SYSTEMATIC COARSE-GRAINING AND DYNAMICAL SIMULATIONS OF ANISOTROPIC MOLECULES WITH APPLICATIONS IN ORGANIC SEMICONDUCTORS

HUONG THI LAN NGUYEN

Supervisors: A/Prof David Huang & A/Prof Tak Kee

A thesis submitted in partial fulfilment of the
requirements for the degree of
Master of Philosophy in Chemical Science

May 2019

Department of Chemistry, School of Physcial Sciences
The University of Adelaide
Adelaide, South Australia 5005

Contents

List of figures

List of tables

List of abbreviations

Abbreviation	**Description**
CG	coarse-grained
OPV	organic photovoltaic
BQR	benzodithiophene quaterthiophene rhodanine
BTR	benzodithiophene terthiophene rhodanine
$PC_{71}BM$	[6,6]-phenyl-C71-butyric acid methyl ester
MD	molecular dynamics
FG	fine-grained
RBF	radial basis function
AFM-CG	anisotropic force-matching coarse-grained
PMF	potential of mean force
MSCG	multi-scale coarse-graining
IBI	iterative Boltzmann inversion
RDF	radial distribution function
PPPM	particle-particle particle-mesh

Abstract

Organic semiconductors are used widely in different applications, including organic photovoltaics (OPVs), devices that convert solar energy to electricity. These devices, if applicable commercially, can help to supply the world's energy needs without requiring complicated manufacturing and maintenance. Moreover, OPVs possess several useful physical properties such as being light weight, highly transparent, and flexible. This makes organic electronic devices advantageous over those made of inorganic hard materials, especially in applications in which these conditions are required.

Although experimental studies show that organic semiconductors can potentially yield high performing devices, the electronic processes that govern the conversion of light to energy are not fully understood. Specifically, how free electrons are created and transferred within the device when a photon is absorbed is strongly debated in the literature. Many experimental and theoretical results have shown that microstructure at the interfaces between the component organic semiconductor materials that make up the device plays an important role in these processes. The microstructure can be induced by directional forces between generally anisotropic organic-semiconductor molecules, combined with translational symmetry breaking at interfaces. In Chapter 3, the interface of a high-performing electron donor–acceptor OPV system consisting of two small organic semiconductors benzodithiophene quaterthiophene rhodanine (BQR) and [6,6]-phenyl-C71-butyric acid methyl ester ($PC_{71}BM$) is studied using classical molecular dynamics (MD). Atomistic simulations at high temperatures indicate that the "face-on" configuration is more favorable at a liquid–solid interface between the materials. In addition, molecules close to the interface are less ordered with respect to one another than those far from the interface. These factors may benefit charge separation and transport, resulting in good device performance.

In general, atomistic simulations are not feasible for studying donor–acceptor interface formation for the typical domain sizes found in devices. A solution to this is to use coarse-grained (CG) models, which increases the simulation efficiency by replacing a collection of atoms as a single interacting site. In Chapter 4, a new systematic methodology to generate CG models for MD simulations is introduced and validated, which constitutes the main result of this thesis. This algorithm is developed so that MD simulations can be simplified but still accurately represent the physical and thermodynamic properties of the simulated materials. More importantly, this method can produce models that capture the anisotropy of molecules, which is especially useful for theoretical studies of organic materials and has not previously been achieved via a systematic algorithm. To validate the method, a CG model of a simple anisotropic organic molecule (benzene) is produced in Chapter 5. Simulations using this model accurately describes the structural and thermodynamic properties of the FG model and is an improvement over previous CG benzene models.

A future application of this method will be the study of the interface structure of materials in OPV systems on realistic time and spatial scales compared to experimental conditions. Ultimately, the studies presented in this thesis work towards the same goal, which is to discover optimal molecular design rules to increase the power conversion efficiency of OPVs.

Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint award of this degree.

I give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship.

Acknowledgements

There are many people who have contributed to make this thesis possible that I want to acknowledge in this section. First and foremost, I want to say thanks to my supervisor Assoc. Prof. David M. Huang for sharing his knowledge and inspirations in doing research, and being thorough and patient when it comes to reading my writing and fixing my grammar. Thank you for giving me opportunities to direct parts of my research projects to my own interests, and for your trust and support at all time. I also want to say thanks to Assoc. Prof. Tak Kee for his occasional share of knowledge and wisdom.

The work in this thesis would have not been carried out without the help from Dr. Lewis Mitchell, who have suggested mathematical algorithms to complete the ideas that we have come up with. Thank you!

I am also grateful for the help I have received from many past and current members of the Huang research group. I want to say thanks to Andrew Tarzia who taught me how to use a computer when I first started my MPhil. Thanks to you that I am a better computer user now, and thank you for all of your help and share, and support. Thanks Belinda Boehm for your help, share of knowledge, research aspects, little cute things and delicious food. Without you going to the office every day would be less interesting. Also, I want to say thanks to the Kee research group members for sharing incredible time with me and our group members.

Thank you my friends in the Lucid Dream Music Club, who have driven me out of the office and help me express and develop my interest in music. Thanks for the time and occasions you have shared with me which have meant a lot to me. Also thanks to my friend in Viet Nam or overseas, you are my incredible people!

To my family. Mum and Dad - you have trusted and given me so much love that helped me go through all difficulties in my life and research work. Without you I would have never achieved anything in my life, and the little time we are together every year always brings me so much joy and happiness. Thank you my little sister, Nguyen, for the love and care you give to me. Thank you my boyfriend Jordan, you have been so patient and willing to help me go through so many hard times. I am incredibly lucky to have such a supporting and caring person in my life and I wouldn't have made it this far without you. Thank you to Mary and Greg too, and Tofu the fat cat, I feel like I have a family when I am with you. Thanks to you all!

Dedication

For my beloved mum and dad
For my little sister
And for Jordan.

1

Introduction

"Imagination is more important than knowledge. For knowledge is limited, whereas imagination embraces the entire world, stimulating progress, giving birth to evolution."

– Albert Einstein

With the rapid increase in the world population, global energy demand is increasing rapidly. The primary source of energy is fossil fuels (coal, oil and gas),[1] which release greenhouse gases and toxic gases, contributing to global climate change and pollution, leading to natural disasters and health problems that threaten the lives and livelihoods of humans and other species. Moreover, the inequitable distribution of fossil fuels has created conflicts between countries over areas rich in these energy resources. Resolving climate change and other problems caused by the use of fossil fuels has become a great challenge, requiring an alternative large source of energy to be found and used.

One of the best candidates is energy from the Sun, which is available almost everywhere on Earth. Transforming solar energy to electricity was indeed successful and first commercialized in the 20th century, using inorganic materials that can generate electric current under sunlight. Although these silicon-based solar panels have been refined and improved to obtain relatively high energy-conversion efficiency, solar energy still only contributes to a small amount of world energy usage, partly due to high manufacturing and maintenance costs required for these panels.

Organic semiconductors including carbon-based small molecules and polymers have been increasingly studied in the last two decades, and have been introduced to the society as efficient and practical materials particularly for electronic display devices.[2,3] These organic materials can be economically advantageous over inorganic materials as organic devices can be produced using simple and low-temperature processes,[4] reducing the manufacturing costs. In addition, organic materials are also lightweight and can be used in tunable,[5] large-area and flexible devices,[6] which are more convenient and require less cost for supporting materials and installation than photovoltaics based on inorganic materials. Successful applications of photovoltaic devices will promote the use of the unlimited solar energy resource instead of fossil fuels, providing promising solutions to environmental and global energy problems.

1.1 Organic semiconductors

Organic semiconductors are non-metallic materials that possess semiconducting properties. These materials can be polymers or small molecules, which consist mostly of carbon and hydrogen atoms, and often together with a smaller proportion of heteroatom(s) such as sulfur, oxygen, nitrogen, and halogens. For organic photovoltaics (OPVs), polymers are often chosen as they can form smooth thin films when deposited from solution. However, small-molecule organic semiconductors have also drawn attention from researchers in this field.[7,8]

Semiconducting properties result from the extensive conjugated backbone of organic semiconductors. Alternating single and double carbon-carbon bonds allow electrons to be delocalized and transferred along the conjugated sections. This conjugated π system also lowers the band gap between the occupied and unoccupied molecular orbitals of the materials, leading to excitation of electrons upon photon absorption typically in the visible part of the electromagnetic spectrum.

In OPV devices, blends of organic semiconductor materials are deposited onto the device substrate in solvents. Methods such as spin coating and inkjet printing are normally used, followed by solvent evaporation to yield phase-separated domains. Even though fabrication is simple and cost-effective, processing from solution allows the materials to be mobile before formation of solid films. The device performance largely depends on the microstructures formed within the films, which may not be ideal for electrons to be delocalized and transferred efficiently. Such conditions create challenges to systematic control of the performance of the devices, but these challenges may also allow opportunities for better development and optimization. Studies of factors that manipulate these properties are needed in order to make organic semiconductors to become a suitable alternative to their inorganic counterparts.

1.2 Bulk-heterojunction organic photovoltaics

The generation of an electric current and voltage in semiconductor materials upon exposure to light is known as the photovoltaic effect, which was observed and studied from the 19th century.[9] From then, the development of photovoltaic devices has made significant achievements in both theoretical and practical aspects, in particular the discovery of inorganic silicon-based solar cells, with the efficiency increasing from less than 1% to approximately 30% today.[10] Organic solar cells have been studied from the 20th century, with the first investigation of photovoltaic effect of some organic dyes in the 1950s[9] and the use of conjugated polymers in devices in the 1980s with power conversion efficiency far below 1%.[9] The idea of putting an electron donor and an electron acceptor in the photo-active layer of an organic solar cell was first introduced by Tang in 1986, which effectively increased the efficiency to 1%.[11]

Today, bulk-heterojuntion solar cells are of most interest, in which the photo-active layer of an organic cell generally consists of a mixture of electron donor and acceptor materials that are phase-separated on the nano scale (Figure 1.1a). Absorption of a photon generates an exciton, which is a bound state of an electron and a hole. This state is more strongly bound in organic materials than in inorganic materials due to the significantly smaller dielectric constant of organic materials.[13] At the interfaces between the domains of different materials, an exciton can be separated into free charges, driven by the difference in energy between the highest occupied molecular orbital of the donor and the lowest unoccupied molecular orbital of the acceptor (Figure 1.1b). This creates flows of charges across the device to the electrodes to generate electricity. Typically, the bulk-heterojunction consists of a polymer electron donor material, and a fullerene-based electron acceptor. However, recently there have been studies on the use of small organic semiconductors as electron donors[14,15] and non-fullerene materials as acceptors.[16,17]

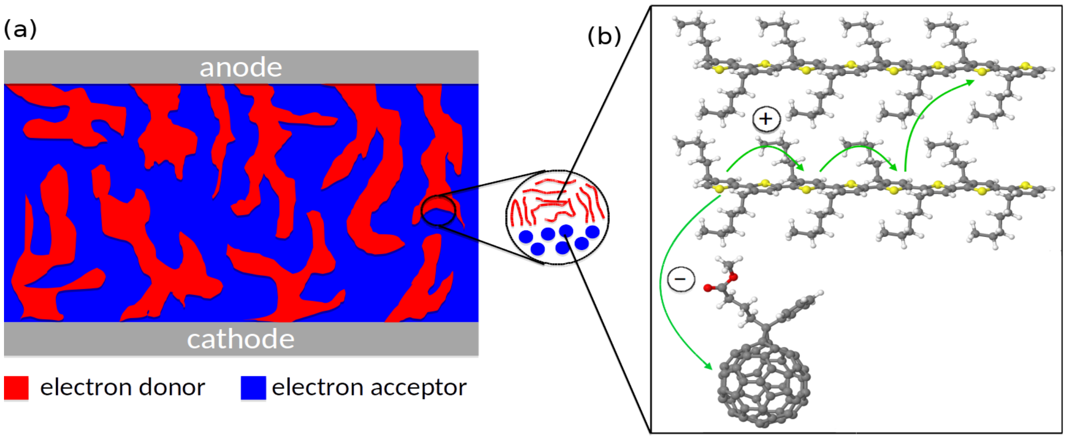


Figure 1.1: (a) Schematic of the structure of the active layer of a bulk-heterojunction solar cell, and (b) schematic of photo-induced charge separation at a donor–acceptor interface for a semiconducting polymer/fullerene (P3HT/PCBM) blend (two adjacent P3HT octamers and one PCBM are shown). Image taken from ref. [12].

1.2.1 Current state-of-the-art technology and challenges in the development of organic semiconductor solar cells

One of the major obstacles to more widespread adoption of organic semiconductor solar cells is the low efficiency compared to conventional inorganic devices. There have been a large number of studies on improving the device performance of organic semiconductor materials. Different aspects such as the choice of materials used (varying from small molecules[18–20] to polymers[21–23]), the chemical structures of component molecules,[24,25] and addition of doping or functional layers[26] in the devices have been investigated widely and have achieved significant device performance improvements. For organic solar cells particularly, the energy-conversion efficiency has improved from less than 1% to more than 10%[10] over the previous decade. However, this is still less efficient compared to inorganic devices, with efficiency close to 30%.[10] Many other promising applications of organic semiconductors have not been made practical due to the lack of general optimization of the materials and manufacturing processes, which requires a complete understanding of controlling the electronic properties of the materials.

A challenge in optimizing OPV devices is that the device performance can not be predicted directly from the chemical structure of the constituent materials, due to the complicated microstructures that may be formed in the active layer of the devices. The variability in the microstructure of organic semiconductor films results from relatively weak intermolecular non-covalent interactions between the materials, which may impact the device performance in a way that is not easy to predict. This is opposite to inorganic semiconductor devices, in which the inorganic semiconductor materials in most cases have ordered crystal structures that allow accurate predictions of the device electronic properties. In particular for OPV devices, variations in the interface structures between constituent materials makes it hard to generalize how to predict the efficiency of charge generation that happens at theses interfaces. Several proposed mechanisms have emerged to explain the charge separation phenomenon, by which several different physical phenomena are argued to play a role in driving the charge separation process. These include morphological variations that induce variations in electronic-state energies at the donor–acceptor interface,[27,28] the role of excess electronic[29] or vibrational energy,[28,30–32] intra- and intermolecular charge delocalisation,[29,33] and the role of quadrupole moments and quadrupole field at the interface.[34] These factors are all affected by the structure at the donor–acceptor interfaces where charges are created. It is necessary to reconcile the different proposed mechanisms by accurately studying interface structural variations, and their impacts on electronic processes, and to determine how the structure and resultant electronics can be optimized.

Although these interfaces are clearly important for understanding electronic processes in organic semiconductor-based devices, experimental characterization of them is generally hard, as they are normally embedded in the active layer. As a consequence, optimization of organic semiconductor devices is often conducted by trial-and-error, in which different materials and conditions are tested until the devices obtain their desired properties and efficiency. Thus, more predictive approaches are necessary, which can be achieved with theoretical scaffolds using computer simulations and calculations. Ideally, the computational approaches should produce models that are capable of realistically describing dynamic and electronic properties of organic materials within the devices, allowing accurate calculations that potentially lead to the optimization of the devices. The work presented in this thesis is entirely theoretical and computational, with the goal to develop computational methods that clarify the structure of organic semiconductors.

1.2.2 The effects of molecular anisotropy on the performance of organic semiconductor devices

Molecular anisotropy, in terms of shape and interactions, is a general and important property of almost all organic semiconductor molecules. In particular, the effect of molecular anisotropy at an interface between two different materials can lead to symmetry breaking and induce preferential orientations of the molecules at the interface.[35,36] An example of this is shown in Figure 1.2, with liquid crystals arranging differently with respect to the interface due to molecular anisotropy. Computer simulations of previous studies have shown either parallel, perpendicular, or other intermediate alignments of model rod-shaped molecules with respect to an interface. These include interfaces with a vapour,[37] a solid substrate[38] or liquid of spherical molecules,[39] depending on the particular molecular anisotropy and interface. Experimental results have also suggested ordered layers of organic compound can be formed using an ordered inorganic template,[40] and the level of ordering at the interface can be higher than in bulk materials.[41] The molecular anisotropy thus can be used to control the structural order at the interfaces in organic semiconductors, which have not been studied systematically.

In addition, for organic electronics in which multiple organic semiconductors can be used, ordering of component molecules at interfaces is extremely important for the device performance. In particular, for OPVs, it has been shown that ordering at the donor–acceptor interfaces and material crystallinity have a great impact on separation of charges and charge transport to the electrodes. A study by Tumbleston et al.[42] using resonant soft X-ray scattering confirmed that the variations in device performance can be correlated to the degree of orientation at the interfaces between the electron donor and acceptor materials. In particular, device performance is improved for polymer–fullerene systems with a face-on arrangement of the aromatic rings of the polymer at material interfaces, yet the correlation can vary depending on the molecular structures of the components. This is also the case for bulk-heterojunction systems of small-molecule donors that have been studied, with the reason behind this phenomenon under debate.[29,43,44] A recent study by Jakowetz et al.[41] using an optical time-resolved method shows that at the active donor–acceptor interfacial sites the en-

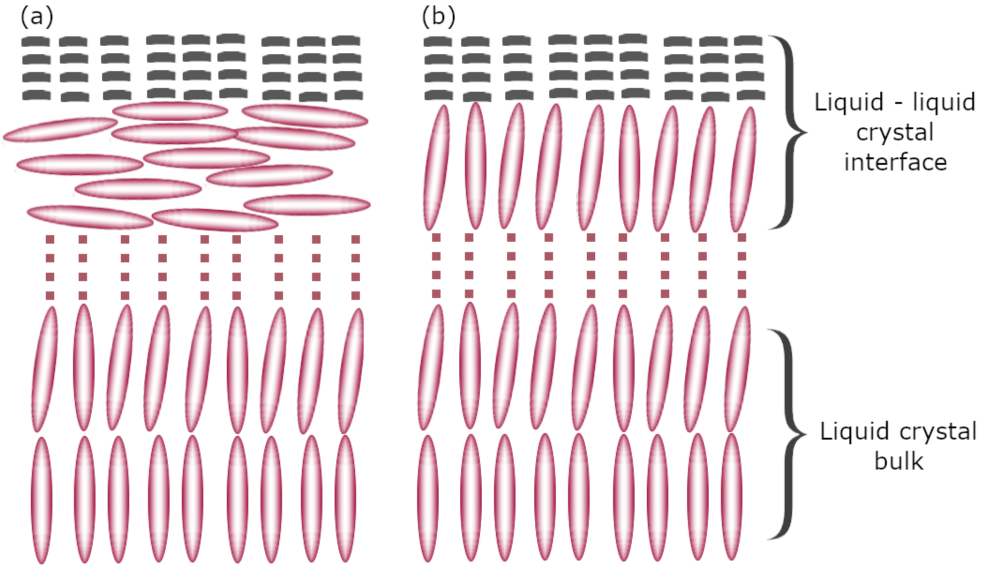


Figure 1.2: At a liquid–liquid crystal interface, liquid crystal molecules can (a) align with the interface, or (b) arrange perpendicular to the interface. Here rod-shaped liquid crystal molecules are shown in pink, aqueous liquid is shown in black.

ergy barrier between charge–transfer and charge separated states is reduced, possibly due to a more ordered interfacial structure, which may be key to efficient charge separation. Also, high crystallinity in the bulk of each material is good for charges to be transported to the electrodes.[28,45] Nevertheless, there are also studies that have shown a disordered interface is preferrable for good charge generation, and ordered interfaces should be avoided.[28,29,46]

1.2.3 A proposed systematic study of using molecular anisotropy to control interface structure and device efficiency

As described above, it is clear that interfacial order or disorder and interfacial alignment of materials can play an important role in the performance of OPV devices. In Chapter 3, the interface structure of a high-performing bulk-heterojunction system will be studied using computational modeling. The system consists of benzodithiophene quaterthiophene rhodanine (BQR) as the electron donor, and a fullerene derivative [6,6]-phenyl-C71-butyric acid methyl ester ($PC_{71}BM$) as the electron acceptor (Figure 1.3), with a reported un-optimized efficiency of 9.4%.[47] This system is chosen because the morphology of these molecules has been characterized and they are known to exhibit the effect of molecular anisotropy, which can result in interfacial characteristics that facilitate charge generation. More importantly, BQR and $PC_{71}BM$ also have many other structural analogues[47,48] which yield different morphological and device properties, allowing systematic testing of the role of molecular shape on interfacial order and electronic properties.

In fact, a study by Geraghty et al.[47] has tested and shown a large variation in the power conversion efficiency between systems of BXR–$PC_{71}BM$, where BXR can be BQR or its analogues in Figure 1.3. These efficiencies are presented in Table 1.1 for the devices fabricated using each different donor material. From the data, it can be seen that increasing the backbone length from BMR to BQR significantly improves the device performance, but BPR with two thiophene rings longer than BQR yields a less efficient device. With the molecules being similar in chemical moieties, OPV devices containing these donor molecules have similar light absorption ranges[47] and interactions with the acceptor molecule. Therefore, it is

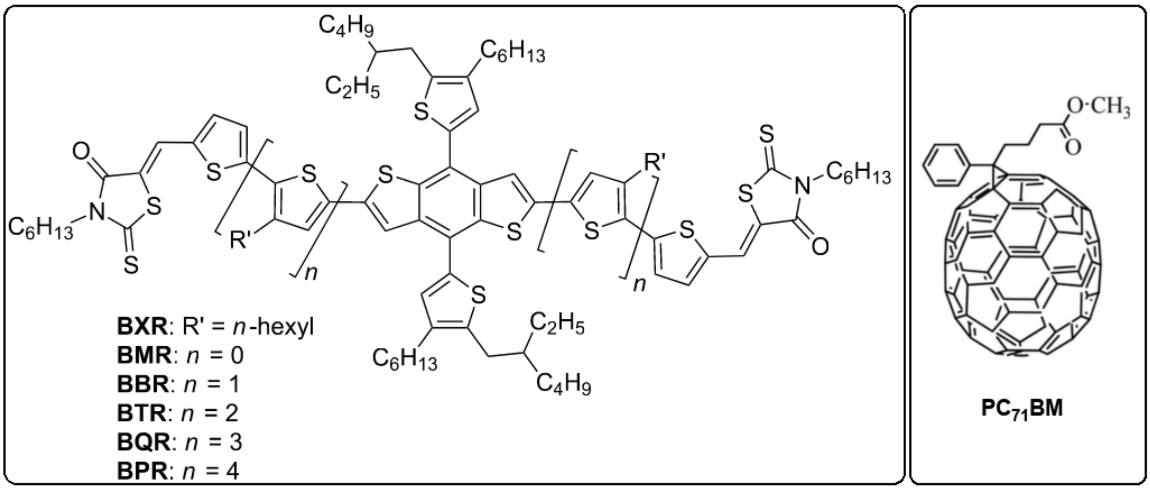


Figure 1.3: Chemical structures of electron donors BQR and analogues (left), and electron acceptor $PC_{71}BM$ (right).

of the most interest to study how variations in molecular shape, in particular the conjugation length for the BXR molecules, influences the structure of the donor–acceptor interface, and possibly within the bulk, so that the electronic processes are consequently affected.

Table 1.1: Photovoltaic performances of BXR–$PC_{71}BM$ BHJ solar cells using solvent vapour annealing, adapted from ref. [47]

Electron Donor	BMR	BBR	BTR	BQR	BPR
Efficiency (%)	3.5	6.0	9.3	9.4	8.7

Chapter 3 includes the work on atomistic molecular modeling of only the BQR–$PC_{71}BM$ interfaces, but the model can be reapplied to analogues of BQR readily in order to understand fully the different morphological properties and their effects on important electronic processes in OPVs. Ideally, studies of these systems, together with computational calculations of charge dynamics, will help reveal the unknown correlation between material variation and device performance, and contribute to develop design rules to optimize organic solar cells.

