IMPROVING COARSE-GRAINED SCHEMES WITH APPLICATION TO ORGANIC MIXED CONDUCTORS

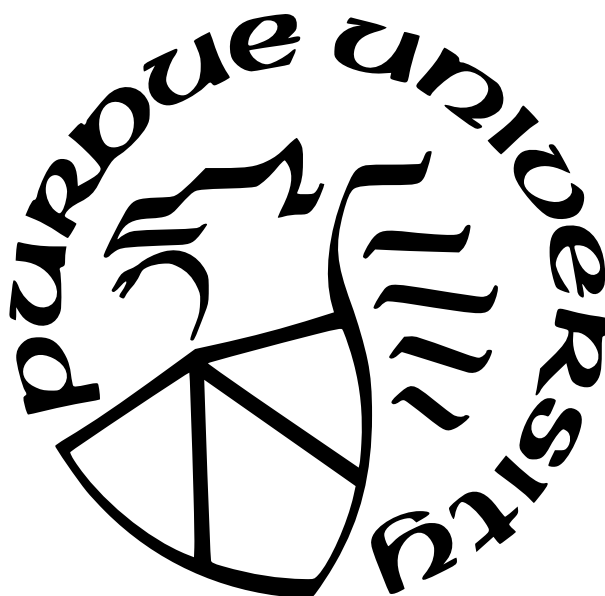by

Aditi Khot

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

Davidson School of Chemical Engineering

West Lafayette, Indiana

May 2022

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Brett M. Savoie, Chair

Department of Chemical Engineering

Dr. Bryan W. Boudouris

Department of Chemical Engineering

Dr. David S. Corti

Department of Chemical Engineering

Dr. Alejandro H. Strachan

Department of Materials Engineering

Approved by:

Dr. John A. Morgan

ACKNOWLEDGMENTS

Graduate school is an important learning experience and I am ever so grateful to have Prof. Brett Savoie as my advisor, who has been a knowledgeable and kind mentor to me. I am thankful for his expertise in mesoscale modeling of materials, his knack for finding valuable meaning in data and beautiful presentation of the data and, importantly, his nice demeanour and genuine support for his students. I am especially thankful to Prof. Pushpavanam for encouraging me to pursue research during my undergraduate studies, and recommending me as a graduate candidate to Purdue University, and to Prof. Ramakrishna for providing me the opportunity to join the graduate program. I would also like to thank my committee members Dr. Boudouris, Dr. Corti and Dr. Strachan for their useful feedback on modifying my approach towards studying mixed conductors after my preliminary presentation. I want to specially thank Dr. Boudouris for his constant encouragement and expert feedback on mixed conducting polymers. I am very thankful to Dr. Stephen Shiring for all his help in getting acquainted with cluster and computational chemistry, his technical contribution to the study on coarse-grained mapping operators, all his inputs during graduate school and for navigating the career search post-graduate school.

Of course, none of the work could have been accomplished without funding support. I want to acknowledge and thank all the funding sources– the Extreme Science and Engineering Discovery Environment (XSEDE) supported by National Science Foundation grant number ACI-1548562, Donors of the American Chemical Society Petroleum Research Fund, the Dreyfus Program for Machine Learning in the Chemical Sciences and Engineering and the National Science Foundation (NSF) Division of Chemical, Bioengineering, Environmental, and Transport Systems (CBET). I am also thankful to the computing resources at Purdue University, ITaP Research Computing for providing the resources and help to run the simulations presented in this thesis. I also want to extend my thanks to graduate administrators, Bev and Robin for making navigating through any formalities during graduate school very easy.

I would not have been able to attend graduate school and learn all the wonderful things, if not for the hard work and support of my family, whom I cannot thank enough. I want to

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

OMIEC	Organic mixed ionic-electronic conductor
CG	Coarse-graining
MD	Molecular Dynamics
CG-MD	Coarse-grained Molecular Dynamics
AA	All atomistic
OECT	Organic Electrochemical Transistor
LJ	Lennard-Jones
GB	Gay-Berne
RDF	Radial distribution function
KMC	Kinetic Monte Carlo
MSD	Mean square displacement
PES	Potential energy surface
IBI	Iterative Boltzmann Inversion
MSCG	Multiscale Coarse-graining
VAC	Velocity Autocorrelation
ACF	Autocorrelation Function
ML	Machine learning
SP	Simple physics
TL	Transfer learning

ABSTRACT

Organic mixed ion-electron conducting (OMIEC) polymers are capable of transporting both electrons and ions. This unique functionality underpins many emerging applications, including biosensors, electrochemical transistors, and batteries. The fundamental operating principles and structure-function relationships of OMIECs are still being investigated. Computational tools such as coarse-grained molecular dynamics (CGMD), which use simpler representations than in atomistic modeling, are ideal to study OMIECs, as they can explore the slow dynamics and large length scale features of polymers. Nevertheless, methods development is still required for CGMD simulations to accurately describe OMIECs.

In this thesis, two CGMD simulation approaches have been adopted. One is a so-called "top-down" approach to develop a generic model of OMIECs. Top-down models are phenomenological but capable of exploring a broad space of materials variables, including backbone anisotropy, persistence length, side-chain density, and hydrophilicity. This newly developed model was used to interrogate the effect of side-chain polarity and patterning on OMIEC physics. These studies reproduce experimentally observed polymer swelling while for the first time clarifying several molecular factors affecting charge transport, including the role of trap sites, polaron delocalization, electrolyte percolation, and suggesting side-chain patterning as a potential tool to improve OMIEC performance.

The second strategy pursued in this thesis is bottom-up CGMD modeling of specific atomistic systems. The bottom-up approach enables CGMD simulations to be quantitatively related to specific materials; yet, the sources of error and methods for addressing them have yet to be systematically established. To address this gap, we have studied the effect of the CG mapping operator, an important CG variable, on the fidelity of atomistic and CGMD simulations. A major observation from this study is that prevailing CGMD methods are underdetermined with respect to atomistic training data. In a separate study, we have proposed a hybrid machine-learning and physics-based CGMD framework that utilizes information from multiple sources and improves on the accuracy of ML-only bottom-up CGMD approaches.

1. INTRODUCTION

Organic conducting polymers have been adopted in many commercial applications and are a widely studied class of materials due to facile polymer processing, cheaper cost, and high synthetic tunability of these materials. Organic mixed ion-electron conducting polymers (OMIEC) are an emerging subclass of organic conducting polymers, which carry these advantages in addition to the ability to conduct both electrons and ions. In order to expedite the search of new OMIEC materials, it is crucial that we not only characterize new mixed conducting materials but also understand the physical phenomena unique to this class of materials. Computational tools, specifically, coarse-grained (CG) methods are an ideal molecular tool to achieve this goal as they encode the chemical information at molecular scale and at the same time, offer much needed mesoscale picture of these disordered materials. To that end, in this thesis, we developed a highly generic framework for a class of mixed conducting polymers and deployed it to systematically explore the implications of side-chain design on its physics. The usefulness of coarse-grained methods are not limited to establishing fundamentals of OMIEC and in future, we envision it as a common high throughput molecular screening tool. The thesis also includes steps in that regard, where we have benchmarked and improved upon existing bottom-up coarse-grained methods. Thus, the thesis can be broadly divided into two sections: (i) development and application of coarse-grained methods to understand and design OMIEC polymers, and (ii) improving techniques to build better coarse-grained models itself.

1.1 Organic Ion-Electron Conducting Polymers

OMIECs are a class of organic polymer that exhibit both electronic and ionic transport [126] (1.1). OMIECs exhibit electron conduction within the polymeric phase while ionic conductivity is mediated by a percolating electrolyte phase. Depending on the degree of mixing and polymer characteristics, these two conduction channels can be strongly coupled, with ion percolation leading to volumetric doping of the polymer. This mechanism has been exploited for signal transduction and charge storage, and is the basis for several technologies, including organic batteries and super capacitors [28, 157, 186], organic electrochemical
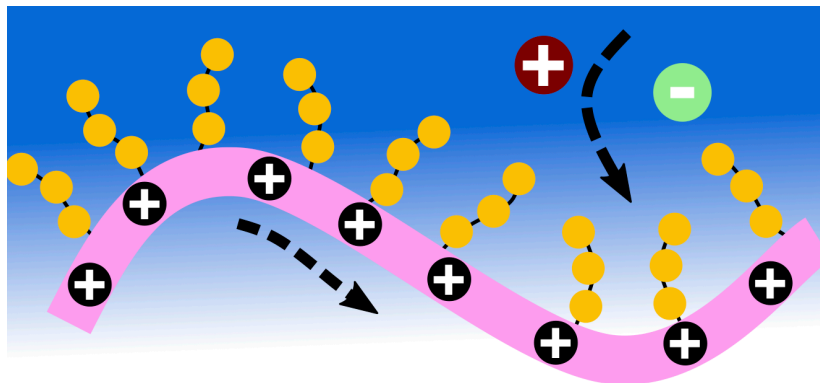
Figure 1.1. Schematic for organic mixed ion electron conductor (OMIEC). The pink film indicates a charge conducting polymer. The polymer film has certain hydrophilic component (in this case, side-chains) which allow the polymer to interface with electrolyte (the blue background). The electrolyte phase in turn conducts ions. Hence, the OMIEC system conducts both electrons and ions.

transistors,[137] neuro-computing devices [137, 161, 167] light-emitting electrochemical cells [185], chemical sensors [170], and electrochromic devices[171]. With the advent of these exciting applications, there is an urgent need for methods capable of characterizing signal transduction and transport in these materials.

To date, the design of OMIECs has been achieved using prior knowledge from organic conducting polymers and polyelectrolytes according to the end-use application. Although this modular approach has been successful in producing materials that exhibit mixed conduction it neglects the fundamental coupling between ion and electronic transport in OMIECs. This coupling manifests in several ways. First, ion transport mediates the reversible volumetric electronic doping that underpins signal transduction in these materials. Second, ion penetration impacts the morphology of these materials, which in turn impacts electronic mobility. Third, strong electrostatic interactions potentially exist between ions and charge carriers that can lead to coupled transport. These three sources of coupling confound purely modular approaches to OMIEC design, and their molecular details have yet to be significantly characterized.

Although several aspects of the performance of OMIECs have been improved by engineering side-chains, dopants and solvent conditions, many questions regarding the underlying

physics of these systems remain unanswered. Systematic material design of polymers requires a fundamental understanding of the physics of the polymeric systems, which leads to development of design rules and hypothesis-driven studies that further extend our knowledge. The aim of this thesis is designing mixed conductors in such a systematic fashion and to achieve this goal we adopt molecular simulation techniques. Molecular tools, particularly coarse-grained methods can elucidate the complex coupling mechanisms in OMIEC systems, and guide the development of design rules as they incorporate various design variables at molecular scale and provide important macroscopic information.

1.2   Coarse-grained Molecular-Dynamics

A prominent challenge in the characterization and design of mixed-conducting polymers is understanding the slow dynamic processes and nanoscale spatial heterogeneity intrinsic to these materials. The timescales of segmental motion and carrier transport (ns-ms) represents a challenge to atomistic simulation methodologies, whereas the short lengthscales (Å-nm) and disordered morphologies are a challenge to experimental microscopy and spectroscopic characterizations.

Coarse-grained Molecular-Dynamics (CG-MD) methods alleviate the sampling and timescale limitations of classical Molecular-Dynamics (MD) methods by reducing the degrees of freedom that are explicitly modeled. In CG-MD methods, individual atoms are coarse-grained into 'beads' and similar to MD methods dynamic trajectories of these beads are simulated to obtain macroscopic properties and improved configurational sampling (Figure 1.2a).

Thus, CG-MD methods retain essential molecular features but require fewer computational resources to integrate and exhibit simplified configurational landscapes that facilitate equilibration and mechanistic interpretation. CG-MD methods have been successfully implemented in the context of organic semiconductors to characterize different structural features, free energy changes, and calculate macroscopic transport properties by back-mapping to atomistic configurations [72]. Despite these successes, CG-MD approaches have not been significantly developed for mixed conducting systems, and the only studies published in literature are limited to PEDOT based materials.[103, 107, 108, 109, 138]
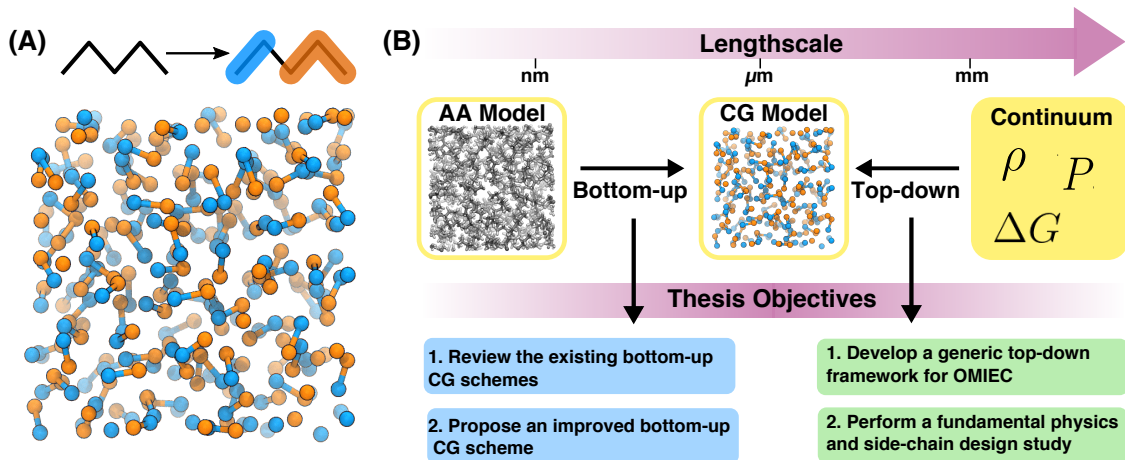
Figure 1.2. (a) Schematic showing CG model: the pentane molecule is represented by 2 CG beads instead of 17 atoms and an MD simulation is performed on this 2-bead CG model. (b) Two approaches to build the coarse-grained model– top-down (using information from bulk properties) and bottom-up (using information from atomistic MD simulations). Both the approaches have been adopted in this work to attain specific objectives as presented.

A crucial step in CG-MD simulations is developing the coarse-grained force-field (FF) that describes the interactions between chemical moieties. The forcefield is developed by reproducing information such as experimental macroscopic measurements (eg. density, pressure, free energy) or descriptors obtained from higher resolution atomistic MD simulations. The former and latter approaches are called top-down and bottom-up, respectively (Figure 1.2b). Use of top-down force-fields has been more common due to the availability of experimental data on bulk properties and easier parameterization. There are already widely used standard CG top-down force-fields such as Martini[99, 100], SAFT-γ[8, 113], etc. Although there are several schemes to incorporate bottom-up information and software packages[36, 79, 93, 106, 141, 184] to construct bottom-up models, no general method for deriving these force-fields exists. Top-down methods are more qualitative than quantitative in nature due to the limitations of target experimental information and the phenomenological relationship to specific atomistic systems. On the contrary, the target information can be systematically controlled if derived from all atomistic simulations. The top-down approach is an excellent choice for modeling generic structure-function relationships as the system variables are changed. However, in order to screen materials in a high-throughput fashion,

we need force-fields that are chemically specific and developed consistently across the chemical space of interest. Hence, the existing bottom-up schemes must be tested, improved and eventually automated to obtain a routine molecular modeling tool.

Considering the availability and easier adoption of top-down models and the severe lack of simulation studies for mixed conductors, we adopted the top-down CG force-field for our immediate investigation of mixed conductor physics. But simultaneously, we also examined the existing bottom-up CG schemes and proposed a new approach to advance bottom-up CG parameterization further. To summarize, there are two overarching objectives of this thesis (Fig. 1.2). First, developing a generic CG framework that enables us to probe key characteristics and design rules of OMIEC using a top-down CG model. Second, to benchmark and extend bottom-up CG methods in a systematic fashion.

1.3 Top-Down Coarse-grained Framework for OMIEC Polymers and OMIEC Side-chain Design

We develop a comprehensive framework for characterizing mixed conducting polymers, which consists of a generic CG model built using the top-down coarse-graining approach. Instead of parameterizing CG models of specific mixed-conducting polymers, we build a general model for the class of conjugated mixed conducting polymers, by varying its potential interactions over a physically-informed parameter space from the coarse-grained and experimental literature.

The generic CG model is composed of four building blocks: an anisotropic backbone bead, isotropic side-chain beads, ionic beads and water. The Martini CG model[101], originally obtained for biomolecules, is the only extensive CG parameterization in literature where the CG beads were classified according to their polarity and hydrogen bonding nature. These potential interactions and chemistry based bead classification have been adopted in our generic CG model to systematically vary the backbone, side-chain and dopant chemistry (Figure 1.3a) and various intermolecular interactions (Figure 1.3b). Moreover, our model takes into account the anisotropic nature of the backbone bead that allows control over the torsional variation, pi-pi packing and aggregation of the polymer (Figure 1.3a). In addition to the CG model, the framework consists of charge transport calculation scheme

19
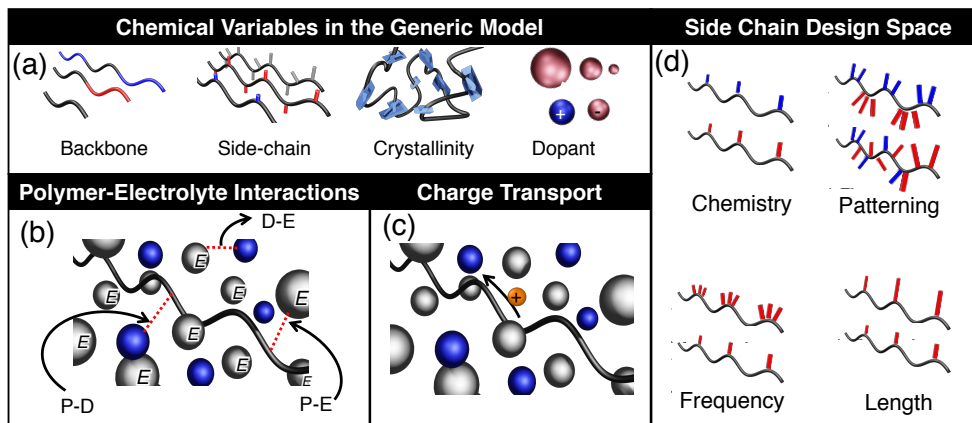
Figure 1.3. Schematic displaying the crucial aspects of OMIEC system captured by the top-down CG framework: (a) Major design variables in OMIEC systems, (b) chemical interactions between polymer (P), dopant (D) and electrolyte (E), and (c) charge transport processes. (d) Various aspects of side-chain design space, from which the top two, i.e. chemistry and patterning have been explored in this work.

using kinetic Monte Carlo (KMC) simulations (Figure 1.3c). The CG model characterizes the morphological arrangement of the polymer and electrolyte percolation in the polymer bulk; while the KMC simulations elucidate how the emergent structure affects charge transport. To validate the model, we performed studies of the effect of hydration and oxidation levels on a representative mixed conductor. The model recapitulates experimental trends related to the macroscopic ionic and electronic conductivities, including the non-linear suppression of the electronic mobility with respect to oxidation level and the direct relationship between ionic mobility and hydration level, while revealing the complex interplay of polymer morphology, ionic-electronic coupling, and electrolyte distribution that govern these relationships.

As a demonstration of the model's ability to elucidate design features, we have also performed study of how side-chain properties affect OMIEC figures-of-merit. Polar side-chains have been used to transform the traditional organic conducting polymer to a mixed conductor by facilitating electrolyte percolation through the polymer film.[50, 150] Remarkably, side-chains offer more design knobs than just polarity such as their length, sterics, frequency and patterning along the backbone (Figure 1.3d). We perform the first dedicated CG-MD study of OMIECs where the side-chain polarity and distribution are systematically varied using a

random copolymer architecture. The simulations recapitulate the non-linear progression of the morphology from an interfacially gated electrolyte when large fractions of hydrophobic side-chains are incorporated, to an electrolyte swelled morphology after crossing a threshold of approximately 40% polar side-chains. Kinetic Monte Carlo simulations were used to characterize the charge transport behaviors in these systems, revealing two interesting maxima in the mobility at 40% and 100% polar side-chain fractions, respectively. With respect to maximizing the charge mobility and conductivity, these simulations suggest that a uniform hydrophilic side-chain distribution is optimal and that there are few advantages to using mixed side-chains in a random copolymer architecture. The immediate question is if there is any advantage to adding mixed side-chains if they are placed in ordered blocks along the backbone. For instance, we have designated polar side-chain blocks to interface with the elctrolyte while apolar blocks are potentially advantageous for efficient charge transport through closely packed polymers.

In order to test this hypothesis, we perform a systematic study of polymers with varying block length of polar and apolar side-chains along the backbone for side-chain compositions of 50%, 60% and 80% polar side-chains. Through these simulations, we observe that block patterning of side-chains definitely result in distinct morphologies as the polar and apolar side-chain blocks tend to stack with side-chain blocks of its own kind. This has desirable implications, that is, for polymers with larger block length, the swelling improve because we have dedicated polar blocks to interface in solvent has hypothesized. It also improves the aggregation and connectivity as blocks of similar nature stack with each other and avoid any energy penalty associated with interaction of apolar block with polar blocks as well as electrolyte. An important finding of this study is that the improvement in morphology is greater when the blocks of polar and apolar side-chains are separated across the backbone than along the backbone. In the scope of the simulations, we do not see a direct impact of the improved polymer morphology on the final mobility estimated. In the large block length configurations, the charge appears to get trapped in the well stacked aggregates reducing the effective mobility despite the increased connections in this networks. This undesirable phenomena of charge tapping may not exist in real systems due to trap filling

and the morphological improvements associated with block patterning in our study offer encouragement for further experimental work on side-chain patterning.

1.4 Bottom-Up Coarse-Graining Framework

Although we obtain important qualitative insights into mixed conductor physics by a top-down approach, in the long run, it is crucial to construct a CG force-field in a systematic and automated fashion to obtain a chemically specific and quantitative analysis across polymers. Again, to enable quantitative comparison across different polymers, the target data must be obtained from equally reliable sources. Due to the variability and limited availability of experimental data, it is desirable to use bottom-up coarse-graining methods which use atomistic MD simulations for obtaining target data. We further ensure the comparability of atomistic force-fields, by using TAFFI force-fields which are developed using consistent levels of quantum-chemistry calculations and retain chemically transferability due to group-theory methods [148].

The prevailing bottom-up CG methods have several unresolved issues, including (*i*) poor reproduction of system properties compared to atomistic simulations, (*ii*) poor transferability of the CG models to other thermodynamic state-points and (*iii*) poor transferability of the CG models to other chemical structures with similar constituent atoms. A bottom-up CG model has three components– a CG mapping operator (i.e., the number of beads, their topology and their relation with the underlying atomistic representation), the complexity of the CG potential surface and, the parameterization procedure. Poor performance of bottom-up CG models can be the result of insufficient complexity of the CG mapping and its potential surface (i.e., 'representability limitations') or the target information used in CG potential paramterization (i.e., 'information limitations').

The CG mapping operators must be sufficiently complex across different polymers to capture chemically-specific information and exhibit correct physical behavior. Identifying the most effective CG mapping scheme is not straightforward, and researchers have typically resorted to heuristic models based on chemical intuition[99]. To address this, we have developed an automated framework which generates all possible CG mappings for a given

molecule, under a few physically motivated constraints. We then performed a systematic investigation of these errors by generating all possible coarse-grained models of three liquids (butane, 1-butanol, 1,3-propane diol) using this framework. Further, the performance and transferability of the CG model depends on the type of target system properties incorporated in it. Hence, we compare two common bottom-up frameworks– Iterative Boltzmann Inversion[134] (IBI) and Multiscale Coarse-graining,[120, 121] (MSCG). For off-target properties, we observe a strong correlation between the accuracy and the resolution of the CG model that suggests the approximations represented by MSCG and IBI deteriorate with decreasing resolution. Conversely, on-target properties exhibit an extremely weak resolution dependence that suggests a limited role of representability errors in model accuracy. Taken together, these results suggest that simple CG models are capable of utilizing more information than is provided by standard parameterization algorithms, and that model accuracy can be improved by algorithm development rather than resorting to more complicated CG models.

Apart from CG mapping operators, the complexity of the potential surface is also essential in order to capture specific interactions such as many-body forces, anisotropy, etc. As a physically intuitive guess, the PES is represented in the same fashion as for atomistic systems, but that may neglect some of the effective interactions originating from coarse-graining of the system. Machine learning (ML) methods have been used across multiple chemical applications to perform a variety of complex tasks in a black-box fashion. The ML architecture provides a complex representationof the PES that potentially eliminates representability errors. While there are some ML based bottom-up parameterization schemes already in literature, whether they are truly advantageous over the physics motivated representations still remains unknown. The physics motivated intuitions are not just reflected in the PES representation but also its parameterization, as they are often coupled. In our second work, we compare a widely used physics motivated CG parameterization scheme called IBI with an ML based CG model called DeePCG. The comparison is again performed across different mappings of a small molecule (butanol) and a polymer (poly-etheylene glycol). We found that the ML model outperforms the former physics motivated approach only minimally and lacks some of the key physics leading to erroneous behavior such as breaking of bonds in

a nonreactive systems. In addition to this comparison, we proposed a new approach for obtaining coarse-grained potentials called difference learning (Δ-models), which is a transfer learning strategy for leveraging physics-based models during neural network training. Basically, in the transfer learning approach we obtain the bulk of PES using a physics based approach as a good starting point and use the ML model only to capture the missing physics. The main finding is that the transfer learning model outperform the reference ML-only CG-MD models in nearly all scenarios. In several cases, the ML-only models also manage to game the objective function leading to qualitatively incorrect dynamics, which is corrected by transfer learning. Given their negligible added cost, transfer learning provide essentially free gains over their ML-only counterparts and warrant further investigation as the search for robust black-box CG-MD models continues.

The outline of the thesis as follows. The thesis is bifurcated into two parts– the work on modeling and understanding mixed conducting polymers, and the development of bottom-up coarse-grained framework. In the second chapter, we provide a detailed review of existing literature for these two subjects. Through this, we explain the fundamental mechanism of these polymers and the existing knowledge gap, their applications and corresponding properties of merit, and how the chemical design search has progressed in that regard. Similarly, we explain the basics of constructing a coarse-grained model in a bottom-up fashion, highlighting the respective work done in each aspect. With this existing understanding of the mixed conducting polymers, we define the objectives of our generic CG-MD model in the third chapter and a brief summary of relevant CG-MD top-down models. Then, we discuss the development of the CG-MD model and several case studies. In the fourth chapter, we report the first side-chain design study, using the CG-MD model to investigate side-chain polarity effect on OMIECs. The chapter follows a basic outline of introducing the system, the methodology, results and discussion. The fifth chapter discusses the effect of block pattering of side-chains along the same lines. In the later half of the thesis, we change gears and describe the work on developing bottom-up CG force-fields. In the sixth chapter, we report our study on mapping schemes and popular parameterization schemes. In the seventh chapter, we propose a new transfer learning approach that is a combination of typical

physics motivated and ML based approaches. Further, we discuss a case-study comparing the shortcomings and benefits of the three approaches.

2. LITERATURE REVIEW

In this chapter, we review the existing work on both fronts of organic mixed ionic-electronic conducting polymers and bottom-up coarse-grained methods. Through the first section, we reintroduce organic mixed ionic-electronic conducting polymers and their fundamental characteristics, as well as the existing experimental studies investigating various candidate mixed conductors. In the second section, we discuss the basics of building a coarse-grained model, and the respective methods reported in literature to develop coarse-grained models.

2.1 Organic Mixed Ionic-Electronic Conducting Polymers

Organic mixed ionic-electronic conductors (OMIEC) [126] are composed of organic semi-conducting polymers and an absorbed electrolyte. OMIECs exhibit electron conduction within the polymeric phase while ionic conductivity is mediated by a percolating electrolyte phase. Depending on the degree of mixing and polymer characteristics, these two conduction channels can be strongly coupled, with ion percolation leading to volumetric doping of the polymer. This mechanism has been exploited for signal transduction and charge storage, and is the basis for several technologies, including organic batteries and super capacitors [28, 157, 186], organic electrochemical transistors,[137] neuro-computing devices [137, 161, 167] light-emitting electrochemical cells [185], chemical sensors [170], and electrochromic devices[171].

Depending on the end-use application, the desired figure-of-merit and the optimal chemical design of these mixed conducting polymers varies.[127] For instance, consider an organic electrochemical transistor (OECT, Figure 2.1), it is comprised of three terminals– source, drain and gate akin to a transistor. The drain and source are connected by a conducting polymer substrate, this substrate is further in contact with an electrolyte solution and the gate electrode is placed in the electrolyte solution. The voltage at the gate terminal (V_{GS}) controls the flow of ions from the electrolyte through the susbtrate, effectively controlling the doping in the polymer substrate and hence, its output current between source and drain. The performance of an OECT device is governed by transductance which is product of the capacitance and mobility of the device. In other words, how effectively the polymer is

(de)doped by the electrolyte (i.e., capacitance) and how well does current transmit through the polymer film (i.e., mobility).[66] On the contrary, for devices such as organic batteries, energy and power stored per unit volume should be higher, while for electrochromic devices the response time should be faster.
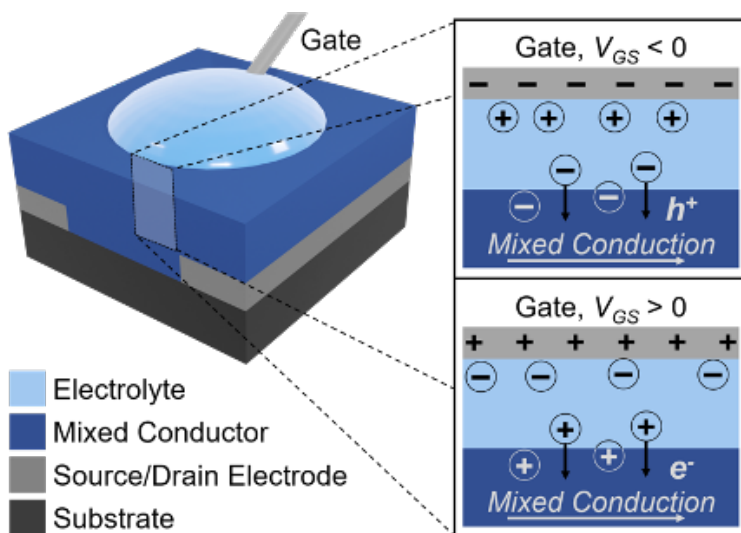
Figure 2.1. Schematic geometry of an organic electrochemical transistor (OECT). The OECT contains an electrolyte gate placed on top of the organic semiconductor layer. The dopant migration into the mixed conducting layer is modulated by the gate voltage and causes penetration of ions into the active layer polymer, which facilitates mixed charge and ion conduction. Reprinted (adapted) with permission from [30]. Copyright 2021 American Chemical Society.

To begin with, the synthesis of materials exhibiting mixed conduction typically has been achieved by combining known ion-conducting and charge-conducting components into a single polymeric material. In this approach, polyelectrolyte species are introduced to facilitate ion transport, while electronically-active moieties are introduced to facilitate charge transport. Given the potential modularity associated with a number of these design approaches, most of the chemical groups introduced for electronic and ionic conduction have been borrowed from adjacent fields in organic electronics and have not been designed specifically for the purpose of mixed ion-electron conduction. Of course, this is a reasonable first approach; however, it does show the significant opportunities that could be had for polymer chemists

as most of the groundbreaking work in the filed has relied on well-known, commercially-sourced materials. For example, the most commonly-used organic mixed conductor, poly(3,4-ethylenedioxythiophene) doped with poly(styrene sulfonate) (PEDOT:PSS), was originally developed as an antistatic coating, rather than for mixed conducting applications.[59, 128] There is no doubt that PEDOT:PSS has served the early stages of this growing community well; however, it is clear that advanced macromolecular designs will be required in order to push the bounds of the field, both in terms of fundamental understanding and in terms of translational devices. This is not unlike other organic electronics applications where a champion material has helped establish the feasibility of soft materials in applications at the outset of the field; then, as time progresses, polymer chemists have joined the field in order to make significant contributions in terms of macromolecular design. For instance, other variants of dioxythiophene,[146, 187] thiophenes [42, 119], and thienothiophenes[52] (groups that are well-represented in organic semiconducting and conducting materials) are extant in mixed conductors exhibiting p-type transport, and naphthalene diimide[51] is prevalent in mixed conductors exhibiting n-type transport (see Fig 2.2b for specific chemical structures for oft-used mixed-conducting polymers). The ion-conduction is achieved through ionic moieties like sulfonate ions [137] or highly polar moieties on side-chains such as alcohol [124], ethylene dioxide [42, 52], sulfate [67].

While there is no doubt that many of the molecules, borrowed from other polymer science subfields, have shown impressive performance metrics in end-use applications[23, 58, 104], this macromolecular implementation strategy neglects the fundamental ways that the ionic and electronic component are intrinsically coupled in these materials that must ultimately be addressed in the design process. In addition to the molecular chemistry, its molecular weight and macroscale aggregation as well as choice of ions and solvent will further complicate OMIEC performance. Thus, it is essential to understand how the coupling manifests between ionic and electronic components and how it is affected by the various chemical entities in the OMIEC system.

The foremost form of ion-electron coupling is through the alteration of doping of the electron conducting polymer due to the presence of an ion (Figure 2.2a left). The effective modulation of polymer conductivity with various ionic and polymer dopants have been
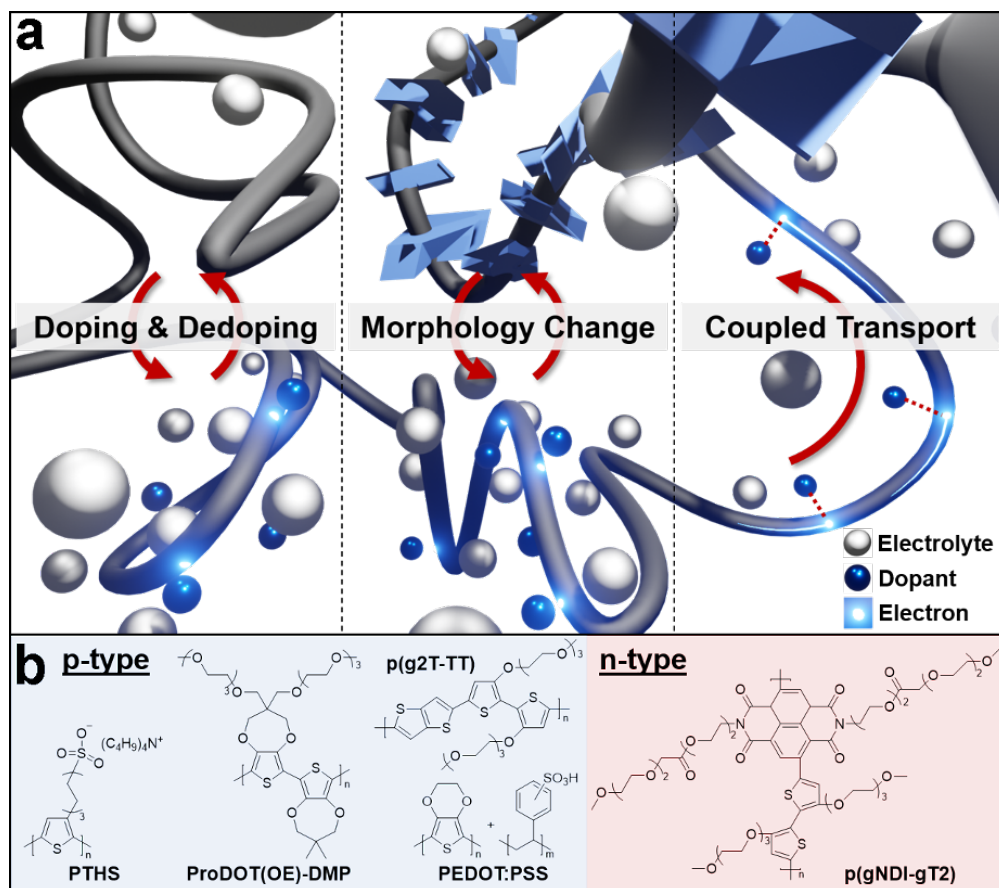
Figure 2.2. (a) Schematic diagram of the three central roles in mixed ion-electron transport processes. These processes are the: (left) doping and dedoping mechanisms; (center) morphological changes with respect to ion penetration; and (right) ion-coupled electron transport. The blue prisms on the polymer chain represent crystalline domains of the polymers. (b) Examples of polymeric materials that are used in mixed conducting applications. These materials can either be hole-transporting (i.e., p-type) or electron-transporting (i.e., n-type) polymers. Reprinted (adapted) with permission from [30]. Copyright 2021 American Chemical Society.

successfully achieved in organic electronics literature [74, 155]. The volumetric doping in OMIEC systems however adds additional challenges. The ions usually percolate the polymer in the ion-conducting domains [136] through water channels, but the prediction of this distribution as a function of chemical nature of polymer, ion concetration and solvent properties is still being studied. Some studies have shown that ions are more populated in the amorphous regions than crystallite but oxidize the crystallite regions easily due to the longer

delocalization lengths in the crystallite region.[35, 166, 187] The extent and rate of doping of anions is found to be higher for bulkier ions due to smaller hydration shell [24, 42, 43]. In [151], they discovered that a large fraction of ions which percolate the polymer does not contribute to the doping change. Also, the relative extent of motion of counter-ion and dopant ion is dependent on the size and concentration of the ions [114, 132]. The resultant electronic mobility varies in a non-monotonic fashion with the dopant concentration as the dopant ions can act as charge traps [86, 174].

Further, the electrolyte percolation is dependent on the polymer morphology as it percolates primarily in the side-chains or the amorphous polymer region (Figure 2.2a center).[42, 150] This morphology dependence on the polymer crystallinity adversely affects the dopant concentration by limiting the electrolyte uptake.[38, 136, 158]. Thus, a trade-off appears between the electronic conductivity and ionic conductivity. However, in [42], they discovered that the water percolation impaired the polymer connectivity in crystalline film leading to lower electronic mobility compared to the amorphous film. Alternatively, in [83], they developed an improved procedure for film synthesis which yields both high ionic and electronic conductivity. The swelling due to water percolation and the recurring ionic transport, in turn, alter the morphology and hence, electron transport characteristics [156, 165, 166] and leads to poor device stability. In general, the electrolyte percolation leads to a decrease in $\pi - \pi$ stacking of polymer and increase in lamellar packing of the system.[35, 42, 166] The stability has been improved using by bulky hydrophobic side-chain groups [183] and novel polymer processing techniques [42]. This cause and effect relationship between ion transport and polymer morphology is another mechanism of ionic coupling.

The third mechanism of ion-electron coupling is the coupled ion-electron transport (Figure 2.2a right). This can be mitigated due to separation of ion and electron transporting domains [136] and electrostatic screening by solvent. However, the extent of this mitigation must be quantified as the ion and electron mobilities can be on similar time-scale in some configurations [126]. The simultaneous charge-ion transport has been primarily studied using drift-diffusion models which incorporate ionic and electronic coupling through macroscopic charge balance [41, 163, 168]. These studies yield useful device level characteristics but do not reveal the ion-electron coupling mechanisms at molecular level.

30

In the scope of this thesis, we try to understand these fundamental mechanisms while advancing OMIEC design, more specifically for OECT devices, by tuning the chemical design of polymer side-chain. In the past few years, extensive efforts have been dedicated to developing OMIECs specifically for organic electrochemical transistors (OECT).[137] For this reason, the structure-function relationships that have been observed for side-chain engineering of OMIECs have been studied in the context of OECTs. Several comparative studies of how side-chain chemistry impacts OECT performance have documented a variety of behaviors depending on the polymer chemistry, electrolyte, and particular side-chain substitution. The first studies introduced polar side-chains on n-type and p-type organic polymers as a way of obtaining volumetric electrolyte percolation.[50, 150] In these studies, they investigated how OECT properties respond to incorporating varying amounts of polar and apolar side-chains and found that all polar side-chains generally maximize the transconductance. Further, the side-chain length, branching, frequency, and placement have also been shown to affect OMIEC figures of merit. These factors have been explored for specific backbone and electrolyte combinations across many recent experimental studies,[29, 35, 61, 98, 111, 112, 118, 122, 129, 153, 162, 187] and to a lesser extent by simulations.[110, 145] Among the salient trends from these studies, are that branched side-chains exhibit higher capacitance[187] but lower conductivity[129], intermediate length side-chains (~3 glycol units) lead to the highest transconductance[111, 112], and tail-to-tail side-chain linkage shows higher transconductance than head-to-head linkage.[61] An additional strategy that has recently gained traction is to include an alkyl linker between polar side-chains and the backbone to limit backbone-electrolyte interactions and improve polymer packing and conductivity.[35, 98, 118, 122, 153] However, these side-chain design rules are not universal and strongly depend on the backbone chemistry.[29, 162]

2.2 Bottom-Up Coarse-grained Methods

As discussed before, coarse-grained (CG) models are attractive for studying biological and soft material processes that occur on time and length scales that are too costly for direct atomistic simulation. By reducing the degrees of freedom that are directly modeled, CG

equations-of-motion require fewer computational resources to integrate and CG models exhibit simplified configurational phase spaces that are easier to sample and interpret. Despite these advantages, both fundamental and practical problems are presented by the dimension-reduction associated with coarse graining. Inherently, coarse-grained models lose information from high resolution atomistic representation and the quality of the coarse-grained model can be improved. The architecture of the coarse-grained model boils down to three factors–the mapping operator, the potential energy surface (PES) and the parameterization scheme (Figure 2.3). The mapping operator defines a mapping between atomistic representation and coarse-grained representation. The potential energy surface describes the nature of interaction between the CG beads and the parameterization scheme entails what and how the information from atomistic simulation is used to obtain the PES for the coarse-grained molecular system.
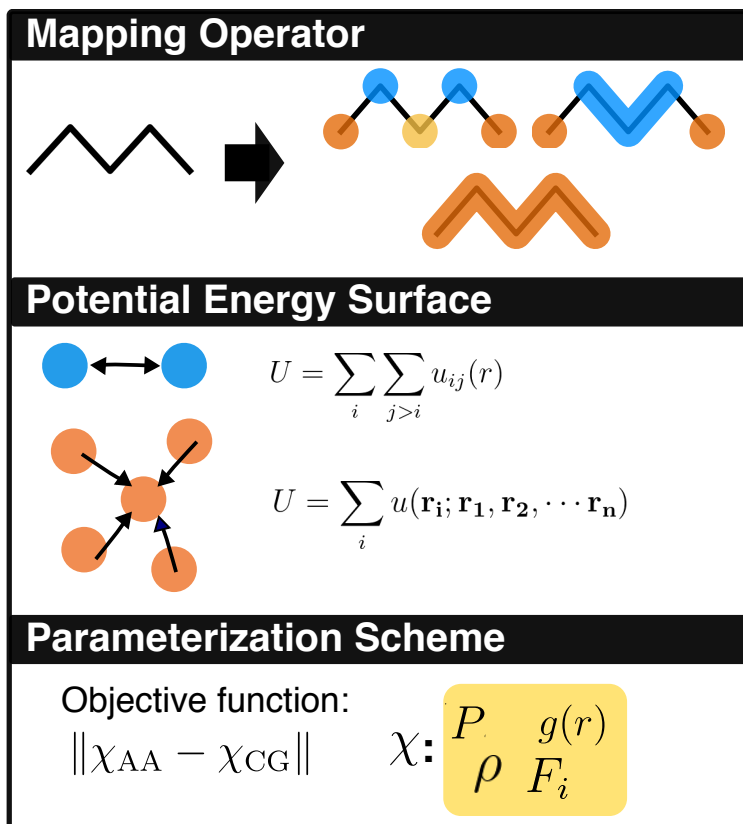
Figure 2.3. The three components which dictate the CG model performance–mapping operator, functional form of potential energy surface, and how the function is parameterized (from top to bottom).

The mapping operator is generally represented by a matrix that relates the positions and momenta of the atoms in the atomistic representation to pseudo-particles in the coarse-grained representation. Since CG models must be rederived for every unique mapping, the choice of mapping operator is typically fixed by chemical intuition without studying the impact of this choice. For example, common mapping operators for water combine 1 to 4 molecules into a single bead.[60, 68] Protein CG models sometimes use multiple operators, one for the backbone (2-4 atoms per bead) and others for specific side-chains (1-2 atoms per bead).[68] Most CG models for nucleic acids separately combine each base, sugar, and phosphate into 1-3 beads [68]. In the MARTINI scheme, 2-4 heavy atoms are beaded together and matched to one of 18 types classified by charge and polarity.[99, 100, 101] Faced with this diversity of *ad hoc* mappings, several groups have attempted to develop systematic mapping procedures that conserve stereochemistry,[105] topology,[178], mass distribution[7] and symmetry.[26]. There have also been attempts to coarse-grain atoms which are connected by high frequency motions[54], or encode the atomistic system to a coarse-grained latent space using autoencoders[177] as well as training a neural network on chemically intuitive mappings annotated by human experts[89].

The performance of the coarse-grained model may depend on the choice of mapping operator. An important question is if this quality of a mapping operator can be roughly described based on its *resolution* (number of atoms in a bead) and number of beads, or more specifically on the exact atoms which define a bead, i.e. conservation of specific features such as symmetry, topology, etc. Hence, we investigated the effect of mapping operator on the performance of coarse-grained model in our work in a systematic fashion (also discussed in Chapter 6).[82] Prior to this work, the studies to understand the effect of mapping operator were limited[26, 32]. We found in our work that the mapping operator has seemingly less effect on the performance of coarse-grained models in capturing properties like structural correlations, forces per bead and system pressure. However, dynamics of the system suffer as one approaches fewer number of beads, irrespective of the resolution. Other works have also reported the resolution independence for structural properties and pressure.[27, 34] Other properties such as information, quality of information, certain thermodynamic properties

and electronic energy have showed poor reproduction with resolution of the coarse-grained model.[32, 44, 45, 73]

The potential energy surface and the target information and algorithm used to parameterize can often be interrelated for obvious reasons. The potential energy surface function is borrowed most often from the atomistic MD simulation literature, where the PES is decomposed into intramolecular (bond, angle and dihedral) and intermolecular interactions. Common functional forms used for intramolecular interactions are harmonic, cosine, OPLS style etc. as well as spline interpolated tabulated potentials. The intermolecular interactions are commonly decomposed further into pairwise repulsive and attractive Van Der Waals contribution and described using functional forms such as Lennard-Jones, Morse, etc. or spline interpolated potentials. Most of these representation schemes are motivated by physical intuition, but recently, a more blackbox approach of using machine learning (ML) to describe PES has emerged. The true PES is a many body function and hence, ML based potentials serve as a better approximation.

The aforementioned physics-based representation for CG models have been parameterized by reproducing various properties from atomistic representation such as, forces between a pair of beads (Effective force coarse-graining)[177] or forces averaged over the entire configuration in condensed phased (Multiscale coarse-graining)[69, 70], effective potential energy (Conditional Reversible Work)[20], or structural correlations (Iterative Boltzmann Inversion, Inverse Monte Carlo)[94, 134]. Other approaches include maximizing the likelihood between atomistic and CG representation (Relative Entropy)[154], and incorporating dynamics information[40, 40, 87], or derivative-free optimization using algorithms like Particle Swarm Optimization [12, 39, 180]. Some of these parameterization schemes can be used for an ML based representation as well.

Typically, machine learned potential architecture consists of some physics informed descriptors which are fed to a black-box ML framework to output the forces or potential for a configuration. The ML models are trained most commonly on the easily differentiable property of force. In DeePCG model, Zhang et al used a beadwise local neighborhood descriptor fed to a deep neural network[184], CGnet model[173] was composed of handpicked features fed to an artificial neural network and was later improved by replacing the curated features

with filter convolutions[64], Wang et al used autoencoders to encode AA coordinates to CG coordinates in latent space and backmap them back to AA coordinates, and this latent space CG representation was fed to a neural network to obtain the potential surface[176], Scherer et al proposed a kernel based approach, where instead the two-body and three-body interactions were transformed through a nonlinear kernel before applying simple linear regression to obtain forces[152], and John et al implemented features decomposed into monomer-based terms fed to a Gaussian process regression framework[75]. In addition to force-matching approach, some works have attempted sampling the correct conformational distribution or free energy surface itself, by using a classfier network[88] and generative adversarial network[37] approach.

3. TOP-DOWN COARSE-GRAINED FRAMEWORK FOR CHARACTERIZING MIXED CONDUCTING POLYMERS

As discussed in the previous chapters, the ultimate goal of this work is understanding the fundamentals of mixed conducting polymers through systematic sweep of the mixed conductor material design space. More specifically, the work is focused on designing conjugated mixed conducting polymers and characterizing their emergent properties in the context of OECT devices. In order to achieve this, we need a model which includes the key physical processes in conjugated polymers, as well as characterizes the fundamental processes in mixed conducting system. Morphology plays a crucial role in organic polymer performance and is governed by the π-π interactions characteristic of conjugated polymers, which must be incorporated. On top of polymer arrangement, for mixed conductors, we need to account for the co-percolation of electronic and conducting phases electrolyte. These phases are coupled both by their intimate mixing (typically <100 nm) and through significant morphology changes that accompany the (de)doping processes.[136, 156, 165, 166] The presence of mobile ions and polarons in the same material substantially alters the description of charge transport relative to typical organic semiconductors and conductors.[86, 114, 132, 151, 174] On the one hand, potentially strong electrostatic interactions between ions and polarons presents the possibility of significant charge trapping and ion-coupled electron transport. On the other hand, depending on the charge densities and characteristics of the electrolyte phase (e.g., aqueous vs. organic) screening may mitigate trapping.[136] Hence, after obtaining the structural features, the effective ability of the polymer to transmit charge must be characterized, as a function of the obtained polymer arrangement and electrostatic interaction with ions and solvent. Finally, our objective is modeling a variety of polymers belonging to the family of conjugated mixed conducting polymers and that capability must be reflected in the model.

Hence, we have developed a generic coarse-grained molecular dynamics (CG-MD) model of OMIECs and coupled it with a kinetic Monte Carlo (KMC)[117] model of charge transport (Figure 3.1). This modeling framework provides a description of both electronic and ionic transport mechanisms at lengthscales of 10-100 nm, which fills a critical characterization gap in contemporary experimental approaches. The coarse-grained models accounts for the π-π stacking and torsional disorder characteristic of conjugated polymers as well as encodes the hydrogen-bonding nature and polarity of side-chains and solvent. The KMC methodology that we develop builds on a relatively mature approach to modeling charge transport in organic semiconductors, but with extensions to account for the distinct ion-electronic coupling mechanisms in OMIECs. These KMC simulations utilize CG-MD configurations with a mean-field treatment of electrons and an explicit description of the electrolyte and polymer nuclear degrees of freedom.

CG-MD models have been previously developed to study the morphology of various organic semiconducting polymers;[72, 135] however, CG-MD modeling of OMIECs has been
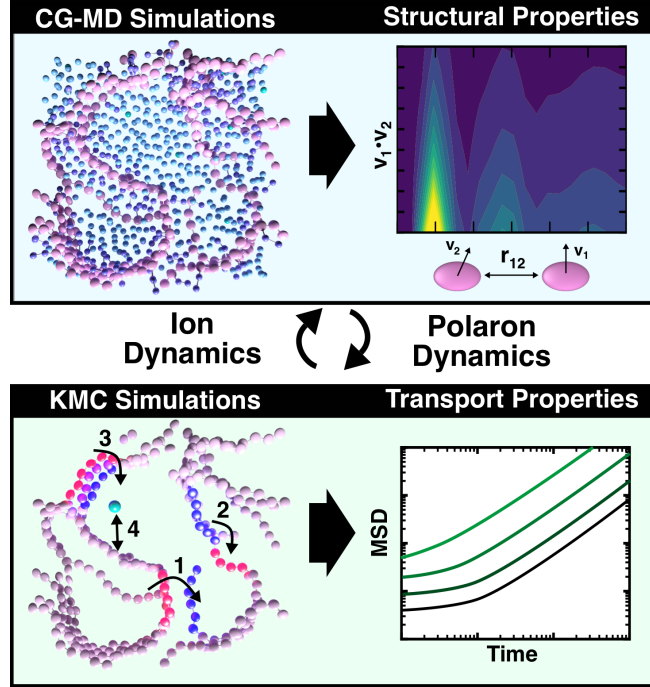
Figure 3.1. Overview of the simulation methodology developed in this work. (Top) Coarse-grained molecular dynamics simulations with anisotropic polymer backbone interactions are used to describe mixed conductor morphology. (Bottom) Kinetic Monte Carlo simulations based on configurationally averaged rates are used to characterize polaron dynamics. The arrows indicate different physical processes relevant to mixed conductors, including 1. interchain charge transfer, 2. intrachain charge transfer, 3. $\pi - \pi$ aggregation, and 4. ion-polaron coupling.

limited to PEDOT-based materials.[103, 107, 108, 109, 138] Although it would be desirable to develop bottom-up CG models for a broader range of OMIECs, this approach is currently impeded by the limited availability of atomistic force-fields for this class of materials. In this respect, the Martini CG force-field is attractive, since it provides access to a library of CG particle types based on water-oil transfer free-energies. Martini has been adapted for simulations of organic bulk-heterojunction morphologies,[3, 4, 5] and simulations of PEDOT-based OMIECs that recapitulate many structure-function relationships within these materials.[103, 107, 108, 109, 138] Here we adopt Martini to describe the side-chain and electrolyte components of the OMIEC, but also introduce a biaxial backbone bead modeled with the Gay-Berne potential to reproduce the anisotropic $\pi - \pi$ interactions germane to conjugated polymers. Coarse-grained Gay-Berne potentials have been previously used to model conjugated polymers[6, 135], but to the best of our knowledge, this work is the first attempt at implementing Gay-Berne governed anisotropic beads in combination with the Martini framework. This phenomenological approach provides a generic CG-MD force-field that can be tuned to systematically interrogate the various designable OMIEC components, including

$\pi - \pi$ interactions, and their effect on morphology and transport processes. As an initial benchmark demonstration of this framework, we have investigated effects of oxidation and hydration levels on the morphology and transport processes of a generic p-type conjugated homopolymer with glycolated side-chains in the presence of an aqueous electrolyte.

3.1 Methods

3.1.1 Mixed Conduction Model

An overview of the mixed conduction workflow is presented in Figure 3.1. The details of each step are described in the subsequent sections but are briefly discussed here to guide the reader. The first step is the simulation of mixed conductor configurations at the coarse-grained level for a given polymer and electrolyte. For this purpose, we have developed a general coarse-grained model that can represent a broad class of OMIEC chemistries (Section 3.1.2). Molecular dynamics is used to characterize the equilibrium behavior of the mixed conductor and ionic diffusivity at different conditions (Sections 3.1.3 and 3.1.4). Based on these configurations, polaron transport is modeled as hopping between localized sites along the conjugated polymer backbones. The rates for hopping are determined by semi-classical Marcus theory in a manner that incorporates the details of charge delocalization, intrachain vs. interchain hopping, and local electrostatic contributions to the site energies (Section 3.1.5). The resulting set of polaron transport processes forms a master equation that we evaluate using the Kinetic Monte Carlo algorithm (Section 3.1.6). This approach yields average polaron mobilities, ionic mobilities, and scattering functions that can be compared with macroscopic measurements, as well as the distribution of transport pathways, their relative resistance, and microscopic ion-polymer configurations that are currently inaccessible by experiments.

3.1.2 Coarse-Grained Molecular Dynamics Model
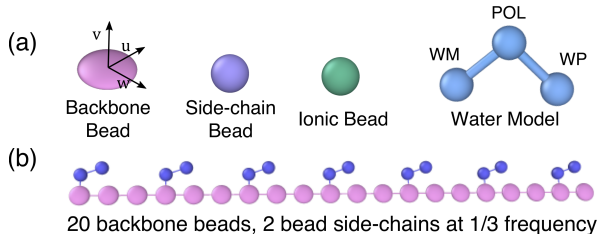
Figure 3.2. Overview of the distinct bead-types used here for coarse-grained simulations of mixed conductors: (a) Building blocks of the generic CG model. For polymers with more complex sequence or side-chain structure, multiple beads might be utilized in each category. (b) The polymer structure used in the present study.

The essential features of a mixed conducting, conjugated polymer electrolyte system consist of four components–conjugated backbone, polar side-chain, ions, and solvent (Figure 3.2). We have developed a flexible top-down framework for modeling each component that can be used to represent a broad range of chemistries relevant to OMIECs. The general features of the model are described here, while parameterization details are described in the following section.

To reproduce the anisotropy of the π-system, the conjugated beads are treated as ellipsoids. The Gay-Berne potential[14] is used to model anisotropic interactions between ellipsoid particles with radii σ_u, σ_v and σ_w and potential minima ϵ_u, ϵ_v and ϵ_w, in the directions $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$, respectively (Figure 3.2a). The Gay-Berne potential is derived by treating these ellipsoids as gaussian volumes, evaluating their overlap integral and mapping the integral on to a Lennard-Jones type potential with some empirical modifications, resulting in the expression

$$
\begin{aligned}
U_{12}(\omega_1, \omega_2, \mathbf{r}_{12}) = \\
4\epsilon_0 \epsilon_{12}^\nu(\omega_1, \omega_2) \epsilon_{12}^{'\mu}(\omega_1, \omega_2, \hat{\mathbf{r}}_{12}) \times \\
\left[\left(\frac{\sigma_c}{r_{12} - \sigma_{12}(\omega_1, \omega_2, \hat{\mathbf{r}}_{12}) + \sigma_c} \right)^{12} \right. \\
\left. - \left(\frac{\sigma_c}{r_{12} - \sigma_{12}(\omega_1, \omega_2, \hat{\mathbf{r}}_{12}) + \sigma_c} \right)^6 \right],
\end{aligned}
\tag{3.1}
$$

where the orientation of the particles is given by their Euler angles, ω_1 and ω_2, and $\mathbf{r}_{12}$ is the displacement vector between ellipsoids. In Eq. 3.1, ϵ_{12} and ϵ_{12}' are energy prefactors, and σ_{12} is the contact radius. The exponents ν and μ are empirical coefficients and σ_c is the minimum contact radius, introduced to control the width of potential well and avoid unphysical overlaps.

In the current work, the conjugated backbone is modeled by uniaxial ellipsoids with the major axes (equivalent $\mathbf{u}$, $\mathbf{w}$) in the plane of backbone and the minor axis ($\mathbf{v}$) perpendicular to the backbone (Figure 3.2a). The parameters ϵ_v and $\epsilon_{u,w}$ govern the interaction minima between conjugated beads along the minor and major axes, respectively. These can be tuned to interrogate OMIECs with variable π-π interactions. The radius σ_v controls the π-π stacking distance and the radii $\sigma_{u,w}$ approximate the conjugated ring radius. The non-bonded interactions of all other components of the system (i.e., side-chains, solvent, and ions) are modeled with isotropic Lennard-Jones potentials based on the Martini force-field. The cross-interactions between the backbone beads and the side-chain, solvent and ions are also modeled as isotropic.

