complex, which arises as a result of the Franck-Condon principle [242]. That is, one must evaluate the energetic penalty of the nuclear coordinates of the precursor complex "reorganizing" to that of the successor complex *before* the ET reaction has the ability to occur. This reorganization of the surroundings is referred to as an outer-sphere reorganization, and is denoted here as λ_s. The reorganization energy can be identified as related to the difference between the vertical excitation ΔE_0 and free energy of reaction ΔA, written $\lambda = \Delta E_0 - \Delta A$, from the Marcus parabolas. With an expression for ΔA on-hand, λ can be similarly defined in terms of ΔE_M as

$$\lambda_s = \frac{1}{2}(\Delta E_0 - \Delta E_1). \tag{4.12}$$

Similarly to ΔA, it is useful to re-write this expression in terms of the constituent energy terms that must be computed,

$$\lambda = \frac{1}{2}(\langle E_1\rangle_0 - \langle E_0\rangle_0 - \langle E_1\rangle_1 + \langle E_0\rangle_1) = \frac{1}{2}([\langle E_1\rangle_0 - \langle E_1\rangle_1] + [\langle E_0\rangle_1 - \langle E_0\rangle_0]).$$

This can be further simplified in the event that entropic contributions can be neglected. That is, if $\eta_0 = \eta_1 = \eta$, then the average $\lambda_s = (1/2)(\eta_0 + \eta_1) = \eta$. This assumption is often in the literature for applications of Marcus theory to MD simulations, but is an approximation that will not be used here because doing so saves little computational expense when all of the necessary energies are already being computed.

When computed via classical simulations, λ_s is corrected by a scalar factor s_{pol} that accounts for the lack of polarizability in classical force fields. Its necessity has been discussed before in aqueous systems and it tends to be less than unity to compensate for over-estimations of λ_o from empirical simulations [243, 244]. Its value is related to the relative magnitudes of the optic and static dielectric constants ϵ_∞ and ϵ_0, respectively. This relationship is [245, 246]

$$s_{\text{pol}} = (\epsilon_\infty^{-1} - \epsilon_0^{-1})/(1 - \epsilon_0^{-1}). \tag{4.13}$$

In aqueous solution, taking $\epsilon_0 = 78.4$ and $\epsilon_\infty = 1.4 - 2.0$, prior works have given $s_{\text{pol}} = 0.5 - 0.7$ [245, 247, 248, 249]. In P3HT, where $\epsilon_0 = 3.75$ [209] and $\epsilon_\infty \approx 2$ [250, 251], the value of s_{pol} could be as low as 0.25. In the simulations conducted here, with the number of thiophene oligomers outnumbering the number of waters, we err on the lower side with $s_{\text{pol}} = 0.3$.

The reorganization energy should also include an internal contribution from underlying electronic structure changes in the donor and acceptor molecules and their differing geometries [211]. Since MD simulations are incapable of accurately capturing this quantum mechanical phenomenon, the "internal" contribution is computed separately with electronic structure theory (DFT, at the same level of theory as before) from *in vacuo* single point energies E^{QM} of donor and acceptor in both ground state geometries [252]. That is, for an energy $E_{\text{D}}^{\text{QM}}(\text{D})$ computed for the charge state in parenthesis and the optimized geometry of the state in subscript [253],

$$\lambda_i = (E_{\mathrm{D}}^{\mathrm{QM}}(\mathrm{D}^+) - E_{\mathrm{D}^+}^{\mathrm{QM}}(\mathrm{D}^+)) + (E_{\mathrm{A}^+}^{\mathrm{QM}}(\mathrm{A}) - E_{\mathrm{A}}^{\mathrm{QM}}(\mathrm{A})). \tag{4.14}$$

In the case of oligothiophenes, where the ground state geometries of the neutral and cationic states differ (and the latter experiences a strong energetic penalty for deviations from planarity), λ_i is found to be just over 75 kJ/mol, which is larger than what tends to be observed for the external contribution λ_s. The total reorganization energy λ is then given by

$$\lambda = \lambda_s + \lambda_i. \tag{4.15}$$

The final piece, electronic coupling H_{DA}, cannot be reliably estimated from classical atomic coordinates without quantum mechanical benchmarking, and must be evaluated from tabulated data obtained at a higher level of theory. In the case of thiophene (C_4H_4S, as a standalone molecule), the intermolecular transfer integral for two planar bodies has been tabulated and an exponential decay function fitted to data [219]. The resulting function is $C\exp(-\alpha d_{ij})$ with $C = 24.5$ eV, $\alpha = 1.455\text{Å}^{-1}$, and ring-ring separation d_{ij}, which is used to estimate electronic coupling here. In aqueous media a "pathway model" is often applied [254, 255] that permits descriptions of electron transfers over hydrogen bonds as well, but in the present work water molecules do not H-bond with the polymers and so these contributions are neglected. Furthermore, in principle a sinusoidally-dependent term could be appended to account for lack of ring-ring co-planarity between donor and acceptor polymers, but is thought to be unnecessary since those rings at closest contact (where transfers will be most likely) have a strong tendency to indeed be planar, as indicated in Fig 4.3E with $\langle(1/2)(3\cos^2(\theta) - 1)\rangle \approx 1$, and those transfers far enough away to be un-correlated will have poor electronic coupling by any measure.

Initial configurations of 32 polymer molecules, one counterion (BF_4^-, PF_6^-, Cl^-, or $TFSI^-$), and some number of waters (as determined by the prior biphasic simulations) were annealed as described previously to equilibrate polymer configurations and the ion/water cluster. These configurations were initialized eight times for each molecular combination and each replicate was taken as an initial configuration from which to compute electron transfer rates. For each replica, 32 short simulations (500 ps in the NVT ensemble) were conducted, each with a different one of the polymers containing a backbone cation to redirect solvent polarization. These short trajectories correspond to some $\langle E_M\rangle_M$ for an initial (0) or final (1) state corresponding to the electron hole being localized on one polymer or another. Then, each trajectory, with coordinates propagated according to the ensemble $\langle\cdot\rangle_M$, was re-evaluated using a different potential energy function that describes the oxidized polymer as being a different molecule than the trajectory was propagated under. Evaluating the system energy $E_{0/1}$ of simulation frames obtained from sampling ensemble $\langle\cdot\rangle_{1/0}$ provides $\langle E_0\rangle_1$ or $\langle E_1\rangle_0$, which is computed for all 32^2 combinations of donors and acceptors.

For any pair of molecules assigned donor D and acceptor A, electronic coupling H_{DA} is estimated for all individual thiophene ring-ring combinations, which for a pair of hexameric molecules includes 36 evaluations. Since the electron hole is thought to localize to as small a volume as one thiophene unit in order to charge transfer [201, 197], only the top-ranking ring-ring pair was used to estimate H_{DA} between molecules. A matrix of molecule-molecule couplings is produced from the trajectory corresponding to the "final" state of the reaction, which is so to align with the no-

tion of reorganization energies and that the electron transfer will not actually occur until the nuclear coordinates are in-line with the product state. With this, all three of the key Marcus parameters have been established within a short MD simulation and electron transfer rates k_{ET} can be computed. Additional information can be extracted using the Marcus formalism by identifying the dominant contributions to k_{ET}, and thereby better understand the underlying principles (free energy, reorganization energy, or electronic coupling) that result in a desirable value for device performance.

4.4 Results and Discussion

The simulations performed here describe two halves of a related problem: In order for the electron holes to diffuse macroscopically across the polymer phase in an electronic device, the polarons' and counterions' motions are coupled with one another, with their combined diffusivity being slower than the expected diffusion rate of either the polaron or counterion independently. Mechanistically, this coupling is two-sided, in which the diffusion rate of a counterion is tied to the current location of the electron hole, and the electron transfer probability is intrinsically related to the current location of the counterion. The simulations presented here address both sides of this coupling: In the biphasic aqueous/polymer simulations conducted, the diffusion of a counterion when residing in the polymer phase may be measured directly, providing a metric of how tightly associated the counterion is with the electron hole when the electron hole is stationary. Conversely, the electron transfer calculations performed assess the likelihood of the electron hole to shift from one chain to another when in the presence of the counterion. That is, the co-diffusion of the counterion and electron hole together is stepwise, in which the counterion localizes in the vicinity of neutralizing polymer charge, which by Coulombic interactions stabilize the polaron in its current location. There are many hypothetical electron transfers which may occur at any instant, but their likelihood is related to the charge-countercharge separation that is imposed on the system upon a successful transfer. By simulating both sides of this co-diffusion problem, insight is shed on the nature of the counterion-polymer properties necessary for an effective device with high bulk conductivity.

4.4.1 Counterion Intercalation and Confined Diffusion

When driving solvent influx into the polymer phase by increasing the polymer doping levels, the charge per thiophene required to observe measurable intercalation is on the order of 1/8 (16/128) hexameric molecules oxidized, equating to 1 unit of charge per 48 thiophene rings, a doping level of ∼2% or 12.5% of molecules with a positive charge. This is gentle doping for many experiments, and illustrates the capability of counterions and solvent to penetrate into the disordered polymer phase even when large (nm-sized) cavities are not present. This is visualized in probability density distributions, a sampling of which is provided in Fig 4.5 for P3HT with regiorandom side chains in the un-doped and doped states. Additional panels can be found in the Supplementary Information, Figure 4.13.

The number of water molecules per counterion that make their way into the polymer phase tends to be independent of regioregularity in this study and follows

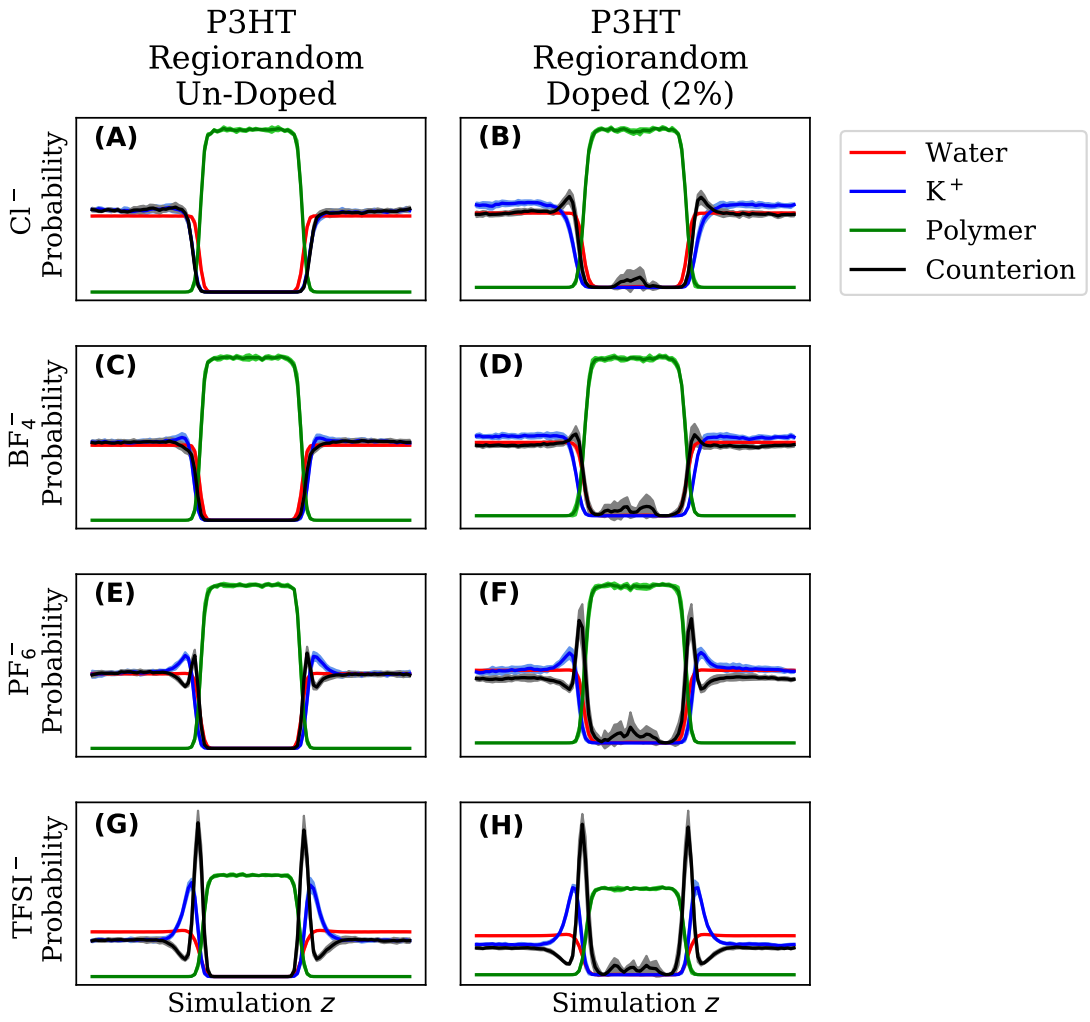

Figure 4.5: Probability density traces for molecular species in each simulation cell. The polymer slab, in green, is at the center of each cell and is flanked on either side by water in red. The cation, K^+, is in blue, while the counterion of interest, denoted on the lefthand-side of each row, is in black. When the polymer phase is un-doped (left column) the polymer and aqueous media remain phase-separated, with some propensity from $TFSI^-$ and PF_6^- to aggregate at the interface, accompanied by some compensating cations. Following doping (right column, at a level of 0.02 positive charge per thiophene ring), all four counterion species intercalate into the polymer phase. The probability density of water in these instances is nonzero, but is fractionally small compared to the aqueous bulk. Upon doping, both BF_4^- and to a lesser extent Cl^- also show some minor propensity to aggregate at the surface.

expected trends in terms of relative hydrophilicity (Fig 4.6A). The lack of distinction between regioregular and regiorandom morphologies likely arises from the highly disordered state simulated, in which structural metrics such as $g(r)$ are hardly distinguishable. The number of total counterions that penetrates the polymer phase tends to be greater for regiorandom side chains than regioregular, but only when hexyl side chains are present. With ether side chains this difference vanishes, which is likely due to increases in the accompanying water molecules due to the more hydrophilic environment (Fig 4.6C).

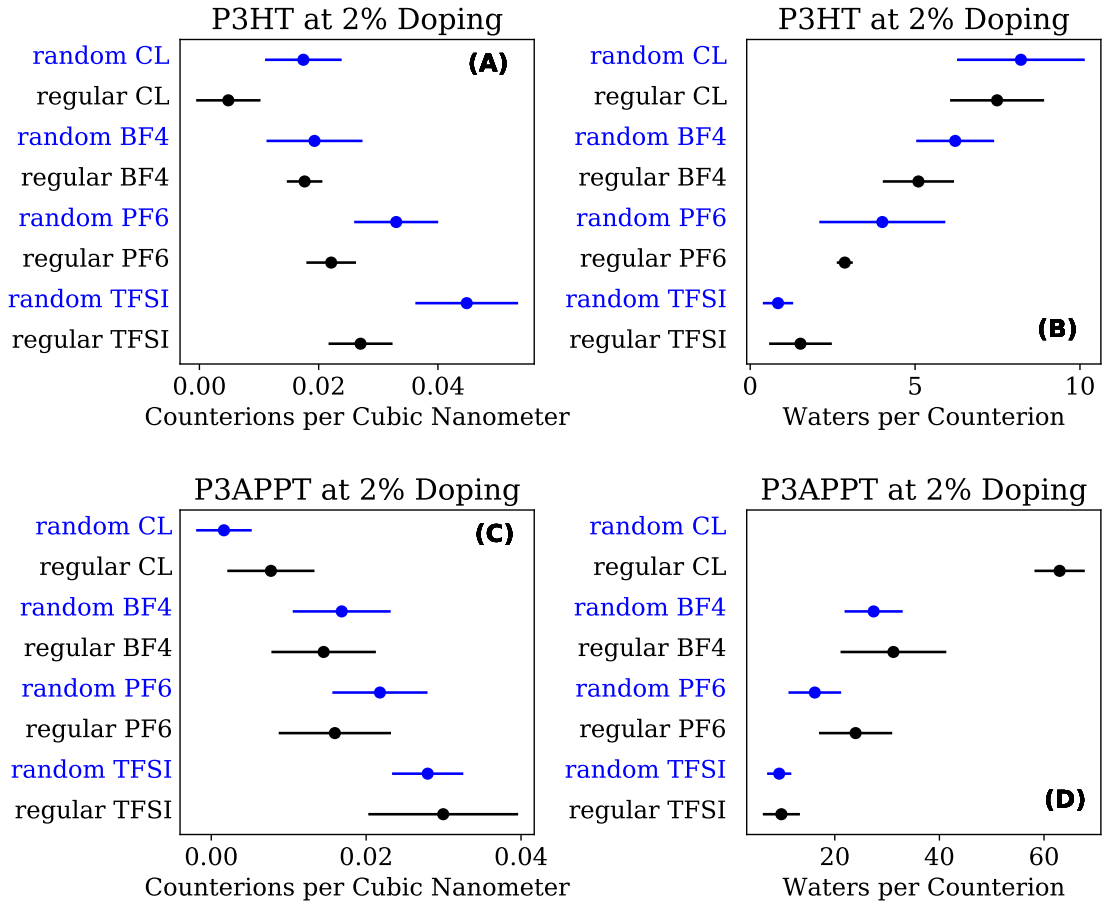

Figure 4.6: Molecular counts within the polymer phase as determined by integrating probability densities over the inner-most 3 nm of the polymer phase averaged over 8 simulation replicas. The left column, (A,C), shows the number of counterions that are within the polymer phase, while the right column (B,D) shows the number of water molecules per counterion present. In panel (D), the water molecules per chloride in the case of regiorandom side chains is omitted because the computed value has error bars larger than the quantity itself, and is off the scale compared to the others.