1.3 Molecular dynamics simulation at interfaces of organic semiconductors

1.3.1 Coarse-grained molecular dynamic simulation

To date, computational modelling and simulations of electronic properties of organic semiconductors have not considered molecular anisotropic effects at the interfaces in a general way. Most models neglect the effects of molecular anisotropy on electronic properties of organic semiconductors, assume the interfaces between the materials are either structureless,[49] or set up specific initial structures,[27] or use atomistic models for the interfaces,[33] which made their outcomes hard to generalize for all systems. Also, computational studies using atomistic models are expensive in terms of time and computer power, limiting the size of the system being studied typically to less than 100,000 atoms and time scales to typically less than a microsecond. This is insufficient for studying meso-scale system, particularly in this case in which time scales needed for interfacial formations and length scales needed for structural variations, important factors for interfacial electronic phenomena, are required.

Therefore, mesoscale methods such as coarse-grained (CG) molecular models (Figure1.4) are a good alternative. CG molecular models integrate out atomistic details but realistically describe molecular shape and connectivity, addressing the challenges of providing sufficiently large-scale systems. It can be seen from Figure 1.4, that in a CG model, many atoms that rigidly move together are grouped into one CG particle/site. Modeling of these sites can be orders of magnitude faster, depending on the number of atoms in each CG site. Results from simulations of CG models are also easier to interpret, as the degrees of freedom in the system is significantly reduced. This provides a scaffold to experimental studies on the optimisation of organic electronic devices. Other studies that have used CG molecular models have shown that the models can accurately describe structural,[50] mechanical,[51] and surface adsorption[52] properties which were not parametrized in the models, proving their capability for providing accurate results in a large range of applications.

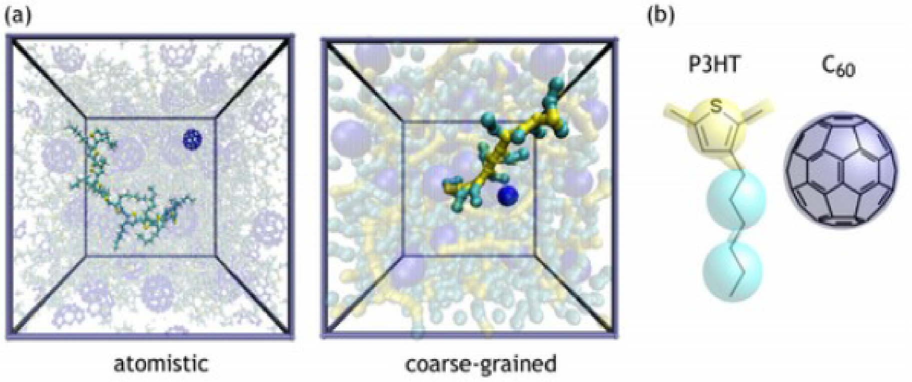


Figure 1.4: (a) Example of small-scale atomistic and CG molecular dynamics simulations used to parametrize a coarse-grained model for a semiconducting polymer/fullerene (P3HT/C60) blend (one P3HT dodecamer and C60 molecule are highlighted). (b) Coarse-graining scheme for P3HT and C60. Image taken from ref. [12].

From here on, the atomistic description of a molecular model is referred to as fine-grained (FG) model, due to its high level of detail. The FG model and simulations of the BQR–$PC_{71}BM$ interfaces in Chapter 3 could still provide useful information of the realistic device structures. However, the ultimate goal in the future is to develop a larger scale model and simulation for this system, as well as its analogous systems described in section 1.2.3. The model and simulation results in Chapter 3 can be used as the reference during the development and validation of a CG model for this system, that will be developed in later projects.

1.3.2 Coarse-grained methodology for molecules with anisotropy

There are a number of "bottom-up" methods that have been commonly used to parametrize CG models from FG descriptions. One of the most intuitive and simplest methods, is the iterative Boltzmann inversion (IBI) method. The IBI method is based on an iterative optimization algorithm to generate probability distributions of intra- and intermolecular structural properties, such as non-bonded pair, bond, or angle distribution functions, that match those produced in a corresponding FG model.[53,54] The method has been successfully applied for systems of liquid and polymer, and biological models.[55] However, in some cases convergence of the CG model can be difficult if there are a lot of complex couplings between interactions and insensitivitity of structural correlations to the CG model.[56,57] This is most likely because the IBI method is not rigorously derivable from statistical mechanical theory. A more rigorous approach is the multi-scale coarse-graining (MSCG) method, a method based on matching of CG and FG forces induced by many-body potential of mean forces (PMFs) in molecular dynamics (MD) simulations.[58,59] The method has also been extensively developed and used to study liquids[60] and biological systems.[61,62] Recently, a CG method based on minimizing the relative entropy of the CG and FG models was introduced,[63] which has been shown to be comparable to the other methods above. Details of these CG "bottom-up" methods together with "top-down" methods, have been reviewed elsewhere.[56,64]

A common theme of these systematic methods is that the CG particles are assumed to have spherical shape, i.e. isotropic CG sites. The interactions in these models depend only on the distances between the CG sites, and orientations are neglected. This simplifies the CG parametrization process, yet may reduce the accuracy of the models. In particular, for CG models of organic semiconductors, many structural quantities will not be captured correctly using spherical particles, because they depend greatly on the directional π-stacking interactions between molecules. Moreover, due to the anisotropic nature of organic semiconductors, more CG sites must be used in spherical models to ensure the accuracy in the representation of shapes and interactions, reducing the speed of the simulations. Anisotropic models using ellipsoidal,[65,66] or disk-shaped[67] particles have been developed, but used non-rigorous and non-systematic methods that may have not been successful in accurately transferring the FG models to the CG level.

In Chapter 4, for the first time, a systematic and rigorous CG method based on force-matching of anisotropic CG particles is introduced, termed the anisotropic force-matching coarse-grained (AFM-CG) method. The method is based largely on the MSCG method for spherical sites,[58,59] with the addition of orientations in the theory and implementation. To validate the performance of the AFM-CG method, in Chapter 5, a disk-shaped model of liquid benzene is developed, with structural and thermodynamic properties compared with those of the FG model simulation. These properties are also compared to those from a similar CG model of benzene that has been developed using a non-systematic approach.[67] If optimized, the AFM-CG method can be used to generate CG models of the BQR–$PC_{71}BM$ and analogous systems described in section 1.2.3. This will allow investigation of the interface structures for these systems on realistic time and spatial scales compared to experimental conditions.

2

Computational methods

"There's plenty of room at the bottom."

– Richard P. Feynman

2.1 Molecular dynamics simulations

Equilibrium states of condense-phase molecular systems normally involve the collective behaviours of many atoms, with structural and dynamic properties resulting mostly from interactions on larger scales than quantum effects. The most readily available methods for simulating dynamics on this scale typically utilize algorithms in the category of classical molecular dynamics (MD). MD is a computational method based on a classical description of the interactions between particles in a system. Generally the particles represent atoms in molecules, in which case the system is a fine-grained (FG) model of the molecules. The physical movements of the molecular system are determined by numerically solving a Newton's equation of motion for each particle, in which forces between particles are defined by molecular mechanics force fields that define the total potential energy of the system. MD simulations can provide atomistic information which might be important to discover mechanistic details that are generally not easy to determine experimentally.

Nuclei are heavy enough to be considered as classical particles and so their movements can be determined by solving the Newton's equation of motion. Basically, for a system of n particles

$$\mathbf{F}_i(\mathbf{r}^n) = m_i \frac{d^2\mathbf{r}_i}{dt^2} = -\frac{\partial U(\mathbf{r}^n)}{\partial \mathbf{r}_i} \tag{2.1}$$

Here $\mathbf{F}_i$ is the force vector acting on particle i, $\mathbf{r}^n$ are position vectors of the n particles, m_i is mass of particle i located at $\mathbf{r}_i$, and U is the potential energy surface that describes interactions between the particles. This function U is also referred to as the force field in MD simulations.

For simulations in the NVE (constant number of particles, volume, and energy) ensemble, the force on each particle can be computed from the force field as described in Eq (2.1), which allows integration of particle positions over time and thus generates a trajectory. The timestep between each position propagation is chosen to be smaller than the fastest process in the system.[68] For molecular systems, the fastest processes are normally stretching vibrations that involve hydrogen, requiring a very small timestep to be used. Therefore, it is often useful to constrain bond lengths that involve hydrogen atoms to their equilibrium values which allows the use of a longer timestep. In this work, these bonds are constrained using the SHAKE algorithm,[69] allowing the use of 1–2 fs timestep for simulations of FG models.

2.1.1 Molecular mechanics force field

Defining the force field is important as this decides how molecular systems evolve over time. The force field is composed of bonded and non-bonded interactions that are specific to different simulation systems. In general, it can be written as

$$U_{\text{total}} = \underbrace{U_{\text{electrostatic}} + U_{\text{vdw}}}_{\text{non-bonded}} + \underbrace{U_{\text{bond}} + U_{\text{angle}} + U_{\text{dihedral}} + U_{\text{improper}}}_{\text{bonded}} \tag{2.2}$$

where $U_{\text{electrostatic}}$ and U_{vdw} are the non-bonded electrostatic (Coulombic) and van der Waals interactions, respectively, U_{bond} is the bond stretching potential, U_{angle} the bond angle potential, U_{dihedral} the dihedral angle potential, and U_{improper} the improper torsional potential.

In Chapter 3, FG models of interfaces between organic semiconductors are investigated using MD simulations. In these simulations, the OPLS-AA force field (version OPLS2005)[70] is used for van der Waals, bond stretching, angle bending, and improper potentials, because it has been found to give reasonably accurate overall structure and energy for organic semiconductors.[71] The dihedral potentials are calculated from high-level quantum calculations, together with atomistic partial charges that define the electrostatic interactions, following the parametrization method of the OPLS-DA force field developed by Jackson et al.[72] that is described in more details in Chapter 3. The additional parameters to the OPLS2005 force field were shown to improve the accuracy of ordering and aggregation structures for conjugated organic semiconductors,[72] which are important for electronic processes in organic photovoltaic (OPV) devices. The OPLS2005 force field was applied for FG model of benzene in Chapter 5 without modification.

2.1.2 Equilibrium simulations

There are a few common simulation techniques that are used in typical MD simulations. Even though molecular models can contain a lot of atoms, their sizes are still far from realistic systems due to limitations on computational time and storage. At the start of a simulation, the coordinates of the simulated particles are defined, and they are contained in a volume called the simulation box. To prevent the outer molecules being pushed out of the simulation box, and minimizing surface effects, periodic boundary conditions are often used. Here, the simulation box is replicated infinitely in all directions to produce a continuous medium. A schematic of this in two dimensions is shown in Figure 2.1, with the simulation box at the center, surrounded by its periodic images. When periodic conditions are applied, generally short-range non-bonded interactions are calculated only between the closest images of two molecules. This is known as the minimum image convention, which greatly reduces the number of interactions that need to be calculated while still giving a good approximation of the bulk environment. In general, the size of the system must be large enough so that the simulation box dimensions are larger than two times the correlation length of the system to avoid finite size effects.

While van der Waals interactions are normally effective mostly at short distances, that is their decay rates are large enough that the interactions can be truncated at a finite cutoff distance without significant error, electrostatic interactions have a much longer range, and cannot be truncated the same way without producing a large error to the calculated thermodynamics. Therefore, an alternative method must be used for calculating the electrostatics for infinite periodic systems. In the MD simulations in Chapter 3 and 5, the particle-particle particle-mesh (PPPM) method[73] was used for this purpose. This method maps atom charge to a mesh

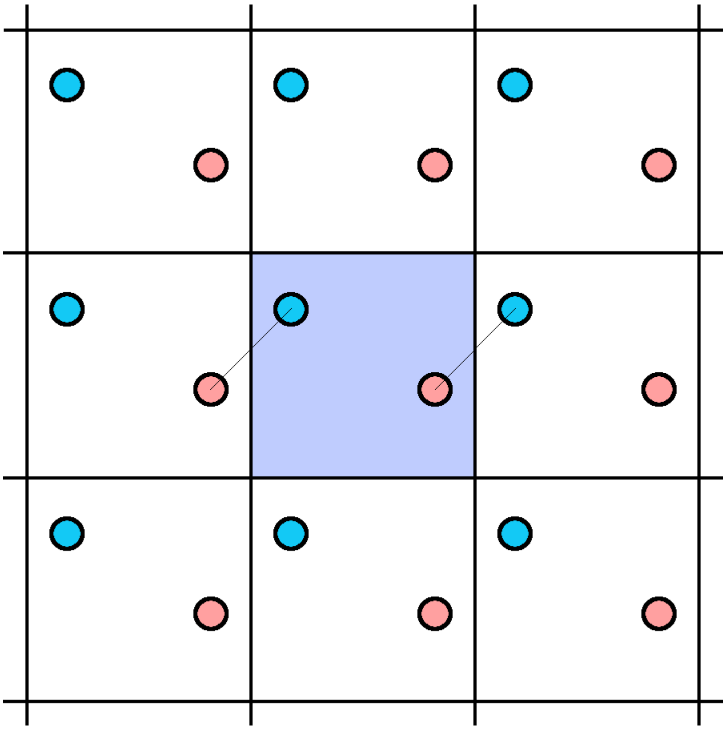


Figure 2.1: Periodic boundary conditions in two dimensions. The true simulation box (colored) is surrounded by its image boxes in all dimensions. Short-ranged non-bonded interactions are calculated between atoms within the cutoff distance, regardless of which box they belong to (examples of these interactions shown by the dotted lines).

in three dimensions and uses fast Fourier transform to evaluate electrostatic forces applied on the atoms beyond the cutoff distance. Within the cutoff, the electrostatics are calculated between pair of molecules explicitly.

The thermodynamics of the simulated systems are controlled by specific methods. MD simulations in this work involve relaxing the system configuration in the canonical (NVT) ensemble, and obtaining equilibrium in the isothermal-isobaric (NPT) ensemble. In the NVT ensemble, the number of atoms N, the system volume V and the temperature T are constant, while in the NPT ensemble, the pressure P is constant instead of the volume. A Nosé-Hoover thermostat and barostat were used to enforce the conditions on the temperature and pressure, in which additional ficticious dynamical variables are added to the system to simulate volume and energy fluctuations respectively, consistent with the specified thermodynamic ensemble.[74]

2.2 Coarse-grained molecular dynamics

In many cases, MD simulations are carried out using FG models, in which all atoms in the systems are defined explicitly. Force fields for these all-atom models are well-developed for different types of molecular systems, allowing accurate simulation results to be obtained. Nevertheless, in general, computational resources allow the size of these models up to 1,000,000 atoms, and simulation time up to a few hundred nanoseconds. A solution to this is to simplify the systems, using their coarse-grained (CG) representations. Beside reducing the number of particles, a CG system also allows a longer integration timestep due to elimination of rapid motions in molecular systems, such as internal vibrations of C-H bonds. MD simulations can be a lot faster with the same computational resources.

In Chapter 5, a CG model is applied to obtain equilibrium structure and thermodynamic properties of liquid benzene at specific conditions. MD simulations of this model were completed with periodic boundary conditions in NVT and NPT ensembles as described in the previous section of this chapter. The CG force field for this CG model was obtained using the anisotropic force-matching coarse-grained (AFM-CG), which is the main result of this work, presented in Chapter 4. The method was developed for anisotropic molecular models using force-matching of CG interaction forces to those in the FG model. Coarse-graining by force matching was successfully implemented for spherical CG sites by Izvekov and Voth,[75] which was termed the multi-scale coarse-graining (MSCG) method. In the remaining part of this section, the theory and basic algorithms of the MSCG method for spherical sites will be recapped, mostly based on the theoretical foundations that have been introduced previously,[75] which will provide a framework for how the AFM-CG method for anisotropic models is developed in Chapter 4.

2.2.1 The theory of force matching coarse-graining for spherical models

Statistical mechanics allows one to study physical systems that have a large number of degrees of freedom, explaining the thermodynamic behaviour of these systems. It can be shown in statistical mechanics that all equilibrium properties of a molecular system can be derived if both the intermolecular potential energy and the phase-space distribution function are known.[76] The consistency of a CG model with the FG model of the same system is built based on these conditions, together with an approximation of classical rigid-body CG sites (see appendix B). Basically, a CG model is "consistent" with a FG model of the same system, if (1) each coordinate and momentum of a CG site are defined as linear combinations of the coordinates and momenta of a group of atoms in the FG model, and (2) the equilibrium phase-space distribution function of the CG coordinates and momenta is equal to that of the FG model.[75] Below, molecular and statistical mechanical quantities used in the FG and CG models of a molecular system will be introduced, followed by derivation of conditions needed for consistency between the two models.

Definition of the fine-grained and coarse-grained models and their statistical mechanical quantities

The following defines the molecular descriptions of the FG and CG models, together with quantities that are important to the thermodynamics in both models such as the intermolecular potential energy functions and the distribution functions.

Consider a FG model at a instantaneous dynamical state, defined by the coordinates $\mathbf{r}^n = \{\mathbf{r}_1, ..., \mathbf{r}_n\}$ and momenta $\mathbf{p}^n = \{\mathbf{p}_1, ..., \mathbf{p}_n\}$ of n component atoms. The Hamiltonian h that describes the energy of all of these atoms, consists of kinetic and potential parts, respectively expressed in the following equation:

$$h = \sum_{i=1}^{n} \frac{\mathbf{p}_i^2}{2m_i} + u(\mathbf{r}^n) \tag{2.3}$$

where m_i is mass of atom i in n atoms, and $u(\mathbf{r}^n)$ is the potential energy operator of the system. Note that momentum is defined as $\mathbf{p}_i = m_i \dot{\mathbf{r}}_i$.

The phase-space distribution function $P(\mathbf{r}^n, \mathbf{p}^n)$ of atom positions and momenta is related to the Hamiltonian as

$$P(\mathbf{r}^n,\mathbf{p}^n) \propto \exp(-\frac{h}{k_\mathrm{B}T}) = \exp(-\sum_{i=1}^{n}\mathbf{p}_i^2/2m_ik_\mathrm{B}T)\times\exp(-\frac{u(\mathbf{r}^n)}{k_\mathrm{B}T}) \tag{2.4}$$

where k_B is the Boltzmann constant and T is the absolute temperature.

As shown in Eq (2.4), the distribution function can be factorized into a translational part that depends on the momenta, a the configuration part that depends on the position coordinates:

$$P(\mathbf{r}^n,\mathbf{p}^n) = P_\mathrm{t}(\mathbf{p}^n)P_\mathrm{c}(\mathbf{r}^n) \tag{2.5}$$

where

$$P_\mathrm{t}(\mathbf{p}^n) \propto \exp(-\sum_{i=1}^{n}\mathbf{p}_i^2/2m_ik_\mathrm{B}T) \tag{2.6}$$

$$P_\mathrm{c}(\mathbf{r}^n) \propto \exp(-u(\mathbf{r}^n)/k_\mathrm{B}T) \tag{2.7}$$

If considering a CG representation of this model consisting of N CG sites, the Hamiltonian and the distribution function of the same dynamical state can be expressed in a similar way. For this CG representation capitalized symbols are used to emphasize the similarity in the derivation to those in the FG model. The CG Hamiltonian is

$$H = \sum_{I=1}^{N}\frac{\mathbf{P}_I^2}{2M_I} + U(\mathbf{R}^N) \tag{2.8}$$

where the coordinates $\mathbf{R}^N = \{\mathbf{R}_1,...,\mathbf{R}_N\}$ and momenta $\mathbf{P}^N = \{\mathbf{P}_1,...,\mathbf{P}_N\}$ of N component sites define the instantaneous state, M_I is mass of site I, and $U(\mathbf{R}^N)$ is the CG potential energy operator. Note it is assumed that the potential energy depends only on the positions of the CG sites, giving an isotropic (spherical) CG model. In Chapter 4, the potential energy depends on both positions and orientations of CG sites, introducing anisotropy into the model.

The CG distribution function is related to this Hamiltonian, similar to the FG model stated in Eq (2.4), and so can also be factorized into kinetic and potential components. Note that the symbols used for a momentum vector $\mathbf{P}$ and the phase-space distribution function P should not be confused, with the former labeled in bold as a vector quantity, and the latter in normal text as a scalar quantity.

$$P(\mathbf{R}^N,\mathbf{P}^N) = P_\mathrm{t}(\mathbf{P}^N)P_\mathrm{c}(\mathbf{R}^N) \tag{2.9}$$

where

$$P_\mathrm{t}(\mathbf{P}^N) \propto \exp(-\sum_{I=1}^{N}\mathbf{P}_I^2/2M_Ik_\mathrm{B}T) \tag{2.10}$$

and

$$P_{\mathrm{c}}(\mathbf{R}^N) \propto \exp(-U(\mathbf{R}^N)/k_{\mathrm{B}}T) \tag{2.11}$$

Bridging fine-grained and coarse-grained models

A CG model must satisfy certain requirements to produce thermodynamic properties consistent with a FG model. The first consistency requirement is that the coordinates and momentum of each CG site are defined as linear combinations of the coordinates and momenta of a group of atoms in the FG model. For instance, for a spherical CG model, the position of a CG site I can be derived as a linear combination of n atoms in the FG model via a mapping operator:

$$\mathbb{M}_{\mathbf{R},I}(\mathbf{r}^n) = \sum_{i=1}^{n} c_{I,i}\mathbf{r}_i \tag{2.12}$$

with $\sum c_{I,i} = 1$ to ensure displacement of the atoms involved results in the same displacement of CG site I.

Similarly, mapping of momenta is via a momentum mapping operator:

$$\mathbb{M}_{\mathbf{P},I}(\mathbf{p}^n) = \sum_{i=1}^{n} c_{I,i}\mathbf{p}_i \tag{2.13}$$

The second consistency requirement is the equilibrium phase-space distribution function of the CG coordinates and momenta is equal to that of the FG model. This means the distribution function defined in Eq (2.9) is equal to the one defined in Eq (2.4), together with the mapping conditions in Eqs (2.12) and (2.13):

$$\begin{aligned} P(\mathbf{R}^N,\mathbf{P}^N) &= P(\mathbf{r}^n,\mathbf{p}^n) \\ &= \int d\mathbf{r}^n \int d\mathbf{p}^n P(\mathbf{r}^n,\mathbf{p}^n)\delta(\mathbb{M}_{\mathbf{R}}^N(\mathbf{r}^n)-\mathbf{R}^N)\delta(\mathbb{M}_{\mathbf{P}}^N(\mathbf{p}^n)-\mathbf{P}^N) \end{aligned} \tag{2.14}$$

In Eq (2.14), Dirac delta functions are used (δ function, see Appendix B) to enforce the mapping conditions specified in Eq (2.12) and (2.13). Substituting Eq (2.9) and Eq (2.4) into Eq (2.14) gives

$$\begin{aligned} P_{\mathrm{c}}(\mathbf{R}^N)P_{\mathrm{t}}(\mathbf{P}^N) &= \int d\mathbf{r}^n \int d\mathbf{p}^n P_{\mathrm{c}}(\mathbf{r}^n)P_{\mathrm{t}}(\mathbf{p}^n)\delta(\mathbb{M}_{\mathbf{R}}^N(\mathbf{r}^n)-\mathbf{R}^N)\delta(\mathbb{M}_{\mathbf{P}}^N(\mathbf{p}^n)-\mathbf{P}^N) \\ &= \int d\mathbf{r}^n P_{\mathrm{c}}(\mathbf{r}^n)\delta(\mathbb{M}_{\mathbf{R}}^N(\mathbf{r}^n)-\mathbf{R}^N) \int d\mathbf{p}^n P_{\mathrm{t}}(\mathbf{p}^n)\delta(\mathbb{M}_{\mathbf{P}}^N(\mathbf{p}^n)-\mathbf{P}^N) \end{aligned} \tag{2.15}$$

where

$$\delta(\mathbb{M}_{\mathbf{R}}^N(\mathbf{r}^n)-\mathbf{R}^N) = \prod_{I=1}^{N} \delta(\mathbb{M}_{\mathbf{R},I}(\mathbf{r}^n)-\mathbf{R}_I) \tag{2.16}$$

$$\delta(\mathbb{M}_{\mathbf{P}}^N(\mathbf{p}^n)-\mathbf{P}^N) = \prod_{I=1}^{N} \delta(\mathbb{M}_{\mathbf{P},I}(\mathbf{p}^n)-\mathbf{P}_I) \tag{2.17}$$

As shown the CG and FG distribution functions are both factorized into coordinate and momentum parts in Eq (2.14) and (2.15), and the following two relationships are apparent:

$$P_{\mathrm{c}}(\mathbf{R}^N) = \int d\mathbf{r}^n P_{\mathrm{c}}(\mathbf{r}^n)\delta(\mathbb{M}_{\mathbf{R}}^N(\mathbf{r}^n) - \mathbf{R}^N) \tag{2.18}$$

$$P_{\mathrm{t}}(\mathbf{P}^N) = \int d\mathbf{p}^n P_{\mathrm{t}}(\mathbf{p}^n)\delta(\mathbb{M}_{\mathrm{P}}^N(\mathbf{p}^n) - \mathbf{P}^N) \tag{2.19}$$

These relationships lead to expressions relating the potential energy and the momentum quantities in the FG and the CG models, using Eq (2.6),(2.7),(2.10), and (2.11):

$$\exp(-U(\mathbf{R}^N)/k_{\mathrm{B}}T) \propto \int d\mathbf{r}^n \exp(-u(\mathbf{r}^n)/k_{\mathrm{B}}T)\delta(\mathbb{M}_{\mathbf{R}}^N(\mathbf{r}^n) - \mathbf{R}^N) \tag{2.20}$$

$$\exp(-\sum_{I=1}^{N}\mathbf{P}_I^2/2M_I k_{\mathrm{B}}T) \propto \int d\mathbf{p}^n \exp(-\sum_{I=1}^{n}\mathbf{p}_i^2/2m_i k_{\mathrm{B}}T)\delta(\mathbb{M}_{\mathbf{P}}^N(\mathbf{p}^n) - \mathbf{P}^N) \tag{2.21}$$

Eq (2.20) and Eq (2.21) are conditions for thermodynamic consistency of the CG model to the FG model. Eq (2.20) indicates the CG potential function for a consistent CG model can be uniquely derived from the many-body potential of mean force (PMF) of the FG model. In the next step, these conditions will be interpreted further to give physical quantities for practical use of this theory.