Three types of bonded potentials are used to control the orientation of the conjugated backbone beads with respect to their neighbors. The first potential governs the dihedral rotation about the bond connecting backbone beads ($\mathbf{r}$) with respect to the orientation vectors of each bead ($\mathbf{v}$), as illustrated in Figure 3.3a.[17] This dihedral is modeled using the standard OPLS form and is the analogue of the inter-ring torsions in conjugated systems that dictate the delocalization of polarons along the backbone.[56] Two additional harmonic potentials are used to model the bending interaction between bonded backbone beads and
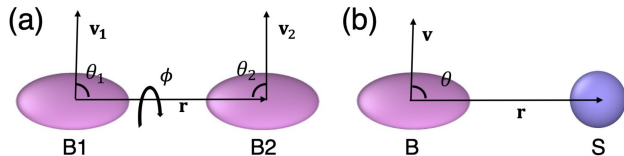
Figure 3.3. Illustration of the bonded interactions governing the orientation of backbone beads. (a) Dihedral potential between two backbone beads based on their orientation vectors ($\mathbf{v}$). (b) Bending potential between the backbone bead normal and the side-chain bead.

between backbone beads and bonded side-chain beads, respectively. In each case, the potential is based on the angle between the bond vector, $\mathbf{r}$, and the orientation vector, $\mathbf{v}$, of the backbone bead (Figure 3.3a-b). The force constant for the backbone-backbone bending term can be modified to vary the stiffness of the backbone. Likewise, the equilibrium angle of the backbone-sidechain bending term can be modified to model side-chains bonded to sp^3 vs sp^2 positions along the backbone. The derivation of the forces and torques associated with these potentials can be found in the 3.A.

The remaining three components, namely, the polymer side-chain, ions, and solvent are modeled as isotropic beads using the Martini force-field.[181] The Martini parameters have been developed to reproduce experimental partition coefficients for biomolecules and organic molecules using a relatively small number of bead types that typically represent four heavy atoms. Martini bead classification is based on polarity, making it straightforward to vary the polarity of side-chains and solvent, which alters the penetration of solvent and ultimately the transport properties. Each Martini bead is modeled as a Lennard-Jones particle with harmonic bonded interactions and cosine bending terms, where applicable.

3.1.3 Parameterization

The CG framework described in the previous section is capable of modeling a broad range of conjugated polymer electrolyte systems through the appropriate choice of Martini bead-types, and at the same time, captures essential features of π-conjugated polymers including anisotropic backbone interactions, controllable planarity, and stiffness. As a baseline benchmark of how this model performs, we have implemented a parameterization representating a p-type conjugated homopolymer with polar side-chains in the presence of an aqueous electrolyte (Figure 3.2b). The Martini force-field parameters were developed to reproduce the experimental oil-water transfer free energy and density of molecules. As the goal for this work is to develop a phenomenological framework to model various designable attributes of OMIECs, the parameters associated with the backbone potentials are treated as tunable variables. As a case study to illustrate the framework, we have chosen representative values for these parameters based on existing literature. A single backbone bead type was utilized with $\sigma_{u,z}$ and σ_v parameters chosen to yield repulsive core dimensions of 5 Å along the major axes, which is the average radius of a thiophene ring, and 3 Å along the minor axis of the ellipsoid to yield a typical $\pi-\pi$ stacking distance. To obtain an anisotropy in the binding en-

ergy of ~5, which is typical of common repeat units,[65] ϵ_v was chosen to yield a 1.2 kcal/mol minimum energy in the π-stacked orientation and $\epsilon_{u,w}$ of 0.25 kcal/mol minimum in side-by-side orientation, similar to coarse-grained Gay-Berne model developed for sexithiophene. [6] These non-bonded parameters result in backbone beads with dimensions and interaction energies comparable to typical aromatic repeat units. For the cross-interactions, the backbone bead was modeled as a "C3" Martini bead of intermediate hydrophobicity. An equilibrium angle of 180° and force constant of 7.5 kcal/mol were used for the bending term associated with bonded backbone beads, which results in a persistence length between 4-60 Å for the range of hydration and oxidation levels simulated here (Table 3.7 & 3.8). The backbone dihedral term was chosen to exhibit two minima at 0° and 180° with an interconversion barrier of 7.5 kcal/mol, which results in dihedral angle distributions with a standard deviation of ~4.63-5.18° in the simulations. Side-chains composed of two Martini beads (~8 heavy atoms in an atomistic representation) were included on every third backbone bead (Figure 3.2b). The side-chain bead of type "P1" was used to approximate the CG force-fields developed for conventional OMIEC polyethylene glycol (PEG) side-chains.[139] An equilibrium angle of 90° and force constant of 3.0 kcal/mol were used for the backbone-sidechain bending term. The equilibrium angle of 90° is meant to mimic a side-chain attached to an sp^2 backbone atom. The equilibrium morphologies of this system were interrogated as a function of hydration level and oxidation level, then subsequently utilized in charge transport studies.[46, 107] These simulations consisted of 100 chains with 20 repeat units. This represents an intermediate degree of polymerization compared with published OMIECs ranging from 10[92, 164] to 60[35, 42] repeat units. Oxidation levels of 100%, 50%, 20%, 10%, and 5%, were simulated by equally distributing a positive charge of 1, 0.5, 0.2, 0.1 and $0.05q$, respectively, along the polymer backbone. The decision to equally distribute the charge along the backbone is based on the assumption that polaron transport is fast relative to polymer dynamics and follows previous work.[107] A more sophisticated treatment of the coupling between polaron and polymer dynamics using fluctuating charge models is a strategy for possibly improving this description in the future. In each case, counter-ions with a unit negative charge and Martini type "Qa" were introduced to maintain charge neutrality. Water is used as the solvent and represented by the Martini polarizable water model.[181] A list of all Martini bead types utilized here and their associated parameters are supplied in Table 5.1.

3.1.4 Molecular Dynamics

LAMMPS was used to perform all molecular dynamics simulations.[130] All simulations used a 10 fs integration time step, Velocity-Verlet integration, and periodic boundary conditions. Long-range electrostatics were modeled using the particle-particle-particle-mesh/stagger (PPPM/stagger) algorithm.[25, 63, 115] Long-range electrostatics, Lennard-Jones interactions and Gay-Berne interactions were truncated at 12 Å. All simulations were initialized from diffuse configurations, using a cubic grid to place molecules in random orientations without overlap. The velocities were initialized from a uniform distribution obtained with a random seed value and scaled to give the correct kinetic energy. The simulations were first relaxed in the NVE ensemble with restrained displacements of 0.1 Å per time step for 10 ps. This was followed by a simulation in the NVT ensemble at 300 K while linearly

rescaling the simulation box over 2 ns to obtain a condensed system with a density of 1 g/cc. The density was then equilibrated in the NPT ensemble for 100 ns at 300 K and 1 atm. After the NPT simulations, the system was further equilibrated at the average NPT density in the NVT ensemble at 300 K for 100 ns, followed by a 100 ns production run in the NPT ensemble at 300 K and 1 atm. These simulation times were chosen to ensure that the polymers diffused 2-3 times its radius of gyration, on average, during the equilibration and production periods, respectively (end-to-end vector autocorrelations are also presented in the 3.B). In both NVT and NPT simulations, the Nosé-Hoover thermostat and barostat were employed using the modified form proposed by Martyna, Tobias, and Klein, as implemented in LAMMPS [102], with a relaxation time constant of 0.15 ps and 1.5 ps for the thermostat and barostat, respectively. In all the simulations, the thermodynamic data and coordinates were sampled at 1 ps. The ASPHERE package in LAMMPS was used for all Gay-Berne force and torque evaluations and numerical integrations.[22] The SHAKE algorithm [143] implemented as a part of the RIGID package in LAMMPS was used to constrain the bonds in the water molecules. The additional bonded potentials to control the orientation of neighboring beads were implemented as custom classes in LAMMPS by our group.

All simulations contained 100 polymers, with 20 backbone beads each. Depending on the backbone bead oxidation level, the number of anionic beads in the simulation box varied to ensure zero net charge on the system. For the oxidation level studies with backbone bead charges of 1.0, 0.5, 0.2, 0.1, and 0.05, the total number of anionic beads in the system were 2000, 1000, 400, 200, and 100, respectively. All the oxidation level studies were carried out at 50 wt% hydration level with 3400 Martini water molecules (i.e., 13600 actual water molecules). For the hydration level studies, the backbone bead oxidation level was fixed at the intermediate value of 0.2 and the simulation box contained 400 anionic beads. The hydration level studies were carried out at 20, 35, 50, 65, and 80 wt/wt% hydration with 850, 1831, 3400, 6134, and 13600 water molecules.

3.1.5 Charge Transfer Rate Parameters

Charge transport was modeled here on the basis of configurations generated from the CG-MD trajectories. In our model, each backbone bead is parameterized to represent a single conjugated unit (i.e., a ring system without internal flexible dihedral degrees of freedom), thus sites consisting of one or more contiguous backbone beads constitute the relevant polaronic basis states for charge transfer. The charge transfer rates between sites were parameterized using semi-classical Marcus theory,[96] along similar lines to what has been described in prior work in the context of conjugated polymers.[18, 55, 131, 144] The rate for charge transfer between sites on different chains was modeled in the weak coupling limit as

$$\omega_{ij} = \frac{2\pi}{\hbar} \frac{J_{ij}^2}{\sqrt{4\pi\lambda_{ij}k_B T}} \exp\left\{-\frac{(\Delta E_{ij} - \lambda_{ij})^2}{4\pi\lambda_{ij}k_B T}\right\}, \tag{3.2}$$

where J_{ij} is the electronic coupling between the sites, λ_{ij} is the reorganization energy, ΔE_{ij} is the energy difference for a charge occupying the site j vs site i, T is the system temperature,

and k_B is the Boltzmann constant. The radial dependence of J_{ij} was modeled as exponentially decaying with distance as

$$J_{ij} = J_{inter}\exp\left\{-\alpha(r_{ij} - r_0)\right\},\tag{3.3}$$

where α is the decay factor, r_0 is the reference separation for J_{inter}, and r_{ij} is the inter-site distance.[9] The inter-site distance was calculated as an average over the nearest-separation for the beads in each site according to

$$r_{ij} = \frac{1}{2}\left(\frac{\sum_{m\in i}\min_{n\in j}(r_{mn})}{l_i} + \frac{\sum_{n\in j}\min_{m\in i}(r_{nm})}{l_j}\right),\tag{3.4}$$

where the indices m and n run over the beads in sites i and j, respectively, and l_i and l_j are the number of beads in each site. This expression captures the correct physical behavior as two sites slide laterally by one another, with a maximum occurring at perfect registration. Likewise, as sites move apart along the center of mass displacement vector, the standard exponential decay of the wavefunction overlap is reproduced.

Eq. 3.2 must be modified for evaluating the charge transfer rates between neighboring sites on the same chain due to the potentially large coupling between bonded conjugated rings. Specifically, the electronic coupling between neighboring backbone beads increases with the alignment in their π-orbitals and can be expressed in terms of cosine function in the dihedrals between these beads ($J_{intra}\cos(\phi_{ij})$).[56] This coupling leads to a modulation of the intramolecular hopping barrier due to orbital mixing,[142] which is approximated as

$$\omega_{ij} = \nu\exp\left\{-\frac{(\Delta E_{ij} - \lambda_{ij})^2}{4\pi\lambda_{ij}k_BT} + \frac{|J_{intra}\cos(\phi_{ij})|}{k_BT}\right\}.\tag{3.5}$$

In this expression, the semi-classical Marcus activation energy is recovered when the backbone beads are perpendicular, whereas the barrier can be lowered by up to a factor of J_{intra} when they are co-planar. The definition of variables λ_{ij} and ΔE_{ij} are the same as in Eq. 3.2, and ν is a prefactor associated with the nuclear frequency,[15] and was set to $10^{15}s^{-1}$.[142]

A basis of hopping sites localized on individual beads neglects potential delocalization effects and is inappropriate for the evaluation of the Marcus parameters associated with Eqs. 3.2-3.4. To account for delocalization, rates calculated using Eq. 3.5 in an individual bead basis were used to partition contiguous beads into sites comprised of one or more backbone beads. We used the rates, rather than the dihedrals, to demarcate sites because electrostatic factors (e.g., a nearby ion) can also contribute to site localization. Details of site formation are covered in the subsequent section, however this detail affects the manner in which the Marcus parameters, λ_{ij} and ΔE_{ij}, are evaluated. The reorganization energy, λ_{ij} is the energy associated with the rearrangement of the nuclear coordinates due to charge transfer and consists of internal, λ_{int}, and external, λ_{ext}, contributions. The internal reorganization energy is associated with the nuclear rearrangement of the sites i and j, given by an average of individual internal reorganization energies λ_i and λ_j, which vary inversely with the size of the site. The values for internal reorganization energy associated with sites of varying number of backbone beads were obtained by fitting a quadratic function to the values for polypyrrole. [142]. Polypyrrole was a adopted as a useful reference from a previous charge

transport study that exhibits similar trends for reorganization energy and ionization potential to polythiophene.[182] More generally, these are potentially tunable quantities that could strongly affect structure-function relationships. The external reorganization energy was approximated using the model introduced by Marcus,[95, 97] obtained by treating the sites as spheres with constant charge density and the solvent as a continuum dielectric,

$$\lambda_{\mathrm{ext,ij}} = \frac{e^2}{4\pi\epsilon_0}(\frac{1}{\epsilon_{\mathrm{opt}}} - \frac{1}{\epsilon_s})(\frac{1}{2r_{\mathrm{i}}} + \frac{1}{2r_{\mathrm{j}}} - \frac{1}{r_{\mathrm{ij}}}), \tag{3.6}$$

where r_{ij}, r_{i}, r_{j}, ϵ_{opt} and ϵ_s are the distance between the sites, the radii of the sites, the optical frequency dielectric constant, and the static dielectric constant, respectively. r_{ij} is calculated based on the distance between the center of mass of the sites and the radius of a site is taken to be an average of the radius along the backbone and perpendicular to the backbone. For sites that significantly extend along the polymer backbone, the spherical approximation fails, and the expression can produce negative values which were set to zero in this study. Thus, the external reorganization energy decreases with increasing site delocalization until reaching zero. Mixed conducting systems are heterogenous, composed of organics and aqueous solvent, and hence, an intermediate dielectric constant of $\epsilon_s = 10$ was used. We note that more sophisticated treatments of the external reorganization energy and local dielectric are compatible with the current KMC framework,[144] however we defer treatment of these to future work.

The site energy difference is given by $\Delta E_{\mathrm{ij}} = E_{\mathrm{j}} - E_{\mathrm{i}}$, where the individual site energies include contributions from the ionization potential and the local electrostatic environment of each site. The contribution from the ionization potential was modeled as a linear function in inverse length fitted to the values for polypyrrole. [142] The electrostatic energy contribution to each super-site was modeled on the basis of the CG-MD configurations as

$$E_{\mathrm{elec,i}} = \sum_{m \subset \mathrm{i}} \sum_{n \not\subset \mathrm{i}} \frac{q_m q_n}{4\pi\epsilon_s\epsilon_0 r_{mn}} \tag{3.7}$$

where m denotes beads within the site and n denotes beads outside the site. In the charged state, the site has an excess unit charge equally distributed along the polymer (i.e., $1/l_{\mathrm{i}}$ for site length l_{i}), while the deficit of unit charge is distributed to all other backbone beads in the system. These electrostatic energies were evaluated in LAMMPS by rerunning the CG-MD configurations with each site evaluated in the corresponding oxidation state with long-range interactions and periodic boundary conditions, including the contributions from all beads (i.e., ions, solvent and other polymers). The charge transfer parameters used in all KMC simulations are summarized in 3.E.

3.1.6 KMC Framework

Charge transfer is modeled here from the perspective of a single explicit polaron in the mean-field of the other polarons that are present on the polymer. This assumption is reflected in the rate expressions described above, where the electrostatic contributions to the site energies (ΔE_{ij}) are calculated with the residual polymer charge equally distributed to all

backbone beads. This approximation neglects polaron correlations and potential coulomb blockade effects that may be relevant at high charge densities, but it is an important reference for establishing how OMIEC morphology and design features affect charge transport. For a single explicit polaron capable of occupying a pre-defined set of polaron sites, the evolution of the charge dynamics is governed by a master equation composed of coupled first-order differential equations

$$\dot{p}_i = \sum_j k_{ji} p_j - \sum_j k_{ij} p_i, \tag{3.8}$$

where p_i is the probability of the polaron occupying site i at time t, and k_{ij} is the charge transfer rate between sites as defined in the previous section. Both transient and steady-state solutions to this master equation can be calculated using the kinetic Monte Carlo (KMC) algorithm, which consists of simulating an ensemble of Markov chains with transitions based on Eqs. 3.8.[117] This approach is advantageous because it yields both correct steady-state occupation statistics and trajectories that can be analyzed to characterize details of the conduction network. For instance, polaron diffusivities are obtained based on the mean-squared displacement of charges, and graphical analysis can be used to identify resistive bottlenecks in the transport network. Likewise, the KMC framework can be extended include the treatment of more than one explicit polaron.

The definitions of available polaron sites, their site energies, and associated rates are required for the application of KMC. For typical conjugated polymers, polarons are delocalized over multiple backbone units, which makes sites composed of one or more backbone beads in our CG model the relevant basis states for KMC. A typical approach for identifying these delocalized sites from MD simulations is to break each polymer backbone into segments based on deviations in the backbone dihedrals that exceed a threshold [142]. However, in OMIECs polaron-ion electrostatic interactions also contribute to site localization. To account for this, we use the rates of intramolecular transport, as predicted by Eq. 3.5 for sites composed of individual backbone beads, as the basis for designating sites. In particular, the intramolecular rates reflect contributions from both dihedral deviations as well as the differences in electrostatics along the backbone. After calculating the charge transfer rates between neighboring beads, the sites are defined based on contiguous groups of backbone beads with a transfer rate above $10^{14} s^{-1}$. These sites are the reference states for Eq. 3.8 and charge transfer between sites is modeled using Eq. 3.2 with size dependent input parameters as described in the previous section.

In a KMC framework, sites and transition rates are held fixed while evaluating charge transport, whereas in the real system, transition rates are a function of time. Rate fluctuations that occur on timescales that are shorter than the characteristic charge transfer rates within the master equation can be incorporated by averaging the rate calculations over a commensurate interval of MD configurations. To characterize relevant short-timescale fluctuations, we evaluated the autocorrelation functions of the dihedrals (ϕ_{ij}), bonds (r_{ij}) and electrostatic energies ($\Delta E_{elec,ij}$), shown in Figure 3.16 and Figure 3.17. We observe that all of these factors decorrelate within 1 ps, and thus use 20 configurations uniformly sampled across a 2 ps interval to average these fluctuations and their contributions to the charge transfer rates.

On timescales longer than the characteristic charge transfer rate, the assumption of a fixed master equation can also break down due to slow diffusion processes or the periodicity of the system (e.g., the polaron traverses many box lengths and the system remains unphysically correlated with the initial state). Webb et al. have recently addressed this issue in the context of polymer electrolytes by incorporating site refresh events into the KMC Markov chains [179]. In our system, we introduced site refresh events when the charge travelled 10 times the box length, which is between 380-500 nm in the systems studied here. The site refresh event consisted of *(a)* rotating the configuration by a random angle about each cartesian axis, in a random order, *(b)* randomly selecting a polaron site from the configuration as the occupied site, and *(c)* translating the occupied site to the position of the originally occupied site. A refresh criteria for addressing polymer diffusion was not used, since we observed that all of our simulations of dense polymer networks reached the Fickian regime (for charge transport) well within the diffusion timescale of the polymers and ions. It is possible that for less viscous systems, a site refresh related to the fastest diffusion timescale would be required.

For each of the hydration and oxidation conditions that were studied, a total of 60 uncorrelated configurations were characterized by KMC. Specifically, 20 configurations were sampled every five ns from three independent MD simulations. These configuration counts exclude the additional 0.1 ps spaced configurations used for averaging charge transfer rates. For each configuration, 20 KMC charge transport trajectories were simulated for 150 ns. The mean square displacement (MSD) of the charge was calculated for each configuration and averaged over all 60 configurations. The self-diffusion coefficient was obtained from the MSD using the relation,

$$D = \lim_{t \to \infty} \frac{1}{6} \frac{d}{dt} \langle \|\mathbf{r}(t) - \mathbf{r}(0)\|^2 \rangle, \tag{3.9}$$

where $\mathbf{r}(t)$ is the position vector of the center of mass of the site occupied by the charge at time t. The angular brackets in Eq. 7.17 indicate an average over all polarons and time origins. Numerically, Eq. 7.17 was evaluated by identifying the Fickian regime of the MSD curves (i.e., where the curves exhibited a consistent log(MSD) vs log(t) of one), then performing a least-squares linear regression to the interval from 10ns-100ns. The mobility was obtained from the diffusivity using the Einstein relation,

$$\mu = \frac{eD}{k_B T}, \tag{3.10}$$

where e is the unit electron charge.

3.2   Results and Discussion

3.2.1   Effect of Hydration Level

To establish the structural changes that accompany hydration of our model OMIEC, we characterized the joint-probability distribution for $\pi - \pi$ stacking and radial separation between the polymer backbones as a function of hydration level while fixing the oxidation level
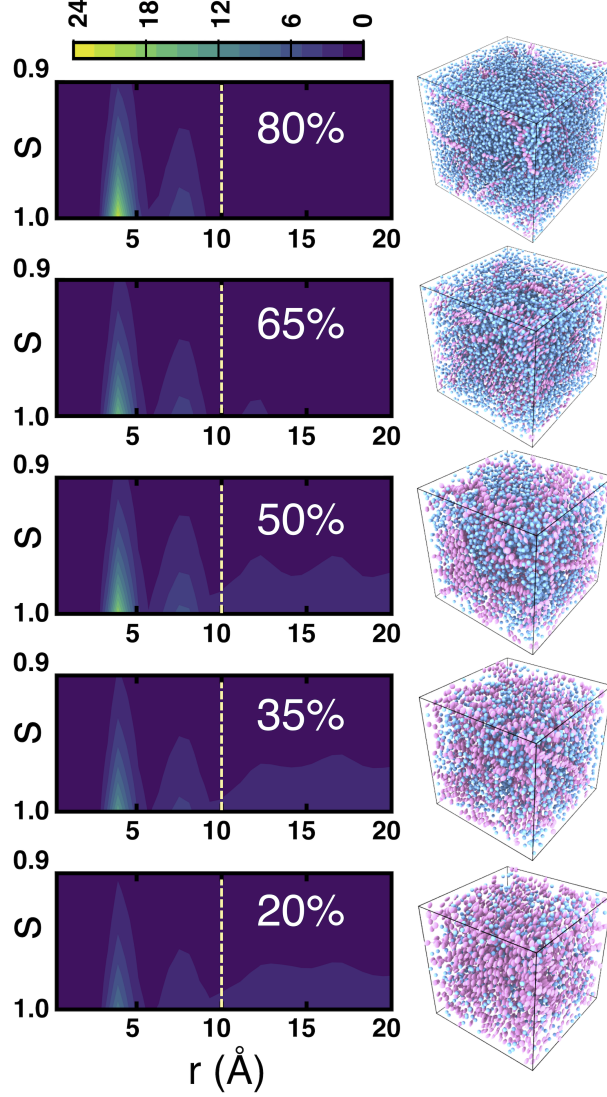
Figure 3.4. $\pi - \pi$ stacking probability presented as a joint probability distribution of distance between the backbone beads on the x-axis and $\pi - \pi$ alignment on the y-axis, as a function of hydration level. The yellow dashed line indicates the peak for lamellar packing. Trajectory snapshots are shown to the right, where the pink and blue beads are backbone beads and water, respectively. Other components are omitted for clarity.

at the intermediate value of $+0.2e$ per backbone bead (Figure 3.4). In these distributions, the x and y axes represent the distance between backbone beads on different chains (r_{ij}) and their $\pi - \pi$ alignment ($s_{ij} = \mathbf{v}_i \cdot \mathbf{v}_j$), respectively, while the color represents their joint probability distribution given by

$$g(r, s) = \frac{V}{4\pi r^2 N_p} \sum_{i=1}^{N} \sum_{j=1}^{N} \delta\left(r - r_{ij}\right) \delta\left(s - s_{ij}\right), \tag{3.11}$$

47

where N_p is the total number of pairs of backbone beads in the system, excluding pairs within the same molecule. These distributions are averaged over configurations parsed from multiple independent trajectories after equilibration of the density, potential energy, and translational diffusion of the chains, and thus reflect an effort to sample authentic equilibrium configurations within the limits of the MD simulations.

Prominent $\pi - \pi$ stacking is apparent at all hydration levels (i.e., $s_{ij} > 0.9$ in all of the peaks), which is consistent with conjugated polymers in general and the backbone anisotropy of our model OMIEC. All of the systems also exhibit peaks at multiples of ~4 Å, which is indicative of aggregates of $\pi - \pi$ stacked polymers. In contrast, some lamellar organization occurs at low hydration levels (~10 Å) but overall it is weak in comparison with the nearby higher-order $\pi - \pi$ stacking peak at this oxidation level. We also observe an abrupt transition between the 50% and 65% hydration levels that corresponds to the dissolution of the polymer aggregates. Specifically, at hydration levels of 65-80%, we do not find any features at distances greater than 1.5 nm, as the polymers are separated by large clusters of water, while at hydration levels $\leq 50\%$ dense percolating polymer networks are formed. We also observe that the $\pi - \pi$ alignment increases with increasing hydration level.[107] This reflects the configurational freedom of the smaller aggregates to reorganize into highly aligned structures, whereas packing defects frustrate chain registration at low hydration levels. Interestingly, at the intermediate hydration level of 50%, both percolation and relatively strong $\pi - \pi$ alignment are achieved. This structural analysis suggests a potential trade-off between chain alignment and the number of conductive pathways as the hydration level is reduced, wherein highly hydrated morphologies (e.g., $\rho_{H_2O} = 50\%$) exhibit relatively few pathways that are highly aligned while less hydrated morphologies exhibit a larger number of conductive pathways that are on average more disordered.

To clarify the distinct spatial distributions of the electrolyte phase within each system, we also analyzed the size and distribution of water clusters (Figure 3.5). Clusters were defined using nearest-neighbor hierarchical clustering with a threshold of 5.8 Å, which is 10 % above the Van der Waals equilibrium distance for the Martini water model (i.e., all water molecules in a cluster are within the threshold of at least one other member of the cluster, and molecules in distinct clusters are separated by at least the threshold distance). The observed cluster distributions consist of either small water clusters with very few molecules, or a single large percolating cluster. Interestingly, no moderately sized clusters are observed, even at low hydration levels. Rather, for all hydration levels $\geq 35\%$, the electrolyte phase percolates throughout the volume of the polymer. This indicates that the electrolyte spontaneously percolates throughout the polymer phase through channel formation, rather than through the coalescence of gradually increasing water clusters. In contrast, a recent computational study of PEDOT:Tos[107] reports a gradual increase of water cluster size through coalescence which can be attributed to the tosylate dopants dispersed throughout the system.

To determine the impact of these distinct morphologies on charge transport, KMC simulations were performed on each system (Figure 3.6). For the 65% and 80% hydration level simulations, sub-diffusive behavior is observed, which is consistent with the dissolution of the polymers and lack of macroscopic percolation. These cases are excluded from further discussion of charge transport, however we highlight the recent description of scaling laws in this concentration regime by Sing et al [13]. Additionally, a comparison was performed of KMC
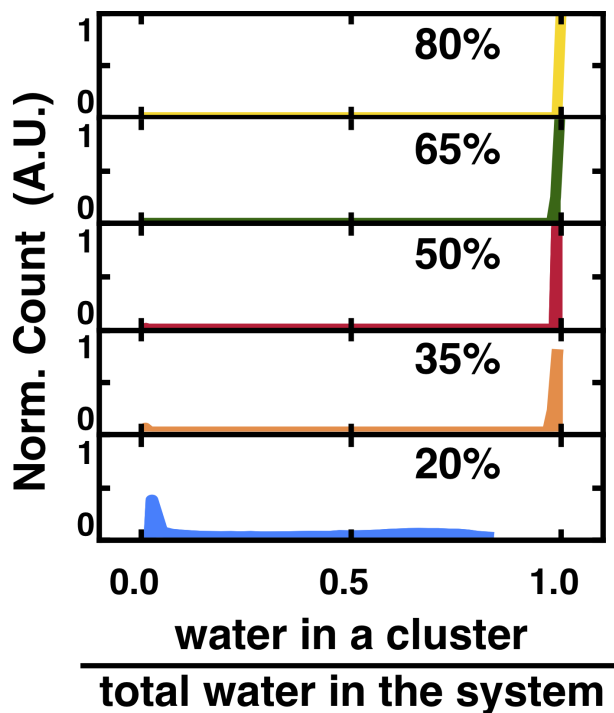
Figure 3.5. Distribution of water clusters for different hydration levels. The number of molecules in a cluster normalized the by total number of molecules is plotted on the x-axis and the number of its occurrences normalized by the maximum number of possible occurrences is plotted on the y axis.
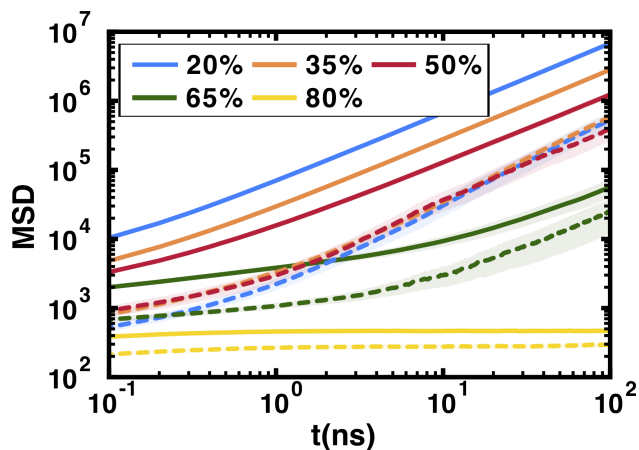
Figure 3.6. Mean square displacement of charge with intermolecular rates excluding (solid lines) and including (dashed lines) electrostatic energy component at different hydration levels. Standard deviations calculated across the sampled configurations are shown in the shaded regions.

results with and without electrostatic contributions to the site energy differences to elucidate

the impact of ion-polaron coupling on the charge transport. In all of the simulations, the polarons are delocalized along the entire polymer chain based on the intramolecular rates (as described in Section II.F) and each polymer served as a single site for intermolecular charge transport. Comparing the polaron mean-squared displacement (MSD) in the two cases (Figure 3.6), illustrates that including ion-polaron electrostatics suppresses charge transport at all hydration levels. This is intuitive, as all else being equal, electrostatics introduce energetic disorder that reduces charge transport rates. However, the results also reveal that the ion-polaron electrostatics suppress charge transport more strongly at low hydration levels, with the consequence that all of the dense systems exhibit comparable polaron mobilities when electrostatics are included. Comparing the quantitative trends within each series of simulations reveals several distinct factors that contribute to the observable polaron mobility.

When ion-polaron electrostatics are excluded from the simulations, the polaron diffusivity monotonically decreases with increasing hydration level. Based on the previous structural analysis, decreasing hydration leads to relatively decreased $\pi - \pi$ alignment, however this is countervailed by an increase in the number of conductive pathways as the polymer network densifies. These two factors are reflected in the distributions of charge transfer rates (Figure 3.7a) and site connectivity (Figure 3.7b) within each system. The impact of $\pi - \pi$ alignment is reflected in a peak at high rates within the rate distributions of the 35% and 50% systems, whereas the impact of density is reflected in an increase in the site connectivity of the 20% system. Based on the observed trends in polaron mobility (Figure 3.6), site connectivity outweighs the impact of $\pi - \pi$ alignment with respect to increasing polaron mobility. As the connectivity increases, the distinct sites visited both at short (Figure 3.7c) and long (Figure 3.7d) timescales increase, representing relatively reduced trapping and percolative transport in the highest mobility 20% system. In contrast, at the 50% hydration level, many trajectories involve only a small fraction of sites which is indicative of poor connectivity within the transport network despite high charge transfer rates within localized subnetworks of the morphology.

The forgoing examination of charge transport when polaron-ion electrostatics are excluded illustrates the complex interplay between morphology and site connectivity that underpins the ostensibly straightforward trend that increasing hydration reduces polaron mobility. When ion-polaron electrostatics are included in the KMC simulations, these factors are also present, but they compete with the site energy disorder associated with the spatial distribution of the electrolyte. Comparing the rate distributions with and without ion-polaron electrostatics, reveals that electrostatic heterogeneity suppresses the charge transport rates in all systems and the peaks at high rates observed for the 35% and 50% hydration levels are greatly reduced (Figure 3.7a). Including ion-polaron electrostatics also results in a dramatic reduction in the site connectivity of the 20% hydrated system, such that all of the dense polymer systems exhibit comparable connectivities (Figure 3.7b). This reduction in overall connectivity also leads to a reduction in the fraction of sites that participate in charge transport relative to when ion-polaron electrostatics are excluded (Figure 3.7c-d). The contraction in the number of sites visited reflects trapping of the polaron to clusters of sites with similar electrostatic energies (Figure 3.20). This is further evidenced by the occurrence of sub-diffusive behavior in all of the dense networks when ion-polaron electrostatics are included in the simulations (Figure 3.6). The consistent picture that emerges
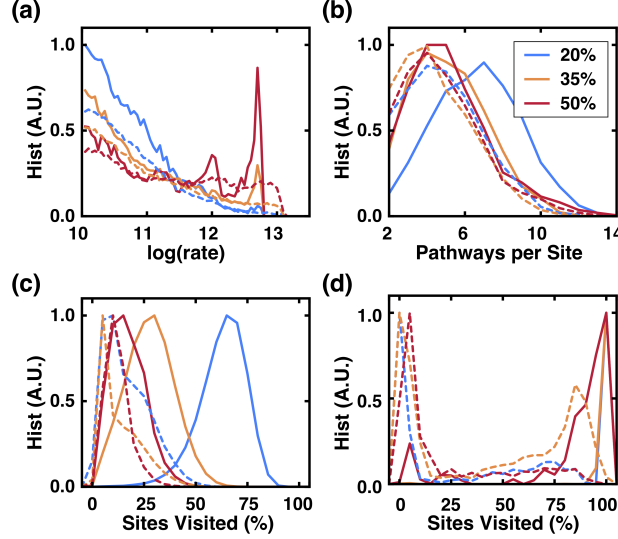
Figure 3.7. Distribution of (a) intermolecular rates and (b) pathways per site, (c) distinct sites visited over 500 jumps and (d) distinct sites visited over last 75ns trajectory, when excluding (solid lines) and including (dashed lines) electrostatic energy component at different hydration levels. The distributions (a-b) are normalized by the maximum value across all plots and (c) is normalized by the maximum value within each plot.

is that ion percolation within the polymer phase results in localization of charge transport to a subnetwork of sites with similar electrostatic environments. Remarkably, these subnetworks are similar in all of the dense systems with the net result that the polaron mobility is relatively constant with respect to hydration level. This effect has also been observed experimentally in some OMIECs, where the measured hole mobility is insensitive to swelling caused by varying the anion in the electrolyte [24] and varying the electrolyte concentration [149].

To directly interrogate the changes in electrostatics that accompany ion percolation within the polymer, we also analyzed the distribution of electrostatic energies of the polaron sites and the ion-polymer backbone RDF at each hydration level (Figure 3.8). Comparing the RDFs, all systems exhibit a prominent nearest-neighbor peak and periodic features with a spacing of ~4.5 Å, which is consistent with ion intercalation within the polymer aggregates. At high hydration levels (i.e., $\rho_{H_2O} \geq 65\%$) a small decrease in the first RDF peak is accompanied by an increase in the later peaks, which is due to a weak preference for ions to remain external to the more highly ordered polymer aggregates after dissolution (Figure 3.8 (c-e)). Although all systems display a similar degree of ion mixing based on the RDFs[111], the 20% system exhibits a much broader distribution of electrostatic site energies compared with the 35% and 50% hydrated systems (Figure 3.8b). This can be rationalized by the earlier observation that the electrolyte phase fails to fully percolate in the 20% system, which leads to electrostatic heterogeneity and limited electrostatic screening. The resulting elec-

trostatic induced site energy disorder suppresses charge transfer rates and is consonant with the reduced site connectivity of the 20% system when ion-polaron electrostatics are included.

In most mixed conducting applications, the mobility of ions is critical to mediate reversible doping of the semiconducting polymer. Moreover, ion diffusion is typically much slower than electron diffusion[78, 150] and is the rate-limiting step in (de)doping. The ionic mobility was thus characterized in each system to determine the complementary impact on ion dynamics as the hydration level is varied (Table 3.1). It is established that Martini dynamics are accelerated by approximately a factor of four in comparison with all-atom simulations [99, 181]. Thus, the reported ionic mobilities have been scaled by a factor of 0.25 relative to the raw MD trajectories. The polaron mobility is between one and two orders of magnitude larger than the ionic mobility for the dense polymer systems (i.e., $\rho_{H_2O} \leq 50\%$), however at higher hydration levels the electronic mobility is limited by the dissolution of polymer aggregates and the loss of macroscopic percolation pathways. In contrast, the ionic mobility monotonically increases with respect to hydration level [146], albeit with distinct behaviors above and below the electrolyte percolation threshold. Specifically, for $\rho_{H_2O} \geq 35\%$, the ion diffusivity increases approximately linearly with respect to hydration level, while the 20% system exhibits an abrupt three-fold reduction in ion mobility. This is consistent with the loss of electrolyte percolation in the 20% system, resulting in macroscopic ion diffusion being restricted by reorganization of the polymer phase. In comparison, a recent computational study of PEDOT:TOS[107] reports an exponential increase in ion diffusivity with hydration level, while a recent experimental study on OMIECs with varying side-chains report a sub-exponential trend.[150] These results highlight the importance of establishing co-percolating electronic and ionic phases within the mixed conductor, and suggest that polymer design strategies focusing on electrolyte management may be critical to striking the optimal balance. We also note that since comparable electronic mobilities were observed in all of the dense systems, we do not observe a significant trade-off between ionic mobility and electronic mobility in this polymer. Specifically, by maintaining hydration levels between 35-50%, through either side-chain or morphology engineering, optimal electronic mobility and ionic mobility within an order of magnitude of the bulk electrolyte can be achieved.

Table 3.1. Summary of transport coefficients at different hydration levels with a fixed oxidation level of 0.2 per backbone unit.

$\rho_{H_2O}^a$	μ_+^b	$\mu_{+,ES}^c$	μ_-^d
20 %	14.30	1.03	0.0081
35 %	6.08	1.33	0.0224
50 %	2.94	0.87	0.041
65 %	0.13	0.07	0.0696
80 %	-	-	0.1004

[a] The hydration level in weight percent.
[b] Polaron mobility with excluded electrostatics, $10^{-3}\mathrm{cm}^2\mathrm{V}^{-1}\mathrm{s}^{-1}$.
[c] Polaron mobility with included electrostatics, $10^{-3}\mathrm{cm}^2\mathrm{V}^{-1}\mathrm{s}^{-1}$.
[d] Anion mobility, $10^{-3}\mathrm{cm}^2\mathrm{V}^{-1}\mathrm{s}^{-1}$.

Figure 3.8. (a) Ion-backbone RDF and (b) histogram of the electrostatic site energy differences as a function of the hydration level
. Snapshots of the system showing ions (green) and polymer (pink) for hydration levels of (c) 20 %, (d) 50 % and (e) 80 %. The site-energy distributions are normalized by the maximum value across all plots.

3.2.2 Effect of Oxidation Level

In this section, we inspect the changes in the structural and transport behaviors of the model OMIEC upon changing the oxidation level of the polymer, while keeping the hydration level fixed at 50 %. To help orient the reader, the same analyses are performed as in the previous section and discussed in the same order.

To characterize the impact of oxidation level on morphology, the joint probability distribution for $\pi - \pi$ stacking and radial separation between the polymer backbones were analyzed (Figure 3.9). No features are observed at very high oxidation levels ($\geq +0.5e$) as the polymer completely dissolves due to repulsive inter-chain electrostatics. Irreversible morphology changes due to OMIEC swelling during constant operation have been reported experimentally.[111] At lower oxidation levels ($\leq +0.2e$), the polymers assemble into well stacked $\pi - \pi$ aggregates with a prominent nearest-neighbor peak and additional peaks at multiples of ~4 Å. Although all of the dense systems are well-ordered, a small but persistent increase in $\pi - \pi$ alignment is apparent in the first two peaks as the oxidation level decreases. Additionally, for oxidation levels $\leq +0.2e$, a prominent lamellar stacking peak occurs at ~10 Å and larger aggregates are formed[107], as indicated by the higher probabilities at large radial separations. These distinct changes in morphology indicate that a combination of repulsive inter-chain electrostatics and electrolyte percolation affect the inter-chain disorder and aggregate dimensions as the oxidation level increases. This interplay is also qualitatively apparent in the swelling behavior of many OMIECs upon doping [35, 53, 123].

The change in OMIEC morphology with respect to oxidation is also reflected in the distribution of the size of water clusters (Figure 3.10). At hydration levels of 50%, all of the systems exhibit a large percolating electrolyte phase; however, the probability of forming a single large percolating cluster decreases as the oxidation level decreases. This reflects the limited percolation of electrolyte through the weakly charged aggregates, which results in a small population of disconnected water clusters. Thus, the solvation and morphology of the
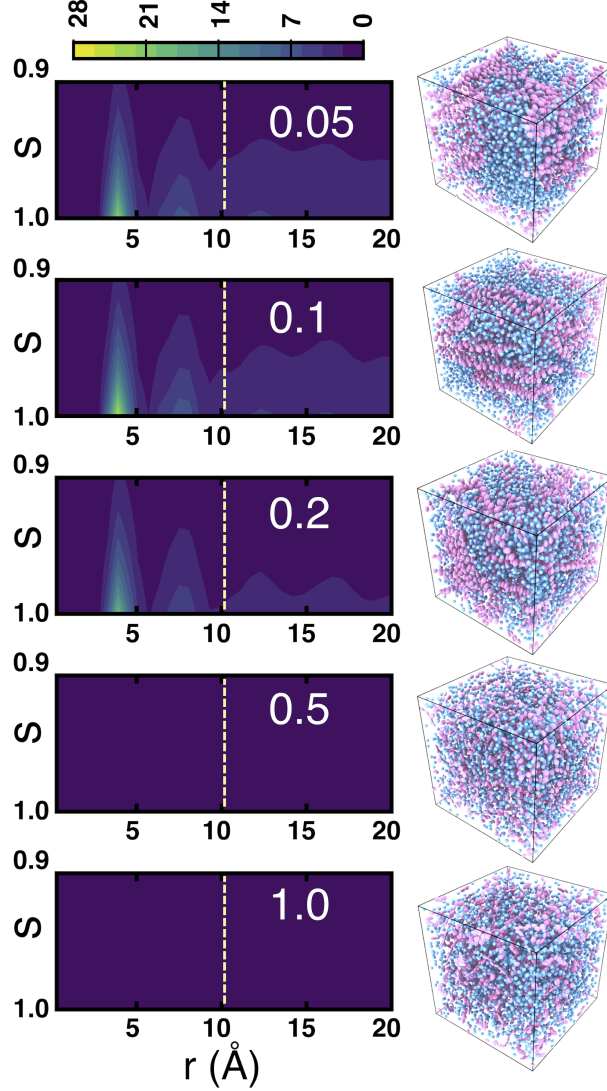
Figure 3.9. $\pi - \pi$ stacking probability presented as a joint probability distribution of distance between the backbone beads on the x-axis and $\pi - \pi$ alignment on the y-axis, as a function of oxidation level. The yellow dashed line indicates the peak for lamellar packing. Trajectory snapshots are shown to the right, where the pink and blue beads are backbone beads and water, respectively. Other components are omitted for clarity.

polymer system, although generally associated with the side-chain chemistry and processing, also exhibits a complex dependence on the extent of polymer doping.

To establish how these morphological differences impact polaron transport, KMC simulations were used to characterize the polaron MSD at each oxidation level (Figure 3.11). To decouple the role of electrostatics and morphology, a comparison of the polaron MSD with and without ion-polymer electrostatics is also presented. In all cases, we observe that the polaron is delocalized along the entire backbone, and thus polaron transport occurs via hop-
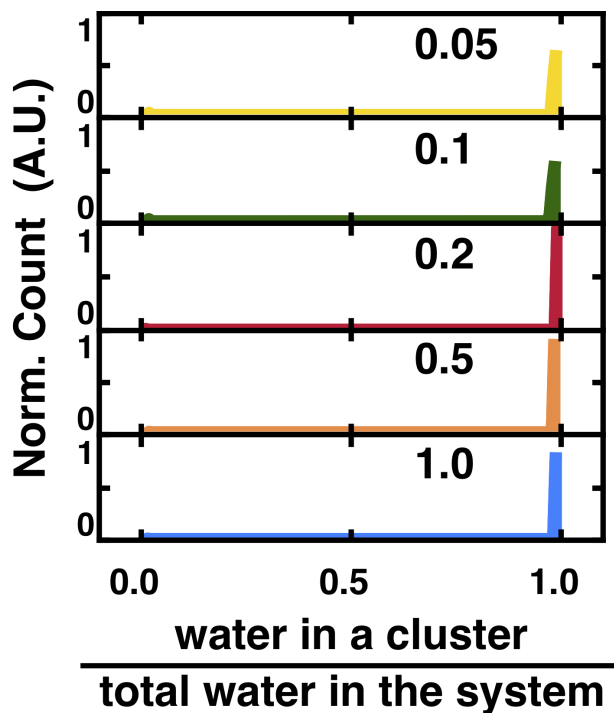
Figure 3.10. Distribution of water clusters for different oxidation levels. The number of molecules in a cluster normalized the by total number of molecules is plotted on the x-axis and the number of its occurrences normalized by the maximum number of possible occurrences is plotted on the y axis.
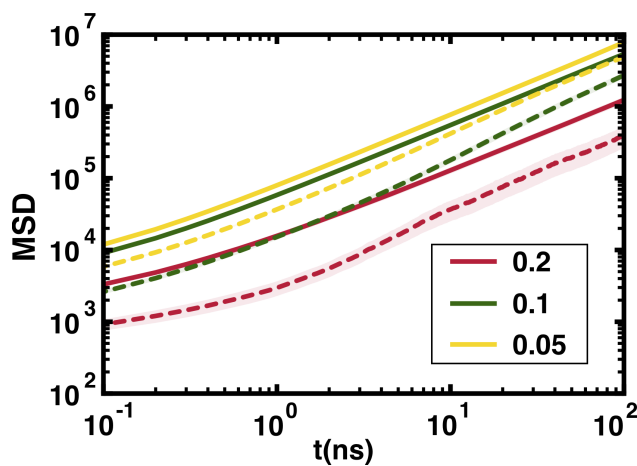
Figure 3.11. Mean square displacement of charge with intermolecular rates excluding (solid lines) and including (dashed lines) electrostatic energy component at different oxidation levels. Standard errors calculated across the sampled configurations are shown in the shaded regions.

ping between distinct chains. For oxidation levels $q \geq +0.5e$, the polymers are completely dissolved and are excluded from further discussion. Similar to the simulations that were performed with respect to varying hydration level, the inclusion of ion-polymer electrostatics suppresses polaron transport in all cases. However, unlike the case of varying hydration level, a monotonic decrease in polaron mobility with respect to oxidation level is observed both with and without electrostatics. A comparison of the quantitative details reveals several distinct mechanisms by which oxidation level modulates the polaron mobility in comparison with hydration level.

When ion-polaron electrostatics are excluded from the simulations, the polaron diffusivity monotonically decreases with increasing oxidation level. Based on the previous structural analysis, decreasing oxidation leads to relatively increased $\pi - \pi$ alignment and the formation of larger lamellar aggregates. These two factors are reflected in the distributions of charge transfer rates (Figure 3.12a) and site connectivity (Figure 3.12b) within each system. All of the dense systems exhibit a peak at high rates, with an increase in site connectivity as oxidation is reduced. We note that despite the relative decrease in $\pi - \pi$ alignment of individual backbone beads with oxidation level, the +0.2e system exhibits the largest unnormalized number of high rate pathways, which suggests that site alignment–versus bead alignment–is in fact marginally improved at higher oxidation levels. Thus, similar to the effect of hydration, charge transfer rates are reduced but site connectivity is increased with reduced oxidation, which leads to the observed increase in polaron mobility. This trend is also consistent with the observed increase in sites visited at both short and long timescales (Figure 3.12c-d) as the oxidation level is reduced.
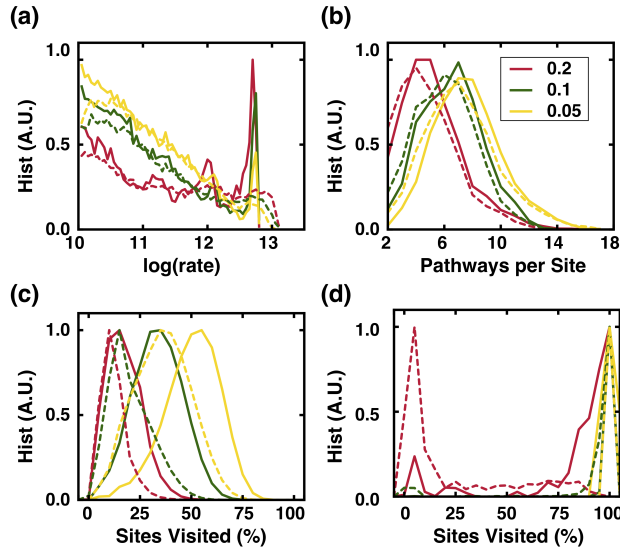
Figure 3.12. Distribution of (a) intermolecular rates and (b) pathways per site, (c) distinct sites visited over 500 jumps and (d) distinct sites visited over last 75ns trajectory, when excluding (solid lines) and including (dashed lines) electrostatic energy component at different oxidation levels. The distributions (a-b) are normalized by the maximum value across all plots and (c) is normalized by the maximum value within each plot.

When ion-polymer electrostatics are included in the KMC simulations, the polaron transport is suppressed in all cases but the behavior of the $q = +0.2e$ system is qualitatively altered. Due to the additional electrostatic site disorder, all of the systems apparently exhibit a similar suppression of fast charge transport rates (Figure 3.12a) and a minor decrease in the mean site connectivity (Figure 3.12b). However, comparing the site visitation histograms, the $q = +0.2e$ system exhibits dramatically increased polaron trapping compared with the less oxidized systems (Figure 3.12c-d). We interpret this result to indicate that although the mean site connectivity is only marginally reduced in the $q = +0.2e$ system, the pathways that are suppressed by the inclusion of electrostatics are inter-aggregate connections. In contrast, the lower oxidation level systems exhibit similar percolative polaron transport both with and without the inclusion of ion-polymer electrostatics. It should be noted that this disparity in transport would be mitigated at lower hydration levels. Thus, the polaron transport is determined by an interplay of oxidation level and hydration level, which are experimentally coupled by electrolyte gating, polymer chemistry, and processing conditions.
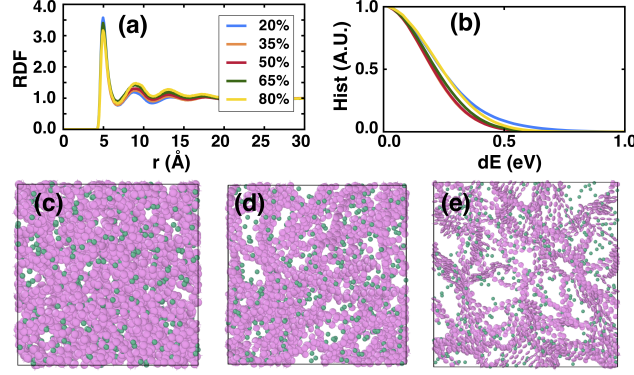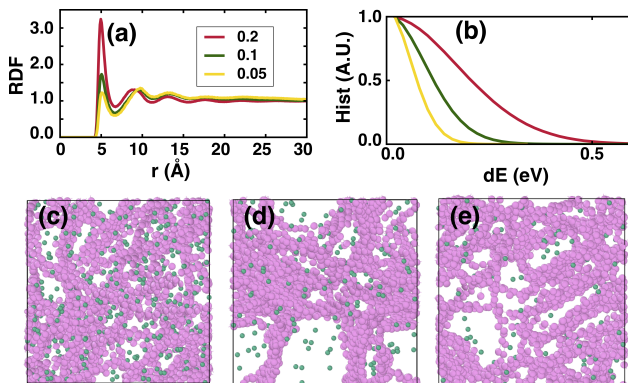

Figure 3.13. (a) Ion-backbone RDF and (b) histogram of the electrostatic
site energy differences as a function of oxidation level
. Snapshots of the system showing ions (green) and polymer (pink) for oxidation levels of
(c) 0.2, (d) 0.1, and (e) 0.05. The site-energy distributions are normalized by the maximum
value across all plots.

Unlike the hydration level series, we find significant differences in ion percolation in the polymer phase as oxidation level changes. Comparing the ion-polymer RDFs, ion intercalation within the polymer aggregates decreases with decreasing oxidation level (Figure 3.13a), which is consistent with the earlier structural and water cluster analyses. Note that the lower oxidation levels have fewer total ions in the simulation box; nevertheless, a smaller fraction of ions intercalate within the polymer phase (Figure 3.13c-e). The distinct ion distributions are also evident in the much higher electrostatic heterogeneity observed in the $q = +0.2e$ system (Figure 3.13b). Due to a combination of increased ion density and the fact that the ions intercalate rather than surround the polymer aggregates, high oxidation levels induce electrostatic disorder that strongly suppresses the polaron mobility (Table 3.2). We note that the electronic conductivity rather than the mobility is often relevant for applications. For the model OMIEC investigated here, the polaron mobility is reduced approximately

eight-fold in the the $q = +0.2$e system compared with the $q = +0.1$e system, suggesting that maximum conductivities would actually be obtained by operating at lower oxidation levels (i.e., $q = 0.05 - 0.1$e).

Table 3.2. Summary of transport coefficients at different oxidation levels with a fixed hydration level of 50%.

q_B^a	μ_+^b	$\mu_{+,ES}^c$	μ_-^d
1.0	-	-	0.0013
0.5	-	-	0.032
0.2	2.94	0.87	0.041
0.1	11.82	6.53	0.0536
0.05	16.31	10.09	0.0551

[a] The oxidation level per backbone unit.
[b] Polaron mobility with excluded electrostatics, $10^{-3}\text{cm}^2\text{V}^{-1}\text{s}^{-1}$.
[c] Polaron mobility with included electrostatics, $10^{-3}\text{cm}^2\text{V}^{-1}\text{s}^{-1}$.
[d] Anion mobility, $10^{-3}\text{cm}^2\text{V}^{-1}\text{s}^{-1}$.

Finally, the ionic mobility was characterized in each system to determine the effect on ion dynamics as the oxidation level is varied (Table 3.2). As the oxidation level increases, the ion mobility monotonically decreases. At the highest oxidation level ($q = +1.0$e), strong ion-polymer coupling leads to an acute non-linear suppression of the ion mobility. However, at the lower oxidation levels that are more representative of experimentally attainable doping, the ion mobility is only weakly sensitive to the oxidation level and also substantially lower than the polaron mobility. This weak sensitivity is notable in light of the distinct morphologies and electrolyte percolation characteristics as oxidation is varied. This reflects the fact that at the 50% hydration level, the electrolyte still exhibits macroscopic percolation in all of the investigated systems. At lower hydration levels it is possible that the oxidation level has a stronger impact on ion mobility.

3.3 Conclusions

Ionic and electronic transport are coupled within OMIECs through a complex interplay of morphological and electrostatic changes that accompany doping. Although the molecular details of these features are consequential for rational OMIEC design, they occur on length and timescales that are difficult to experimentally probe. In this contribution, we have presented a modeling framework that addresses this characterization gap while retaining sufficient chemical flexibility and molecular physics to resolve structure-function relationships. Specifically, polymer and electrolyte dynamics are modeled at the coarse-grained level using the Martini force-field supplemented by an anisotropic Gay-Berne description of the conjugated OMIEC backbone, while polaron dynamics are modeled as hopping between localized sites with rates and site energies determined by sampling configurations from the CG trajectories. In combination, these approaches are capable of simultaneously probing the molecular details of morphology and macroscopic conduction as a function of polymer identity, electrolyte concentration, and doping level.

As a demonstration of this framework, we have presented a comprehensive characterization of ion and polaron transport in a model OMIEC as a function of oxidation and hydration levels. In each case, qualitative differences in polaron dynamics occurred when ion-polaron electrostatics were excluded from the model, which reinforces the critical role of ion-polaron coupling on local transport behaviors. The simulations also recapitulate several intuitive trends, including the suppression of polaron transport by electrolyte induced morphological disorder and the increase in ion mobility with hydration level, while revealing several mechanisms that can mitigate trade-off between ion transport and polaron transport. For instance, swelling by the electrolyte is accompanied by increased screening and a decrease in deep electrostatic traps that results in comparable polaron mobilities at low to moderate hydration levels, which is a trend that has also been empirically observed in experiments.[24, 149] Additionally, we observed a non-linear suppression of polaron mobility as a function of oxidation level that indicates an optimum doping level for this system. This latter result occurs despite the fact that the polaron simulations are performed in a mean-field approximation of other polarons, and thus exclude polaron-polaron correlations and Coulomb blockade effects that are likely to further suppress transport at high oxidation levels.

With respect to the design of OMIECs, this work elucidates several molecular scale factors that are potentially subject to rational design. The results highlight near-term opportunities explored in subsequent chapters such as tuning side-chain hydrophilicity, density, and sequence to achieve electrolyte percolation without suppressing inter-aggregate polaron transport, and possibly to tune equilibrium polymer-ion configurations. On the longer term, engineering polymer aggregate dimensions and connectivity are potentially achievable by exploring electrolyte additives or tuning polymer-polymer interactions. The presented computational framework is capable of decoupling morphological features of the system such as $\pi - \pi$ overlap and connectivity of the polymer network, and further account for their relative contribution to charge transport in the system. Although the current work has focused on a case study of a single model system, the flexibility of the phenomenological framework will be valuable to elucidate these structure-function relationships.

APPENDIX

3.A Derivation of Forces and Torques for Additional Bonded Potentials

As described in the main text, our model incorporates three potentials that govern the interactions between bonded backbone beads, side-chain beads, and their relative orientation (Figure 3.3).