It is clear that the anion TFSI has a high affinity for the polymer phase due to its hydrophobic nature. This manifests itself both with un-doped films, in which case the TFSI ion aggregates at the aqueous-polymer interface to shield its hydrophobic CF_3 groups from water, and can also be seen when the polymer phase has been doped, in which case the number of TFSI ions that enter the polymer phase tends to be greater than the other species considered (Fig 4.6A). The relatively hydrophobic

nature of TFSI means that it brings relatively few water molecules with it into the polymer phase.

At the other end of the spectrum is Cl^-, which is the most hydrophilic ion studied having a relatively strong interaction energy with water. As a result, at low doping levels where the polymer phase remains rather hydrophobic, few counterions enter the polymer phase and when such penetrance does occur, the volume swelling observed is largely due to the accompanying water molecules, which may be present alongside chloride at greater than a 50:1 ratio (Fig 4.6D). The counterions BF_4 and PF_6 fall in an intermediate regime, again with an anti-correlation between the number of counterions that penetrate the polymer phase and both the number of water molecules that accompany them.

Counterion mobility in the polymer phase was assessed via diffusion constants computed via Mean Square Displacement (MSD), with control simulations containing no polymer chains as a benchmark for aqueous mobilities. For each 3 ns subdivision of the simulation, counterions were assigned membership to the set of molecules in the polymer phase if at time $t = 0$ the counterion center of mass was within the polymer phase, using the same spatial criteria as previously. In aqueous bulk, counterion mobilities are influenced by their place on the Hofmeister series. That is, those counterions that reside on the more chaotropic "structure-breaking" end of the Hofmeister ranking tend to increase the viscocity of their aqueous solutions [256], which intuitively correlates with overall diffusivity. Additional influences on molecular diffusion include total mass and the number of degrees of freedom. That is, complex molecular structures with many degrees of freedom can absorb energy from molecular collisions as rotations and oscillations rather than center of mass motion. While the molecules rank on the Hofmeister series (in terms of increasing chaotropism) as $Cl^- < BF_4^- < PF_6^- < TFSI^-$, molecular mass and number of chemical bonds increase with the same trend. As a result, the observed diffusion constants of the four counterions in aqueous bulk are very near one another, with little apparent trend and $TFSI^-$ having the slowest motions in aqueous bulk. The magnitudes observed align with the diffusion constant of the solvent, TIP3P water (with HMR), $\sim 5 * 10^{-5}$ cm^2/s.

The counterions' diffusion constants decrease substantively while residing in the polymer phase, changing more than an order of magnitude (to $\sim 10^{-6}$ cm^2/s) but still remaining faster than the polymers themselves (at $\sim 10^{-8}$ cm^2/s). In the P3HT phase, regioregular or regiorandom, the mobility of counterion species do not differ in a statistically significant way. In P3APPT, the same is true with the possible exception of Cl^-. The number of chloride anions in the polymer phase is low compared to the other counterions and so the number of observations are fewer, contributing to statistical uncertainty. It is interesting that the number of water molecules per counterion increases from P3HT to P3APPT side chains, which from the counterion's perspective makes the polymer phase appear more aqueous-like, and yet counterions in P3APPT appear to have slower mobilities than those in P3HT (where statistically significant). This is attributed to the polymers' diffusivity decreasing with ether side chains compared to alkyl and an enhanced side chain-counterion interaction ability. At the doping levels used here, the number of water molecules present is insufficient to truly emulate the aqueous bulk.

The surprising takeaway from these counterion diffusion studies is that the classical MD model applied here does not differentiate the ions' mobilities well either in

aqueous medium or the polymer phase. Experimentally, the electron hole/counterion pair diffuse with orders of magnitude different rates than one another, and these MD data suggest that these differences cannot be simply explained by counterion diffusion rates about a stationary electron hole. However, counterion diffusion is substantively slowed in the polymer phase compared to aqueous, and the proportion of counterions in the two phases could underpin the observed experimental results.

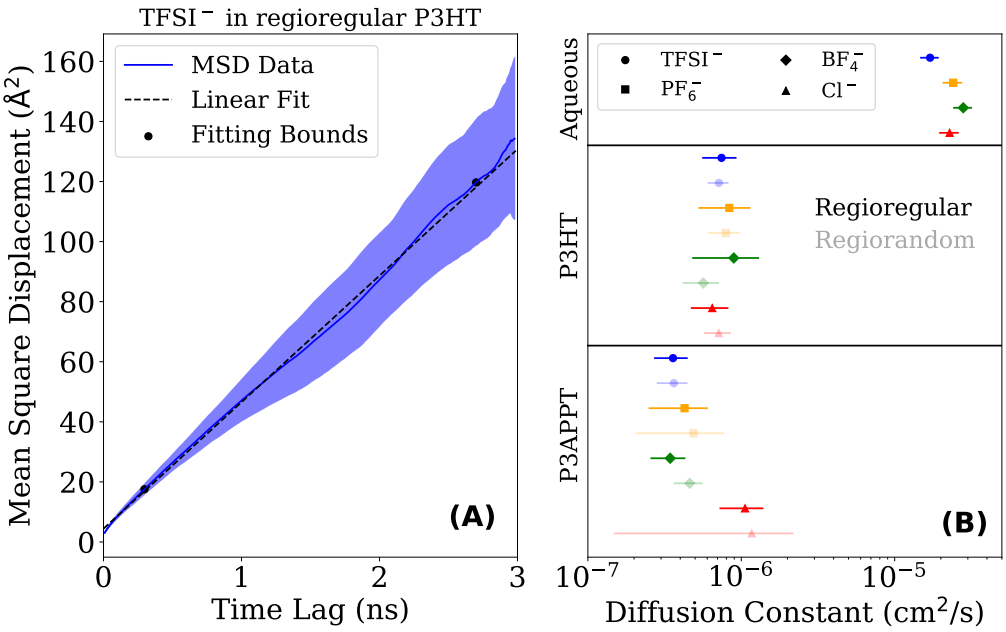

Figure 4.7: (A) Demonstration of an MSD plot for a counterion ($TFSI^-$) within the polymer phase (regioregular P3HT, doped at 0.02 positive charge per thiophene ring) to illustrate the goodness of fit and demonstrate the regular diffusion behavior. (B) Diffusion constants as obtained for each type of counterion in the polymer phase of P3HT, P3APPT, or in aqueous bulk. Counterions have distinctive markers ($TFSI^-$ in blue, circles; PF_6^- in orange, squares; BF_4^- in green, diamonds; Cl^- in red, triangles) and regiorandom side chain conditions are indicated by transparency. All diffusion constants were computed at the 2% doping level.

4.4.2 Electron Transfers

The "mobility" of an electron hole is assessed in two ways: its *maximum* electron transfer rate, the singular most probable reaction, and the *breadth* of available transfers. While the former is enticing to focus on because it is the most likely to be observed, the single most favorable ET reaction tends to be one that moves the electron hole to a polymer immediately adjacent to the counterion. This motion is not especially productive, since the electron hole is thus confined to a proximity around the counterion, which itself has a limited diffusion rate in the confines of the bulk polymer phase. The latter metric of electron hole diffusion, measuring the breadth of the distribution of favorable electron transfers, is perhaps of greater interest (so long as the maximum rate does not substantively suffer) because it illustrates the capability of the electron hole to move independently of the counterion. If the cation-anion pair remain tightly associated at all times, and any move to separate them is penalized, then diffusion of the pair will be hindered.

Interestingly, the rate of electron transfer is rather unaffected by either the identity of a counterion or polymer regioregularity, as measured by either maximum transfer rate or by sorting the top 10 transfers. This is observed in Figure 4.8, in which the electron transfer rate k_{ET} (obtained from the Marcus rate equation, Eq. 4.1) is computed between every pair of polymers for a single initial molecular configuration and then ranked greatest to least and plotted. While deviations are observed replica to replica, with independent initial conditions, the spread is nearly identical across counterion intercalants and side chain regioregularities for a given polymer side chain chemistry. On average, the ether side chain chemistry on P3APPT tends to result in slower ET rates than the alkyl side chains on P3HT, although with the variance observed from replica to replica there is still substantive overlap in their distributions. To understand why these observations are being made, it is pertinent to deconstruct the top-ranking ET rates for each trajectory and evaluate the components of Marcus Theory (free energy, reorganization energy, and electronic coupling) that contribute to reaction favorability.

Considering that there are hundreds of possible electron transfers in each system here, a distribution of free energies, reorganization energies, and electronic couplings are observed. For free energy and reorganization energy, these distributions are approximately Gaussian in nature, centered about zero for reaction free energy and reorganization energies are shifted in the positive energy direction due to the internal contribution $\lambda_i > 0$ in the overall reorganization energy $\lambda = \lambda_s + \lambda_i$. In each case, the top-rated ET rates in terms of k_{ET} magnitude correspond to ΔA values that tend to fall on the lower-energy half of the distribution curve, but are not necessarily the most extremely favorable (Figure 4.9). This is because the resulting k_{ET} is also dependent on electronic coupling H_{DA}, which across all possible electron transfers shows a roughly exponentially decaying probability distribution with the maximum observed coupling $\sim$0.1 eV. When the values of H_{DA} that correspond to the top-ranked ET reaction for each replica are highlighted along the distribution of all couplings, a wide spread is observed (see Supplementary Information, Figure 4.17), illustrating that no one component of the Marcus rate equation seems to dominate the resulting rate. That is, the top-ranked ET reactions must achieve a compromise between the energetics of ET and electronic coupling.

It is possible to convert ET rates from Marcus theory into effective diffusion constants in organic crystals via the relation [257]

$$D = \frac{1}{6}\langle k_{ij} r_{ij}^2 \rangle. \tag{4.16}$$

Averaging over all possible transfers, the reaction rates k_{ET} reported convert to diffusion constants on the order of 10^{-5} cm^2/s. This is comparable to the diffusion of liquid water, and much faster than the diffusion of the polymer or counterion molecules themselves. This would suggest that counterion diffusion is rate-limiting in the electron hole/counterion pair mobility.

Remarkably, all of the ion species and side chain regioregularities analyzed (for a given side chain identity) displayed the same behavior: Not only did they show the same maximum ET rate k_{ET} and the same distribution among the top-ranked ET reactions, the analysis of these top reactions in terms of contributions from ΔA, λ, and H_{DA} were also indistinguishable. This observation is surprising in light of experimental observations that counterion identity and regioregularity have profound

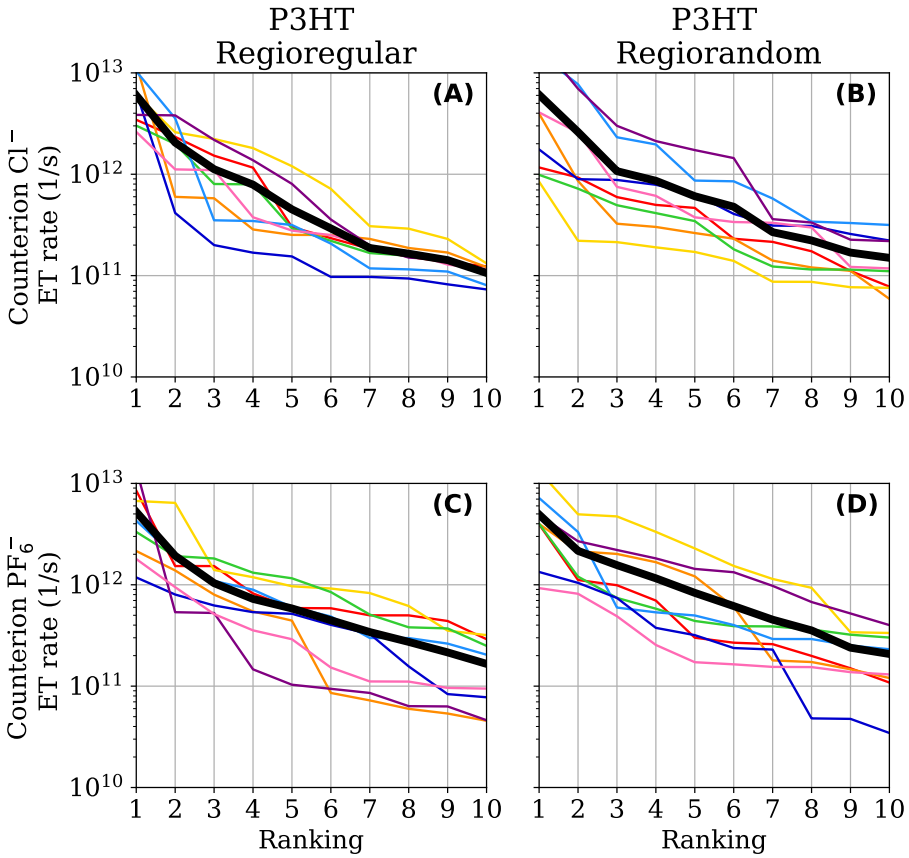

Figure 4.8: Distribution of electron transfer (ET) rates for select counterion species and both regioregular (left column) and regiorandom (right column) polymers with hexyl side chains. Of the hundreds of ET reactions for any given molecular configuration, the top 10 greatest magnitude ET rates in units of 1/s are ranked in descending order from 1 to 10. Each thin colored line corresponds with an independent replicate of the calculation, while the bold black line is the average of the eight replicates. ET rates are statistically indistinguishable in each case, both in terms of maximum rate and in distribution of the secondary rates.

impacts on the resulting coupled electron transfer rates. In as far as this coarse model is to be trusted to describe the quantum mechanical ET process using classical simulations, these data suggest that the influence of an individual counterion on the polymers' ET rate is rather insensitive to the exact nature of the counterion, but rather it is the mere presence of a counter-charge that permits a diversity of ET reactions in the vicinity of neutralization. If this is true, it would suggest that it is the number of counterions that penetrates the polymer phase that is more important for facilitating charge transfer than it is specific ion-polymer interactions that promote ETs.