Consistency in configuration space

Eq (2.20) shows how to determine a CG potential energy function for the model to be consistent with the FG model in configuration space. Taking natural logarithm of both sides of the equation and rearranging gives

$$U(\mathbf{R}^N) = -k_{\mathrm{B}}T\big(\ln g(\mathbf{R}^N)\big) + (\mathrm{const}) \tag{2.22}$$

where (const) was added as the result of the proportionality in Eq (2.20), and

$$g(\mathbf{R}^N) = \int d\mathbf{r}^n \exp\big(-u(\mathbf{r}^n)/k_{\mathrm{B}}T\big)\delta(\mathbb{M}_{\mathbf{R}}^N(\mathbf{r}^n) - \mathbf{R}^N) \tag{2.23}$$

The CG force vector field is readily determined as gradients of this potential $U(\mathbf{R}^N)$. The force vector exerts on a specific CG site I is

$$
\begin{aligned}
\mathbf{F}_{\mathrm{CG},I}(\mathbf{R}^N) &= -\frac{\partial U(\mathbf{R}^N)}{\partial \mathbf{R}_I} = -\frac{\partial\Big(-k_{\mathrm{B}}T\big(\ln g(\mathbf{R}^N)\big)+(\mathrm{const})\Big)}{\partial \mathbf{R}_I} \\
&= \frac{k_{\mathrm{B}}T}{g(\mathbf{R}^N)}\frac{\partial g(\mathbf{R}^N)}{\partial \mathbf{R}_I} = \frac{k_{\mathrm{B}}T}{g(\mathbf{R}^N)}\frac{\partial\Big(\int d\mathbf{r}^n \exp\big(-u(\mathbf{r}^n)/k_{\mathrm{B}}T\big)\delta\big(\mathbb{M}^N_{\mathbf{R}}(\mathbf{r}^n)-\mathbf{R}^N\big)\Big)}{\partial \mathbf{R}_I} \\
&= \frac{k_{\mathrm{B}}T}{g(\mathbf{R}^N)}\int d\mathbf{r}^n \exp\big(-u(\mathbf{r}^n)/k_{\mathrm{B}}T\big)\prod_{J\neq I}^{N}\delta\big(\mathbb{M}_{\mathbf{R},J}(\mathbf{r}^n)-\mathbf{R}_J\big)\frac{\partial}{\partial \mathbf{R}_I}\delta\big(\mathbb{M}_{\mathbf{R},I}(\mathbf{r}^n)-\mathbf{R}_I\big) \\
&= \frac{k_{\mathrm{B}}T}{g(\mathbf{R}^N)}\int d\mathbf{r}^n \exp\big(-u(\mathbf{r}^n)/k_{\mathrm{B}}T\big)\prod_{J\neq I}^{N}\delta\big(\mathbb{M}_{\mathbf{R},J}(\mathbf{r}^n)-\mathbf{R}_J\big)\frac{\partial}{\partial \mathbf{R}_I}\delta\Big(\sum_{c_{I,I}\neq 0} c_{I,i}\mathbf{r}_i-\mathbf{R}_I\Big)
\end{aligned}
\tag{2.24}
$$

In the third line of Eq (2.24), the delta function $\delta(\mathbb{M}^N_{\mathbf{R}}(\mathbf{r}^n)-\mathbf{R}^N)$ was factorized as in Eq (2.16). The partial derivative $\partial/\partial\mathbf{R}_I$ applies only to the delta function that involves $\mathbf{R}_I$, whereas the FG potential $u(\mathbf{r}^n)$ and delta functions mapping other sites J for $J\neq I$ can be taken out of the derivative. In the next line, $\mathbb{M}_{\mathbf{R},I}(\mathbf{r}^n)$ is substituted using Eq (2.12).

Using the chain rule on the delta function $\delta(\sum c_{I,i}\mathbf{r}_i-\mathbf{R}_I)$, the following holds for any atom k that has $c_{I,k}\neq 0$:

$$
\frac{\partial}{\partial \mathbf{R}_I}\delta\Big(\sum_{c_{I,i}\neq 0} c_{I,i}\mathbf{r}_i-\mathbf{R}_I\Big) = -\frac{1}{c_{I,k}}\frac{\partial}{\partial \mathbf{r}_k}\delta\Big(\sum_{c_{I,i}\neq 0} c_{I,i}\mathbf{r}_i-\mathbf{R}_I\Big)
\tag{2.25}
$$

In the MSCG theory paper,[58] a general definition of CG sites was used, in which atoms could contribute to several CG sites. Later it was stated that with specific intramolecular constraints in the FG model, it was convenient to assume each atom only contributed to one CG site. For simplicity, this condition is applied directly here with atoms contributing to site I defined as $i\in I$. Now note that $\sum c_{I,k}=1$ (the dummy variables k and i are equivalent) so multiplying the right hand site of Eq (2.25) to this cancels the denominator:

$$
\frac{\partial}{\partial \mathbf{R}_I}\delta\Big(\sum_{i\in I} c_{I,i}\mathbf{r}_i-\mathbf{R}_I\Big) = -\sum_{k\in I}\frac{\partial}{\partial \mathbf{r}_k}\delta\Big(\sum_{i\in I} c_{I,i}\mathbf{r}_i-\mathbf{R}_I\Big)
\tag{2.26}
$$

Substituting this in to the CG force equation (2.24) followed by integration by parts simplifying:

$$\begin{aligned}
\mathbf{F}_{\mathrm{CG},I}(\mathbf{R}^N) &= -\frac{k_{\mathrm{B}}T}{g(\mathbf{R}^N)} \int d\mathbf{r}^n \exp\big(-u(\mathbf{r}^n)/k_{\mathrm{B}}T\big) \prod_{J\neq I}^{N} \delta\big(\mathbb{M}_{\mathbf{R},J}(\mathbf{r}^n) - \mathbf{R}_J\big) \\
&\quad \times \sum_{k\in I} \frac{\partial}{\partial \mathbf{r}_k} \delta\big(\sum_{i\in I} c_{I,i}\mathbf{r}_i - \mathbf{R}_I\big) \\
&= \frac{k_{\mathrm{B}}T}{g(\mathbf{R}^N)} \int d\mathbf{r}^n \left(\sum_{k\in I} \frac{\partial}{\partial \mathbf{r}_k} \exp\big(-u(\mathbf{r}^n)/k_{\mathrm{B}}T\big) \right) \\
&\quad \times \prod_{J\neq I}^{N} \delta\big(\mathbb{M}_{\mathbf{R},J}(\mathbf{r}^n) - \mathbf{R}_J\big) \delta\big(\sum_{i\in I} c_{I,i}\mathbf{r}_i - \mathbf{R}_I\big) \\
&= -\frac{k_{\mathrm{B}}T}{g(\mathbf{R}^N)} \int \frac{1}{k_{\mathrm{B}}T} d\mathbf{r}^n \exp\big(-u(\mathbf{r}^n)/k_{\mathrm{B}}T\big) \prod_{J\neq I}^{N} \delta\big(\mathbb{M}_{\mathbf{R},J}(\mathbf{r}^n) - \mathbf{R}_J\big) \\
&\quad \times \delta\big(\mathbb{M}_{\mathbf{R},I}(\mathbf{r}^n) - \mathbf{R}_I\big) \sum_{i\in I} \frac{\partial}{\partial \mathbf{r}_i}\big(-u(\mathbf{r}^n)\big) \\
&= -\frac{1}{g(\mathbf{R}^N)} \int d\mathbf{r}^n \exp\big(-u(\mathbf{r}^n)/k_{\mathrm{B}}T\big) \prod^{N} \delta\big(\mathbb{M}_{\mathbf{R},J}(\mathbf{r}^n) - \mathbf{R}_J\big) \sum_{i\in I} \frac{\partial}{\partial \mathbf{r}_i}\big(-u(\mathbf{r}^n)\big)
\end{aligned} \tag{2.27}$$

The force on an atom i in the FG model is related to the potential as $\mathbf{f}_i = \frac{\partial}{\partial \mathbf{r}_i}\big(-u(\mathbf{r}^n)\big)$, and substituting in $g(\mathbf{R}^N)$ from Eq (2.23) yields

$$\begin{aligned}
\mathbf{F}_{\mathrm{CG},I}(\mathbf{R}^N) &= \frac{\int d\mathbf{r}^n \exp\big(-u(\mathbf{r}^n)/k_{\mathrm{B}}T\big)\delta\big(\mathbb{M}_{\mathbf{R}}^N(\mathbf{r}^n) - \mathbf{R}^N\big)\big(\sum_{i\in I} \mathbf{f}_i(\mathbf{r}^n)\big)}{\int d\mathbf{r}^n \exp\big(-u(\mathbf{r}^n)/k_{\mathrm{B}}T\big)\delta\big(\mathbb{M}_{\mathbf{R}}^N(\mathbf{r}^n) - \mathbf{R}^N\big)} \\
&= \langle \mathbf{F}_{\mathrm{FG},I}(\mathbf{r}^n) \rangle_{\mathbf{R}^N}
\end{aligned} \tag{2.28}$$

where $\mathbf{F}_{\mathrm{FG},I}(\mathbf{r}^n) = \sum_{i\in I} \mathbf{f}_i(\mathbf{r}^n)$ is the force in the FG model, and the angle brackets denote averaging of the function at thermal equilibrium over FG configurations in which each is consistent with a CG configuration. Eq (2.28) is the major result of the MSCG theory, expressing that, for a CG model to be consistent with the FG model in configuration space, the force acting on a CG site is equal to the sum of the forces acting on the atoms within that site in the FG model at every time in an equilibrium state.

Consistency in momentum space

Defining the CG model so that every atom belongs to one CG site, Eq (2.19) can be factorized out for each CG site I, and the CG site mass can be derived:

$$M_I = \left(\sum_{i\in I} \frac{c_{I,i}^2}{m_i} \right)^{-1} \tag{2.29}$$

This ensures the momenta are consistent for the two models.

Summary about consistency of spherical coarse-grained models

To summarize, the following conditions are required to transform a FG model to its CG description:

(1) The coordinates of each CG site are a linear combination of a set of FG atom coordinates including at least one atom, and each atom are defined to contribute to only one CG site and not the others.

(2) The CG potential is defined such that for each site, the CG force is equal to the average sum of the FG forces acting on the atoms involved in that site in a thermal equilibrium state in the equivalent configuration, as expressed in Eq (2.28).

(3) The CG masses are as defined in Eq (2.29).

2.2.2 Variational algorithm for matching coarse-grained to fine-grained forces for spherical models

Above, three conditions for which a CG model is consistent with a FG model were defined. Conditions (1) and (3) can be readily satisfied when one builds a CG model based on the atomistic description, whereas condition (2) requires mathematical algorithms that are able to produce the CG forces from the FG forces, which result from the PMF of many-body interactions in typical MD simulations. These FG forces $\mathbf{F}_{\mathrm{FG}}(\mathbf{r}^n))$ are readily computed from atomistic potentials during a simulation, and serve as input information for the coarse-graining process. The MSCG method[58] uses the variational principle to perform this task for spherical CG models, which aims to find a solution for the CG forces $\mathbf{F}_{\mathrm{CG}}(\mathbf{R}^N)$ in order to minimize the sum of squared residuals of the difference between the CG and FG forces:

$$\chi^2(\mathbf{F}_{\mathrm{CG}}) = \frac{1}{3N}\left\langle \sum_{I=1}^{N} |(\mathbf{F}_{\mathrm{FG},I}(\mathbf{r}^n)) - \mathbf{F}_{\mathrm{CG},I}(\mathbf{R}^N)|^2 \right\rangle \tag{2.30}$$

with the CG configuration $\mathbf{R}^N$ corresponds to the FG configuration $\mathbf{r}^n$ under the mapping rules of coordinates and masses.

In real MD simulations, the forces can result from different types of interactions, in which the potential energy has different forms. For simplicity, only one particular type of interaction is considered here, and defined so that the CG force $\mathbf{F}_{\mathrm{CG}}$ is a real continuous function. This does not affect generalizing the algorithm to the situation in which many CG interaction functions are included in the force field, as the resulting forces are sum of the component forces computed from these interaction functions.

In implementation of the MSCG method, a finite set of linearly independent vectors (called basis vectors) are defined to span a subspace of the CG force function, where the interaction potential and forces are non-zero. If there are N_{B} of these basis vector, this subspace $\mathfrak{F}_{\mathrm{CG}}$ of the CG force $\mathbf{F}_{\mathrm{CG}}$ can be written as

$$\mathfrak{F}_{\mathrm{CG}}(\mathbf{R}^N, \mathbf{\Omega}^N) = \sum_{b=1,\dots,N_{\mathrm{B}}} \lambda_b \mathbf{v}_b(\mathbf{R}^N, \mathbf{\Omega}^N) \tag{2.31}$$

where $\mathbf{v}_b$ is a basis vector with corresponding real coefficient λ_b. With this limitation on the force field space, Eq (2.30) becomes

$$\chi^2(\mathfrak{F}_{\mathrm{CG}}) = \chi^2(\boldsymbol{\lambda}) = \frac{1}{3Nt}\sum_{I=1}^{N} |(\mathbf{F}_{\mathrm{FG},I}(\mathbf{r}^n)) - \mathfrak{F}_{\mathrm{CG},I}(\mathbf{R}^N, \boldsymbol{\lambda})|^2 \tag{2.32}$$

where the angle brackets have been replaced with averaging over all t recorded FG simulation snapshots. Here, if the CG coordinates $\mathbf{R}^N$ are defined based on the FG coordinates $\mathbf{r}^n$ and the consistency conditions, the residual quantity can be written as a function of $\boldsymbol{\lambda}$, the set of coefficients of the basis functions constructing the CG force field subspace in Eq (2.31). In the variational principle, if the amount of FG data is sufficient, determining the coefficients to minimize $\chi^2(\boldsymbol{\lambda})$ can be achieved using regression algorithms such as least squares approximation. The desired result from this is a force field function that can reproduce the FG many-body PMF in its CG representation, given the FG simulation information and the basis set.

In practice, the CG PMF is assumed to be contributed by CG interactions that depend only a limited set of particles. These can be represented as CG non-bonded and bonded interactions that can be expressed in terms of a vector or scalar quantity representing the relative positions between the set of interacting particles. This allows basis functions of a single variables to be defined, that construct the interactions within the CG potential. In the MSCG method, choices of these single-variable basis functions and least squares algorithms were tested and discussed.[77] This will be reviewed and examined later in Chapter 4, together with adaptation of the algorithm to the case of CG models with anisotropy.

3

All-atom simulations of bulk-heterojunction interface of BQR–PC_{71}BM

> *"Nothing in life is to be feared, it is only to be understood. Now is the time to understand more, so that we may fear less."*
>
> – Marie S. Curie

3.1 Abstract

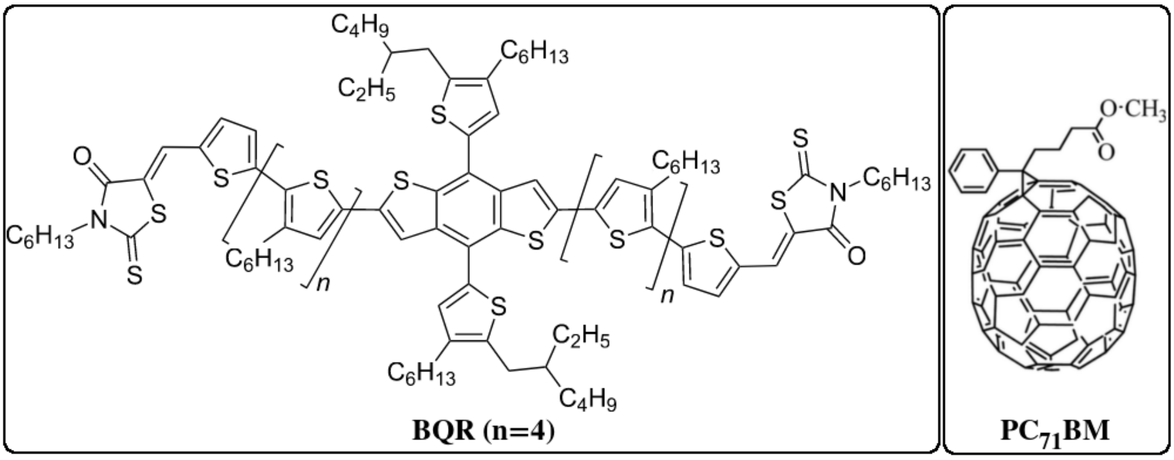
Organic solar cells using of organic semiconductors have achieved great success over the last twenty years of development. However, for large-scale application of this technology further improvement and better predictability of device efficiency are required. Structural features at the interfaces between materials composing the photo-absorbing layer in organic solar cells play a crucial role in important electronic processes that generate electric current in these devices. In this work, the interface structure of an experimentally high-performing electron donor–acceptor blend consisting of small-molecule organic semiconductors benzodithiophene quaterthiophene rhodanine (BQR) and [6,6]-phenyl-C71-butyric acid methyl ester (PC_{71}BM) is investigated for the first time using all-atom molecular dynamics (MD) simulations. Analyses of molecular orientation indicate BQR molecules have a great tendency to align with the interface, and a face-on configuration appears to be the preferred arrangement between phase-separated material interfaces. A higher level of disorder is also observed for molecules at the interface. These findings have implications for generalization of the role of structural variations at the donor–acceptor interface in device performance of organic photovoltaics (OPVs).

3.2 Introduction

Organic photovoltaics (OPVs) have gained attention in research as a potential application of organic semiconductors materials. OPV devices are famous for simple fabrication methods such as roll-to-roll printing and spray-coating. Moreover, these devices also possess desirable physical properties such as being light-weight, highly transparent and flexible, which are properties that are generally absent in inorganic solar cells. The photo-active layer for most OPVs consists of a mixture of two or more organic semiconductor materials that are phase-separated on a nano scale. Until now, commercialization of this technology has been limited by the device performance, which is significantly affected by charge generation and recombination processes at the electron donor–acceptor interfaces. These processes can vary depending on the micro-structure including molecular orientations at the interfaces, which are not yet fully understood and generalized for applications.

There have been a number of mechanisms proposed for optimizing charge generation and transport that involve the structure at donor–acceptor interfaces. In particular, electron–hole separation has been shown in theoretical studies to be enhanced by charge delocalization over molecules adjacent to the interfaces, induced by molecular alignment.[29,78] Moreover, computational studies have indicated increased order at the interfaces creates a quadrupole field that reduces the electron–hole binding energy, driving the charges from the interfaces. In contrast, charge separation can also be facilitated in systems with disordered interfaces caused by an increase in free energy due to energetic disorder.[46] Interfaces with disorder are also recommended in a number of studies to avoid non-radiative recombination and voltage loss[79,80] for better efficiency. In addition to the level of ordering, interfacial alignment is also an important factor to be considered. For a number of studied systems, a face-on arrangement of the aromatic rings in the materials is believed to increase electronic coupling between the donor and acceptor, promoting charge separation at their interfaces.[43,44,78] Though in some cases contradictory to each other, these factors may all have potential roles in controlling the efficiency of charge generation and transport to a degree which may vary from system to system.

In order to have a better understanding of how the interface structure influences the device performance of OPVs, it is necessary to have a good description of the donor–acceptor interfaces that yield different efficiencies. A study by Geraghty et al.[47] showed a large variation in the device performance between systems using various analogues of the small-molecule electron donor benzodithiophene quaterthiophene rhodanine (BQR) studied in this chapter. Interesting results were obtained, with increasing electron donor backbone length apparently increasing the power conversion efficiency, but with the highest efficiency of 9.4% for an intermediate backbone length. The best system was the bulk-heterojunction device of BQR as the electron donor and [6,6]-phenyl-C71-butyric acid methyl ester (PC$_{71}$BM) as the electron acceptor (chemical structure shown in Figure 3.1).



BQR (n=4)

PC$_{71}$BM

Figure 3.1: Chemical structures of electron donor BQR (left), and electron acceptor PC$_{71}$BM (right). Geraghty et al.[47] found systematic variations in the power conversion efficiency of OPV devices made with PC$_{71}$BM as the electron acceptor and a BQR analogue (with varying number, n, of hexylthiophene units) as the electron donor.

In this work, two different interfacial structures between BQR and PC$_{71}$BM are investigated, together with their pure systems as references. In particular, the interfacial alignment of BQR molecules with respect to the interface between the two materials, and the level of ordering at the interfaces are studied.

3.3 Computational methods

3.3.1 Force field parametrization

Force field parametrizations of BQR and $PC_{71}BM$ are necessary in order to accurately capture their structural properties such as molecular ordering or crystallinity in molecular dynamics (MD) simulations. Parametrization of the fine-grained (FG) models of both molecules used a combination of quantum chemistry calculations and MD simulations, together with adoption of most parameters from the OPLS-AA force field (version OPLS2005).[70] This force field has previously been shown to accurately capture the structural and thermodynamic properties of organic semiconductors.[12,72,81] Below an outline of the parametrization methods is presented. More details on specific force field parameters can be found in Appendix A.

Force field parametrization for BQR

Extended π-conjugation in organic semiconductors means that the default parameters in general force fields, in particular those for dihedral potentials and atomic charges can be inadequate. The parametrization procedure for BQR follows the method from Jackson et. al,[72] used to determine dihedral parameters for conjugated ring systems in organic semiconductors. This allows a more accurate atomistic description for BQR to be constructed. In particular, each of the four dihedral angles in BQR was calculated separately by computation of torsional potential energies at 10° intervals from -180° to 180°. For each calculation, the dihedral angle of interest was constrained while all other degrees of freedom were unconstrained, and geometry optimization using density functional theory at the B3LYP/6-31+G** level was performed. The optimized geometries were then used in RI-MP2/ccpVTZ single-point energy calculations. The resulting dihedral potential curve was constructed from the individual energy calculations at each dihedral angle, and fitted to a four-term OPLS dihedral potential form. To reduce the computational cost, all geometry optimizations were completed with the central benzodithiophene unit and one arm of the molecule, with all side chains replaced by methyl groups. Intramolecular geometric parameters such as equilibrium bond lengths and angles were recorded for an unconstrained geometry optimized at the B3LYP/6-31+G** level of theory. Bond and angle stretching constants, and pairwise non-bonded Lennard-Jones parameters were adopted from the OPLS-AA force field (version OPLS2005)[70] and from Jackson et. al.[72] Atomic charges were also calculated using the "Charges from the Electrostatic Potential on a Grid" (ChelpG) method for electrostatic potential fitting with B3LYP/6-31+G** level of theory.