The position vectors and the orientation vectors of bonded backbone beads are given by $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{v}_1$, $\mathbf{v}_2$, respectively. The displacement vector, unit vector and the distance between the bonded beads are given by $\mathbf{r} = \mathbf{x}_1 - \mathbf{x}_2$, $\hat{\mathbf{r}} = \dfrac{\mathbf{r}}{\|\mathbf{r}\|}$ and $r = \|\mathbf{r}\|$, respectively. The dihedral potential (ϕ) is defined with respect to the displacement vector $\mathbf{r}$ and the orientation axes of the beads $\mathbf{v}_1$ and $\mathbf{v}_2$ (Figure 3.3) and described as,

$$U_{\text{B-B},d}(\phi) = k_1(1 + \cos(\phi)) + k_2(1 - \cos(2\phi)) + k_3(1 + \cos(3\phi)) + k_4(1 - \cos(4\phi)). \tag{3.12}$$

The expression for dihedral angle, ϕ, and the derivation of forces and torque associated with this potential is implemented as in reference [17]. The specific parameters, k, used in the present work are presented in Table 5.2 with the other force-field terms.

A bending potential to controls the angle between the interbead vector $\mathbf{r}$ and the orientation axis of each bead (i.e., $\mathbf{v}_1$ or $\mathbf{v}_2$ in Figure 3.2a) as given by

$$U_{\text{B-B},a_1}(\theta_1) = k(\theta_1 - \theta_0)^2, \tag{3.13}$$

and

$$U_{\text{B-B},a_2}(\theta_2) = k(\theta_2 - \theta_0)^2, \tag{3.14}$$

where,

$$\cos(\theta_1) = \mathbf{v}_1 \cdot \hat{\mathbf{r}}, \tag{3.15}$$

and

$$\cos(\theta_2) = -\mathbf{v}_2 \cdot \hat{\mathbf{r}}, \tag{3.16}$$

and the parameters, k and θ_0, for angles involving $\mathbf{v}_1$ and $\mathbf{v}_2$ are same in this case due to symmetry. The forces $\mathbf{f}_1$ and $\mathbf{f}_2$ on beads 1 and 2, respectively, due to the potential $U_{\text{B-B},a_1}$, are given by

$$\mathbf{f}_1 = 2k\frac{(\theta_1 - \theta_0)}{\sin\theta_1 r}(\mathbf{v}_1 - \cos\theta_1\hat{\mathbf{r}}), \text{ and}$$
$$\mathbf{f}_2 = -2k\frac{(\theta_1 - \theta_0)}{\sin\theta_1 r}(\mathbf{v}_1 - \cos\theta_1\hat{\mathbf{r}}). \tag{3.17}$$

The torques τ_1 and τ_2 on beads 1 and 2 respectively due to potential $U_{\mathrm{B-B},a_2}$, are given by

$$\tau_1 = 2k\frac{(\theta_1 - \theta_0)}{\sin\theta_1}\mathbf{v}_1 \times \hat{\mathbf{r}}, \text{ and}$$
$$\tau_2 = 0. \tag{3.18}$$

The forces and torques due to the potential $U_{\mathrm{B-B},a_1}$ are similar, and given by

$$\mathbf{f}_1 = -2k\frac{(\theta_2 - \theta_0)}{\sin\theta_2 r}(\mathbf{v}_2 + \cos\theta_2\hat{\mathbf{r}}),$$
$$\mathbf{f}_2 = 2k\frac{(\theta_2 - \theta_0)}{\sin\theta_2 r}(\mathbf{v}_2 + \cos\theta_2\hat{\mathbf{r}}),$$
$$\tau_1 = 0, \text{ and}$$
$$\tau_2 = -2k\frac{(\theta_2 - \theta_0)}{\sin\theta_2}\mathbf{v}_2 \times \hat{\mathbf{r}}. \tag{3.19}$$

Similarly, harmonic bending potentials are used to control the alignment of bonded side-chain beads with respect to the backbone (Figure 3.3b). The positions of the backbone bead and side-chain bead are given by $\mathbf{x}_1$ and $\mathbf{x}_2$, respectively, and the orientation of the backbone bead is given by $\mathbf{v}$. The displacement vector, unit vector and the distance between the beads are given by $\mathbf{r} = \mathbf{x}_1 - \mathbf{x}_2$, $\hat{\mathbf{r}} = \dfrac{\mathbf{r}}{\|\mathbf{r}\|}$, and $r = \|\mathbf{r}\|$, respectively. The harmonic bending potential is defined in terms of the angle between the interbead vector $\mathbf{r}$ and the backbone bead orientation vector $\mathbf{v}$,

$$U_{\mathrm{B-S}}(\theta) = k(\theta - \theta_0)^2, \text{ where}$$
$$\cos(\theta) = \mathbf{v} \cdot \hat{\mathbf{r}}. \tag{3.20}$$

The forces $\mathbf{f}_1$ and $\mathbf{f}_2$ on beads 1 and 2 respectively due to this potential, are given by

$$\mathbf{f}_1 = 2k\frac{(\theta - \theta_0)}{\sin\theta r}(\mathbf{v} - \cos\theta\hat{\mathbf{r}}), \text{ and}$$
$$\mathbf{f}_2 = -2k\frac{(\theta - \theta_0)}{\sin\theta r}(\mathbf{v} - \cos\theta\hat{\mathbf{r}}). \tag{3.21}$$

The torques τ_1 and τ_2 on beads 1 and 2 respectively, are given by

$$\tau_1 = 2k\frac{(\theta - \theta_0)}{\sin\theta}\mathbf{v} \times \hat{\mathbf{r}}, \text{ and}$$
$$\tau_2 = 0. \tag{3.22}$$

3.B   Polymer End-to-End Autocorrelation Decay
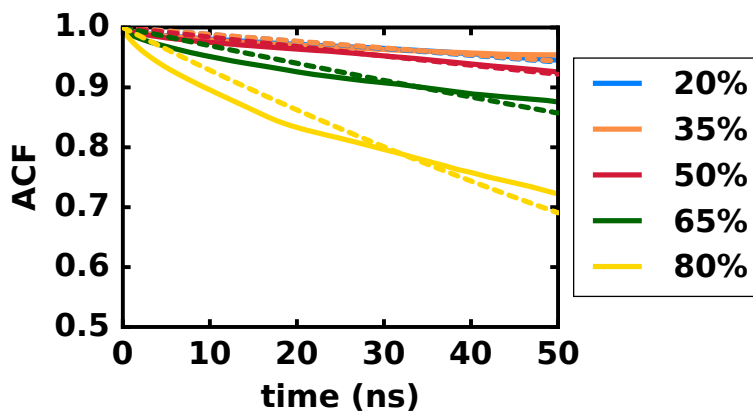
Figure 3.14. Autocorrelation function of polymer end-to-end vector across hydration levels (fixed oxidation at +0.2e per repeat unit) and exponential fits to the series (dashed lines).
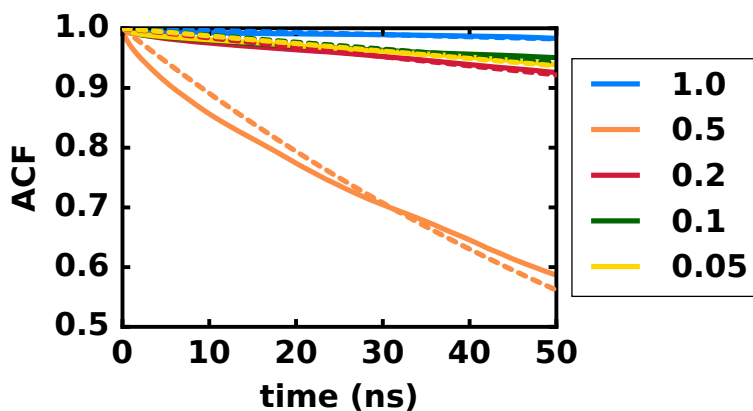
Figure 3.15. Autocorrelation function of polymer end-to-end vector across oxidation levels (fixed hydration level of 50%) and exponential fits to the series (dashed lines).

The autocorrelation of the end-to-end vectors reports on the degree to which the polymers tumble over the course of the simulation (Figure 3.14 and 3.15). The general trend is that increased electrolyte swelling leads to more rapid decorrelation. In the highly oxidized +1.0e per repeat unit case, the strong coulombic repulsion between chains limits rearrangement. The decorrelation timescales from the exponential fits are reported below.

Table 3.3. Timescale of decay of end-to-end vector autocorrelation at different hydration levels with a fixed oxidation level of 0.2 per backbone unit.

[a] The hydration level in weight percent.

[b] Timescale of decay of end-to-end vector autocorrelation in ns.

$\rho_{H_2O}^a$	τ^b
20 %	840.80
35 %	862.70
50 %	615.99
65 %	324.22
80 %	135.04

Table 3.4. Timescale of decay of end-to-end vector autocorrelation at different oxidation levels with a fixed hydration level of 50 %.

[a] The oxidation level per backbone bead.

[b] Timescale of decay of end-to-end vector autocorrelation in ns.

q_B^a	τ^b
1.0	2780.64
0.5	86.62
0.2	614.99
0.1	842.96
0.05	768.911

3.C Time-scale of Decay of Various Rate Parameters

As described in the main text, the autocorrelation decay timescales of several variables are used to determine the averaging period for parameters involved in the master equation as well as the timescale of a configuration refresh to avoid stale rates. The autocorrelation functions (ACF) for the various quantities described in the main text are shown below. We find the autocorrelation of bonded modes and electrostatic energies decay very fast due to the intramolecular fluctuations, and diffusion and fluctuation of ions and water, respectively. On the contrary, the autocorrelation of the intermolecular coupling decays very slowly as it depends on the relative positions of the polymers which have much slower dynamics.
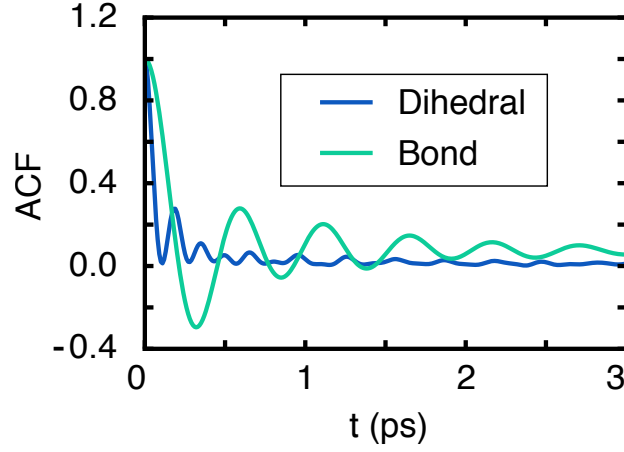
Figure 3.16. ACF of bond lengths and dihedrals between neighboring backbone beads
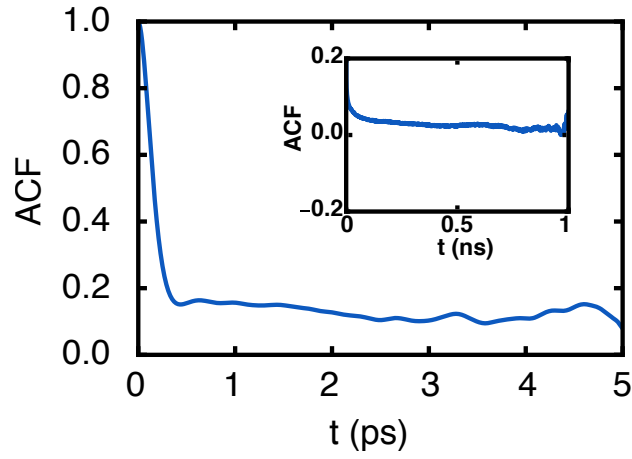
Figure 3.17. ACF of electrostatic energy difference between distinct backbone beads.
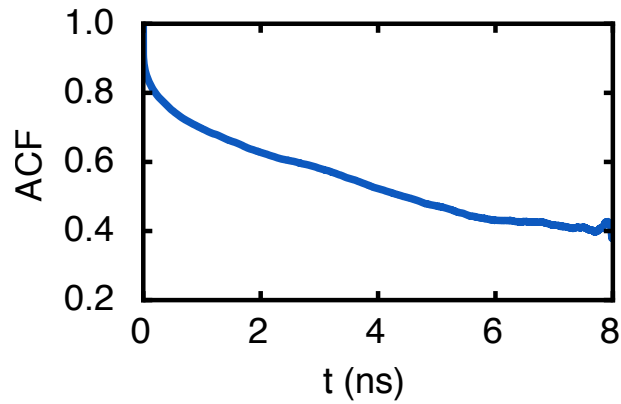
Figure 3.18. ACF of electronic coupling between polaron sites.
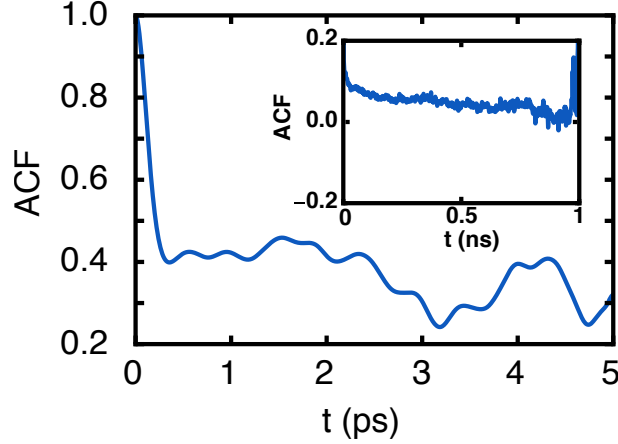
Figure 3.19. ACF of electrostatic energy difference between polaron sites.

3.D   Parameters for Non-bonded and Bonded potentials

The backbone bead, side-chain bead, and ion are denoted by B, S and Q respectively. The beads in water molecules are denoted by their respective MARTINI bead types POL, WM and WP. The Gay-Berne parameters are in LAMMPS notations, where the axes **x**, **y** and **z** correspond to **u**, **v** and **w** in the main text, respectively.

Table 3.5. Parameters used for bonded interactions in the simulations. Units of all parameters are consistent with LAMMPS 'real' units (kcal mol^{-1}, Å etc.) and potential styles. The bonds in water molecules WM-POL and WP-POL are constrained using the SHAKE algorithm [143].

Mode	Mode Type	Parameters
Bond	B-B	$k = 3.0$, $r_0 = 5.0$
Bond	B-S	$k = 1.5$, $r_0 = 4.7$
Bond	S-S	$k = 1.5$, $r_0 = 4.7$
Bond	POL-WM(WP)	$r_0 = 1.4$
Angle	B-B-B	$k = 7.5$, $\theta_0 = 180.0$
Angle	B-B-S	$k = 3.0$, $\theta_0 = 90.0$
Angle	B-S-S	$k = 3.0$, $\theta_0 = 180.0$
Angle	WM-POL-WP	$k = 0.5019$, $\theta_0 = 0.0$
Dihedral	B-B-B-B	$k_1 = 0.0$, $k_2 = 0.6$, $k_3 = 0.0$, $k_4 = 0.0$
Other	B-B, a_1	$k_a = 7.5$, $\theta_0 = 90.0$
Other	B-B, a_2	$k_a = 7.5$, $\theta_0 = 90.0$
Other	B-S, a	$k_a = 3.0$, $\theta_0 = 90.0$
Other	B-B, d	$k_1 = 0.0$, $k_2 = 7.5$, $k_3 = 0.0$, $k_4 = 0.0$

Table 3.6. Parameters used for non-bonded interactions in the simulations. Units of all parameters are consistent with LAMMPS 'real' units (kcal mol^{-1}, Å etc.) and potential styles. The water particles WM and WP do not experience any Van der Waal interactions.

	B	S	POL	Q
B	$\sigma_x = \sigma_z = 5.0$, $\epsilon_x = \epsilon_z = 0.25$, $\sigma_y = 3.0$, $\epsilon_y = 1.2$, $\sigma_c = 3.0$	$\epsilon = 0.84$, $\sigma = 4.7$	$\epsilon = 0.61$, $\sigma = 4.7$	$\epsilon = 0.65$, $\sigma = 4.7$
S		$\epsilon = 1.08$, $\sigma = 4.7$	$\epsilon = 1.02$, $\sigma = 4.7$	$\epsilon = 1.34$, $\sigma = 4.7$
POL			$\epsilon = 0.96$, $\sigma = 4.7$	$\epsilon = 1.19$, $\sigma = 4.7$
Q				$\epsilon = 0.84$, $\sigma = 4.7$

3.E Parameters for Charge Transport Simulations

The following parameters were used for the Marcus expressions discussed in the main text: $J_{\text{inter}} = 0.02$ eV, $r_0 = 3$ Å, $\alpha = 0.25$ Å^{-1}, $J_{\text{intra}} = 0.4$ eV, $\epsilon_{\text{opt}} = 3$, and $\epsilon_s = 10$. The expressions for external reorganization, λ, and oxidation energy, E, as a function of oligimer length, l, were obtained from fits to reported data for polypyrrole[142]: $\lambda = 0.0003l^2 - 0.0275l + 0.5253$ and $E = 1.56/l + 4.31$, respectively.

3.F Persistence Length of Polymer for All Systems

Table 3.7. Persistence lengths at different hydration levels with a fixed oxidation level of 0.2 per backbone unit.

$\rho_{H_2O}^a$	l_p^b
20 %	17.35
35 %	41.25
50 %	59.78
65 %	60.27
80 %	52.96

^{a} The hydration level in weight percent.
^{b} Persistence length in Å.

The persistence lengths of organic conducting polymers are generally reported for solutions with organic solvent [85], rather than aqueous conditions, making a direct comparison difficult. However, we find the values reported here show a sensible trend of increasing with hydration level due to elimination of packing defects and chain extension as the system swells. For oxidation levels up to +0.2e, we observe an increase in persistence length with

Table 3.8. Persistence lengths at different oxidation levels with a fixed hydration level of 50 %.

q_B^a	l_p^b
1.0	4.85
0.5	24.43
0.2	59.78
0.1	34.29
0.05	20.18

[a] The oxidation level per backbone bead.
[b] Persistence length in Å.

oxidation level due to increased clustering and stiffening of the polymers. The +1.0e and +0.5e oxidation-level cases exhibit complete dissolution due to Coulombic repulsion, leading to a lower persistence length.

3.G   Additional Results Referenced in the Main Text
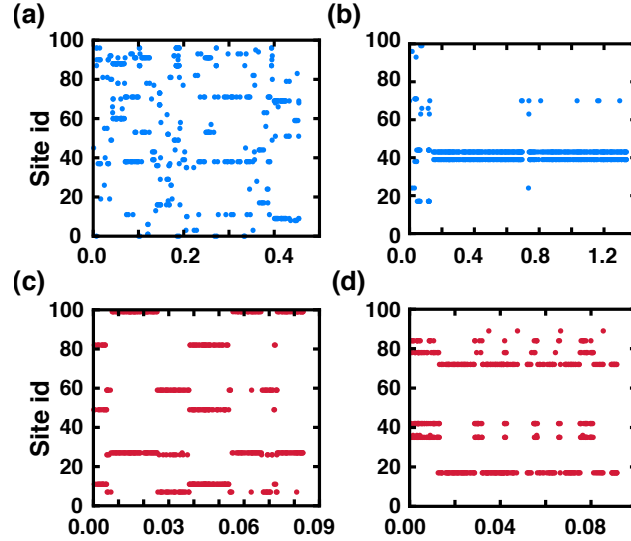
Figure 3.20. Polaron site IDs of first 500 sites visited during a sample KMC trajectory for (a) hydration level 20% without electrostatic energy component, (b) hydration level 20% with electrostatic energy component, (c) hydration level 50% without electrostatic energy component, and (d) hydration level 50% with electrostatic energy component.
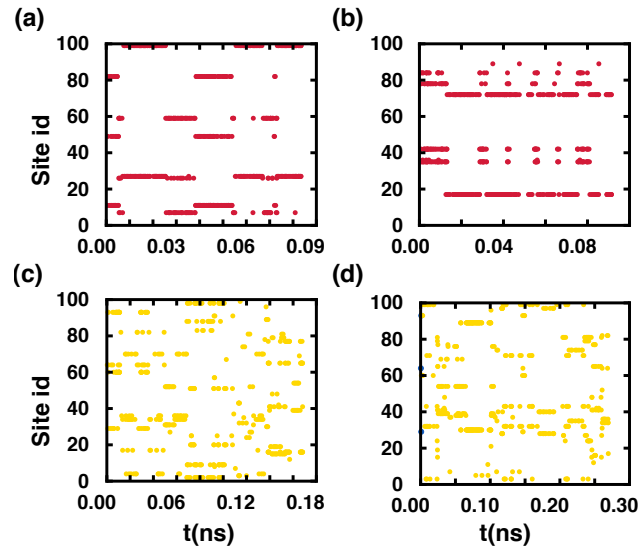
Figure 3.21. Polaron site IDs of first 500 sites visited during a sample KMC trajectory for (a) oxidation level 0.2 without electrostatic energy component, (b) oxidation level 0.2 with electrostatic energy component, (c) oxidation level 0.05 without electrostatic energy component, and (d) oxidation level 0.05 with electrostatic energy component.

4. HOW SIDE-CHAIN HYDROPHILICITY MODULATES MORPHOLOGY AND CHARGE TRANSPORT IN MIXED CONDUCTING POLYMERS

Reprinted (adapted) partially with permission from [80]. Copyright Wiley Periodicals, LLC.

As discussed in the previous chapters, the goal of this thesis is to elucidate the structure-function relationships for side-chain engineering of OMIECs, specifically, for conjugated mixed conductors, in the context of OECT devices.

In an OECT, the output source-drain current is the current though the polymer film, which is modulated by changing the doping within the polymer via a gate electrode in the electrolyte. The magnitude of the current change is related to the transconductance—the product of capacitance and carrier mobility—of the OMIEC and is thus a key figure-of-merit.[66] However, relating transconductance to side-chain identity is complicated by how doping efficiency (i.e., carriers created in the polymer per electrolyte dopant) and carrier mobility potentially respond to increased electrolyte uptake. For example, increased electrolyte uptake straightforwardly increases capacitance, while it has an indeterminant effect on doping efficiency and mobility. Several comparative studies of how side-chain chemistry impacts OECT performance have documented a variety of behaviors depending on the polymer chemistry, electrolyte, and particular side-chain substitution. Two recent studies investigated how OECT properties respond to incorporating varying amounts of polar and apolar side-chains on different OMIEC polymer backbones.[50, 150] For both p-type and n-type conducting polymers, these studies observed that all polar side-chains generally maximize the transconductance. Both studies observed maximum electrolyte uptake for the all polar side-chain cases, suggesting that the net increase in doping offsets any potentially negative effects on doping efficiency or mobility.

However, the molecular details of these trends and the degree to which mobility and doping efficiency might be simultaneously optimized with electrolyte uptake remain undetermined. To provide molecular insight into some of these potential trade-offs, we have performed the first molecular dynamics (MD) study of the effect of systematically varying the side-chain polarity on the morphology and mixed conduction of a model OMIEC based on a thiophene backbone with either alkyl or glycolated side-chains. These simulations provide molecular level details on electrolyte percolation, morphology, and charge transport to supplement the macroscopic picture that has been previously established in experimental studies performed in literature. Additionally, this study acts as a validation that our model correctly captures the effect of side-chain design variables to the extent that experimental comparisons are possible, and opens the door for future side-chain design studies.

4.1 Methods

The coarse-grained force-field and kinetic Monte Carlo were modified when necessary as compared to the original version in Chapter 3. In the following sections, we briefly review the details and highlight the changes with respect to the originally reported implementation. In the appendix, all force-field parameters and simulation conditions are reported.

4.1.1 Coarse-grained Model

The OMIEC molecular dynamics model used in this study was composed of an anisotropic conjugated backbone, and isotropic side-chain, ions, electrolyte solvent, and a processing solvent. Unless specified otherwise, the isotropic bonded and nonbonded interactions refer to bead types from Martini version 2.[100, 181] To better approximate film processing, all simulations were first equilibrated in chloroform, modeled by a C4 Martini bead, followed by equilibration in an aqueous electrolyte modeled using a chloride counter-ion and coarse-grained water. The polarizable Martini forcefield was used for water[181] and chloride ions were represented by Qa bead.

A comparison of polar and apolar side-chains was performed using isotropic Martini based beads that are representative of alkyl and polyethylene glycol (PEG) chains, respectively, which are commonly used side-chains in experiments. The alkyl side-chain was modeled with the SC3 Martini bead that represents a propyl group (Figure 5.1) as has previously been implemented in the Martini model of poly(3-hexyl thiophene) (P3HT).[5] For the PEG chain, we adopted the parameters from the Martini CG model for polyethelyene glycol[57], where an EO bead represents the $CH2 - O - CH2$ group (Figure 5.1). However, the EO bead was originally parameterized with the water force-field from Martini version 2.2. To make it compatible with the polarizable water model, we modified the interaction of EO with the POL water bead by setting $\epsilon = 3.325$ kJ/mol, to obtain a hydration free energy of -14.77 kJ/mol, which is comparable to the value of -14.73 kJ/mol obtained by the original Grunewald model[57]. The bonded parameters for EO were also retained from the Grunewald model except for the restricted bending potential and instead cosine-squared potential was used similar to the original Martini model. As the purpose of the restricted-bonding potential was numerical stability and we are modeling short PEG oligomers as side-chains, this modification should be inconsequential.

As in the original version, the potential interaction between backbone beads was modeled using an anisotropic Gay-Berne potential[14] and the orientation of consecutive backbone bead was controlled using additional torsional and stretching potentials. The Gay-Berne potential parameters were selected to be representative of thiophene and follow our previous report. Simulations of a 3-hexyl-thiophene (3HT) monomer liquid based on this backbone and two SC3 side-chain beads (i.e., the same Martini model that has previously been used for P3HT[5]) resulted in a density of 0.97 g/cc that was in close agreement with the experimental density of 0.95 g/cc. The backbone bonded potential parameters were modified to obtain a persistence length of 7.34 repeat units (RU) for P3HT in toluene, comparable to the experimental persistence length for polythiophene of 6.75-7.43 RU.[85] As in our previous work, the backbone bead interacts with other side-chain and solvent beads with an isotropic
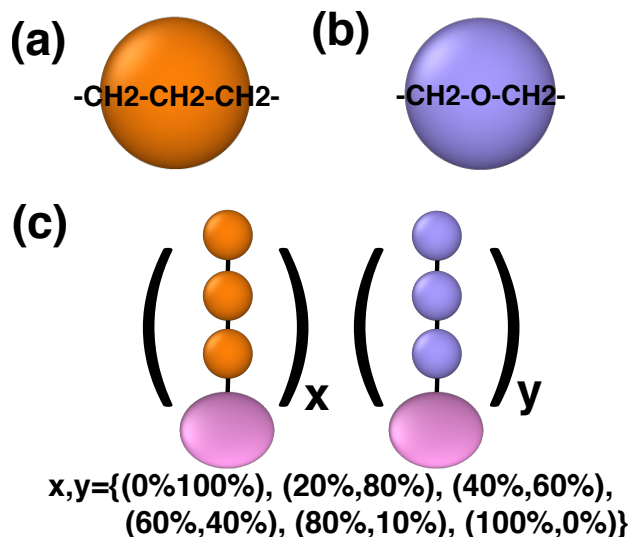
Figure 4.1. The coarse-grained mapping of the apolar (a) and polar (b) beads used to model the hydrophobic and hydrophilic polymer side-chains, respectively. (c) The materials simulated here consist of random copolymers with varying fractions of the two types of side-chain.

potential, but is modeled as a Martini C4 bead here instead of C3 bead from our previous model. The Martini C4 bead reproduces the hexadecane-water partition free energy of a thiophene ring as 9.27 kJ/mol in better agreement than C3 with the experimental partition free energy of 10 kJ/mol[1].

All simulations consisted of monodisperse solutions containing polymers of 50 backbone beads. Six polymers were studied (Figure 5.1(c)) with varying amounts of polar and apolar side-chains: 0 % polar (100 % apolar), 20 % polar (80 % apolar), 40 % polar (60 % apolar), 60 % polar (40 % apolar), 80 % polar (20 % apolar), and 100 % polar (0 % apolar). Each side-chain contained 3 CG beads representing roughly 9 heavy atoms. The side-chains were attached on each backbone bead in a regioregular fashion. The cases with mixed amounts of polar and apolar side-chains were modeled as random copolymers by stochastically assigning the identity of each side-chain. The degree of polymerization[35, 42] and side-chain length[52, 90, 112] were selected to be comparable to that of typical experimental OMIECs.

4.1.2 Molecular Dynamics

LAMMPS was used to perform all molecular dynamics simulations.[130] All simulations used a 10 fs integration time step, Velocity-Verlet integration, and periodic boundary conditions. Lennard-Jones, electrostatic and Gay-Berne interactions were truncated at 12 Å. The Lennard-Jones and Coulombic interactions were shifted using the standard GROMACS shift function over a length of 9-12 Å and 0-12 Å, respectively. Film equilibration occurred over two distinct simulations. First, polymer film formation was simulated by dissolving the neu-

tral polymers in chloroform then evaporating the solvent. Subsequently, the polymers were doped (i.e., charged) and electrolyte was added to the simulation and allowed to equilibrate.

For the polymer film formation simulations, the polymer and chloroform solvent molecules were initialized from diffuse configurations, using a cubic grid to place molecules in random orientations without overlap. Each simulation box contained 50 polymers and was initially surrounded by 50,000 chloroform molecules. The velocities were initialized from a uniform distribution obtained with a random seed value and scaled to give the correct kinetic energy. The simulations were first relaxed in the NVE ensemble with restrained displacements of 0.1 Å per time step for 10 ps. This was followed by a simulation in the NVT ensemble at 300 K while linearly rescaling the simulation box over 2 ns to obtain a condensed system with a density of 1 g/cc. The density was then equilibrated in the NPT ensemble for 100 ns at 300 K and 1 atm. After the NPT simulation, 20 wt% of the chloroform beads were randomly selected and removed from the system, followed by an equilibration step in the NPT ensemble for 100 ns at 300 K and 1 atm. This process of chloroform removal followed by NPT equilibration was performed five times, resulting in 28,750 (80 wt% of solvent), 10,781 (60%), 4,792 (40%), 1,797 (20%), and 0 (0%) chloroform molecules, respectively. This procedure is intended to qualitatively mimic the solvent processing and evaporation during experimental film formation (e.g., spin-coating) and has been employed in other coarse-grained simulations of organic polymers.[107]

The final polymer film configuration was unwrapped in accordance with the periodic boundary conditions, the backbone beads were charged and then surrounded by randomly placed solvent and ion beads on the outside of the polymer film, each within a separate cubic box of length 6 Å to avoid overlaps. At this stage, 10,781 CG water beads were added, which correspond to 43124 real water molecules and 60 wt% of solvent. Each backbone unit carried a charge of +0.2 (i.e., a doping induced charge every five backbone beads), and 500 counterions with a charge of −1.0 were added to achieve charge neutrality. These systems were relaxed in the NVE ensemble with restrained displacements of 0.1 Å per time step for 10 ps, followed by an NPT equilibration for 100 ns, and an NPT production run for an additional 100 ns at 300 K and 1 atm. In all of the NVT and NPT simulations, the Nosé-Hoover thermostat and barostat were employed using the modified form proposed by Martyna, Tobias, and Klein, as implemented in LAMMPS [102], with a relaxation time constant of 0.15 ps and 1.5 ps for the thermostat and barostat, respectively. In all of the simulations, the thermodynamic data and coordinates were sampled at 1 ps. The ASPHERE package in LAMMPS was used for all Gay-Berne force, torque evaluations, and numerical integrations.[22] The SHAKE algorithm [143] implemented as a part of the RIGID package in LAMMPS was used to constrain the bonds in the water molecules. Additional bonded potentials to control the orientation of neighboring beads were implemented as custom classes in LAMMPS by our group. OVITO was used for visualizing the final trajectories and generating snapshots of MD configurations.[160]

4.1.3 Charge Transport Calculations

Charge transport within the OMIEC polymer network was modeled on the basis of configurations generated from the CG-MD trajectories. The expressions for charge transport

rates, the physics behind the rate parameters, the choice of these rate parameters, the kinetic Monte Carlo procedure and the final calculation of mobility is explained in detail Chapter 3. In brief, each backbone bead was parameterized to represent a single conjugated unit (i.e., a ring system without internal flexible dihedral degrees of freedom), thus sites consisting of one or more contiguous backbone beads constitute the relevant polaronic basis states for charge transfer. The extent of polaron delocalization (i.e., the number of contiguous beads involved in a single polaron site) was determined based on a threshold for the charge transfer rate for contiguous beads ($10^{-14}\mathrm{s}^{-1}$). Contiguous beads with charge transfer rates above this threshold were combined together to form a delocalized polaron site; contrarily, smaller polaron sites reflect poor torsional alignment or electrostatic disorder along the polymer chains. The charge transfer rates between polaron sites were parameterized using semi-classical Marcus theory[96], whereas the charge transport between contiguous beads was modeled using a modified Marcus rate in the adiabatic limit.[56, 142] The charge transfer rate parameters capture the appropriate dependence on the delocalization of electrons along the backbone, the electronic coupling due to orbital overlap, the reorganization energy in the presence of solvent, and the energetic disorder due to electrostatic interactions with the solvent.

The expression used to calculate the intermolecular electronic coupling was updated to be calculated on the basis of individual pair-wise couplings between beads, instead of an average inter-site distance over all beads, as was done previously. The new expression for coupling is

$$J_{\mathrm{ij}} = J_{\mathrm{inter}} \sum_{\mathrm{m}\in\mathrm{i}\cup\mathrm{j}} \exp\left\{-\alpha(r_{\mathrm{ij}} - r_{\mathrm{min,m}})\right\}, \tag{4.1}$$

where α is a decay factor, r_0 is the reference separation for J_{inter}, and $r_{\mathrm{m,min}}$ is the minimum distance of bead m in site i (or j) from beads in the other site j (or i). This expression resolves coupling at each backbone unit and assigns a nonzero coupling for polymers which partially overlap, which was not captured by the earlier expression based on the average intersite distance. The intermolecular coupling value J_{inter}, the reference separation r_0 and the decay factor α were set to 0.01 eV, 3.5 Å and 0.8 Å^{-1}, respectively to match with the new coupling expression and commonly reported values in organic electronic literature.[9] For the length 50 chain, the values for HOMO energy levels and reorganization energy as a function of site length were obtained from the length-dependence study by Zade et al. on polythiophenes.[182] The intramolecular reference coupling, J_{intra}, used in the adiabatic rate expression for charge transfer between contiguous beads was set to a value of 0.31 eV. This parameter raises or lowers the mean intra-chain charge transfer rates and thus affects the delocalization length of the polaron sites. Here, J_{intra} was selected to obtain an average polaron delocalization length of ~13 backbone beads for the 100% polar side-chain case (Figure 4.4). The delocalization lengths ranged from 1-50 (polarons localized to a single backbone bead vs the entire polymer chain) depending on the side-chain composition, degree of backbone order and local electrostatic interactions with electrolyte.

A common numerical problem in KMC simulations of processes with kinetics spanning multiple orders of magnitude, is that it the slow processes are necessarily rarely sampled. In the current systems, intra-chain charge transfer rates tend to be much larger than inter-chain rates, with the effect that the charge can spend a large number of KMC steps hopping

within a small set of spatially localized sites while waiting for a relatively slow interchain hop to occur. To address this problem, we use the Aggregate Monte Carlo (AMC) procedure proposed by Brereton et al.[19] In the AMC procedure, the "trap sites" are aggregated together based on their connectivity and relatively high rates for transfer, then the charge transport is simulated similar to KMC as a Markov chain of jumps between these aggregates. An important distinction is that the transport out of an aggregate is modeled as a Markov jump process from the trap to any of its neighbors which are treated as absorbing barriers. In doing this, the long physical time spent in the trap site is correctly accounted for and the exit probabilities depend on the transport probability to the neighbors as well as the transport within the trap region. In order to determine the aggregates, we use the decomposition algorithm described in Brereton et al with the hyperparameters $\alpha = 0.2$, $\beta = 0.2$ and $\gamma = 0.03$.[19]

The key steps involved in the charge transport characterization are summarized below. For each configuration sampled at time t obtained from the CG-MD simulation,

1. **Calculation of intramolecular rates.** The charge transfer rates between contiguous backbone beads are calculated in preparation for defining polaron sites. These rates are averaged over 20 configurations over t± 1 ps.

2. **Definition of polaronic basis states.** The contiguous beads with rates above 10^{14}s^{-1} are combined into a single site.

3. **Calculation of intermolecular rates.** The charge transfer rates between polaron sites are averaged over 20 configurations uniformly sampled from t± 1 ps.

4. **Determination of Aggregate States.** Polaron sites with fast transfer rates are aggregated and their exit probabilities are calculated based on the decomposition algorithm.

5. **KMC simulations of charge hopping trajectories.** For each configuration, 40 trajectories are generated by randomly choosing different starting sites.

The above procedure to estimate MSD is performed for 20 configurations sampled every 5ns and subsequently, the diffusivity and mobility were estimated from the MSD. The MSD and mobility calculation was performed for 3 independent MD simulations, to obtain the net average mobility and error reported below. Likewise, distributions of various quantities such as electronic coupling, energies, site length, etc. were obtained as averages over the 3 independent MD simulations.

4.2 Results and Discussion

4.2.1 Morphology Characterization

The simulated series of OMIECs comprise a case study on the effects of varying the side chain hydrophilicity and composition on the morphology and transport of random copolymers. We first focus on the distinct morphological features across the series, since it provides an interpretive basis for the later charge transport results.

The density of polar side-chains has a predictably large impact on the extent of electrolyte percolation in each OMIEC. Representative snapshots for each system demonstrate the onset of electrolyte percolation around 40-60% of polar side-chain incorporation (Figure 4.2). The trend is notably non-linear, as the morphology qualitatively changes from a relatively dense organic film with excluded electrolyte to a swelled network around a concentration of 60% polar side-chains. Neglecting other factors, the free energy of water absorption should be linear in the fraction of polar side-chains, meaning that the likelihood of water percolation would be exponential in the polar fraction. This is consistent with the apparent phase transition around 60% polar side-chain fraction, and also consistent with experimental observations of super-linear swelling with respect to the fraction of polar side-chains.[50, 150] For the simulated electrolyte densities, a bulk electrolyte phase persists until ≥80% hydrophilic side-chain fractions, at which point the polymers completely swell (Additional details can be observed in Figure 4.7).

The side-chain composition also has several subtle effects on the distribution of electrolyte within the polymer phase (Figure 4.2). First, the much larger probabilities for finding water and ions near polar side-chains compared with apolar side-chains (Figure 4.2b) confirms the intuition that the former drive electrolyte percolation within the OMIEC. We also observe that the the effective (a)polarity of the side-chains is function of side-chain concentration. Specifically, as the fraction of polar chains increases, the likelihood of finding water and ions near the apolar chains also increases. Conversely, as the fraction of apolar chains increases the likelihood of observing electrolyte near the polar chains decreases. At polar fractions ≤ 40% this is explainable by limited electrolyte percolation, which correspondingly limits electrolyte association with buried polar chains. However, this trend is also observable in the percolating morphologies, which we interpret to reflect the effect of neighboring side-chains on the local hydrophilicity. Specifically, at higher polar side-chain fractions, both polar and apolar side-chains are more likely to have a polar neighbor that promotes local association with the electrolyte. The reverse is also true as the fraction of apolar chains increases. It isn't obvious whether this effect is immediately useful, but it is a distinct prediction of the simulations and we would hypothesize that it is more pronounced in random copolymers (as simulated here), compared with block copolymers.

A key feature of the employed OMIEC model is that it explicitly describes the π-orientation of the polymer backbone. The $\pi - \pi$ organization between chains was characterized using the joint-distribution function of backbone bead separation and the dot product of the π normal vectors, S (Figure 4.3). All of the systems exhibit a high degree of $\pi - \pi$ organization (i.e., S > 0.8 in all cases) between neighboring chains; however, in the progression from low to high polar side-chain fraction, we observe that the $\pi - \pi$ organization decreases for intermediate fractions (≥20%) of polar side-chains. We ascribe this to the combination of side-chain heterogeneity and partial percolation of the electrolyte, both of which frustrate chain packing at the intermediate fractions. This trend is also consistent with the organization at the second and third nearest neighbor chain positions (i.e., ~7 and 11 Å, respectively). Interestingly, the 100% polar case with the greatest degree of electrolyte swelling shows the highest overall organization. In contrast, the 100% apolar case shows weaker organization, but we caution that this system forms a thin layer (~5 nm) that is dominated by the polymer-electrolyte interface and is not representative of a bulk organic
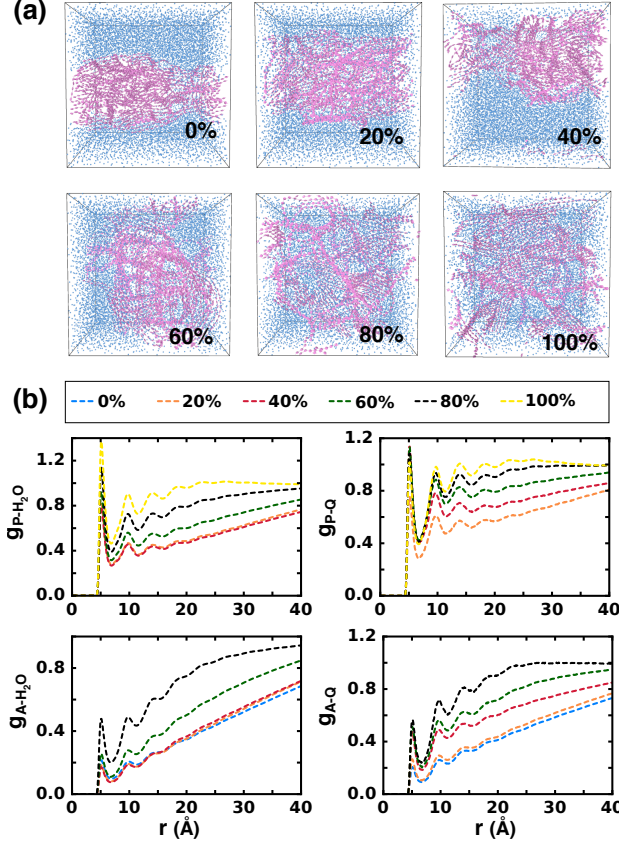
Figure 4.2. (a) Snapshots showing polymer (pink) and water (blue) for different side-chain compositions. All other bead types are omitted for clarity. (b) Radial distribution functions g_{x-y} for the polar side-chains and water (P-H_2O), polar side-chains and ions (P-Q), apolar side-chains and water (A-H_2O), and apolar side-chains and ions (A-Q).

semiconducting polymer. The one-dimensional radial distribution functions for the backbone beads also reveal several subtle features (Figure 4.7d). For instance there is only weak lamellar organization as evidenced by a small peak at ~17Åthat is suppressed with increasing electrolyte percolation.[42] This length is consistent with experimentally reported length of 14-19Å[42, 90, 150], roughly two times the side-chain length, and thus also indicates that limited side-chain interdigitation occurs in these systems.

4.2.2 Charge Transport Characterization

Next, we examined how charge transport within the polymer is affected by the changes in morphology and electrostatic environment attendant to changing the side-chain composition. Charge transport was modeling using KMC to generate ensembles of single polaron hopping trajectories across a range of representative MD configurations for each system. The
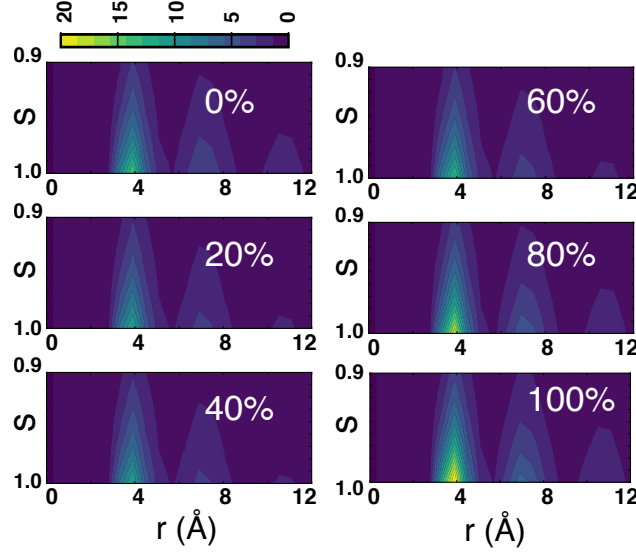
Figure 4.3. $\pi - \pi$ stacking probability presented as a joint probability distribution of distance between the backbone beads on the x-axis and $\pi - \pi$ alignment on the y-axis, as a function of polymer side-chain composition.

mobilities extracted from these simulations reflect the low carrier density limit and do not capture potential finite-concentration effects (e.g., trap filling). The interpretive advantage of the low density limit is that the mobility changes solely reflect the changes in morphology and energy distribution. We note that finite concentration simulations are compatible with the KMC approach, but consider a comparison in the context of OMIECs to be beyond the current scope.

The trend in simulated polaron mobilities shows an increase for the cases with electrolyte percolation ($\geq$40%), and a maximum for the 40% and 100% polar side-chain fractions (Figure 4.4). Nevertheless, these trends are minor in comparison with the similar order of magnitude of the mobility in all of the cases. Experimentally, the electronic mobility generally increases with polar side-chain concentration, but a direct comparison is difficult since the experimental charge carrier densities generally increase with polar side-chain concentration rather than remaining fixed as simulated here.[50, 150] Considering that the mobility typically increases with respect to increasing charge density, the present prediction of relatively constant mobility in the low carrier density limit as the polar side-chain fraction increases is broadly consistent with the experimental trend.

We note that the error bars illustrate the variance across the simulated KMC trajectories, but there is also uncertainty associated with some of the charge transfer parameters used in the KMC simulations (e.g., the electronic coupling, dielectric, and reorganization energy). Thus, we performed a sensitivity analysis by conducting the KMC simulations for several choices of the charge transfer parameters to assess the extent to which the observed trend survives changes in the model parameters (Figure 4.6). This analysis confirms that the mobilities of the 40% and 100% polarity cases are consistently the highest among the systems,
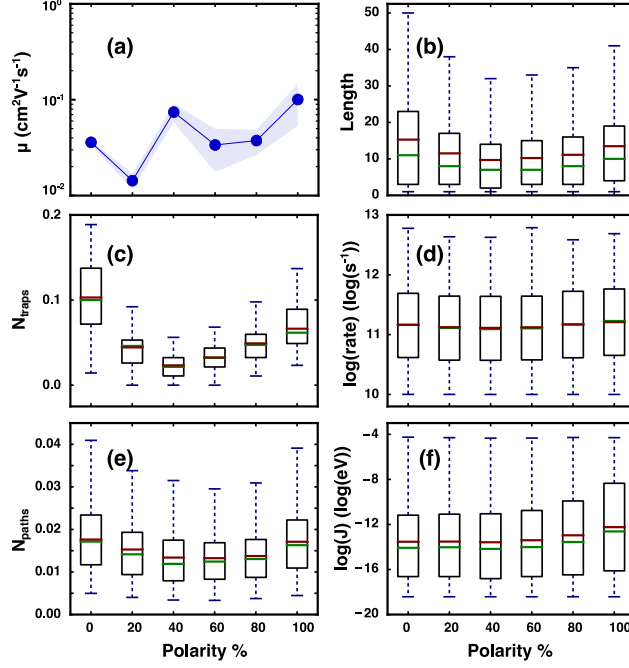
Figure 4.4. (a) charge mobility, (b) polaron delocalization length (number of backbone beads per polaron site), (c) number of trap sites (N_{traps}) normalized by total number of aggregated sites, (d) log distribution of the intersite charge transfer rate , (e) the number of pathways per site (N_{paths}) normalized by the total number of pathways, and (f) log distribution of the intersite electronic coupling, across different side-chain compositions. The green and red line in the whisker plots indicate the median and mean, respectively.

whereas the mobility difference between these two cases is small and within the uncertainty of the various parameters.

To understand the relatively constant mobility as the polar side-chain increases, and secondarily, why the 40% and 100% cases show the highest mobility (~2 times larger that the others), we parsed the distributions of several key quantities across side-chain compositions and KMC configurations, including the polaron delocalization length, number of trap sites (N_{traps}), charge transfer rates, pathways per site (N_{paths}), and electronic couplings (Figure 4.4). The number of trap sites are defined here as polaron sites for which all outgoing charge transfer rates are $< 100ps^{-1}$, which is slower than the relaxation time of the local structure at most operating temperatures. The trap sites are determined after assigning aggregate states (as described in the methods section), and thus, aggregated sites with exit times $> 100ps$ are also considered traps. We emphasize that aggregates and trap sites are not synonymous, since aggregation is only based on whether one or more polaron sites exhibit rapid charge exchange. We find a pronounced trend in polaron delocalization length and the number of trap sites (Figure 4.4b-c). The polaron delocalization length is distinctly smaller for mixed-chain polymers (i.e, 20-80% polar fraction) and can be attributed to the increased torsional disorder reducing the electronic coupling along the backbone (Figure 4.8). This also follows

from the frustrated packing discussed in the previous section for the mixed chain polymers. However, the smaller polarons show more uniformity in rate distributions and fewer trap sites (Figure 4.4c) leading to the non-monotonic trend with higher mobility for an intermediate polarity of 40%.

In contrast, the higher mobility for the 100% polar case is driven by a combination of higher average charge transfer rates and a better connected charge transfer network. The 100% polarity case shows the highest mean charge transfer rate out of the simulated systems (Figure 4d). The increased rates are more apparent in the rate distribution (Figure 4.9), where more interchain pathways ($10^{10} - 10^{13} \text{s}^{-1}$) are seen for the higher polarity cases. This trend is also consistent with the monotonic increase in intersite couplings with the polar side-chain fraction (Figure 4.4f). The second major driver of increased mobility in the 100% scenario is better connectivity within the charge transfer network. This is quantified in an average sense by the distribution of the number of charge transfer pathways for each polaron site within the network (Figure 4.4c). With respect to this measure, there is a monotonic increase in the number of pathways (i.e., connections per node using network terminology) as the polar fraction of side-chains increase. These two factors compete with an increasing number of traps as the polarity increases (Figure 4.4c), with the result that intermediate fractions of polar side-chains exhibit lower mobilities but the homogeneous 100% case consistently shows the highest mobility.

To better quantify the assertion that the charge transport networks are better connected in the more polar systems, we also performed a percolation analysis for the 40-100% polar side-chain fractions that showed electrolyte swelling. This analysis follows earlier work where the number of distinct subnetworks are parsed based on a charge transfer threshold.[147] Specifically, polaron sites were aggregated into subnetworks based on their charge transfer rates, such that each site within a network was connected to at least one other site in that network by the stated rate threshold. Since the charge transfer rates for these systems are not symmetric, a site was incorporated in a subnetwork if there existed at least one charge transfer pathway both to and from the polaron site in the cluster. Interestingly, the 40% apolar side-chain fraction shows the smallest number of subnetworks at all values of the rate threshold (Figure 4.5a), meaning that on average there are more polaron sites in each of its subnetworks. Nevertheless, the spatial extent of the largest charge transport network monotonically increases with polar side-chain fraction (Figure 4.5b), meaning that while the 40 % case has larger networks in terms of the average number of sites per network, they are relatively collapsed and do not effectively percolate the system volume. In contrast, the largest network for the 100% polar side-chain system consistently spans >80% of the system, as measured by its radius of gyration, at all investigated rate thresholds. Thus, the high mobility of the 100% polar system is driven by a combination of higher average charge transfer rates and a more percolative charge transfer network topology.

4.3   Conclusions

The results shown here are consistent with prior experimental observations of the effect of side-chain polarity while providing a detailed picture of the morphology evolution and charge transport. For the random copolymer topology simulated here, we predict that the
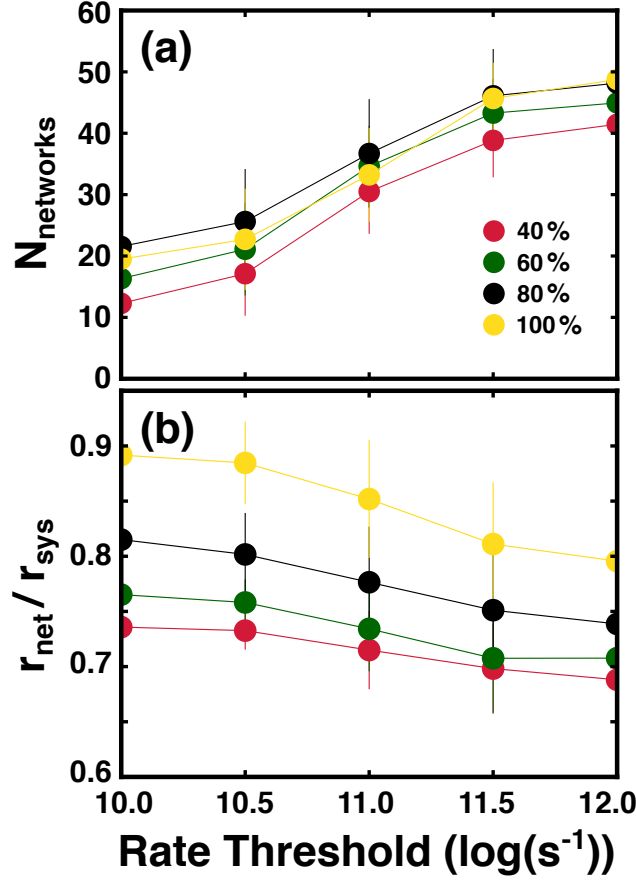
Figure 4.5. The number of distinct subnetworks of polaron sites (a) and the radius of gyration of the largest subnetwork (b) calculated as a function of the side-chain polarity. Subnetworks are defined such that each site is connected to at least one other site within the subnetwork by a rate above the threshold. The radius of gyration of the largest subnetwork is normalized by the radius of gyration the total system (r_{net}/r_{sys}).

highest charge mobility would be achieved for the polymers with the highest fraction of polar side-chains. Moreover, since the conductivity is a figure of merit for OMIECs, the increased doping associated with the higher polar side-chain fraction is an additional advantage. Within the limitations of the simulation conditions, there is no advantage to including a mixture of polar and apolar side-chains; whereas, there are several disadvantages, including increased dihedral disorder and reduced interchain charge transfer rates. A key limitation of these simulations is that the fixed water concentration impedes complete chain dissolution, which is a real problem in actual devices. Thus, the guidance resulting from the present simulations are that the polar fraction of side-chains should be increased up to the point that the film does not dissolve in the electrolyte during operation.

The current study has intentionally excluded from consideration the effect of polar side-chain fraction on the doping efficiency. Instead, the doping efficiency was held constant

for all cases. This is an important topic for understanding the upper limits of achievable transconductance, for optimizing OMIEC morphology, and designing dopants. Indeed, it is obvious from the current simulations that at high apolar side-chain incorporation that the doping efficiency would necessarily be lower due to limited electrolyte penetration; however, for the intermediate-to-high polar side-chain fractions it is unclear how the changes in polymer-dopant configurations would affect the doping efficiency. This is a phenomenon of consequence for side-chain engineering that subsequent simulations studies could foreseeably resolve.

There are also several side-chain engineering strategies that were not assessed in this study that may have more merit than the mixed random copolymer strategy. For instance, block copolymers with separate contiguous blocks of polar and apolar side-chains may potentially provide control over electrolyte percolation without harming the torsional distributions and polymer connectivity. Likewise, it may be possible to introduce apolar and polar segments into individual side-chains. Such a strategy is similar to the contemporary apolar linker concept and side-steps the heterogeneity of random copolymers. The current study is also limited with respect to the exploring the degree of hydrophilicity of the side-chains. It is possible that a lower frequency of more hydrophilic side-chains, or alternatively a higher frequency of less hydrophilic side-chains, might be more optimal than the current best case scenario. The molecular trade-offs for block vs. random copolymer are investigated in the next chapter.

APPENDIX

4.A   Parameters for Non-bonded and Bonded potentials

The LJ parameters used in the simulations are listed in Table 5.1 and on the corresponding Martini bead type. The Gay-Berne parameters are specified according to the LAMMPS notation. The parameters for the Gay-Berne interactions are unchanged from our previous publication, but the y and z axes parameters have been interchanged to comply with the original Gay-Berne functional form.[14] Note that this modification only alters the energy coefficient by a factor of 0.91 and has no significant affect on the physics of the system. The intramolecular parameters used for the simulations are listed in Table 5.2.

Table 4.1. Parameters used for non-bonded interactions in the simulations. Units of all parameters are consistent with LAMMPS 'real' units (kcal mol^{-1}, Å etc.) and potential styles. The bead-type for backbone, polar side-chain, apolar side-chain, and chloroform are C4, EO, SC3 and C4, respectively. Water is made of POL, WM, and WP beads, from which the beads WM and WP do not experience any Van der Waal interactions. The interaction of backbone bead with other backbone beads is modelled as Gay-Berne potential (and not C4-C4 interaction) and the parameter values are reported next to "B-B".

B-B	$\sigma_x = \sigma_y = 5.0,\ \epsilon_x = \epsilon_y = 0.25,\ \sigma_z = 3.0,\ \epsilon_z = 1.2,\ \sigma_c = 3.0$				
	C4	EO	SC3	POL	Qa
C4	$\epsilon = 0.84,$ $\sigma = 4.7$	$\epsilon = 0.70,$ $\sigma = 4.7$	$\epsilon = 0.84,$ $\sigma = 4.7$	$\epsilon = 0.61,$ $\sigma = 4.7$	$\epsilon = 0.74,$ $\sigma = 4.7$
EO		$\epsilon = 0.61,$ $\sigma = 4.3$	$\epsilon = 0.53,$ $\sigma = 4.3$	$\epsilon = 0.79,$ $\sigma = 4.7$	$\epsilon = 0.84,$ $\sigma = 4.7$
SC3			$\epsilon = 0.63,$ $\sigma = 4.3$	$\epsilon = 0.61,$ $\sigma = 4.7$	$\epsilon = 0.65,$ $\sigma = 4.3$
POL				$\epsilon = 0.96,$ $\sigma = 4.7$	$\epsilon = 1.19,$ $\sigma = 4.7$
Qa					$\epsilon = 0.84,$ $\sigma = 4.7$

4.B   Parameters for Charge Transport Simulations

The following parameters were used for the Marcus expressions discussed in the main text: $J_{\text{inter}} = 0.01$ eV, $r_0 = 3.5$ Å, $\alpha = 0.8$ Å^{-1}, $J_{\text{intra}} = 0.31$ eV, $\epsilon_{\text{opt}} = 3$, and $\epsilon_s = 10$. The expressions for external reorganization, λ, and oxidation energy, E, as a function of oligomer

82

Table 4.2. Parameters used for bonded interactions in the simulations. Units of all parameters are consistent with LAMMPS 'real' units (kcal mol^{-1}, Å etc.) and potential styles. The bonds in water molecules between the WM-POL and WP-POL beads are constrained using the SHAKE algorithm [143].