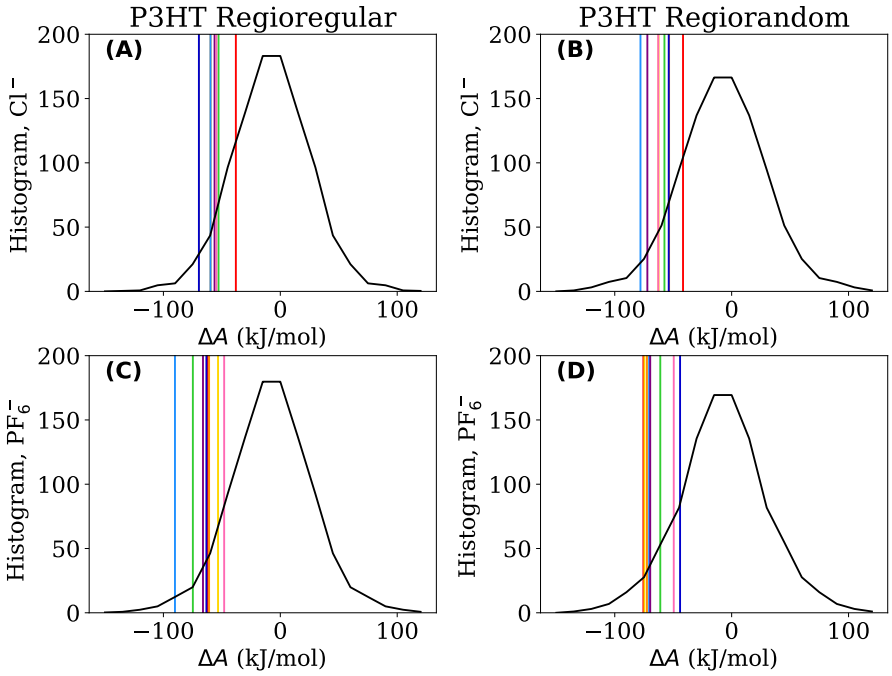

Figure 4.9: Of all inter-polymer ET reaction free energies ΔA measured, which are histogrammed in black, those reactions that correspond to a top-rated reaction (giving the greatest magnitude of k_{ET}) are marked by vertial lines, one for each replica. Select results compare the same counterions and P3HT regioregularities as Fig 4.8, with regioregular (left column), regiorandom (right column), and two counterions Cl^- (top row) and PF_6^- (bottom row). The ET reactions that occur most quickly in terms of k_{ET} for each case are not the most energetically favorable in terms of ΔA, because a compromise must be made with other terms in the Marcus rate equation, reorganization energy λ and electronic coupling H_{DA}.

This conclusion regarding the inability of a counterion to influence the ET rate between polymer chains is one that is made in the "dilute limit" – that is, only a single counterion is present with a single electron hole among 192 thiophene rings. At this low doping level, cooperative effects between multiple charge reorganizations cannot be discerned. Indeed these may be pertinent at higher doping levels, as even single chains of P3HT are known to take on bipolarons [258, 259], a doubly cationic state along a single backbone. This is in addition to polaron-polaron effects from nearby chains. Such effects become increasingly difficult to assess in simulation due to the extra layer of combinatoric complexity this places on the number of states to consider: For each inter-chain ET reaction already considered, they must be

performed in replicate for each instance of a second oxidized polymer. It is for this reason that we do not investigate beyond the dilute limit presented here.

The picture presented from this data is incomplete, because the polymer phase in question is highly disordered and therefore the differences between regioregular and regiorandom side chains are subtle. In more crystalline polymer phases, the regioregular side chain functionalities are highly interdigitated and more well-ordered than their regiorandom counterparts [214], and so are expected to display differences in quantities like reorganization energy and ring-ring coupling strength. As reflected in the characterization of the bulk polymer in this work from Fig 4.3, it is clear that the disordered nature of the polymer slab results in similar structural character between the two regioregularities, which explains the lack of differences observed in ET rate distributions. The polymer phase used in this study is only representative of a fraction of the polymer bulk in experiments, which contain crystalline and disordered regions, along with vacancies that flood with aqueous solution [231]. Further studies involving these other phases are warranted.

4.4.3 Potential for Model Improvements

The present work has endeavored to capture ion-polymer and ion-mediated polymer-polymer interactions via a classical two-body force field, which relies on a number of approximations in order to achieve the high-throughput necessary to identify trends among molecular configurations. While trends are ultimately what this work sought to produce, it is worthwhile to consider further modeling advancements to increase the accuracy of the predictions made. For example, the model at hand does not incorporate molecular polarizability, which is of great importance when delocalized electrons have an energetic driving force to transfer. For instance, the Coulombic interaction between a polymer and an adjacent cation is not accurately captured in this model, since the electron hole is liable to shift in response to a counter-charge. The lack of polarizability manifests itself in water molecule and *p*-block atoms as well, albeit to a lesser extent. While polarizable water and ion models have been proposed previously, achieving such an effect with the polymers is a much more challenging feat. While many polarizable atomic models, for instance crystalline silver [260], employ movable psuedo-atoms within the atom's van Der Waals radius to permit anisotropic charge distributions (via the Drude model [261]), such minute adjustments in atomic partial charge may not be capable of capturing the (comparatively) large fractional changes necessary to capture a unit e^- charge localizing on different places of the polymer backbone. As a result, capturing polarizability in P3HT is best achieved through quantum-mechanics-based methods, which for the system size being considered might be as detailed as the DFT level of electronic structure theory. Typically scaling with $\mathcal{O}(N_{e^-}^3)$ for number of electrons N_{e^-} [226], a Kohn-Sham approach scales among the best of electronic structure theory methods, but for each molecule containing more than 1,000 electrons each propagated time step is prohibitively lengthy.

Less expensive models that still incorporate multipolar electrons are semiempirical models such as the recently published GFN2-xTB functional [262]. By incorporating density-dependent dispersion contributions, electron localization along the polymer backbone could be easily captured, and improvements in recent years to the empirical training set have vastly improved the torsional distributions of conjugated

organic molecules such as P3HT. By still using an atom-centered minimal valence basis set of contracted Gaussian functions and parameterizing only on an element-by-element basis (rather than explicitly parameterizing any pairwise interactions), modern semiempirical methods like GFN2-xTB remain somewhat inexpensive (built for linear scaling [263]) while also building in multipolar electrostatic interactions and exchange-correlation effects (up to second order in the multipolar expansion). Such an intermediate approach, using inexpensive parameterizations like classical MD but incorporating a Kohn-Sham theoretical basis, provides a potentially attractive alternative approach to probe P3HT systems of many atoms.

Ultimately these methodologies are problematic for systems at hand, however, in which the disordered polymer phase is highly heterogeneous and clean crystal lattices cannot be taken advantage of. The number of atoms in the system, at least tens of thousands, are prohibitive to obtaining molecular diffusion constants with any sort of statistical certainty. Even for electron transfer reactions, which are inherently a quantum mechanical phenomenon, the present system with its high degree of disorder presents a manifold of available electron transfer reactions, many of which are potentially relevant to a given molecular configuration. Even if the liberties are taken to reduce computational cost, such as using short oligomers of P3HT as in the present work, the number of molecules that must be considered in the ET problem leads to combinatorically many possible electron transfers. It is therefore imperative that the model used to describe these processes scales well enough that all of the computations can be made, even if sacrificing in accuracy.

4.5 Conclusion

The present work has endeavored to model the intercalation of aqueous ions into disordered oligothiophene derivatives for the purposes of understanding the polaron/counterion pair diffusion rates that are observed experimentally, which is complicated by a tight coupling between the two species' mobilities. To better understand the rate-limiting steps in this collective motion, we have modified a classical empirical force field that has been well-benchmarked to model the disordered polymer phase to include a delocalized electron hole along the backbone, thereby providing a driving force for ion intercalation into an otherwise hydrophobic phase that prior theoretical works have established remains phase segregated from aqueous when no doping has been performed. With this new model, we observe the extent of counterion intercalation stoichiometrically, in terms of the number of counterions that penetrate the polymer phase and the number of water molecules that accompany each. These counterions are also characterized kinetically, measuring diffusion constants when in and out of the polymer phase to assess the ability of polymers to localize different counterion species. Finally, borrowing methodology from the biochemical community, we assess the rates of Marcus electron transfer between different polymer chains via classical simulation and tabulated quantum mechanical data, to assess the ability to each counterion to localize the electron hole in its vicinity.

Simulated findings from counterion diffusion studies and electron hole transfer studies agree in that neither detects specific counterion-dependent interactions that have direct ramifications on charge mobility. Exchanging one counterion species for another is not found to change the diffusion constant of the counterion in the

polymer phase nor change the distribution of ET reactions that occur adjacent to the counterion. However, the very presence of a counterion, any counterion (of those studied here), does have an impact on observable rates: Counterions in the polymer phase are slowed by more than an order of magnitude compared to aqueous bulk, and the most favorable ET reactions occur in such a way that charge is localized near the counterion. The key difference between counterion treatments, then, is the number of counterions that penetrate the polymer phase. The hydrophobic ion $TFSI^-$ intercalates most readily, with BF_4^- and PF_6^- rather comparable, and Cl^- penetrates the least. This trend both aligns with published schematics that represent Cl^- as being averse to intercalation, and agrees with trends in macroscopic counterion/hole mobility in moving front experiments. Under this premise, the most successful devices will be those that allow the greatest extent of polymer intercalation into the polymer phase, namely those with sizeable hydrophobic domains to drive influx with polymers like P3HT. The same can be said for more hydrophilic polymers like P3APPT due to the hydrophobic thiophene backbone, although the trend is somewhat diminished as water enters the polymer phase much more readily.

Further studies are warranted to confirm this finding, as the system studied here uses classical fixed point charge models to describe highly polarizable molecules, and the disordered morphology of the polymer phase is not representative of the whole polymer bulk used experimentally. This work does illustrate, however, that even for hydrophobic polymers like P3HT the influx of aqueous counterions can be measured in simulation at low doping levels, even with non-polarizable thiophenes in dense configurations. Such findings are sufficient premise to further investigate the counterion / electron hole pair mobility using multiple charged sites to better reflect experimental doping levels and explicitly measure counterion diffusion and hole mobility in tandem.

4.6 Supplementary Information

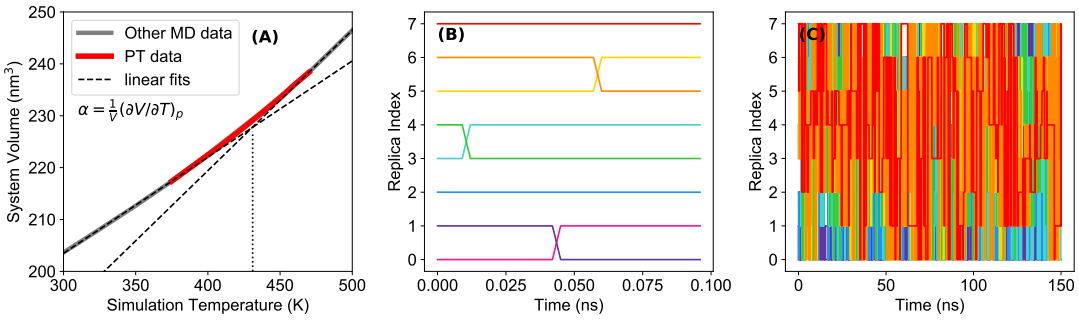

Figure 4.10: Demonstration of applying Parallel Tempering to obtain system volume as a function of temperature. (A) The end result of the calculation provides a thoroughly-sampled volume/temperature relationship in the vicinity of the glass transition point, T_g (dotted line), which is well above 300 K. (B) Zoom-in of simulation replicas, indexed 1 through 16 (with temperatures assigned according to an exponential distribution, $T_i = (370\text{K})\exp(0.02 * i)$ from $i = 0$ to $i = 15$), illustrating that as time proceeds, the coordinates of one simulation replica exchange with an adjacent replica. Panel (C) shows the same data, but along the whole simulation trajectory, illustrating that over time replicas will thoroughly exchange and explore all temperature settings.

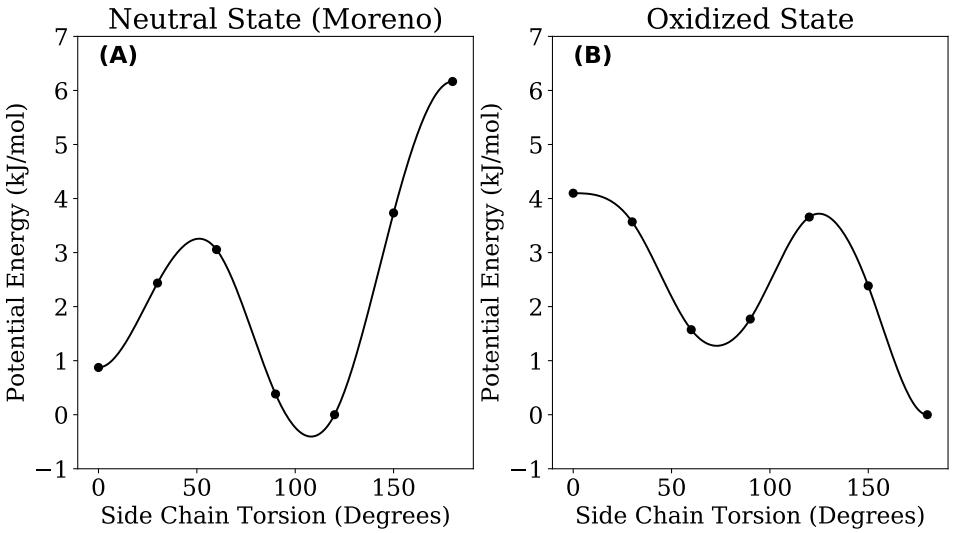

Figure 4.11: The side chain dihedral scan performed by Moreno was done on the molecule 3-ethylthiophene, a thiophene monomer which does not capture the steric effects of ratating the side chain in the presence of an adjacent linked thiophene ring. To rationalize the dihedral distribution obtained in the main text, in which a multidimensional dihedral scan was performed on an oligothiophene, the study of Moreno *et al* is reproduced here using DFT and B3LYP 6-311G** as in the main text. A dihedral scan about the 3-ethylthiophene side chain in the neutral (A) and cationic (B) states are shown, with 0° corresponding to the ethyl CH_3 pointing in the direction of the thiophene alpha carbon to which a neighboring thiophene would be attached in an oligomer. In the neutral state, 3-ethylthiophene shows an energetic minimum at 0°, while in the cationic state a change in electron localization on the thiophene ring makes this same position an energetic maximum. This difference, observed for a monomeric thiophene, illustrates that the presence of a maximum at 0° in the cationic state is not due to the presence of a neighboring thiophene ring. In the main text, where parameterization is performed on a thiophene oligomer, the 0°maxima is even more pronounced due to steric hindrance with the neighboring ring.

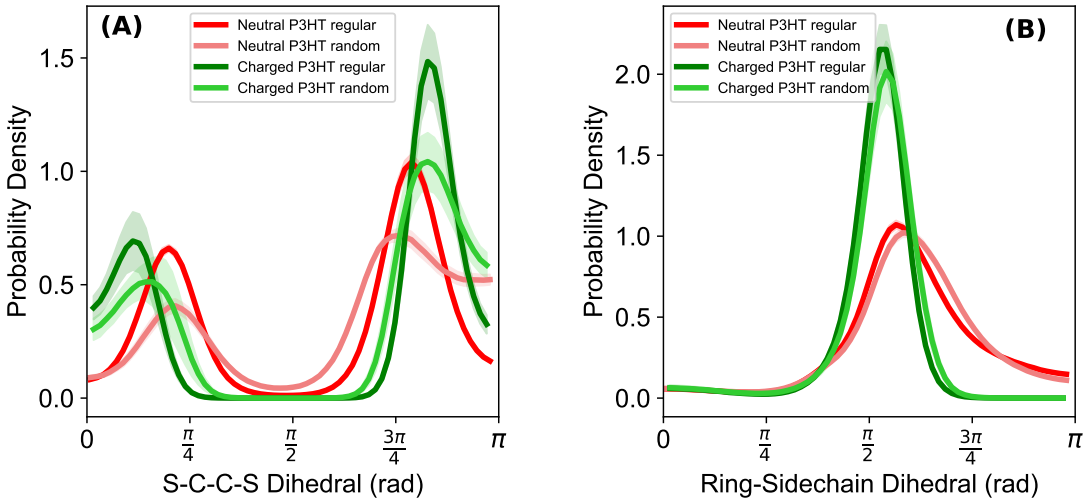

Figure 4.12: Dihedral characterization of charged oligothiophenes compared to their neutral counterparts, validating the parameterization performed here. (A) The backbone S-C-C-S dihedral is shifted in the directions of 0 or 180 ° in accordance with an increased barrier to breaking planarity. (B) The side chain dihedral between the backbone thiophene and the first two carbons of the side chain (where 0 ° points toward an adjacent thiophene ring and 180 ° points toward the neighboring hydrogen atom) is also narrower in distribution, reflecting the steeper energetic barrier now present.