Force field parametrization for $PC_{71}BM$

Parametrization of $PC_{71}BM$ was not trivial due to the asymmetric shape of the molecule. In particular, it was found that the C_{70} ball in the molecule could not maintain its shape when simulated at temperatures needed for this work if only one type of C atom was defined for the ball. Using only one type of C atom resulted in only one bond length and angle defined in C_{70}, which were not adequate to describe the ellipsoidal shape of C_{70} and the effect of the asymmetric side chain bonded to it. Eventually, six different types of C atoms were defined for the C_{70} part of $PC_{71}BM$ to ensure the molecules were not significantly distorted during MD simulations. Equilibrium bond lengths and angles, and atomic charges were calculated similarly for BQR. Bond and angle stretching constants, and pairwise non-bonded parameters for all atoms, except C atoms in the C_{70}, were adopted from the OPLS-AA force field (version OPLS2005).[70] Pairwise non-bonded parameters for C in the C_{70} ball were taken from Frigerio et al.,[82] which were based on Girifalco's values.[83] These parameters have been shown to give more consistent results to experimental data on miscibility with another organic semiconductor,[84] when being used together with OPLS-AA parameters. Parameters for dihedral angles in the side chain of $PC_{71}BM$ are from the OPLS-AA force field without modification.

3.3.2 Molecular dynamics simulations

For clarity and conciseness, the simulated systems are labeled as following:

- BQR–O: simulation of pure BQR of 125 molecules initially arranged in an **ordered** fashion in the simulation box, shown in Figure 3.2a. The ordered structure is based on a previously published crystal structure of benzodithiophene terthiophene rhodanine (BTR), an analogue of BQR, in which the molecules form a continuous π-stacking system.[7]

- $PC_{71}BM$–O: simulation of pure $PC_{71}BM$ of 539 molecules initially arranged in an **ordered** fashion with orthorhombic crystal structure to allow an orthogonal simulation box, shown in Figure 3.2b.

- BQR–$PC_{71}BM$–F: simulation of BQR–$PC_{71}BM$–F bilayer interface with the BQR molecules arranged **face-on** to $PC_{71}BM$. The face-on configuration is shown in Figure 3.3a, which is a part of the initial configuration of BQR–$PC_{71}BM$–F simulation box.

- BQR–$PC_{71}BM$–E: simulation of BQR–$PC_{71}BM$–E bilayer interface with the BQR molecules arranged **end-on** to $PC_{71}BM$. The end-on configuration is shown in Figure 3.3b, which is a part of the initial configuration of BQR–$PC_{71}BM$–E simulation box.

Simulations of pure BQR and $PC_{71}BM$ systems

Simulation of two different initial configurations of the BQR–$PC_{71}BM$ interface were carried out, together with simulations of the pure systems for reference. All simulations were performed using MD simulation package LAMMPS (version 22Aug18),[85] with the force fields obtained from the parametrization methods described above. The cutoff distance for van der Waals and real-space part of the calculation of Coulombic interactions was set to 11.0 Å, and the scaling factors for intramolecular 1–4 electrostatic and Lennard Jones interactions were set to 0.5 as specified in the OPLS-AA force field.[86] Long-ranged electrostatic interactions were calculated using the particle-particle particle-mesh (PPPM) method.[73]

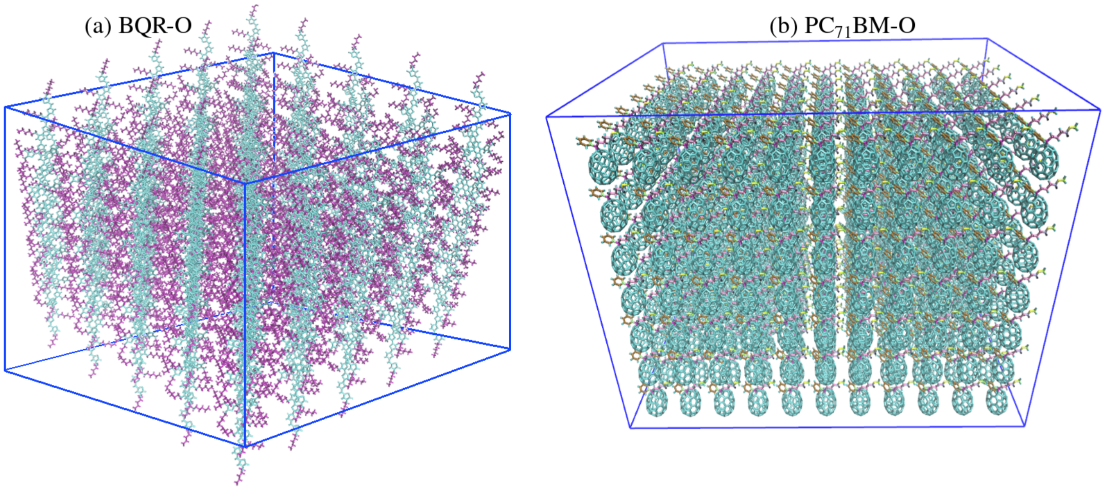


Figure 3.2: Initial configurations of pure BQR (left) and $PC_{71}BM$ (right) simulation systems

Two systems of pure BQR–O and $PC_{71}BM$–O arranged in an ordered fashion, as shown in Figure 3.2, were simulated for comparison to the heterogeneous systems at interfaces. The BQR–O consisted of 125 molecules, and the $PC_{71}BM$–O had 539 molecules. The systems were first relaxed with iterative energy minimizations using a conjugate gradient algorithm, then simulated in the NVT ensemble at temperature $T = 20\,\mathrm{K}$ with increasing timesteps (0.01 to 1 fs) for less than 0.05 ns. This was to ensure that the systems were relaxed to a mechanically stable state before being simulated with pressure control. NPT simulations at 1 atm pressure were carried out for all systems at a timestep of 0.5 fs, with the temperature increases from 50 K to 460 K (experimental liquid crystal temperature of BQR[47]) with 30 K intervals every 0.25 ns. The systems were allowed to equilibrate at 460 K for 5 ns. The densities recorded at this temperature is shown in Table 3.1. Simulated density for $PC_{71}BM$–O is within 4% difference compared to a previously reported density for simulation of $PC_{71}BM$ at the same temperature.

Table 3.1: Density of BQR and $PC_{71}BM$ from simulations and for simulated data $PC_{71}BM$ from literature at 460 K

Model	Density ($\mathrm{g\,cm^{-3}}$)
BQR–O	0.976
$PC_{71}BM$–O	1.479
$PC_{71}BM^{(a)}$	1.541

(a) Simulated data from ref. [87]

Simulations of BQR–$PC_{71}BM$ interface systems

The BQR–$PC_{71}BM$–F and BQR–$PC_{71}BM$–E interface systems with different configurations were simulated as illustrated in Figure 3.3. Similarly, the systems were relaxed using energy minimization with a conjugate gradient algorithm, then simulations in the NVT ensemble were carried out with increasing timesteps (0.01 to 1 fs), with a total simulation time of 0.1 ns. Again, this was to ensure that the systems were relaxed to a mechanically stable state before being simulated with pressure control. NPT simulations with simulation box lengths in three dimensions fluctuating to maintain a pressure of 1 atm were carried out for all systems at timestep of 0.5 fs, with the temperature increases from 50 K to 460 K with

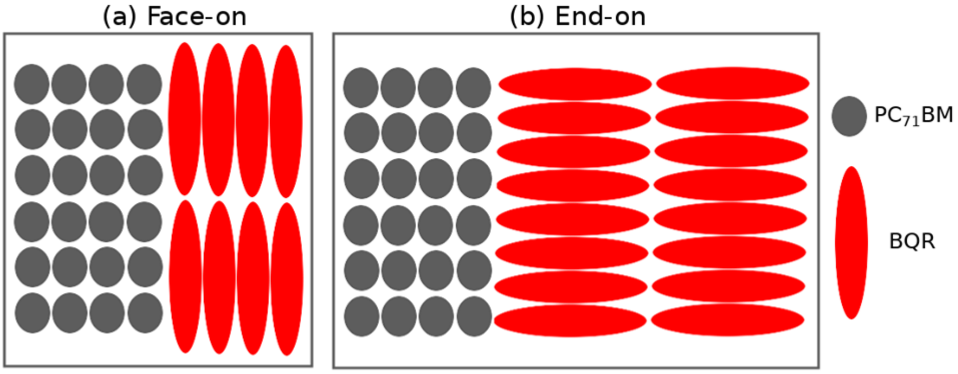


Figure 3.3: Representations of the initial configurations of the (a) face-on BQR–$PC_{71}BM$–F, and (b) end-on BQR–$PC_{71}BM$–E interface systems. $PC_{71}BM$ molecules shown in gray, and BQR molecules shown in red.

Table 3.2: Details of simulations of BQR–PC$_{71}$BM interface systems at high temperatures

Temperature (K)	Timestep (fs)	Run time (ns)
460	0.75	7.5
466	0.75	3.75
473	0.75	3.75
480	1.0	5.0
490	1.0	5.0
500	1.0	10.0

30 K intervals every 0.25 ns. NPT simulations near the experimental melting temperature of BQR[47] of 475 K were completed, with details shown in Table 3.2. The time steps used at these temperatures are higher than the heating process (0.75 to 1 fs) as the models have reached quite stable states at this point, and the longer time steps allow simulations for longer time to be obtained.

With the simulation times as presented, these systems are not likely to be at equilibrium, but the evolution of the systems with time at each temperature provides an indication of the microstructure of the equilibrium state that the system is evolving towards.

3.4 Structural analyses of BQR–PC$_{71}$BM interfaces

3.4.1 Molecular alignment at BQR–PC$_{71}$BM interfaces

Face-on BQR–PC$_{71}$BM interface system

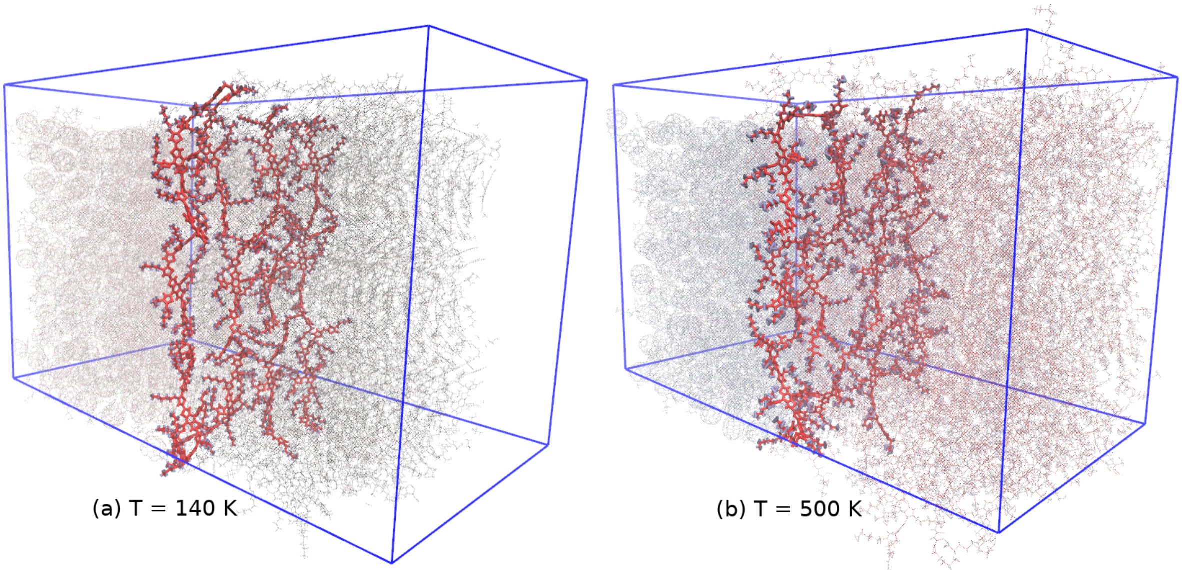


Figure 3.4: Simulation snapshots of the face-on BQR–PC$_{71}$BM–F interface with BQR molecules adjacent to the interfaces highlighted, at (a) a low temperature $T = 140\,\text{K}$, and (b) a high temperature $T = 500\,\text{K}$

As illustrated qualitatively in Figure 3.4, which shows the alignment of BQR molecules close to the BQR–PC$_{71}$BM–F interface, the molecules mostly remain in a face-on configuration relative to the $PC_{71}BM$ layer during the simulation at different temperatures. To quantify the degree of alignment of the molecules during the simulation, the average orientational order parameters $S_2(r)$ were measured as functions of distance r to the interface over different temperatures. In particular, for each temperature, the angle $\theta(r)$ between one of the molecular axis of BQR and a suitable interface axis were measured, and a histogram of distances to the interface r with weights as the angles was generated, with the angles for each histogram bin averaged to give a mean value. The orientation parameter $S_2(r)$ is the average value of the second Legendre polynomial of $\cos\theta(r)$:

$$S_2(r) = \big\langle P_2\big(\cos\theta(r)\big)\big\rangle = \Big\langle \frac{3\cos^2\theta(r) - 1}{2} \Big\rangle \tag{3.1}$$

The average orientational order parameter $S_2(r)$ shows the degree of alignment with respect to the reference axis. A value of 1 indicates perfect parallel alignment. Values of close to 0 indicate random alignment to the reference axis, and values close to -0.5 indicate perpendicular alignment. As shown in Figure 3.5, two types of angles were measured: one is between the long axis of BQR, $\mathbf{n}_1$, and an axis in the plane of the interface, $\mathbf{i}_1$ (angle α in Figure 3.5a), and the other is between the short axis of BQR, $\mathbf{n}_2$, and the axis normal to the interface, $\mathbf{i}_2$ (angle β in Figure 3.5 b). The angle α defined in Figure 3.5a shows whether the molecules are parallel or perpendicular to the interface, while β defined in Figure 3.5b allows differentiation between face-on and end-on configuration to the interface. It is necessary to measure both of them for sufficient information on the alignment with respect to the interface plane. It is evident from Figure 3.5a that, during the course of the simulation from low to high temperature, the long axis of the BQR molecules remains parallel to the interface, with $S_2(r)$ close to a value of 0.8 near the interface. There is a small drop and large variation at distances of 20.0 Å and further from the interface. This shows a loss of correlation between the interface and the molecular axis parallel to the interface as the distance from the interface increases. However, it can be seen from Figure 3.5b that the face-on molecular alignment of the system is enhanced as the temperature increases both near and far away from the interface. At temperature of 500 K, $S_2(r)$ is about 0.7 both close to and far from the interface, increasing from 0.25 near the interface and from -0.25 far from the interface, respectively, at the lowest temperature. This qualitatively agrees with experimental results showing the nematic phase of BQR in the bulk.[47] The arguments above indicate that the face-on configuration is thermodynamically stable at the interface between BQR and $PC_{71}BM$ near the melting temperature of BQR.

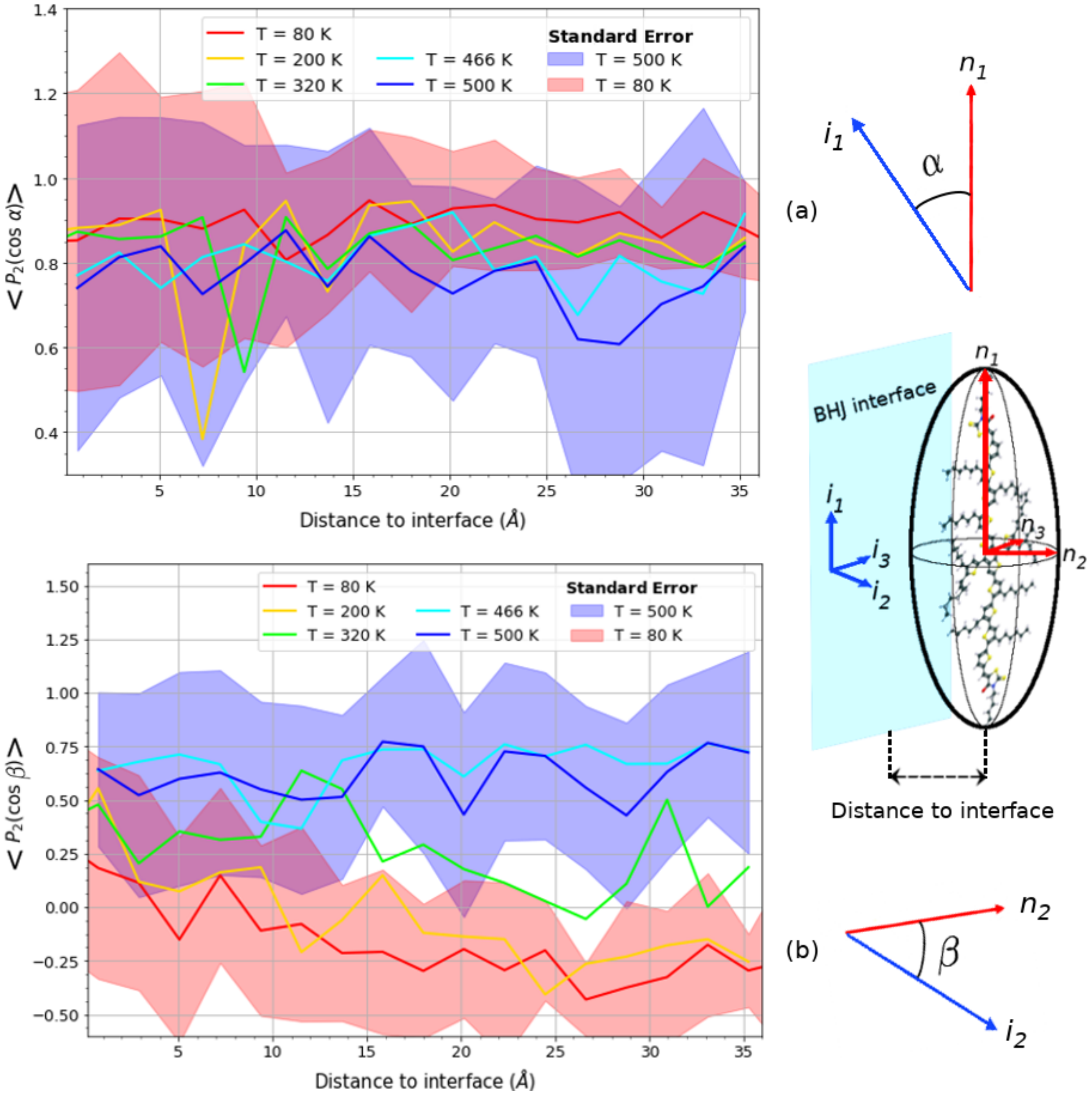


Figure 3.5: Relative orientations between BQR molecules and BQR–PC$_{71}$BM–F interface at different temperatures. (a) Average order parameter versus distance to interface $S_2(r) = \left\langle P_2\left(\cos\alpha(r)\right)\right\rangle$ that describes the angle α between the long axis of BQR $\mathbf{n}_1$ and axis $\mathbf{i}_1$ of the interface, and (b) average order parameter versus distance to interface $S_2(r) = \left\langle P_2\left(\cos\beta(r)\right)\right\rangle$ that describes the angle β between the short axis of BQR $\mathbf{n}_2$ and axis $\mathbf{i}_2$ normal to the interface. Positive values of $S_2(r)$ indicate there is parallel alignment, while values close to -0.5 indicate perpendicular alignment. Standard errors of $S_2(r)$ for the lowest and highest temperatures 80 K and 500 K are shown as filled areas.

End-on BQR–PC$_{71}$BM interface system

The simulation of the BQR–PC$_{71}$BM–E system over different temperatures shows contrasting behaviour to the BQR–PC$_{71}$BM–F system, with the structure at the heterojunction interface changing significantly. An example of this is shown in Figure 3.6. The top image shows the initial configuration with all BQR molecules end-on to the interface with PC$_{71}$BM, and the bottom shows a snapshot of the system after 30 ns when simulated at temperature $T = 500$ K. The three same molecules adjacent to the interface are highlighted in both to illustrate the orientational changes. It can be seen that the molecules become perpendicular to their original orientations and parallel to the interface.

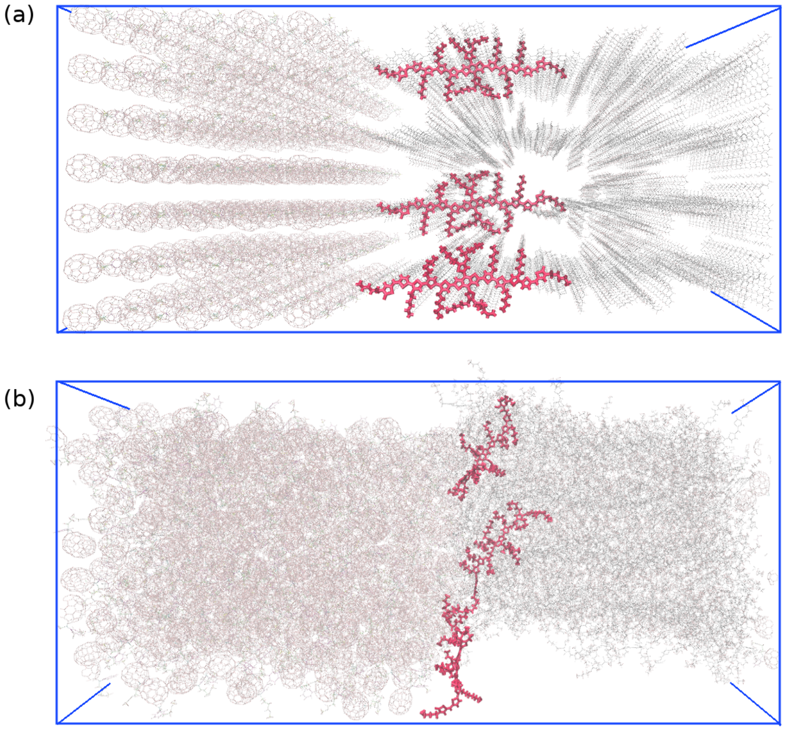


Figure 3.6: Simulation snapshots of BQR–PC$_{71}$BM–E interface with BQR molecules adjacent to the interfaces highlighted (a) initial end-on interface configuration, and (b) at high temperature $T = 500$ K when molecules near the interface rotate to align parallel with the interface.

Orientational order parameters were also calculated for different temperatures along the simulation, which are shown in Figure 3.7. Again, there are two angles of interest, the α angle between the long axis $\mathbf{n}_1$ of BQR and the axis $\mathbf{i}_1$ in the plane of the interface which determines whether the molecules are parallel or perpendicular to the surface, and the β angle between the short axis $\mathbf{n}_2$ of BQR and axis $\mathbf{i}_2$ normal to the interface. With the setup for BQR–PC$_{71}$BM–E, at the start of the simulation these angles are close to 90°, giving average orientational order parameters close to -0.5. This is the case for simulation at low temperature (80 K in Figure 3.7). As the temperature increases, more BQR molecules close to the interface rotate to align with the interface, as illustrated for the three representative molecules in Figure 3.6. This increases $S_2(r)$ near the interface to a value of 0.3, indicating there is parallel alignment to the interface. $S_2(r)$ near the interface in Figure 3.7(b) also increases with temperature but remains negative. Overall, the end-on configuration at BQR–PC$_{71}$BM interface appears not to be favorable near the melting point of BQR, and BQR molecules near the interface appear to be evolving towards a parallel alignment.