Bonds		Angles	
B-B	$k = 25.0,$ $r_0 = 4.3$	B-B-B	$k = 1.0,$ $\theta_0 = 160.0$
B-SP(SA)	$k = 1.5,$ $r_0 = 4.7$	B-B-SP(SA)	$k = 3.0,$ $\theta_0 = 90.0$
SP-SP	$k = 8.35,$ $r_0 = 3.22$	B-SP(SA)-SP(SA)	$k = 3.0,$ $\theta_0 = 180.0$
SA-SA	$k = 1.5,$ $r_0 = 3.6$	SP-SP-SP	$k = 6.0,$ $\theta_0 = 135.0$
POL-WM	$r_0 = 1.4$	SA-SA-SA	$k = 3.0,$ $\theta_0 = 180.0$
POL-WP	$r_0 = 1.4$	WM-POL-WP	$k = 0.5019,$ $\theta_0 = 0.0$
Dihedrals		Other	
B-B-B-B	$k_1 = 0.5,$ $k_2 = 0.6,$ $k_3 = 0.0,$ $k_4 = 0.0$	B-B, d	$k_1 = 0.0,$ $k_2 = 2.0,$ $k_3 = 0.0,$ $k_4 = 0.0$
SP(SA)-B-B-SP(SA)	$k_1 = 3.0,$ $k_2 = 0.0,$ $k_3 = 0.0,$ $k_4 = 0.0$	B-B, a_1	$k_a = 7.5,$ $\theta_0 = 90.0$
		B-B, a_2	$k_a = 7.5,$ $\theta_0 = 90.0$
		B-SP(SA), a	$k_a = 3.0,$ $\theta_0 = 90.0$

length, l, were obtained from fits to reported data for polythiophene.[182] The governing expression for the external reorganization energy as a function of l is

$$\lambda = \begin{cases} -0.0973\sqrt{l} + 0.4922, & \text{if } l \leq 15 \\ \frac{1.9489}{l} - 0.0092, & \text{otherwise} \end{cases}.$$

The governing expression for the oxidation energy as a function of l is

$$E = 4.1336 \, l^{-0.75} + 4.67.$$

4.C Sensitivity Analysis of KMC Simulations

To assess the impact of the charge transport parameters on the results of the KMC simulations, the simulations were rerun with single factor sensitivity analysis. Specifically, a positive and negative deviation from the values used in the main text simulations was performed for each variable, including the intermolecular coupling decay rate ($\beta = 0.6$ Å^{-1} and 1.0 Å^{-1} are the low and high values, respectively), reference intermolecular coupling strength ($J_{\text{inter}} = 0.005$ eV and 0.02 eV), delocalization length dependence of the HOMO energy level ($E+$ and $E-$ modeled as $3.1l^{-0.75} + 4.67$ eV and $5.167l^{-0.75} + 4.67$ eV, respectively), and the dielectric constant ($\epsilon = 7.5$ and 12.5) while keeping the other parameters at their mean value (Figure 4.6).

For all cases, we observe a mobilities within approximately the same order of magnitude for all side-chain polarities. When parameters are selected that increase the mean charge transport rate (e.g., by increasing the reference coupling J_{inter}, decreasing the coupling decay rate β, or reducing the HOMO energy level decay), we observe the expected increase in mobility, while still observing a non-monotonic trend in the mobility across side-chain polarity composition. The mobility at 100% polarity also increases and becomes predominant compared to the other cases as the charge transport rate increases (Figure 4.6a-b). This corroborates the previous explanation that the mobility is higher at 100% side-chain polarity due to the faster charge transport rate. The the length-dependence of the HOMO energies (Figure 4.6c) more strongly impacts the 40% case than 100% as the latter has longer polaron sites for which posses similar HOMO energies across all three HOMO energy curves tested. Interestingly, we find the mobility to be highest for an intermediate dielectric of 10. Increasing the dielectric constant increases the reorganization energy for charge transfer while reducing the electrostatically induced energy disorder. We refrain from overinterpreting this result, since the treatment of a spatially constant dielectric is already a strong, albeit common, approximation. However we note that both beneficial effects could be achieved by locally reducing the dielectric along the backbone, which is similar to the contemporary strategy of adding an apolar linking group at the base of the side-chain.

4.D Additional Figures Referenced in the Main Text

In the main text we reported the superlinear swelling of the polymer with respect to the polar side-chain percentage (Figure 4.2 in the main text). In Figure 4.7 we report several corroborating datasets that were parsed from the CG-MD trajectories, including the net volume of the swelled polymer at each side-chain composition (Figure 4.7a), the water-water radial distribution function (rdf; Fig 4.7b), and the polymer backbone-ion rdf (Figure 4.7c). The onset of water percolation at 40% polar side-chains followed by superlinear growth is clearly visible from the polymer volumes. Likewise, electrolyte percolation is accompanied by a reduction in bulk water concentration, as evidenced by the monotonic decay in the height of the first peak in the water-water rdf with increasing polar side-chain fraction. The onset of percolation is even more clearly observed in the monotonic increase in the first peak of the ion-backbone rdf with polar side-chain fraction, and the large jump between the 20% and 40% curves. The polymers show very weak lamellar packing for all polar side-chain
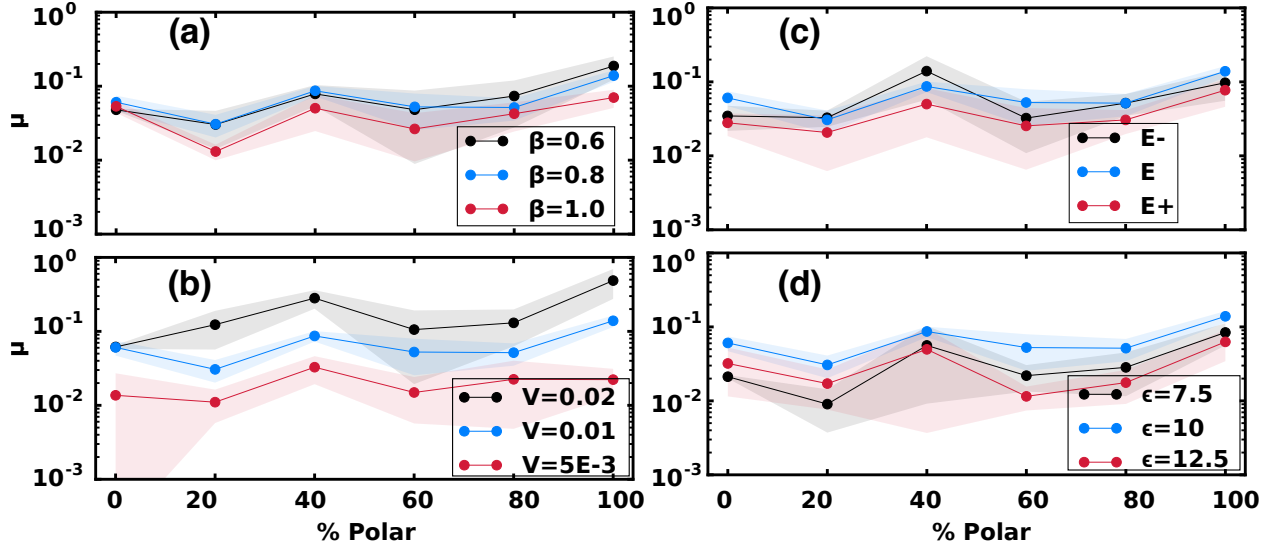
Figure 4.6. Single factor sensitivity analysis of the KMC simulation results for mobility (cm²/Vs) as a function of different side-chain compositions. The effect of the intermolecular coupling decay rate (β[Å^{-1}]; panel a), the reference intermolecular coupling strength (J_{inter} [eV];b), delocalization length dependence of the HOMO energy level (E;c), and the dielectric constant (ϵ;d) are shown in the corresponding panels.

fractions (Figure 4.7d) evidenced by a small peak around 17 Å. This is approximately twice the side-chain length, meaning there is very little interdigitation of the chains, and this lamellar packing is further suppressed as electrolyte enters the polymer film.

As reported in the main text, polaron delocalization is significantly affected by the polar side-chain fraction. This is also reflected in the intramolecular coupling distributions (Figure 4.8), which skew to the lowest values for the 40% case. We ascribe this to the increased torsional disorder caused by side-chain heterogeneity.

The distribution of rates between polaronic sites are reported below before (Figure 4.9a) and after (Figure 4.9b) incorporating the effect of polaron site aggregates. The higher rates of $> 10^{13}$s^{-1} represent the relatively fast transport between localized polarons along the same chain and appear similar for all side-chain compositions before including the effect of aggregates. In contrast, the interchain rates $10^{12} - 10^{13}$s^{-1} are faster for more polar cases due to the improved packing. After aggregating sites where most time is spent, the fast and slow pathways decrease and increase, respectively for the 100% polar case. Note that the intrachain rates above the polaron threshold of 10^{14}s^{-1} represent pairs of intrachain sites where the rate is faster one way than the reverse due to electrostatic disorder.

For completeness, the distributions for the reorganization energy (λ) and site energy differences (ΔE) of charge transfer are also reported in Figure 4.10. As pointed out in the main text, the interchain rate increases with polarity due to improved electronic coupling. The reorganization energy decreases with polarity from 40-100%, due to the increase in polaron delocalization length with increasing side-chain polarity. This trend also partially
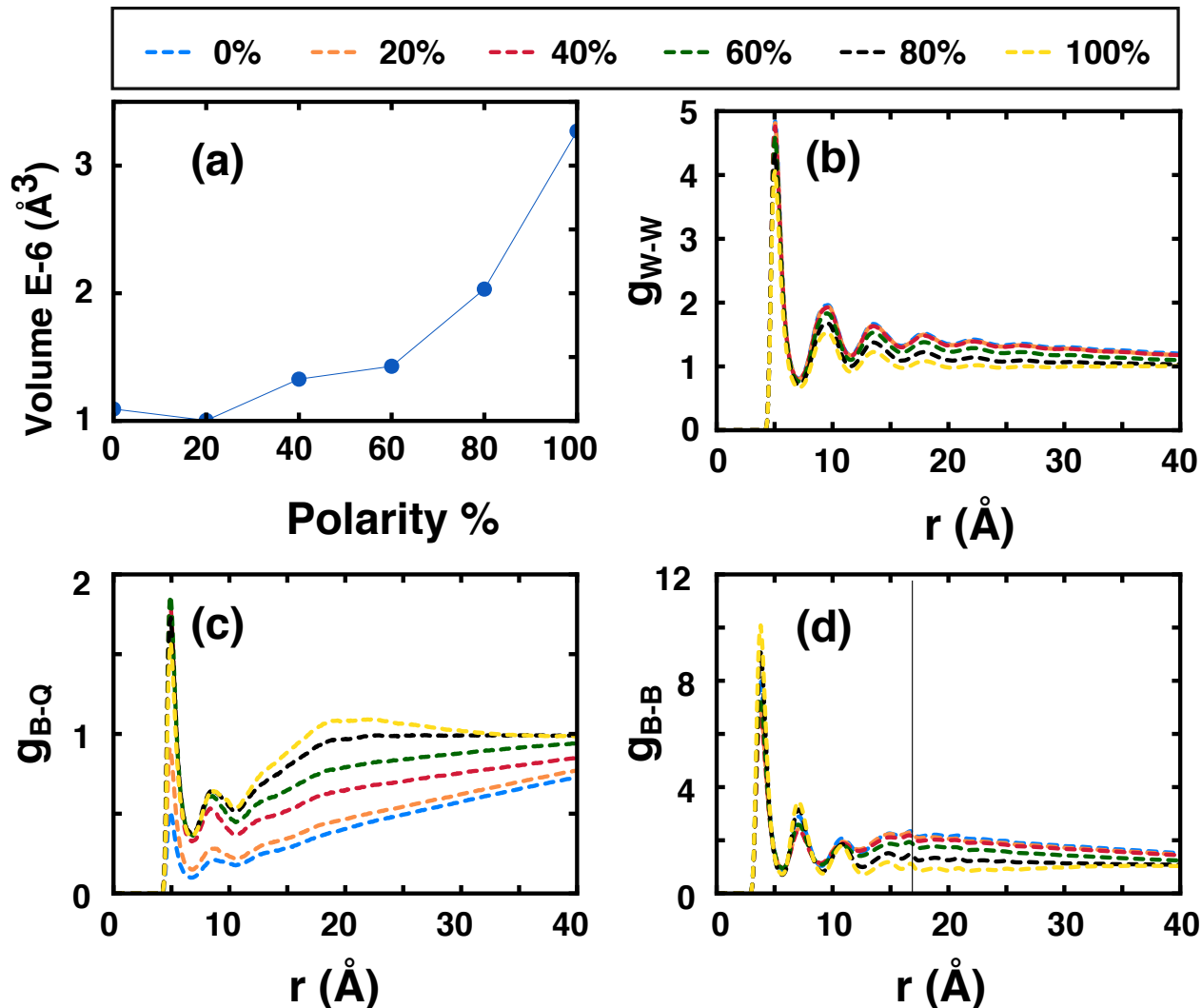
Figure 4.7. (a) The approximate volume of the swelled polymer as calculated from the convex hull of the polymer coordinates parsed from the CG-MD trajectories. (b) The water-water radial distribution function (rdf) as a function of the polar fraction of side-chains. (c) The backbone-ion rdf as a function of the polar fraction of side-chains (d) The backbone-backbone rdf, with the weak lamellar packing around 17 Å indicated by a vertical line.

explains the increase in the mean charge transfer rate with the polar side-chain fraction. There is not a significant correlation between the energies and the observed mobilities with the exception that the 100% polar side-chain scenario shows the lowest average site disorder.

Ion mobility (Figure 4.11) is another important figure of merit for mixed conductor systems. Although the ion mobility is easy to parse from the CG-MD simulations, the interpretation is confounded by the presence of bulk electrolyte in the simulations with low polar side-chain fractions. Specifically, between 0-60% polar side-chain fractions, bulk electrolyte is still present in the simulations which results in the highest ion mobilities in
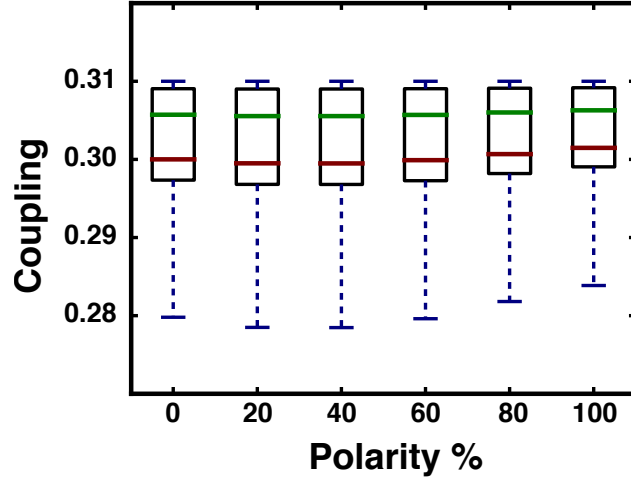
Figure 4.8. Distributions of the intramolecular couplings between polaron sites for different side-chain compositions. The green and red lines indicate the median and mean, respectively.
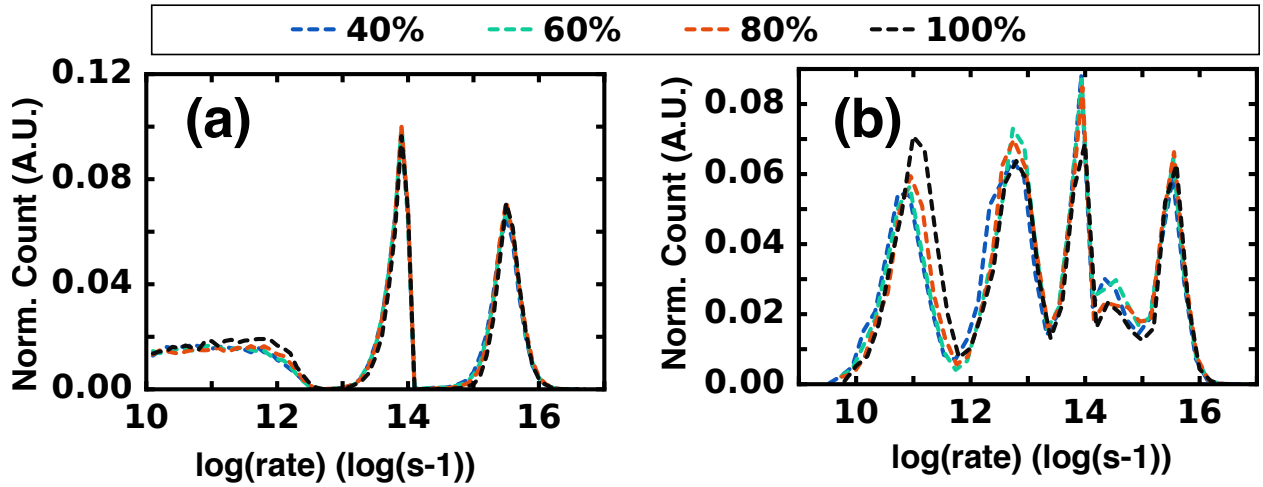
Figure 4.9. Distributions of the polaronic site charge transfer rates (a) before and (b) after incorporating the effect of aggregates.

these scenarios. This is caused by the fact that all of the simulations were equilibrated with sufficient electrolyte to saturate the polymer, leading to bulk electrolyte still existing in the cases of high apolar side-chain loadings and partial percolation. For this reason we have included these results for completeness, but have not significantly discussed them in the main text.
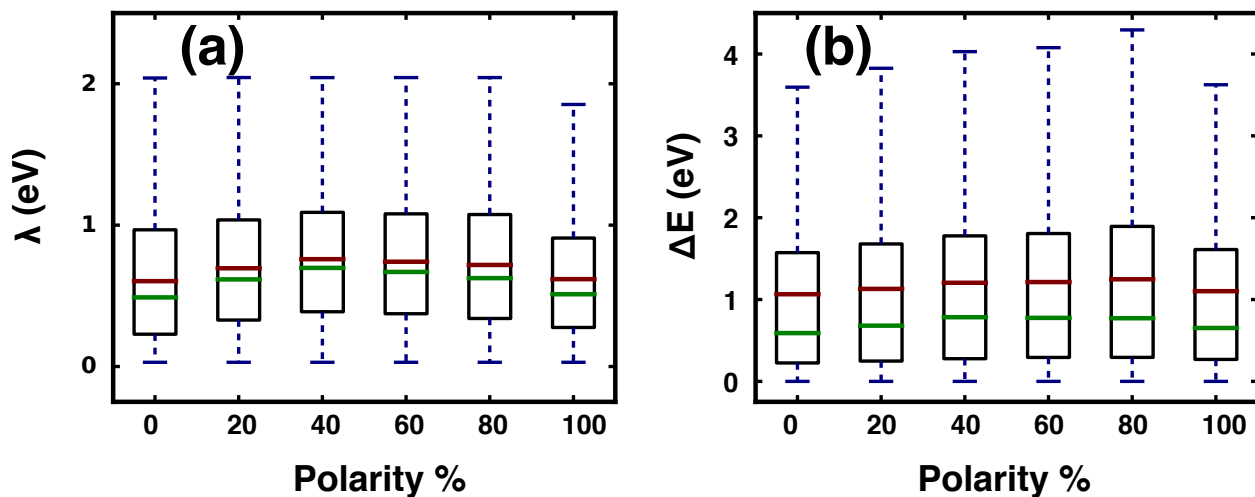
Figure 4.10. Distribution of (a) reorganization energy and (b) energy differences between polaron sites across different side-chain compositions. The green and red lines indicate the median and mean, respectively.
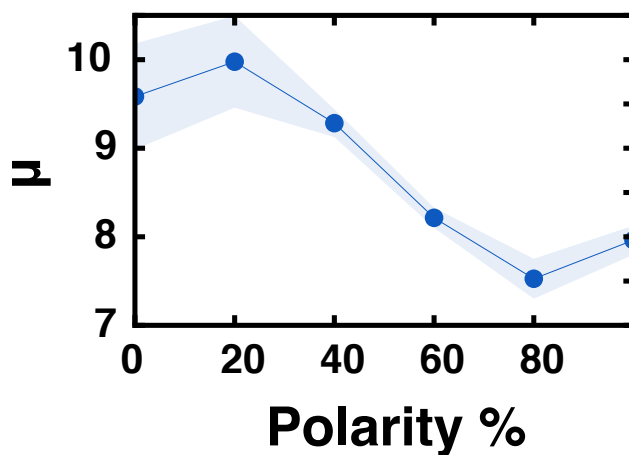
Figure 4.11. The ion mobility ($\times 10^{-5} \mathrm{cm}^2/\mathrm{Vs}$) as a function of side-chain composition. These mobilities were parsed from the CG-MD simulations based on the ionic mean-squared displacements versus time and using the Einstein relation. The standard errors across the trajectories are reported as the shaded regions.

5. DO BLOCK PATTERNED SIDE-CHAINS HELP MIXED CONDUCTOR PERFORMANCE?

In the fourth chapter, we saw that there was no advantage of including mixed polar and apolar side-chains and the backbone with all polar side-chains performed the best both in terms of ion doping and mobility. However, the side-chains were attached in random fashion to all the polymers tested in the previous chapter, similar to the mixed conductors reported in experimental literature. Block patterning or block copolymers have showed interesting properties by obtaining a combination of desired properties using blocks of appropriate chemical moieties. In the context of mixed conductor, there is often a challenge as to maximizing the electrolyte percolation, or in other words, doping while maintaining the connections and aggregation of polymers which allow superb charge conduction. The polar chains have been used to facilitate the electrolyte percolation in mixed conductor. While the random copolymer did not show an advantage of adding apolar chains due to frustrated packing, the immediate question arises if addition of blocks of apolar chains and polar chains may alleviate this disturbance of polymer aggregation. The polar chain blocks can be dedicated to interfacing with electrolyte, while the apolar chain blocks lead to closely packed cluster of polymer with efficient charge transport. Additionally, the polymer stacking and connectivity can be controlled as the polar blocks would interface with polar blocks over apolar blocks, and vice versa. We expect the morphology could unfold in two ways– either the larger blocks of similar side-chains will help larger regions of polymers to aggregate together without any conflict between chemical affinities of neighbors chains, or instead the larger size of the blocks can lead to frustrated packing as the polymer tries to avoid the energy penalty for a large block of apolar chain interfaced in water.

In this chapter, we simulate polymers with varying length of blocks of polymer with polar and apolar side-chains. We perform this analysis while keeping the overall ratio of polar vs. apolar side-chains constant for three compositions– 60% polar (40% apolar), 80% polar (20% apolar), and 50% polar (50% apolar) side-chains. Each OMIEC system was characterized by parsing structural features such as polymer stacking and connectivity, swelling and ion percolation, as well as its ability to transport charge.

5.1 Methods

The coarse-grained force-field, Molecular Dynamics procedure and kinetic Monte Carlo procedure were exactly same as that used in chapter 4, except 5 independent simulations were performed for each case instead of 3. In the following sections, we describe the specific block polymers simulated in this study and very briefly touch upon the methods used.

5.1.1 Coarse-grained Model

In this work, we simulate block patterned polymers with the same proportion of total apolar vs. polar side-chains. Each polymer has a backbone of 50 repeat units, with side-chains attached on every backbone in a regioregular fashion. Starting from a random copolymer as simulated in the previous chapter, we increase the block length. We vary the block length for three ratios of polar to apolar side-chains– 60% polar (40% apolar), 80% polar (20% apolar), and 50% polar (50% apolar) polymers (Figure 5.1). The 60% polar case is simulated as it has almost equal polar and apolar side-chains, to realize the effect of block-length, and the random copolymer for this ratio is not one of the high performing models with both poor swelling and mobility. We also modulate the block length for the 80% polar (20% apolar) case as it showed swelling sufficient to be characterized as a mixed conductor but not the highest mobility. For these two compositions, we simulate four polymers each. The first being a random copolymer, followed by block copolymers of increasing length and a fourth polymer where all the apolar chains are sandwiched between polar chain blocks. For instance, for the 60% case (Figure 5.1c), for the 50 repeat unit backbone simulated here, we have a total of 30 polar and 20 apolar side-chains. The random copolymer is referred as random. The block copolymer with smaller blocks is referred as p7a5 and contains consecutive blocks of 7 polar, 5 apolar, 8 polar, 5 apolar, and again, 7 polar, 5 apolar, 8 polar, 5 apolar chains. The next block copolymer p15a10 contains larger blocks of 15 polar, 10 apolar, 15 polar and 10 apolar side-chains. The fourth polymer, p15a20p15, contains a block of 20 apolar side-chains sandwiched between two blocks of 15 polar side-chains. We hypothesize that adding chemically similar (polar) blocks at both the ends will improve polymer connectivity by allowing more connections. We have analogously designed and named polymers

random, p10a2, p20a5, p20a10p20, for the 80% case (Figure 5.1d). In addition to these two ratios, we also simulate the 50% polar (50% apolar) but we do not vary the block length for this composition and instead the relative placement of the side-chains across the backbone (Figure 5.1b). The p1a1 case shown in Figure 5.1b, has a regioregular polymer with polar and apolar side-chains on alternate backbone beads, but on the opposite side of backbone, providing a new mechanism for improving stacking and connectivity.
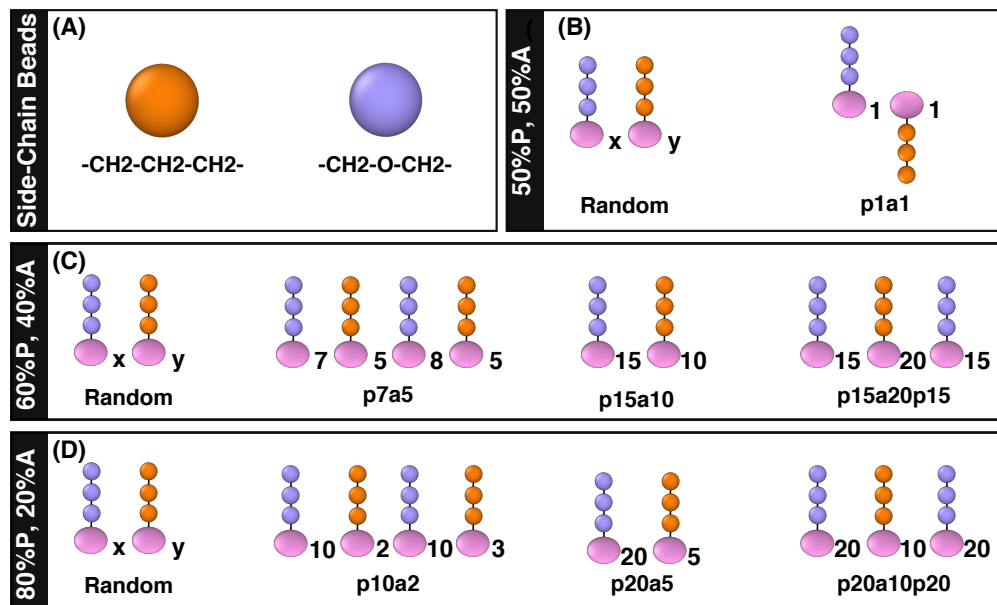
Figure 5.1. (a) The coarse-grained mapping of the apolar and polar beads used to model the hydrophobic and hydrophilic polymer side-chains, respectively. Polymers are simulated in this study with different patterning of polar and apolar side-chains for (b) 50% polar and 50 % apolar side-chains, (c) 60% polar and 40% apolar side-chains and, (d) 80% polar and 20% apolar side-chains.

The coarse-grained model and parameters is same as used for the side-chain composition study in Chapter 4. The backbone bead roughly represents a thiophene ring, and is represented by a single anisotropic bead. The side-chains are placed on each backbone. The polar side-chain and apolar side-chains are represented by a polyetheylene glycol (PEG) chain of roughly 3 repeat units and nonyl chain, respectively. First, the polymer film is formed by processing it in chloroform solvent, and then the polymer is charged and equilibrated with an electrolyte solution of water and chloride ion. The backbone-backbone interactions are

given by Gay-Berne potential parameterized in Chapter 3. Rest of the non-bonded interactions are isotropic and are either adopted from Martini or parameterized in Chapter 3-4 to reproduce experimental data similar to Martini. For the cross interactions between backbones and other beads, backbone bead is treated as Martini bead of type C4. The polar and apolar side-chains are represented by 3 beads, each of type EO and SC3, respectively. Each chloroform molecule is represented by a C4 bead, 4 water molecules are represented by a 3 bead molecule from Martini polarizable water model, and chloride ion is represented by a Qa bead. Bonded interactions are a combination of Martini force-fields and additional force-fields to modulate the torsional order along backbone and side-chain regioregularity. The force-fields used in this study are reported in Section 5.A.

5.1.2 Molecular Dynamics

LAMMPS was used to perform all molecular dynamics simulations.[130] All simulations used a 10 fs integration time step, Velocity-Verlet integration, and periodic boundary conditions. Lennard-Jones, electrostatic and Gay-Berne interactions were truncated at 12 Å. The Lennard-Jones and Coulombic interactions were shifted using the standard GROMACS shift function over a length of 9-12 Å and 0-12 Å, respectively. Film equilibration occurred over two distinct simulations. First, polymer film formation was simulated by dissolving the neutral polymers in chloroform then evaporating the solvent. Subsequently, the polymers were doped (i.e., charged) and electrolyte was added to the simulation and allowed to equilibrate.

For the polymer film formation simulations, the polymer and chloroform solvent molecules were initialized from diffuse configurations, using a cubic grid to place molecules in random orientations without overlap. Each simulation box contained 50 polymers and was initially surrounded by 50,000 chloroform molecules. The velocities were initialized from a uniform distribution obtained with a random seed value and scaled to give the correct kinetic energy. The simulations were first relaxed in the NVE ensemble with restrained displacements of 0.1 Å per time step for 10 ps. This was followed by a simulation in the NVT ensemble at 300 K while linearly rescaling the simulation box over 2 ns to obtain a condensed system with a density of 1 g/cc. The density was then equilibrated in the NPT ensemble for 100

ns at 300 K and 1 atm. After the NPT simulation, 20 wt% of the chloroform beads were randomly selected and removed from the system, followed by an equilibration step in the NPT ensemble for 100 ns at 300 K and 1 atm. This process of chloroform removal followed by NPT equilibration was performed five times, resulting in 28,750 (80 wt% of solvent), 10,781 (60%), 4,792 (40%), 1,797 (20%), and 0 (0%) chloroform molecules, respectively. This procedure is intended to qualitatively mimic the solvent processing and evaporation during experimental film formation (e.g., spin-coating) and has been employed in other coarse-grained simulations of organic polymers.[107]

The final polymer film configuration was unwrapped in accordance with the periodic boundary conditions, the backbone beads were charged and then surrounded by randomly placed solvent and ion beads on the outside of the polymer film, each within a separate cubic box of length 6 Å to avoid overlaps. At this stage, 10,781 CG water beads were added, which correspond to 43124 real water molecules and 60 wt% of solvent. Each backbone unit carried a charge of +0.2 (i.e., a doping induced charge every five backbone beads), and 500 counterions with a charge of −1.0 were added to achieve charge neutrality. These systems were relaxed in the NVE ensemble with restrained displacements of 0.1 Å per time step for 10 ps, followed by an NPT equilibration for 100 ns, and an NPT production run for an additional 100 ns at 300 K and 1 atm. In all of the NVT and NPT simulations, the Nosé-Hoover thermostat and barostat were employed using the modified form proposed by Martyna, Tobias, and Klein, as implemented in LAMMPS [102], with a relaxation time constant of 0.15 ps and 1.5 ps for the thermostat and barostat, respectively. In all of the simulations, the thermodynamic data and coordinates were sampled at 1 ps. The ASPHERE package in LAMMPS was used for all Gay-Berne force, torque evaluations, and numerical integrations.[22] The SHAKE algorithm [143] implemented as a part of the RIGID package in LAMMPS was used to constrain the bonds in the water molecules. Additional bonded potentials to control the orientation of neighboring beads were implemented as custom classes in LAMMPS by our group. OVITO was used for visualizing the final trajectories and generating snapshots of MD configurations.[160]

5.1.3 Charge Transport Calculations

The charge transport within the OMIEC polymer network is simulated in the configurations obtained from MD simulations using kinetic Monte Carlo (KMC) algorithm. In order to simulate the trajectory, essentially, we account for three key transport processes. First and the fastest being delocalization of electrons along the backbone. The intramolecular charge transport rate is determined between consecutive repeat units along the backbone using Marcus equation in adiabatic limit. If this intramolecular rate is above a threshold along consecutive repeat units, the charge can be assumed to be delocalized along this section of polymer throughout the charge transport simulation. These section of repeat units then form individual polarons, and the charge can transfer between polarons through hopping mechanism. The charge hopping is our second transport process of interest, and the rate of charge transfer between polarons is calculated using Marcus rate equation. Among these hopping transport pathways, we can have significant timescale separation, where a charge may remain for a significant time on an individual polaron or subnetworks of polarons, before escaping to less favorable pathway or until the polymers reorganize themselves. This separation of timescale is the third transport process of interest. We account for these aggregates and the time the polaron spends in these aggregates. Once all these quantities are determined, for a given configuration, one can simulate the hopping of charge across polarons throughout the polymer network. Multiple charge trajectories are simulated for each MD configuration by choosing the first site at random, to sample all the subnetworks within the OMIEC polymer network. The mean square displacement (MSD) of the charge is then calculated using multiple trajectories for multiple MD configurations. From the MSD, the mobility of the charge is parsed using Einstein equation. This analysis is performed over five independent simulations and an average charge mobility is obtained. The Marcus rate equations, their physical interpretation, their specific parameter as well as the sampling details can be found in Chapter 4.

5.2 Results and Discussion

In this section, we compare the morphology and charge transport features across various side-chain block patterning for given polar vs. apolar side-chain compositions. We start the discussion with the 60% polar (40 % apolar) side-chain composition due to the comparable number of polar and apolar side-chains, and hence, it can be expected to be strongly affected by patterning. Afterwards, we discuss the 80% case, for which the random copolymer itself shows significant swelling from Chapter 4 and check if altering block length has any further consequences on polymer performance. Finally, we discuss the 50% case, for which a completely different approach for patterning is proposed.

5.2.1   60 % Polar - 40 % Apolar

In this study, we tune the block size of polar and apolar side-chains to control the swelling of polymer, ion percolation and polymer aggregation, which consequently alter the charge conducting properties of the polymer network. Hence, first we report the morphological characterization for the polymers studied.

Morphology Characterization

We hypothesized in Section 1 that the polar (apolar) blocks would preferentially stack with chemically similar polar (apolar) blocks. In Figure 5.2a, we report representative snapshots of all the polymers, and we find that this is indeed the case. As we move up in the block length, larger regions corresponding to each side-chain type denoted by the orange and purple blocks appear. For the random case, very small orange and purple regions distributed all across the polymer network, indicating polymer blocks containing apolar and polar chains, respectively. The polymer also appears to form a relatively closed off globule with moderate swelling (Figure 5.12). For p7a5, the size of the two side-chain blocks increases and we find some region of polymer extending into the solvent. For the case p15a10 and p15a20p15, we find very large regions of apolar (orange) side-chain attached polymer blocks surrounded by polar side-chain (purple) regions. For p15a20p15, we observe these polar side-chain attached

polymer segments extending out from the globule into the solvent. Thus, the polar regions at the end selectively interface with solvent increasing the area in contact with solvent as expected.
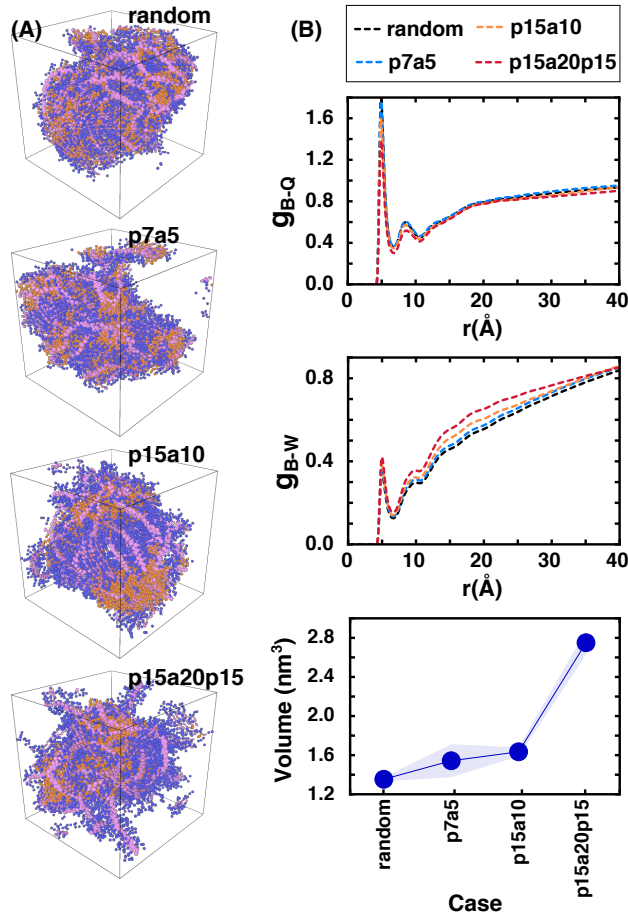
Figure 5.2. (a) Snapshots of representative configurations of random, p7a5, p15a10 and p15a20p15 configurations showing how the polar and apolar blocks of backbone have come together. Only the backbone bead, polar and apolar side-chains are shown for clarity, and are indicated by the pink, purple and orange bead, respectively. (b) Characterization of electrolyte percolation: (top, center) The radial distribution function (RDF) of backbone with ion and water to quantify the electrolyte percolation in the polymer film. The y-axis g_{x-y} denotes RDF between moiety x and y and, the backbone bead, ion and water are denoted by B, Q and W, respectively. (bottom) Volume of the swollen polymer after electrolyte percolation.

Next, we inspect how this distinct morphologies affect the percolation of water and more importantly, that of ions. The radial distribution function (RDF) for backbone bead with

water and ion is reported in Figure 5.2b. We find that the RDF for backbone and water increases with block length due to the separate large polar blocks and in agreement with the extended polymer network discussed above. These changes in percolation are however minimal and all configurations except the p15a20p15 case show similar swelling (Figure 5.2b). The p15a20p15 case shows swollen volume of the swollen film more than the 80% random configuration (discussed below) and close to the 100% case from previous chapter. The entire polymer network doesn't swell similar to the 80-100% cases and hence, the changes in RDF are quite weak but we see these extended fibers of polar blocks which lead to similar effective volume. Thus, a definite increase in polymer-electrolyte interaction can be achieved compared to the random case by adding polar blocks at the end, as they extend into the solvent from the polymer clusters. However, the block length has an opposite affect on net ion percolation as seen in the backbone-ion RDF in Figure 5.2b, which shows a decrease with increase in block length. To inspect this further, we parse the RDF of backbone bead attached to apolar and polar side-chains with ion separately (Figure 5.13). More ions percolate close to the backbone bead attached to polar side-chains for polymers with larger block length, i.e. p15a20p15 and p15a10. This is congruent with higher contact between these polar blocks and solvent seen in morphologies and backbone-water RDF. On the contrary, more ions percolate next to the backbone beads attached to the apolar side-chains, smaller the block length. The presence of polar(apolar) side-chains next to side-chains increases(decreases) their effective polarity. This was observed in the side-chain composition study in Chapter 4 and can also be observed from polar and apolar side-chain RDF with ion. The polar and apolar side-chain RDF increases and decreases, respectively for polymers with higher block length due to higher and lower respective probability of presence of neighboring polar side-chains (Figure 5.13). The increase in percolation in polar regions is limited and cannot compensate for the reduced percolation in shielded blocks of backbone connected to apolar side-chains, leading to an overall reduction in ion percolation. The percolation difference is higher for apolar blocks than polar blocks; for example, the difference between random and p15a20p15 case is ~ 1 and ~ 0.4 for apolar block-ion and polar block-ion RDFs, respectively. We stress that the difference in both solvent and ion percolation across the polymers is weak but there is a definite qualitative difference, which may get amplified in a real device.
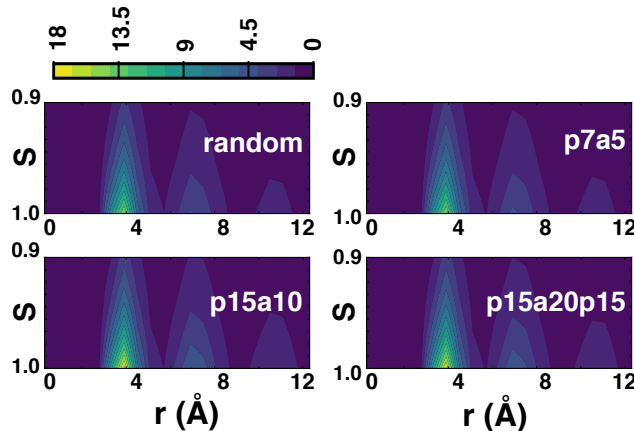
Figure 5.3. $\pi - \pi$ stacking probability presented as a joint probability distribution of distance between the backbone beads on the x-axis and $\pi - \pi$ alignment on the y-axis, as a function of polymer side-chain patterning

Finally, we analyze how the polymer π-π stacking and connectivity varies across these distinct morphologies. In Figure 5.3, we report the joint probability distribution of distance between backbone beads on x-axis and the π-π alignment between them, given by dot product of the orientation vector of backbone beads, on the y-axis. We see higher π-π alignment as indicated by the brighter yellow color for S ∼ 1.0 as well as improved connectivity at distanced of 12 Å for the polymer cases p15a10 and p15a20p15. The larger blocks of polar and apolar blocks lead to stacking of larger backbone regions which is reflected in the improved π-π stacking between polymer backbones. These large blocks also allow for more connections between polymer clusters reflected in the higher structural order seen at farther distances. The p15a20p15 polymer has polar blocks on both the ends and should result in more connections. It does show extended segments of polymers in solvent (Figure 5.2a) but this is not reflected in the polymer connectivity. Also, the differences in the polymer stacking and aggregation across all polymers are minor, similar to the percolation and swelling trends.

Charge Transport Characterization

In this section, we analyze how these emergent side-chain distribution and effective polymer aggregation alter the charge transport through the polymer network. A trajectory of polymer jumps is simulated using KMC algorithm for individual MD configurations. We

explore the hops of a single hole in a network of polarons with fixed charge and limited dopants sufficient to satisfy the zero net charge constraint in MD simulations. As discussed in Chapter 4, this is representative of a low carrier density limit, and does not account for trap filling. However, the analysis provides us useful insights such as trends in electronic coupling, number of pathways and trap sites with respect to the block size of side-chains.
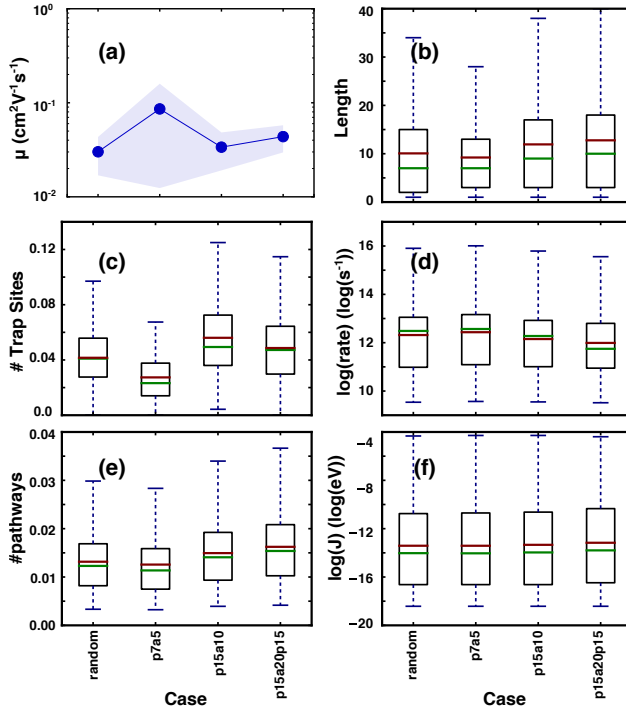
Figure 5.4. (a) charge mobility, (b) polaron delocalization length (number of backbone beads per polaron site), (c) number of trap sites (N_{traps}) normalized by total number of sites, (d) log distribution of the intersite charge transfer rate , (e) the number of pathways per site (N_{paths}) normalized by the total number of sites, and (f) log distribution of the intersite electronic coupling, across different side-chain patterning. The green and red line in the whisker plots indicate the median and mean, respectively.

In Figure 5.4a, we see the averaged mobility across polymer block configurations and the maximum is observed for p7a5 case while all other cases are comparable. Overall, the average mobility are on the same order and all the mobility values are within uncertainity. However, to rationalize the observed trend we parse relevant quantities such as the polaron delocalization length, number of pathways and trap sites, charge transfer rate and electronic coupling. We find that the average polaron delocalization length (Figure 5.4b) is smaller and

99

number of trap sites are fewer for the p7a5 case (Figure 5.4c) which shows the maximum mobility. A strong correlation between the these three quantities was also observed in our previous study on side-chain composition. The 40% polar side-chain composition also showed the smallest average polaron length, the least number of trap sites and one of the highest mobility. The delocalized polymer segments form highly coupled clusters which can act as trap sites from which the hole never escapes or spends significant time before escaping. In the previous work, we observed higher mobility for the 100% polar case as well, due to a combination of multiple effects such as higher electronic coupling, better connected network and longer spatial extent of charge jumps. We saw an interplay of two competing factors for the 100% case– it shows polymer connectivity and delocalization highly conducive to charge transport but also very well stacked polymer clusters which act as trap sites. The p15a20p15 case is the analogous to 100% polar case in this study. It has similarly higher number of pathways (Figure 5.4e) due to improved connectivity as evident from the morphological characterizations discussed above. However, the improvement in connectivity is not prominently high and not reflected in the electronic coupling, and consequently, the charge transport rate, leading to an overall low mobility. We find the charge transport rate (Figure 5.4d) is rather slightly lower for p15a20p15 case and this is again due to the presence of aggregates. In Figure 5.14c, we report the charge transport rate before determining the aggregates and the rates are quite similar across all cases. These charge transport rates are primarily a reflection of the coupling which remains uniform across the polymers. We have also reported the free energy differences between polarons and reorganization energy which both decrease with block length due to increased average delocalization length, but do not affect the pre-aggregation charge transfer rate (Figure 5.14). After accounting for the extensive time spent in certain aggregates, we find the average charge transfer rate drops for p15a20p15 as shown in Figure 5.4d.

We also performed analysis on the quality of percolation in the polymer network by parsing the disconnected subnetworks in the polymer network for varying charge transfer rate threshold (Figure 5.5). The analysis was performed as described in Chapter 4. Higher number of subnetworks imply a broken polymer network with poor charge transport capability. We observe lowest number of subnetworks for p7a5 while other cases show uniformly higher
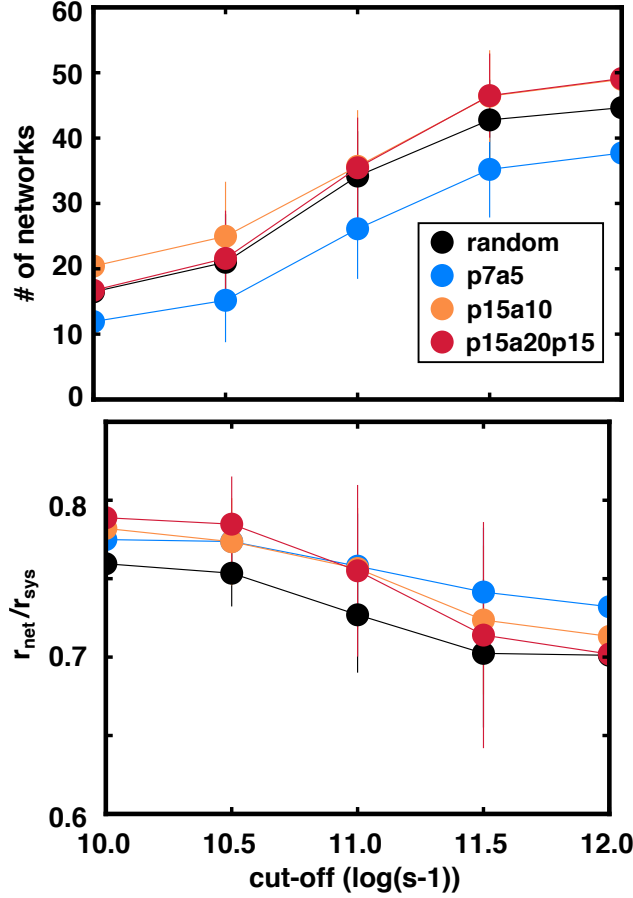
Figure 5.5. (a)The number of distinct subnetworks of polaron sites and (b) the radius of gyration of the largest subnetwork calculated as a function of the side-chain patterning. Subnetworks are defined such that each site is connected to at least one other site within the subnetwork by a rate above the threshold. The radius of gyration of the largest subnetwork is normalized by the radius of gyration the total system (r_{net}/r_{sys}).

subnetworks across different rate thresholds. This is consistent with the observed trend for mobility and indicates that the trap sites significantly affect the percolation through the network. The spatial extent of the largest network is also reported to compare the spatial span available to the charge. For smaller rate threshold, we find the p15a20p15 case has the highest spatial extent attendant to its better morphology which drops as we approach higher threshold at 11 s^{-1}. Thus, while the p15a20p15 case shows longer delocalization along the polymer and farther connections, the network is not resistant enough as it loses some connections to the trap sites. In Chapter 4 as well as this study, we found trap sites

have serious impact on charge transport capability of the polymer but in a real system, this impact may dampen due to trap-filling which cannot be accounted for by our simulations. Thus, the p15a20p15 may end up being the case with highest mobility, but it still doesn't compare with the completely polar case discussed in Chapter 4 both in terms of mobility and capacitance.

5.2.2 80 % Polar - 20 % Apolar

In this subsection, we repeat the analysis above over the polymer series for 80% polar and 20% apolar side-chain compostion.

Morphology Characterization

For the 80% case, the changes in structural arrangement of polymer with respect to patterning of side-chains are exactly similar to the 60% polar case but much weaker (Figure 5.6a), as most of the backbone is covered by polar side-chains. We observe regions of similar side-chains blocks stacked together and the size of these regions increases with block size. For the case of p20a5p20, with apolar block sandwiched between polar blocks, we see extending fibers of polar blocks of polymer from the polymer network, similar to the p15a20p15 case. Similar to the 60% case, these changes in morphology are reflected weakly in the backbone-water RDF, and we see a definite rise in the increase in volume of the swollen polymer (Figure 5.6b). The volume of the p20a10p20 case is comparable to the 100% case studied in previous chapter. We also see a non-monotonic trend with respect to the block length, as the p10a2 configuration performs better than p20a5. The percolation of ion also remains largely similar across all the cases, except for a slight decrease in the immediate neighbors for p20a10p20 (Figure 5.6b). As in the 60% case, we attribute this to reduced ion percolation near apolar blocks with minimal changes in both swelling and ion percolation near polar regions (Figure 5.15). The neighboring side-chains locally alter the effective affinity of the polymer with electrolyte leading to changes in overall polymer morphology and solvent percolation. For instance, the water is more uniformly distributed across the polymer for a smaller block length as the apolar chains are more likely surrounded by polar chains. The p10a2 and

Figure 5.6. (a) Snapshots of representative configurations of random, p10a2, p20a5 and p20a10p20 configurations showing how the polar and apolar blocks of backbone have come together. Only the backbone bead, polar and apolar side-chains are shown for clarity, and are indicated by the pink, purple and orange bead, respectively. (b) Characterization of electrolyte percolation: (top, center) The radial distribution function (RDF) of backbone with ion and water to quantify the electrolyte percolation in the polymer film. The y-axis g_{x-y} denotes RDF between moiety x and y and, the backbone bead, ion and water are denoted by B, Q and W, respectively. (bottom) Volume of the swollen polymer after electrolyte percolation.

p20a5 swelling behavior are quite similar and large differences are seen only for p20a10p20 which suggests that beyond a certain number of side-chains (5-10), modulations in effective hydrophilicity are dampened. Further, the differences in morphology are not reflected in both π-π stacking and polymer connectivity (Figure 5.7). In Figure 5.7, we see the color map at higher alignment ($S \sim 0.1$) and peak size at distances of 12 Å is similar for all four

103

Figure 5.7. $\pi - \pi$ stacking probability presented as a joint probability distribution of distance between the backbone beads on the x-axis and $\pi - \pi$ alignment on the y-axis, as a function of polymer side-chain patterning

cases. Hence, the polymer network and water distribution around it remains largely similar irrespective of how the side-chains are connected for the net composition of 80% polar case. Only p20a5p20 case stands out with slightly higher swelling and lower overall ion percolation.

Charge Transport Characterization

In Figure 5.8a, we report the mobility across the different block copolymers. Surprisingly, the highest mobilty is observed for the random configuration while all other cases are comparable. Again, the differences in the mobility are minor and within the error across all cases. On closer investigation, we do not find any strong indicator as to why the random configuration outperforms even in an average sense. The average number of trap sites (Figure 5.8c) is lowest for the random configuration, but marginally, and the range of number of trap sites for it is broader than p10a2 and p20a5. The polaron delocalization length (Figure 5.8b) is only slightly lower for random case than p10a2, whereas the charge transport rate (Figure 5.8d) is similar and the number of pathways (Figure 5.8e) are slightly lesser than the p10a2 and p20a5 cases. Really, the p20a10p20 case is the one which stands out through all these comparisons. It shows the highest delocalization of polaron, most number of pathways but also most number of trap sites. However, these differences are not reflected in the net mobility. One explanation can be that any increase in number of pathways is rendered

Figure 5.8. (a) charge mobility, (b) polaron delocalization length (number of backbone beads per polaron site), (c) number of trap sites (N_{traps}) normalized by total number of sites, (d) log distribution of the intersite charge transfer rate , (e) the number of pathways per site (N_{paths}) normalized by the total number of sites, and (f) log distribution of the intersite electronic coupling, across different side-chain patterning. The green and red line in the whisker plots indicate the median and mean, respectively.

fruitless due to the presence of trap sites. However, in order to rationalize why the random configuration has highest mobility, we compare only the cases with highest mobility for all polymers in Figure 5.16. This comparison shows a more prominent trend in trap sites with minimum trap sites for the random configuration, followed by p20a5, p10a2 and p20a10p20 with an identical trend for mobility. Thus, we see, that trap sites again play a huge role in dictating the overall charge transport and the random case displays lowest of it. However, in terms of intramolecular charge delocalization and intermolecular polymer connectivity, the p20a10p20 case stands out due to its extended polymer fibers obtained by large polar blocks at the end. For the sake of completion, we also report the percolation analysis for this polymer series (Figure 5.17). Neither the number of networks nor the spatial extent stand out for the random configuration but they do demonstrate the adverse impact of trap

sites for the p20a10p20 case. It has a broken network as indicated by the most number of subnetworks and for lower charge rate cut-off, it has the most extended polymer network but the networks quickly breaks down into these subnetworks.

5.2.3 50 % Polar - 50 % Apolar

Finally, we discuss a new type of side-chain patterning, where the placement of side-chains across the backbone is controlled instead of along the backbone, for 50% polar and 50% apolar side-chain compostion.

Morphology Characterization

Both random and p1a1 polymer configurations do not show very high swelling which is the key characteristic of mixed conductors and form a closed off polymer globule (Figure 5.9a). However, within the globule we find notable differences in the morphology, where p1a1 shows distinct polar (purple) and apolar (orange) regions and are highly elongated, while the polar and apolar regions are mixed and disordered in the random configuration. These elongated regions in p1a1 have important consequences to the polymer packing, as they denote well stacked sides of polymers, as marked by the red circles in Figure 5.9a. Another difference is the outmost layer of the polymer network is covered with mostly polar side-chains in the p1a1 case whereas a mixture of chains is seen for random case. The side of the polymer p1a1 covered with polar side-chains solely interfaces with the solvent while the apolar regions remain inside and stack with each other.