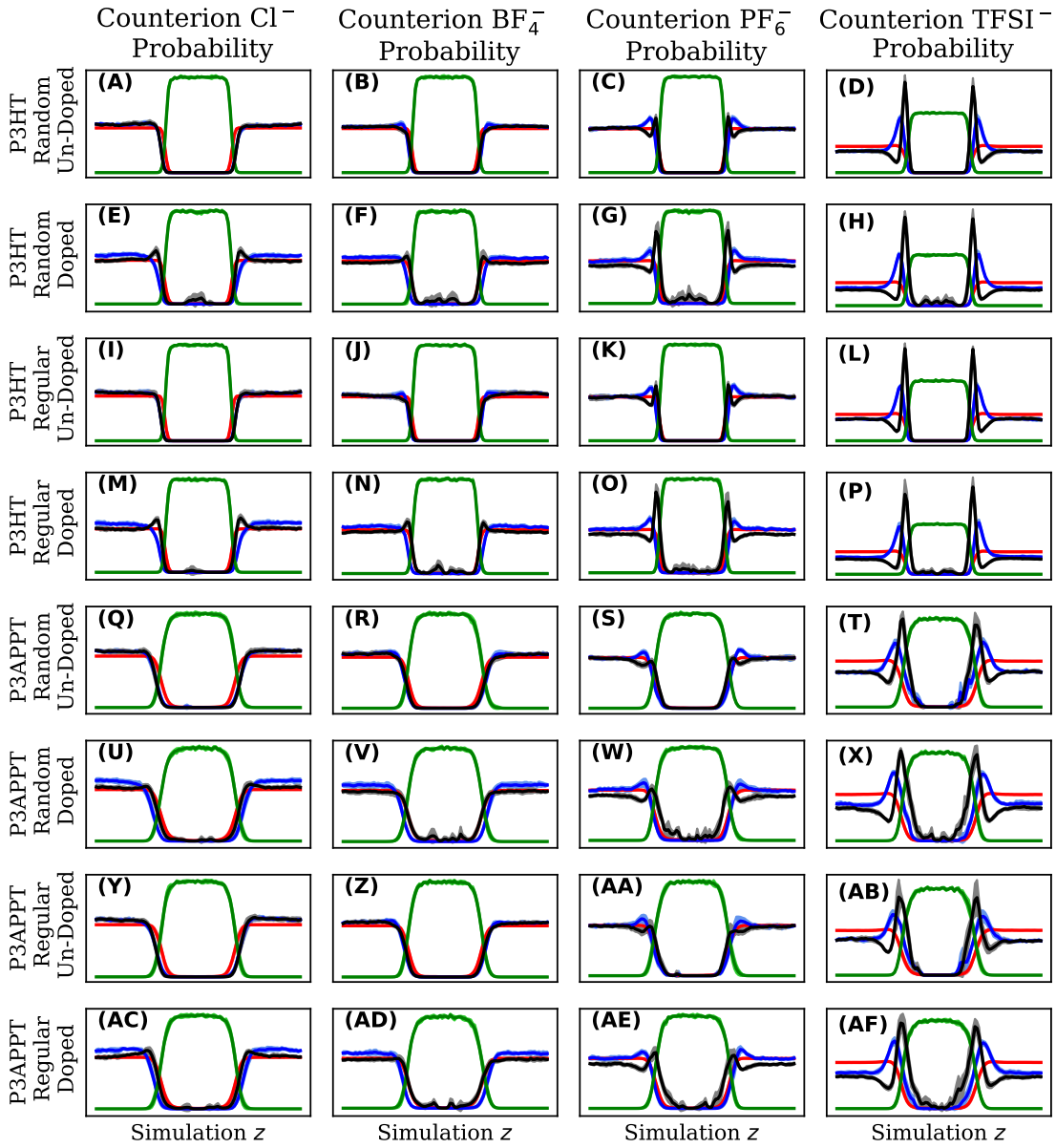

Figure 4.13: Density plots for all combinations of counterions, polymer side chains, and regioregularities, for un-doped polymer (all polymers electrically neutral) and doped at the level of 16/128 hexamers, or ∼2% of thiophene rings carrying a positive charge. Columns are Cl^-, BF_4^-, PF_6^-, and $TFSI^-$, while rows are changing the polymer identity, doping level, or side chain regioregularity as marked.

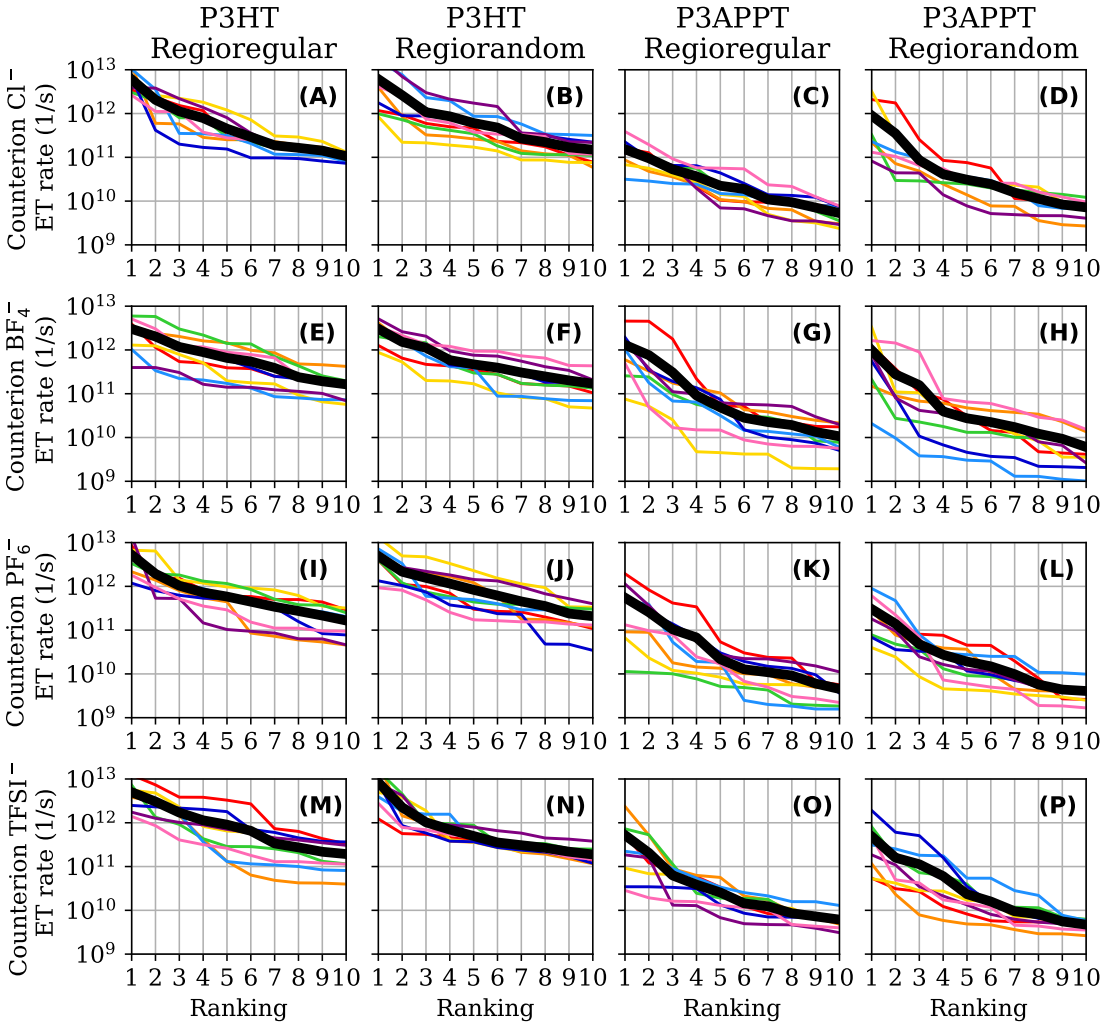

Figure 4.14: For all possible ET reactions from a given initial condition, ET rates (in units of 1/s) are ranked from greatest to least and the top 10 displayed here in descending order. Each row corresponds to a different counterion, $TFSI^-$, PF_6^-, BF_4^-, or Cl^-, while each column corresponds to a different polymer setup, either varying regioregularity or side chain identity. Thin colored lines are different replicas of the same system with independent initial conditions, while the bold black line is an average over the eight replicas. Each row corresponds to a different counterion identity, while each column indicates a different polymer setup, with different side chain regioregularity of side chain functionality.

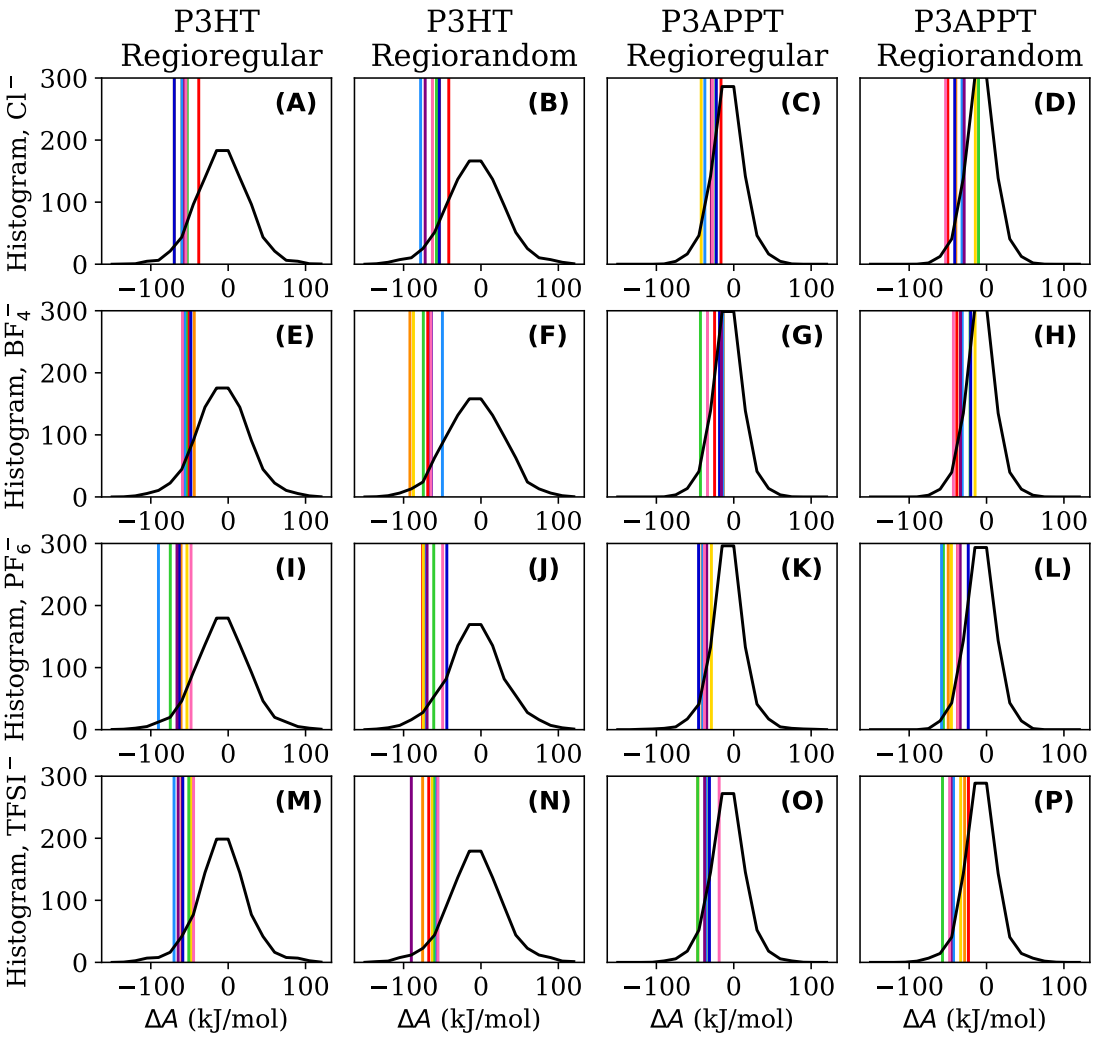

Figure 4.15: For all possible ET reactions from a given initial condition, the free energy of reaction (in kJ/mol) is histogrammed, summing over replicas, and displayed in a black trace as an un-normalized probability density. Eight vertical lines mark the reaction free energy of the top-rated ET reaction (that with the greatest rate) from each of eight replicas to illustrate that the most favorable ET reactions for each replica do not exhibit the lowest free energies of all possible reactions. Each row corresponds to a different counterion identity, while each column indicates a different polymer setup, with different side chain regioregularity of side chain functionality.

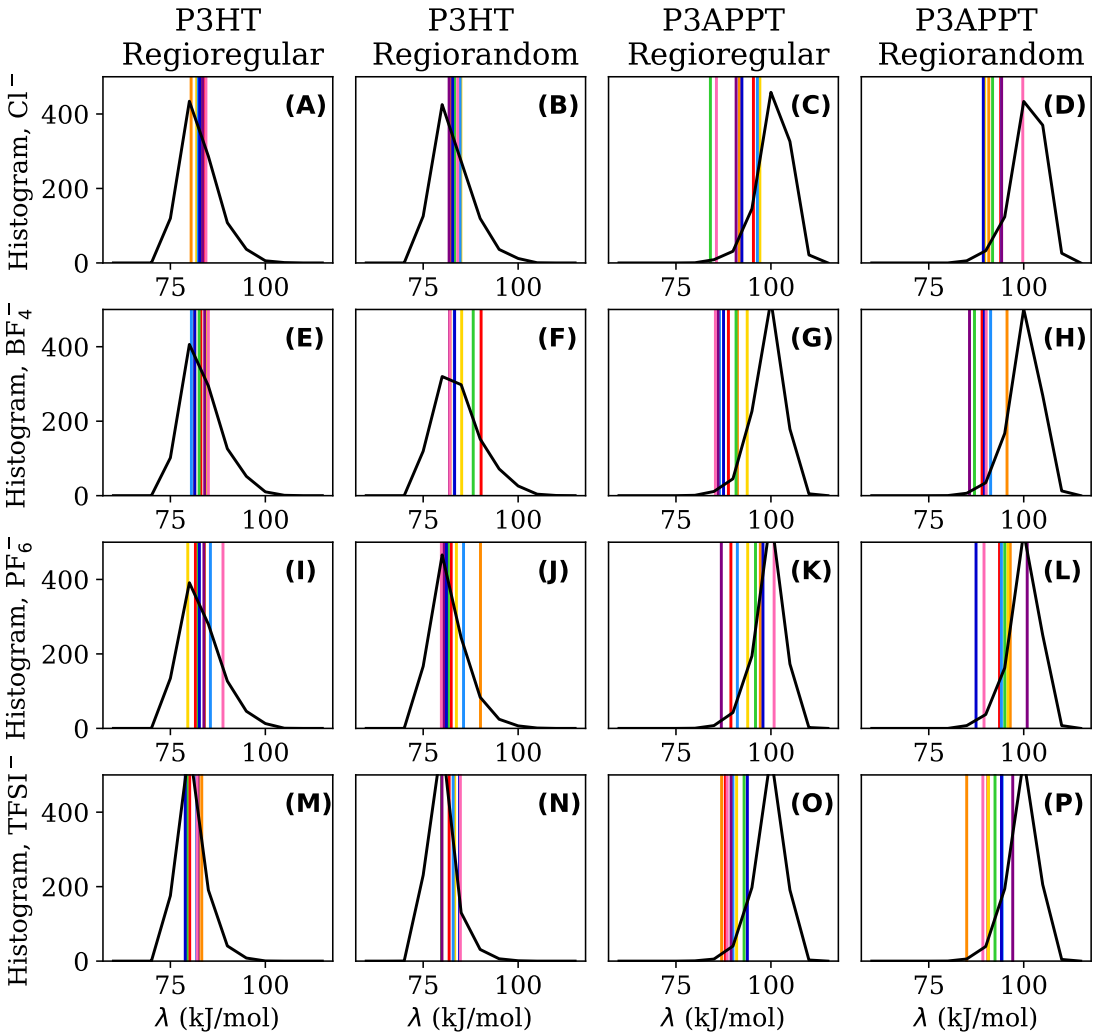

Figure 4.16: For all possible ET reactions from a given initial condition, the corresponding reorganization energy (in kJ/mol) is histogrammed, summing over replicas, and displayed in a black trace as an un-normalized probability density. Eight vertical lines mark the reaction free energy of the top-rated ET reaction (that with the greatest rate) from each of eight replicas to illustrate that the most favorable ET reactions for each replica do not exhibit the lowest reorganization energies of all observed reactions. Each row corresponds to a different counterion identity, while each column indicates a different polymer setup, with different side chain regioregularity of side chain functionality.