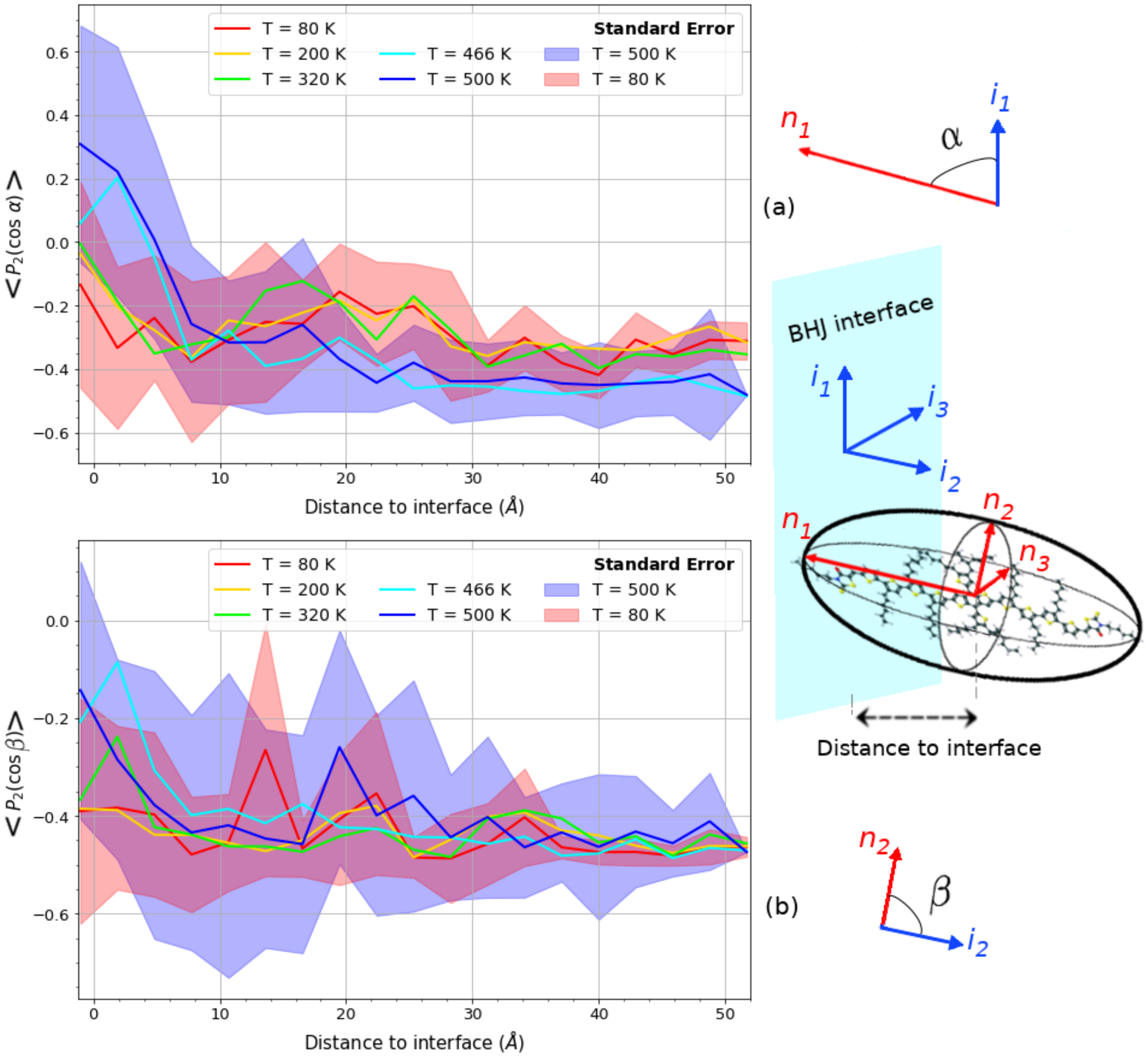


Figure 3.7: Relative orientations between BQR molecules and BQR–PC$_{71}$BM–F interface at different temperatures.(a) Average order parameter versus distance to interface $S_2(r) = \langle P_2(\cos\alpha(r)) \rangle$ that describes the angle α between the long axis of BQR $\mathbf{n}_1$ and axis $\mathbf{i}_1$ of the interface, and (b) average order parameter versus distance to interface $S_2(r) = \langle P_2(\cos\beta(r)) \rangle$ that describes the angle β between the short axis of BQR $\mathbf{n}_2$ and axis $\mathbf{i}_2$ normal to the interface. Positive values of $S_2(r)$ indicate there is parallel alignment, while values close to -0.5 indicate perpendicular alignment. Standard errors of $S_2(r)$ for the lowest and highest temperatures 80 K and 500 K are shown as filled areas.

3.4.2 Investigation of degree of ordering at BQR–PC$_{71}$BM interfaces

Besides orientation of the molecules with respect to the interface, the degree of order is also important for performance of OPV devices.[28] Alignment of molecules with respect to one another at the interface has been shown to induce electric fields that promote charge separation for OPVs.[27] In order to measure these structural features of the simulated systems, a different order parameter can be measured. These parameters also have the form as defined in Eq (3.1), but the angles θ are measured between pairs of molecular axes, and averaged over all molecules. These angles are also averaged over molecular separation distances. The results are order parameters as a function of distance between molecules, showing the local ordering between molecules close to each other. In this section, the analysis is for the highest temperature studied $T = 500\,\mathrm{K}$ in region close to or far from the interface, with molecules within 15 Å from the interface considered to be in the "near" interface region and other molecules are in the "far" region. Data for the BQR–O system is for a temperature of 460 K where nematic liquid crystal phase has been observed experimentally.[47]

Figure 3.8 shows the average intermolecular orientational order parameter for BQR molecules in the BQR–PC$_{71}$BM–F system, close to or far from the interface with PC$_{71}$BM. Ordering in the nematic direction along the long axis of BQR and along the π-stacking direction was measured. The same quantity was measured for the pure BQR–O simulation for comparison. It can be seen that for the BQR–PC$_{71}$BM–F system the ordering is high for both near and far regions, and is comparable to the bulk BQR–O simulation. A noticeable feature is molecules near the interface have reduced ordering in both the nematic and π-stacking directions with their neighboring molecules, evidenced by a drop of $S_2(r)$ at distance closer to 4 Å in Figure 3.8(a) and (b). The negative value of $S_2(r)$ normally indicates perpendicular alignment between pairs of BQR molecules. However as molecules near the interface are likely to be bent, the molecular axes may not be the best to describe the correct orientational correlation between the molecules and may underestimate the order parameter. In any case, it is sufficient to conclude that molecules at the interface are more disordered than in the bulk, especially for short-range packing.

Variation of the average intermolecular orientational order parameter for BQR molecules in the BQR–PC$_{71}$BM–E system close and far away from the interface is shown in Figure 3.9. For this system, the parameters and so ordering are overall higher for molecules far from the interface in both the nematic and π-stacking directions. The same feature as for the BQR–PC$_{71}$BM–F is observed, that is a drop in short-ranged ordering for molecules closer to the interface. These observations for both the face-on and end-on interface systems support a reduced degree of molecular ordering at the interface, indicating π stacking is less effective at close proximity to a different environment. A possible reason a possible reason is that BQR and PCBM are very different in shape, so packing at the interface might be inefficient. This interesting result may be a factor that contributes to the high performance of solar cells made of BQR–PC$_{71}$BM, as disorder at the donor-acceptor interface has been proposed to provide a driving force for charge separation.[46] In order to clarify this, future work can focus on the consequences of the interface microstructure in this BQR–PC$_{71}$BM system on the electronic structure and charge separation, together with comparison to the reported analogous systems that yield different efficiencies in their OPV devices.[47]

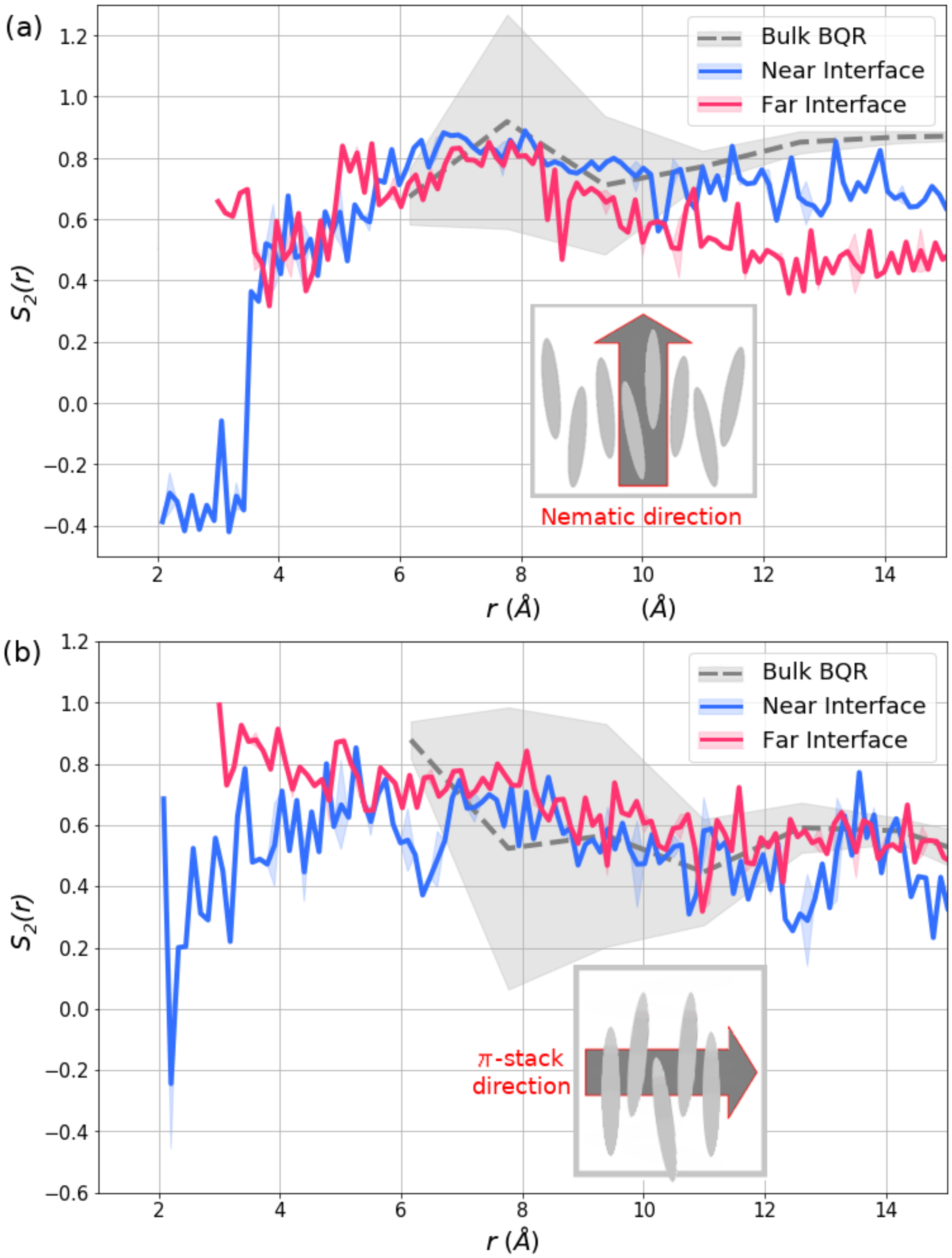


Figure 3.8: Relative orientation of BQR molecules versus intermolecular distance r for near to (blue) or far from (pink) the BQR–PC$_{71}$BM–F interface (a) in the nematic direction and (b) in the π-stacking direction. Standard errors are shown as filled areas, and the error for the BQR–O system is larger due to larger intermolecular distance intervals used to calculate the average order parameter for easier visualization.

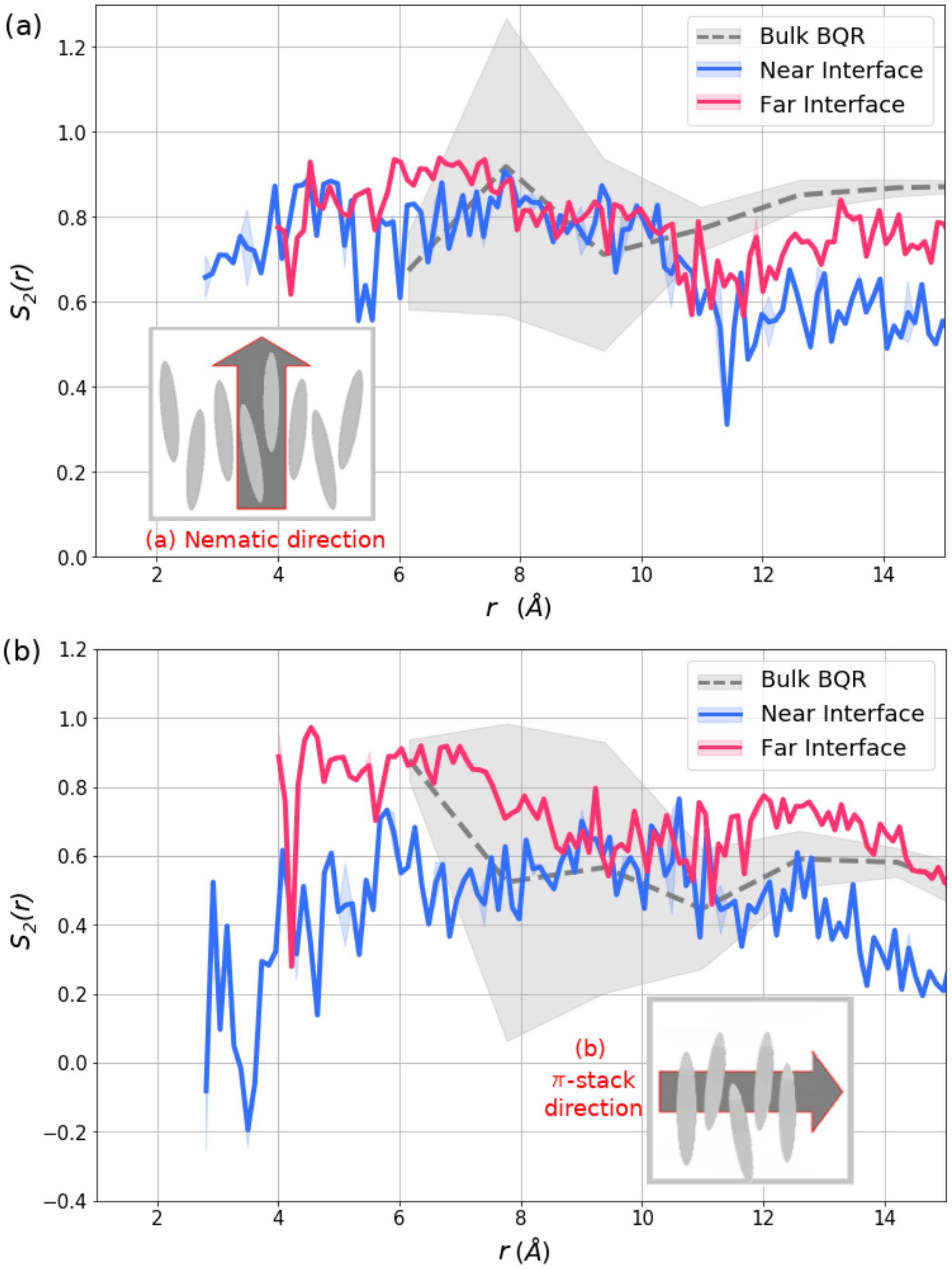


Figure 3.9: Relative orientation of BQR molecules versus intermolecular distance r for near to (blue) or far from (pink) the BQR–$PC_{71}BM$–E interface (a) in the nematic direction and (b) in the π-stacking direction. Standard errors are shown as filled areas, and the error for the BQR–O system is larger due to larger intermolecular distance intervals used to calculate the average order parameter for easier visualization.

3.5 Conclusions

In this work, atomistic molecular models of a high-performing donor–acceptor bulk heterojunction BQR–PC$_{71}$BM were used to study the molecular-level structural variation of the molecules at the interface. The results show that for both systems with initial face-on or end-on configurations, BQR molecules have a tendency to be parallel rather than being end-on to the PC$_{71}$BM molecules. The face-on configuration being more stable at the interface, may indicate this is the preferred configurations for nano-scaled phase-separated domains of materials in a real bulk-heterojunction with PC$_{71}$BM. This preferred microstructure may be related to the outstanding performance of this system compared to others with similar molecular chemical moieties.[47] Furthermore, BQR molecules close to the interface are less ordered with respect to one another than those far from the interface with PC$_{71}$BM, possibly due to the incompatible shapes of the molecules. The preferred alignment of the long conjugated donor molecules parallel to the interface and the higher level of disorder at the interface found for BQR–PC$_{71}$BM can be factors that benefit charge generation and transport. Subsequent work can be completed for analogous systems that give lower efficiencies for comparison. This will allow a full investigation of how the shape of these molecules affects the structure of the interfaces, and consequently the electronic processes in their OPV devices.

4

The anisotropic force-matching coarse-grained method: A systematic coarse-grained algorithm for molecules with anisotropy

"Oh, it's delightful to have ambitions. I'm so glad I have such a lot. And there never seems to be any end to them– that's the best of it. Just as soon as you attain to one ambition you see another one glittering higher up still. It does make life so interesting."
(Anne Shirley)

– Lucy M. Montgomery, *Anne of Green Gables*

4.1 Abstract

Molecular anisotropy plays an important role in controlling the microstructure of organic semiconductor materials. Currently, computer modeling of this microstructure is limited, either by the small sizes of simulation systems using atomistic/fine-grained (FG) models, or neglect of molecular anisotropy when traditional spherical coarse-grained (CG) models are employed. The present work addresses the above challenge with the development of a novel CG theory and parametrization algorithm for anisotropic molecular models, named the anisotropic force-matching coarse-grained (AFM-CG) method. The method is based on a rigorous statistical mechanical framework, which produces rules on definition of thermodynamically consistent CG models. The method can be employed to generate CG models of conjugated organic semiconductors, with the potential to allow realistic structures on large spatial and time scales to be studied.

4.2 Introduction

Organic electronic devices possess a number of characteristics that can be advantageous compared with traditional inorganic counterparts.[22,23] These devices are fabricated using organic semiconductors, which are often processed in solvent[20] and yield thin flexible devices. However, these organic thin films are still not commonly used, especially in organic photovoltaic (OPV) technology, because the performance is not comparable to that of silicon-based devices. Experimental studies of organic electronic devices suggest that morphology at material interfaces substantially affects the device performance,[19,42,88] yet this is still not fully understood due to difficulties in visualizing these structural details. More predictive approaches are necessary, which can be achieved with theoretical scaffolds using computer simulations and calculations.

To date, computational modelling and simulations of electronic properties of organic semiconductors have not been completed in a general and efficient way. While fine-grained (FG) simulations with atomistic details provide specific interactions at material interfaces, they are normally limited by time and length scales needed for interfacial formations and structural variations. A solution to this is to use coarse-grained (CG) molecular simulations, where atoms in rigid groups are mapped to a single site, speeding up the simulations by orders of magnitude. Molecular simulations of large systems including polymers have been obtained with the use of spherical CG models with enough accuracy in the case when the overall polymer aggregates are isotropic.[89] Spherical models are mainly used due to their simple and readily developed CG parametrization methods.[54,59]

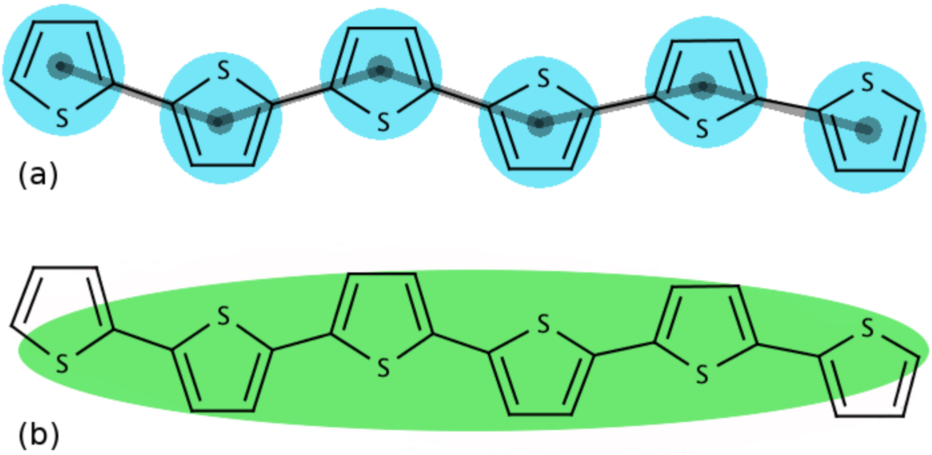


Figure 4.1: Schematics of (a) a spherical model and (b) a ellipsoidal CG model of small conjugated molecule sexithiophene. The spherical model represents each thiophene unit as a sphere connected by their centers of mass (black lines to black dots), while the whole thiophene molecule is coarse-grained into a single site in the ellipsoid model

However, in molecular systems with significant molecular anisotropy in their structures and interactions, such as organic semiconductors, spherical CG models may not be able to produce accurate and useful information on molecular structures and interactions. A solution to this is the use of anisotropic CG models, which can capture the molecular shape more accurately. Although simulations of these models are slower than for isotropic models, they can still provide computational saving, if they contain fewer particles than the spherical model of the same molecular system. An example of potential saving in computational cost is shown by comparing two CG models of a small conjugated molecule sexithiophene in Figure 4.1. It can be seen clearly that in order to remain anisotropic, a number of spherical CG sites must be used, while the molecule can be readily coarse-grained into an anisotropic particle such as an ellipsoid. Molecular dynamics (MD) simulations of these anisotropic CG models are also assisted with development of a number of analytical non-bonded potentials. The most commonly used potential forms are the Gay-Berne,[90] and the RE-squared potentials for ellipsoidal particles.[91] Moreover, structural analysis of anisotropic models is generally simpler, as the models often contain sufficient configuration degrees of freedom to quantify both the positions and orientations of molecules in simulation systems.

Despite the advantages listed above, there is not yet a general and rigorous method to parametrize anisotropic CG models. The reason behind this may be because the short-range non-bonded potential function becomes very complex when orientations of CG sites are considered. There are common systematic “bottom-up” coarse-graining approaches available such as the iterative Boltzmann inversion (IBI) method which matches probability distribution functions,[53,54] the multi-scale coarse-graining (MSCG) method which matches forces,[58,59] and the relative entropy method which minimizes the relative entropy between the CG and FG systems.[63] However, general algorithm and implementation for parametrization of anisotropic CG models using these methods have not yet been generated. Consequently, CG parametrization for anisotropic systems in a small numbers of previous studies normally lacked a strong theory base on the method and a systematic algorithm that can be reused for other systems.[65,66,92,93] It is thus necessary to develop such a method, which will allow the use of anisotropic CG models in studies of molecules such as organic semiconductors.

In this present work, a CG parametrization method with a strong theory basis using rigorous statistical mechanical principles and computational algorithms is introduced. The method is called the anisotropic force-matching coarse-grained (AFM-CG) method, which uses force matching to transfer FG models to their CG descriptions. In this chapter, the theory of CG force matching method for spherical models is briefly summarized. Then the configuration and force matching conditions for anisotropic CG models are derived. Finally, practical algorithms of the method are introduced, which addresses the use of variation principle and multivariate approximation methods.

4.3 Coarse-grained force matching for spherical models

Rigorous theoretical backgrounds and algorithms were introduced previously in the MSCG method[58,59,77] of force matching for spherical CG models. As described in Chapter 2, it was shown that for a CG model that is thermodynamically consistent with a FG model, the force acting on a CG site is equal to the sum of the forces acting on the atoms within that site in the FG model with the same configuration.[94] Considering a FG system of n atoms that is mapped to a CG system of N sites, the above statement can be expressed for any site I in the CG system as

$$\mathbf{F}_{\mathrm{CG},I}(\mathbf{R}^N) = \left\langle \mathbf{F}_{\mathrm{FG},I}(\mathbf{r}^n) \right\rangle_{\mathbf{R}^N} = \langle \sum_{i \in I} \mathbf{F}_i(\mathbf{r}^n) \rangle_{\mathbf{R}^N} \tag{4.1}$$

where $\mathbf{r}^n$ and $\mathbf{R}^N$ define the positions of the atoms/sites in the FG and CG systems, respectively, $\mathbf{F}_{\mathrm{CG},I}$ is the force acting on CG site I in the CG system, while $\mathbf{F}_{\mathrm{FG},I}$ is the force in the FG system, which is computed as the sum of the forces acting on any atom i belonging to CG site I, $\mathbf{F}_i(\mathbf{r}^n)$, and the angle brackets denote average over FG configurations consistent with the CG configuration.