The morphological reflections further reflect in the percolation of solvent and ion through the network, the polymer aggregation and connectivity. These polymers are largely phase separated into a polymer and a solvent phase with limited swelling (Figure 5.9b). However, at large distances both solvent and ion percolation is higher for the p1a1 case than the random configuration (Figure 5.9b). The primarily hydrophilic outer surface of the p1a1 polymers will attract more water and can conceivably explain the increase in RDF only at longer distances. This is further corroborated by the lower volume of the swollen film, which indicate the solvent-polymer interaction is maximized only on the outer surface of polymer
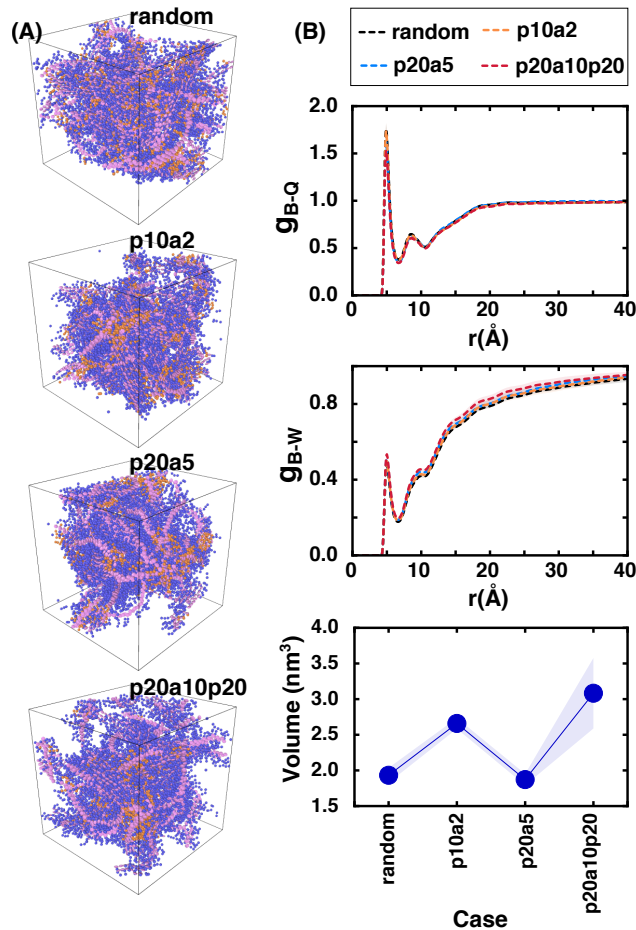
106
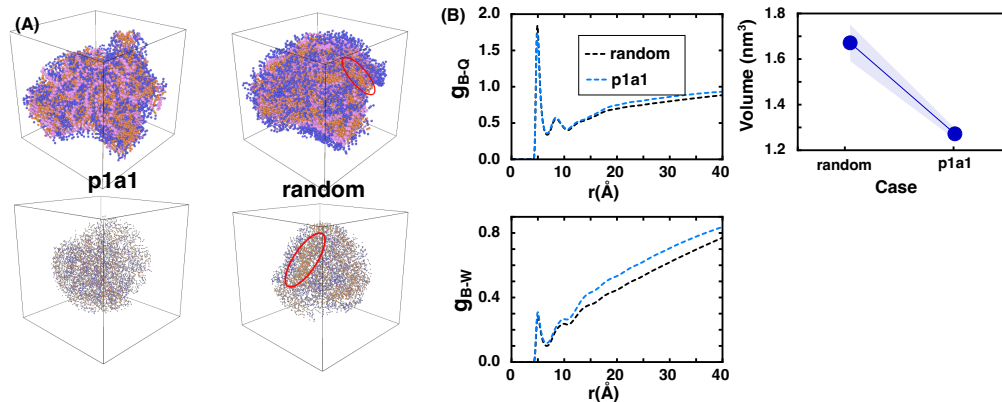
Figure 5.9. (a) Snapshots of representative configurations of random, and p1a1 configurations showing how the polar and apolar blocks of backbone have come together. In the top snapshots, only the backbone bead, polar and apolar side-chains are shown for clarity, and are indicated by the pink, purple and orange bead, respectively. In the bottom one, even the backbone beads are removed and only a portion of polymer network is shown to highlight the stacking of similar side-chain blocks. These stacks are marked in the snapshots by red circles. (b) Characterization of electrolyte percolation: (left) The radial distribution function (RDF) of backbone with ion and water to quantify the electrolyte percolation in the polymer film. The y-axis g_{x-y} denotes RDF between moiety x and y and, the backbone bead, ion and water are denoted by B, Q and W, respectively. (right) Volume of the swollen polymer after electrolyte percolation.

network wherease the net polymer swelling remains limited. The polar and apolar side-chain RDF with ions is higher and lower for p1a1, respectively (Figure 5.18). The effective hydrophilicity, as discussed for the 60% and 80% cases, is higher for the apolar side-chains in random case than p1a1. Even though the polar and apolar side-chains appear on alternate backbone beads for p1a1, the apolar blocks being on the other side of the polymer, are shielded from the water. The dominant impact of this kind of patterning is however seen for the polymer morphology (Figure 5.10). We observe a significant rise in the $\pi - \pi$ stacking between the immediate and alternate neighbors around 3.8 Å and 7 Å, respectively. The connectivity is also improved in p1a1 configuration as seen by a larger peak around 11 Å than for random case. The overall backbone morphology looks more ordered in the p1a1 than the random case and we also see higher number of individual polymer π-aggregates (Figure 5.10). In the p1a1 case, large sections of polymer are available for favorable packing
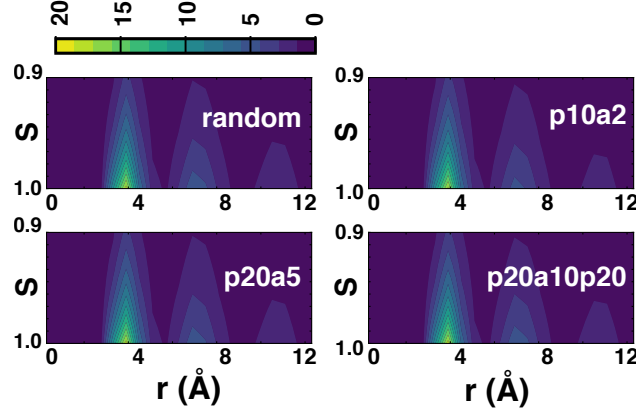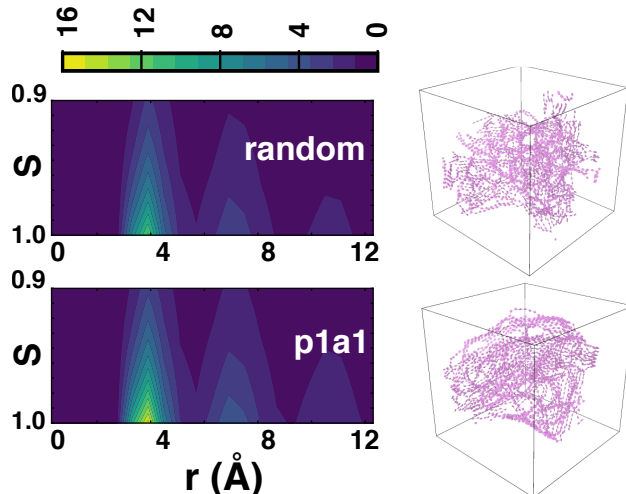
Figure 5.10. $\pi - \pi$ stacking probability presented as a joint probability distribution of distance between the backbone beads on the x-axis and $\pi - \pi$ alignment on the y-axis, as a function of polymer side-chain patterning. On the right of the joint probability distributions, the respective configuration snapshots of just the backbone beads are shown to illustrate the backbone aggregation.

due to homogeneous chemically similar side-chains along these sections. We see huge stacks of polar and apolar regions, and the stacking between apolar blocks is even more beneficial to shield them water which is primarily observed in the polar regions.

Charge Transport Characterization

Now, we investigate how the improved stacking and selective percolation of water near polar side-chain alters the polymer mobility. The average mobility is higher for p1a1 case than the random case, but within error tolerance (Figure 5.11a). The higher mobility follows from the superior connectivity and polymer aggregation mediated by the separate polar and apolar blocks in the p1a1 polymer. We find higher delocalization length (Figure 5.11b) and accordingly, lowered polaron free energy differences and reorganization energy (Figure 5.19(a-b)) for p1a1. The improvement in polymer morphology is significant which has raised the number of charge transport pathways (Figure 5.11e). and improved electronic coupling between polaron sites (Figure 5.11f)., further reflected in the charge transport rates (Figure 5.19d shown before determining aggregates). After accounting for the time spent in the
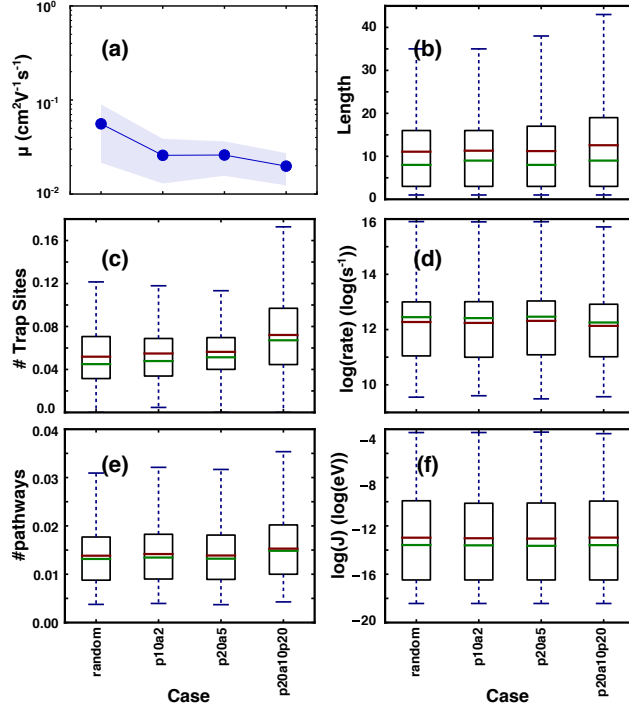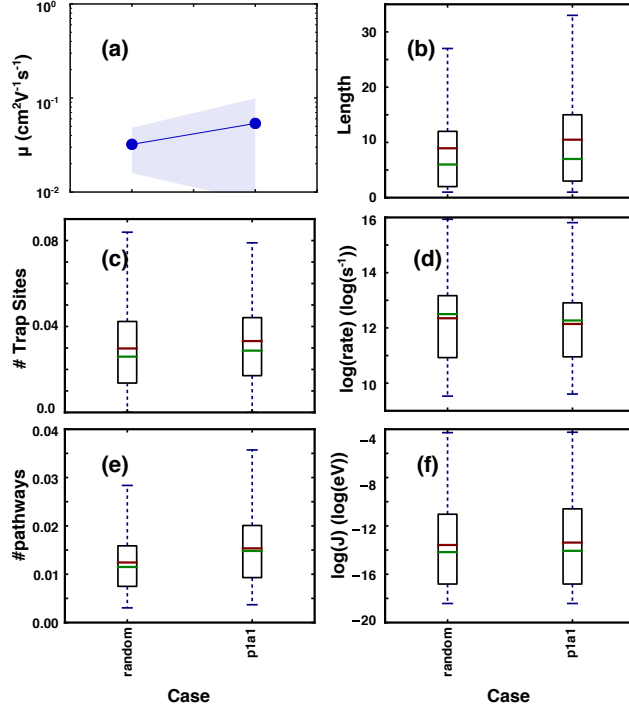
Figure 5.11. (a) charge mobility, (b) polaron delocalization length (number of backbone beads per polaron site), (c) number of trap sites (N_{traps}) normalized by total number of sites, (d) log distribution of the intersite charge transfer rate , (e) the number of pathways per site (N_{paths}) normalized by the total number of sites, and (f) log distribution of the intersite electronic coupling, across different side-chain patterning. The green and red line in the whisker plots indicate the median and mean, respectively.

aggregates, we see a lower charge transport rate in p1a1 than the random configuration. The improved π-π stacking in p1a1 polymers leads to aggregates where the charge spends significant amount of time before escaping by a less favorable pathway. This is also reflected in the higher number of trap sites (Figure 5.11c) for this case. A powerful effect of trap sites has already been discussed for all of the polymers. Well connected polymers such as p15a20p15 for the 60% case, p20a10p20 for the 80% case and 100% case (from Chapter 4), also showed more number of trap sites due to better packing, resulting in reduction in mobility. Similar to the 100% case, however, for the p1a1 case, we see a significant rise in polymer connectivity and stacking to outcompete the effect of trap sites to give an overall high mobility.

5.3 Conclusion

In this study, we have extensively characterized the influence of placing side-chains in block fashion compared to experimentally studied random copolymers on the OMIEC performance. We performed this analysis by simulating polymers with varying side-chain pattern for three compositions of polar vs. apolar side-chains– 60% polar (40% apolar), 80% polar (20% apolar), and 50% polar (50% apolar) side-chains. The hypothesis behind the study is having dedicated blocks of polar side-chains for attracting the solvent in the polymer film and apolar blocks for stacking backbones closely and efficiently. Also, we test whether the large blocks of chemically similar side-chains promote packing of larger polymer segments, due to their similar hydrophilicity.

We observe distinct structural distribution across the various block configurations, where the polar blocks stack mainly with polar blocks than apolar blocks and vice versa for apolar blocks. This is conducive to the polymer packing, as the regions of backbones which stack together do not have to reorganize to minimize interactions between polar and apolar side-chains. It leads to an increase in polymer delocalization with reduced torsional disorder, higher π-π stacking, and more connections between polymer clusters across the network. For the block configurations, which have polar blocks on both ends, we find extended polymer chains through the solvent leading to increased polymer-solvent interaction. However, the best block patterning scheme was separating the polar and apolar side-chains across the polymer than along the polymer. This configuration allows for even larger and well separated stack of polymers. For the cases with polar blocks at the end (p15a20p15 and p20a10p20) as well as p1, we find an increased fraction of polymer in contact with solvent. Although, the entire polymer is not homogeneously swollen, we see blocks of polar side-chain attached backbone interfacing with the solvent. These cases show an effective film volume comparable to the random configurations with higher polar side-chains. The overall doping, however, reduces for these well arranged configurations, because the solvent primarily percolates in the polar region while the backbone segments connected to apolar side-chains do not have enough dopants. We reconfirmed that an effective hydrophilicity exists for polymer side-chains encoded by the nature of neighboring side-chains. These large

block length configurations show an increased and decreased hydrophilicity for the polar and apolar side-chains, respectively while the opposite is true for random configuration. We also saw that the influence of neighboring chains is rendered ineffective when the number of neighbors drop, somewhere around 5-10 side-chains. Interestingly, even though for the p1a1 case, the polar and apolar side-chains are on consecutive backbone beads, they are well separated and do not influence each others' polarity unlike random configurations.

We tested how these morphological features translate into the conductive properties of the polymer network. While the improved stacking and connectivity for higher block length polymers lead to higher delocalization length, more number of electronic pathways and improved electronic coupling, it doesn't translate necessarily to net mobility. The KMC simulations performed here are representative of a low carrier density limit and the trap sites seem to have a huge contribution to the reduction of mobility. The most ordered systems showed most trap sites, due to the presence of well stacked clusters and energetically favorable sites which can act as huge time sinks for the hole and reduce the effective mobility. The competition of number of electronic pathways vs. trap sites determined the best performing case in these simulations. Hence, although well connected, the p15a20p15 case for 60% polar composition and p20a10p20 case for 80% polar composition performed poorly but the p1a1 case for 50% polar composition outperformed the random case. Considering the limitations and qualitative nature of the simulations, the mobility trends should be taken with a grain of salt. The study however does provide with important trends and mechanisms in the context of block patterning of OMIEC polymers. To summarize, we find an improved structural and electronic connectivity and swelling with rise in block lengths, especially when both polymer ends are polar and side-chains blocks are placed across the backbone. In the trap-filling regime, these polymers may even emerge as winning polymer candidates and we hope the study inspires further experimental investigation in that regard.

APPENDIX

5.A   Parameters for Molecular Dynamics Simulations

The LJ parameters used in the simulations are listed in Table 5.1 and the corresponding Martini bead type.

Table 5.1. Parameters used for non-bonded interactions in the simulations. Units of all parameters are consistent with LAMMPS 'real' units (kcal mol^{-1}, Å etc.) and potential styles. The bead-type for backbone, polar side-chain, apolar side-chain, and chloroform are C4, EO, SC3 and C4, respectively. Water is made of POL, WM, and WP beads, from which the beads WM and WP do not experience any Van der Waal interactions. The interaction of backbone bead with other backbone beads is modelled as Gay-Berne potential (and not C4-C4 interaction) and the parameter values are reported next to "B-B".

B-B	$\sigma_x = \sigma_y = 5.0,\ \epsilon_x = \epsilon_y = 0.25,\ \sigma_z = 3.0,\ \epsilon_z = 1.2,\ \sigma_c = 3.0$				
	C4	EO	SC3	POL	Qa
C4	$\epsilon = 0.84,$ $\sigma = 4.7$	$\epsilon = 0.70,$ $\sigma = 4.7$	$\epsilon = 0.84,$ $\sigma = 4.7$	$\epsilon = 0.61,$ $\sigma = 4.7$	$\epsilon = 0.74,$ $\sigma = 4.7$
EO		$\epsilon = 0.61,$ $\sigma = 4.3$	$\epsilon = 0.53,$ $\sigma = 4.3$	$\epsilon = 0.79,$ $\sigma = 4.7$	$\epsilon = 0.84,$ $\sigma = 4.7$
SC3			$\epsilon = 0.63,$ $\sigma = 4.3$	$\epsilon = 0.61,$ $\sigma = 4.7$	$\epsilon = 0.65,$ $\sigma = 4.3$
POL				$\epsilon = 0.96,$ $\sigma = 4.7$	$\epsilon = 1.19,$ $\sigma = 4.7$
Qa					$\epsilon = 0.84,$ $\sigma = 4.7$

The intramolecular parameters used for the simulations are listed in Table 5.2.

Table 5.2. Parameters used for bonded interactions in the simulations. Units of all parameters are consistent with LAMMPS 'real' units (kcal mol^{-1}, Å etc.) and potential styles. The bonds in water molecules between the WM-POL and WP-POL beads are constrained using the SHAKE algorithm [143].

Bonds		Angles	
B-B	$k = 25.0,$ $r_0 = 4.3$	B-B-B	$k = 1.0,$ $\theta_0 = 160.0$
B-SP(SA)	$k = 1.5,$ $r_0 = 4.7$	B-B-SP(SA)	$k = 3.0,$ $\theta_0 = 90.0$
SP-SP	$k = 8.35,$ $r_0 = 3.22$	B-SP(SA)-SP(SA)	$k = 3.0,$ $\theta_0 = 180.0$
SA-SA	$k = 1.5,$ $r_0 = 3.6$	SP-SP-SP	$k = 6.0,$ $\theta_0 = 135.0$
POL-WM	$r_0 = 1.4$	SA-SA-SA	$k = 3.0,$ $\theta_0 = 180.0$
POL-WP	$r_0 = 1.4$	WM-POL-WP	$k = 0.5019,$ $\theta_0 = 0.0$

Dihedrals		Other	
B-B-B-B	$k_1 = 0.5,$ $k_2 = 0.6,$ $k_3 = 0.0,$ $k_4 = 0.0$	B-B, d	$k_1 = 0.0,$ $k_2 = 2.0,$ $k_3 = 0.0,$ $k_4 = 0.0$
SP(SA)-B-B-SP(SA)	$k_1 = 3.0,$ $k_2 = 0.0,$ $k_3 = 0.0,$ $k_4 = 0.0$	B-B, a_1	$k_a = 7.5,$ $\theta_0 = 90.0$
		B-B, a_2	$k_a = 7.5,$ $\theta_0 = 90.0$
		B-SP(SA), a	$k_a = 3.0,$ $\theta_0 = 90.0$

5.B Parameters for Charge Transport Simulations

The following parameters were used for the Marcus expressions discussed in the main text: $J_{\text{inter}} = 0.01$ eV, $r_0 = 3.5$ Å, $\alpha = 0.8$ Å^{-1}, $J_{\text{intra}} = 0.31$ eV, $\epsilon_{\text{opt}} = 3$, and $\epsilon_s = 10$. The expressions for external reorganization, λ, and oxidation energy, E, as a function of oligomer

length, l, were obtained from fits to reported data for polythiophene.[182] The governing expression for the external reorganization energy as a function of l is

$$\lambda = \begin{cases} -0.0973\sqrt{l} + 0.4922, & \text{if } l \leq 15 \\ \frac{1.9489}{l} - 0.0092, & \text{otherwise} \end{cases}.$$

The governing expression for the oxidation energy as a function of l is

$$E = 4.1336 \, l^{-0.75} + 4.67.$$

5.C   Additional Figures Referenced in the Main Text

5.C.1   60 % Polar - 40 % Apolar

The snapshots of backbone bead and water are shown below (Figure 5.12). The polymers are not very swollen and the network is closed off to the solvent. The highest block length p15a20p15 case shows the maximum swelling due to the extended polymer blocks with polar chains.
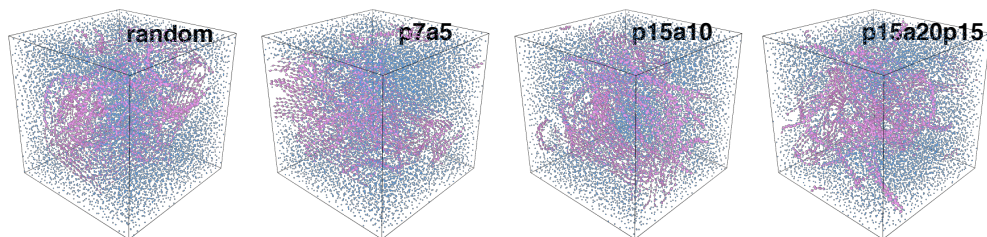
Figure 5.12. Snapshots illustrating water percolation in polymer networks for all polymers case. Only backbone and water are shown for clarity, represented by pink and blue backbone, respectively.

The radial distribution function (RDF) for the polar and apolar side-chains with ions is compared across all block copolymers below (Figure 5.13). We find that for larger block length, the effective polarity of polar side-chains increases and more solvent, and consequently ions are found near the side-chains as confirmed by the increased RDF. The opposite trend is observed for apolar side-chains, whose effective polarity drops for larger blocks lengths as

the apolar side-chains are away from polar side-chains and hence, shielded from electrolyte. The RDFs of corresponding backbone beads attached to either polar side-chain (BP) or apolar side-chain (BA) are also given, and they exhibit a similar trend. As discussed in the main text, the net RDF of all backbone beads with ions shows a drop with block length, as the increased electrolyte in polar blocks of backbone do not compensate for the missing electrolyte in apolar blocks.
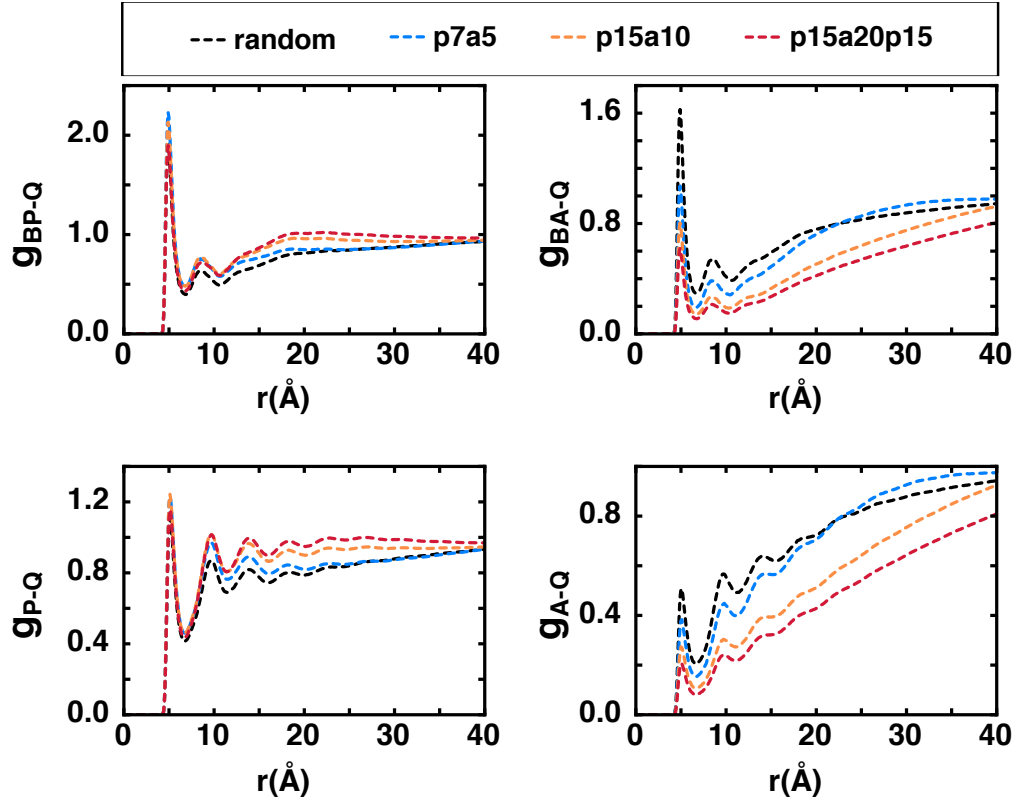
Figure 5.13. The radial distribution function (RDF) of backbone and side-chains with ions to quantify the electrolyte percolation in the polymer film, parsed separately for polar and apolar blocks. That is, the backbone attached to polar and apolar chains are treated differently and their individual RDFs are reported. The y-axis g_{x-y} denotes RDF between moiety x and y and, the backbone bead attached to polar and apolar side-chains is denoted by BP and BA, respectively, while polar side-chain, apolar side-chain and ion are denoted by P, A and Q, respectively.

For completeness, the distributions for the reorganization energy (λ) and site energy differences (ΔE) of charge transfer are reported in Figure 5.14. As pointed out in the main

text, the delocalization length increases with block length and it is reflected in the form of decrease in reorganization energy and site energy differences. Additionally, the charge transfer rate, before incorporating the time spent in aggregates is reported in Figure 5.14c. All polymer block configurations show a similar charge transfer rate.
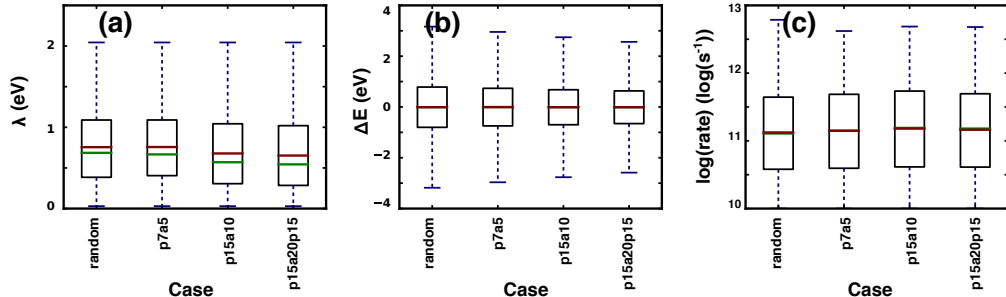
Figure 5.14. Distribution of (a) reorganization energy and (b) energy differences between polaron sites and (c) inter-site charge transfer rate (before determining aggregates) across different polymer block lengths. The green and red lines indicate the median and mean, respectively.

5.C.2 80 % Polar - 20 % Apolar

Similar to the 60% polar case, the RDF of polar and apolar side-chains with ions as well as RDF of the respective backbone beads attached to polar or apolar side-chains with ions are reported. For larger block lengths, both side-chain and by consequence, the backbone attached to it find more neighboring ions. On the contrary, the opposite is true for the apolar blocks, as witnessed by the reduced A-Q and BA-Q RDF with increasing block length.

The distribution of mobility and number of trap sites in the polymer network are reported only for the simulations showing maximum mobility for each block copolymer. We find the mobility trend is inversely correlated with the number of trap sites.

The percolation analysis performed for this series of polymer block configurations. We find that the case p20a10p20 with largest block size, shows most fragmented network due to highest trap sites (as discussed in the main text). It shows spatially extended networks at low rate threshold, in conjunction with highly delocalized sites but the spatial extent drops with increase in rate threshold as the network is not resistant to the loss of connections.

116
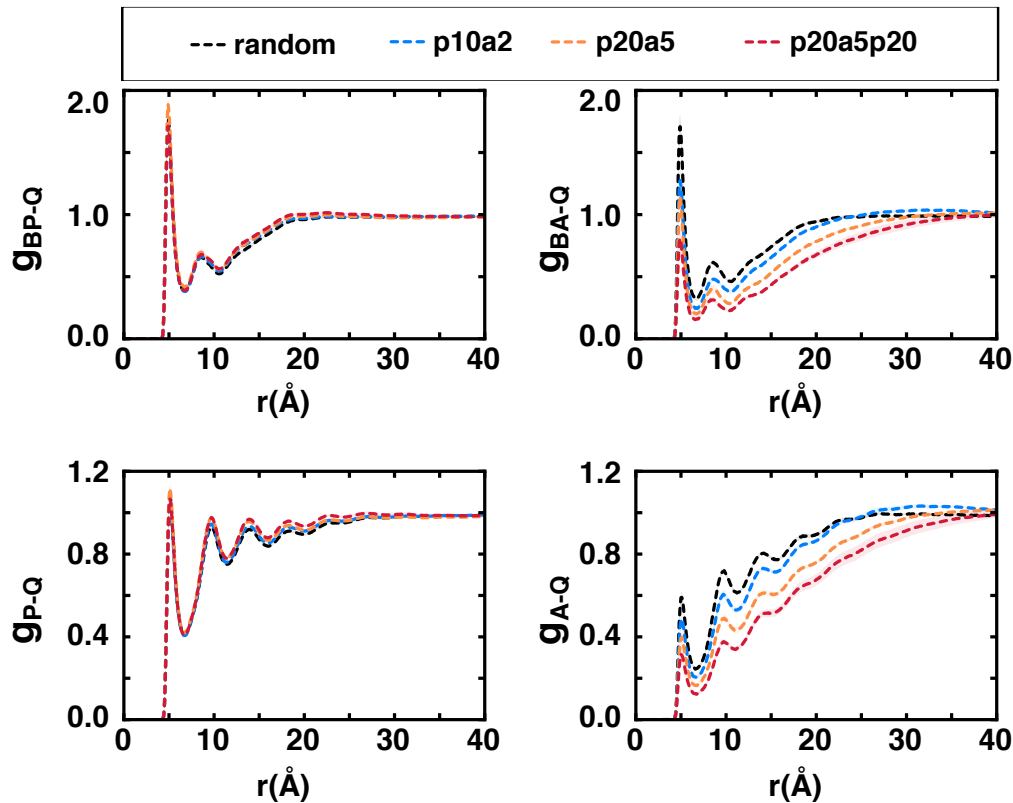
Figure 5.15. The radial distribution function (RDF) of backbone and side-chains with ions to quantify the electrolyte percolation in the polymer film, parsed separately for polar and apolar blocks. That is, the backbone attached to polar and apolar chains are treated differently and their individual RDFs are reported. The y-axis g_{x-y} denotes RDF between moiety x and y and, the backbone bead attached to polar and apolar side-chains is denoted by BP and BA, respectively, while polar side-chain, apolar side-chain and ion are denoted by P, A and Q, respectively.

5.C.3   50 % Polar - 50 % Apolar

The RDF of polar and apolar side-chains with are reported below. For p1a1, the RDF of polar side-chain and ion is higher while that of apolar side-chain is lower. The blocks of polar and apolar side-chains remain well separated in p1a1 configuration and electrolyte selectively percolates in the polar region.

The distributions for the reorganization energy (λ) and site energy differences (ΔE) of charge transfer are reported in Figure 5.19. As pointed out in the main text, the delocalization length is higher for p1a1 and it is reflected in the form of decrease in reorganization
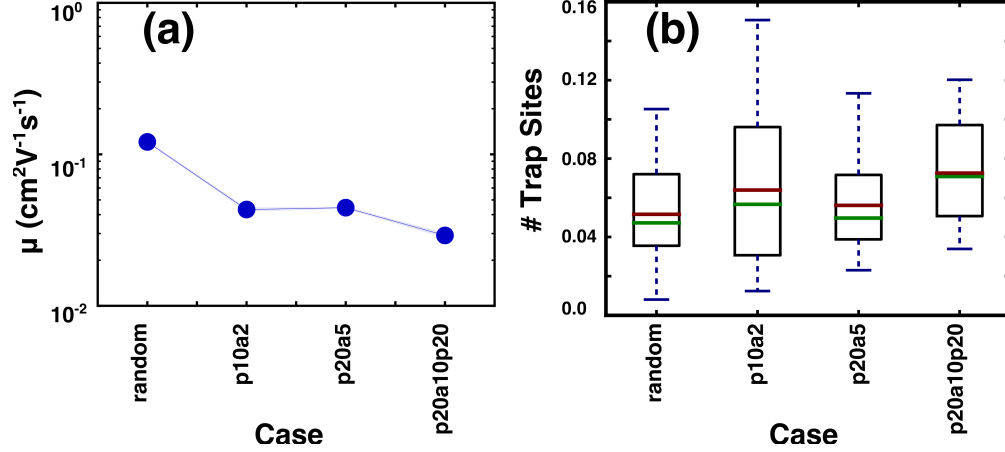
Figure 5.16. (a) Mobility and (b) distribution of trap sites are reported only for one simulation which shows maximum mobility (among the five independent simulations), for each block copolymer.

energy and site energy differences. Additionally, the charge transfer rate, before incorporating the time spent in aggregates is reported in Figure 5.19(c-d). Considering only the slower rates corresponding to inter-site hopping (i.e., excluding fast transfers between consecutive sites), we see faster pathways for p1a1 than random configuration which is in agreement with higher delocalization and connectivity.
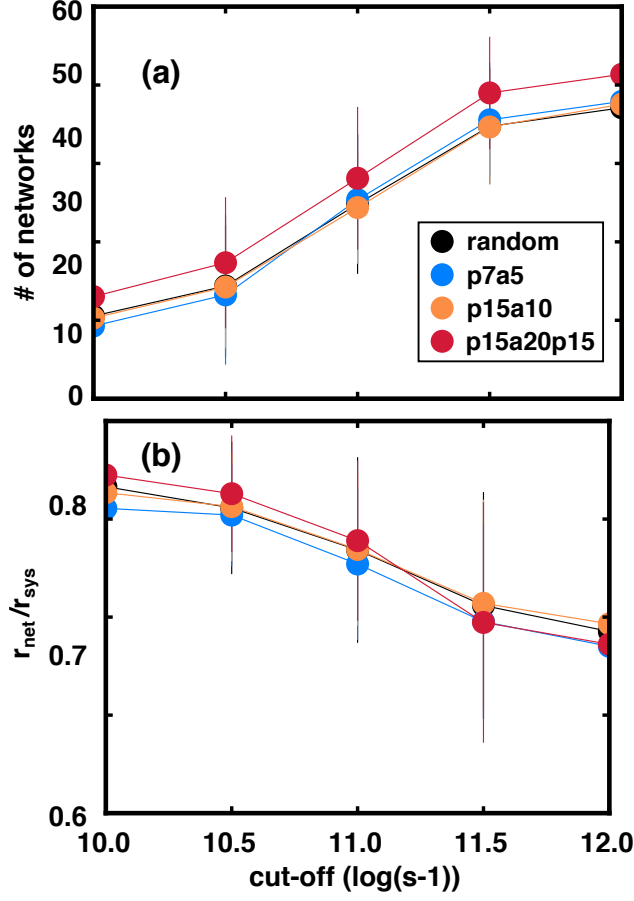
Figure 5.17. (a)The number of distinct subnetworks of polaron sites and (b) the radius of gyration of the largest subnetwork calculated as a function of the side-chain patterning. Subnetworks are defined such that each site is connected to at least one other site within the subnetwork by a rate above the threshold. The radius of gyration of the largest subnetwork is normalized by the radius of gyration the total system (r_{net}/r_{sys}).
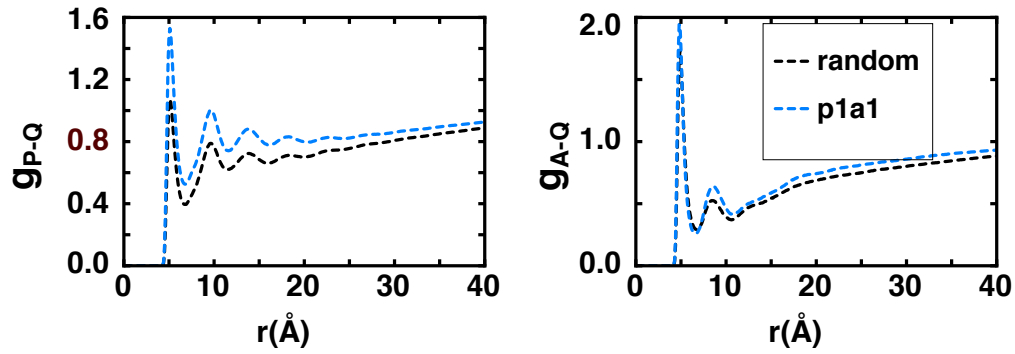
Figure 5.18. The radial distribution function (RDF) of side-chains with ions to quantify the electrolyte percolation in the polymer film. The y-axis g_{x-y} denotes RDF between moiety x and y and the polar side-chain, apolar side-chain and ion are denoted by P, A and Q, respectively.
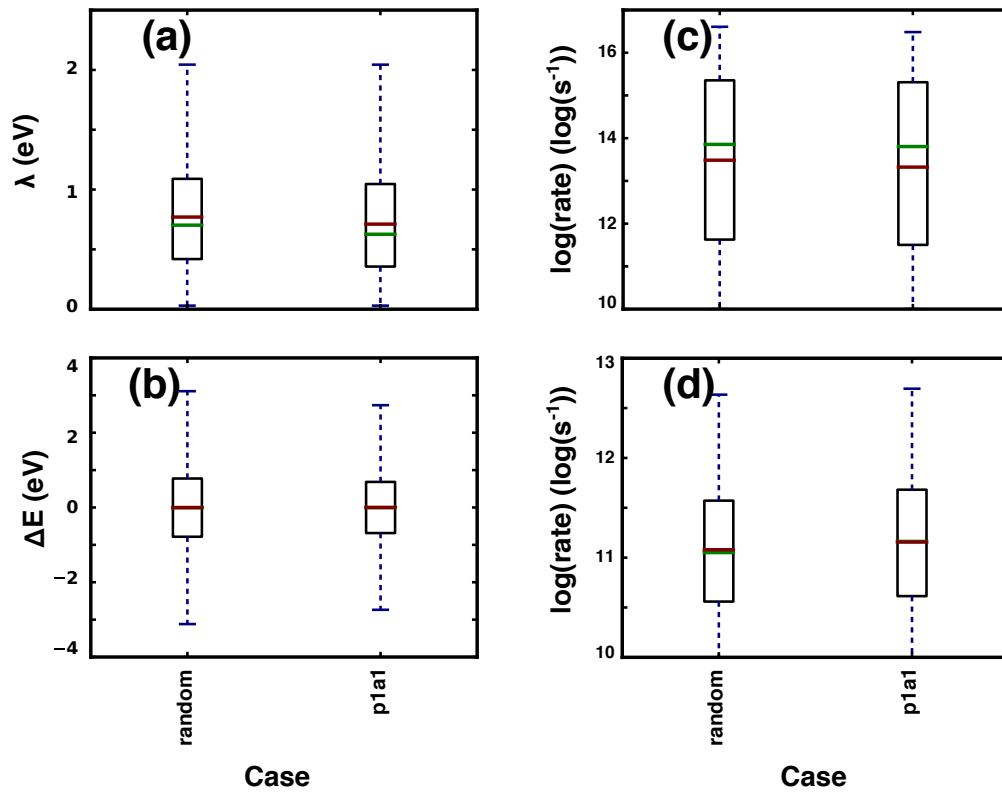
Figure 5.19. Distribution of (a) reorganization energy and (b) energy differences between polaron sites, inter-site charge transfer rate (before determining aggregates) (c) with and (d) without neighboring polaron jumps, across different polymer block lengths. The green and red lines indicate the median and mean, respectively.

6. EVIDENCE OF INFORMATION LIMITATIONS IN COMMON BOTTOM-UP COARSE-GRAINING METHODS

Reproduced from [82], with the permission of AIP Publishing

As discussed in Chapter 2, the CG model constitutes of three components which can limit the performance of the CG model– the mapping operator, functional form of potential energy surface (PES) and the target information used to obtain the PES. The former two components essentially determine the retained degrees of freedom (i.e., the "representability" problem) while the latter accounts for the information that is incorporated into standard parameterization algorithms (i.e., the "information" problem). In this work, we try to investigate the extent of these limitations in most common bottom-up CG methods and find them to be information limited. The attribution of these errors is important for future methods development as a representability-limited model can only be improved by increasing the CG degrees of freedom or by using a more complex form of interaction potentials, whereas an information-limited model can be improved by incorporating more information into the parameterization algorithm. Although representability limitations and information limitations are two distinct factors impacting the performance; they have been typically addressed under the umbrella of representability errors. Previous work has demonstrated that CG information-loss prohibits perfect thermodynamic state transferability, [76] that exact preservation of all macroscopic observables is impossible with reduced resolution [91, 159], and that thermodynamic expressions may become invalid in the CG representation [76, 169]. Under the dichotomy maintained here, most of these issues pertain to information limitations rather than representability limitations of CG models.

The practical question regarding these limitations is whether *effective* CG approximations are obtainable at different levels of resolution with existing algorithms [91]. For chemical systems, the answer is assumed to be positive, but relatively little work has been done to systematically quantify the sources and effects of these CG errors [33, 47, 48, 76, 172]. For instance, CG dimension reduction can (*i*) increase the required complexity of the CG potentials, (*ii*) reduce chemical transferability (i.e., accuracy of using the same CG potentials for different molecules and mixtures with similar underlying atoms) (*iii*) limit the accuracy and number of properties that can be simultaneously reproduced, and (*iv*) limit thermodynamic state transferability. Establishing the extent to which these tradeoffs occur due to representability-limitations versus information-limitations in practical scenarios is critical for guiding methods development and applying CG models.

In this chapter, we present the results of a systematic study of the accuracy tradeoffs associated with dimension reduction and mapping operator selection in bottom-up CG modeling. In general, these tradeoffs depend on the intrinsic dimensionality of the specific fine-grained system being modeled, as well as the choice of evaluated properties. For systems of practical complexity, such as liquids, mixtures, and soft materials, it is likely not possible to establish these trade-offs analytically; however, with modern simulation throughput we can empirically elucidate this tradeoff with systematic studies of mapping-related errors. To this

end, we have systematically generated all possible coarse-grained models that conserve a set of chemically-motivated locality and topology relationships for three liquids (pentane, 1-butanol, 1,3-propanediol). In turn, standard algorithms (iterative Boltzmann inversion with pressure correction and multiscale coarse-graining) were used to parameterize the models, and CG predictions for the 54 total models were compared with AA results for on-target properties (i.e., properties that are directly utilized in the parameterizations) and off-target properties.

Since on-target properties are directly supplied during parameterization, they exhibit no information limitations. Hence, the accuracy of on-target properties is diagnostic of representability limitations in the CG models. Likewise, a weak mapping dependence of the accuracy of on-target properties also indicates minimal representability limitations with respect to the mapping operator alone. In contrast, information limitations are evaluated based on the performance of the CG models in reproducing off-target properties. Since off-target properties are not included in the parameterization directly, their accuracy is limited by the amount of indirect information supplied by the on-target AA parameterization data. We observe poor reproduction of off-target properties and a weak mapping dependence of on-target properties which leads us to conclude that the current CG models are *information* limited. In turn, this raises the possibility that with better parameterization algorithms it will be possible to develop CG models of comparable complexity to those presented here that reproduce additional fine-grained properties.

6.1 Methods

In this work, we investigate the performance of 54 different CG models of three liquids– pentane, 1-butanol, 1,3-propanediol parameterized with the two standard bottom-up CG parameterization algorithms– Iterative Boltzmann Iversion (IBI) and Multiscale Coarse-graining (MSCG) in comparison to all atomistic models. In this section, the methods used to generate mapping operators and parameterize the CG potentials are discussed. The details of Molecular Dynamics simulations are listed and the procedure for different property evaluation is described.

6.1.1 Mapping Operator Generation

A resolution-based algorithm was developed for generating the mapping operators in this study.

Here, resolution (R) is an integer corresponding to the number of non-hydrogen atoms ("heavy atoms") to be combined into each CG bead. Mapping proceeds by (i) combining the indices of all hydrogen atoms with their bonded heavy atoms, (ii) identifying the molecular backbone based on the pair of heavy atoms with the largest graphical separation, then, starting at one end of the backbone, (iii) combining R consecutive heavy atoms (and their hydrogens) into each bead, until reaching the end of the backbone or running out of consecutive heavy atoms divisible by R, (iv) separately combining the remaining groups of connected
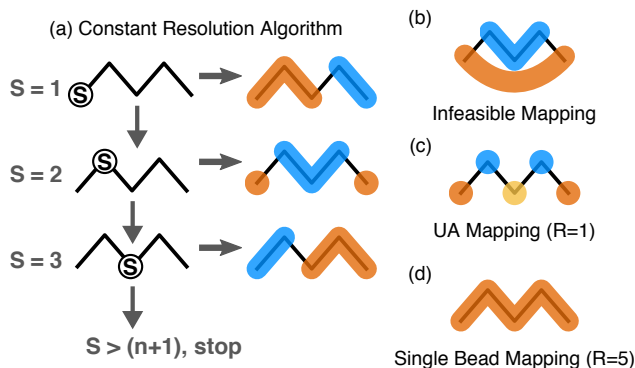
Figure 6.1. (a) Illustration of generating the R=3 mappings for pentane. CG beads comprised of R heavy-atoms are generated from consecutive atoms along the backbone, starting with the first backbone atom (S). New mappings are generated by incrementing S until it exceeds n+1, where n (2, for pentane) is the remainder upon dividing the chain length (5) by R (3). (b) An example of an excluded mapping since the atoms in orange bead do not form a connected subgraph. (c) A simple UA mapping obtained from R=1. (d) A single bead mapping from R=5.

heavy atoms (i.e., groups smaller than R) to complete the mapping, (v) incrementing the starting atom, and repeating (iii-v) as shown in Figure 6.1a.

Mappings are generated in this way until n (remainder on dividing backbone length by resolution R) increments have been performed, in which case no new mappings will be obtained. The final mapping operators are obtained by converting these index-based mappings into matrices for calculating the center-of-mass of each group of atoms.

The mapping algorithm is depicted schematically in Figure 6.1 for the case of pentane and $R = 3$. The algorithm has the feature that it will not combine heavy atoms into a bead if they do not form a connected subgraph (Figure 6.1b); this is consistent with local realism and also reduces the number of possible mappings. For cases where the number of heavy atoms in a molecule is not divisible by R, the algorithm will return mappings with mixed resolution. As described, the algorithm will always place these higher resolution beads at the edge of the molecular graph. For large macromolecules this is physically desirable, since it ensures that monomer units are consistently mapped. In this framework, common "united-atom" CG models are obtained with $R = 1$ (Figure 6.1c) and single-bead CG models correspond to $R = 5$ for all of the studied liquids (Figure 6.1d).

In the current study, we parameterized models for pentane, 1-butanol, and 1,3-propanediol based on all mappings for $R = 1 - 5$. These molecules were chosen because they have the same chain length but varying degree of anisotropic behavior depending on the presence of alcohol groups. For each molecule, there are a total of 9 mappings for the possible resolutions from 1 to 5. In a few cases, these mappings are symmetry equivalent, but are reported as independent models as an additional consistency check on our parameterization procedures. The structure of the mappings is similar across all molecules (Figure 6.2), the only difference is that the end group beads may have a methyl or hydroxyl group depending on the molecule.

123

CG beads that consist of the same number, topology, and types of atoms are assigned to be of the same CG bead type and are described by the same parameters in the models. In all of the presented figures, the unique bead types for each mapping are indicated by color.
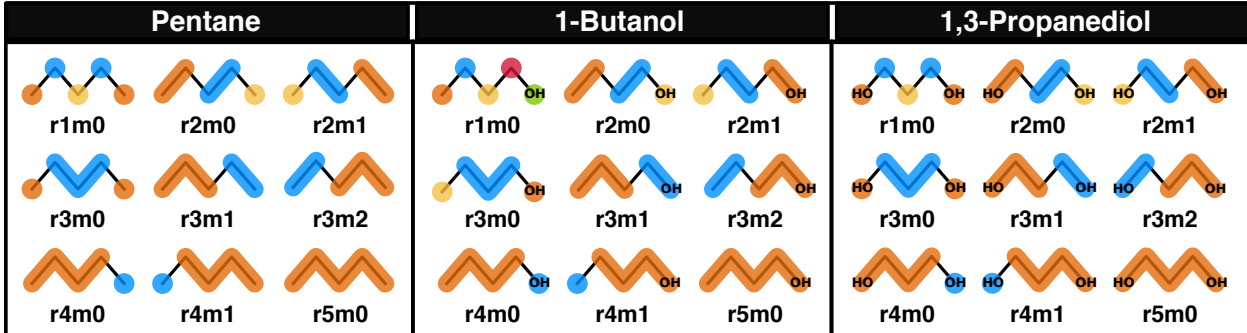
Figure 6.2. Mapping operators from resolution 1 to 5 obtained using constant resolution mapping scheme pentane, 1-butanol, and 1,3-propanediol from top to bottom

Although the molecules in this study are simple enough to enumerate these mappings by hand, the outlined algorithm is part of a more general procedure that will be described elsewhere. We note that additional modifications are required for molecular systems that contain rings and side-chains.

6.1.2 Iterative Boltzmann Inversion

Iterative Boltzmann Inversion (IBI) is an iterative method to generate CG potentials that reproduce the radial distribution functions ($g(r)$) of high-resolution, typically atomistic, models [134]. The target $g(r)$ is sampled by coarse-graining a previously generated atomistic trajectory at a fixed thermodynamic state. The initial guess for each pair potential is the potential of mean force obtained by Boltzmann Inversion,

$$U_{\alpha\beta}(r) = -k_B T \ln g_{\alpha\beta}(r) \tag{6.1}$$

$$g_{\alpha\beta}(r) = \frac{V}{4\pi r^2 N_\alpha N_\beta} \sum_{i \in \alpha} \sum_{j \in \beta} \delta\left(r - r_{ij}\right) \tag{6.2}$$

where, $U_{\alpha\beta}(r)$ is the pair potential between CG particles of type α and type β, T is the simulation temperature, $g_{\alpha\beta}(r)$ is the pair-specific radial distribution function, $V/N_\alpha N_\beta$ corresponds to the density of pairs in an ideal gas, δ is the Dirac delta, r_{ij} is the distance between particles i and j, and dividing the summation by $4\pi r^2$ normalizes the density of pairs at each separation r. Pairs separated by less than four bonds are neglected in both the summation and denominator calculations. This guess potential is simultaneously corrected for both the

structural property (i.e., $g(r)$) and simulation pressure using an iterative procedure according to

$$U(r)^{\mathrm{k}} = U(r)^{\mathrm{k}-1} + \frac{1}{\mathrm{n}} \sum_{\mathrm{i}=1}^{\mathrm{n}} \ln \frac{g(r)^{\mathrm{k}-1}}{g(r)^{\mathrm{AA}}} + \frac{\omega}{\mathrm{n}} \sum_{\mathrm{i}=1}^{\mathrm{n}} A \left(1 - \frac{r}{r_{cut}} \right) \qquad (6.3)$$

where $U(r)^{\mathrm{k}}$ is the potential at the kth iteration, the α and β subscripts are omitted for clarity, and the r-dependence is discretized using tabluated potentials. The first correction approaches zero as the radial distribution function obtained from the CG simulations $(g(r)^{\mathrm{k}-1})$ approaches the target distribution $(g(r)^{\mathrm{AA}})$. The second correction goes to zero at the potential cutoff (r_{cut}) and is updated proportional to a weight parameter (ω) and A, which is calculated according to

$$A = -\frac{3V r_{cut} (P^{\mathrm{k}} - P^{\mathrm{AA}})}{2\pi N \rho \int_0^{r_{cut}} r^3 g(r)^{\mathrm{k}} dr}. \qquad (6.4)$$

In Eq. 6.4, P^{k} and P^{AA} correspond to the pressure in the CG simulation and AA simulation respectively, N is the number of particles in the AA simulation, and ρ is the mass density of the CG simulation. A approaches zero as the pressure from the CG simulation approaches the target atomistic pressure [172]. Both correction terms are averaged over n independent CG simulations performed at each iteration, denoted by the summation over index i in the Eq. 7.11.

IBI is similarly implemented for bond, angle, and dihedral potentials. The corresponding initial guess and update equations are identical to Eqs. 7.10 and Eq. 7.11, with the pair radial distribution function $(g_{\alpha\beta}(r))$ being replaced by the distribution function $(P_\gamma(\kappa))$ of the corresponding interaction type, γ, and omission of the pressure correction term. The distribution functions for bonds, angles and dihedrals are given by $P_b(r) = \frac{H_b(r)}{4\pi r^2}$, $P_a(\theta) = \frac{H_a(\theta)}{\sin\theta}$ and $P_d(\phi) = H_d(\phi)$ respectively, where $H_\gamma(\kappa)$ is the histogram of the quantity κ governing the interaction type γ. Additional IBI implementation details, including treatment of undersampled regions, cutoffs, bin sizes, and convergence behavior are included in the Appendix 7.A.

6.1.3 Multiscale Coarse-Graining

The Multiscale Coarse-graining (MSCG) framework was developed in [120], where the error between the bead forces in the atomistic representation (i.e., calculated by mapping the atomistic forces) and the coarse-grained representation (i.e., using the CG potentials on the mapped atomistic configurations) was proposed as a variational functional. It was shown that minimizing this functional resulted in equivalent phase distributions of the coarse-grained system when sampled from the atomistic or coarse-grained representations. The system of equations can be further linearized using spline potentials yielding

$$\sum_{\gamma_{\mathrm{i}}} f_{\gamma_{\mathrm{i}}}(\kappa) \boldsymbol{\nabla}_{\mathrm{i}} \kappa(\{\mathbf{r}_{\mathrm{k,t}}\}) = \mathbf{F}_{\mathrm{i,t}}^{\mathrm{AA}}, \qquad (6.5)$$

where $f_{\gamma_i}(\kappa)\boldsymbol{\nabla}_i\kappa(\{\mathbf{r}_{k,t}\})$ is the force in CG representation on the ith site due to γ_i type of interaction (e.g., pair potential, bond potential, etc.) and κ is the value of the variable associated with that interaction (eg. pair radial distance, bond radius, etc. which is a function of the position of site i) in snapshot t [121, 141]. Summation over γ_i gives the net force on the site i, which is parameterized to equal the force obtained for that site in the atomistic representation ($\mathbf{F}_{i,t}^{AA}$). The function $f_{\gamma_i}(\kappa)$ is the negative of the derivative of the potential $U_{\gamma_i}(\kappa)$ with respect to the variable κ and $\boldsymbol{\nabla}_i\kappa(\{\mathbf{r}_{k,t}\})$ is the gradient of the variable κ with respect to the position vector of site i. The potentials $U_{\gamma_i}(\kappa)$ have the form of spline polynomials and their coefficients are obtained by solving this linear system of equations. To alleviate memory constraints, the atomistic trajectory is divided into blocks, the linear system of equations Eq. 6.5 is solved for each of the blocks, and the final forces are given by an average over forces obtained in each block.

In this study, MSCG was only used to parameterize non-bonded potentials, whereas the bonded potentials were obtained in all cases from IBI. So, when we discuss errors in certain properties for IBI vs MSCG generated potentials, it is should be interpreted as IBI nonbonded + IBI bonded vs. MSCG nonbonded + IBI bonded. This hybrid method of using MSCG only for non-bonded potentials and using distributions for obtaining bonded potentials has been shown to perform better than using MSCG for all potentials [140]. All MSCG models were parameterized using the VOTCA package [141]. Additional details of the specific VOTCA settings used for the MSCG procedure are discussed in the Appendix 7.A.

6.1.4 Property Calculation

Five properties—forces, $g(r)$, self-diffusion coefficient (D), velocity autocorrelation function (VAC), and enthalpy of vaporization (ΔH_v)—were calculated for each CG model and compared with the corresponding atomistic results. Here, force and $g(r)$ are used as on-target properties (i.e., properties that are explicitly included in the model parameterizations), while D, VAC, and ΔH_v are off-target properties (i.e., properties that are not included in the model parameterizations). The mathematical expressions and numerical details involved in the computation of these properties and corresponding errors are presented in the Appendix 7.A. In brief, for forces the norm squared error is computed for all beads and averaged over all beads. For $g(r)$ and VAC, the absolute error at each radial distance and time-difference, respectively, is computed and averaged. The $g(r)$ error is additionally weighted by the exponential of the atomistic radial distribution function at each radial distance. For D and ΔH_v, the signed error is reported. In all cases, the errors are normalized with respect to atomistic representation, to obtain a normalized metric that is comparable across different mappings and molecules.

6.1.5 Molecular Dynamics Simulations

LAMMPS was used to perform all molecular dynamics simulations [130]. All atomistic reference simulations used a one fs integration time step, Velocity-Verlet integration, and

periodic boundary conditions. Long-range electrostatics were modelled using the particle-particle-particle-mesh (PPPM) algorithm [21] and Lennard-Jones interactions were truncated at 14 Å. All simulations were initialized from diffuse configurations containing at least 2000 atoms, using a cubic grid to place molecules in random orientations without overlap. The velocities were initialized from a uniform distribution obtained with a random seed value and scaled to give the correct kinetic energy. The simulations were first relaxed in the NVE ensemble with restrained atomic displacements of 0.1 Å per time step for 10 ps, followed by a 1 ns NPT equilibration at 298K and 1 atm. The simulations were extended at 298K and 1 atm for 2 ns in the NPT ensemble to obtain the average density. In the NPT simulations, the Nosé-Hoover thermostat and barostat were employed using the modified form proposed by Martyna, Tobias, and Klein, as implemented in LAMMPS [102], with a relaxation time constant of 0.1 ps and 1 ps for the thermostat and barostat, respectively. From these initial simulations, five independent NVT simulations were generated with velocities reinitialized as described above. These simulations were first relaxed in the NVE ensemble with restrained atomic displacements of 0.1 Å per time step for 10 ps, then the simulation box was rescaled over 10 ps to obtain the correct average density, followed by a 1 ns NVT equilibration at 298K. The production data was obtained by extending these simulations for 10 ns. The thermodynamic data and coordinates were sampled at 0.1 ps and 1 ps for simulations in NPT and NVT ensemble respectively.

The coarse-grained simulations used to evaluate the structural and dynamic properties of each model were performed in the NVT ensemble. The initial configuration for these simulations was obtained by starting from a diffuse configuration of 1000 particles and equilibrating to the known atomistic density by the same NVT equilibration procedure as in the atomistic simulations. The production data for each property was obtained by then extending the simulations to appropriate lengths to obtain statistical convergence (See Appendix 7.A for property specific trajectory lengths). Where independent trajectories were simulated, the velocities were reinitialized for each simulation, and an equilibration run of 1 ns was performed before starting the production run. All CG simulations used tabulated potentials and excluded long-range electrostatic interactions. Experimentation with models that included a long-range electrostatic component seemed to have no effect (not shown).

To provide consistent parameterizations of each liquid, all atomistic force-fields in this study were derived from quantum chemistry calculations using the topology automated force-field interactions methodology (TAFFI) [148]. In brief, all force field terms were parameterized on the basis of density functional theory (DFT) quantum chemistry calculations, using the B3LYP-D3/def2-TZVP level of theory computed via the Orca software package [116]. The force fields were parameterized using the OPLS force-field functional form [77], except that 1-4 pairwise interactions were excluded in the non-bonded interaction computation. Bond, angle, and dihedral force-field terms were derived from potential energy curves computed for internal degrees of freedom for each molecule in vacuum, optimizing the other degrees of freedom as a function of the mode scan. The resulting energy curves were self-consistently fit to obtain the corresponding force-constant parameters and equilibrium displacement parameters in the force field. Partial charges were obtained from CHELPG calculations on molecular configurations sampled from the liquid, and Lennard-Jones parameters

were parameterized on the basis of counterpoise-corrected interaction energies calculated for molecular pairs sampled from the liquid state.

The atomistic molecular dynamics simulations for pentane, 1-butanol and 1,3-propanediol contained 143, 134 and 154 molecules (equivalent to at least 2000 atoms) and an average density of 0.65 g/cc, 0.81 g/cc and 1.04 g/cc was obtained, respectively. Additional details pertaining to the sampling and details of the property calculations are included in the Appendix 7.A.