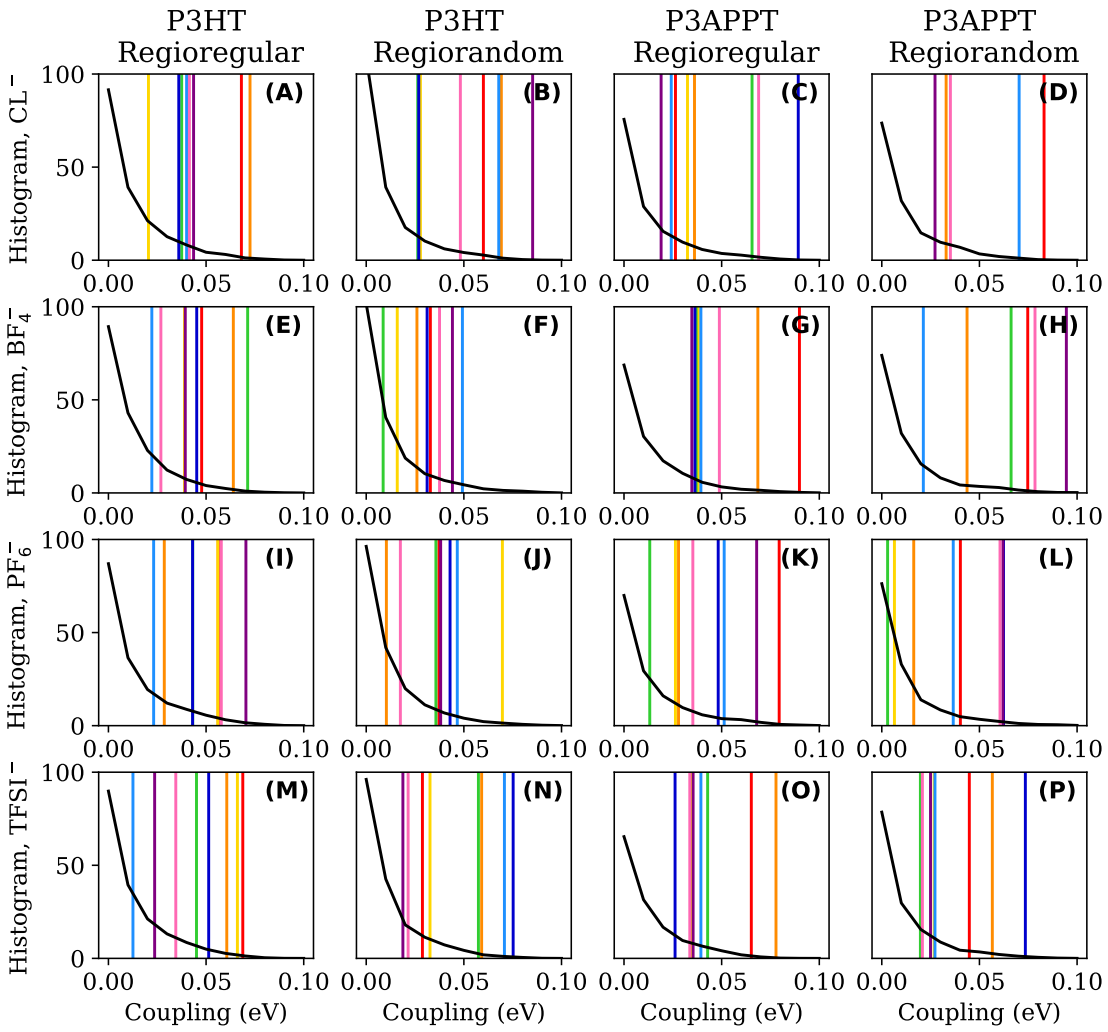

Figure 4.17: For all possible ET reactions from a given initial condition, the electronic coupling (in eV) is histogrammed, summing over replicas, and displayed in a black trace as an un-normalized probability density. Eight vertical lines mark the reaction free energy of the top-rated ET reaction (that with the greatest rate) from each of eight replicas to illustrate that the most favorable ET reactions for each replica exhibit a breadth of electronic couplings. Each row corresponds to a different counterion identity, while each column indicates a different polymer setup, with different side chain regioregularity of side chain functionality.

Chapter 5

Conclusion

This body of work has examined three case studies of aqueous interfaces and the ramifications of including different ion species on interfacial phenomena such as molecular orientation and relative affinities. These phenomena occur on the length scale of nanometers, and so can be challenging to capture with *ab initio* techniques that scale poorly with the number of electrons, but are also below experimentally resolvable resolutions and so are inaccessible to optical techniques or even sophisticated surface force apparatuses. MD is an excellent tool for accessing systems on the scale of many thousands of atoms, because it is computationally inexpensive enough to sample many configurations of the system for a rigorous treatment via statistical mechanics. Particularly in soft matter systems where many configurations are accessible, the ensemble properties measured macroscopically are only predicted reliably from an atomistic scale when thorough sampling of the relevant degrees of freedom is accomplished.

The first project presented here, "Decoupling of the Water-Ion Density Fields Adjacent to Molecular-Sized Cavities", belabors thorough solution sampling of an aqueous-hard wall interface via an Umbrella Sampling scheme. Technical challenges are addressed with respect to calculating free energies of a hard wall system in a framework where potential energies must be differentiable, relying on the notion that two distinct ensembles (the hard-wall system of interest and a soft-walled system that was sampled) can be related to one another by clever manipulations of probability. The curved hard-wall aqueous interface considered, studied via Umbrella Sampling with two-dimensional WHAM, revealed that the local density fields adjacent to the cavity wall undergo qualitative changes at different levels of curvature depending on the molecular species considered. While water molecules retain their close-packed hydrogen bonded network around small radius (< 0.5 nm) cavities, the density fields of ions are highly sensitive to the underlying curvature and begin to deplete from the interface at different cavity radii. Such ion depletion is known to occur in the limit of a planar interface (an infinitely large spherical cavity), and a similar phenomenon is observed here for cavity sizes that water molecules respond to as being "small". Conversely, in the limit of very small cavities, Cl^- is found to accumulate at the interface, and these findings are relevant to understanding the generic problem of small molecule solvation, especially hydrophobic groups. While some anion adsorption is known to occur near a water liquid-vapor planar interface, water molecules in the small cavity limit have not yet formed such a liquid-vapor coexistence. Furthermore, the model used here to ascertain this adsorption behav-

ior does not allow for explicit ion polarization, which in the planar limit is known to diminish adsorption tendencies. This presents the ion adsorption and depletion behavior found near small cavities as distinct from adsorption to the planar interface, and is likely relevant to simulation studies both with and without polarizability explicitly included. In an addendum, a novel umbrella sampling alignment scheme is proposed based upon minimization of residuals, and its performance is found to be comparable to the standard algorithm, WHAM. Ultimately, the newly developed method did not out-shine the tried-and-true WHAM for the application at hand, but is potentially useful for circumventing the curse of dimensionality. In particular, the least-squares fitting procedure suggested is generally applicable to simulations where many one-dimensional biases are placed on the many bodies in the system, and may be advantageous in situations where the dimensionality of the biasing problem exceeds 2.

The second project discussed here, "Ion-Dependent Protein-Surface Interactions from Intrinsic Solvent Response", considers the role that the aqueous environment has on the long-range attraction between two solvated planar bodies. Spatial patterning of highly-organized molecular building blocks polarize solvent, giving rise to an electrostatic signal that permeates far into bulk solution. This signal is measured atomistically via MD simulations that were carefully catered to match experimental observations of self-assembling protein-mineral systems, and the far-field component of the polarization was isolated via a perturbation theory-like method called Local Molecular Field theory. By relating far-field polarization to a Landau energy density expression, the polarization-induced pressure between two parallel planar bodies can be estimated, indicating attractive or repulsive forces depending upon solvent interference patterns arising from the chemical nature of the planar interface. Particularly, we found that different adsorption tendencies of Na^+ and K^+ to the muscovite surface results in two qualitatively different far-field electrostatic responses that we attribute to the presence or lack of a well-ordered electric double layer. This has ramifications for face-specific protein-mineral attractions, and suggests that exchanging K^+ for Na^+ diminishes far-field orientational preferentiality and ultimately gives rise to a less well-ordered bound layer of protein. The theoretical platform presented is found to give quantitatively comparable results to other far-field theories such as DLVO (Derjaguin, Landau, Verwey, and Overbeek) reported in literature but has the advantage that key model parameters such as a characteristic decay length are determined directly from molecular simulation. This is in contrast to directly solving Poisson-Boltzmann equation because the necessary surface potentials are not easily obtained from a molecular model. Furthermore, our LMF-Landau interaction theory reveals the exquisite sensitivity of polarization-induced interactions on the electrostatic screening length, which can result in several orders of magnitude difference in attractive strength. This points to the necessity of developing more accurate (but still computationally inexpensive) atomistic models including features like atomic polarizability in order to accurately predict the nature of many-body interactions.

Finally, in the third project discussed, "Coupled Charge Carrier / Counterion Mobility in Organic Electrochemical Cells", a soft-matter aqueous interface is presented in the context of organic semiconductors with an emphasis on understanding their tunable properties as a consequence of solvent intercalation. Of particular interest is the co-diffusion of electron holes along the semiconducting polymers'

backbone and the aqueous counterions that intercalate into the polymer film for net neutrality. To maximize pair diffusivity and improve the performance of electronic devices that utilize these materials, we attempt to identify attributes of polymer/counterion pairings that are rate-limiting for bulk transport. The problem was approached from two sides: We assessed the ability of stationary electron holes in the polymer phase to localize counterions in their vicinity and conversely analyzed the ability of a localized counterion to steer which electron transfer reactions are favorable between polymers. This project necessitated force field modification to drive the solvent influx that facilitates charge transfer experimentally. With these parameter modifications, counterion intercalation into the polymer phase has been achieved, an advance on previous works, permitting experimental observables such as volume swelling and water molecules per counterion in the polymer phase to be directly observed in simulation. Furthermore, being able to observe the local environment of a counterion while inserted into the polymer phase permits local dynamics to be probed, providing key insight into the polaron/counterion co-diffusion problem. Then, using mole ratios of polymer / water / counterions from solvent influx simulations, a separate set of simulations were conducted to estimate the rates of electron transfer between pairs of polymers in the presence of a single counterion in a classical framework. Interestingly, counterion diffusion while in the polymer phase was found to be largely independent of counterion species, a trend that held for two different polymer side chains and different side chain attachment patterns. Additionally, the identity of a counterion failed to produce significant differences in the distributions of electron transfer rates. While individual counterion molecules behave similarly to one another in their interactions with the polymer, it was observed that the total number of counterions that intercalate into the polymer phase differs substantively depending on the species, and it is likely this attribute that gives rise to the experimentally observed differences in pair mobility. This work suggests that to develop more successful devices, one should maximize the number of counterions in the polymer phase, with the caveat that the influx of water does not become inhibitory. This is in alignment with experimental observations that hydrophobic counterions have good bulk charge transfer properties, because the counterions do not need water accompaniments to penetrate the hydrophobic polymer phase.

In all three of the case studies presented, there are opportunities for model advancement: (1) In the study of cavity formation in water, polarizable force fields and careful comparisons of different water models would lead to a more complete view; (2) In the development of LMF-Landau theory for macromolecule-surface interactions, describing the hydration of highly-charged interfaces at low salinity would be improved by considering the self-ionization of water; and (3) The study of counterion-polymer interactions via classical simulation neglects important quantum mechanical effects pertaining to electronic coupling. In each case, the considered improvements require substantively more computational effort to implement, with only modest gains to be made. The classical molecular dynamics framework has been immensely successful in describing a breadth of soft-matter systems, as benchmarked against available experiments and high-level electronic structure calculations. While the models omit some real-world physics, the framework permits the study of complex systems that may be otherwise inaccessible. In the words of statistician George Box, “All models are wrong, but some are useful” [264].

Bibliography

[1] N. I. Zheludev. "What diffraction limit?" In: *Nature Materials* 7.6 (2008), pp. 420–422.

[2] M. Heilemann. "Fluorescence microscopy beyond the diffraction limit". In: *Journal of Biotechnology* 149.4 (2010), pp. 243–251.

[3] D. K. Gramotnev and S. I. Bozhevolnyi. "Plasmonics beyond the diffraction limit". In: *Nature Photonics* 4.2 (2010), pp. 83–91.

[4] M. L. M. L. Klein and W. Shinoda. "Large-scale molecular dynamics simulations of self-assembling systems". In: *Science* 321.5890 (2008), pp. 798–800.

[5] K. Kadau, T. C. Germann, and P. S. Lomdahl. "Large-scale molecular-dynamics simulation of 19 billion particles". In: *International Journal of Modern Physics C* 15.1 (2004), pp. 193–201.

[6] R. C. Tolman. *The Principles of Statistical Mechanics*. Courier Corporation, 1979.

[7] F. Reif. *Fundamentals of Statistical and Thermal Physics*. Waveland Press, 1998.

[8] C. Kittel, P. McEuen, and P. McEuen. *Introduction to olid state physics*. New York: Wiley, 1996.

[9] R. Zwanzig. *Nonequilibrium Statistical Mechanics*. Oxford University Press, 2001.

[10] G. Gallavotti and E. G. D. Cohen. "Dynamical ensembles in nonequilibrium statistical mechanics". In: *Physical Review Letters* 74.14 (1995), p. 2694.

[11] H. B. Callen. *Thermodynamics and an Introduction to Thermostatistics*. Wiley, 1998.

[12] W. Feller. *An introduction to probability theory and its applications, vol 2*. John Wiley and Sons, 2008.

[13] J. C. Mauro, P. K. Gupta, and R. J. Loucks. "Continuously broken ergodicity". In: *The Journal of Chemical Physics* 126.18 (2007), p. 184511.

[14] F. Borgonovi et al. "Broken ergodicity in classically chaotic spin systems". In: *Journal of Statistical Physics* 116.5 (2004), pp. 1435–1447.

[15] Y. Wang et al. "Evidence for broken ergodicity in strain glass". In: *Physical Review B* 76.13 (2007), p. 132201.

[16] R. J. Baxter. *Exactly solved models in statistical mechanics*. Elsevier, 2016.

[17] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives. "Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids". In: *Journal of the American Chemical Society* 118.45 (1996), pp. 11225–11236.

[18] B. R. Brooks et al. "CHARMM: a program for macromolecular energy, minimization, and dynamics calculations". In: *Journal of Computational Chemistry* 4.2 (1983), pp. 187–217.

[19] J. Wang et al. "Development and testing of a general amber force field". In: *Journal of Computational Chemistry* 25.9 (2004), pp. 1157–1174.

[20] L. Monticelli and D. P. Tieleman. *Force fields for classical molecular dynamics*. Springer, 2013.

[21] C. Vega et al. "What ice can teach us about water interactions: a critical comparison of the performance of different water models". In: *Faraday Discussions* 141 (2009), pp. 251–276.

[22] J. R. Espinosa et al. "Homogeneous ice nucleation evaluated for several water models". In: *The Journal of Chemical Physics* 141.18 (2014), p. 18C529.