This suggests that an "optimal" CG model is one with parameters $\boldsymbol{\lambda}$ chosen to minimize a cost function χ^2 that is related to the difference between the FG and CG forces:

$$\chi^2(\mathbf{F}_{\mathrm{CG}}) = \chi^2(\boldsymbol{\lambda}) = \frac{1}{3Nt} \sum_{I=1}^{N} |(\mathbf{F}_{\mathrm{FG},I}(\mathbf{r}^n)) - \mathbf{F}_{\mathrm{CG},I}(\mathbf{R}^N, \boldsymbol{\lambda})|^2 \tag{4.2}$$

Here the cost function is divided by $3Nt$, which denotes averaging over three dimensions, N CG sites and t atomistic simulation snapshots from which the FG forces are taken.

In implementations of the method,[59,95,96] the CG force is constructed as a linear combination of many basis functions with scalar parameters $\boldsymbol{\lambda}$. These parameters are solved in a least-squares sense, with $\mathbf{F}_{\mathrm{CG}}$ equal to the known $\mathbf{F}_{\mathrm{FG}}$ for all simulation snapshots. When a least-squared solution is found, a CG force function can be constructed from the basis functions and the resulting parameters, rigorously transforming the FG to the CG system. The method has been improved to allow better accuracy and different simulation conditions by including additional algorithms such as pressure matching,[97] three-body interaction force matching,[98] and state-point dependent force matching.[99]

4.4 Definition of anisotropic coarse-grained models and statistical mechanics quantities

Similar to the case of spherical CG models, an anisotropic CG model needs to satisfy the consistency requirements to ensure that the thermodynamic behavior is the same as that of the FG model in MD simulations. Here we follow a similar procedure as that described in Chapter 2, to determine consistency conditions and to prove the force-matching equations for CG models that include molecular anisotropy for the AFM-CG method.

For a spherical CG model, potential energy functions are assumed to depend only on the positional coordinates of the CG sites, each of them is a linear combination of the coordinates of a group of atoms in the atomistic model. For a CG model of N sites that includes molecular anisotropy in terms of shape and interactions, the potential energy functions depend on both the positions $\mathbf{R}^N = \{\mathbf{R}_1, ..., \mathbf{R}_N\}$, and orientations $\mathbf{\Omega}^N = \{\mathbf{\Omega}_1, ..., \mathbf{\Omega}_N\}$ of the CG sites, which depend on how these quantities are defined with respect to the FG model. Again, molecular vibration and internal rotation are neglected in this classical rigid body approximation (see Appendix B).

As described in Chapter 2, a FG model at an instantaneous dynamical state is defined by the coordinates $\mathbf{r}^n = \{\mathbf{r}_1, ..., \mathbf{r}_n\}$ and momenta $\mathbf{p}^n = \{\mathbf{p}_1, ..., \mathbf{p}_n\}$ of n component atoms. The Hamiltonian $h(\mathbf{r}^n, \mathbf{p}^n)$ that describes the energy of all of these atoms, consists of kinetic and potential parts, respectively expressed in the following equation:

$$h = \sum_{i=1}^{n} \frac{\mathbf{p}_i^2}{2m_i} + u(\mathbf{r}^n) \tag{4.3}$$

where m_i is mass of atom i in n atoms, and $u(\mathbf{r}^n)$ is the potential energy operator of the system. Note that momentum is defined physically as $\mathbf{p}_i = m_i \dot{\mathbf{r}}_i$.

The phase-space distribution function $P(\mathbf{r}^n, \mathbf{p}^n)$ is related to the Hamiltonian as

$$P(\mathbf{r}^n, \mathbf{p}^n) \propto \exp(-h/k_\mathrm{B}T) = \exp(-\sum_{i=1}^{n} \mathbf{p}_i^2/2m_i k_\mathrm{B}T) \times \exp(-u(\mathbf{r}^n)/k_\mathrm{B}T) \tag{4.4}$$

where k_B is the Boltzmann constant and T is the absolute temperature. The FG Hamiltonian $h(\mathbf{r}^n, \mathbf{p}^n)$ is substituted using Eq (4.3).

As shown in Eq (4.4), the phase-space distribution function can be factorized into the kinetic component that depends on the momenta, and the potential component that depends on the position coordinates:

$$P(\mathbf{r}^n, \mathbf{p}^n) \propto P_\mathrm{t}(\mathbf{p}^n) P_\mathrm{c}(\mathbf{r}^n) \tag{4.5}$$

where

$$P_{\mathrm{t}}(\mathbf{p}^n) \propto \exp\left(-\sum_{i=1}^{n} \mathbf{p}_i^2/2m_i k_{\mathrm{B}} T\right) \tag{4.6}$$

and

$$P_{\mathrm{c}}(\mathbf{r}^n) \propto \exp(-u(\mathbf{r}^n)/k_{\mathrm{B}} T) \tag{4.7}$$

In the case where molecular orientations are important, a CG representation of this model consisting of N CG sites is defined in a more precise way by a set of configuration variables $(\mathbf{R}^N, \mathbf{P}^N, \mathbf{\Omega}^N, \mathbf{P}_{\Omega}^N)$. For this CG representation we use capitalized symbols of those used for the FG model to emphasize the correspondence in the derivation. Here $\mathbf{P}^N = \{\mathbf{P}_1, ..., \mathbf{P}_N\}$ are the translational momenta conjugate to the positions $\mathbf{R}^N = \{\mathbf{R}_1, ..., \mathbf{R}_N\}$ of the CG sites, $\mathbf{\Omega}^N = \{\mathbf{\Omega}_1, ..., \mathbf{\Omega}_N\}$ describes the orientations of the sites, and $\mathbf{P}_{\Omega}^N = \{\mathbf{P}_{\Omega_1}, ..., \mathbf{P}_{\Omega_N}\}$ are angular momentum conjugate to $\mathbf{\Omega}^N$. The Hamiltonian is constituted by the translational energy K_{t}, rotational energy K_{r} and potential energy U:

$$\begin{aligned} H &= K_{\mathrm{t}} + K_{\mathrm{r}} + U(\mathbf{R}^N, \mathbf{\Omega}^N) \\ &= \sum_{I=1}^{N} \frac{\mathbf{P}_I^2}{2M_I} + \sum_{I=1}^{N} \sum_{q=x,y,z} \mathbf{P}_{Iq}^2/2\mathbb{I}_{Iq} + U(\mathbf{R}^N, \mathbf{\Omega}^N) \end{aligned} \tag{4.8}$$

where $\mathbf{P}_{Iq}$ is the component of the angular momentum in reference to the body-fixed principal axis q for CG site I, $\mathbb{I}_{Iq}$ is the moment of inertia tensor with respect to the q axis, and M_I is the mass of the site.

The probability density is related to the Hamiltonian similarly to the expression in Eq (4.4) and can be separated into translational, rotational and configuration parts, respectively, as

$$P(\mathbf{R}^N, \mathbf{P}^N, \mathbf{\Omega}^N, \mathbf{P}^N) = P_{\mathrm{t}}(\mathbf{P}^N) P_{\mathrm{r}}(\mathbf{P}^N) P_{\mathrm{c}}(\mathbf{R}^N, \mathbf{\Omega}^N) \tag{4.9}$$

where

$$P_{\mathrm{t}}(\mathbf{P}^N) \propto \exp\left(-\sum_{I=1}^{N} \mathbf{P}_I^2/2M_I k_{\mathrm{B}} T\right) \tag{4.10}$$

$$P_{\mathrm{r}}(\mathbf{P}^N) \propto \exp\left(-\sum_{I=1}^{N} \sum_{q=x,y,z} \mathbf{P}_{Iq}^2/2\mathbb{I}_{Iq} k_{\mathrm{B}} T\right) \tag{4.11}$$

$$P_{\mathrm{c}}(\mathbf{R}^N, \mathbf{\Omega}^N) \propto \exp(-U(\mathbf{R}^N, \mathbf{\Omega}^N)/k_{\mathrm{B}} T) \tag{4.12}$$

Here the body-fixed angular momenta $\mathbf{P}^N$ are used to allow separation of the probability and partition functions. Proof of this can be found elsewhere.[76]

4.5 Consistency conditions for anisotropic coarse-grained models

For anisotropic models, the CG coordinates and translational momenta are mapped from those of the FG models via the following linear mapping operators:

$$\mathbb{M}_{\mathbf{R}}^{N}(\mathbf{r}^{n}) : \mathbf{r}^{n} \rightarrow \mathbf{R}^{N} \quad \text{with} \quad \mathbb{M}_{\mathbf{R},I}(\mathbf{r}^{n}) = \sum_{i=1}^{n} c_{I,i}\mathbf{r}_{i} \quad \text{for site } I \tag{4.13}$$

$$\mathbb{M}_{\mathbf{P}}^{N}(\mathbf{p}^{n}) : \mathbf{p}^{n} \rightarrow \mathbf{P}^{N} \quad \text{with} \quad \mathbb{M}_{\mathbf{P},I}(\mathbf{p}^{n}) = \sum_{i=1}^{n} c_{I,i}\mathbf{p}_{i} \quad \text{for site } I \tag{4.14}$$

Mapping operators of CG quantities that depend on the orientations can be more complicated and nonlinear in general. We shall define the configuration mapping operator for the CG orientations as $\mathbb{M}_{\mathbf{\Omega}}^{N} : \mathbf{r}^{n} \rightarrow \mathbf{\Omega}^{N}$, and mapping operator for the CG angular momenta as $\mathbb{M}_{\mathbf{P}_{\omega}}^{N} : \mathbf{p}^{n} \rightarrow \mathbf{P}^{N}$

As probability distribution functions describe the thermodynamics of a system, for the FG and CG models to be consistent, these functions in the two descriptions of the system must be equal. With separation into configurational and kinetic components as shown in Eq (4.5) and (4.9), the following equations are required:

$$\begin{aligned} P_{\mathrm{c}}(\mathbf{R}^{N}, \mathbf{\Omega}^{N}) &= P_{\mathrm{c}}(\mathbf{r}^{n}) \\ &= \int d\mathbf{r}^{n} \int d\mathbf{r}^{n} P(\mathbf{r}^{n}) \delta(\mathbb{M}_{\mathbf{R}}^{N}(\mathbf{r}^{n}) - \mathbf{R}^{N}) \delta(\mathbb{M}_{\mathbf{\Omega}}^{N}(\mathbf{r}^{n}) - \mathbf{\Omega}^{N}) \end{aligned} \tag{4.15}$$

$$\begin{aligned} P_{\mathrm{t}}(\mathbf{P}^{N}) P_{\mathrm{r}}(\mathbf{P}^{N}) &= P_{\mathrm{t}}(\mathbf{p}^{n}) \\ &= \int d\mathbf{p}^{n} \int d\mathbf{p}^{n} P(\mathbf{p}^{n}) \delta(\mathbb{M}_{\mathbf{P}}^{N}(\mathbf{p}^{n}) - \mathbf{P}^{N}) \delta(\mathbb{M}_{\mathbf{P}_{\omega}}^{N}(\mathbf{p}^{n}) - \mathbf{P}^{N}) \end{aligned} \tag{4.16}$$

Dirac delta functions (δ function, see Appendix B) are used to enforce the momentum and configuration mappings for each CG site:

$$\delta(\mathbb{M}_{\mathbf{R}}^{N}(\mathbf{r}^{n}) - \mathbf{R}^{N}) = \prod_{I=1}^{N} \delta(\mathbb{M}_{\mathbf{R},I}(\mathbf{r}^{n}) - \mathbf{R}_{I}) \tag{4.17}$$

$$\delta(\mathbb{M}_{\mathbf{\Omega}}^{N}(\mathbf{r}^{n}) - \mathbf{\Omega}^{N}) = \prod_{I=1}^{N} \delta(\mathbb{M}_{\mathbf{\Omega},I}(\mathbf{r}^{n}) - \mathbf{\Omega}_{I}) \tag{4.18}$$

$$\delta(\mathbb{M}_{\mathbf{P}}^{N}(\mathbf{p}^{n}) - \mathbf{P}^{N}) = \prod_{I=1}^{N} \delta(\mathbb{M}_{\mathbf{P},I}(\mathbf{p}^{n}) - \mathbf{P}_{I}) \tag{4.19}$$

$$\delta(\mathbb{M}_{\mathbf{P}_{\omega}}^{N}(\mathbf{p}^{n}) - \mathbf{P}^{N}) = \prod_{I=1}^{N} \delta(\mathbb{M}_{\mathbf{P}_{\omega},I}(\mathbf{p}^{n}) - \mathbf{P}_{I}) \tag{4.20}$$

where $\mathbb{M}_{\mathbf{R},I}$,$\mathbb{M}_{\mathbf{\Omega},I}$,$\mathbb{M}_{\mathbf{P},I}$ and $\mathbb{M}_{\mathbf{P}_{\omega},I}$ are individual mapping operators for each CG site I as defined in the definitions of the mapping operators.

Eq. (4.15) and (4.16) describes the general consistency conditions of a CG model compared to the atomistic description. These equations can be more useful once expressed in termed of physical quantities such as forces, masses and inertia tensors of rigid CG sites, presented in the next section.

4.6 Expressions of force-matching rules for anisotropic coarse-grained models

4.6.1 Consistency conditions in configuration space and expressions for coarse-grained forces

It is convenient to derive a relationship between the CG and FG potential energies and the resulting forces, as these quantities are readily extracted from MD simulations, given the potentials are defined completely for all interactions. CG equations can be determined from this relationship that are useful in implementation of the AFM-CG method, as in the case of the MSCG method for spherical CG sites. Given the relations of the potentials to the configurational probability distributions in Eq (4.7) and (4.12), and Eq (4.15), it can be seen that:

$$\exp(-U(\mathbf{R}^N,\mathbf{\Omega}^N)/k_\mathrm{B}T) \propto \int d\mathbf{r}^n \int d\mathbf{r}^n \exp(-u(\mathbf{r}^n)/k_\mathrm{B}T)\delta(\mathbb{M}_\mathbf{R}^N(\mathbf{r}^n)-\mathbf{R}^N)\delta(\mathbb{M}_\mathbf{\Omega}^N(\mathbf{r}^n)-\mathbf{\Omega}^N) \tag{4.21}$$

Taking natural logarithm of both sides of the equation and rearranging gives

$$U(\mathbf{R}^N,\mathbf{\Omega}^N)) = -k_\mathrm{B}T\big(\ln g(\mathbf{R}^N,\mathbf{\Omega}^N)\big) + (\text{const}) \tag{4.22}$$

where (const) was added as the result of the proportionality in Eq (4.21), and

$$g(\mathbf{R}^N,\mathbf{\Omega}^N) = \int d\mathbf{r}^n \int d\mathbf{r}^n \exp\big(-u(\mathbf{r}^n)/k_\mathrm{B}T\big)\delta(\mathbb{M}_\mathbf{R}^N(\mathbf{r}^n)-\mathbf{R}^N)\delta(\mathbb{M}_\mathbf{\Omega}^N(\mathbf{r}^n)-\mathbf{\Omega}^N) \tag{4.23}$$

The CG forces are readily determined as gradients of this CG potential $U(\mathbf{R}^N)$. The force vector exerted on a specific CG site I is

$$\begin{aligned}
\mathbf{F}_{\mathrm{CG},I}(\mathbf{R}^N,\mathbf{\Omega}^N) &= -\frac{\partial U(\mathbf{R}^N,\mathbf{\Omega}^N)}{\partial \mathbf{R}_I} \\
&= -\frac{\partial\Big(-k_\mathrm{B}T\big(\ln g(\mathbf{R}^N,\mathbf{\Omega}^N)\big)+(\text{const})\Big)}{\partial \mathbf{R}_I} = \frac{k_\mathrm{B}T}{g(\mathbf{R}^N,\mathbf{\Omega}^N)}\frac{\partial g(\mathbf{R}^N,\mathbf{\Omega}^N)}{\partial \mathbf{R}_I} \\
&= \frac{k_\mathrm{B}T}{g(\mathbf{R}^N,\mathbf{\Omega}^N)}\frac{\partial\Big(\int d\mathbf{r}^n \int d\mathbf{r}^n \exp\big(-u(\mathbf{r}^n)/k_\mathrm{B}T\big)\delta\big(\mathbb{M}_\mathbf{R}^N(\mathbf{r}^n)-\mathbf{R}^N\big)\delta\big(\mathbb{M}_\mathbf{\Omega}^N(\mathbf{r}^n)-\mathbf{\Omega}^N\big)\Big)}{\partial \mathbf{R}_I} \\
&= \frac{k_\mathrm{B}T}{g(\mathbf{R}^N,\mathbf{\Omega}^N)}\int d\mathbf{r}^n \int d\mathbf{r}^n \exp\big(-u(\mathbf{r}^n)/k_\mathrm{B}T\big)\delta(\mathbb{M}_\mathbf{\Omega}^N(\mathbf{r}^n)-\mathbf{\Omega}^N) \\
&\quad \times \prod_{J\neq I}^{N}\delta\big(\mathbb{M}_{\mathbf{R},J}(\mathbf{r}^n)-\mathbf{R}_J\big)\frac{\partial}{\partial \mathbf{R}_I}\delta\big(\mathbb{M}_{\mathbf{R},I}(\mathbf{r}^n)-\mathbf{R}_I\big)
\end{aligned} \tag{4.24}$$

In the third line of Eq (4.24), the delta function $\delta(\mathbb{M}^N_{\mathbf{R}}(\mathbf{r}^n)-\mathbf{R}^N)$ was factorized as in Eq (4.17). The partial derivative $\partial/\partial\mathbf{R}_I$ applies only on the delta function that involves $\mathbf{R}_I$, whereas the FG potential $u(\mathbf{r}^n)$, delta functions of mapping to other sites J different from I, and delta functions of mapping to orientations can be taken out of the derivative.

Using the chain rule on the delta function $\delta(\sum c_{I,i}\mathbf{r}_i-\mathbf{R}_I)$, the following holds for any atom k that has $c_{I,k}\neq 0$:

$$\frac{\partial}{\partial\mathbf{R}_I}\delta\Big(\sum_{c_{I,i}\neq 0}c_{I,i}\mathbf{r}_i-\mathbf{R}_I\Big)=-\frac{1}{c_{I,k}}\frac{\partial}{\partial\mathbf{r}_k}\delta\Big(\sum_{c_{I,i}\neq 0}c_{I,i}\mathbf{r}_i-\mathbf{R}_I\Big)\tag{4.25}$$

It is convenient here to assume that only atoms specific to a CG site has non-zero mapping coefficients. We apply this assumption here with atoms specific to site I defined as $i\in I$. Now note that $\sum c_{I,k}=1$ (the dummy variables k and i are equivalent) so multiplying the right hand site of Eq (4.25) to this cancels the denominator:

$$\frac{\partial}{\partial\mathbf{R}_I}\delta\Big(\sum_{i\in I}c_{I,i}\mathbf{r}_i-\mathbf{R}_I\Big)=-\sum_{k\in I}\frac{\partial}{\partial\mathbf{r}_k}\delta\Big(\sum_{i\in I}c_{I,i}\mathbf{r}_i-\mathbf{R}_I\Big)\tag{4.26}$$

Substituting this in to the CG force equation (4.24) followed by integration by parts and cleaning up the equation give

$$\begin{aligned}
\mathbf{F}_{\mathrm{CG},I}(\mathbf{R}^N,\mathbf{\Omega}^N)&=-\frac{k_{\mathrm{B}}T}{g(\mathbf{R}^N,\mathbf{\Omega}^N)}\int d\mathbf{r}^n\int d\mathbf{r}^n\exp\big(-u(\mathbf{r}^n)/k_{\mathrm{B}}T\big)\delta(\mathbb{M}^N_{\mathbf{\Omega}}(\mathbf{r}^n)-\mathbf{\Omega}^N)\\
&\quad\times\prod_{J\neq I}^{N}\delta\big(\mathbb{M}_{\mathbf{R},J}(\mathbf{r}^n)-\mathbf{R}_J\big)\sum_{k\in I}\frac{\partial}{\partial\mathbf{r}_k}\delta\Big(\sum_{i\in I}c_{I,i}\mathbf{r}_i-\mathbf{R}_I\Big)\\
&=\frac{k_{\mathrm{B}}T}{g(\mathbf{R}^N,\mathbf{\Omega}^N)}\int d\mathbf{r}^n\int d\mathbf{r}^n\left(\sum_{k\in I}\frac{\partial}{\partial\mathbf{r}_k}\exp\big(-u(\mathbf{r}^n)/k_{\mathrm{B}}T\big)\right)\delta(\mathbb{M}^N_{\mathbf{\Omega}}(\mathbf{r}^n)-\mathbf{\Omega}^N)\\
&\quad\times\prod_{J\neq I}^{N}\delta\big(\mathbb{M}_{\mathbf{R},J}(\mathbf{r}^n)-\mathbf{R}_J)\delta\Big(\sum_{i\in I}c_{I,i}\mathbf{r}_i-\mathbf{R}_I\Big)\\
&=-\frac{k_{\mathrm{B}}T}{g(\mathbf{R}^N,\mathbf{\Omega}^N)}\int d\mathbf{r}^n\int d\mathbf{r}^n\frac{1}{k_{\mathrm{B}}T}\exp\big(-u(\mathbf{r}^n)/k_{\mathrm{B}}T\big)\delta(\mathbb{M}^N_{\mathbf{\Omega}}(\mathbf{r}^n)-\mathbf{\Omega}^N)\\
&\quad\times\prod_{J\neq I}^{N}\delta\big(\mathbb{M}_{\mathbf{R},J}(\mathbf{r}^n)-\mathbf{R}_J\big)\delta\big(\mathbb{M}_{\mathbf{R},I}(\mathbf{r}^n)-\mathbf{R}_I\big)\sum_{i\in I}\frac{\partial}{\partial\mathbf{r}_i}\big(-u(\mathbf{r}^n)\big)\\
&=-\frac{1}{g(\mathbf{R}^N,\mathbf{\Omega}^N)}\int d\mathbf{r}^n\int d\mathbf{r}^n\exp\big(-u(\mathbf{r}^n)/k_{\mathrm{B}}T\big)\delta(\mathbb{M}^N_{\mathbf{\Omega}}(\mathbf{r}^n)-\mathbf{\Omega}^N)\\
&\quad\times\prod^{N}\delta\big(\mathbb{M}_{\mathbf{R},J}(\mathbf{r}^n)-\mathbf{R}_J\big)\sum_{i\in I}\frac{\partial}{\partial\mathbf{r}_i}\big(-u(\mathbf{r}^n)\big)
\end{aligned}\tag{4.27}$$

We know that the FG model force on atom i is related to the potential as $\mathbf{f}_i=\frac{\partial}{\partial\mathbf{r}_i}\big(-u(\mathbf{r}^n)\big)$, and substituting in $g(\mathbf{R}^N,\mathbf{\Omega}^N)$ from Eq (4.23) yields

$$\begin{aligned}\mathbf{F}_{\mathrm{CG},I}(\mathbf{R}^N,\mathbf{\Omega}^N) &= \frac{\int d\mathbf{r}^n \int d\mathbf{r}^n \exp\big(-u(\mathbf{r}^n)/k_{\mathrm{B}}T\big)\delta(\mathbb{M}^N_{\mathbf{R}}(\mathbf{r}^n)-\mathbf{R}^N)\delta(\mathbb{M}^N_{\mathbf{\Omega}}(\mathbf{r}^n)-\mathbf{\Omega}^N)\big(\sum_{i\in I}\mathbf{f}_i(\mathbf{r}^n)\big)}{\int d\mathbf{r}^n \int d\mathbf{r}^n \exp\big(-u(\mathbf{r}^n)/k_{\mathrm{B}}T\big)\delta(\mathbb{M}^N_{\mathbf{\Omega}}(\mathbf{r}^n)-\mathbf{\Omega}^N)\delta(\mathbb{M}^N_{\mathbf{R}}(\mathbf{r}^n)-\mathbf{R}^N)} \\ &= \langle \mathbf{F}_{\mathrm{FG},I}(\mathbf{r}^n)\rangle_{\mathbf{R}^N,\mathbf{\Omega}^N}\end{aligned} \tag{4.28}$$

where $\mathbf{F}_{\mathrm{FG},I}(\mathbf{r}^n) = \overset{i\in I}{\sum}\mathbf{f}_i(\mathbf{r}^n)$ is the force in the FG model, and the angle brackets denote averaging of the function at thermal equilibrium with the constraint on FG configuration being consistent with the CG configuration. Therefore, Eq (4.28) shows that for a CG model to be consistent with the FG model in configuration space, one of the conditions is the force acting on a CG site is equal to the average sum of the forces acting on the atoms within that site in the FG model in an equilibrium state for a mapped FG configuration that matches the CG configuration.