6.2   Results and Discussion

The CG simulation results are presented separately for on-target (i.e., properties that are explicitly included in the model parameterizations) and off-target (i.e., properties that are not included in the model parameterizations) properties. Since on-target properties are directly supplied during parameterization, they exhibit no information limitations. Hence, the accuracy of on-target properties is diagnostic of representability limitations in the CG models and their performance is indicative of the representability limitations incurred by the choice of mapping operator and the isotropic potential interaction. In contrast, information limitations are evaluated based on the performance of the CG models in reproducing off-target properties. Since off-target properties are not included in the parameterization directly, their accuracy is limited by the amount of indirect information supplied by the on-target AA parameterization data and their performance has separate significance with respect to how the approximations represented by each parameterization algorithm are affected by the CG dimension reduction. We observe qualitatively different behavior for these two classes of properties, with implications for the representability and information limitations of CG models.

6.2.1   On-Target Properties
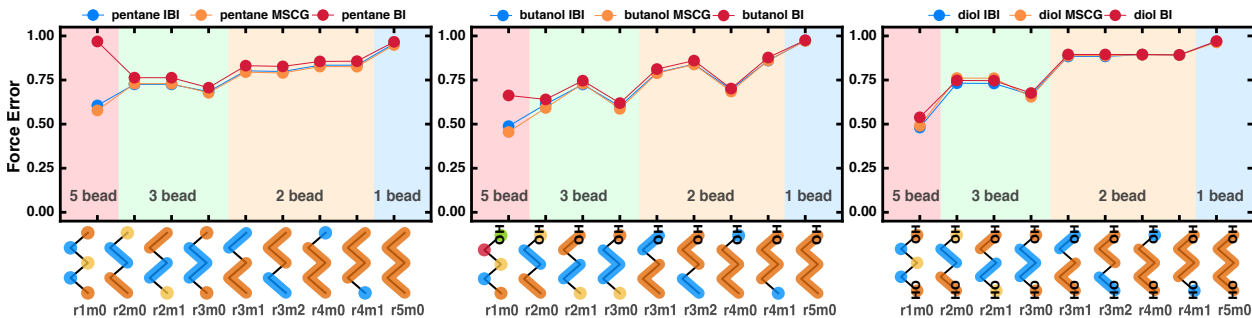
Figure 6.3. Normalized average error in squared norm of forces per bead using IBI, MSCG and BI methods across all mappings for pentane, 1-butanol and 1,3-propanediol, from left to right. All three algorithms exhibit similar errors for most mappings. There is a consistent effect of mapping on the error, but it is small compared with the overall high errors.

Figure 6.3 presents the results for the average errors in bead forces for each CG model and the reference atomistic data. These values have been normalized to the average force in the atomistic representation to provide a measure of the relative error (See Appendix 7.A for additional details). The bead forces are directly parameterized in the MSCG algorithm via a least-squares minimization, whereas the forces are only indirectly included in the IBI algorithm through the potential of mean force.

First, we note that IBI and MSCG exhibit very similar error residuals, despite the fact that IBI does not utilize the force information directly in the parameterization. Although the potentials derived from IBI and MSCG show significant differences in all cases (Figure 6.9), comparable force residuals are obtained. Considered together, these observations imply that the force error optimization surface is non-convex and relatively shallow with a broad manifold of CG potentials that are capable of approximately minimizing the residual. As a further illustration of this, even the Boltzmann inverted (BI) potentials, which are distinct from both IBI and MSCG potentials (Figure 6.9), show comparable mean force errors to MSCG and IBI for lower resolutions.
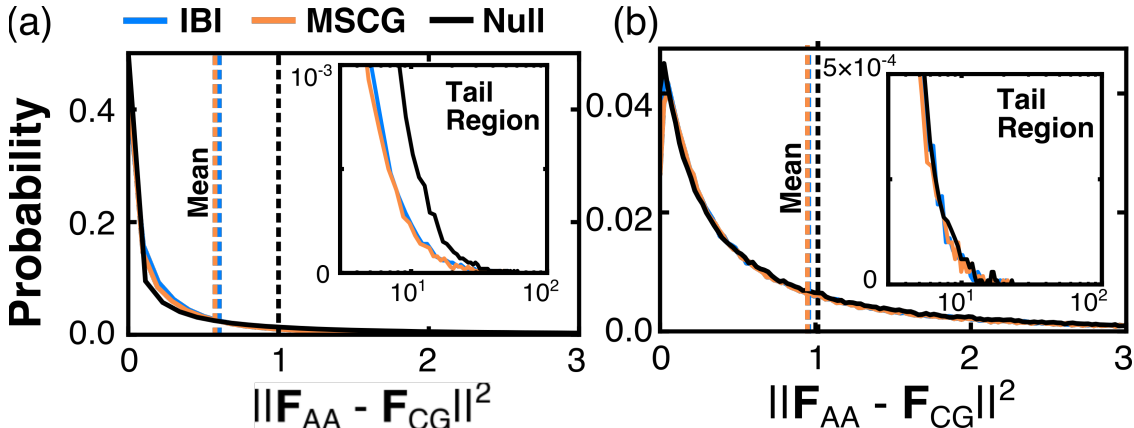
Figure 6.4. Distribution of normalized errors in squared norm of bead forces for IBI, MSCG, and null (zero CG potentials) for (a) pentane r1m0 and (b) pentane r5m0. The vertical dotted lines represent the mean (as reported in Figure 6.3) for each case. The IBI and MSCG distributions are very similar to the null model for both of the mappings but with a small improvement in mean error. This results from lower errors in the extreme values for MSCG and IBI as shown in the insets. The insets are drawn with log-linear axes to visualize the tail.

Second, the force errors are remarkably large for all resolutions, models, and algorithms. We note that the error residuals for MSCG are rarely reported [84, 121]. In the current case, the residual errors across all models vary between 1- 38 kcal^{2}mol^{-2}Å^{-2}, which are smaller or equivalent χ^2 values than those reported in Ref. [121] for methanol and a model ionic liquid ($\sim$ 21 and 163 kcal^{2}mol^{-2}Å^{-2}, respectively). For additional clarification, we also present the distribution of errors in the force magnitude in Figure 6.4 for two illustrative cases. Comparison with the null (i.e., zeroed out CG potentials) confirms that IBI and MSCG

are capturing some atomistic force information, but this primarily manifests in minimizing the extreme values of the training forces, as seen in the inset of Figure 6.4. In particular, the null model yields comparable median force errors to either of the CG models, with a slightly higher mean force error. The relatively small improvement over the null model with respect to mean force error for IBI and MSCG provides further evidence that the force minimization exhibits a shallow minimum. One interpretation of these high overall errors is that the chosen functional form for the CG models, specifically the use of isotropic pair-wise interactions, leads to representability limitations in all cases. Perhaps most surprising is the large errors for R=1, which is the mapping that corresponds to typical united-atom force fields that have been broadly adopted in experimentally parameterized force fields. These errors suggest that even at the united atom level, reproduction of the atomistic forces is representation limited. As discussed below, errors in the molecular forces do not necessarily imply that the CG models do not reproduce other significant properties of the atomistic models; however, it does bear on the question of whether molecular forces supply sufficient information in practice to parameterize CG models.

Third, there is a consistent increase in the force error as the model resolution decreases. However, this trend is negligible in comparison with the high overall errors. This suggests that all of the models are representability-limited due to the insufficiency of isotropic CG potentials in reproducing the atomistic forces, while the representability limitations associated with the mapping operator are small. Likewise, there is no effect on the force error upon the addition of anisotropic moieties (i.e., comparing pentane to 1-butanol to 1,3-propanediol). For a more expressive basis of force field functions, the mapping effects in each of these cases could be more significant.

The $g(r)$ errors, averaged across all pair types and radial distances, are shown in Figure 6.5 for IBI and MSCG. For these data, a probability weighted average has been utilized for averaging over the radial component to avoid poorly sampled regions from dominating the average. To better visualize the $g(r)$ mismatch at each radial distance, the $g(r)$ distributions for IBI and MSCG models are compared with target AA distributions for their respective minimum and maximum error cases in Figure 6.10. $g(r)$ is used as the target property to fit potentials in IBI but are not directly utilized in MSCG.

First, we note the qualitative difference between the IBI and MSCG results. IBI exhibits near perfect reproduction of the $g(r)$ for all molecules and mappings, whereas MSCG exhibits worse overall reproduction in most cases. Moreover, we observe that the MSCG models show decreasing performance upon addition of anisotropic moieties (i.e., the errors for 1,3-propanediol and 1-butanol are greater than that of pentane). This comparison suggests a potential asymmetry in information content supplied by the atomistic $g(r)$ and forces. We have observed that the IBI algorithm, which utilizes only $g(r)$ information, also results in force errors that are comparable to MSCG. However, MSCG, which utilizes only force-information, has much higher $g(r)$ errors than IBI. This asymmetry suggests that the atomistic $g(r)$ is either more information rich or numerically expedient as a CG parameterization target.

Second, we stress the significance that IBI is capable of accurately reproducing the $g(r)$ for all mappings and molecules. At face value, this is simply evidence that IBI is doing what it purports to do. However, the absence of a mapping dependence also indicates that these
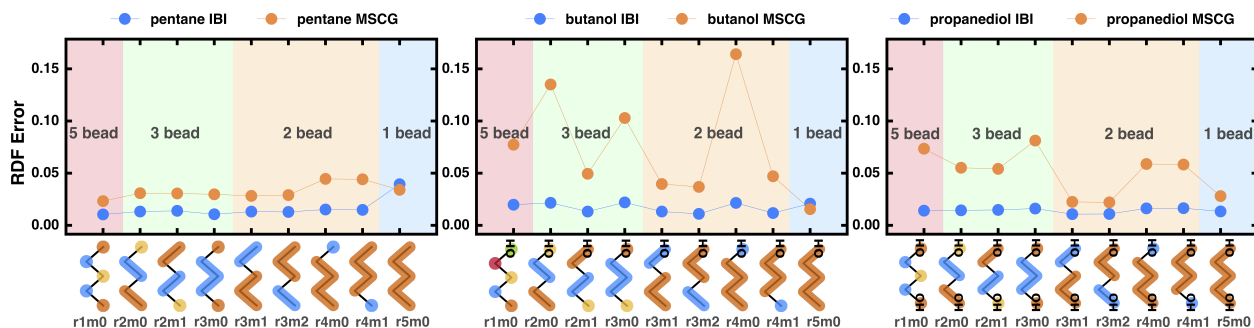
Figure 6.5. Error in $g(r)$ averaged across all pair types and radial distances obtained using IBI and MSCG methods across all mappings for pentane, 1-butanol, and 1,3-propanediol, from left to right. Errors for IBI are very low across all mappings and molecules. MSCG shows higher errors than IBI, especially for 1-butanol and 1,3-propanediol.

isotropic CG models are not representability limited with respect to reproducing the $g(r)$, whereas they were representability limited with respect to reproducing the atomic forces. Thus, representability limitations can strongly depend on the target properties of the CG model.
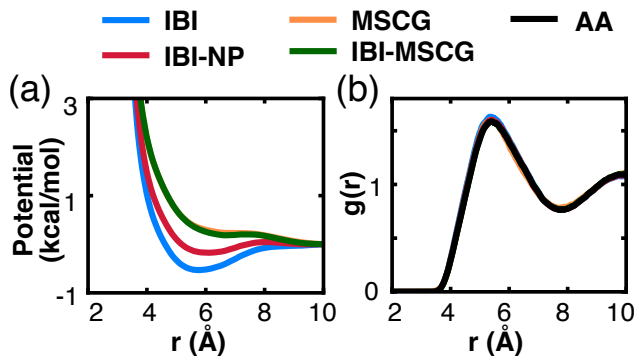
Figure 6.6. Comparison of (a) potentials and (b) corresponding RDFs for IBI, MSCG, IBI without pressure correction (IBI-NP) and IBI with MSCG potential (IBI-MSCG) as initial guess for the 1-butanol r5m0. Even though the potentials are qualitatively different, the simulations result in indistinguishable RDFs.

Third, we also note that in the cases where IBI and MSCG similarly reproduce the $g(r)$ (e.g., comparing the $g(r)$ results for the 1-butanol r5m0 mapping), they also exhibit significant differences in the associated potentials (Figure 6.6). Obtaining similar $g(r)$ from qualitatively different potentials suggests that the $g(r)$ optimization surface, like the force optimization surface, is non-convex and relatively shallow. We additionally report the potentials for IBI without pressure correction (IBI-NP) and IBI initialized from MSCG potential as the initial guess (IBI-MSCG) in Figure 6.6. These additional models also exhibit differ-

ent converged potentials and yet reproduce $g(r)$. Several other groups have made a similar observation, in that although the Henderson theorem [62] formally guarantees the uniqueness of the potential that produces an $g(r)$ up to a constant, in practice the potential is underdetermined by the $g(r)$ alone [47, 49, 172]. The implication of a non-convex $g(r)$ optimization surface is that isotropic pair-wise potentials are potentially capable of supporting more physical information.

6.2.2 Off-Target Properties

The self-diffusion coefficient and enthalpy of vaporization were calculated for each CG model and compared with the atomistic simulations. These serve as off-target properties for both MSCG and IBI. By reproductng key target properties (i.e., forces and $g(r)$), MSCG and IBI potentially transfer physical information for additional system properties to the CG models. Our goal here is to quantify the extent of this information transfer as the model resolution decreases and to highlight the implications for the information limitations of the models.
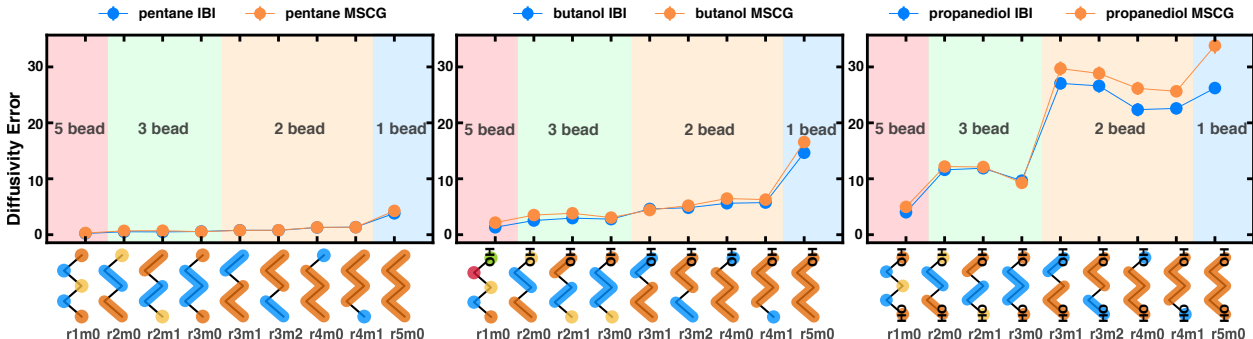
Figure 6.7. Normalized error in self-diffusion coefficient using IBI and MSCG methods across all mappings for pentane, 1-butanol, and 1,3-propanediol, from left to right. The standard errors are within the size of the markers in all cases. Both IBI and MSCG have similar errors, and the errors increase with number of hydroxyl groups. Overall the error is dominated by the number of beads and increases with a decrease in the number of beads.

The relative error in self-diffusion coefficients, averaged over all trajectories and normalized with respect to the atomistic coefficients, are shown for each mapping and molecule in Figure 6.7. It is generally known that CG models parameterized with MSCG and IBI substantially overestimate diffusivity, and several strategies have been proposed to correct this [33, 71]. Here, we likewise observe a dramatic overestimation of the diffusivity for both MSCG and IBI. The two algorithms perform similarly in all cases, with a small but consistent trend of higher accuracy for IBI. Izvekov and Voth presented a general argument for this diffusivity overestimation, whereby the missing degrees of freedom in the CG model manifest as a stochastic friction in the atomistic model that is missing from CG dynamics [71]. Following this argument, it is unsurprising that the diffusivity errors also show a trend

of increasing error with decreasing resolution. However, several quantitative aspects of this comparison are relevant to the accuracy of each CG approximation with respect to dimension reduction.

First, we note that the diffusivity errors show a strong increase with the number of anisotropic moieties in each molecule (i.e., the errors are ranked as 1,3-propanediol > 1-butanol > pentane). However, this trend is the reverse of the number of degrees of freedom that are lost during the CG procedure for each liquid (i.e., for the same mapping, the lost degrees of freedom are ranked as pentane > 1-butanol > 1,3-propanediol). Thus, the diffusivity overprediction is a non-trivial function of the chemical nature of the coarse-grained degrees of freedom, with those associated with more strongly interacting groups having a disparate impact on the CG dynamics. These results also suggest that there is no generally applicable rescaling of the CG dynamics based on resolution, since the associated speedup is a function of the specific degrees of freedom being coarse-grained.

Second, the diffusivity errors correlate more strongly with the number of CG beads in each model than with the resolution. For a fixed number of beads, a mapping effect is observable in some cases (e.g., for 1,3-propanediol, r3m0, which retains a high-resolution bead on both hydroxyl groups, performs better than r2m0 and r2m1, which do not retain a high-resolution hydroxyl bead), however it is minute in comparison with the effect of changing the number of beads. This weak mapping dependence provides justification for the *ad hoc* mappings that are typical in the CG literature and indicates that the number of beads affects the accuracy of the CG approximation much more strongly than the specific mapping operator.

The relative error in the VAC, averaged over bead types and times less than 500 fs, and normalized with respect to the atomistic values, are shown for each mapping and molecule in Figure 6.11. Whereas the diffusivity corresponds to the long-timescale integral of the VAC, a comparison of the VAC errors indicates differences in the time-resolved dynamics between the CG models and atomistic simulations on the short timescale. The same qualitative error trends observed for diffusivity are observed for the VAC errors. Increasing errors are exhibited with respect to increasing number of anisotropic groups in the molecule, and the correlation between accuracy and number of beads is stronger than the effect of mapping changes for a fixed number of beads. The overall conclusion is that the short-timescale fidelity of the CG dynamics follows a similar dependence to the diffusivity.

The relative error in ΔH_v is presented in Figure 6.8 for all molecules and mappings. The relative error is calculated as the difference between the CG and atomistic enthalpies of vaporization, normalized by the atomistic value, and averaged over multiple independent trajectories. We observe that IBI exhibits consistently lower errors compared to MSCG, while both underestimate ΔH_v. As with diffusivity, several quantitative aspects of this comparison are relevant to the accuracy of each CG approximation with respect to dimension reduction.

First, we note that IBI consistently shows a lower error for ΔH_v than MSCG that can be attributed to the additional attractive interactions in IBI CG models resulting from pressure corrections. This is confirmed by comparison with the ΔH_v errors for IBI without pressure correction (IBI-NP), which are larger than IBI and, in some cases, MSCG. Thus, IBI with pressure correction is not only capable of reproducing two on-target properties (the $g(r)$ and pressure), but also is able to capture additional physical information which is reflected in an off-target property. This supports the underlying premise of bottom-up coarse-graining (i.e.,
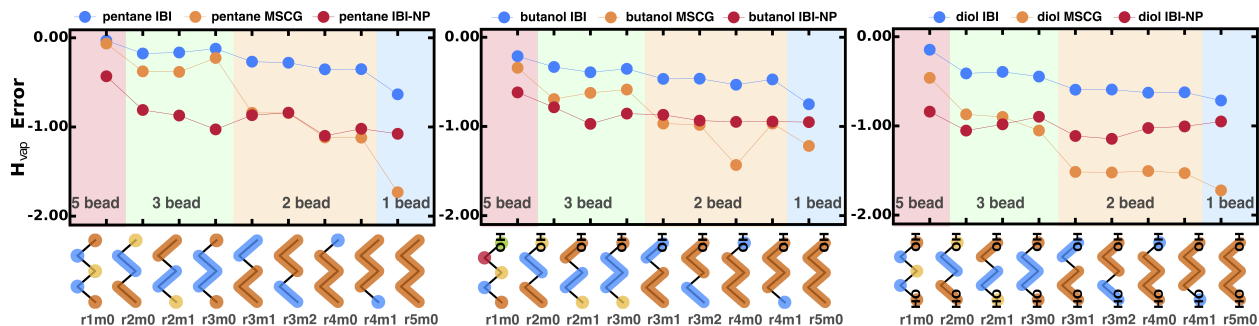
Figure 6.8. Normalized error in enthalpy of vaporization for IBI, MSCG and IBI without pressure correction (IBI-NP) methods across all mappings for pentane, 1-butanol and 1,3-propanediol from left to right. The standard errors are within the size of the markers in all cases. All three methods underestimate the error, but IBI with pressure corrections performs the best.

that reproducing a subset of critical atomistic properties will simultaneously reproduce other off-target properties), but also suggests that the existing algorithm is likely underdetermined with respect to the target information supplied.

Second, we note that the ΔH_v error generally increases with respect to the number of anisotropic moieties in each molecule (i.e., the absolute error ranked as 1,3-propanediol > 1-butanol > pentane). For IBI, this dependence is weak, but can still be seen for intermediate mappings (e.g., comparing the 2 and 3 bead mappings across the liquids). Whereas for MSCG, the underestimation increases with the number of anisotropic moieties, while also exhibiting a stronger resolution dependence.

Finally, we note that the ΔH_v errors correlate more strongly with the number of beads, rather than the resolution of the model. However, the mapping dependence of errors is weaker for IBI models compared to the MSCG models, and no consistent trend is observable for IBI-NP. The ΔH_v error for IBI-NP is inversely correlated with the final pressure for the CG model (not shown), which explains the lack of mapping dependence and is also consistent with the interpretation that the pressure information supplied in IBI results in its low ΔH_v errors. The weak mapping dependence of IBI for enthalpy of vaporization (similar to $g(r)$) coupled with the correlation between the simulation pressure and ΔH_v error, suggests that the IBI models are information limited, rather than representability limited.

6.3 Conclusions

We have systematically investigated the effect of dimension-reduction on the ability of CG models to reproduce the on-target and off-target properties of fine-grained atomistic systems. The weak dimension-dependence of the on-target property accuracy provides evidence that these models are information limited, rather than representability limited. We note that the question of representability limitations for off-target properties cannot be addressed here, because by virtue of being off-target these properties were not included in the objective function of the two parameterization algorithms (IBI and MSCG). Presumably at some point

simple isotropic pair-potentials will have to make compromises in the number and accuracy of atomistic properties that they can reproduce. However, until algorithms are developed that can flexibly incorporate multiple properties into CG parameterization, the general representability limitations of isotropic CG models will be left unanswered. In contrast, for on-target properties, we have observed that the reproduction of atomistic forces is deeply in the representability limited regime for all of the models that we investigated, while $g(r)$ reproduction is far from the representability limited regime. The former observation suggests that the atomistic forces are poorly conditioned in practice for CG model parameterization, whereas the latter observation suggests that pair-wise potentials are capable of supporting more physical information than is provided by the $g(r)$ alone.

The results presented here also demonstrate that CG model accuracy for off-target properties is strongly sensitive to dimension reduction. One interpretation of these results is that the low-resolution models are representability limited, with the consequence that fewer properties can be reproduced in the low-resolution limit. As pointed out above, the representability limitations of off-target properties cannot be assessed by the current study. An alternative interpretation that may be useful in guiding further studies is that the IBI and MSCG approximations become less useful at low-resolutions for reproducing off-target properties. In this latter interpretation, the low-resolution models are information limited, rather than representability limited. This latter interpretation is supported by the non-convex optimization surfaces of both $g(r)$ and the atomistic force errors, which admit a potentially large range of parameterizations. Additionally, for IBI, the capability to reproduce the target $g(r)$ is nearly unaffected by resolution, and the accuracy is uniformly high in all cases. Thus, the IBI models are not representability limited for their target property, and can potentially accommodate more information by including additional atomistic data into the parameterization algorithm. In the case of MSCG, the force errors are uniformly high in all cases, leading us to conclude that all of the models are representability limited with respect to the force reproduction, while cautioning that these models do still reproduce off-target properties to a variable extent. Thus, while the MSCG models are representability limited for their target property, they too can potentially accommodate more information by including additional atomistic data into the parameterization algorithm.

While atomistic molecular dynamics yields an abundance of potential information for CG models to reproduce, existing algorithms utilize only a small fraction of it during parameterization. The results presented here suggest that isotropic pair potentials are likely capable of supporting additional chemical detail beyond reproducing the atomistic $g(r)$, forces, and pressure. Moreover, we have observed that what have been previously classified as representability limitations are more likely information limitations that could be addressed with algorithm development and without increasing model complexity.

APPENDIX

6.A   Simulation Details for Bottom-Up CG-MD Study

6.A.1   IBI Procedure for Bonded Potentials

All IBI parameterizations were performed using a custom implementation developed by our group. The target distribution functions for bonds, angles, and dihedrals were sampled from a single atomistic trajectory and histogrammed with bin sizes of 0.1 Å, 1°, and 1° respectively. The potentials were extrapolated to a range of 0-10 Å, 0-180° and -180-180° for bonds, angles and dihedrals respectively. Quadratic extrapolation was utilized to obtain potential values outside the sampled regions for bonds and angles. Linear extrapolation was utilized to obtain potential values outside the sampled region for dihedrals.

For each bonded IBI iteration, single-molecule CG simulations with shrink-wrap boundary conditions were performed to sample the intramolecular distributions and avoid confounding condensed phase non-bonded interactions. The CG simulations were first relaxed in the NVE ensemble with restrained atomic displacements of 0.1 Å per time step for 10 ps, followed by a 0.1 ns NVT equilbriation at 298K. The production data for sampling the intramolecular distributions were obtained by extending this trajectory for 10 ns and sampling configurations every 0.1 ps. The IBI procedure was run for 60 iterations, and the tabulated potentials from the iteration with the smallest distribution error (Eq. 7.8) was used for the subsequent non-bonded IBI parameterizations and CG property calculations. The average distribution error was given by

$$\varepsilon_{\text{IBI,b}} = \frac{1}{N_{\text{g}}} \sum_{\gamma}^{N_{\text{g}}} \frac{\sum_{\kappa} \|P_{\gamma}^{\text{AA}}(\kappa) - P_{\gamma}^{\text{CG}}(\kappa)\|}{\sum_{\kappa} P_{\gamma}^{\text{AA}}(\kappa)}, \tag{6.6}$$

where γ are different bonds, angles and dihedrals, κ are the values of these quantities, $P_{\gamma}(\kappa)$ is the value of distribution function of γ at κ and N_{g} is the total number of these interactions.

6.A.2   IBI Procedure for Non-bonded Potentials

All IBI parameterizations were performed using a custom implementation developed by our group. The target RDFs were sampled from five independent trajectories and histogrammed with a bin size of 0.01 Å and a range of 0-10 Å. An exponential function was used to extrapolate the potential values for the inner region where the RDF was not sampled. The fitted potentials were further interpolated to a finer grid of 0.001 Å when utilized in CG simulations.

For the CG sampling necessary to iteratively refine the pair-wise interactions, a system containing at least 1000 beads was used for all mappings. This system was equilibrated for 1 ns in an NVT ensemble, as discussed in the Molecular Dynamics Simulations section of the main text, and the last frame was used as an initial configuration for the start of the IBI

procedure. For subsequent iterations, the last frame from the previous iteration was used as the initial configuration.

At each iteration of IBI, 5 independent CG trajectories were simulated to obtain average correction terms in Eq. 7.11,

$$U(r)^{\mathrm{k}} = U(r)^{\mathrm{k}-1} + \frac{1}{\mathrm{n}} \sum_{\mathrm{i}=1}^{\mathrm{n}} \ln \frac{g(r)^{\mathrm{k}-1}}{g(r)^{\mathrm{AA}}} + \frac{\omega}{\mathrm{n}} \sum_{\mathrm{i}=1}^{\mathrm{n}} A \left(1 - \frac{r}{r_{cut}}\right). \qquad (6.7)$$

For the first 19 iterations of IBI, only the RDF correction was applied (i.e., $\omega = 0$ in Eq. 7.11). From the 20th iteration onwards, both RDF and pressure corrections were applied starting with $\omega = 0.01$. After that, the weighing coefficient ω was dynamically changed depending on the pressure error trend over the previous 5 iterations. For the first 20 iterations, each CG trajectory was of length 0.1 ns and sampled at 0.1 ps. After turning on the pressure correction, a longer trajectory of 1 ns with sampling of 0.1 ps was used to obtain accurate pressures.

The IBI procedure ran for a maximum of 80 iterations. The procedure was allowed to terminate earlier if the pressure error fell below 10 atm. From all iterations, the tabulated potentials from the iteration with the lowest value of the objective function

$$\varepsilon_{\mathrm{IBI,nb}} = \frac{1}{\mathrm{N}_{\mathrm{P}}} \sum_{\mathrm{p}=1}^{\mathrm{N}_{\mathrm{P}}} \frac{\sum_{r=0}^{r_{cut}} \|g_{\mathrm{p}}^{\mathrm{AA}}(r) - g_{\mathrm{p}}^{\mathrm{CG}}(r)\|}{\sum_{r=0}^{r_{cut}} \|g_{\mathrm{p}}^{\mathrm{AA}}(r) + g_{\mathrm{p}}^{\mathrm{CG}}(r)\|} + 0.05 * \frac{\|P^{\mathrm{AA}} - P^{\mathrm{CG}}\|}{P^{\mathrm{AA}}} \qquad (6.8)$$

were retained and utilized in subsequent production simulations. The first and second terms on the RHS of 7.12 indicate the normalized error in RDF ($g_p(r)$) and pressure (P), respectively. The weight of 0.05 was applied to the pressure errors due to their inherently higher variance than the RDF error.

6.A.3   MSCG procedure for Non-bonded Potentials

All MSCG models were parameterized with the VOTCA package [141]. 10 ns of atomistic data with 1 ps sampling was used as training data for each liquid. Only non-bonded potentials were parameterized with MSCG, while all intramolecular modes were taken from the corresponding IBI models. This choice was motivated by previous work indicating that this hybrid approach results in more accurate potentials than models where all terms are parameterized with MSCG [140]. The forces were fit with a spline grid spacing of 0.1 Å and a range up to 10 Å. The lower bound of the interval was taken to be the first radial distance where the RDF exceeds 0.05. The frames per block for each mapping were chosen such that the number of equations (number of beads × number of frames) were 50 times the total number of spline functions over all pair-types. The block average of forces was integrated to obtain the non-bonded MSCG potentials. The final CG potentials were further interpolated to a spacing of 0.05 Å for use in production simulations. An exponential function was used to extrapolate the potential as per default VOTCA settings unless it failed to produce a repulsive core, in which case a quadratic function was used.

6.A.4 Property Calculations

Since the properties that are evaluated in this work have distinct sampling requirements, independent trajectories were simulated for each property. The number of trajectories and sampling details are described below for each property. In all cases, the initial configuration was obtained from the pre-equilibrated trajectories described in the main text. To obtain statistically independent trajectories, the velocities were reinitialized in each case, followed by a 10 ps NVE simulation with constrained atomic displacements of 0.1 Å per time step and a 1 ns equilibration in the NVT ensemenble. These trajectories were then extended by variable durations to collect production data, as described below.

The force error per CG bead (ε_F) was calculated as the square of the magnitude of the difference in force calculated in the atomistic ($\mathbf{F}_{\mathrm{j,t}}^{\mathrm{AA}}$) and coarse-grained representations ($\mathbf{F}_{\mathrm{j,t}}^{\mathrm{CG}}$), averaged over all coarse-grained beads (j) and atomistic snapshots (t). The former quantity was calculated according to

$$\mathbf{F}_{\mathrm{j}}^{\mathrm{AA}} = \sum_{\mathrm{i} \in \mathrm{j}} \mathbf{F}_{\mathrm{i}}, \tag{6.9}$$

where the index i runs over the atoms in bead j for a given mapping. The final expression for the force error,

$$\varepsilon_\mathrm{F} = \frac{\sum_{\mathrm{j,t}} \left\| \mathbf{F}_{\mathrm{j,t}}^{\mathrm{AA}} - \mathbf{F}_{\mathrm{j,t}}^{\mathrm{CG}} \right\|^2}{\sum_{\mathrm{j,t}} \left\| \mathbf{F}_{\mathrm{j,t}}^{\mathrm{AA}} \right\|^2}, \tag{6.10}$$

was further normalized with respect to the average force magnitude in the atomistic representation to provide a fractional measure of the error.

Only non-bonded forces are used in evaluating Eq. 6.9, as these form the target property for the MSCG parameterization. Additionally, poorly sampled short-range interactions are omitted in the MSCG training by fixing a lower bound of radial distance (as described in the MSCG section above). The same radial bound is used for error evaluation to avoid the error being dominated by these points. The numerator in Eq. 6.9 is proportional to the residual of the MSCG system of linear equations as defined in [121].

The RDF was calculated for each CG pair according to Eq. 2 in the main text. The error in the RDF, (ε_RDF), is calculated as an average over errors for all pair types (p) according to

$$\varepsilon_\mathrm{RDF} = \frac{1}{\mathrm{N_P}} \sum_{\mathrm{p}=1}^{\mathrm{N_P}} \frac{\sum_{r=0}^{r_{cut}} \exp(g(r)_\mathrm{p}^{\mathrm{AA}}) \| g(r)_\mathrm{p}^{\mathrm{AA}} - g(r)_\mathrm{p}^{\mathrm{CG}} \|}{\sum_{r=0}^{r_{cut}} \exp(g(r)_\mathrm{p}^{\mathrm{AA}}) \| g(r)_\mathrm{p}^{\mathrm{AA}} \|}. \tag{6.11}$$

where $g(r)_\mathrm{p}^{\mathrm{AA}}$ is the RDF in the atomistic representation, $g(r)_\mathrm{p}^{\mathrm{CG}}$ is the RDF in the coarse-grained representation, and the summations run from $r = 0$ to the radial cutoff for the pairwise interactions, r_{cut}. To avoid having the poorly sampled regions dominate the error measure, the difference in RDFs is weighted by $\exp g(r)_\mathrm{p}^{\mathrm{AA}}$ and the error at each r is included in the summation only if $g(r)_\mathrm{p}^{\mathrm{AA}} > 0.1$. In the denominator, the error is normalized by the average value of atomistic RDFs over radial distances, with the same exponential weighting.

The calculation details for the atomistic systems RDFs, which serve as the target RDF for the IBI parameterization, are described in the IBI procedure. For calculating the RDF of

each CG system, an MD simulation was performed in the NVT ensemble for 0.5 ns with 0.1 ps sampling. The initial configuration for the MD simulation was same as the pre-equilibrated configuration used for the IBI procedure.

The self-diffusion coefficient was obtained from the mean square displacement of each molecule using the Einstein relation

$$D = \lim_{t \to \infty} \frac{1}{6} \frac{d}{dt} \langle \|\mathbf{r}_{\mathrm{m}}(t) - \mathbf{r}_{\mathrm{m}}(0)\|^2 \rangle, \tag{6.12}$$

where $\mathbf{r}_{\mathrm{m}}(t)$ is the position vector of the molecule center of mass at time t. The angular brackets in Eq. 7.17 indicate an average over all the molecules and time origins. The fractional error in the diffusion coefficient was calculated according to

$$\varepsilon_{\mathrm{D}} = \frac{1}{\mathrm{N_t}} \sum_{\mathrm{i}}^{\mathrm{N_t}} \frac{D^{\mathrm{CG}} - D^{\mathrm{AA}}}{D^{\mathrm{AA}}} \tag{6.13}$$

where $\mathrm{N_t}$ is the number of independent trajectories.

Five independent simulations were performed for both atomistic and coarse-grained representation to obtain the self-diffusion coefficient. The trajectories were extended to collect 400000 molecular samples at a sampling frequency of 1 ps in NVT ensemble. For each trajectory, the numerical derivative of the first continuous 10 ps interval after 100 ps which exhibited a linear slope (0.95- 1.05) of log(MSD) versus time were used to obtain the diffusion coefficient. The diffusion coefficient utilized in Eq. 7.18 was obtained as the average over all five trajectories.

The VAC of each molecule, C_{m} was calculated as

$$C_{\mathrm{m}}(t) = \frac{\langle \mathbf{v}_{\mathrm{m}}(0) \cdot \mathbf{v}_{\mathrm{m}}(t) \rangle}{\langle \mathbf{v}_{\mathrm{m}}(0) \cdot \mathbf{v}_{\mathrm{m}}(0) \rangle}, \tag{6.14}$$

where $\mathbf{v}_{\mathrm{m}}(t)$ is the velocity vector of the molecule center of mass at time t, t has a range of 0-500 fs, and the angular brackets indicate an average over all molecules and time origins. The error for a single trajectory is the absolute difference in VAC for the atomistic and coarse-grained representation averaged over all time differences and is normalized by the absolute value of atomistic VAC averaged over all time differences. The net error for VAC, similar to diffusion coefficient, is an average over the error for $\mathrm{N_t}$ independent trajectories, given by

$$\varepsilon_{\mathrm{VAC}} = \frac{1}{\mathrm{N_t}} \sum_{\mathrm{i}}^{\mathrm{N_t}} \frac{\sum_{\mathrm{t}} \|C_{\mathrm{m,t}}^{\mathrm{AA}} - C_{\mathrm{m,t}}^{\mathrm{CG}}\|}{\sum_{\mathrm{t}} \|C_{\mathrm{m,t}}^{\mathrm{AA}}\|}. \tag{6.15}$$

Five independent simulations were performed for both atomistic and coarse-grained representations to obtain the VAC. 10 ps of production data was collected from each trajectory at a sampling frequency of 1 fs in NVT ensemble. The VAC was calculated for time differences from 0 fs to 500 fs.

The molar enthalpy of vaporization, ΔH_v, is calculated by

$$\Delta H_v = U_v + k_B T - U_c, \tag{6.16}$$

where U_v and U_c are the mean molar potential energies in the gas phase and condensed phase, respectively. U_v was calculated by single-molecule MD simulation. The net fractional error in the enthalpy of vaporization was calculated according to

$$\varepsilon_H = \frac{1}{N_t} \sum_i^{N_t} \frac{\Delta H_v^{CG} - \Delta H_v^{AA}}{\Delta H_v^{AA}} \tag{6.17}$$

where N_t is the number of trajectories.

Five independent simulations were performed in both the gas and condensed phases in the coarse-grained representation to obtain the potential energies for the ΔH_v calculation. 1 ns of production data was collected from each trajectory at a sampling frequency of 1 ps in NVT ensemble. An identical procedure was followed to obtain the gas phase atomistic potential energies for the ΔH_v calculation. For calculating the condensed phase potential energy in atomistic representation, thermodynamic data from 1 ns of the parameterization trajectories described in the main text were used.

6.B Additional Results Referenced for Bottom-Up CG-MD Study

(a) pentane   (b) 1-butanol   (c) 1,3-propanediol
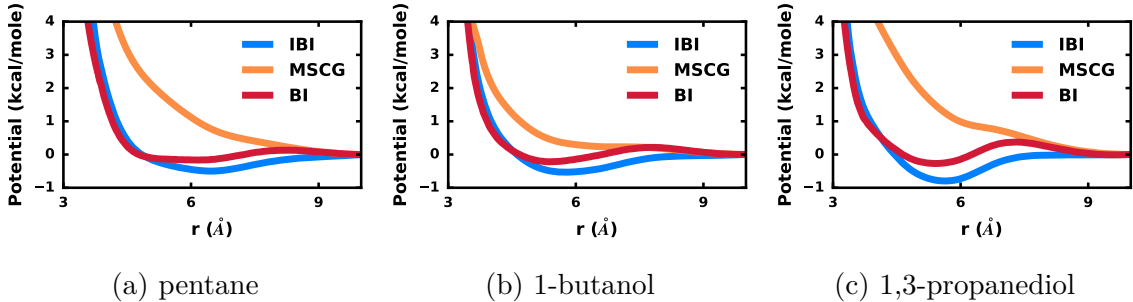
Figure 6.9. IBI, MSCG and BI potentials for pentane, 1-butanol and 1,3-propanediol for the r5m0 mapping. Even though the potentials for IBI, MSCG and BI are very different for these mappings, their force errors are very similar: 0.956, 0.948 and 0.966 for pentane-IBI, pentane-MSCG and pentane-BI respectively, 0.973, 0.971 and 0.976 for 1-butanol-IBI, 1-butanol-MSCG and 1-butanol-BI respectively, 0.967, 0.964 and 0.971 for 1,3-propanediol-IBI, 1,3-propanediol-MSCG and 1,3-propanediol-BI respectively.

(a) pentane r1m0 1-1   (b) pentane r1m0 1-2   (c) pentane r1m0 1-3

(d) pentane r1m0 2-2   (e) pentane r1m0 2-3   (f) pentane r1m0 3-3

(g) pentane r5m0 1-1

(h) 1-butanol r5m0 1-1

(i) 1-butanol r4m0 1-1   (j) 1-butanol r4m0 1-2   (k) 1-butanol r4m0 2-2
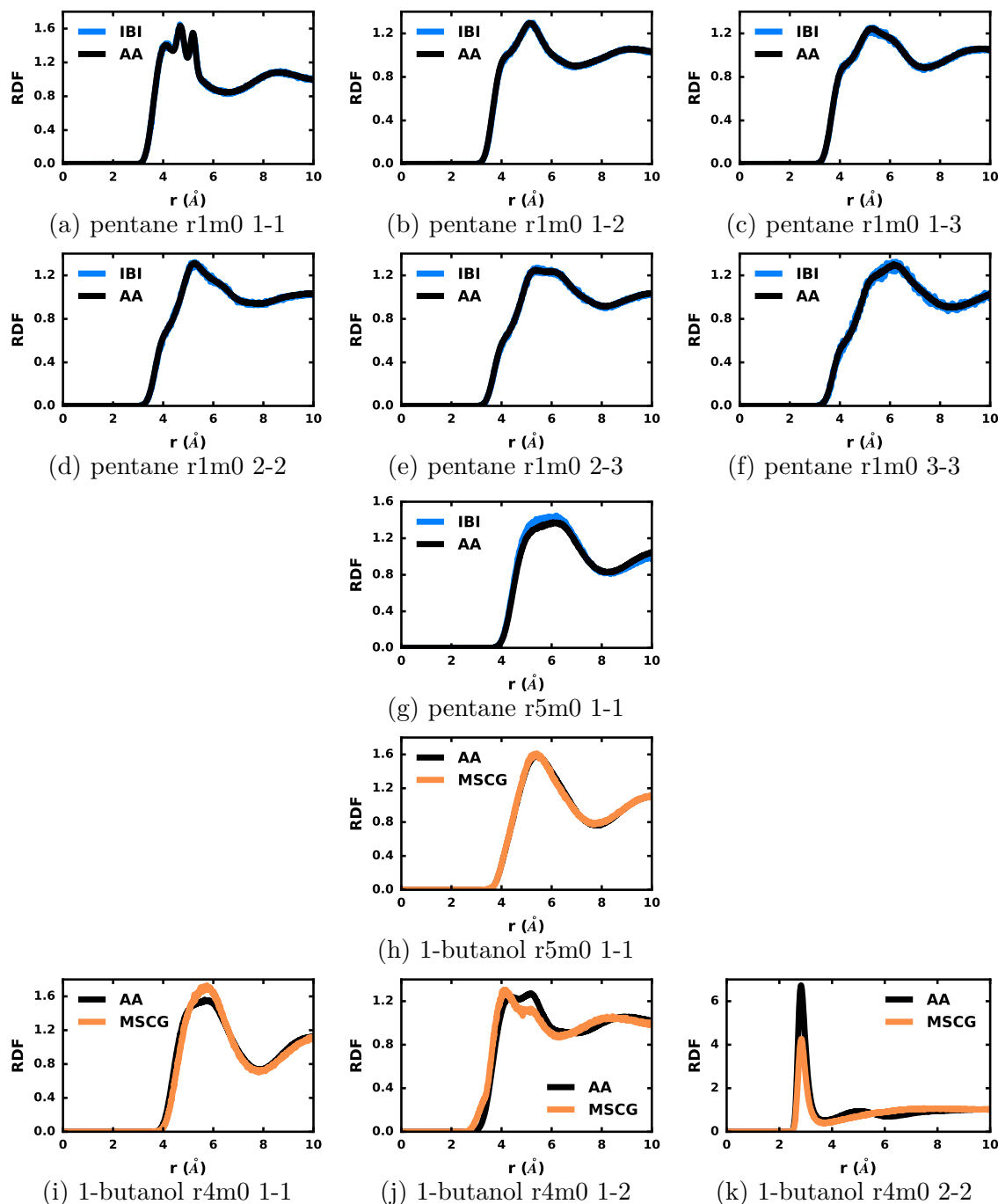
Figure 6.10. RDFs for IBI and MSCG compared with target AA RDFs for the lowest and highest error cases among all mappings and molecules. For IBI, the RDF error is lowest and highest for the pentane r1m0 (0.010) and pentane r5m0 (0.039) mappings, respectively. The IBI RDFs for all pairs for pentane r1m0 and pentane r5m0 are shown in (a-f) and (g), respectively. For MSCG, the RDF error is lowest and highest for the 1-butanol r5m0 (0.015) and 1-butanol r4m0 (0.16) mappings, respectively. The MSCG RDFs for all pairs for 1-butanol r5m0 and 1-butanol r4m0 are shown in (h) and (i-k), respectively.
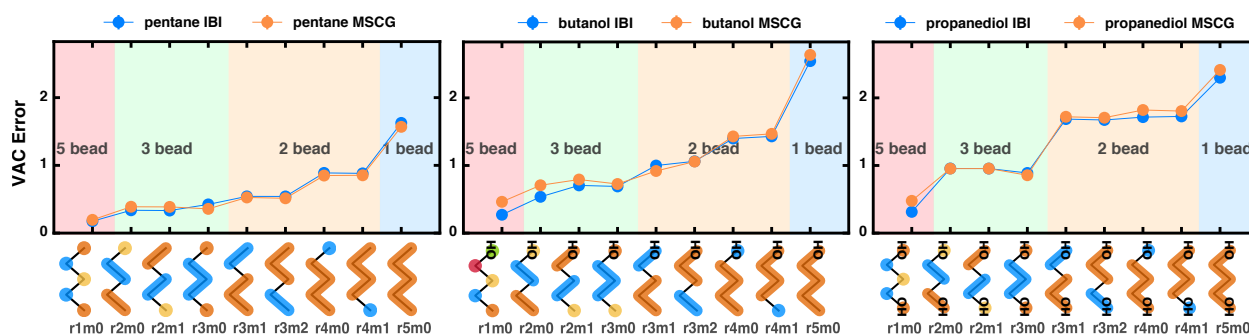
Figure 6.11. Error in VAC obtained using IBI and MSCG methods across all mappings for pentane, 1-butanol, and 1,3-propanediol from left to right. Standard errors are within the marker in all cases. The error trend is very similar to that obtained for diffusion coefficients. Both IBI and MSCG exhibit similar errors and the error increases with number of hydroxyl groups. Overall the error is dominated by the number of beads, with very little effect of mapping.

7. TRANSFER LEARNING FROM PHYSICS-INSPIRED MODELS TO MACHINE LEARNING BASED COARSE-GRAINED POTENTIALS

We investigated if the coarse-grained (CG) model is representability limited or information limited in Chapter 6, by systematically comparing two common bottom-up parameterization schemes across various mapping operators for three molecules. Throughout this previous study, the potential energy surface (PES) was represented as a combination of spline extrapolated bond stretching, angle bending, and pairwise nonbonded interactions. In this chapter, we investigate the role PES as well as its parameterization scheme plays in the CG model performance.

More specifically, we propose a new distinction in how the potential energy surface is represented – physics vs black-box. The former constitutes of a CG representation motivated by a physical intuition. Examples of physics based representations are decomposition into stretching, bending and torsional interactions; pair-wise non-bonded decomposition or functional forms of potential interaction such as Lennard-Jones, Morse, etc. The latter black-box approach on the other hand, constitutes representations which do not have a strong physical intuition but provide the flexibility to capture more complex interactions, such as many-body forces. The most obvious example of black-box representations is Machine Learning (ML) based coarse-grained potentials, but use of tabulated spline based potentials over Lennard-Jones can also be considered a black-box approach. Even for the ML based potentials, physics is incorporated in the choice of input features to satisfy physical constraints such as rotational and translational invariance. Hence, we find a continuum of physics-inspired choices and black-box schemes across various CG parameterization schemes to represent the atomistic system in the best possible way. In addition to the representation choice, the parameterization procedure can also be inspired by physical intuition. For instance, in Iterative Boltzmann Inversion (IBI), the PES is corrected over discretized radial separation between beads in order to obtain the correct structural correlation at that radial separation. The correction scheme is inspired by the fact that in canonical ensemble, independent degrees of freedom show a Boltzmann distribution.

A physics based representation offer simpler interpretation, and can be easier to parameterize in comparison to a black-box approach (Figure 7.1) but loses some information due to simplified representation. For instance, the pairwise decomposition of non-bonded interactions can be inadequate in reproducing many-body interactions and inherent anisotropy in the system. In this chapter, we propose a hybrid approach, where the bulk of the potential interactions are captured by a simple physics based model (SP) and the remainder of interactions are captured by a blackbox machine-learning based model (ML). In doing so, we retain the physics captured by the SP model, and reduce the work to be done by ML model while still leveraging its complexity. We call this new model– Transfer Learning (TL) model as we are transferring the information learnt by the SP model and assisting the training of ML model (Figure 7.1). This approach is applicable for any combination of SP and ML model, as long as the ML model is trained on forces as a target property and the SP model preferably uses some property other than forces. In the following sections, we describe the TL approach by providing the equations to construct the TL model in detail and finally, compare all three approaches by coarse-graining selected molecules.

7.1 Methods

7.1.1 Construction of the Transfer Learning Model

The true forces $\mathbf{F}^{\mathrm{AA}}$ on any CG bead in a given configuration which serve as target are obtained by performing an all atomistic MD simulation and then calculating the net force on each CG bead (j) as the sum of forces on all the atoms (i $\in$ j) constituting it, $\mathbf{F}^{\mathrm{AA,j}} = \sum_{\mathrm{i}\in\mathrm{j}} \mathbf{f}^{\mathrm{AA,j}}$. For a general ML model, the objective function is typically the root mean square error in the forces averaged over all beads (j) and configurations (t),

$$L = \sqrt{\frac{1}{N_{\mathrm{j}} N_t} \sum_{\mathrm{j},t} \|\mathbf{F}_{\mathrm{AA,j},t} - \mathbf{F}_{\mathrm{CG,j},t}\|} \tag{7.1}$$

In order to construct the TL force-field, first, we calculate the force $\mathbf{F}^{\mathrm{CG}}_{\mathrm{SP}}$ on a CG bead as obtained by the SP model. Due to the simplicity of model (such as pair-wise assumptions), it does not completely reproduce the target force $\mathbf{F}^{\mathrm{AA}}$. Note that a CG model will also lose

144
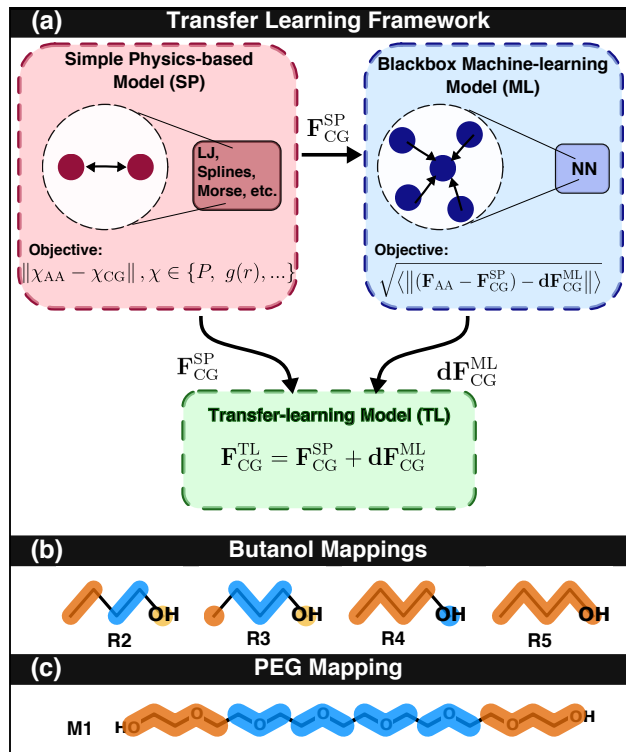
Figure 7.1. (a) Schematic explaining the transfer learning approach proposed here. The force on a bead is obtained by a sum of force by the simple physics and a correction term predicted by the blackbox machine learning based model. The circles in the schematic represent atoms and the pairwise and many-body forces among the atoms for SP and ML, model respectively. The objective functions for SP and ML model are also shown. NN and LJ indicate neural network and Lennard-Jones potential, respectively. (b) Mappings of butanol of resolution R 2-5 and (c) two 5-bead mappings of PEG with different CG bead definition, simulated to compare all the models

some information due to the lost degrees of freedom. However, we can correct for the forces ($\mathbf{dF}_{\mathrm{ML}}^{\mathrm{CG}}$) which are lost due to simple architecture through a more complex ML architecture. Thus, we obtain a more accurate description by using a cheaper yet physics motivated framework for a good guess and correcting it using machine learning. This approach has been successfully implemented in electronic level theory, where quantum chemical accuracy is achieved by correcting semiemperical DFT level prediction using machine learning.[16, 31, 133] We calculate the SP model force error $\mathbf{dF}^{\mathrm{AA,j}}$ for each CG bead j as $\mathbf{dF}_{\mathrm{SP}}^{\mathrm{AA,j}} =$

$\mathbf{F}^{\text{AA,j}} - \mathbf{F}_{\text{SP}}^{\text{CG,j}}$. Now, we train the ML model to reproduce this error $\mathbf{dF}_{\text{SP}}^{\text{AA}}$ rather than the complete force $\mathbf{F}^{\text{AA}}$. The new objective function for the ML architecture becomes,

$$L = \sqrt{\frac{1}{N_{\text{j}} N_t} \sum_{\text{j},t} \|\mathbf{dF}_{\text{AA,j,t}} - \mathbf{dF}_{\text{CG,j,t}}^{\text{ML}}\|}, \tag{7.2}$$

instead of Eq.7.1. The final force on any CG bead (j) is given by two components, the bulk of the force $\mathbf{F}_{\text{SP}}^{\text{CG,j}}$ by the SP model and the correction force $\mathbf{dF}_{\text{ML}}^{\text{CG,j}}$ by the ML model, $\mathbf{F}^{\text{CG,j}} = \mathbf{F}_{\text{SP}}^{\text{CG,j}} + \mathbf{dF}_{\text{ML}}^{CG,\text{j}}$.

In this chapter, for the SP model, we use tabulated potentials for bonds, angles, non-bonded pairwise interactions which are obtained using Iterative Boltzmann Inversion (IBI)[134] with and without pressure correction[172], and the DeePCG framework for the ML model[184]. IBI with and without pressure correction has been extensively used for various applications[2, 10, 125], and incorporates structural information in contrast to force information and hence, was chosen for SP model; whereas the choice of DeePCG model was purely due to ease of adoption.

7.1.2 Iterative Boltzmann Inversion

In the IBI process, all potential functions are corrected iteratively using a correction proportional to the logarithmic error of radial distribution function (RDF) as below,

$$U_{ab}(r)^{\text{k}} = U_{ab}(r)^{\text{k}-1} + \alpha k_B T \ln \frac{g_{ab}(r)^{\text{k}-1}}{g_{ab}(r)^{\text{AA}}} \tag{7.3}$$

where $U_{ab}(r)^{\text{k}}$ and $g_{ab}(r)^{\text{k}}$ (Eq. 7.9) are the potential function and RDF at the kth iteration, respectively, for the pair of CG beads of type a and b. The $g_{ab}(r)^{\text{AA}}$ is the target atomistic RDF for the same pair type, T is the simulation temperature, k_B is the Boltzmann constant and α is a damping coefficient. Through the second term on RHS, the potential is corrected iteratively so that the CG RDF $(g(r)^{\text{k}-1})$ approaches the target RDF $(g(r)^{\text{AA}})$. Pairs separated by less than four bonds do not experience nonbonded interactions but rather bond, angle and torsional interactions, which are also parameterized iteratively using IBI (Section 8.B) to obtain the respective distributions.

The Boltzmann inverted (Eq. 7.10) potential is used as an initial guess for the potential and hence, crude information about the local structure is incorporated apriori and iteratively corrected to obtain the accurate pairwise correlations. When pressure correction is included, an additional attractive potential correction of the form,

$$\beta A \left(1 - \frac{r}{r_{cut}} \right), \tag{7.4}$$

is included where the coefficient A is a function of the error in pressure as proposed in Wang et al[172] and β is a damping coefficient. Thus, the converged SP potential reproduces both pair-wise RDF and net system pressure. The IBI parameterization procedure for bonded and nonbonded potentials with additional explanatory equations are provided in detail in Section 8.B-8.C.

7.1.3 DeePCG

In the DeePCG framework, the potential energy is decomposed into energy contribution of each bead similar to various atomistic ML forcefields.[11] The per-bead energy is obtained as an output of a deep neural network (DNN) with an input capturing the local neighborhood of the bead. More specifically, this descriptor is a vector of interbead radial and angular information between the bead and its short-distance neighbor and just radial information between the bead and its long-distance neighbors (Section 8.D). In order to satisfy the translational and rotational invariance, the interbead distances are expressed in terms of a local coordinate frame determined for each bead formed by the bead with its two nearest neighbors (Section 8.D). This descriptor hence sufficiently captures the local neighborhood of a bead and the DNN calculates the energy of the bead as function of this descriptor. The DNN is trained to reproduce the forces as sampled in atomistic MD simulation (Eq. 7.1), where the CG forces are obtained by the derivative of the potential energy function. The transfer-learning model is constructed by first obtaining the IBI potentials, evaluating IBI force errors per bead for different configurations, and training the DeePCG model for predicting these force errors (Eq. 7.2) instead. All the details of training of the neural

147

network for both ML and TL model are explained and the optimized hyperparameters are reported in Section 8.D.

7.1.4 Model System

In this chapter, we construct and compare a total of five models using IBI (SP), IBI with pressure correction (SP_{PC}), DeePCG (ML), a transfer-learning model which is a combination of IBI and DeePCG (TL) and another transfer-learning model which is a combination of IBI with pressure-correction and DeePCG (TL_{PC}). In doing so, we provide first demonstration of using transfer learning model for CG forcefield and test its performance in comparison with the simple physics model and blackbox model. The performance for on-target properties of RDF and pressure is compared to test if the transfer learning retains the physics from the SP model. Another important question is if transferring this information from the SP model is necessary or the ML model is itself capable of reproducing RDF and pressure. We also compare the reproduction of forces to test if transfer learning hurts the direct force reproduction in the ML model. Further, we test if the inclusion of forces, RDF and pressure translates to other off-target properties such as molecular enthalpy of vaporization (H_v) and self-diffusion coefficient. Additionally, this chapter provides a comparison of traditional physics based SP model and ML model for various mappings, which has been performed in only one study for a star polymer[175]. The error for RDF, pressure, H_v and diffusivity are normalized with respect to the atomistic values, whereas the force error reported is same as the objective function, i.e., the net RMSE over all configurations and beads. The procedure for simulations performed to evaluate these errors as well as the error expressions used are described in Section 8.E.