[23] C. J. Tainter et al. "Robust three-body water simulation model". In: *The Journal of Chemical Physics* 134.18 (2011), p. 184501.

[24] J. Pérez-Ríos et al. "Comparison of classical and quantal calculations of helium three-body recombination". In: *The Journal of Chemical Physics* 140.4 (2014), p. 044307.

[25] R. J. Woods and R. Chappelle. "Restrained electrostatic potential atomic partial charges for condensed-phase simulations of carbohydrates". In: *Journal of Molecular Structure: THEOCHEM* 527.1-3 (2000), pp. 149–156.

[26] D. X. Zhao et al. "Development of a polarizable force field using multiple fluctuating charges per atom". In: *Journal of Chemical Theory and Computation* 6.3 (2010), pp. 795–804.

[27] F. Martin and H. Zipse. "Charge distribution in the water molecule—a comparison of methods". In: *Journal of Computational Chemistry* 26.1 (2005), pp. 97–105.

[28] R. S. Mulliken. "Electronic population analysis on LCAO–MO molecular wave functions. I." In: *The Journal of Chemical Physics* 23.10 (1955), pp. 1833–1840.

[29] J. S. Gomez-Jeria. "An empirical way to correct some drawbacks of mulliken population analysis". In: *Journal of the Chilean Chemical Society* 54.4 (2009), pp. 482–485.

[30] C. I. Bayly et al. "A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model". In: *The Journal of Physical Chemistry* 97.40 (1993), pp. 10269–10280.

[31] R. M. Costa Filho et al. "Morse potential derived from first principles". In: *Europhysics Letters* 10.1 (2013), p. 10009.

[32] D. C. Ghosh et al. "The wave mechanical evaluation of the absolute radii of atoms". In: *Journal of Molecular Structure: THEOCHEM* 865.1-3 (2008), pp. 60–67.

[33] A. D. MacKerell Jr and M. Karplus. "Importance of attractive van der Waals contribution in empirical energy function models for the heat of vaporization of polar liquids". In: *The Journal of Physical Chemistry* 95.25 (1991), pp. 10559–10560.

[34] D. Yin and A. D. MacKerell Jr. "Combined ab initio/empirical approach for optimization of Lennard–Jones parameters". In: *Journal of Computational Chemistry* 19.3 (1998), pp. 334–348.

[35] R. D. Levine. "Representation of one-dimensional motion in a Morse potential by a quadratic Hamiltonian". In: *Chemical Physics Letters* 95.2 (1983), pp. 87–90.

[36] D. M. Ferguson. "Parameterization and evaluation of a flexible water model". In: *Journal of Computational Chemistry* 16.4 (1995), pp. 501–511.

[37] M. Reppert and A. Tokmakoff. "Electrostatic frequency shifts in amide I vibrational spectra: Direct parameterization against experiment". In: *The Journal of Chemical Physics* 138.13 (2013), 04B603.

[38] M. Woińska et al. "Hydrogen atoms can be located accurately and precisely by x-ray crystallography". In: *Science Advances* 2.5 (2016), e1600192.

[39] M. Moreno et al. "Molecular modeling of crystalline alkylthiophene oligomers and polymers". In: *The Journal of Physical Chemistry B* 114.4 (2010), pp. 1591–1602.

[40] J. C. Phillips et al. "Scalable molecular dynamics with NAMD". In: *Journal of Computational Chemistry* 26.16 (2005), pp. 1781–1802.

[41] H. Fujitani et al. "High-level ab initio calculations to improve protein backbone dihedral parameters". In: *Journal of Chemical Theory and Computation* 5.4 (2009), pp. 1155–1165.

[42] G. Marc and W. G. McMillan. "The virial theorem". In: *Advances in Chemical Physics* 58 (1985), pp. 209–361.

[43] E. de Miguel and G. Jackson. "The nature of the calculation of the pressure in molecular simulations of continuous models from volume perturbations". In: *The Journal of Chemical Physics* 125.16 (2006), p. 164109.

[44] P. J. H unenberger. "Thermostat algorithms for molecular dynamics simulations". In: *Advanced Computer Simulation* (2005), pp. 105–149.

[45] G. J. Martyna, D. J. Tobias, and M. L. Klein. "Constant pressure molecular dynamics algorithms". In: *The Journal of Chemical Physics* 101.5 (1994), pp. 4177–4189.

[46] S. Alavi. *Molecular simulations: fundamentals and practice*. John Wiley and Sons, 2020.

[47] M. Heier et al. "Equation of state for the Lennard-Jones truncated and shifted fluid with a cut-off radius of 2.5 σ based on perturbation theory and its applications to interfacial thermodynamics". In: *Molecular Physics* 116.15-16 (2018), pp. 2083–2094.

[48] J. Vrabec et al. "Comprehensive study of the vapour–liquid coexistence of the truncated and shifted Lennard-Jones fluid including planar and spherical interface properties". In: *Molecular Physics* 104.9 (2006), pp. 1509–1527.

[49] M. Saito. "Molecular dynamics simulations of proteins in solution: artifacts caused by the cutoff approximation". In: *The Journal of Chemical Physics* 101.5 (1994), pp. 4055–4061.

[50] O. N. De Souza and R. L. Ornstein. "Effect of periodic box size on aqueous molecular dynamics simulation of a DNA dodecamer with particle-mesh Ewald method". In: *Biophysical Journal* 72.6 (1997), pp. 2395–2397.

[51] T. Darden, D. York, and L. Pedersen. "Particle mesh Ewald: An N log (N) method for Ewald sums in large systems". In: *The Journal of Chemical Physics* 98.12 (1993), pp. 10089–10092.

[52] H. G. Petersen. "Accuracy and efficiency of the particle mesh Ewald method". In: *The Journal of Chemical Physics* 103.9 (1995), pp. 3668–3679.

[53] R. D. Engle, R. D. Skeel, and M. Drees. "Monitoring energy drift with shadow Hamiltonians". In: *Journal of Computational Physics* 206.2 (2005), pp. 432–452.

[54] R. D. Skeel and D. J. Hardy. "Practical construction of modified Hamiltonians". In: *SIAM Journal on Scientific Computing* 23.4 (2001), pp. 1172–1188.

[55] A. K. Mazur. "Hierarchy of fast motions in protein dynamics". In: *The Journal of Physical Chemistry B* 102.2 (1998), pp. 473–479.

[56] C. D. Christ, A. E. Mark, and W. F. Van Gunsteren. "Basic ingredients of free energy calculations: a review". In: *Journal of Computational Chemistry* 31.8 (2010), pp. 1569–1582.

[57] B. Li. *Error Propagation of Verlet Algorithm*. Tech. rep. University of California, Berkeley, 2014.

[58] C. W. Hopkins et al. "Long-time-step molecular dynamics through hydrogen mass repartitioning". In: *Journal of Chemical Theory and Computation* 11.4 (2015), pp. 1864–1874.

[59] Glenn Torrie and John Valleau. "Nonphysical sampling sistributions in Monte Carlo free-energy estimation: Umbrella sampling". In: *Journal of Computational Physics* 23 (1977), pp. 187–199.

[60] A. Barducci, M. Bonomi, and M. Parrinello. "Metadynamics". In: *Wiley Interdisciplinary Reviews: Computational Molecular Science* 1.5 (2011), pp. 826–843.

[61] J. F. Dama, M. Parrinello, and G. A. Voth. "Well-tempered metadynamics converges asymptotically". In: *Physical Review Letters* 112.24 (2014), p. 240602.

[62] M. Bonomi, A. Barducci, and M. Parrinello. "Reconstructing the equilibrium Boltzmann distribution from well-tempered metadynamics". In: *Journal of Computational Chemistry* 30.11 (2009), pp. 1615–1621.

[63] J. Pfaendtner and M. Bonomi. "Efficient sampling of high-dimensional free-energy landscapes with parallel bias metadynamics". In: *Journal of Chemical Theory and Computation* 11.11 (2015), pp. 5062–5067.

[64] Y. Sugita and Y. Okamoto. "Replica-exchange molecular dynamics method for protein folding". In: *Chemical Physics Letters* 314.1-2 (1999), pp. 141–151.

[65] L. S. Stelzl and G. Hummer. "Kinetics from replica exchange molecular dynamics simulations". In: *Journal of Chemical Theory and Computation* 13.8 (2017), pp. 3927–3935.

[66] B. H. Mahan. "Microscopic reversibility and detailed balance. An analysis". In: *Journal of Chemical Education* 52.5 (1975), p. 299.

[67] K. Hukushima and K. Nemoto. "Exchange Monte Carlo method and application to spin glass simulations". In: *Journal of the Physical Society of Japan* 65.6 (1996), pp. 1604–1608.

[68] K. A. Dill et al. "Modeling water, the hydrophobic effect, and ion solvation". In: *Annual Review of Biophysics and Biomolecular Structure* 34 (2005), pp. 173–199.

[69] P. L. Geissler. "Water interfaces, solvation, and spectroscopy". In: *Annual Review of Physical Chemistry* 64 (2013), pp. 317–337.

[70] N. N. Rajput et al. *Elucidating solvation structures for rational design of multivalent electrolytes—a review*. Springer, 2018, pp. 79–124.

[71] David Chandler. "Interfaces and the driving force of hydrophobic assembly". In: *Nature* 437.7059 (2005), pp. 640–647.

[72] L. R. Pratt and D. Chandler. "Theory of the hydrophobic effect". In: *Proceedings of the National Academy of Sciences* 67.8 (1977), pp. 3683–3704.

[73] Joyce Noah-Vanhoucke and Phillip Geissler. "On the fluctuations that drive small ions toward, and away from, interfaces between polar liquids and their vapors". In: *Proceedings of the National Academy of Sciences* 8 (2009), pp. 15125–15130.

[74] Yan Levin, Alexandre Dos Santos, and Alexandre Diehl. "Ions at the air-water interface: An end to a hundred-year-old mystery?" In: *Physics Review Letters* 103.257802 (2009), pp. 1–4.

[75] Pavel Jungwirth and Douglas Tobias. "Ions at the air/water interface". In: *Journal of Physical Chemistry B* 106.25 (2002), pp. 6361–6373.

[76] Lars Onsager. "Deviations from Ohm's Law in weak electrolytes". In: *Journal of Chemical Physics* 2 (1934), p. 599.

[77] Pavel Jungwirth and Douglas Tobias. "Specific Ion Effects at the Air/Water Interface". In: *Chemical Reviews* 106.4 (2006), pp. 1259–1281.

[78] David Huang and David Chandler. "Cavity formation and the drying transition in the Lennard-Jones fluid". In: *Physical Reviews E* 61.2 (2000), pp. 1501–1506.

[79] S. Rajamani, T. M. Truskett, and S. Garde. "Hydrophobic hydration from small to large lengthscales: Understanding and manipulating the crossover". In: *Proceedings of the National Academy of Sciences* 102.27 (2005), pp. 9475–9480.

[80] David Huang, Phillip Geissler, and David Chandler. "Scaling of hydrophobic solvation free energies". In: *Journal of Physical Chemistry B* 105.28 (2001), pp. 6704–6709.

[81] Gustavo Chapela, Laurence Scriven, and Ted Davis. "Molecular dynamics for discontinuous potential. IV. Lennard-Jonesium". In: *Journal of Chemical Physics* 91.7 (1989), pp. 4307–4313.

[82] A. B. Adib. "Algebraic perturbation theory for dense liquids with discrete potentials". In: *Physical Review E* 75.6 (2007), p. 061204.

[83] G. Hummer and S. Garde. "Cavity expulsion and weak dewetting of hydrophobic solutes in water". In: *Physical Review Letters* 80.19 (1998), p. 4193.

[84] G. Shekhar et al. "Temperature dependence of the solubility of non-polar gases in water". In: *Biophysical Chemistry* 78.1-2 (1999), pp. 21–32.

[85] F. M. Floris et al. "Free energy and entropy for inserting cavities in water: comparison of Monte Carlo simulation and scaled particle theory results". In: *The Journal of Chemical Physics* 107.16 (1997), pp. 6353–6365.

[86] Y. Zhou, B. Mei, and K. S. Schweizer. "Integral equation theory of thermodynamics, pair structure, and growing static length scale in metastable hard sphere and Weeks-Chandler-Andersen fluids". In: *Physical Review E* 101.4 (2020), p. 042121.

[87] M. J. Abraham et al. "GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers". In: *SoftwareX* 1 (2015), pp. 19–25.

[88] Gareth Tribello et al. "PLUMED 2: New feathers for an old bird". In: *Computer Physics Communications* 185.2 (2014), pp. 604–613.

[89] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma. "The missing term in effective pair potentials". In: *Journal of Physical Chemistry* 91.24 (1987), pp. 6269–6271.

[90] Andrew Pohorille and Lawrence Pratt. "Cavities in molecular liquids and the theory of hydrophobic solubilities". In: *Journal of the American Chemical Society* 112 (1990), pp. 5066–5074.

[91] F. Sittel and G. Stock. "Perspective: Identification of collective variables and metastable states of protein dynamics". In: *The Journal of Chemical Physics* 149.15 (2018), p. 150901.

[92] Francis Starr, Johannes Nielsen, and Eugene Stanley. "Hydrogen-bond dynamics for the extended simple point-charge model of water". In: *Physical Reviews E* 62.1 A (2000), pp. 579–587.

[93] Marc Souaille and Benoit Roux. "Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations". In: *Computer Physics Communications* 135 (2001), pp. 40–57.

[94] Johannes Kastner. "Umbrella sampling". In: *Wires Computational Molecular Science* 1.6 (2011), pp. 932–942.

[95] Benoit Roux. "The calculation of the potential of mean force using computer simulations". In: *Computer Physics Communications* 91 (1995), pp. 275–282.

[96] S. Awasthi, V. Kapil, and N. N. Nair. "Sampling free energy surfaces as slices by combining umbrella sampling and metadynamics". In: *Journal of Computational Chemistry* 37.16 (2016), pp. 1413–1424.

[97] Ramona Taylor, Liem Dang, and Bruce Garrett. "Molecular dynamics simulations of the liquid/vapor interface of SPC/E water". In: *Journal of Physical Chemistry* 100.28 (1996), pp. 11720–11725.

[98] Feng Chen and Paul E. Smith. "Simulated surface tensions of common water models". In: *Journal of Chemical Physics* 126.22 (2007), p. 221101.

[99] Carlos Vega and Enrique de Miguel. "Surface tension of the most popular models of water by using the test-area simulation method". In: *Journal of Chemical Physics* 126.15 (2007), p. 154707.

[100] Grinnell Jones and Wendell Ray. "The surface tension of solutions of electrolytes as a function of the concentration. III. Sodium chloride a saturated solution". In: *Journal of the American Chemical Society* 19 (1937), p. 26.

[101] Markus Deserno. "How to calculate a three-dimensional g (r) under periodic boundary conditions". In: 1 (2004), p. 55128.

[102] J. P. Postma, H. J. Berendsen, and J. R. Haak. "Thermodynamics of cavity formation in water. A molecular dynamics study". In: *Faraday Symposia of the Chemical Society* 17 (1982), pp. 55–67.

[103] Debra McCaffrey et al. "Mechanism of ion adsorption to aqueous interfaces: Graphene/water vs. air/water". In: *Proceedings of the National Academy of Sciences* (2017), p. 201702760.

[104] M. Neumann. "Dipole moment fluctuation formulas in computer simulations of polar systems". In: *Molecular Physics* 50.4 (1983), pp. 841–858.

[105] Lubos Vrbka et al. "Propensity of soft ions for the airywater interface". In: *Current Opinion in Colloid and Interface Science* 9 (2004), pp. 67–73.

[106] C. V. Nguyen et al. "Adsorption of sodium iodine at air/water interface". In: *Journal of Molecular Liquids* 298 (2020), p. 112076.

[107] A. C. Stern et al. "Thermodynamics of iodide adsorption at the instantaneous air-water interface". In: *The Journal of Chemical Physics* 138.11 (2013), p. 114709.

[108] J. Zhang, J. Wang, and Y. "Micro-droplets formation during the deliquescence of salt particles in atmosphere". In: *Corrosion* 61.12 (2005), pp. 1167–1172.