4.6.2 Consistency condition in momentum space and expressions for coarse-grained masses and inertia tensors

The consistency condition in momentum space is described by Eq (4.16), showing the relation between the momentum probability distribution functions for CG models that include translational and rotational motions being consistent with the FG momentum probability function. The left side of this equation can be factorized out to distribution functions for each CG site I independently such as $P_{\mathrm{t}}(\mathbf{P}_I)P_{\mathrm{r}}(\mathbf{P}_I)$. The equation can only be satisfied using the same assumption, that no atom in the FG model is shared between two or more CG sites, so that the right side can be factorized to momentum probability functions for groups of atoms corresponding to each CG site separately.

With the above assumption, using Eq (4.16), it can be shown that the mass and inertia tensor of CG site I have to match those of the groups of atoms that belong to site I. In a simplified way, one can write

$$M_I = \sum_{i\in I} m_i \tag{4.29}$$

$$\mathbb{I}_{I,\mathrm{c}} = \sum_{i\in I} m_i(\mathbf{r}_{I,i}\otimes\mathbf{r}_{I,i}) \tag{4.30}$$

In words, the consistency condition in momentum space is fulfilled if atoms assembled into a CG site move and rotate together, as if they are fixed into a rigid body represented by that site. The mass M_I of CG site I can be defined as sum of the masses of atoms represented by that site m_i for $i \in I$, and the moment of inertia relative to the center of mass $\mathbb{I}_{I,\mathrm{c}}$ can be defined as the second moment of masses m_i with respect to the orthogonal distance from an axis fixed to the body, related to the distance from the atom position to the center of mass of the CG site $\mathbf{r}_{I,i}$. This ensures the linear and angular momenta are consistent between the two models.

4.6.3 Summary about consistency of coarse-grained models with anisotropy

To summarize, to transform a FG model to its CG description in which molecular anisotropy is important, the necessary conditions are

(1) The coordinate of each CG site is a linear combination of a set of FG atom coordinates of at least one atom, and each atom is only involved in one individual CG site.

(2) The CG potential is defined such that for each site the CG force is equal to average sum of the FG forces acting on the atoms involved in that site in a thermal equilibrium state for an FG configuration that is equivalent to the CG configuration, as expressed in Eq (4.28).

(3) The CG masses are as defined in Eq (4.29).

(4) The CG moments of inertia relative to the center of masses are as defined in Eq (4.30).

The major results of this section are Eqs (4.28), (4.29) and (4.30), in which the configurations and forces of CG sites are expressed in terms of those in the atomistic description.

4.7 Numerical implementation of coarse-grained force matching for anisotropic coarse-grained models

4.7.1 Variational principle for matching potential of mean force

In section 4.6.3, four conditions for which a CG model is consistent with a FG model were defined for the AFM-CG method. Conditions (1), (3) and (4) can readily be satisfied when one builds a CG model based on the atomistic description, whereas condition (2) requires mathematic algorithms that are able to produce the CG forces from the FG forces, that result from the potential of mean force (PMF) of many-body interactions in typical MD simulations. These FG forces $\mathbf{F}_{\mathrm{FG}}(\mathbf{r}^n)$ are readily computed from atomistic potentials during a simulation, and serve as input information for the coarse-graining process.

As for the to the MSCG method,[58] the variational principle can be used to perform this task for CG models that have directional dependence, which aims to find a solution for the CG forces $\mathbf{F}_{\mathrm{CG},I}(\mathbf{R}^N,\mathbf{\Omega}^N)$ in order to minimize the sum of squared residuals of the difference between the CG and FG forces:

$$\chi^2(\mathbf{F}_{\mathrm{CG}}) = \frac{1}{3N}\left\langle \sum_{I=1}^{N} |(\mathbf{F}_{\mathrm{FG},I}(\mathbf{r}^n)) - \mathbf{F}_{\mathrm{CG},I}(\mathbf{R}^N,\mathbf{\Omega}^N)|^2 \right\rangle \tag{4.31}$$

where $\mathbf{R}^N$ and $\mathbf{\Omega}^N$ are the CG coordinates and orientations mapped from $\mathbf{r}^n$ using the configuration mapping conditions developed in the AFM-CG method.

For simplicity, we only consider a particular type of interaction for which the CG force $\mathbf{F}_{\mathrm{CG},I}(\mathbf{R}^N,\mathbf{\Omega}^N)$ is a real continuous function. This does not affect generalizing the algorithm to the situation in which many CG interaction functions are included in the force field, as the resulting forces are the sum of the component forces computed from these interaction functions.

In implementing of the variational principle, a finite set of linearly independent vectors (called basis vectors) are defined to span a subspace of a functional CG force vectors, where the interaction potential and forces are non-zero. If it there are N_{B} of these basis vectors, this subspace $\mathfrak{F}_{\mathrm{CG}} \subset \mathbf{F}_{\mathrm{CG}}$ can be written as:

$$\mathfrak{F}_{\mathrm{CG}}(\mathbf{R}^N,\mathbf{\Omega}^N) = \sum_{b=1,\ldots,N_{\mathrm{B}}} \lambda_b \mathfrak{G}_b(\mathbf{R}^N,\mathbf{\Omega}^N) \tag{4.32}$$

where $\mathfrak{G}_b$ is a basis vector with corresponding real coefficient λ_b. With this approximation, Eq (4.31) becomes

$$\chi^2(\mathfrak{F}_{\mathrm{CG}}) = \chi^2(\boldsymbol{\lambda}) = \frac{1}{3Nt}\sum_{I=1}^{N} |(\mathbf{F}_{\mathrm{FG},I}(\mathbf{r}^n)) - \mathfrak{F}_{\mathrm{CG},I}(\mathbf{R}^N, \mathbf{\Omega}^N, \boldsymbol{\lambda})|^2 \tag{4.33}$$

where the angle brackets have been replaced with averaging over all t recorded atomistic MD snapshots. Here, if the CG coordinates $\mathbf{R}^N$ are defined based on the FG coordinates $\mathbf{r}^n$ and the consistency conditions, we can write the residual quantity as a function of $\boldsymbol{\lambda}$, the set of coefficients of the basis functions constructing the CG force field subspace in Eq (4.32). In the variational principle, if the amount of FG data is sufficient, determining the coefficients to minimize $\chi^2(\boldsymbol{\lambda})$ can be achieved using regression algorithms such as least squares approximation. The desired result from this is a force field function that can reproduce the FG many-body PMF in its CG representation, given the FG simulation information and the basis set.

In real MD simulations, the forces can result from different types of interactions, in which the potential energy has different forms. These interactions can be intra- or intermolecular, and can be between bonded or non-bonded CG sites:

$$\begin{aligned} U_{\mathrm{total}}(\mathbf{R}^N, \mathbf{\Omega}^N) = {} & \underbrace{U_{\mathrm{electrostatic}}(\mathbf{R}^N) + U_{\mathrm{vdw}}(\mathbf{R}^N, \mathbf{\Omega}^N)}_{\text{non-bonded}} \\ & + \underbrace{U_{\mathrm{bond}}(\mathbf{R}^N, \mathbf{\Omega}^N) + U_{\mathrm{angle}}(\mathbf{R}^N, \mathbf{\Omega}^N) + U_{\mathrm{dihedral}}(\mathbf{R}^N, \mathbf{\Omega}^N)}_{\text{bonded}} \end{aligned} \tag{4.34}$$

Here, the total electrostatic potential between pairs of CG sites can be readily evaluated from the charges of the individual particles and their positions using

$$U_{\mathrm{electrostatic}}(\mathbf{R}^N) = \sum_{J\neq I} \frac{1}{4\pi\varepsilon_0} \frac{Q_I Q_J}{|\mathbf{R}_I - \mathbf{R}_J|} \tag{4.35}$$

where $\mathbf{R}_I$ is the position, and Q_I is the charge assigned for CG site I as the sum of the charges of the individual atoms within that site. This potential can be evaluated directly for each simulation snapshot, in which the CG positions and charges are known. Thus, it is convenient to subtract this from U_{total} before using the AFM-CG method for other interactions.

Even though the bonded interactions depend on both positions and orientations of the CG sites, in general these interactions have simpler forms than short-range intermolecular interactions U_{vdw}. In many cases, the bonded potentials can be conveniently obtained using other CG methods, such as the IBI method,[100] to eliminate some problems that are associated with insufficient sampling. This can be done separately from evaluation of the more complicated potential functions U_{vdw} in which case the AFM-CG method is more suitable. This is particularly useful for CG models that consider molecular anisotropy. There have been a number of studies for spherical CG models of chemical and biological systems that have applied this hybrid approach, for which the method has been shown to give good structural accuracy.[60,61,64]

In summary, in applying the variational principle to construct CG many-body PMF using the AFM-CG method, it is convenient to use the method to compute only CG short-range intermolecular interactions U_{vdw} from the FG potentials. Bonded interactions can be obtained more readily using other CG methods such as IBI. The non-zero part of $\mathbf{F}_{\text{vdw}}$ can be constructed as a linear combination of a set of N_{B} basis vectors that depend on both positions and orientations of the CG sites $\{\mathcal{G}_1(\mathbf{R}^N, \boldsymbol{\Omega}^N), ..., \mathcal{G}_{N_{\text{B}}}(\mathbf{R}^N, \boldsymbol{\Omega}^N)\}$. Finding the coefficients $\{\lambda_1, ..., \lambda_{N_{\text{B}}}\}$ using linear regression to minimize a residue quantity of the difference of the FG and CG forces is the ultimate goal of the AFM-CG method.

4.7.2 Basis force vectors for coarse-grained forces and solving the force matrix

Above, it has been shown that the variational principle can be used to obtain the CG force field. This is completed by using least-squares methods to solve for the coefficients of basis functions that make up the force vector fields for all particles in the CG model. The basis functions may be as defined in Eq (4.32). However, if there are many CG sites in a model, the number of basis functions can be enormous. It is more convenient to define one set of basis functions for each type of interaction. For instance, if there is only one zero-charged CG site type in the system and none of them are bonded to the others, then U_{vdw} is the only type of interaction. In practice, CG interactions are normally assume to depend only a limited set of particles. These can be represented as CG non-bonded and bonded interactions that can be expressed in terms of a vector or scalar quantity representing the relative positions and orientations between the set of interacting particles for an anisotropic CG model. In this case, if U_{vdw} is assumed to be pairwise, it only depends on the relative positions, and orientations of the two interacting sites regardless of which sites they are.

As the force-matching equation (4.28) is for each CG site, it is simpler to consider the interactions with respect on one CG site I. The interactions for other sites can be expressed in a similar way. Let us define a set of CG sites that interact with I via van der Waals interaction as Γ. Any interaction potentials between I and any site J in Γ has a common form $U_{\text{vdw}}(\boldsymbol{\tau}_{IJ})$, where $\boldsymbol{\tau}_{IJ}$ is some set of variables that describe the relative position and orientation between the two interacting sites I and J. Expressing $U_{\text{vdw}}(\boldsymbol{\tau}_{IJ})$ as a linear combination of N_{B} scalar functions $\phi_b(\boldsymbol{\tau}_{IJ})$:

$$U_{\text{vdw}}(\boldsymbol{\tau}_{IJ}) = \sum_{b=1,...,N_{\text{B}}} \lambda_b \phi_b(\boldsymbol{\tau}_{IJ}) \tag{4.36}$$

Then the force acting on I due to a CG site J in Γ is

$$\mathbf{F}_{\text{vdw},IJ}(\boldsymbol{\tau}_{IJ}) = -\frac{\partial U_{\text{vdw}}(\boldsymbol{\tau}_{IJ})}{\partial \mathbf{R}_{IJ}} = -\sum_{b=1,...,N_{\text{B}}} \lambda_b \frac{\partial \phi_b(\boldsymbol{\tau}_{IJ})}{\partial \mathbf{R}_{IJ}} \tag{4.37}$$

The total force acting on I is the sum of the individual forces $\mathbf{F}_{\text{vdw},IJ}(\boldsymbol{\tau}_{IJ})$:

$$\mathbf{F}_I(\mathbf{R}^N, \boldsymbol{\Omega}^N) = \sum_{b=1,...,N_{\text{B}}} \lambda_b \mathcal{G}_{b,I}(\mathbf{R}^N, \boldsymbol{\Omega}^N) \tag{4.38}$$

where

$$\mathcal{G}_{b,I}(\mathbf{R}^N, \boldsymbol{\Omega}^N) = -\sum_{J\in\Gamma} \frac{\partial \phi_b(\boldsymbol{\tau}_{IJ})}{\partial \mathbf{R}_{IJ}} \tag{4.39}$$

Eq (4.39) is the form for the set of force basis vectors used to describe the force acting on each CG site I. This set of basis vectors can vary for different CG sites, but they can be derived from the same set of scalar basis functions of the potential energy defined in Eq (4.36). Solving for the N_{B} coefficients $\lambda_b = \{\lambda_1, ..., \lambda_{N_{\mathrm{B}}}\}$ gives a solution for $U_{\mathrm{vdw}}(\boldsymbol{\tau}_{IJ})$, which is the purpose of the AFM-CG method. The problem to be solved can be written in the following matrix equation:

$$\begin{bmatrix} \mathcal{G}_{1,I_x}(\mathbf{R}^N,\boldsymbol{\Omega}^N) & \mathcal{G}_{2,I_x}(\mathbf{R}^N,\boldsymbol{\Omega}^N) & \cdots & \cdots & \mathcal{G}_{N_{\mathrm{B}},I_x}(\mathbf{R}^N,\boldsymbol{\Omega}^N) \\ \mathcal{G}_{1,I_y}(\mathbf{R}^N,\boldsymbol{\Omega}^N) & \mathcal{G}_{2,I_y}(\mathbf{R}^N,\boldsymbol{\Omega}^N) & \cdots & \cdots & \mathcal{G}_{N_{\mathrm{B}},I_y}(\mathbf{R}^N,\boldsymbol{\Omega}^N) \\ \mathcal{G}_{1,I_z}(\mathbf{R}^N,\boldsymbol{\Omega}^N) & \mathcal{G}_{2,I_z}(\mathbf{R}^N,\boldsymbol{\Omega}^N) & \cdots & \cdots & \mathcal{G}_{N_{\mathrm{B}},I_z}(\mathbf{R}^N,\boldsymbol{\Omega}^N) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathcal{G}_{1,N_x}(\mathbf{R}^N,\boldsymbol{\Omega}^N) & \mathcal{G}_{2,N_x}(\mathbf{R}^N,\boldsymbol{\Omega}^N) & \cdots & \cdots & \mathcal{G}_{N_{\mathrm{B}},N_x}(\mathbf{R}^N,\boldsymbol{\Omega}^N) \\ \mathcal{G}_{1,N_y}(\mathbf{R}^N,\boldsymbol{\Omega}^N) & \mathcal{G}_{2,N_y}(\mathbf{R}^N,\boldsymbol{\Omega}^N) & \cdots & \cdots & \mathcal{G}_{N_{\mathrm{B}},N_y}(\mathbf{R}^N,\boldsymbol{\Omega}^N) \\ \mathcal{G}_{1,N_z}(\mathbf{R}^N,\boldsymbol{\Omega}^N) & \mathcal{G}_{2,N_z}(\mathbf{R}^N,\boldsymbol{\Omega}^N) & \cdots & \cdots & \mathcal{G}_{N_{\mathrm{B}},N_z}(\mathbf{R}^N,\boldsymbol{\Omega}^N) \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \vdots \\ \vdots \\ \vdots \\ \vdots \\ \lambda_{N_{\mathrm{B}}} \end{bmatrix} = \begin{bmatrix} \mathbf{F}_{1x} \\ \mathbf{F}_{1y} \\ \mathbf{F}_{1z} \\ \vdots \\ \vdots \\ \mathbf{F}_{Nx} \\ \mathbf{F}_{Ny} \\ \mathbf{F}_{Nz} \end{bmatrix} \tag{4.40}$$

Here, F_I is the force acting on the CG site I, which is obtained from the forces acting on atoms belonging to that site in the FG simulation. The basis force vectors $\mathcal{G}_{b,I}$ are 3D vectors as the result of the differentiation of the scalar potential basis functions with respect to a 3D vector in Eq (4.39). In the above matrix equation, the directional component of the basis force vectors are extracted, to match with the respective force component in the same direction. Thus, matching the CG force for each site contributes three rows in the force matrix. The same matching rules apply for all CG sites $1, ..., N$ in the system as shown in the matrix equation, and the matrix can also extend to different FG simulation snapshots by adding rows of basis force vectors to match the corresponding CG force vectors. If there are more force values to match than the number of coefficients N_{B}, then the matrix equation becomes an over-determined system where there are more equations to solve than unknowns. Solutions to this type of matrix systems are least squares solutions, which can be solved using a range of different regression methods. In this particular case, because the number of basis functions is normally large, up to thousands of functions, pre-processing of the force matrix is often needed in order to save computer storage. In the AFM-CG method, sequential Householder triangularization was applied to transform the over-determined problem to an equivalent full-rank problem,[101] following by solving the matrix using QR decomposition. In principle, the more simulation snapshots are used to construct the force matrix, the better convergence of the force solution.

This can also be applied when there is more than one potential interaction types in the system, which can all be represented using different sets of basis functions. The basis vectors $\mathcal{G}_{b,I}(\mathbf{R}^N, \boldsymbol{\Omega}^N)$ can be derived in a similar fashion as the sum of all basis potential functions for all interaction types.

4.7.3 Basis functions for coarse-grained potential

Above, the force matrix equation that is the main problem to solve in the AFM-CG method is described. Solving the matrix returns the coefficients for basis force vectors that best match the CG forces, which are derived from the force-matching equation, Eq (4.28). The basis force vectors are derived from scalar potential basis functions that construct a non-zero subset of a interaction potential type, as described in Eq (4.36). The form of these potential basis functions are thus important, as they initialize the whole process of the AFM-CG method.

In the MSCG method, different types of basis functions were tested. These include discrete delta functions, linear spline basis functions, and piecewise continuous cubic polynomials,[77] which are all functions of one variable. This is because the interaction potentials in the MSCG method are assumed isotropic, and depend only on the distance between CG sites. In the AFM-CG method, the interaction potentials depend on both relative position and orientation of the CG sites. Thus, multivariate basis functions must be used to perform this task.

Radial basis functions (RBFs) were chosen as the class of basis functions in the AFM-CG method. These are commonly used in approximations of multivariate functions in many different fields.[102] The approximate functions are usually linear combinations of translates of a scalar radially symmetric basis function, $\phi(||.||)$, where $||.||$ denotes the Euclidean norm,

$$||\mathbf{x}||_2 := \sqrt{x_1^2 + ... + x_n^2} \tag{4.41}$$

where $\mathbf{x} = (x_1, x_2, ..., x_n)$ is an n-dimensional vector.

RBFs have radial symmetry, which means that their values only depend on the Euclidean distance of an argument vector to the origin. For different purposes, the origin can be translated to a reference point **b**, which is called a RBF center. An approximate scalar function generally has a form of

$$f(\mathbf{x}) = \sum_{\mathbf{b} \in \mathbf{B}} \lambda_{\mathbf{b}} \phi(||\mathbf{x} - \mathbf{b}||) \quad \text{for} \quad \mathbf{x} \in \mathbb{R}^n \tag{4.42}$$

where $\lambda_{\mathbf{b}}$ are real coefficients, and **B** is the set of reference points at which the RBFs are constructed.

Variables of potential basis functions

For applications of RBFs in the AFM-CG method, the center of each basis function represents a different relative position and orientation between two interacting particles. It is not convenient to use the actual positions and orientations of the two sites $(\mathbf{R}_I, \mathbf{\Omega}_I, \mathbf{R}_J, \mathbf{\Omega}_J)$ as variables for RBF centers, as each position vector has three components and each orientation matrix has 9 components. The relative position and orientation can be represented by the following variables:

- Centre to centre distance: $R_{IJ} = ||\mathbf{R}_I - \mathbf{R}_J||$

- Three Euler angles that represent relative orientations between two coordinate frames $\mathbf{\Omega}_I$ and $\mathbf{\Omega}_J$ of the respective site I and J. These angles $(\alpha_{IJ}, \beta_{IJ}, \gamma_{IJ})$ can be computed from the orientations of the two sites $\mathbf{\Omega}_I = (\mathbf{\Omega}_{Ix}, \mathbf{\Omega}_{Iy}, \mathbf{\Omega}_{Iz})$ and $\mathbf{\Omega}_J = (\mathbf{\Omega}_{Jx}, \mathbf{\Omega}_{Jy}, \mathbf{\Omega}_{Jz})$ (see Appendix C for more details). For CG sites that has only one unique rotation axis, such as an uniaxial ellipsoid, only β_{IJ} is needed to describe the relative orientation.

• Scalar parameters that describe the angles between the center-to-center vector and a principal axis for each site: $A_{IJ} = |\cos(\mathbf{R}_{IJ}, \mathbf{\Omega}_{Ix})|$ and $B_{IJ} = |\cos(\mathbf{R}_{IJ}, \mathbf{\Omega}_{Jx})|$ where $|.|$ denotes absolute values.

In total, to construct the potential basis functions, there are 4 variables for CG sites with only one unique rotation axis $(R_{IJ}, \beta_{IJ}, A_{IJ}, B_{IJ})$, and 6 variables for CG sites without this uniaxial symmetry $(R_{IJ}, \alpha_{IJ}, \beta_{IJ}, \gamma_{IJ}, A_{IJ}, B_{IJ})$. The RBF centers can be uniformly distributed grid points of these variables in four or six dimensions.

In implementation of the AFM-CG method, compactly supported RBFs were used, in particular a class of functions called the Wendland's RBFs.[103] Evaluation of these RBFs built on a center at a particular position and orientation depends on the Euclidean distance between to the reference point. For example, in the case of uniaxial CG sites with center $\xi = (R_c, \beta_c, A_c, B_c)$, the RBF built on ξ, evaluated at relative position and orientation (R, β, A, B) is

$$f_\xi(d) = \begin{cases} 0, & \text{if } d > 1. \\ \phi(d), & \text{otherwise.} \end{cases} \tag{4.43}$$

where

$$d = ||(R, \beta, A, B) - (R_c, \beta_c, A_c, B_c)|| \tag{4.44}$$

and $\phi(d)$ is a univariate polynomial function of d.

The RBF in Eq 4.43 are supported in a unit sphere with radius $d = 1$, but the radius can be scaled when use in applications. More details about different types of RBFs and examples of RBFs on grid points can be found in Appendix D.

4.8 Conclusions and outlook

To conclude, the AFM-CG method was successfully developed for parametrization of CG models considering molecular anisotropy. Several mapping conditions, together with the force matching equation were proved, setting up rules for constructing consistent CG models. In particular, each CG site contains a separate set of atoms, with the mass and inertia tensor of the CG site matching that of the FG group of atoms. With these mapping conditions, it was shown that the force acting on a CG site is the sum of the individual forces in the corresponding FG model with the same configuration. The variational principle was invoked to make the force matching equation a regression problem in practical implementation of the method. Due to the complicated form of the short-ranged non-bonded pairwise potential, the AFM-CG method should be used to parametrize this potential function alone. Other simpler potentials can be obtained using simpler methods to complete an anisotropic CG model with both bonded and non-bonded interactions. The general form of the basis force vectors that are used in constructing the CG forces was specified, which are derived from scalar potential basis functions with multiple variables to describe the relative orientations and positions between a pair of interacting particles. These potential functions can take the form of RBFs, a class of functions that are commonly used for approximations of multivariate functions. Solving for the force function can be completed using numerical methods for least squares regression.