We perform this comparison for butanol molecule for four mappings from resolution 2-5 (Figure 7.1b). We chose butanol due to its chemical anisotropy and four mapping operators which retained the hydroxyl group as a separate CG bead. These mappings support chemical intuition and displayed slightly reduced force error over other mapping operators, in our previous study[82]. Additionally, as coarse-graining is relevant primarily for large molecules, we also included a polymeric system of 7-mer of poly-ethelyene glycol (PEG) (Figure 7.1c).

The PEG polymer was coarse-grained with 4 CH2-O-CH2 beads and 2 end beads of OH-CH2-O-CH2-CH2, which are chemically distinct groups with typically used 3-5 heavy atoms per bead. All the MD simulations for the target atomistic system, the IBI training as well as simulations for performance evaluation of coarse-grained models were performed using LAMMPS[130] at 298K and 1 atm pressure and the details of simulation procedure are reported in Section 8.A.

7.2   Results
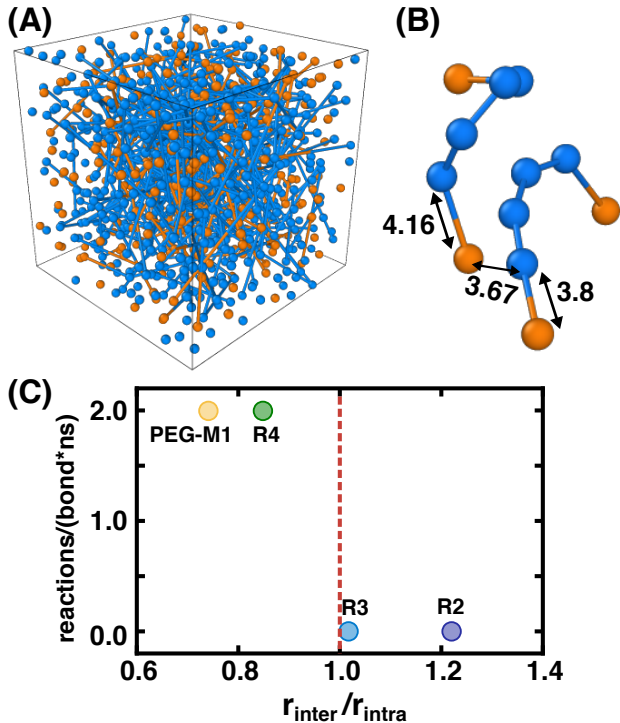
Figure 7.2. (a) A configuration from PEG-M1 MD simulation which shows the bonded beads drift apart as indicated by the stretched (broken) bonds. (b) A snapshot of two PEG-M1 molecules which show lower intermolecular separation than intramolecular bonded separation between orange and blue beads. (c) Intermolecular and intramolecular lengthscale separation in molecule on x-axis vs the rate of bond breaking on y-axis for all mappings. The lengthscale separation is given by the ratio of minimum spearation at which intermolecular beads can be found (r_{inter}) and the maximum spearation at which bonded beads can be found (r_{intra}).

The final error for five properties– RDF, force, pressure, H_v and diffusivity are reported in Figure 7.3. We report the errors only for high and low resolution butanol mappings R3 and R5, respectively and PEG M1 mapping, as we find that these capture the key mapping dependencies in the CG model performance. The performance for the mappings R2 and R4 and any other minor mapping differences are discussed in 8.G. First thing to note is that the PEG M1 mapping is denoted by hatched bar to differentiate this mapping which exhibited unphysical behavior. As the simulation progresses, the intramolecular neighbors of a bead drift away as indicated by the extended bonds in the PEG-M1 configuration (Figure 7.2a). The bead replaces its intramolecular neighbors falsely by other beads forming a "pseudo" molecule. Now, the intramolecular constraining forces act between the beads of this pseudo molecule. Hence, we obtain the correct force distribution and overall structural correlations and will be discussed further below. This unphysical behavior stems from overlap between two length-scales of the system– intramolecular bond length and distance between first shell neighbors. If the distance between bead and its intermolecular neighbor is less than its bonded neighbor, the bead mistakenly applies constraining force on the intermolecular neighbor instead of the bonded neighbor and in the absence of these forces, the real bonded neighbor keep moving away. A representative snapshot of two PEG M1 molecules is shown as an example (Figure 7.2 b), where the distance between blue and orange beads in different molecules (3.67 Å) is less than from their actual bonded neighbors (3.8 and 4.16 Å). This behavior was also observed for R4 mapping but not for R2 and R3 mappings (Figure 7.2c). On y-axis of Figure 7.2c, the reaction rate normalized over total number of bonds is reported. The PEG-M1 and R4 mapping show breaking of all bonds, whereas no bonds are broken for R2 and R3 mappings. The corresponding length-scale separation, i.e. ratio of minimum separation at which intermolecular beads can be founds (r_{inter}) and maximum separation at which bonded beads can be found (r_{intra}) is reported on x-axis. For PEG-M1 and R4 mappings, the ratio is less than 1, indicating the possibility of smaller intermolecular separation than intramolecular seperation. However, for R2 and R3, this ratio is greater than 1 and hence, the bead does not mistake intermolecular neighbor for bonded neighbors. The values for reaction rate and (r_{inter}/r_{intra}) are averaged over all bond types. In addition to the lengthscale ratio, the RDF for intermolecular beads and bonded beads are shown

in Figure 7.5 to display the lengthscale overlap more clearly. PEG-M1 and R4 are more representative of commonly used coarse mappings of 3-5 heavy atoms per bead, but these coarser beads have overlapping intermolecular and intramolecular lengthscales and hence, exhibit unphysical behavior with the DeePCG framework.
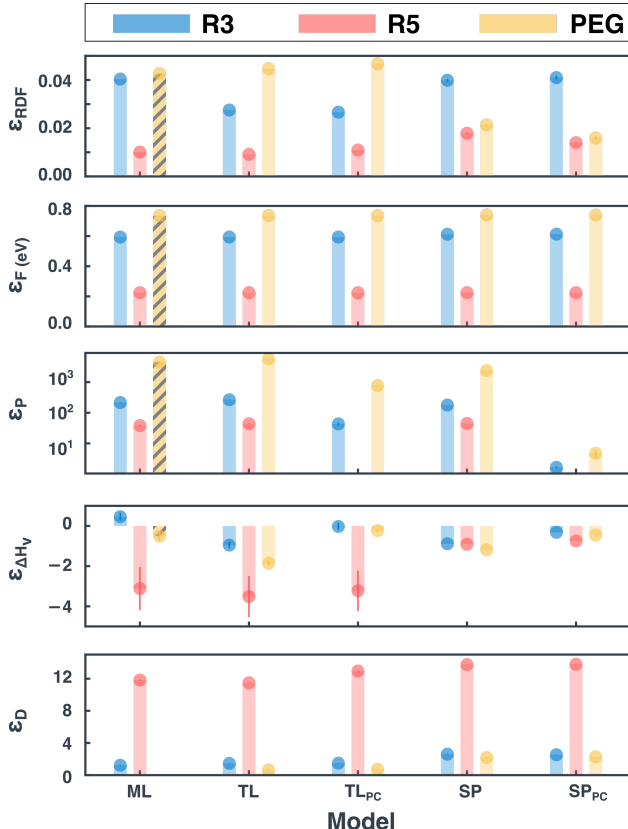
Figure 7.3. Error for RDF, force, pressure, ethalpy of vaporization, and self-diffusion coefficient (from top to bottom), for mapping butanol R3 (pink), butanol R5 (blue) and PEG M1 (yellow). The bar for PEG mapping is hatched to caution the reader that the simulation shows unphysical breaking and formation of bonds. The standard errors are denoted by the thin line passing through the marker. All errors except that for forces are normalized.

A comparison of RDF of bonded beads vs. intermolecular beads clearly show this overlap for R4 and PEG M1 mapping, but not for R2 and R3 mapping which do not exhibit this behavior (Figure 7.5). The TL approach proposed here however corrects this issue because it incorporates direct information about what constitutes a molecule and bending, stretching and torsional potentials. Thus, transfer learning molecular information improves the pure

ML model. Next, we discuss the performance of various models for different properties of interest. It should be noted that, although, the error for various properties is reported for PEG M1 mapping ML model in Figure 7.3, we do not include it in our discussion on comparison of different models as it violates basic physics during the simulation.

The force error appears very similar across all models for all mappings (Figure 7.3 and 7.6). Although, the SP models do not have any direct force information they reproduce force similarly as TL and ML models. On a closer inspection (Figure 7.7), the SP models do have slightly higher error due to exclusion of many-body effects but compared to the absolute error we find these differences to be minor. This was reported in our previous work[82] as well, where IBI and MSCG produced forces with similar accuracy across all mappings. Nevertheless, the comparably high force errors in the ML model suggests that the irreducible error dominates the representability related errors for the forces at these CG resolutions. In particular, the ML model has sufficient complexity to describe many-body interactions and anisotropic interactions; nevertheless, this additionally complexity does not result in lower force errors which is explicable by high irreducible errors. We find that structural information captures the essential forces equally well as supplying the force information directly. Also, it is important to note that the force error for the ML model for PEG M1 and R4 mapping is comparable, even slightly lower than other models. This is a good example of how the ML model can trick the model with low objective function value but undesired characteristics, and serves as a warning of implementing ML models carefully.

On the contrary for RDF, we do see some effect of the model depending on the mapping. For R5 mapping (Figure 7.3), although we see a trend, the RDF errors are very small and the RDF is perfectly reproduced across all models. For R3 mapping (Figure 7.3), we see lower errors for TL models than SP and ML. However, for R2 mapping, we find TL to perform poorly compared to ML and SP and for R4 mapping, TL and SP models are comparable (Figure 7.6). Thus, we do not see a common winner for butanol mappings but two important findings come to light. First, ML outperform the SP IBI based models by reproducing intramolecular structural correlations better. This was seen for R2 and R3 mapping, and the intramolecular RDFs for R3 mapping are reported in (Figure 7.8). Moreover, even though IBI perfectly reproduces the bond and angle correlations (Figure 7.8), it doesn't

translate to structural correlations among all pairs in the molecule and explicit many-body forces through ML like functional form are needed. However, the second finding is IBI-based SP model always outperform in capturing the intermolecular strutural distributions (Figure 7.8). These strengths of ML and SP model do not always translate into the TL approach. For instance, for R2 mappings TL model does not reproduce intramolecular distributions but does very well for R3 mapping. We want to emphasize that these differences are minor. Most of the errors are associated with hydroxyl beads which may be a much finer coarse-graining than commonly used CG models. For R5 mapping, which is very representative of CG models, we find all models reproduce RDF equally well (Figure 7.9).

For the more relevant system of PEG M1 mapping, we find that the SP models clearly outperform the TL models and transfer learning hurts the reproduction of structure (Figure 7.10). One possible reason may be that for the very coarse mapping of PEG M1, with intramolecular constraints, the variance in forces is high and hence, supplying forces to the TL model hurts more than helps. Another possible reason is the overlapping length-scales of bonded and first shell neighboring beads, discussed above in the context of breaking of molecules. Even for the TL model, the NN may be applying the forces to be applied to the bonded neighbor instead to the bead's first shell neighbor. Thus, it is unclear yet if the poor reproduction of structural features by TL model can be generalized. Also, note that, similar to the force error, the RDF error, which serves as the testing objective function, is low for R4 (Figure 7.6) and PEG M1 (Figure 7.3). Thus, despite the breaking and forming of molecules through the simulation, these mappings maintain an accurate overall structure.

Coarse-grained models are known to suffer from the problem of high pressure unless explicitly trained for, which we find in our comparisons as well. The models which account for pressure correction SP_{PC} and TL_{PC} naturally outperform other models (Figure 7.3). The only exception to this was the TL_{PC} model for PEG M1 mapping. We note that pressure can be included in any force-matching framework as a constraint as done by Izvekov et al in the MSCG framework[70] and can be applied to the DeePCG (ML) model as well. However, here we show that transfer-learning with pressure corrected IBI is another approach to do so.

Now, we look at two off-target properties for which none of the models received any information– another thermodynamic property of enthalpy of vaporization (H_v) and a dynamic property of molecular self-diffusion coefficient (D). In our previous work[82], we found that H_v is usually underestimated in CG models due to reduced attraction among beads in the condensed phase than all atomistic system, leading to a negative errror. Hence, we saw pressure-corrected potentials, which essentially incorporate additional attractive forces reproduce H_v, as seen from a comparison of SP vs SP_PC (Figure 7.3 and 7.6). However, a greater effect of inclusion of forces is seen for the ML model which shows positive H_v error mappings R2 and R3. Overall, we find that H_v is better reproduced by ML and pressure-corrected models $\text{TL}_\text{PC}, \text{SP}_\text{PC}$ than SP, TL. Thus, in the context of H_v, for these mappings, the ML model does a good job on its own and transfer learning is not needed, but it doesn't hurt much either. For the R5 mapping, since the vapor phase energy is zero and the potential learnt by the neural network has no reference, we find large error and deviation in error for ML, TL and TL_PC models. Whereas, since in the IBI procedure, we use shifted potential, we find a much better performance. This problem with the neural network based models can be easily resolved by adding an external reference. Another problem in calculation of H_v was found for R4 mapping, where it is not possible to simulate a vapor phase system as the DeePCG framework uses a local frame of reference which requires two neighbors when R4 mapping has only two beads in a molecule. This is not a criticism of DeePCG framework, which we believe was primarily designed for condensed phase simulations. This is merely to illustrate that lack of a physical structure may hurt evaluation of certain properties and in this case, transfer learning also does not help.

Coarse-grained models are known to perform poorly for dynamic properties due to the loss of degrees of freedom and we see that clearly across all models (Figure 7.3 and 7.6). We find the models with many-body forces (ML, TL and TL_PC) show slower dynamics closer to the atomistic systems for all the mappings. However, the errors for all the models are still close and we cannot claim that a better reproduction of an averaged quantity such as diffusivity is sufficient to reproduce dynamics.

7.3 Conclusions

The expectation behind implementation of transfer learning approach is two-fold. First, it should be an improvement on the SP model due to its complex architecture which can potentially capture certain missing features. For butanol R3 mapping, we saw that the inclusion of many-body forces through machine learning framework improves intramolecular structural correlation and dynamics. Secondly, it should be an improvement on the ML model, due to the inclusion of certain physical aspects missing in blackbox approach. In the mappings studied here, we found the TL models does a very good job in fulfilling this expectation. For instance, for the butanol R4 and PEG mapping, the TL models preserve the molecular structure lost in the ML model. The TL model also corrects for pressure for all mappings and allows for calculation of vapor phase enthalpy for the butanol R4 mapping. It does not however solve all the problems such as arbitrary enthaly of vaporization values for one bead mappings such as R5 mapping.

Through these comparisons for different properties across various mappings, we don't see a clear winning model for the butanol mappings whereas for PEG M1 mapping, the SP model shows slightly better performance. However, for both molecules, we find the ML model exhibits certain issues due to lack of physics in the framework; whereas these issues are absent in the SP model and resolved for some cases by the TL model. In the context of the parameterization schemes and mappings studied here, SP model does seem the best choice followed by TL and then ML model. However, the goal of this chapter is not to present a new model of IBI + DeePCG, but rather transfer learning as a mechanism to incorporate sufficiently complex infrastructure and physics in the CG model. We demonstrate this through the reproduction of pressure, minor improvements in structural features, and constraining of beads within a molecule. We hope that this serves as another mechanism the community uses to build better CG models, using combination of different bottom-up approaches or even combination of bottom-up and top-down potentials.

APPENDIX

7.A Molecular Dynamics Simulation Details

LAMMPS was used to perform all molecular dynamics simulations [130]. All atomistic reference simulations used a one fs integration time step, Velocity-Verlet integration, and periodic boundary conditions. Long-range electrostatics were modelled using the particle-particle-particle-mesh (PPPM) algorithm[21] and Lennard-Jones interactions were truncated at 14 Å. All simulations were initialized from diffuse configurations containing 2010 atoms and 5200 atoms for butanol and PEG, respectively, using a cubic grid to place molecules in random orientations without overlap. The velocities were initialized from a uniform distribution obtained with a random seed value and scaled to give the correct kinetic energy. The simulations were first relaxed in the NVE ensemble with restrained atomic displacements of 0.1 Å per time step for 10 ps, followed by a 1 ns NPT equilibration at 298K and 1 atm. The simulations were extended at 298K and 1 atm for 2 ns in the NPT ensemble to obtain the average density. In the NPT simulations, the Nosé-Hoover thermostat and barostat were employed using the modified form proposed by Martyna, Tobias, and Klein, as implemented in LAMMPS[102], with a relaxation time constant of 0.1 ps and 1 ps for the thermostat and barostat, respectively. From these initial simulations, NVT simulations were generated with velocities reinitialized as described above. These simulations were first relaxed in the NVE ensemble with restrained atomic displacements of 0.1 Å per time step for 10 ps, then the simulation box was rescaled over 10 ps to obtain the correct average density, followed by a 1 ns NVT equilibration at 298K. The production data was obtained by extending these simulations for 10 ns. The thermodynamic data and coordinates were sampled at 0.1 ps and 1 ps, respectively.

To provide consistent parameterizations of each liquid, all atomistic force-fields in this study were derived from quantum chemistry calculations using the topology automated force-field interactions methodology (TAFFI)[148]. In brief, all force field terms were parameterized on the basis of density functional theory (DFT) quantum chemistry calculations, using the B3LYP-D3/def2-TZVP level of theory computed via the Orca software package [116].

The force fields were parameterized using the OPLS force-field functional form [77], except that 1-4 pairwise interactions were excluded in the non-bonded interaction computation. Bond, angle, and dihedral force-field terms were derived from potential energy curves computed for internal degrees of freedom for each molecule in vacuum, optimizing the other degrees of freedom as a function of the mode scan. The resulting energy curves were self-consistently fit to obtain the corresponding force-constant parameters and equilibrium displacement parameters in the force field. Partial charges were obtained from CHELPG calculations on molecular configurations sampled from the liquid, and Lennard-Jones parameters were parameterized on the basis of counterpoise-corrected interaction energies calculated for molecular pairs sampled from the liquid state.

The coarse-grained simulations used in IBI parameterization as well as to evaluate the thermodynamic and dynamic properties of each model were performed in the NVT ensemble. The same initial configuration was used for these simulations and was obtained by starting from a diffuse configuration of at least 1000 particles and equilibrating to the known atomistic density by the same NVT equilibration procedure as in the atomistic simulations. For butanol R2, R3, R4, R5 mappings and PEG 1002, 1002, 1000, 1000 and 1002 particles were simulated, respectively. Tabulated, DeePCG and a combination of tabulated and DeePCG potentials were used for CG simulations for the SP, ML and TL model, respectively. In our previous work[82], we found electrostatic interactions to have no effect on system physics and hence, were not included in the simulations. The pairwise tabulated interactions were excluded from 1-4 neighbors within a molecule, which experience tabulated bond, angle, and dihedral interactions instead. However, since DeePCG interactions are not pairwise, all particles within the chosen cutoff interact with each other.

7.B   IBI Procedure for Bonded Potentials

All IBI parameterizations were performed using a custom implementation developed by our group. The target distribution functions for bonds, angles, and dihedrals were sampled

from first 5ns of the atomistic trajectory and histogrammed with bin sizes of 0.1 Å, 1°, and 1° respectively. The distributions are obtained using equations,

$$P_b(r) = \frac{\delta(r - r_b)}{4\pi r^2}, \ P_a(\theta) = \frac{\delta(\theta - \theta_a)}{\sin\theta}, \ P_d(\phi) = \delta(\phi - \phi_d), \tag{7.5}$$

where the subscript "b", "a" and "d" indicate a specific type of bond, angle and dihedral, respectively and the value of an instance is given by r_b, θ_a and ϕ_d, respectively.

The initial guess potential for the IBI procedure for a intramolecular mode γ is obtained by a Boltzmann inversion of the corresponding distributions by,

$$U_\gamma(\kappa) = -k_B T \ln g_\gamma(\kappa), \tag{7.6}$$

where k_B and T are the Boltzmann constant and the simulation temperature, respectively. The potentials were extrapolated to a range of 0-20 Å, 0-180° and -180-180° for bonds, angles and dihedrals respectively. Quadratic extrapolation was utilized to obtain potential values outside the sampled regions for bonds and angles. Linear extrapolation was utilized to obtain potential values outside the sampled region for dihedrals. The guess potential for were corrected iteratively, according to the Eq. 7.7,

$$U_\gamma(r)^k = U_\gamma(r)^{k-1} + \alpha k_B T \ln \frac{P_\gamma(\kappa)^{k-1}}{P_\gamma(\kappa)^{\text{AA}}}, \tag{7.7}$$

where $U_\gamma(r)^k$ is the potential of mode γ at k^{th} iteration and P_γ^{AA} is the distribution obtained from atomistic trajectory.

For butanol R4 mapping, for each bonded IBI iterations, single-molecule CG simulations with shrink-wrap boundary conditions were performed to sample the intramolecular distributions. Whereas for all other mappings, we performed condensed phase simulations. We found that for butanol mappings R2 and R3, the potentials parameterized for single-molecule CG simulations were not able to reproduce the intramolecular distributions in the condensed phase and the effective potentials had to be parameterized from condensed phase simulations itself. The initial configuration for the condensed phase simulation was obtained by mapping a frame from the atomistic trajectory. For both single-molecule and condensed phase, CG

simulations were first relaxed in the NVE ensemble with restrained atomic displacements of 0.1 Å per time step for 10 ps, followed by a 0.1 ns NVT equilibration at 298K. The production data for sampling the intramolecular distributions were obtained by extending this trajectory for 10 ns and 0.5ns for single-molecule and condensed phase simulations, respectively and sampling configurations every 0.1 ps. The IBI procedure was run for 50 iterations, and the tabulated potentials from the iteration with the smallest distribution error (Eq. 7.8) were used for the subsequent non-bonded IBI parameterizations and CG property calculations. The average distribution error was given by

$$\varepsilon_{\text{IBI,b}} = \frac{1}{N_g} \sum_{\gamma}^{N_g} \frac{\sum_{\kappa} \| P_{\gamma}^{\text{AA}}(\kappa) - P_{\gamma}^{\text{CG}}(\kappa) \|}{\sum_{\kappa} P_{\gamma}^{\text{AA}}(\kappa)}, \tag{7.8}$$

where γ are different bonds, angles and dihedrals, κ are the values of these quantities, $P_{\gamma}(\kappa)$ is the value of distribution function of γ at κ and N_g is the total number of these interactions.

7.C IBI Procedure for Non-bonded Potentials

The target RDFs were sampled from first 5ns of the atomistic trajectory and histogrammed with a bin size of 0.01 Å and a range of 0-10 Å according to the equation,

$$g_{ab}(r) = \frac{V}{4\pi r^2 N_a N_b} \sum_{i \in a} \sum_{j \in b} \delta \left(r - r_{\text{ij}} \right) \tag{7.9}$$

where a and b are the types of the pair of CG particles, $V/N_a N_b$ corresponds to the density of pairs in an ideal gas, δ is the Dirac delta, r_{ij} is the distance between particles i and j, and dividing the summation by $4\pi r^2$ normalizes the density of pairs at each separation r. Pairs separated by less than four bonds are neglected in both the summation and denominator calculations.

The initial guess potential for each pair of CG particle type is the potential of mean force obtained by Boltzmann Inversion,

$$U_{ab}(r) = -k_B T \ln g_{ab}(r). \tag{7.10}$$

An exponential function was used to extrapolate the potential values for the inner region where the RDF was not sampled. The fitted potentials were further interpolated to a finer grid of 0.001 Å when utilized in CG simulations.

For the CG sampling necessary to iteratively refine the pair-wise interactions, a system containing at least 1000 beads was used for all mappings. This system was equilibrated for 1 ns in an NVT ensemble, as discussed in the Section 8.A, and the last frame was used as an initial configuration for the start of IBI procedure. For subsequent iterations, the last frame from the previous iteration was used as the initial configuration.

At each iteration of IBI, the correction term is obtained as in Eq. 7.11,

$$U_{ab}(r)^{\mathrm{k}} = U_{ab}(r)^{\mathrm{k}-1} + \alpha k_B T \ln \frac{g_{ab}(r)^{\mathrm{k}-1}}{g_{ab}(r)^{\mathrm{AA}}} + \beta A \left(1 - \frac{r}{r_{cut}}\right). \tag{7.11}$$

For the first 20 iterations of IBI, only the RDF correction was applied (i.e., $\beta = 0$ in Eq. 7.11). The potentials from the iteration with the lowest value of the objective function

$$\varepsilon_{\mathrm{IBI,nb}} = \frac{1}{\mathrm{N_P}} \sum_{\mathrm{p}=1}^{\mathrm{N_P}} \frac{\sum_{r=0}^{r_{cut}} \|g_{\mathrm{p}}^{\mathrm{AA}}(r) - g_{\mathrm{p}}^{\mathrm{CG}}(r)\|}{\sum_{r=0}^{r_{cut}} \|g_{\mathrm{p}}^{\mathrm{AA}}(r) + g_{\mathrm{p}}^{\mathrm{CG}}(r)\|} \tag{7.12}$$

were used to represent the SP model, i.e. the IBI without no pressure correction model. These potentials were also used for training and additional production simulations for the TL model, i.e. the transfer learning model based on IBI.

From the 20$^{\mathrm{th}}$ iteration onwards, both RDF and pressure corrections were applied starting with $\beta = 0.01$. After that, the weighing coefficient β was dynamically changed depending on the pressure error trend over the previous 5 iterations. For the first 20 iterations, each CG trajectory was of length 0.1 ns and sampled at 0.1 ps. After turning on the pressure correction, a longer trajectory of 1 ns with sampling of 0.1 ps was used to obtain accurate pressures.

The IBI procedure ran for a maximum of 80 iterations. The procedure was allowed to terminate earlier if the pressure error fell below 10 atm. From all iterations, the tabulated potentials from the iteration with the lowest value of the objective function

$$\varepsilon_{\text{IBI,nb}} = \frac{1}{N_p} \sum_{p=1}^{N_p} \frac{\sum_{r=0}^{r_{cut}} \|g_p^{\text{AA}}(r) - g_p^{\text{CG}}(r)\|}{\sum_{r=0}^{r_{cut}} \|g_p^{\text{AA}}(r) + g_p^{\text{CG}}(r)\|} + 0.05 * \frac{\|P^{\text{AA}} - P^{\text{CG}}\|}{P^{\text{AA}}} \qquad (7.13)$$

were retained and utilized in subsequent production simulations for the model SP_{PC} , i.e. IBI with pressure correction. These potentials were also used for the training and production simulations of the TL_{PC} model, i.e. the transfer learning model based on IBI with pressure correction. The first and second terms on the RHS of Eq. 7.13 indicate the normalized error in RDF ($g_p(r)$) and pressure (P), respectively. The weight of 0.05 was applied to the pressure errors due to their inherently higher variance than the RDF error.

7.D Training and testing of Neural Network

The blackbox (ML) and both transfer learning (TL) models consist of a fully connected feedforward neural network (NN). The former blackbox DeePCG framework decomposes the net energy of the configuration as a summation of energy of each bead. The later TL model instead models the remainder energy not captured in the IBI model. This bead-wise energy is an output of an NN and its input is the local neighborhood of the bead. The local neighborhood is captured in the form of a descriptor vector which consists of interbead distance information between bead and its short-distance (sel_a) and long-distance (sel_r) neighbors within a cut-off radius of R_c. For the short-distance neighbor j of bead i, the vector contains both interbead radial and angular information given by 4 components– $1/R_{\text{ij}}, x_{\text{ij}}/R_{\text{ij}}^2, y_{\text{ij}}/R_{\text{ij}}^2, z_{\text{ij}}/R_{\text{ij}}^2$. The net distance between the two beads is indicatd by R and x, y and z are the components of the interbead vector. The interbead vector is defined in a local axes frame of the bead determined by the bead and its two nearest neighbors of given bead types. Whereas for the long-distance neighbors, only the inverse of net distance $1/R_{\text{ij}}$ is supplied. The user supplies the above inputs of the cut-off radius R_c, the number of

short-distance sel_a and long-distance sel_r neighbors for each bead type, and the beads to be used for determining the local axes frame for each bead type (axis_rule).

For the butanol and PEG mappings, a cutoff radius of 10 and 12 was used, respectively. The "axis_rule", i.e. the rules used for determining the local axes for each CG representation are given in Table 7.1. For each bead type, a total of six values, with three value per axes must be listed. Of the three values per axes, the first entry describes if both radial and angular information must be included, and is always set to 0, i.e. both radial and angular information is included. The second and third entry are used for selecting the beads used for determining the axes, and indicate the bead type and index of the neighbor (0 meaning first neighbor and so on), respectively. For butanol mappings R2 and R3, which have 3 beads per molecule, the two intramolecular neighboring beads were used to determine the local axes as a natural choice. The type of the first and second neighbor in the local axes frames are determined based on the distance between the bead and the neighbors. For instance, for R3 mapping, for bead the CH2-CH2-CH2, although both CH3 and OH are immediate neighbors, OH is placed as the first neighbor as it is closer than CH3. For R4 mapping, first neighbor for each type is the intramolecular neighboring bead whereas the type of the second neighbor was similarly chosen to be the type of closest bead. Hence, OH bead was chosen as the second neighbor for both OH and CH3-CH2-CH2-CH2 beads. For R5, the choice is straightforward to be the two nearest neighbors as there is only one type of bead.

For PEG M1 mapping, immediate neighbors were chosen for all bead types with one exception. For M1 mapping, for CH2-O-CH2 beads, the first and second neighbors in the axis_rule are chosen to be first and second neighbors of bead type CH2-O-CH2. For the middle CH2-O-CH2 beads (CH2-O-CH2$_m$), this corresponds to its immediate CH2-O-CH2 neighbors on opposite sides (Figure 7.4a). But for the CH2-O-CH2 beads closer to the end group (CH2-O-CH2$_e$), this corresponds to the CH2-O-CH2 neighbors on the same side (Figure 7.4b). However, despite this distinction both bead types have similar local axes frame as the first axis is formed by the bead of interest and its nearest CH2-O-CH2 bead and the second axis is given by the normal to the plane formed by the three consecutive CH2-O-CH2 beads. However, to make sure this definition of axis_rule doesn't affect the system physics, we also experimented with M2 mapping (Figure 7.4c) where the middle and

end CH2-O-CH2 group are treated as different bead types with separate axes rules, but the overall results obtained were very similar to M1 mapping. The the cutoff radius (R_c) and axis_rule were held fixed but the number of short-distance and long-distance neighbors were treated as tunable hyperparameters in the model training. Additional hyperparameters involve the number of configurations per batch (batch_size), number of batches used in training (stop_batch), and the neural network structure (n_neuron). The learning rate for the neural network starts with learning rate (start_lr) and decays exponentially with a rate of (decay_rate) on the scale of (decay_steps).
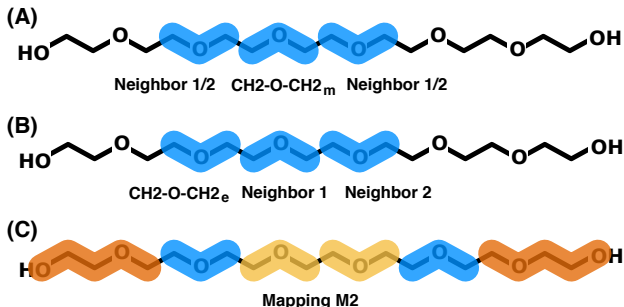
Figure 7.4. The difference in the neighbors used to determine local axes rule for bead of type CH2-O-CH2 depending on its position along the chain, i.e. if it is a (a) middle bead (CH2-O-CH2$_m$) vs. (b) closer to chain end (CH2-O-CH2$_e$). (c) A new mapping which treats (CH2-O-CH2$_m$) and (CH2-O-CH2$_e$) beads as different bead types.

The ML model is trained for the net force on each bead, while the TL models are trained for the error in forces on each bead for the potentials parameterized using IBI. For all the models, a total of 5000 configurations sampled every 1ps from a 5ns production run were used. In the training process, we start with varying the hyperparameters sel_a and sel_r, i.e., the number of short-distance and long-distance neighbors as these parameters affect the physics significantly. For a molecule with N CG beads, we start sel_a from 2N-1, i.e. its intramolecular neighbors (N-1) and beads in closest neighboring molecule (N), and sel_r from zero, i.e. only information about short-distance neighbors is supplied. Both sel_a and sel_r are incremented by total number of beads in the molecule N, i.e. adding a new molecule as neighbor. For instance for butanol mapping R3, with N=3, we vary sel_a as {5 ,8 ,11 ...} and sel_r as {0, 3, 6 ...}. While training, we try all combinations of sel_a and sel_r, i.e. {(5,

Table 7.1. Rules for determining local axes of a given bead type. For each CG mapping, the second column contains the "axis_rule" in the format is as entered in the input json file of the DeePCG module. The corresponding constituent atoms and the type number for each bead type (as supplied to the DeePCG module) are given in the third column for reference.

CG mapping	axis_rule	Bead types
butanol R2	[0, 1, 0, 0, 2, 0, 0, 2, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0]	CH3-CH2: 0, $(CH2)_2$: 1, OH: 2
butanol R3	[0, 1, 0, 0, 2, 0, 0, 2, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0]	CH3: 0, CH3-$(CH2)_3$: 1, OH: 2
butanol R4	[0, 1, 0, 0, 1, 1, 0, 0, 0, 0, 1, 0]	CH3-$(CH2)_3$: 0, OH: 1
butanol R5	[0, 0, 0, 0, 0, 1]	CH3-$(CH2)_3$-OH: 0
PEG M1	[0, 1, 0, 0, 1, 1, 0, 1, 0, 0, 1, 1]	OH-CH2-CH2-O-CH2: 0, CH2-O-CH2: 1

0), (5, 3), ... (8, 0), (8, 3), ...}. We start the training with only two values of sel_a and four values of sel_r, and add more molecules till a local minima is obtained. For mapping PEG M1, we started with N-1 number of sel_a neighbors. During this optimization for number of neighbor, other hyperparameters were held fixed to the values reported in Zhang et al or in the DeePCG tutorial (Table 7.2).

Table 7.2. The values of hyperparameter held constant while optimizing number of neighbors, as entered in the DeePCG input file.

Argument	Value
stop_batch	100000
batch_size	4
start_lr	0.0001
decay_steps	5000
decay_rate	0.95
n_neuron	[120, 60, 30, 15]

After obtaining the optimal number of neighbors, we experimented further with three hyperparameters of the starting learning rate (start_lr), the number of batches used in training (stop_batch), and the number of layers and nodes in the neural network (n_neuron). The following additional hyperparameters were tested– a lower and higher learning rate (start_lr =

164

$0.001, 0.00001$), additional training data (stop_batch $= 500000$) and neural networks with varying number of layers and nodes n_neuron $= [60, 30, 15], [120, 60, 30], [240, 120, 60, 30, 15], [160, 80, 40, 20], [80, 40, 20, 10]$).

For each set of hyperparameters, 3 models were trained with different initialization. After training, a short MD simulation of 0.5 ns was run and the last 0.25ns configurations sampled at 1ps were used to parse the radial distribution function (RDF) of all pair types in the system. The RDF was calculated for each CG pair according to the Eq. 7.9 but all pairs even 1-4 intramolecular neighbors were included contrary to the description in Section 8.C. The net error in the RDF ($\varepsilon_{\mathrm{RDF}}$) was averaged over RDF errors for $N_t = 3$ models as,

$$\varepsilon_{\mathrm{RDF}} = \frac{1}{N_t} \sum_{i}^{N_t} \frac{1}{N_p} \sum_{p=1}^{N_p} \frac{\sum_{r=0}^{r_{cut}} \exp(g(r)_{i,p}^{\mathrm{AA}}) \| g(r)_{i,p}^{\mathrm{AA}} - g(r)_{i,p}^{\mathrm{CG}} \|}{\sum_{r=0}^{r_{cut}} \exp(g(r)_{i,p}^{\mathrm{AA}}) \| g(r)_{i,p}^{\mathrm{AA}} \|}. \tag{7.14}$$

The RDF error for a model i is calculated as an average over errors for all pair types (p), given by the difference in the RDF in the atomistic representation ($g(r)_{i,p}^{\mathrm{AA}}$), and the RDF in the coarse-grained representation ($g(r)_{i,p}^{\mathrm{CG}}$), and the summations run from $r = 0$ to the radial cutoff for the pairwise interactions, $r_{cut} = 10$. To avoid having the poorly sampled regions dominate the error measure, the difference in RDFs is weighted by $\exp g(r)_{i,p}^{\mathrm{AA}}$ and the error at each r is included in the summation only if $g(r)_{i,p}^{\mathrm{AA}} > 0.1$. In the denominator, the error is normalized by the average value of atomistic RDFs over radial distances, with the same exponential weighting.

This RDF error was used as a testing score to select the best set of hyperaparameters throughout the hyperaparameter optimization process. Generally, for tuning the hyperparameters of a force-based ML model, the test loss function is the force error over new configurations not used in training, which tests the ability of model to capture the potential energy surface for unseen configurations. Here, instead, we test the ability of the model to predict structural features which are not directly seen through the training data. In doing so, we select the hyperparamaters with highest transferability for the purely ML model for a fairer comparison with the TL models which already have the structural information. The final hyperparameters for the best model for each mapping are as below,

Table 7.3. For each CG mapping and for the ML and TL models, the second column contains the optimized hyperparameters in the format as entered in the input json file of the DeePCG module.

CG mapping	Model	Hyperparameters
butanol R2	ML	sel_a : 5, sel_r : 6, start_lr : E-4, stop_batch : 500000, n_neuron : $[120, 60, 30, 15]$
	TL	sel_a : 5, sel_r : 3, start_lr : E-3, stop_batch : 100000, n_neuron : $[120, 60, 30, 15]$
	TL$_{PC}$	sel_a : 5, sel_r : 9, start_lr : E-3, stop_batch : 100000, n_neuron : $[120, 60, 30, 15]$
butanol R3	ML	sel_a : 5, sel_r : 9, start_lr : E-4, stop_batch : 100000, n_neuron : $[240, 120, 60, 30, 15]$
	TL	sel_a : 5, sel_r : 6, start_lr : E-4, stop_batch : 100000, n_neuron : $[240, 120, 60, 30, 15]$
	TL$_{PC}$	sel_a : 5, sel_r : 6, start_lr : E-4, stop_batch : 100000, n_neuron : $[240, 120, 60, 30, 15]$
butanol R4	ML	sel_a : 3, sel_r : 10, start_lr : E-4, stop_batch : 100000, n_neuron : $[240, 120, 60, 30, 15]$
	TL	sel_a : 3, sel_r : 0, start_lr : E-4, stop_batch : 100000, n_neuron : $[60, 30, 15]$
	TL$_{PC}$	sel_a : 3, sel_r : 0, start_lr : E-4, stop_batch : 100000, n_neuron : $[120, 60, 30, 15]$
butanol R5	ML	sel_a : 12, sel_r : 16, start_lr : E-4, stop_batch : 500000, n_neuron : $[120, 60, 30, 15]$
	TL	sel_a : 16, sel_r : 4, start_lr : E-4, stop_batch : 500000, n_neuron : $[120, 60, 30, 15]$
	TL$_{PC}$	sel_a : 8, sel_r : 16, start_lr : E-4, stop_batch : 100000, n_neuron : $[240, 120, 60, 30, 15]$
PEG M1	ML	sel_a : 11, sel_r : 12, start_lr : E-4, stop_batch : 100000, n_neuron : $[120, 60, 30]$
	TL	sel_a : 5, sel_r : 12, start_lr : E-4, stop_batch : 100000, n_neuron : $[240, 120, 60, 30, 15]$
	TL$_{PC}$	sel_a : 5, sel_r : 0, start_lr : E-4, stop_batch : 100000, n_neuron : $[240, 120, 60, 30, 15]$

7.E   Property Calculation

The performance of the different models for various properties namely RDF, pressure, force, diffusion coefficient and molecular enthalpy of vaporization, was evaluated as discussed above. In order to parse these properties (except for forces), for all the SP models, 3

independent MD simulations were performed. For the ML and TL models, 3 independent simulations were obtained by performing an MD simulation for the three independently trained models. The initial configuration for the MD simulation was obtained from the pre-equilibrated trajectories as described in Section 8.A. The velocities were reinitialized for each independent simulation, followed by a 10 ps NVE simulation with constrained atomic displacements of 0.1 Å per time step and a 1 ns equilibration in the NVT ensemble. Finally, production runs of 1 ns in NVT ensemble were performed with coordinate and thermodynamic data sampled at 1 ps frequency.

The net error in the RDF (Eq. 7.14) was calculated as described in Section 8.D but for a longer production run of 1ns.

The net force error is the root mean square error of the force over N_k configurations and N_j beads, given by

$$\varepsilon_F = \sqrt{\frac{1}{N_k} \sum_k^{N_k} \frac{1}{N_j} \sum_j^{N_j} \left\| \mathbf{F}_{k,j}^{CG} - \mathbf{F}_{k,j}^{AA} \right\|^2}, \tag{7.15}$$

where $\mathbf{F}_{k,j}^{CG}$ and $\mathbf{F}_{k,j}^{AA}$ are the forces on bead j in k configuration as given by the coarse-grained model and atomistic mapping, respectively. This error is calculated over 1000 configurations obtained from mapping the last 1ns of the atomistic production run.

The pressure was obtained directly from an average over the 1ns of LAMMPS thermodynamic output. The fractional error in pressure was calculated according to

$$\varepsilon_P = \frac{1}{N_t} \sum_i^{N_t} \frac{P_i^{CG} - P_i^{AA}}{P_i^{AA}} \tag{7.16}$$

where N_t is the number of independent trajectories.

The self-diffusion coefficient was obtained from the mean square displacement of each molecule using the Einstein relation

$$D = \lim_{t \to \infty} \frac{1}{6} \frac{d}{dt} \langle \| \mathbf{r}_m(t) - \mathbf{r}_m(0) \|^2 \rangle, \tag{7.17}$$

167

where $\mathbf{r}_m(t)$ is the position vector of the molecule center of mass at time t. The angular brackets in Eq. 7.17 indicate an average over all the molecules and time origins. The fractional error in the diffusion coefficient was calculated according to

$$\varepsilon_D = \frac{1}{N_t} \sum_i^{N_t} \frac{D_i^{CG} - D_i^{AA}}{D_i^{AA}} \tag{7.18}$$

where N_t is the number of independent trajectories. For each trajectory, the numerical derivative of the first continuous 10 ps interval after 100 ps which exhibited a linear slope (0.95- 1.05) of log(MSD) versus time were used to obtain the diffusion coefficient.

The molar enthalpy of vaporization, ΔH_v, is calculated by

$$\Delta H_v = U_v + k_B T - U_c, \tag{7.19}$$

where U_v and U_c are the mean molar potential energies in the gas phase and condensed phase, respectively. U_v was calculated by single-molecule MD simulation. The net fractional error in the enthalpy of vaporization was calculated according to

$$\varepsilon_{H_v} = \frac{1}{N_t} \sum_i^{N_t} \frac{\Delta H_{v_i}^{CG} - \Delta H_{v_i}^{AA}}{\Delta H_{v_i}^{AA}} \tag{7.20}$$

where N_t is the number of trajectories. Similar to the condensed phase, three independent gas phase simulations were performed by starting with 10 ps of NVE simulation with displacement constrained to 0.1 , followed by a 1 ns equilibration and finally, a production run of 1ns. The energy U_v was calculated by averaging the potential energy over the 1 ns trajectory sampled at every ps.

7.F   Molecule Breaking in Butanol R4 and PEG M1 Mapping

As discussed above, for butanol R4 and PEG M1 Mapping, the atoms in the molecule drift apart and move closer to new atoms forming a pseudo molecule. Thus, the bonds between atoms in the originally defined molecule stretch to large lengths as the simulation progresses (Figure 7.5(a-b)). However, overall, the structure remains similar as the original intramolecular neighbors are replaced by new atoms. On closer inspection, we find this results from overlap in the length-scale of a bond and first shell of intermolecular neighbors (Figure 7.5(a-b)). The pair of intermolecular neighbors behave as bonded beads and exhibit constraining force on each other, while the actual bonded bead drifts apart. The overlap in length-scale is very apparent in the 0-1 RDF distribution (Figure 7.5b) for PEG M1 mapping, where the 0-1 bond RDF and 0-1 intermolcular RDF show a large extent of overlap between 3-5 Å. The maximum separation between bonded beads ($r_{intra} = 4.4$ Å) is higher than the minimum separation between intermolecular beads ($r_{inter} = 3.2$ Å). Smaller yet evident overlap is observed for 0-1 RDF for butanol R4 mapping and 1-1 RDF for PEG M1 mapping. The bond and intermolecular RDF were calculated using Eq.7.9, where only pairs of beads forming 0-1 bond and pairs of 0-1 beads not in same molecule were counted for the numerator, respectively; whereas for the all 0-1 RDF all 0-1 pairs were included. The denominator was same across all three RDFs, i.e. total number of all 0-1 pairs was used, to provide the same normalization. The other mappings of R2 and R3 do not exhibit this phenomena due to disparate length-scales of bonded and intermolecular RDF as seen from 0-1 and 1-2 RDFs in Figure 7.5c for R2 mapping.
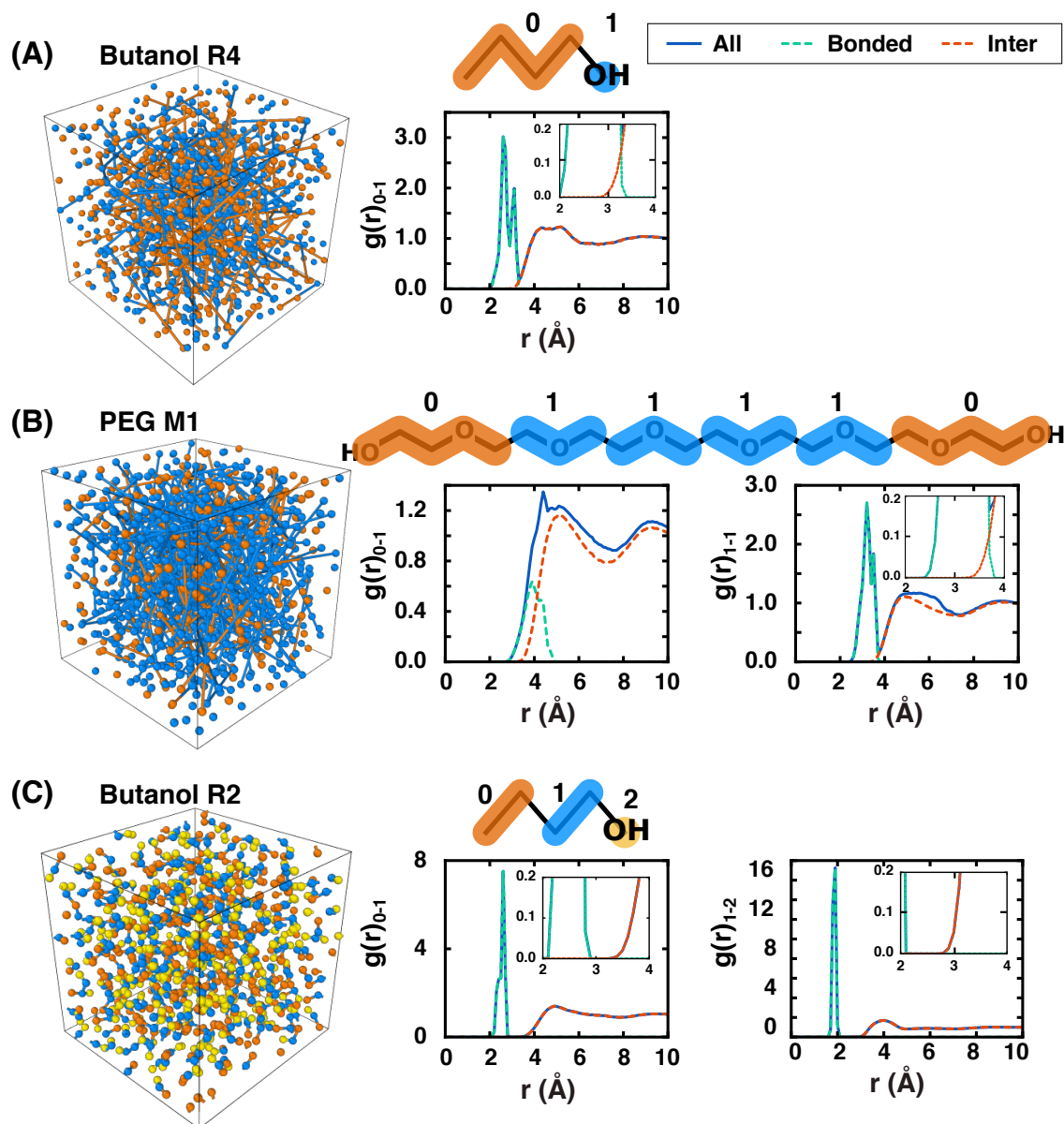
Figure 7.5. Configuration snapshot (left) and RDFs (right) for mappings which experience breaking of molecules, i.e. (a) butanol R4 and (b) PEG M1 as well as mapping which doesn't, i.e. (c) butanol R2.

7.G   Additional Results Referenced in Main Text

Error for different properties for butanol R2 and R4 mappings discussed above are reported in Figure 7.6
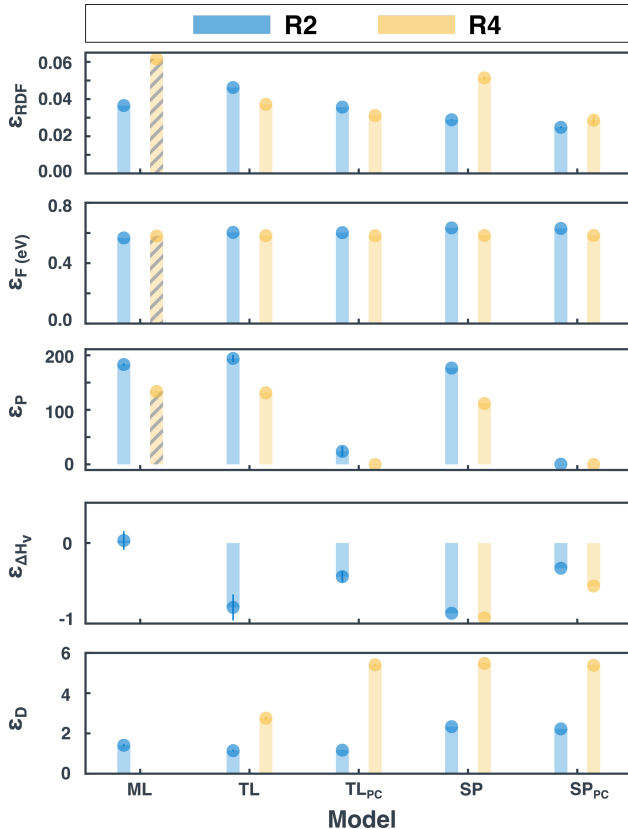
Figure 7.6. Error for RDF, force, pressure, ethalpy of vaporization, and self-diffusion coefficient (from top to bottom), for butanol R2 (pink) and R4 mapping(yellow). The bar for R4 mapping is hatched to caution the reader that the simulation shows unphysical breaking and formation of bonds. The standard errors are denoted by the thin line passing through the marker. All errors except that for forces are normalized.

The error in forces is slightly lower in ML and TL models compared to SP models, due to inclusion of additional many-body interactions, as seen from Figure 7.7. However, this reduction in error is insignificant and the error in force is similar across different models across all mappings as discussed above and reported in Figure 7.6 and Figure 7.3.

As discussed above, we find two main trends for butanol mappings. Intermolecular structural distributions are well reproduced by SP models compared to ML models, whereas
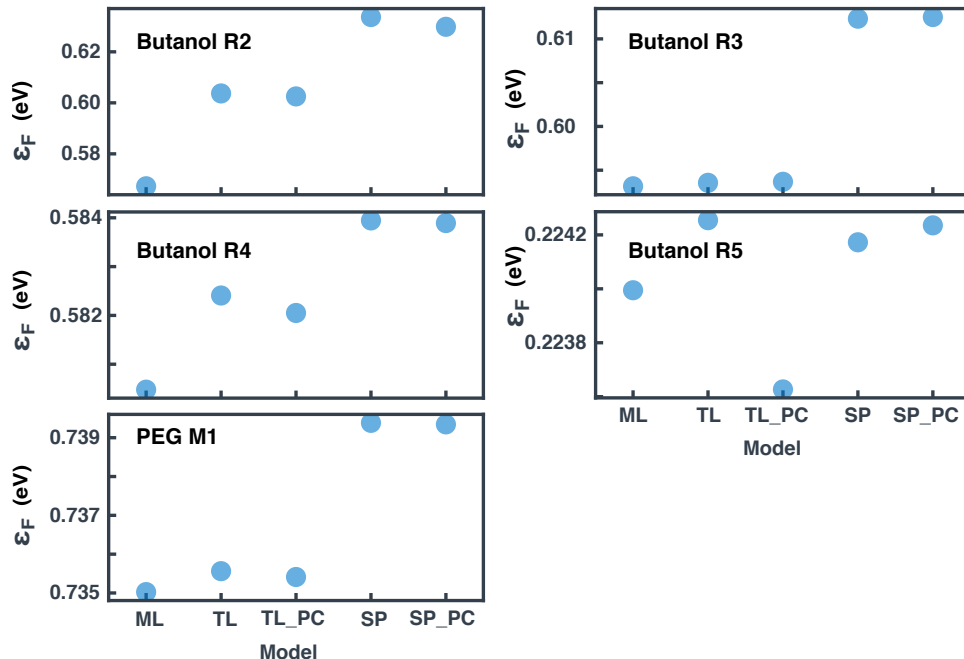
Figure 7.7. Error in forces across different models for all 5 CG mappings simulated in this work.

intramolecular structural distributions are well reproduced by ML models compared to SP models. As an example of this, RDFs for all pair types for butanol R3 mappings are reported below in Figure 7.8. As seen from RDFs for pairs 0-2, 1-2 and 2-2, the intermolecular structural distributions are better reproduced by the SP and SP_{PC} models (as well as TL models) at distance 6-8 , 4-6 and 4-6 , respectively. The intramolecular structural distributions are however better reproduced by the ML (as well as TL models) for pairs of type 0-2 between 3-5 . SP models fail to reproduce the intramolecular radial features despite reproducing the bond and angle distributions very well (bottom row of Figure 7.8).

However, all the differences are very minute and mostly for RDFs involving OH bead. OH bead is a very fine mapping than practically used CG models. For the more practical mapping of R5, we find all models capture the structural distributions perfectly (Figure 7.9).

For PEG M1 mapping, the differences between TL and SP models are more significant than butanol mappings. We find the TL models perform poorly as compared to the SP models as seen below (Figure 7.10).

Figure 7.8. Comparison of RDFs across all models with respect to target atomistic (AA) RDF, are reported for all pair types for butanol R3 mappings, in the top two rows. The intramolecular distributions for R3 mappings are reported in the bottom row.

Figure 7.9. Comparison of RDFs across all models with respect to target atomistic (AA) RDF, are reported for all pair types for butanol R5 mappings.

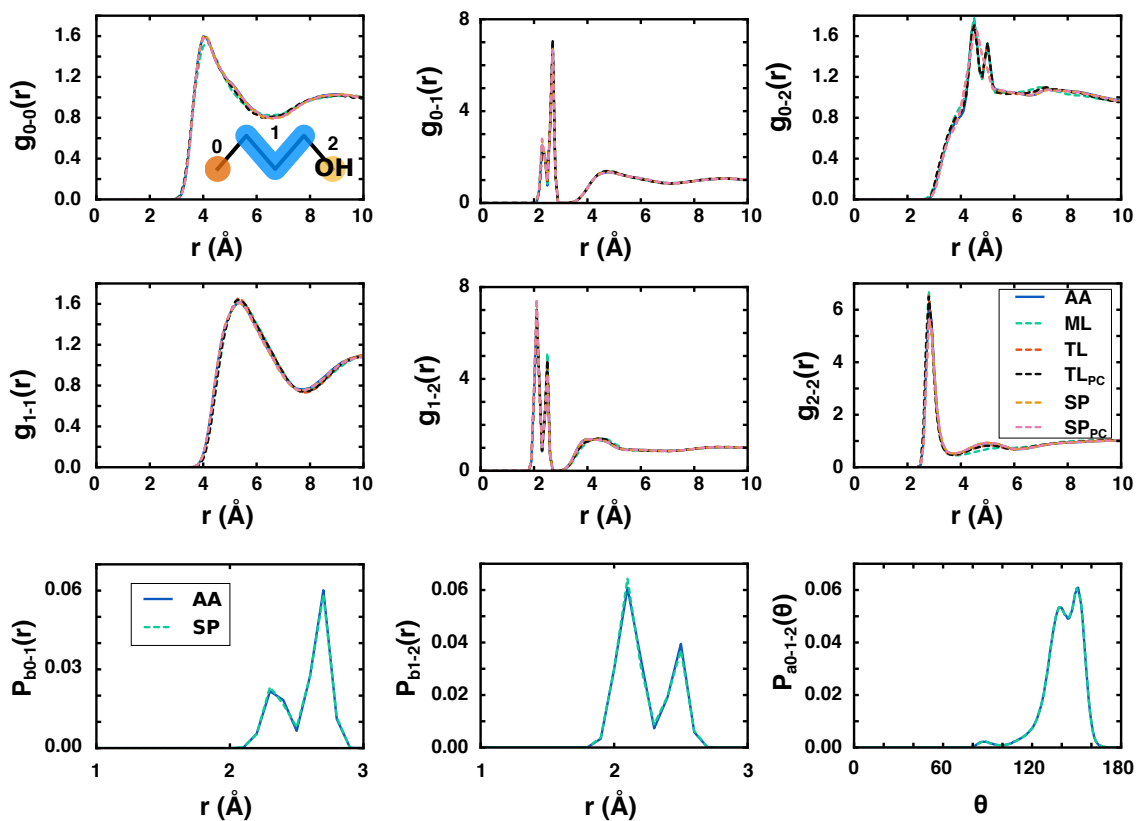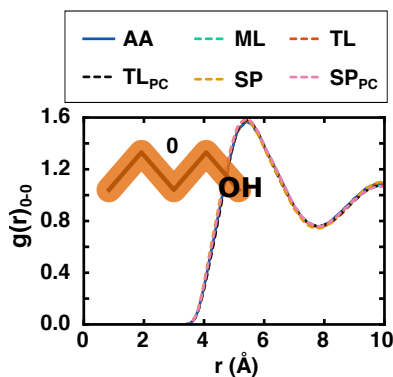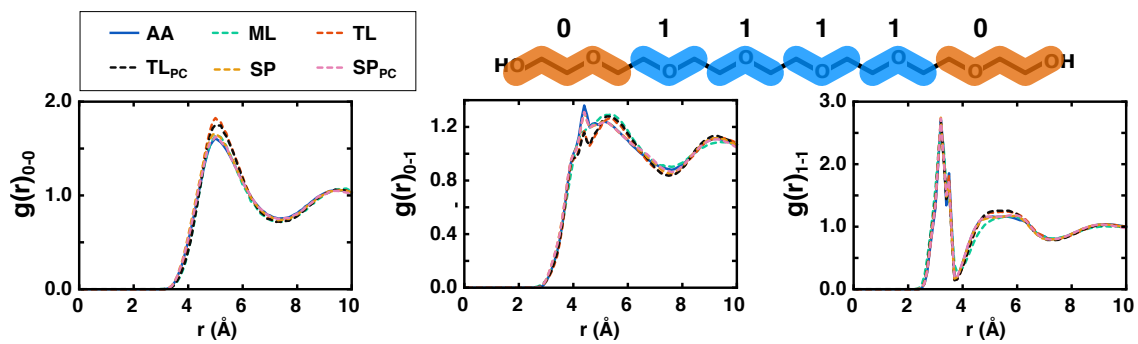**Figure 7.10.** Comparison of RDFs across all models with respect to target atomistic (AA) RDF, are reported for all pair types for PEG M1 mappings.