[109] C. J. Gaston et al. "Laboratory studies of the cloud droplet activation properties and corresponding chemistry of saline playa dust". In: *Environmental Science & Technology* 51.3 (2017), pp. 1348–1356.

[110] P. Jungwirth and D. J. Tobias. "Molecular structure of salt solutions: a new view of the interface with implications for heterogeneous atmospheric chemistry". In: *The Journal of Physical Chemistry B* 105.43 (2001), pp. 10468–10472.

[111] Grigore Albeanu. "On the generalized Halley method for solving nonlinear equations". In: *Romai* 4.2 (2008), pp. 1–6.

[112] John Chodera et al. “Use of the weighted histogram analysis method for the analysis of simulated and parallel tempering simulations”. In: *Journal of Chemical Theory and Computation* 3.1 (2007), pp. 26–41.

[113] M. Frank and D. Drikakis. “Thermodynamics at solid-liquid interfaces”. In: *Entropy* 20.5 (2018), p. 362.

[114] C. Sievers et al. “Phenomena affecting catalytic reactions at solid-liquid interfaces”. In: *ACS Catalysis* 6.12 (2016), pp. 8286–8307.

[115] Robert M. Hazen and Dimitri A. Sverjensky. “Mineral Surfaces, Geochemical Complexities, and the Origins of Life”. In: *Cold Spring Harbor Perspectives in Biology* 2.5 (2010), a002162.

[116] G W achtersh auser. “Before enzymes and templates: theory of surface metabolism”. In: *Microbiological Reviews* 52.4 (1988), p. 452.

[117] F. Nudelman and N. A. Sommerdijk. “Biomineralization as an inspiration for materials chemistry”. In: *Angewandte Chemie International Edition* 51.27 (2012), pp. 6582–6596.

[118] Y. Chen et al. “Biomineralization forming process and bio-inspired nanomaterials for biomedical application: a review”. In: *Minerals* 9.2 (2019), p. 68.

[119] J. Jiang, K. Hirano, and K. Sakurai. “Interface-sensitive imaging by an image reconstruction aided X-ray reflectivity technique”. In: *Journal of Applied Crystallography* 50.3 (2017), pp. 712–721.

[120] L. Dalstein, K. Y. Chiang, and Y. C. Wen. “Direct Quantification of Water Surface Charge by Phase-Sensitive Second Harmonic Spectroscopy”. In: *The Journal of Physical Chemistry Letters* 10.17 (2019), pp. 5200–5205.

[121] T. Fukuma and R. Garcia. “Atomic-and Molecular-resolution mapping of solid-liquid interfaces by 3D atomic force microscopy”. In: *ACS Nano* 12.12 (2018), pp. 11785–11797.

[122] X. Zhang et al. “Direction-specific interaction forces underlying zinc oxide crystal growth by oriented attachment”. In: *Nature Communications* 8.1 (2017), pp. 1–8.

[123] M. Ricci et al. “Direct observation of the dynamics of single metal ions at the interface with solids in aqueous solutions”. In: *Scientific Reports* 7.1 (2017), pp. 1–12.

[124] A. M. Smith, A. A. Lee, and S. Perkin. “The electrostatic screening length in concentrated electrolytes increases with concentration”. In: *Journal of Physical Chemistry Letters* 7.12 (2016), pp. 2157–2163.

[125] Lili Liu et al. “Connecting energetics to dynamics in particle growth by oriented attachment using real-time observations”. In: *Nature Communications* 11.1 (2020), pp. 1–11.

[126] M. Goedert. “Alzheimer’s and Parkinson’s diseases: The prion concept in relation to assembled AB, tau, and α-synuclein”. In: *Science* 349.- (2015), p. 6248.

[127] D. Li et al. “Direction-specific interactions control crystal growth by oriented attachment”. In: *Science* 336.- (2012), p. 6084.

[128] J. J. McManus et al. "The physics of protein self-assembly". In: *Current Opinion in Colloid and Interface Science* 242.- (2016), pp. 73–79.

[129] Eric N. Salgado, Robert J. Radford, and F. Akif Tezcan. "Metal-directed protein self-assembly". In: *Accounts of Chemical Research* 43.5 (2010), pp. 661–672.

[130] Quan Luo et al. "Protein Assembly: Versatile Approaches to Construct Highly Ordered Nanostructures". In: *Chemical Reviews* 116.22 (2016), pp. 13571–13632.

[131] Kenneth Curry. "Mica". In: *U.S. Department of Interior Geological Survey* 24.- (2017), pp. 49.1–49.10.

[132] Sungwook Woo and Paul W. K. Rothemund. "Self-assembly of two-dimensional DNA origami lattices using cation-controlled surface diffusion". In: *Nature Communications* 5.1 (2014), pp. 1–11.

[133] Wee Wen Leow and Wonmuk Hwang. "Epitaxially Guided Assembly of Collagen Layers on Mica Surfaces". In: *Langmuir* 27.17 (2011), pp. 10907–10913.

[134] TJ Brunette et al. "Exploring the repeat protein universe through computational protein design". In: *Nature* 528.7583 (2015), pp. 580–584.

[135] Harley Pyles et al. "Controlling protein assembly on inorganic crystals through designed protein interfaces". In: *Nature* 571.7764 (2019), pp. 251–256.

[136] S. Zhang et al. "Engineering Biomolecular Self-Assembly at Solid–Liquid Interfaces". In: *Advanced Materials* -.- (2020), p. 1905784.

[137] Michael A. Boles, Michael Engel, and Dmitri V. Talapin. "Self-Assembly of Colloidal Nanocrystals: From Intricate Structures to Functional Materials". In: *Chemical Reviews* 116.18 (2016), pp. 11220–11289.

[138] Stephen Whitelam et al. "Common Physical Framework Explains Phase Behavior and Dynamics of Atomic, Molecular, and Polymeric Network Formers". In: *Physical Review X* 4.1 (2014), p. 011044.

[139] W. Leow and W. Hwang. "Epitaxially Guided Assembly of Collagen Layers on Mica Surfaces". In: *Langmuir* 27.- (2011), pp. 10907–10913.

[140] D. M. Czajkowsky et al. "Heteroepitaxial Streptavidin Nanocrystals Reveal Critical Role of Proton "Fingers" and Subsurface Atoms in Determining Adsorbed Protein Orientation". In: *ACS Nano* 6.- (2012), pp. 190–198.

[141] R. W. Loo and M. C. Goh. "Potassium Ion Mediated Collagen Microfibril Assembly on Mica". In: *Langmuir* 24.- (2008), pp. 13276–13278.

[142] Shuai Zhang et al. "Assembly of a Patchy Protein into Variable 2D Lattices via Tunable, Multiscale Interactions". In: *Nature Communications* 11.- (2020), p. 3770.

[143] Margaret Robson Wright. "Debye-Hückel and Fuoss-Onsager equations". In: *An Introduction to Aqueous Electrolyte Solutions*. 2007, p. 24.

[144] Pedro Tarazona2 Daniel Martin-Jimenez1 Enrique Chacon1 and Ricardo Garcia1. "Atomically resolved three-dimensional structures of electrolyte aqueous solutions near a solid surface". In: *Nature Communications* 7.1 (2016), pp. 1–7.

[145] Daniel Martin-Jimenez and Ricardo Garcia. "Identification of Single Adsorbed Cations on Mica-Liquid Interfaces by 3D Force Microscopy". In: *Journal of Physical Chemistry Letters* 8.23 (2017), pp. 5707–5711.

[146] Yng-Gwei Chen and John D. Weeks. "Local molecular field theory for effective attractions between like charged objects in systems with strong Coulomb interactions". In: *Proceedings of the National Academy of Sciences* 103.20 (2006), pp. 7560–7565.

[147] J. M. Rodgers and J. D. Weeks. "Local molecular field theory for the treatment of electrostatics". In: *Journal of Physics: Condensed Matter* 20.49 (2008), p. 494206.

[148] R. C. Remsing and J. D. Weeks. "Hydrophobicity scaling of aqueous interfaces by an electrostatic mapping". In: *The Journal of Physical Chemistry B* 119.29 (2015), pp. 9268–9277.

[149] Richard C Remsing, Shule Liu, and John D Weeks. "Long-ranged contributions to solvation free energies from theory and short-ranged models". In: *Proceedings of the National Academy of Sciences* 113 (2016), pp. 2819–2826.

[150] S Marčelja and N Radić. "Repulsion of interfaces due to boundary water". In: *Chemical Physics Letters* 41.1 (1976), pp. 129–130.

[151] M. Kanduč et al. "Interactions between biological membranes: theoretical concepts". In: *Soft Interfaces: Lecture Notes of the Les Houches Summer School* 98.- (2017), p. 385.

[152] Hendrik Heinz et al. "Thermodynamically Consistent Force Fields for the Assembly of Inorganic, Organic, and Biological Nanostructures: The INTERFACE Force Field". In: *Langmuir* 29.6 (2013), pp. 1754–1765.

[153] Arushi Prakash et al. "Quantifying the Molecular-Scale Aqueous Response to the Mica Surface". In: *The Journal of Physical Chemistry C* 121.34 (2017), pp. 18496–18504.

[154] R. T. Cygan, J. J. Liang, and A. G. Kalinichev. "Molecular models of hydroxide, oxyhydroxide, and clay phases and the development of a general force field." In: *The Journal of Physical Chemistry B* 108.4 (2004), pp. 1255–1266.

[155] S. N. Kerisit and J. J. De Yoreo. "Effect of Hydrophilicity and Interfacial Water Structure on Particle Attachment". In: *The Journal of Physical Chemistry C* 124.9 (2020), pp. 5480–5488.

[156] J. D. Weeks. "Connecting local structure to interface formation: A molecular scale van der Waals theory of nonuniform liquids". In: *Annual Review of Physical Chemistry* 53.1 (2002), pp. 533–562.

[157] J. M. Rodgers, Z. Hu, and J. D. Weeks. "On the efficient and accurate short-ranged simulations of uniform polar molecular liquids". In: *Molecular Physics* 109.7-10 (2008), pp. 1195–1211.

[158] J. D. Jackson. "Multipoles, Electrostatics of Macroscopic Media, Dielectrics". In: *Classical Electrodynamics*. 1999, pp. 145–150.

[159] C. Zhang and M. Sprik. "Electromechanics of the liquid water vapour interface". In: *Physical Chemistry Chemical Physics* 22.- (2020), pp. 10676–10686.

[160] Shenglin Jin et al. “Use of Ion Exchange to Regulate the Heterogeneous Ice Nucleation Efficiency of Mica”. In: *Journal of the American Chemical Society* 142.42 (2020), pp. 17956–17965.

[161] Nurun Nahar Lata et al. “Multivalent Surface Cations Enhance Heterogeneous Freezing of Water on Muscovite Mica”. In: *The Journal of Physical Chemistry Letters* 11.- (2020), pp. 8682–8689.

[162] M. Kanduč et al. “Hydration repulsion between membranes and polar surfaces: Simulation approaches versus continuum theories”. In: *Advances in Colloid and Interface Science* 208.- (2014), pp. 142–152.

[163] Dongsheng Li et al. “Trends in mica–mica adhesion reflect the influence of molecular details on long-range dispersion forces underlying aggregation and coalignment”. In: *Proceedings of the National Academy of Sciences* 114.29 (2017), pp. 7537–7542.

[164] Sebastien N Kerisit and James J De Yoreo. “Effect of Hydrophilicity and Interfacial Water Structure on Particle Attachment”. In: *Journal of Physical Chemistry C* 124.9 (2020), pp. 5480–5488.

[165] Yuta Suzuki et al. “Self-assembly of coherently dynamic, auxetic, two-dimensional protein crystals”. In: *Nature* 533.7603 (2016), pp. 369–373.

[166] Robert Alberstein et al. “Engineering the entropy-driven free-energy landscape of a dynamic nanoporous protein assembly”. In: *Nature Chemistry* 10.7 (2018), pp. 732–739.

[167] X. Shen and I. C. Bourg. “Molecular dynamics simulations of the colloidal interaction between smectite clay nanoparticles in liquid water”. In: *Journal of Colloid and Interface Science* 584 (2021), pp. 610–621.

[168] R. M. Pashley and J. N. Israelachvili. “Molecular layering of water in thin films between mica surfaces and its relation to hydration forces”. In: *Journal of Colloid and Interface Science* 101.2 (1984), pp. 511–523.

[169] D. Li et al. “Trends in mica–mica adhesion reflect the influence of molecular details on long-range dispersion forces underlying aggregation and coalignment”. In: *Proceedings of the National Academy of Sciences* 114.29 (2017), pp. 7537–7542.

[170] M. L. Sushko and K. M. Rosso. “The origin of facet selectivity and alignment in anatase TiO 2 nanoparticles in electrolyte solutions: implications for oriented attachment in metal oxides”. In: *Nanoscale* 8.47 (2016), pp. 19714–19725.

[171] S. N. Kerisit and J. J. De Yoreo. “Effect of Hydrophilicity and Interfacial Water Structure on Particle Attachment”. In: *The Journal of Physical Chemistry C* 124.9 (2020), pp. 5480–5488.

[172] A. M. Smith, A. A. Lee, and S. Perkin. “The electrostatic screening length in concentrated electrolytes increases with concentration”. In: *The Journal of Physical Chemistry Letters* 7.12 (2016), pp. 2157–2163.

[173] I. C. Bourg et al. “Stern layer structure and energetics at mica–water interfaces”. In: *Journal of Physical Chemistry C* 121.17 (2017), pp. 9402–9412.

[174] Dezs o Bodaa and Douglas Henderson. “The effects of deviations from Lorentz-Berthelot rules on the properties of a simple mixture”. In: *Molecular Physics* 106.20 (2008), pp. 2367–2370.

[175] I. S. Joung and T. E. Cheatham III. “Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations”. In: *The Journal of Physical Chemistry B* 112.30 (2008), pp. 9020–9041.

[176] J. Yoo and A. Aksimentiev. “New tricks for old dogs: improving the accuracy of biomolecular force fields by pair-specific corrections to non-bonded interactions”. In: *Physical Chemistry Chemical Physics* 20.13 (2008), pp. 8432–8449.

[177] M. J. Abraham et al. “GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers.” In: *SoftwareX* 1.- (2015), pp. 19–25.

[178] C. Kutzner et al. “More bang for your buck: Improved use of GPU nodes for GROMACS 2018”. In: *Journal of Computational Chemistry* 40.27 (2019), pp. 2418–2431.

[179] Arushi Prakash et al. “Biasing smarter, not harder, by partitioning collective variables into families in parallel bias metadynamics”. In: *Journal of Chemical Theory and Computation* 14.10 (2018), pp. 4985–4990.

[180] J. Sampath, S. Alamdari, and J. Pfaendtner. “Closing the Gap Between Modeling and Experiments in the Self-assembly of Biomolecules at Interfaces and in Solution”. In: *Chemistry of Materials* 32.- (2020), pp. 8043–8059.

[181] J. Israelachvili and H. Wennerstr om. “Role of hydration and water structure in biological and colloidal interactions”. In: *Nature* 379 (1996), p. 6562.

[182] F. H. Stillinger. “Interfacial Solutions of the Poisson-Boltzmann Equation”. In: *The Journal of Chemical Physics* 35.5 (1961), pp. 1584–1589.

[183] A. P. dos Santos and Y. Levin. “Electrolytes between dielectric charged surfaces: Simulations and theory”. In: *The Journal of Chemical Physics* 142.19 (2015), p. 194104.

[184] C. J. Brabec et al. “Organic photovoltaic devices produced from conjugated polymer/methanofullerene bulk heterojunctions”. In: *Synthetic Metals* 121.1-3 (2001), pp. 1517–1520.

[185] K. Feron et al. “Organic bioelectronics: materials and biocompatibility”. In: *International Journal of Molecular Sciences* 19.8 (2018), p. 2382.

[186] C. Liao and F. Yan. “Organic semiconductors in organic thin-film transistor-based chemical and biological sensors”. In: *Polymer Reviews* 53.3 (2013), pp. 352–406.

[187] A. Giovannitti et al. “Controlling the mode of operation of organic transistors through side-chain engineering”. In: *Proceedings of the National Academy of Sciences* 113.43 (2016), pp. 12017–12022.

[188] C. M. Pacheco-Moreno et al. “The importance of materials design to make ions flow: toward novel materials platforms for bioelectronics applications”. In: *Advanced Materials* 29.4 (2017), p. 1604446.

[189] A. Pron et al. “Electroactive materials for organic electronics: preparation strategies, structural aspects and characterization techniques”. In: *Chemical Society Reviews* 39.7 (2010), pp. 2577–2632.

[190] J. Rivnay, R. M. Owens, and G. G. Malliaras. “The rise of organic bioelectronics”. In: *Chemistry of Materials* 26.1 (2014), pp. 679–685.

[191] P. Matyba et al. “The dynamic organic p–n junction”. In: *Nature Materials* 8.8 (2009), pp. 672–676.

[192] E. M. Thomas et al. “Effects of Counter-Ion Size on Delocalization of Carriers and Stability of Doped Semiconducting Polymers”. In: *Advanced Electronic Materials* 6.12 (2020), p. 2000595.

[193] V. G. Kozlov et al. “Optical properties of molecular organic semiconductor thin films under intense electrical excitation”. In: *Applied Physics Letters* 74.8 (1999), pp. 1057–1059.

[194] L. Q. Flagg et al. “Anion-dependent doping and charge transport in organic electrochemical transistors”. In: *Chemistry of Materials* 30.15 (2018), pp. 5380–5389.

[195] S. Inal, G. G. Malliaras, and J. Rivnay. “Optical study of electrochromic moving fronts for the investigation of ion transport in conducting polymers”. In: *Journal of Materials Chemistry C* 4.18 (2016), pp. 3942–3947.

[196] C. G. Bischak et al. “A Reversible Structural Phase Transition by Electrochemically-Driven Ion Injection into a Conjugated Polymer”. In: *Journal of Materials Chemistry C* 4.18 (2016), pp. 3942–3947.

[197] C. Poelking et al. *P3HT Revisited–From Molecular Scale to Solar Cell Devices. Morphology and charge transport in P3HT: a theorist's perspective*. Springer, 2014, pp. 139–180.

[198] N. Di Césare et al. “A conformational study of ethyl-substituted bithiophenes. Semi-empirical versus ab initio methods”. In: *Synthetic Metals* 94.3 (1998), pp. 291–298.

[199] G. Raos, A. Famulari, and V. Marcon. “Computational reinvestigation of the bithiophene torsion potential”. In: *Chemical Physics Letters* 379.3-4 (2003), pp. 364–372.

[200] E. Johansson and S. Larsson. “Electronic structure and mechanism for conductivity in thiophene oligomers and regioregular polymer”. In: *Synthetic Metals* 144.2 (2003), pp. 183–191.

[201] R. Ghosh et al. “Anisotropic polaron delocalization in conjugated homopolymers and donor–acceptor copolymers”. In: *Chemistry of Materials* 31.17 (2019), pp. 7033–7045.

[202] P. C. T. Souza et al. “Martini 3: a general purpose force field for coarse-grained molecular dynamics”. In: *Nature Methods* 18.1-7 (2021), pp. 382–388.

[203] T.F. Harrelson, A. J. Moulé, and R. Faller. “Modeling organic electronic materials: bridging length and time scales”. In: *Molecular Simulation* 43.10-11 (2017), pp. 730–742.

[204] T. T. To and S. Adams. "Modelling of P3HT: PCBM interface using coarse-grained forcefield derived from accurate atomistic forcefield". In: *Physical Chemistry Chemical Physics* 16.10 (2014), pp. 4653–4663.

[205] C. K. Lee and C. W. Pao. "Nanomorphology evolution of P3HT/PCBM blends during solution-processing from coarse-grained molecular simulations". In: *The Journal of Physical Chemistry C* 118.121 (2014), pp. 11224–11233.

[206] A. Khot and B. M. Savoie. "Top–Down Coarse-Grained Framework for Characterizing Mixed Conducting Polymers". In: *Macromolecules* 54.10 (2021), pp. 4889–4901.

[207] N. Vukmirovic and L. W. Wang. "Charge carrier motion in disordered conjugated polymers: A multiscale ab initio study". In: *Nano Letters* 9.12 (2009), pp. 3996–4000.

[208] M. Matta et al. "Ion Coordination and Chelation in a Glycolated Polymer Semiconductor: Molecular Dynamics and X-Ray Fluorescence Study". In: *Chemistry of Materials* 32.17 (2020), pp. 7301–7308.

[209] C. Wang et al. "High dielectric constant semiconducting poly (3-alkylthiophene) s from side chain modification with polar sulfinyl and sulfonyl groups". In: *Macromolecules* 51.22 (2018), pp. 9368–9381.

[210] H. X. Zhou and A. Szabo. "Microscopic formulation of Marcus' theory of electron transfer". In: *The Journal of Chemical Physics* 103.9 (1995), pp. 3481–3494.

[211] A. Heck et al. "Charge transfer in model peptides: Obtaining Marcus parameters from molecular simulation". In: *The Journal of Physical Chemistry B* 116.7 (2012), pp. 2284–2293.

[212] H. L. Tavernier and M. D. Fayer. "Distance dependence of electron transfer in DNA: the role of the reorganization energy and free energy". In: *The Journal of Physical Chemistry B* 104.48 (2000), pp. 11541–11550.

[213] R. F. Giridharagopal et al. "Electrochemical strain microscopy probes morphology-induced variations in ion uptake and performance in organic electrochemical transistors". In: *Nature Materials* 16.7 (2017), pp. 737–742.

[214] O. Alexiadis and V. G. Mavrantzas. "All-atom molecular dynamics simulation of temperature effects on the structural, thermodynamic, and packing properties of the pure amorphous and pure crystalline phases of regioregular P3HT". In: *Macromolecules* 46.6 (2013), pp. 2450–2467.

[215] W. Ten Hoeve et al. "Substituted 2, 2': 5', 2": 5", 2"': 5"', 2"": 5"", 2""': 5""', 2""": 5""", 2"""': 5"""', 2""""": 5""""", 2"""""': 5"""""', 2"""""": 5""""""-undecithiophenes, the longest characterized oligothiophenes". In: *Journal of the American Chemical Society* 113.15 (1991), pp. 5887–5889.

[216] H. Meier, U. Stalmach, and H. Kolshorn. "Effective conjugation length and UV/vis spectra of oligomers". In: *Acta Polymerica* 48.9 (1997), pp. 379–384.

[217] J. Lawrence et al. "A versatile and efficient strategy to discrete conjugated oligomers". In: *Journal of the American Chemical Society* 139.39 (2017), pp. 13735–13739.

[218] M. C. Heiber and A. Dhinojwala. "Estimating the magnitude of exciton delocalization in regioregular P3HT". In: *The Journal of Physical Chemistry C* 117.42 (2013), pp. 21627–21634.

[219] C. M. Wolf et al. "Assessment of molecular dynamics simulations for amorphous poly (3-hexylthiophene) using neutron and X-ray scattering experiments". In: *Soft Matter* 15.25 (2019), pp. 5067–5083.

[220] K. Do et al. "A comparative MD study of the local structure of polymer semiconductors P3HT and PBTTT". In: *Physical Chemistry Chemical Physics* 12.44 (2010), pp. 14735–14739.

[221] A. J. Pearson et al. "Rationalizing phase transitions with thermal annealing temperatures for P3HT: PCBM organic photovoltaic devices". In: *Macromolecules* 45.3 (2012), pp. 1499–1508.

[222] Y. Sugita and Y. Okamoto. "Replica-exchange molecular dynamics method for protein folding". In: *Chemical Physics Letters* 314.1-2 (1999), pp. 141–151.

[223] A. Malavika et al. "Quantitative Analysis of Doping-Induced Polarons and Charge-Transfer Complexes of Poly (3-hexylthiophene) in Solution". In: *The Journal of Physical Chemistry B* 124.35 (2020), pp. 7694–7708.

[224] J.-M. Y. Carrillo et al. "Petascale simulations of the morphology and the molecular interface of bulk heterojunctions". In: *ACS Nano* 10.7 (2016), pp. 7008–7022.

[225] U. C. Singh and P. A. Kollman. "An approach to computing electrostatic charges for molecules". In: *Journal of Computational Chemistry* 5.2 (1984), pp. 129–145.

[226] W. Koch and M. C. Holthausen. *A chemist's guide to density functional theory*. 2001.

[227] M. J. Frisch et al. *Gaussian 16 Revision C.01*. Gaussian Inc. Wallingford CT. 2016.

[228] L. Zhang, G. H. Peslherbe, and H. M. Muchall. "A general measure of conjugation in biphenyls and their radical cations". In: *Canadian Journal of Chemistry* 88.11 (2010), pp. 1175–1185.

[229] J. Oxgaard and O. Wiest. "Rehybridized 1, 3-butadiene radical cations: How far will a radical cation go to maintain conjugation?" In: *The Journal of Physical Chemistry A* 106.15 (2002), pp. 3967–3974.

[230] J. Vymetal and J. Vondrášek. "Metadynamics As a Tool for Mapping the Conformational and Free-Energy Space of Peptides: The Alanine Dipeptide Case Study". In: *The Journal of Physical Chemistry B* 114.16 (2010), pp. 5632–5642.

[231] L. Q. Flagg et al. "Polymer crystallinity controls water uptake in glycol sidechain polymer organic electrochemical transistors". In: *Journal of the American Chemical Society* 141.10 (2019), pp. 4345–4354.

[232] C. Zhang et al. "Molecular Dynamics Simulation of Distribution and Diffusion Behaviour of Oil–Water Interfaces". In: *Molecules* 24.10 (2019), p. 1905.

[233] J. T. Bullerjahn, S. von B ulow, and G. Hummer. "Optimal estimates of self-diffusion coefficients from molecular dynamics simulations". In: *The Journal of Chemical Physics* 153.2 (2020), p. 024116.

[234] N. S. Hush. "Adiabatic theory of outer sphere electron-transfer reactions in solution". In: *Transactions of the Faraday Society* 57 (1961), pp. 557–580.

[235] S. Roy et al. "Marcus theory of ion-pairing". In: *Journal of Chemical Theory and Computation* 13.8 (2017), pp. 3470–3477.

[236] K. J. Laidler and M. C. King. "Development of transition-state theory". In: *The Journal of Physical Chemistry* 87.15 (1983), pp. 2657–26640.

[237] C. Zhu and H. Nakamura. "Theory of nonadiabatic transition for general two-state curve crossing problems. II. Landau–Zener case". In: *The Journal of Chemical Physics* 102.19 (1995), pp. 7448–7461.

[238] R. Vuilleumier et al. "Extension of Marcus picture for electron transfer reactions with large solvation changes". In: *Journal of the American Chemical Society* 134.4 (2012), pp. 2067–2074.

[239] Y. Tateyama et al. "Density-functional molecular-dynamics study of the redox reactions of two anionic, aqueous transition-metal complexes". In: *The Journal of Chemical Physics* 122.23 (2005), p. 234505.

[240] R. W. Zwanzig. "High-temperature equation of state by a perturbation method. I. Nonpolar gases". In: *The Journal of Chemical Physics* 22.8 (1954), pp. 1420–1426.

[241] H. Essén. "The physics of the born–oppenheimer approximation". In: *International Journal of Quantum Chemistry* 12.4 (1977), pp. 721–735.

[242] W. F. Libby. "Electron transfer among the transition elements; the controlling role of the Franck—condon principle on rates". In: *The Journal of Chemical Physics* 38.2 (1963), pp. 420–425.

[243] J. Blumberger and G. Lamoureux. "Reorganization free energies and quantum corrections for a model electron self-exchange reaction: Comparison of polarizable and non-polarizable solvent models." In: *Molecular Physics* 106.12-13 (2008), pp. 1597–1611.

[244] R. A. Kuharski et al. "Molecular model for aqueous ferrous–ferric electron transfer". In: *The Journal of Chemical Physics* 89.5 (1988), pp. 3248–3257.

[245] K. Ando. "Solvent nuclear quantum effects in electron transfer reactions. III. Metal ions in water. Solute size and ligand effects". In: *The Journal of Chemical Physics* 114.21 (2001), pp. 9470–9477.

[246] H. J. Kim and J. T. Hynes. "Equilibrium and nonequilibrium solvation and solute electronic structure. III. Quantum theory". In: *The Journal of Chemical Physics* 96.7 (1992), pp. 5088–5110.

[247] M. Ceccarelli and M. Marchi. "Simulation and modeling of the Rhodobacter sphaeroides bacterial reaction center II: Primary charge separation". In: *The Journal of Physical Chemistry B* 107.23 (2003), pp. 5630–5641.

[248] R. G. Alden et al. "Calculations of electrostatic energies in photosynthetic reaction centers". In: *Journal of the American Chemical Society* 117.49 (1995), pp. 12284–12298.

[249] K. Schulten and M. Tesch. "Coupling of protein motion to electron transfer: Molecular dynamics and stochastic quantum mechanics study of photosynthetic reaction centers". In: *Chemical Physics* 158.2-3 (1991), pp. 421–446.

[250] L. Hrostea et al. "On the Electrical and Optical Properties Stability of P3HT Thin Films Sensitized with Nitromethane Ferric Chloride Solutions". In: *Coatings* 10.11 (2020), p. 1074.

[251] J. Jaglarz, A. Malek, and J. Sanetra. "Thermal dependence of optical parameters of thin polythiophene films blended with PCBM". In: *Polymers* 10.4 (2018), p. 454.

[252] S. C. Blackstock S. F. Nelsen and Y. Kim. "Estimation of inner shell Marcus terms for amino nitrogen compounds by molecular orbital calculations". In: *Journal of the American Chemical Society* 109.3 (1987), pp. 677–682.

[253] M. Sato et al. "Can classical Marcus theory describe hole transfer in polyethylene?" In: *IEEE Transactions on Dielectrics and Electrical Insulation* 23.5 (2016), pp. 2978–2984.

[254] D. N. Beratan et al. "Electron-tunneling pathways in proteins". In: *Science* 258.5089 (1992), pp. 1740–1742.

[255] A. de la Lande and D. R. Salahub. "Derivation of interpretative models for long range electron transfer from constrained density functional theory". In: *Journal of Molecular Structure: THEOCHEM* 943.1-3 (2010), pp. 115–120.

[256] A. S. Parmar and M. Muschol. "Hydration and hydrodynamic interactions of lysozyme: effects of chaotropic versus kosmotropic ions". In: *Biophysical Journal* 97.2 (2009), pp. 590–598.

[257] V. Stehr et al. "Singlet exciton diffusion in organic crystals based on Marcus transfer rates". In: *Journal of Chemical Theory and Computation* 10.3 (2014), pp. 1242–1255.

[258] C. Enengl et al. "Doping-Induced Absorption Bands in P3HT: Polarons and Bipolarons". In: *The Journal of Chemical Physics* 17.23 (2016), pp. 3836–3844.

[259] M. Nowak et al. "Polarons and bipolarons on a conducting polymer in solution". In: *Macromolecules* 20.5 (1987), pp. 965–968.

[260] T. Balankura, X. Qi, and K. A. Fichthorn. "Solvent effects on molecular adsorption on ag surfaces: Polyvinylpyrrolidone oligomers". In: *The Journal of Physical Chemistry C* 122.26 (2018), pp. 14566–14573.

[261] I. Vorobyov et al. "Additive and classical drude polarizable force fields for linear and cyclic ethers". In: *Journal of Chemical Theory and Computation* 3.3 (2007), pp. 1120–1133.

[262] C. Bannwarth, S. Ehlert, and S. Grimme. "An accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions". In: *Journal of Chemical Theory and Computation* 15.3 (2019), pp. 1652–1671.

[263] J. Kussmann and C. Ochsenfeld. "Pre-selective screening for matrix elements in linear-scaling exact exchange calculations". In: *The Journal of Chemical Physics* 138.13 (2013), p. 134114.

[264] G. E. Box. “Science and statistics”. In: *Journal of the American Statistical Association* 71.356 (1976), pp. 791–799.