8. OUTLOOK

Organic mixed ionic-electronic conductors bring a new wave of applications with their unique ability to conduct ions and electrons. In the last decade or so, researchers have worked hard to extend the possible applications of these polymers and in the near future, we can expect further discoveries both in applications and design of novel mixed conductors. In order to achieve this goal, we need combined efforts not only in testing new material candidates but in understanding the coupling between ionic and electronic conducting channels, and complex interplay between morphology and electrostatic interactions. Simulation tools, specifically, coarse-grained simulations offer a great advantage as they can characterize the mesoscale features of these disordered systems inaccessible to experimental techniques while accounting for chemical features at the molecular scale. The simulation studies have been limited for this class of polymers and primarily targeted to a few polymer candidates.

To progress the field further, we needed a method which can model the class of mixed conducting polymers and capture the features unique to them. Hence, through this thesis, we have made efforts to build such a general framework for the class of conjugated mixed conducting polymers. The model captures features unique to mixed conductors such as π-π stacking, torsional disorder in polymers, intramolecular charge delocalization, intermolecular hopping transport, and electrolyte-polaron electrostatic interactions. The model is also designed for a family of mixed conductors and can be used to test effect of a wide range of diverse molecular variables on the OMIEC physics.

Accordingly, we have implemented the model to perform the first ever systematic computational investigation of the effect of side-chain hydrophilicity and block patterning of side-chains on the OMIEC physics. In the hydrophilic study, the polymer with all hydrophilic side-chains showed the highest dopant percolation as well as charge mobility. The mixed side-chain configurations with both hydrophilic and hydrophobic side-chains showed frustrated packing due to the competing interest of the two side-chains. We found two factors essential for good charge conduction– well connected and torsionally aligned polymers, and absence of trap sites. We also reconfirmed the expected effect of the surrounding dielectric–

a higher dielectric screens the electrostatic interactions but also increases the reorganization energy associated with polaron hopping.

In the block-chain patterning study, we discovered advantages to adding separate larger blocks of hydrophilic and hydrophobic side-chains. The dedicated block of hydrophilic side-chains interfaces with electrolyte improving the polymer-electrolyte interactions, stacking of polymer and connections through the polymer network. Although it does not manifest into significantly high mobility, it does show higher number of electronic pathways per polaron sites. We also find that the effective polarity of side-chains is a function of its local neighborhood and can alter the swelling and morphology of the polymer network. These findings suggest some near term strategies which can be tested experimentally to tune the mixed conductors further.

An obvious future extension of this thesis is continuing the exploration of side-chain design space such as tuning the hydrophilicity of side-chains, their frequency of occurrence and their length and sterics. The side-chain patterning study showed that there is an advantage to dedicated blocks of polar side-chains for interfacing with solvent and maintaining connectivity. This finding suggests a possibility of tuning this compartmentalization further by adopting side-chains of different degrees of polarity. Can we obtain sufficient electrolyte percolation with fewer but extremely hydrophilic side-chains? And further can we place these hydrophilic chains more sparsely along the backbone to achieve the same swelling without disturbing the polymer packing due to sterical hindrance between side-chains. Hence, a dedicated study of side-chains of different polarities placed at varying frequency along the backbone would be a worthwhile effort.

Experimental works have shown that the presence of optimal length of side-chains exists for all polar side-chain mixed conductor, but the length can further be tuned for other interesting side-chain configurations as well. For instance, tuning the length of individual side-chains composed of both polar and apolar beads would make for an interesting study. While we have studied the bifurcation of polar and apolar moieties along the backbone, many experimental studies have shown an advantage to have this bifurcation along the side-chain. That is, have an alkyl linker between the backbone bead and polar moieties. The polar moieties at the end of the side-chain facilitate electrolyte percolation but just till a distance

176

sufficient to dope the polymer but not disturb the polymer packing. The mechanistic details of this side-chain design can be investigated through simulations. Further, we can see the extent to which the electrolyte percolation and packing is tuned by systematically varying the respective length of each of these blocks along with the degree of their hydrophilicity. Would we need fewer(shorter) polar moieties if the polar moiety is highly hydrophilic to achieve the same extent of swelling with better polymer packing? In addition to the side-chains, the backbone and dopant chemistry can also be explored as well as effect of additives on the OMIEC physics can be tested. Essentially, we have a model like LEGO blocks, and we can continue to create new polymer designs with these blocks to conduct innovative hypothesis driven studies and expand the bounds of OMIEC design.

In terms of methodology, the top-down framework developed in this work can be applied not just to mixed conductors but also to other fields where conjugated molecules and polymers may be of interest. In the context of developing the methodology further, the most crucial improvement would be enabling the use of graphic processing units (GPUs) in LAMMPS software for this ellipsoid particle systems. The true power of CG models in computational speed-up is not tapped currently and the GPUs can offer over 3 times speed-up over GPUs. In terms of the analysis itself, this framework can be easily extended to mimic the movement of solvent in and out of the polymer bulk and employed to test the stability of various mixed conductors which is a major concern. The simulation procedure proposed in this study is highly representative of experimental procedure to test this polymers, i.e. first the polymer film is formed by processing with a solvent, and then the polymer film is charged and interfaced with an electrolyte. Hence, we can also investigate the effect of processing solvent and procedure on the polymer arrangement, and resultant electrolyte percolation and performance as an OMIEC.

The top-down framework not only includes a chemically flexible coarse-grained model but also a detailed kinetic Monte Carlo framework for charge transport which is a culmination of organic electronics literature. This framework can be employed for a qualitative study of other organic electronic materials as well. While we fix the system relevant parameters to reasonable estimates from literature, the framework can be further enriched by interfacing it with polymer specific values obtained from systematic electronic level studies. For instance,

177

the polaron HOMO energy levels and reorganization energies used in this study are a function of oligomer length, as described a fundamental study by Zade et al.[182] Similarly, the HOMO energies, for example, could be encoded as a function of the side-chain identity to account for the hole side-chain interactions. Additionally, mixed conductors are a heterogeneous systems, and the electrostatic interactions play a critical role such as trapping of hole due to dopants, interactions between polarons, screening of some of these interactions due to the presence of polar solvent, etc. Right now, we have modeled the surrounding in a mean field fashion with an intermediate dielectric constant representative of the mixed polar and apolar phases, but a more rigorous treatment of the dielectric can serve fruitful in future.

While the top-down approach provides qualitative trends characterizing OMIEC systems, a more quantitative picture can be obtained with coarse-grained modeled developed in a bottom-up fashion, i.e. by ensuring uniform standards of the target information using atomistic MD simulations. As it stands, the design of materials is being continuously expedited by using novel approaches to generate, process and understand massive amounts of data. In the future, we can hope to see material design in an automated and high-throughput fashion for various applications. In order to achieve this vision, we must have systematic computational tools to model and characterize the materials, and bottom-up coarse-grained methods can serve as an important molecular tool. While there are many approaches to bottom-up parameterization of CG models and new methods are being proposed in the field, the community is still trying to quantify and understand the inherent limitations of these methods and coarse-graining itself.

In that regard, our first study was to systematically investigate the effect of mapping operator, i.e. the number of coarse-grained beads, and the constituents of these beads. We also studied this in the context of three molecules with diverse chemistry and two common CG bottom-up parameterization schemes. We found that the current CG models perform similarly across all mappings from high to low resolution, in the context of on-target properties. That is, if given the information in the parameterization procedure, even the least complex CG representations capture the information very well. On the contrary, for properties not seen in the parameterization scheme, we see decline in performance with resolution of the coarse-grained model. Moreover we see that qualitatively different potential functions

reproduce the on-target properties equally well and hence, there is leeway to incorporate additional information.

In a subsequent study, we investigated the performance of physics motivated simpler potential energy surfaces (PES) vs. those based on machine learning (ML). In the cases studies, the increased complexity offered by ML potentials did not result in any major improvements in the performance of CG model. Rather, we found erroneous behavior exhibited by the ML model such as breaking of bonds in the system. We proposed a new approach called the transfer learning approach which potentially rectifies the shortcomings of ML model. We use the physics based approach for the better chunk of the PES and use ML model only as a Δ-correction to the PES, i.e. to capture the atomistic interactions (if any) not captured by the physics based approach. This hybrid approach contains key physics due to the physics based model avoiding the erroneous behavior by the ML-only model and additionally corrects it by incorporating the complexity and (minimal) advantages provided by the ML model.

We found in our first study that essentially, the coarse-grained models are information limited, i.e. the models can capture more information than what is being supplied to these models while training. Naturally, the extent to which information can be incorporated without running into representability limitations must be tested. In order to achieve this, we had adopted the derivative free optimization technique of particle swarm optimization (PSO) to fit for structural as well as thermodynamic properties, but found that structural features were a difficult property to be optimized using PSO. But in the light of success of our transfer learning approach, a new possibility surfaces. We can use Iterative Boltzmann Inversion (IBI) to obtain the first approximation to the PES as it reproduces the structural features quite well. And then we can use PSO like optimization scheme to add a correction potential which reproduces additional thermodynamic properties. An inverse approach of first obtaining PES which reproduce thermodynamic properties (using PSO or even top-down approach) and using this PES as an initial guess to IBI instead of the Boltzmann inverted potential is also worth exploring.

In our second study, we found that transfer learning from the physics based model improves the ML model by accounting for missing physics, but the use of ML model doesn't bring significant improvement over the traditional approach. Atomistic forces, which is the

property corrected using the ML framework, are only minimally improved and slightly better performance in diffusion coefficient is observed due to inclusion of additional interactions. The comparable errors in atomistic forces across all models indicate that we might be hitting the limit of average forces which can be captured and the errors seen are associated with missing degrees of freedom. First, this claim must be tested by designing a study where the atomistic configurations are sampled such that true average force is obtained. The assumption is that with enough number of atomistic configurtions used in ML training, we are obtaining the true average force but whether this is true has not been tested. If we prove that the coarse-grained models are reproducing the average forces perfectly, the missing forces are due to inherent loss of degrees of freedom and must be introduced as additional frictional and random forces. While the framework for doing so exists, based on Mori-Zwanzig, the final implementation is often simplified with approximations to the memory kernel. One possible way of accounting for the frictional forces is using probabilistic frameworks such as Gaussian Process Regression and dropout neural networks.

Overall, the thesis makes contributions to coarse-grained methods development in the context of both top-down and bottom-up approaches. The top-down model is constructed in the context of OMIEC and has provided some useful insights on side-chain design. We believe this model can be employed for similar design studies for advancing the understanding of OMIEC systems and other conjugated systems. On the bottom-up approach front, the study performed by us was the first attempt at incorporating transfer learning in the context of coarse-grained models and we hope that similar schemes are explored in the future to improve bottom-up coarse-grained schemes.

REFERENCES

[1] ABRAHAM, M. H., CHADHA, H. S., WHITING, G. S., AND MITCHELL, R. C. Hydrogen bonding. 32. an analysis of water-octanol and water-alkane partitioning and the δlog p parameter of seiler. *Journal of pharmaceutical sciences 83*, 8 (1994), 1085–1100.

[2] AGRAWAL, V., ARYA, G., AND OSWALD, J. Simultaneous iterative boltzmann inversion for coarse-graining of polyurea. *Macromolecules 47*, 10 (2014), 3378–3389.

[3] ALESSANDRI, R., GRÜNEWALD, F., AND MARRINK, S. J. The martini model in materials science. *arXiv preprint arXiv:2012.07194* (2020).

[4] ALESSANDRI, R., SAMI, S., BARNOUD, J., DE VRIES, A. H., MARRINK, S. J., AND HAVENITH, R. W. Resolving donor–acceptor interfaces and charge carrier energy levels of organic semiconductors with polar side chains. *Advanced Functional Materials 30*, 46 (2020), 2004799.

[5] ALESSANDRI, R., UUSITALO, J. J., DE VRIES, A. H., HAVENITH, R. W., AND MARRINK, S. J. Bulk heterojunction morphologies with atomistic resolution from coarse-grain solvent evaporation simulations. *Journal of the American Chemical Society 139*, 10 (2017), 3697–3705.

[6] ALMUTAIRI, A. *Investigation of sexithiophene properties with Monte Carlo simulations of a coarse-grained model.* PhD thesis, University of Akron, 2016.

[7] ARKHIPOV, A., FREDDOLINO, P. L., AND SCHULTEN, K. Stability and dynamics of virus capsids described by coarse-grained modeling. *Structure 14*, 12 (2006), 1767–1777.

[8] AVENDAÑO, C., LAFITTE, T., ADJIMAN, C. S., GALINDO, A., MÜLLER, E. A., AND JACKSON, G. Saft-γ force field for the simulation of molecular fluids: 2. coarse-grained models of greenhouse gases, refrigerants, and long alkanes. *J. Phys. Chem. B 117*, 9 (2013), 2717–2733.

[9] BARBARA, P. F., MEYER, T. J., AND RATNER, M. A. Contemporary issues in electron transfer research. *The Journal of Physical Chemistry 100*, 31 (1996), 13148–13168.

[10] BAYRAMOGLU, B., AND FALLER, R. Coarse-grained modeling of polystyrene in various environments by iterative boltzmann inversion. *Macromolecules 45*, 22 (2012), 9205–9219.

[11] BEHLER, J., AND PARRINELLO, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Physical review letters 98*, 14 (2007), 146401.

[12] BEJAGAM, K. K., SINGH, S., AN, Y., AND DESHMUKH, S. A. Machine-learned coarse-grained models. *The journal of physical chemistry letters 9*, 16 (2018), 4667–4672.

[13] BELLO, L., AND SING, C. E. Mechanisms of diffusive charge transport in redox-active polymer solutions. *Macromolecules 53*, 18 (2020), 7658–7671.

[14] BERARDI, R., FAVA, C., AND ZANNONI, C. A gay–berne potential for dissimilar biaxial particles. *Chemical physics letters 297*, 1-2 (1998), 8–14.

[15] BIXON, M., AND JORTNER, J. Electron transfer—from isolated molecules to biomolecules. *Advances in Chemical Physics: Electron Transfer—from Isolated Molecules to Biomolecules. Part 1 106* (1999), 35–202.

[16] BOGOJESKI, M., VOGT-MARANTO, L., TUCKERMAN, M. E., MÜLLER, K.-R., AND BURKE, K. Quantum chemical accuracy from density functional approximations via machine learning. *Nature communications 11*, 1 (2020), 1–11.

[17] BOWEN, A. S., JACKSON, N. E., REID, D. R., AND DE PABLO, J. J. Structural correlations and percolation in twisted perylene diimides using a simple anisotropic coarse-grained model. *Journal of chemical theory and computation 14*, 12 (2018), 6495–6504.

[18] BRÉDAS, J.-L., BELJONNE, D., COROPCEANU, V., AND CORNIL, J. Charge-transfer and energy-transfer processes in π-conjugated oligomers and polymers: a molecular picture. *Chemical reviews 104*, 11 (2004), 4971–5004.

[19] BRERETON, T., STENZEL, O., BAUMEIER, B., ANDRIENKO, D., SCHMIDT, V., AND KROESE, D. Efficient simulation of markov chains using segmentation. *Methodology and Computing in Applied Probability 16*, 2 (2014), 465–484.

[20] BRINI, E., MARCON, V., AND VAN DER VEGT, N. F. Conditional reversible work method for molecular coarse graining applications. *Physical Chemistry Chemical Physics 13*, 22 (2011), 10468–10474.

[21] BROWN, W. M., KOHLMEYER, A., PLIMPTON, S. J., AND THARRINGTON, A. N. Implementing molecular dynamics on hybrid high performance computers–particle–particle particle-mesh. *Comput. Phys. Commun. 183*, 3 (2012), 449–459.

[22] BROWN, W. M., PETERSEN, M. K., PLIMPTON, S. J., AND GREST, G. S. Liquid crystal nanodroplets in solution. *The Journal of chemical physics 130*, 4 (2009), 044901.

[23] CASADO, J., ORTIZ, R. P., AND NAVARRETE, J. T. L. Quinoidal oligothiophenes: new properties behind an unconventional electronic structure. *Chemical Society Reviews 41*, 17 (2012), 5672–5686.

[24] CENDRA, C., GIOVANNITTI, A., SAVVA, A., VENKATRAMAN, V., MCCULLOCH, I., SALLEO, A., INAL, S., AND RIVNAY, J. Role of the anion on the transport and structure of organic mixed conductors. *Advanced Functional Materials 29*, 5 (2019), 1807034.

[25] CERUTTI, D. S., DUKE, R. E., DARDEN, T. A., AND LYBRAND, T. P. Staggered mesh ewald: an extension of the smooth particle-mesh ewald method adding great versatility. *Journal of chemical theory and computation 5*, 9 (2009), 2322–2338.

[26] CHAKRABORTY, M., XU, C., AND WHITE, A. D. Encoding and selecting coarse-grain mapping operators with hierarchical graphs. *J. Chem. Phys. 149*, 13 (2018), 134106.

[27] CHAKRABORTY, M., XU, J., AND WHITE, A. D. Is preservation of symmetry necessary for coarse-graining? *Physical Chemistry Chemical Physics 22*, 26 (2020), 14998–15005.

[28] Chen, W., Lei, T., Wu, C., Deng, M., Gong, C., Hu, K., Ma, Y., Dai, L., Lv, W., He, W., et al. Designing safe electrolyte systems for a high-stability lithium–sulfur battery. *Advanced Energy Materials 8*, 10 (2018), 1702348.

[29] Chen, X., Marks, A., Paulsen, B. D., Wu, R., Rashid, R. B., Chen, H., Alsufyani, M., Rivnay, J., and McCulloch, I. n-type rigid semiconducting polymers bearing oligo (ethylene glycol) side chains for high-performance organic electrochemical transistors. *Angewandte Chemie International Edition 60*, 17 (2021), 9368–9373.

[30] Chung, J., Khot, A., Savoie, B. M., and Boudouris, B. W. 100th anniversary of macromolecular science viewpoint: Recent advances and opportunities for mixed ion and charge conducting polymers. *ACS Macro Letters 9*, 5 (2020), 646–655.

[31] Collins, E. M., and Raghavachari, K. Effective molecular descriptors for chemical accuracy at dft cost: Fragmentation, error-cancellation, and machine learning. *Journal of Chemical Theory and Computation 16*, 8 (2020), 4938–4950.

[32] Dallavalle, M., and van der Vegt, N. F. Evaluation of mapping schemes for systematic coarse graining of higher alkanes. *Phys. Chem. Chem. Phys. 19*, 34 (2017), 23034–23042.

[33] Depa, P., Chen, C., and Maranas, J. K. Why are coarse-grained force fields too fast? a look at dynamics of four coarse-grained polymers. *J. Chem. Phys. 134*, 1 (2011), 014903.

[34] Dinpajooh, M., and Guenza, M. G. On the density dependence of the integral equation coarse-graining effective potential. *The Journal of Physical Chemistry B 122*, 13 (2017), 3426–3440.

[35] DONG, B. X., NOWAK, C., ONORATO, J. W., STRZALKA, J., ESCOBEDO, F. A., LUSCOMBE, C. K., NEALEY, P. F., AND PATEL, S. N. Influence of side-chain chemistry on structure and ionic conduction characteristics of polythiophene derivatives: a computational and experimental study. *Chemistry of Materials 31*, 4 (2019), 1418–1429.

[36] DUNN, N. J., LEBOLD, K. M., DELYSER, M. R., RUDZINSKI, J. F., AND NOID, W. Bocs: Bottom-up open-source coarse-graining software. *The Journal of Physical Chemistry B 122*, 13 (2017), 3363–3377.

[37] DURUMERIC, A. E., AND VOTH, G. A. Adversarial-residual-coarse-graining: Applying machine learning theory to systematic molecular coarse-graining. *The Journal of chemical physics 151*, 12 (2019), 124110.

[38] ELMAHMOUDY, M., INAL, S., CHARRIER, A., UGUZ, I., MALLIARAS, G. G., AND SANAUR, S. Tailoring the electrochemical and mechanical properties of pedot: Pss films for bioelectronics. *Macromolecular Materials and Engineering 302*, 5 (2017), 1600497.

[39] EMPEREUR-MOT, C., PESCE, L., DONI, G., BOCHICCHIO, D., CAPELLI, R., PEREGO, C., AND PAVAN, G. M. Swarm-cg: automatic parametrization of bonded terms in martini-based coarse-grained models of simple to complex molecules via fuzzy self-tuning particle swarm optimization. *ACS omega 5*, 50 (2020), 32823–32843.

[40] ESPANOL, P., AND ZUNIGA, I. Obtaining fully dynamic coarse-grained models from md. *Physical Chemistry Chemical Physics 13*, 22 (2011), 10538–10545.

[41] FEICHT, S. E., DEGEN, G. D., AND KHAIR, A. S. Moving ion fronts in mixed ionic-electronic conducting polymer films. *AIChE Journal 61*, 4 (2015), 1447–1454.

[42] FLAGG, L. Q., BISCHAK, C. G., ONORATO, J. W., RASHID, R. B., LUSCOMBE, C. K., AND GINGER, D. S. Polymer crystallinity controls water uptake in glycol side-chain polymer organic electrochemical transistors. *Journal of the American Chemical Society 141*, 10 (2019), 4345–4354.

[43] FLAGG, L. Q., GIRIDHARAGOPAL, R., GUO, J., AND GINGER, D. S. Anion-dependent doping and charge transport in organic electrochemical transistors. *Chemistry of Materials 30*, 15 (2018), 5380–5389.

[44] FOLEY, T. T., KIDDER, K. M., SHELL, M. S., AND NOID, W. Exploring the landscape of model representations. *Proceedings of the National Academy of Sciences 117*, 39 (2020), 24061–24068.

[45] FOLEY, T. T., SHELL, M. S., AND NOID, W. The impact of resolution upon entropy and information in coarse-grained models. *The Journal of chemical physics 143*, 24 (2015), 12B601_1.

[46] FRANCO-GONZALEZ, J. F., AND ZOZOULENKO, I. V. Molecular dynamics study of morphology of doped pedot: from solution to dry phase. *The Journal of Physical Chemistry B 121*, 16 (2017), 4299–4307.

[47] FU, C.-C., KULKARNI, P. M., SCOTT SHELL, M., AND GARY LEAL, L. A test of systematic coarse-graining of molecular dynamics simulations: Thermodynamic properties. *J. Chem. Phys. 137*, 16 (2012), 164106.

[48] FU, C.-C., KULKARNI, P. M., SHELL, M. S., AND LEAL, L. G. A test of systematic coarse-graining of molecular dynamics simulations: Transport properties. *J. Chem. Phys. 139*, 9 (2013), 094107.

[49] GANGULY, P., AND VAN DER VEGT, N. F. Representability and transferability of kirkwood–buff iterative boltzmann inversion models for multicomponent aqueous systems. *J. Chem. Theory Comput. 9*, 12 (2013), 5247–5256.

[50] GIOVANNITTI, A., MARIA, I. P., HANIFI, D., DONAHUE, M. J., BRYANT, D., BARTH, K. J., MAKDAH, B. E., SAVVA, A., MOIA, D., ZETEK, M., ET AL. The role of the side chain on the performance of n-type conjugated polymers in aqueous electrolytes. *Chemistry of Materials 30*, 9 (2018), 2945–2953.

[51] GIOVANNITTI, A., NIELSEN, C. B., SBIRCEA, D.-T., INAL, S., DONAHUE, M., NIAZI, M. R., HANIFI, D. A., AMASSIAN, A., MALLIARAS, G. G., RIVNAY, J., ET AL. N-type organic electrochemical transistors with stability in water. *Nature communications 7* (2016), 13066.

[52] GIOVANNITTI, A., SBIRCEA, D.-T., INAL, S., NIELSEN, C. B., BANDIELLO, E., HANIFI, D. A., SESSOLO, M., MALLIARAS, G. G., McCULLOCH, I., AND RIVNAY, J. Controlling the mode of operation of organic transistors through side-chain engineering. *Proceedings of the National Academy of Sciences 113*, 43 (2016), 12017–12022.

[53] GIRIDHARAGOPAL, R., FLAGG, L., HARRISON, J., ZIFFER, M., ONORATO, J., LUSCOMBE, C., AND GINGER, D. Electrochemical strain microscopy probes morphology-induced variations in ion uptake and performance in organic electrochemical transistors. *Nature materials 16*, 7 (2017), 737–742.

[54] GOHLKE, H., AND THORPE, M. F. A natural coarse graining for simulating large biomolecular motion. *Biophysical Journal 91*, 6 (2006), 2115–2120.

[55] GROVES, C. Simulating charge transport in organic semiconductors and devices: a review. *Reports on Progress in Physics 80*, 2 (2016), 026502.

[56] GROZEMA, F. C., VAN DUIJNEN, P. T., BERLIN, Y. A., RATNER, M. A., AND SIEBBELES, L. D. Intramolecular charge transport along isolated chains of conjugated polymers: effect of torsional disorder and polymerization defects. *The Journal of Physical Chemistry B 106*, 32 (2002), 7791–7795.

[57] GRUNEWALD, F., ROSSI, G., DE VRIES, A. H., MARRINK, S. J., AND MONTI-CELLI, L. Transferable martini model of poly (ethylene oxide). *The Journal of Physical Chemistry B 122*, 29 (2018), 7436–7449.

[58] GUO, X., FACCHETTI, A., AND MARKS, T. J. Imide-and amide-functionalized polymer semiconductors. *Chemical reviews 114*, 18 (2014), 8943–9021.

[59] GUSTAFSSON, J., LIEDBERG, B., AND INGANÄS, O. In situ spectroscopic investigations of electrochromism and ion transport in a poly (3, 4-ethylenedioxythiophene) electrode in a solid state electrochemical cell. *Solid state ionics 69*, 2 (1994), 145–152.

[60] HADLEY, K. R., AND MCCABE, C. Coarse-grained molecular models of water: a review. *Mol. Simul. 38*, 8-9 (2012), 671–681.

[61] HALLANI, R. K., PAULSEN, B. D., PETTY, A. J., SHEELAMANTHULA, R., MOSER, M., THORLEY, K. J., SOHN, W., RASHID, R. B., SAVVA, A., MORO, S., ET AL. Regiochemistry-driven organic electrochemical transistor performance enhancement in ethylene glycol-functionalized polythiophenes. *Journal of the American Chemical Society 143*, 29 (2021), 11007–11018.

[62] HENDERSON, R. A uniqueness theorem for fluid pair correlation functions. *Phys. Lett. A 49*, 3 (1974), 197–198.

[63] HOCKNEY, R. W., AND EASTWOOD, J. W. *Computer simulation using particles.* Bristol: Hilger, 1988, 1988.

[64] HUSIC, B. E., CHARRON, N. E., LEMM, D., WANG, J., PÉREZ, A., MAJEWSKI, M., KRÄMER, A., CHEN, Y., OLSSON, S., DE FABRITIIS, G., ET AL. Coarse graining molecular dynamics with graph neural networks. *The Journal of Chemical Physics 153*, 19 (2020), 194101.

[65] HUTCHISON, G. R., RATNER, M. A., AND MARKS, T. J. Intermolecular charge transfer between heterocyclic oligomers. effects of heteroatom and molecular packing on hopping transport in organic semiconductors. *Journal of the American Chemical Society 127*, 48 (2005), 16866–16881.

[66] INAL, S., MALLIARAS, G. G., AND RIVNAY, J. Benchmarking organic mixed conductors for transistors. *Nature communications 8*, 1 (2017), 1–7.

[67] INAL, S., RIVNAY, J., LELEUX, P., FERRO, M., RAMUZ, M., BRENDEL, J. C., SCHMIDT, M. M., THELAKKAT, M., AND MALLIARAS, G. G. A high transconductance accumulation mode electrochemical transistor. *Advanced Materials 26*, 44 (2014), 7450–7455.

[68] INGÓLFSSON, H. I., LOPEZ, C. A., UUSITALO, J. J., DE JONG, D. H., GOPAL, S. M., PERIOLE, X., AND MARRINK, S. J. The power of coarse graining in biomolecular simulations. *Wiley Interdiscip. Rev. Comput. Mol. Sci. 4*, 3 (2014), 225–248.

[69] IZVEKOV, S., AND VOTH, G. A. A multiscale coarse-graining method for biomolecular systems. *The Journal of Physical Chemistry B 109*, 7 (2005), 2469–2473.

[70] IZVEKOV, S., AND VOTH, G. A. Multiscale coarse graining of liquid-state systems. *J. Chem. Phys. 123*, 13 (2005), 134105.

[71] IZVEKOV, S., AND VOTH, G. A. Modeling real dynamics in the coarse-grained representation of condensed phase systems. *J. Chem. Phys. 125*, 15 (2006), 151101.

[72] JACKSON, N. E. Coarse-graining organic semiconductors: The path to multiscale design. *The Journal of Physical Chemistry B 125*, 2 (2020), 485–496.

[73] JACKSON, N. E., BOWEN, A. S., ANTONY, L. W., WEBB, M. A., VISHWANATH, V., AND DE PABLO, J. J. Electronic structure at coarse-grained resolutions from supervised machine learning. *Science advances 5*, 3 (2019), eaav1190.

[74] JACOBS, I. E., AND MOULÉ, A. J. Controlling molecular doping in organic semiconductors. *Advanced Materials 29*, 42 (2017), 1703063.

[75] JOHN, S., AND CSÁNYI, G. Many-body coarse-grained interactions using gaussian approximation potentials. *The Journal of Physical Chemistry B 121*, 48 (2017), 10934–10949.

[76] JOHNSON, M. E., HEAD-GORDON, T., AND LOUIS, A. A. Representability problems for coarse-grained water potentials. *J. Chem. Phys. 126*, 14 (2007), 144509.

189

[77] JORGENSEN, W. L., MAXWELL, D. S., AND TIRADO-RIVES, J. Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc. 118*, 45 (1996), 11225–11236.

[78] KAPHLE, V., PAUDEL, P. R., DAHAL, D., KRISHNAN, R. K. R., AND LÜSSEM, B. Finding the equilibrium of organic electrochemical transistors. *Nature communications 11*, 1 (2020), 1–11.

[79] KARIMI-VARZANEH, H. A., QIAN, H.-J., CHEN, X., CARBONE, P., AND MÜLLER-PLATHE, F. Ibisco: A molecular dynamics simulation package for coarse-grained simulation. *Journal of computational chemistry 32*, 7 (2011), 1475–1487.

[80] KHOT, A., AND SAVOIE, B. M. How side-chain hydrophilicity modulates morphology and charge transport in mixed conducting polymers. *Journal of Polymer Science* (2021).

[81] KHOT, A., AND SAVOIE, B. M. Top–down coarse-grained framework for characterizing mixed conducting polymers. *Macromolecules* (2021).

[82] KHOT, A., SHIRING, S. B., AND SAVOIE, B. M. Evidence of information limitations in coarse-grained models. *The Journal of chemical physics 151*, 24 (2019), 244105.

[83] KIM, S.-M., KIM, C.-H., KIM, Y., KIM, N., LEE, W.-J., LEE, E.-H., KIM, D., PARK, S., LEE, K., RIVNAY, J., ET AL. Influence of pedot: Pss crystallinity and composition on electrochemical transistor performance and long-term stability. *Nature communications 9*, 1 (2018), 1–9.

[84] KRISHNA, V., NOID, W. G., AND VOTH, G. A. The multiscale coarse-graining method. iv. transferring coarse-grained potentials between temperatures. *J. Chem. Phys. 131*, 2 (2009), 024103.

[85] KUEI, B., AND GOMEZ, E. D. Chain conformations and phase behavior of conjugated polymers. *Soft matter 13*, 1 (2017), 49–67.

[86] KUNUGI, Y., HARIMA, Y., YAMASHITA, K., OHTA, N., AND ITO, S. Charge transport in a regioregular poly (3-octylthiophene) film. *Journal of Materials Chemistry 10*, 12 (2000), 2673–2677.

[87] LEI, H., YANG, X., LI, Z., AND KARNIADAKIS, G. E. Systematic parameter inference in stochastic mesoscopic modeling. *Journal of Computational Physics 330* (2017), 571–593.

[88] LEMKE, T., AND PETER, C. Neural network based prediction of conformational free energies-a new route toward coarse-grained simulation models. *Journal of chemical theory and computation 13*, 12 (2017), 6213–6221.

[89] LI, Z., WELLAWATTE, G. P., CHAKRABORTY, M., GANDHI, H. A., XU, C., AND WHITE, A. D. Graph neural network based coarse-grained mapping prediction. *Chemical science 11*, 35 (2020), 9524–9531.

[90] LIU, J., QIU, L., ALESSANDRI, R., QIU, X., PORTALE, G., DONG, J., TALSMA, W., YE, G., SENGRIAN, A. A., SOUZA, P. C., ET AL. Enhancing molecular n-type doping of donor–acceptor copolymers by tailoring side chains. *Advanced Materials 30*, 7 (2018), 1704630.

[91] LOUIS, A. Beware of density dependent pair potentials. *J. Phys. Condens. Matter 14*, 40 (2002), 9187.

[92] LU, J., PINTO, N. J., AND MACDIARMID, A. G. Apparent dependence of conductivity of a conducting polymer on an electric field in a field effect transistor configuration. *Journal of Applied Physics 92*, 10 (2002), 6033–6038.

[93] LU, L., IZVEKOV, S., DAS, A., ANDERSEN, H. C., AND VOTH, G. A. Efficient, regularized, and scalable algorithms for multiscale coarse-graining. *Journal of chemical theory and computation 6*, 3 (2010), 954–965.

[94] LYUBARTSEV, A. P., AND LAAKSONEN, A. Calculation of effective interaction potentials from radial distribution functions: A reverse monte carlo approach. *Phys. Rev. E 52*, 4 (1995), 3730.

[95] MARCUS, R. A. Chemical and electrochemical electron-transfer theory. *Annual review of physical chemistry 15*, 1 (1964), 155–196.

[96] MARCUS, R. A. Electron transfer reactions in chemistry. theory and experiment. *Reviews of Modern Physics 65*, 3 (1993), 599.

[97] MARCUS, R. A., AND SUTIN, N. Electron transfers in chemistry and biology. *Biochimica et Biophysica Acta (BBA)-Reviews on Bioenergetics 811*, 3 (1985), 265–322.

[98] MARIA, I. P., PAULSEN, B. D., SAVVA, A., OHAYON, D., WU, R., HALLANI, R., BASU, A., DU, W., ANTHOPOULOS, T. D., INAL, S., ET AL. The effect of alkyl spacers on the mixed ionic-electronic conduction properties of n-type polymers. *Advanced Functional Materials 31*, 14 (2021), 2008718.

[99] MARRINK, S. J., DE VRIES, A. H., AND MARK, A. E. Coarse grained model for semiquantitative lipid simulations. *J. Phys. Chem. B 108*, 2 (2004), 750–760.

[100] MARRINK, S. J., RISSELADA, H. J., YEFIMOV, S., TIELEMAN, D. P., AND DE VRIES, A. H. The martini force field: coarse grained model for biomolecular simulations. *J. Phys. Chem. B 111*, 27 (2007), 7812–7824.

[101] MARRINK, S. J., AND TIELEMAN, D. P. Perspective on the martini model. *Chem. Soc. Rev. 42*, 16 (2013), 6801–6822.

[102] MARTYNA, G. J., TOBIAS, D. J., AND KLEIN, M. L. Constant pressure molecular dynamics algorithms. *J. Chem. Phys. 101*, 5 (1994), 4177–4189.

[103] MEHANDZHIYSKI, A. Y., AND ZOZOULENKO, I. Computational microscopy of pedot: pss/cellulose composite paper. *ACS Applied Energy Materials 2*, 5 (2019), 3568–3577.

[104] MEI, J., DIAO, Y., APPLETON, A. L., FANG, L., AND BAO, Z. Integrated materials design of organic semiconductors for field-effect transistors. *Journal of the American Chemical Society 135*, 18 (2013), 6724–6746.

[105] MILANO, G., AND MÜLLER-PLATHE, F. Mapping atomistic simulations to meso-scopic models: A systematic coarse-graining procedure for vinyl polymer chains. *J. Phys. Chem. B 109*, 39 (2005), 18609–18619.

[106] MIRZOEV, A., AND LYUBARTSEV, A. P. Magic: Software package for multiscale modeling. *Journal of chemical theory and computation 9*, 3 (2013), 1512–1520.

[107] MODARRESI, M., FRANCO-GONZALEZ, J. F., AND ZOZOULENKO, I. Morphology and ion diffusion in pedot: Tos. a coarse grained molecular dynamics simulation. *Physical Chemistry Chemical Physics 20*, 25 (2018), 17188–17198.

[108] MODARRESI, M., FRANCO-GONZALEZ, J. F., AND ZOZOULENKO, I. Computational microscopy study of the granular structure and ph dependence of pedot: Pss. *Physical Chemistry Chemical Physics 21*, 12 (2019), 6699–6711.

[109] MODARRESI, M., MEHANDZHIYSKI, A., FAHLMAN, M., TYBRANDT, K., AND ZO-ZOULENKO, I. Microscopic understanding of the granular structure and the swelling of pedot: Pss. *Macromolecules 53*, 15 (2020), 6267–6278.

[110] MOSER, M., GLADISCH, J., GHOSH, S., HIDALGO, T. C., PONDER JR, J. F., SHEELAMANTHULA, R., THIBURCE, Q., GASPARINI, N., WADSWORTH, A., SALLEO, A., ET AL. Controlling electrochemically induced volume changes in conjugated polymers by chemical design: from theory to devices. *Advanced Functional Materials* (2021), 2100723.

[111] MOSER, M., HIDALGO, T. C., SURGAILIS, J., GLADISCH, J., GHOSH, S., SHEELAMANTHULA, R., THIBURCE, Q., GIOVANNITTI, A., SALLEO, A., GASPARINI, N., ET AL. Side chain redistribution as a strategy to boost organic electrochemical transistor performance and stability. *Advanced Materials 32*, 37 (2020), 2002748.

[112] MOSER, M., SAVAGIAN, L. R., SAVVA, A., MATTA, M., PONDER JR, J. F., HIDALGO, T. C., OHAYON, D., HALLANI, R., REISJALALI, M., TROISI, A., ET AL. Ethylene glycol-based side chain length engineering in polythiophenes and its impact on organic electrochemical transistor performance. *Chemistry of Materials 32*, 15 (2020), 6618–6628.

[113] MÜLLER, E. A., AND JACKSON, G. Force-field parameters from the saft-γ equation of state for use in coarse-grained molecular simulations. *Annual review of chemical and biomolecular engineering 5* (2014), 405–427.

[114] NAOI, K., LIEN, M., AND SMYRL, W. H. Quartz crystal microbalance study: ionic motion across conducting polymers. *Journal of The Electrochemical Society 138*, 2 (1991), 440–445.

[115] NEELOV, A., AND HOLM, C. Interlaced p3m algorithm with analytical and ik-differentiation. *The Journal of chemical physics 132*, 23 (2010), 234103.

[116] NEESE, F. Software update: the orca program system, version 4.0. *Wiley Interdiscip. Rev. Comput. Mol. Sci. 8*, 1 (2018), e1327.

[117] NELSON, J., AND CHANDLER, R. E. Random walk models of charge transfer and transport in dye sensitized systems. *Coordination Chemistry Reviews 248*, 13-14 (2004), 1181–1194.

[118] NICOLINI, T., SURGAILIS, J., SAVVA, A., SCACCABAROZZI, A. D., NAKAR, R., THUAU, D., WANTZ, G., RICHTER, L. J., DAUTEL, O., HADZIIOANNOU, G., ET AL. A low-swelling polymeric mixed conductor operating in aqueous electrolytes. *Advanced Materials 33*, 2 (2021), 2005723.

[119] NIELSEN, C. B., GIOVANNITTI, A., SBIRCEA, D.-T., BANDIELLO, E., NIAZI, M. R., HANIFI, D. A., SESSOLO, M., AMASSIAN, A., MALLIARAS, G. G., RIVNAY, J., ET AL. Molecular design of semiconducting polymers for high-performance organic electrochemical transistors. *Journal of the American Chemical Society 138*, 32 (2016), 10252–10259.

[120] Noid, W., Chu, J.-W., Ayton, G. S., Krishna, V., Izvekov, S., Voth, G. A., Das, A., and Andersen, H. C. The multiscale coarse-graining method. i. a rigorous bridge between atomistic and coarse-grained models. *J. Chem. Phys. 128*, 24 (2008), 244114.

[121] Noid, W., Liu, P., Wang, Y., Chu, J.-W., Ayton, G. S., Izvekov, S., Andersen, H. C., and Voth, G. A. The multiscale coarse-graining method. ii. numerical implementation for coarse-grained molecular models. *J. Chem. Phys. 128*, 24 (2008), 244115.

[122] Ohayon, D., Savva, A., Du, W., Paulsen, B. D., Uguz, I., Ashraf, R. S., Rivnay, J., McCulloch, I., and Inal, S. Influence of side chains on the n-type organic electrochemical transistor performance. *ACS Applied Materials & Interfaces 13*, 3 (2021), 4253–4266.

[123] Onorato, J. W., and Luscombe, C. K. Morphological effects on polymeric mixed ionic/electronic conductors. *Molecular Systems Design & Engineering 4*, 2 (2019), 310–324.

[124] Pacheco-Moreno, C. M., Schreck, M., Scaccabarozzi, A. D., Bourgun, P., Wantz, G., Stevens, M. M., Dautel, O. J., and Stingelin, N. The importance of materials design to make ions flow: toward novel materials platforms for bioelectronics applications. *Advanced Materials 29*, 4 (2017), 1604446.

[125] Pasi, M., Lavery, R., and Ceres, N. Palace: A coarse-grain protein model for studying mechanical properties. *Journal of chemical theory and computation 9*, 1 (2013), 785–793.

[126] Paulsen, B. D., Tybrandt, K., Stavrinidou, E., and Rivnay, J. Organic mixed ionic–electronic conductors. *Nature materials* (2019), 1–14.

[127] Paulsen, B. D., Tybrandt, K., Stavrinidou, E., and Rivnay, J. Organic mixed ionic–electronic conductors. *Nature materials 19*, 1 (2020), 13–26.

[128] PEI, Q., ZUCCARELLO, G., AHLSKOG, M., AND INGANÄS, O. Electrochromic and highly stable poly (3, 4-ethylenedioxythiophene) switches between opaque blue-black and transparent sky blue. *Polymer 35*, 7 (1994), 1347–1351.

[129] PITTELLI, S. L., DE KEERSMAECKER, M., PONDER JR, J. F., ÖSTERHOLM, A. M., OCHIENG, M. A., AND REYNOLDS, J. R. Structural effects on the charge transport properties of chemically and electrochemically doped dioxythiophene polymers. *Journal of Materials Chemistry C 8*, 2 (2020), 683–693.

[130] PLIMPTON, S. Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys. 117*, 1 (1995), 1–19.

[131] POELKING, C., DAOULAS, K., TROISI, A., AND ANDRIENKO, D. Morphology and charge transport in p3ht: a theorist's perspective. In *P3HT Revisited–From Molecular Scale to Solar Cell Devices*. Springer, 2014, pp. 139–180.

[132] QIU, Y.-J., AND REYNOLDS, J. R. Dopant anion controlled ion transport behavior of polypyrrole. *Polymer Engineering & Science 31*, 6 (1991), 417–421.

[133] RAMAKRISHNAN, R., DRAL, P. O., RUPP, M., AND VON LILIENFELD, O. A. Big data meets quantum chemistry approximations: the δ-machine learning approach. *Journal of chemical theory and computation 11*, 5 (2015), 2087–2096.

[134] REITH, D., PÜTZ, M., AND MÜLLER-PLATHE, F. Deriving effective mesoscale potentials from atomistic simulations. *J. Comput. Chem. 24*, 13 (2003), 1624–1636.

[135] RICCI, M., ROSCIONI, O. M., QUERCIAGROSSA, L., AND ZANNONI, C. Molc. a reversible coarse grained approach using anisotropic beads for the modelling of organic functional materials. *Physical Chemistry Chemical Physics 21*, 47 (2019), 26195–26211.

[136] RIVNAY, J., INAL, S., COLLINS, B. A., SESSOLO, M., STAVRINIDOU, E., STRAKOSAS, X., TASSONE, C., DELONGCHAMP, D. M., AND MALLIARAS, G. G. Structural control of mixed ionic and electronic transport in conducting polymers. *Nature communications 7*, 1 (2016), 1–9.

[137] RIVNAY, J., INAL, S., SALLEO, A., OWENS, R. M., BERGGREN, M., AND MALLIARAS, G. G. Organic electrochemical transistors. *Nature Reviews Materials 3*, 2 (2018), 1–14.

[138] ROLLAND, N., MODARRESI, M., FRANCO-GONZALEZ, J. F., AND ZOZOULENKO, I. Large scale mobility calculations in pedot (poly (3, 4-ethylenedioxythiophene)): Backmapping the coarse-grained martini morphology. *Computational Materials Science 179* (2020), 109678.

[139] ROSSI, G., FUCHS, P., BARNOUD, J., AND MONTICELLI, L. A coarse-grained martini model of polyethylene glycol and of polyoxyethylene alkyl ether surfactants. *The Journal of Physical Chemistry B 116*, 49 (2012), 14353–14362.

[140] RÜHLE, V., AND JUNGHANS, C. Hybrid approaches to coarse-graining using the votca package: Liquid hexane. *Macromol. Theory Simulations 20*, 7 (2011), 472–477.

[141] RUHLE, V., JUNGHANS, C., LUKYANOV, A., KREMER, K., AND ANDRIENKO, D. Versatile object-oriented toolkit for coarse-graining applications. *J. Chem. Theory Comput. 5*, 12 (2009), 3211–3223.

[142] RÜHLE, V., KIRKPATRICK, J., AND ANDRIENKO, D. A multiscale description of charge transport in conjugated oligomers. *The Journal of chemical physics 132*, 13 (2010), 134103.

[143] RYCKAERT, J.-P., CICCOTTI, G., AND BERENDSEN, H. J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of computational physics 23*, 3 (1977), 327–341.

[144] RÜHLE, V., LUKYANOV, A., MAY, F., SCHRADER, M., VEHOFF, T., KIRKPATRICK, J., BAUMEIER, B., AND ANDRIENKO, D. Microscopic simulations of charge transport in disordered organic semiconductors. *Journal of chemical theory and computation 7*, 10 (2011), 3335–3345.

[145] Sahalianov, I., Hynynen, J., Barlow, S., Marder, S. R., Müller, C., and Zozoulenko, I. Uv-to-ir absorption of molecularly p-doped polythiophenes with alkyl and oligoether side chains: Experiment and interpretation based on density functional theory. *The Journal of Physical Chemistry B 124*, 49 (2020), 11280–11293.

[146] Savagian, L. R., Österholm, A. M., Ponder Jr, J. F., Barth, K. J., Rivnay, J., and Reynolds, J. R. Balancing charge storage and mobility in an oligo (ether) functionalized dioxythiophene copolymer for organic-and aqueous-based electrochemical devices and transistors. *Advanced Materials 30*, 50 (2018), 1804647.

[147] Savoie, B. M., Kohlstedt, K. L., Jackson, N. E., Chen, L. X., De La Cruz, M. O., Schatz, G. C., Marks, T. J., and Ratner, M. A. Mesoscale molecular network formation in amorphous organic materials. *Proceedings of the National Academy of Sciences 111*, 28 (2014), 10055–10060.

[148] Savoie, B. M., Webb, M. A., and Miller III, T. F. Enhancing cation diffusion and suppressing anion diffusion via lewis-acidic polymer electrolytes. *J. Phys. Chem. Lett. 8*, 3 (2017), 641–646.

[149] Savva, A., Cendra, C., Giugni, A., Torre, B., Surgailis, J., Ohayon, D., Giovannitti, A., McCulloch, I., Di Fabrizio, E., Salleo, A., et al. Influence of water on the performance of organic electrochemical transistors. *Chemistry of Materials 31*, 3 (2019), 927–937.

[150] Savva, A., Hallani, R., Cendra, C., Surgailis, J., Hidalgo, T. C., Wustoni, S., Sheelamanthula, R., Chen, X., Kirkus, M., Giovannitti, A., et al. Balancing ionic and electronic conduction for high-performance organic electrochemical transistors. *Advanced Functional Materials 30*, 11 (2020), 1907657.

[151] Savva, A., Wustoni, S., and Inal, S. Ionic-to-electronic coupling efficiency in pedot: Pss films operated in aqueous electrolytes. *Journal of Materials Chemistry C 6*, 44 (2018), 12023–12030.

[152] SCHERER, C., SCHEID, R., ANDRIENKO, D., AND BEREAU, T. Kernel-based machine learning for efficient simulations of molecular liquids. *Journal of chemical theory and computation 16*, 5 (2020), 3194–3204.

[153] SCHMODE, P., SAVVA, A., KAHL, R., OHAYON, D., MEICHSNER, F., DOLYNCHUK, O., THURN-ALBRECHT, T., INAL, S., AND THELAKKAT, M. The key role of side chain linkage in structure formation and mixed conduction of ethylene glycol substituted polythiophenes. *ACS applied materials & interfaces 12*, 11 (2020), 13029–13039.

[154] SHELL, M. S. The relative entropy is fundamental to multiscale and inverse thermodynamic problems. *J. Chem. Phys. 129*, 14 (2008), 144108.

[155] SHI, H., LIU, C., JIANG, Q., AND XU, J. Effective approaches to improve the electrical conductivity of pedot: Pss: a review. *Advanced Electronic Materials 1*, 4 (2015), 1500017.

[156] SMELA, E., AND GADEGAARD, N. Surprising volume change in ppy (dbs): an atomic force microscopy study. *Advanced Materials 11*, 11 (1999), 953–957.

[157] SNOOK, G. A., KAO, P., AND BEST, A. S. Conducting-polymer-based supercapacitor devices and electrodes. *Journal of power sources 196*, 1 (2011), 1–12.

[158] STAVRINIDOU, E., LELEUX, P., RAJAONA, H., KHODAGHOLY, D., RIVNAY, J., LINDAU, M., SANAUR, S., AND MALLIARAS, G. G. Direct measurement of ion mobility in a conducting polymer. *Advanced Materials 25*, 32 (2013), 4488–4493.

[159] STILLINGER, F. H., SAKAI, H., AND TORQUATO, S. Statistical mechanical models with effective potentials: Definitions, applications, and thermodynamic consequences. *J. Chem. Phys. 117*, 1 (2002), 288–296.

[160] STUKOWSKI, A. Visualization and analysis of atomistic simulation data with ovito–the open visualization tool. *Modelling and Simulation in Materials Science and Engineering 18*, 1 (2009), 015012.

[161] Sunarso, J., Baumann, S., Serra, J., Meulenberg, W., Liu, S., Lin, Y., and Da Costa, J. D. Mixed ionic–electronic conducting (miec) ceramic-based membranes for oxygen separation. *Journal of membrane science 320*, 1-2 (2008), 13–41.

[162] Surgailis, J., Savva, A., Druet, V., Paulsen, B. D., Wu, R., Hamidi-Sakr, A., Ohayon, D., Nikiforidis, G., Chen, X., McCulloch, I., et al. Mixed conduction in an n-type organic semiconductor in the absence of hydrophilic side-chains. *Advanced Functional Materials 31*, 21 (2021), 2010165.

[163] Szymański, M. Z., Tu, D., and Forchheimer, R. 2-d drift-diffusion simulation of organic electrochemical transistors. *IEEE Transactions on Electron Devices 64*, 12 (2017), 5114–5120.

[164] Takano, T., Masunaga, H., Fujiwara, A., Okuzaki, H., and Sasaki, T. Pedot nanocrystal in highly conductive pedot: Pss polymer films. *Macromolecules 45*, 9 (2012), 3859–3865.

[165] Thelen, J. L., Wu, S.-L., Javier, A. E., Srinivasan, V., Balsara, N. P., and Patel, S. N. Relationship between mobility and lattice strain in electrochemically doped poly (3-hexylthiophene). *ACS Macro Letters 4*, 12 (2015), 1386–1391.

[166] Thomas, E. M., Brady, M. A., Nakayama, H., Popere, B. C., Segalman, R. A., and Chabinyc, M. L. X-ray scattering reveals ion-induced microstructural changes during electrochemical gating of poly (3-hexylthiophene). *Advanced Functional Materials 28*, 44 (2018), 1803687.

[167] van de Burgt, Y., Lubberman, E., Fuller, E. J., Keene, S. T., Faria, G. C., Agarwal, S., Marinella, M. J., Talin, A. A., and Salleo, A. A non-volatile organic electrochemical device as a low-voltage artificial synapse for neuromorphic computing. *Nature materials 16*, 4 (2017), 414–418.

[168] Volkov, A. V., Wijeratne, K., Mitraka, E., Ail, U., Zhao, D., Tybrandt, K., Andreasen, J. W., Berggren, M., Crispin, X., and Zozoulenko, I. V. Understanding the capacitance of pedot: Pss. *Advanced Functional Materials 27*, 28 (2017), 1700329.

[169] Wagner, J. W., Dama, J. F., Durumeric, A. E., and Voth, G. A. On the representability problem and the physical meaning of coarse-grained models. *J. Chem. Phys. 145*, 4 (2016), 044108.

[170] Wang, G., Morrin, A., Li, M., Liu, N., and Luo, X. Nanomaterial-doped conducting polymers for electrochemical sensors and biosensors. *Journal of Materials Chemistry B 6*, 25 (2018), 4173–4190.

[171] Wang, H., Barrett, M., Duane, B., Gu, J., and Zenhausern, F. Materials and processing of polymer-based electrochromic devices. *Materials Science and Engineering: B 228* (2018), 167–174.

[172] Wang, H., Junghans, C., and Kremer, K. Comparative atomistic and coarse-grained study of water: What do we lose by coarse-graining? *Eur. Phys. J. E 28*, 2 (2009), 221–229.

[173] Wang, J., Olsson, S., Wehmeyer, C., Pérez, A., Charron, N. E., De Fabritiis, G., Noé, F., and Clementi, C. Machine learning of coarse-grained molecular dynamics force fields. *ACS Cent. Sci. 5*, 5 (2019), 755.

[174] Wang, S., Ha, M., Manno, M., Frisbie, C. D., and Leighton, C. Hopping transport and the hall effect near the insulator–metal transition in electrochemically gated poly (3-hexylthiophene) transistors. *Nature communications 3*, 1 (2012), 1–7.

[175] Wang, S., Ma, Z., and Pan, W. Data-driven coarse-grained modeling of polymers in solution with structural and dynamic properties conserved. *Soft Matter 16*, 36 (2020), 8330–8344.

[176] Wang, W., and Gómez-Bombarelli, R. Coarse-graining auto-encoders for molecular dynamics. *npj Computational Materials 5*, 1 (2019), 1–9.

[177] WANG, Y., NOID, W. G., LIU, P., AND VOTH, G. A. Effective force coarse-graining. *Physical Chemistry Chemical Physics 11*, 12 (2009), 2002–2015.

[178] WEBB, M. A., DELANNOY, J.-Y., AND DE PABLO, J. J. Graph-based approach to systematic molecular coarse-graining. *J. Chem. Theory Comput. 15*, 2 (2018), 1199–1208.

[179] WEBB, M. A., SAVOIE, B. M., WANG, Z.-G., AND MILLER III, T. F. Chemically specific dynamic bond percolation model for ion transport in polymer electrolytes. *Macromolecules 48*, 19 (2015), 7346–7358.

[180] YANG, S., CUI, Z., AND QU, J. A coarse-grained model for epoxy molding compound. *The Journal of Physical Chemistry B 118*, 6 (2014), 1660–1669.

[181] YESYLEVSKYY, S. O., SCHÄFER, L. V., SENGUPTA, D., AND MARRINK, S. J. Polarizable water model for the coarse-grained martini force field. *PLoS computational biology 6*, 6 (2010).

[182] ZADE, S. S., ZAMOSHCHIK, N., AND BENDIKOV, M. From short conjugated oligomers to conjugated polymers. lessons from studies on long conjugated oligomers. *Accounts of Chemical Research 44*, 1 (2011), 14–24.

[183] ZEGLIO, E., ERIKSSON, J., GABRIELSSON, R., SOLIN, N., AND INGANÄS, O. Highly stable conjugated polyelectrolytes for water-based hybrid mode electrochemical transistors. *Advanced Materials 29*, 19 (2017), 1605787.

[184] ZHANG, L., HAN, J., WANG, H., CAR, R., AND E, W. Deepcg: Constructing coarse-grained models via deep neural networks. *The Journal of chemical physics 149*, 3 (2018), 034101.

[185] ZHANG, Q., LI, B., HUANG, S., NOMURA, H., TANAKA, H., AND ADACHI, C. Efficient blue organic light-emitting diodes employing thermally activated delayed fluorescence. *Nature Photonics 8*, 4 (2014), 326.

[186] Zhao, F., Shi, Y., Pan, L., and Yu, G. Multifunctional nanostructured conductive polymer gels: synthesis, properties, and applications. *Accounts of chemical research 50*, 7 (2017), 1734–1743.

[187] Österholm, A. M., Ponder Jr, J. F., De Keersmaecker, M., Shen, D. E., and Reynolds, J. R. Disentangling redox properties and capacitance in solution-processed conjugated polymers. *Chemistry of Materials 31*, 8 (2019), 2971–2982.