It is necessary to validate the performance of the method via CG parametrization and simulations of anisotropic molecular systems. In Chapter 5, the method is applied to produce a uniaxial ellipsoidal model of benzene. The success of the method can be confirmed if the CG model is capable of reproducing structural and thermodynamic properties for the FG model. Once validated, the method can be applied to produce CG models of organic semiconductors interfaces of the systems discussed in Chapter 3.

5

Applications of coarse-grained modelling with anisotropy

5.1 Abstract

The anisotropic force-matching coarse-grained (AFM-CG) method, proposed in Chapter 4, is a method for systematically determining the potential energy function for a coarse-grained (CG) model with anisotropy that is thermodynamically consistent with a fine-grained (FG) model. The method uses matching of forces produced from molecular dynamics (MD) simulation of the FG model. A systematic algorithm using variational principle for multivariate functions was also developed, allowing practical parametrization of anisotropic CG models. In order to validate the theory and performance of the algorithm, a CG model of liquid benzene is developed, and simulation results using this model are compared to those of the FG model and other CG models of benzene. By comparing several structural and thermodynamic properties, the AFM-CG model is shown to give good agreement with the FG model and to be an improvement over other CG models parametrized using other methods. The AFM-CG method is validated as a systematic method for parametrizing CG models of anisotropic molecules, and as such can be particularly using for MD simulations of materials for which molecular anisotropy plays an important role in controlling mesoscale structural properties.

5.2 Introduction

Molecular dynamics (MD) simulations allow the structure and dynamics of disordered condensed-phase molecular systems to be studied, which in general can be hard to achieve experimentally. However, realistic molecular systems normally contain an enormous number of atoms, and simulations of such systems require large time scales to achieve useful results, making simulations of every atom in atomistic/fine-grained (FG) models computationally expensive and often infeasible. A solution to this problem is to use coarse-grained (CG) models, in which many atoms are grouped and simulated together as single particles. The reduced number of degrees of freedom in these models allow one to simulate a system consisting of millions of atoms, on time scales of of microseconds or more, allowing simulations of mesoscale self-assembly under experimental relevant conditions. Many systems have been successfully investigated using CG molecular models, including systems of liquids,[60] polymers,[12] and biological molecules.[55]

A number of "bottom-up" parametrization methods for CG models have been developed in the literature. These include the iterative Boltzmann inversion (IBI) method, which matches probability distribution functions,[53,54] the multi-scale coarse-graining (MSCG) method,[58,59] which matches forces, and the relative entropy method,[63] which minimizes the relative entropy between the CG and FG simulation systems. Improvement of these parametrization methods has been achieved over the years,[99,104] producing more accurate CG models. Nevertheless, the potential energy functions for non-bonded CG interactions produced by these methods have generally been isotropic, making each CG site spherical bead. In systems that have anisotropy, more CG sites are required in order to represent this anisotropic nature accurately. This increases the degrees of freedom, increasing the computational cost of MD simulations and making the analysis of results more complicated.

CG models that have inherent molecular anisotropy have been developed and used in simulations of organic semiconductor materials,[65,67] as variations in structure and interactions in different directions are important for this class of materials. In CG models with anisotropy, CG sites normally represent a group of atoms that move together, resulting in defined rigid shapes and directions during simulations. This allows more direct mapping schemes from the FG to CG level with fewer sites than isotropic models, and simpler analysis of the CG simulation results due to the anisotropic nature of the models. The CG method used for these models often involve sampling of the potential for various relative orientations, and fitting these to an analytical form.[67,105] This method in general can produce physically meaningful potential functions, but the accuracy cannot be guaranteed due to the potential neglect of the thermal fluctuations and many-body interactions that occur in condensed-phase systems.

In Chapter 4, a systematic CG method termed the anisotropic force-matching coarse-grained (AFM-CG) method was introduced, with a proof of mapping conditions from detailed FG models to CG models using rigorous statistical mechanical theory. In implementing the AFM-CG method, FG simulation forces are used as input data for constructing the directional CG forces using a variational principle for multivariate functions. This guarantees systematic matching of the force functions, producing thermodynamically consistent potential of mean force between the two models. In the following section, the AFM-CG theory and algorithm are used to parametrize of a single-site disk-shaped CG model of liquid benzene. Simulation results of this model are compared with those from FG simulations under the same conditions in order to validate the accuracy of the method. Comparison with previously reported spherical and disk-shaped models of benzene is also completed to emphasize the improvements of the AFM-CG method over other available CG methods.

5.3 Computational details

5.3.1 Atomistic simulation of liquid benzene

A FG model of 500 benzene molecules was simulated using the MD simulation package LAMMPS (version 22Aug18).[85] Interaction parameters used were from the OPLS-AA force field (version OPLS2005)[70] without modification, with a cutoff distance for the van der Waals interactions and the short-ranged real-space part of the calculation of Coulombic interactions of 10.0 Å. The long-ranged component of the electrostatic interactions were calculated using the particle-particle particle-mesh (PPPM) method.[73] The simulation was started with the molecules arranged in a lattice in a cubic simulation box with periodic boundary conditions, with the initial volume sufficiently large to avoid particle overlaps. The energy was minimized, then the system was simulated in the NVT ensemble with temperature $T = 300$K for 20 ps with initially small timesteps for relaxation of the system, followed by simulation in the NPT ensemble for 5 ns to obtain equilibrium. During the 10 ns production time, the simulation was performed with a timestep of 1 fs in the NPT ensemble with the temperature $T = 300$ K and pressure $p = 1$ atm maintained by the Nose-Hoover thermostat and barostat.[74,106] The coordinates and forces for each atom in the system were outputted every 1000 time steps (1 ps) during this production time, to produce a total of $t = 10000$ simulation snapshots as input data for parametrization of the CG model. The same procedure was used to carry out simulations at T = 350 K and p = 1 atm.

5.3.2 Representation of an anisotropic coarse-grained model of benzene

As proved in Chapter 4, for an anisotropic CG model to be consistent with the FG model in configuration space, the coordinates of each CG site must be a linear combination of the coordinates of at least one atom, and each site must represent a specific group of atoms that do not overlap with those of other sites. The masses and moment of inertia of the CG have to be able to reproduce those in the FG model.

An example of such configuration and geometric mappings from the FG to the CG model is shown in Figure 5.1. Here, each benzene molecule is represented as an ellipsoid in the CG model, with CG mass for a molecule/site equal to sum of the masses of its n atoms, as below:

$$M = \sum_{i=1}^{n} m_i \tag{5.1}$$

The ellipsoidal coordinate is defined as the center of mass of the n atoms:

$$\mathbf{R}_\mathrm{c} = \frac{\sum_{i=1}^{n} m_i \mathbf{r}_i}{\sum_{i=1}^{n} m_i} \tag{5.2}$$

Here, M and $\mathbf{R}_\mathrm{c}$ are the CG mass and coordinates of the molecule/site, determined from the component masses $(m_1, m_2, ..., m_n)$, and the coordinates $(\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_n)$ of the atoms in a benzene molecule.

The inertia tensor of the CG site should match the inertia tensor with respect to the center of mass of benzene. This is computed as the second moment of masses $m_{i,I}$ with respect to the orthogonal distances from an axis fixed to the body. These orthogonal distances are calculated as elements of the outer products of the vectors $(\mathbf{r}_{\mathrm{c},1}, \mathbf{r}_{\mathrm{c},2}, ..., \mathbf{r}_{\mathrm{c},n})$ to themselves. These vectors connect the center of mass to each atom. The inertia tensor of the CG site is

$$\mathbb{I}_\mathrm{c} = \sum_{i=1}^{n} m_i (\mathbf{r}_{\mathrm{c},i} \otimes \mathbf{r}_{\mathrm{c},i}) \tag{5.3}$$

with $\mathbf{r}_{\mathrm{c},i} = \mathbf{r}_i - \mathbf{R}_\mathrm{c}$ connecting each atom to the center of mass.

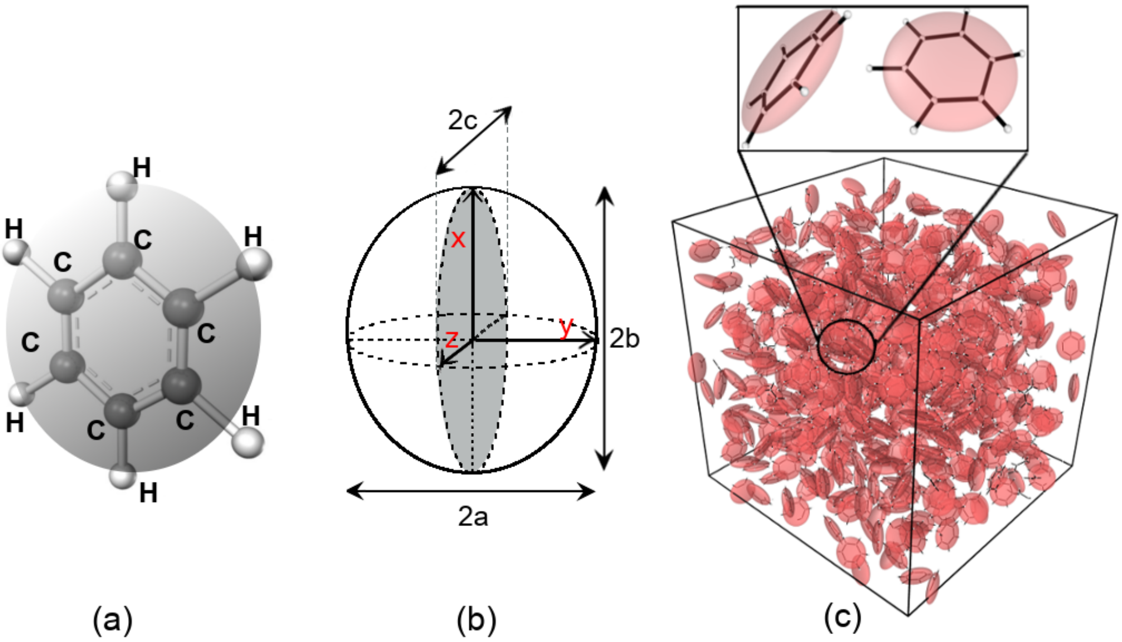


Figure 5.1: Geometric mapping of atomistic benzene to an ellipsoidal model. (a) Mapping of an atomistic benzene molecule to an ellipsoid, (b) axial lengths of benzene CG ellipsoid, and (c) atomistic system of 500 benzene molecules overlayed with their ellipsoidal representations, with their positions, orientations, and geometric properties calculated from the anisotropic CG algorithm.

The inertia tensor expressed in Eq (5.3) is useful, as it allows derivation of principal axes and axial lengths of the ellipsoid. The principal axes of the ellipsoid are determined as the three vectors $(\mathbf{x}, \mathbf{y}, \mathbf{z})$ (see Figure 5.1b) that diagonalize the inertia tensor.[107] This is the same with solving an eigenvalue problem, where the three eigenvectors are the principal axes required. The eigenvalues with respect to these eigenvectors (I_1, I_2, I_3) also provide the axial lengths $(2a, 2b, 2c)$ for ellipsoids, such as

$$a = \sqrt{\frac{5(-I_1+I_2+I_3)}{2M}}, \quad b = \sqrt{\frac{5(I_1-I_2+I_3)}{2M}}, \quad c = \sqrt{\frac{5(I_1+I_2-I_3)}{2M}} \tag{5.4}$$

where $M = \sum_{i=1}^{n} m_i$ is the CG mass. The relative lengths of $2a, 2b$, and $2c$ are as illustrated in Figure 5.1b, if (I_1, I_2, I_3) are sorted in ascending order before evaluating the semi-axial lengths.

5.3.3 Coarse-grained parametrization for ellipsoidal model of benzene

For simplicity, each benzene molecule is coarse-grained into one uniaxial disk, for which two of the three rotational axes of the molecules are equivalent. The interactions between the particles in this CG model are described as non-bonded interactions between pairs of identical uniaxial disks, generally denoted as $U_{\text{vdw}}(\boldsymbol{\tau}_{IJ})$, where $\boldsymbol{\tau}_{IJ}$ is a set of variables that describe the relative positions and orientations of the particles I and J. Also, there is only one type of potential of mean force (PMF) function in this model because the disks are all identical. This is the simplest case in which the AFM-CG method can be used directly.

This pair potential depends on the coordinates and orientations of the pair of interacting particles. In the AFM-CG method, the non-zero part of the resulting short-ranged pairwise non-bonded potential is expressed as a linear combination of a set of N_{B} scalar basis functions:

$$U_{\mathrm{vdw}}(\boldsymbol{\tau}_{IJ}) = \sum_{b=1,\ldots,N_{\mathrm{B}}} \lambda_b \phi_b(\boldsymbol{\tau}_{IJ}) \tag{5.5}$$

where $\phi_b(\boldsymbol{\tau}_{IJ})$ are the basis functions with real coefficients λ_b.

The resulting pairwise force is

$$\mathbf{F}_{\mathrm{vdw},IJ}(\boldsymbol{\tau}_{IJ}) = -\frac{\partial U_{\mathrm{vdw}}(\boldsymbol{\tau}_{IJ})}{\partial \mathbf{R}_{IJ}} = -\sum_{b=1,\ldots,N_{\mathrm{B}}} \lambda_b \frac{\partial \phi_b(\boldsymbol{\tau}_{IJ})}{\partial \mathbf{R}_{IJ}} \tag{5.6}$$

where I and J denote the pair of interacting uniaxial disks, and $\mathbf{R}_{IJ} = \mathbf{R}_I - \mathbf{R}_J$ is the vector connecting the centers of the two particles located at $\mathbf{R}_I$ and $\mathbf{R}_J$. If the above is the force vector from particle J acting on particle I, summation of all of the forces between I and a set of particles interacting with I, denoted by Γ, yields the total force vector acting on I:

$$\mathbf{F}_I(\mathbf{R}^N, \mathbf{\Omega}^N) = \sum_{b=1,\ldots,N_{\mathrm{B}}} \lambda_b \mathfrak{G}_{b,I}(\mathbf{R}^N) \tag{5.7}$$

where

$$\mathfrak{G}_{b,I}(\mathbf{R}^N) = -\sum_{J\in\Gamma} \frac{\partial \phi_b(\boldsymbol{\tau}_{IJ})}{\partial \mathbf{R}_{IJ}} \tag{5.8}$$

Eq (5.8) is the form for the set of force basis vectors used to describe the force acting on each CG site I. This set of basis vectors can vary for different CG sites, but they can be derived from the same set of scalar basis functions of the potential energy defined in Eq (5.5). Solving for the N_{B} coefficients $\lambda_1, \ldots$, and $\lambda_{N_{\mathrm{B}}}$ gives solutions for $U_{\mathrm{vdw}}(\boldsymbol{\tau}_{IJ})$ and $\mathbf{F}_{\mathrm{vdw}}(\boldsymbol{\tau}_{IJ})$.

As described in Chapter 4, $(\mathbf{R}_I, \mathbf{\Omega}_I, \mathbf{R}_J, \mathbf{\Omega}_J)$ is a set of vectors and matrices that describe the specific positions and orientations of sites I and J in 3D, and is not a good choice for the variables of the basis functions. Instead, as the potential depends on the relative position and orientation of one particle with respect to the other, a smaller set of variables can be used to defined the force. In particular, for uniaxial CG sites such as assumed for the model in this work, where there is only one unique axis of rotation, only four variables need to be used in the basis functions to define pairwise potential and forces. This reduces the number of basis functions and computational cost of the AFM-CG method.

One set of basis functions $\phi_b(\boldsymbol{\tau}_{IJ})$ is needed to represent the pairwise non-bonded force function. The basis functions used were discrete radial basis functions (RBFs) of four variables that describe the relative position and orientation between two CG sites. The basis vectors were Wendland basis functions, a class of compactly supported/piece-wise basis functions.[103] These functions were built on 7200 uniformly distributed grid points that span the non-zero potential function $U_{\mathrm{vdw}}(\boldsymbol{\tau}_{IJ})$. More about RBFs used in the AFM-CG method can be found in Appendix D. The results were solved using the optimization tool linalg.lstsq available from Numpy python package,[108] with the force matrix prepared using a sequential QR factorization routine.[101] Details of these numerical methods can be found in Chapter 4.

5.3.4 Coarse-grained simulations for uniaxial disk-shaped models of benzene

The AFM-CG force results were fitted to an analytical form of pairwise non-bonded interaction for disk-shaped molecules, the S-function expansion model implemented by Bowen et al.[67] This allows simplicity for running MD simulation in LAMMPS,[85] but in principle the AFM-CG forces constructed by expansion of the basis functions can be used directly with additional implementation in the software. Fitted parameters can be found in Appendix E, with a cut-off distance of 10 Å. The system contained 500 uniaxial disk-shaped molecules packed together with an initial density of 0.96 g/cm^3. Simulation in the NPT ensemble maintained by the Nose-Hoover barostat and thermostat,[74,106] at temperature T = 300K and pressure p = 1 atm using small timesteps initially allows the system to relax slowly, before reaching an equilibrium. Simulation at equilibrium is done with a production time of 20 ns, using a timestep of 5 fs. Simulations of the parametrized CG model were also carried out at T = 350 K and p = 1 atm. The same procedure was used to simulate the CG model of benzene published by Bowen et al.[67] for comparison of structural and thermodynamic properties between the two models. S-function parameters for this model can be found in Appendix E.

5.4 Results and discussion

The major goal of this chapter is to validate the performance of the AFM-CG method presented in Chapter 4, using the uniaxial disk model of benzene described in the previous section. A successful CG parametrization should give a model that can reproduce structural and thermodynamics properties of the FG model.

5.4.1 Pairwise interaction force for coarse-grained uniaxial disk-shaped benzene

Results of the uniaxial disk-shaped model of benzene are the coefficients of the basis functions in Eq (5.6). The force vector field $\mathbf{F}_{\mathrm{vdw},IJ}(\boldsymbol{\tau}_{IJ})$ can be derived as the linear combination of the basis functions vectors with the coefficients found. This force vector function covers the pairwise interactions between a CG benzene dimer at any relative orientation between the two molecules. Therefore, this function is a complicated function of four variables for uniaxial particles that is difficult to graphically represent. It is easier to extract a part of the force function at some special configurations of the dimer, as shown in Figure 5.2. In Figure 5.2, force functions resulting from parametrization using different amounts of input data from the FG simulation described in section 5.3 are shown. Here, the number of snapshots used are 4000, 6000, and 10000. It can be seen that the results converge for the three dimer configurations using 6000 or more snapshots.

As the number of RBFs used is large, building the force matrix before solving for the least-squares solution of the function is time-consuming. Computational time to build the force matrix is proportional to the number of snapshots used, with an average of 10 minutes for one snapshot on one computer processor. Parallel computing helps to reduce the time needed for this process by the factor of the number of computer processors available. Once the force matrix is built, solving for the least-squares solution can be less computationally expensive with prior sequential QR factorizations to the matrix. After optimizing this process, only one hour of computational time was required for solving the solution for 10000 snapshots, using only a single computer processor.

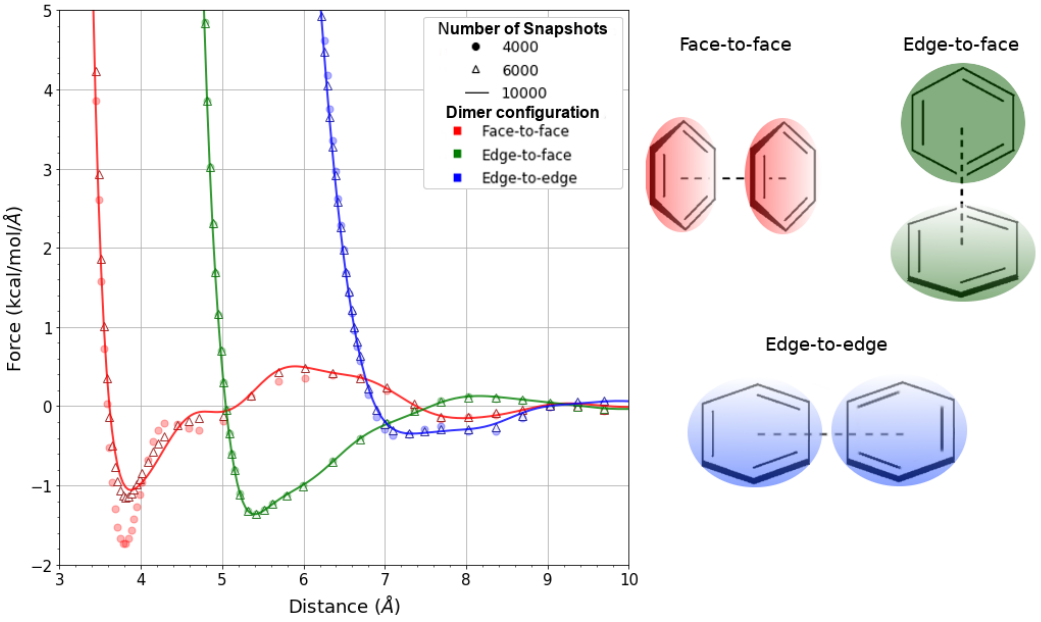


Figure 5.2: Force curves versus separation distance for several different configurations of CG benzene dimers (left) for different amounts of input data, and the corresponding configurations of the benzene dimers (right). Note that the direction of the forces aligns with the dimer's center-to-center vectors, shown by dashed lines connecting the molecules.

5.4.2 Fitting to an analytical potential form

The pairwise non-bonded force function obtained from the AFM-CG method for benzene was fitted to a formulated CG potential form, termed the S-function expansion potential model implemented by Bowen et al.[67] This potential model is designed for disk-shaped CG particles, and has been shown to better describe the pairwise non-bonded interaction between these molecules over other potential models[67] such as the Gay-Berne potential for ellipsoids.[109]

Figure 5.3 shows the force functions obtained using the AFM-CG method and those fitted to an S-function potential at various configurations as specified. It can be seen that fitting the calculated force results to the S-function form is challenging, resulting in deviations in many configurations. Also, some features that appear in the AFM-CG force curves cannot be represented using an analytical potential, such as for the long-range positive regions that appear in almost all dimer configurations.

5.4.3 Structural and thermodynamic properties of coarse-grained uniaxial disk-shaped benzene simulations

An S-function potential that produced the force curves shown in Figure 5.3 was used for CG MD simulations for the AFM-CG model. Figure 5.4 compares the S-function fitted AFM-CG force functions to those of other CG models. In particular, the comparison is between three CG models for benzene: the AFM-CG model, the Bowen model,[67] and a spherical model parametrized using isotropic force-matching.[110] All three models were parametrized for liquid benzene at a temperature of 300 K. Bowen et al.[67] developed their anisotropic CG model of benzene by fitting the CG potential to the FG potential, which was calculated as the sum of pairwise non-bonded potentials between all atom pairs in a benzene dime for a number of randomly selected dimer configurations. The spherical model of benzene was

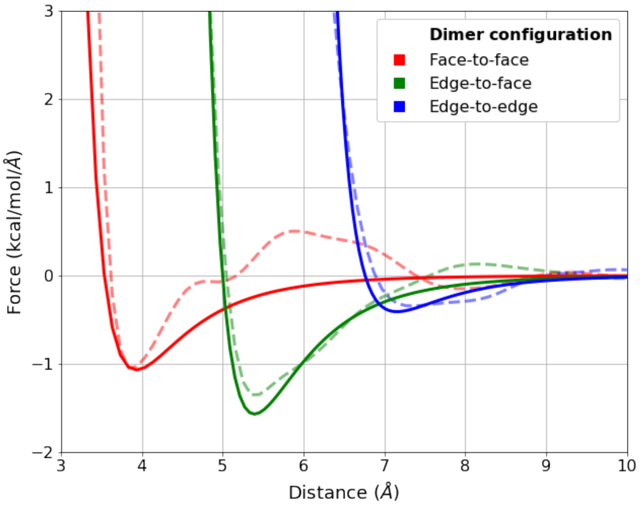


Figure 5.3: Fitting of AFM-CG forces to the S-function form used by Bowen et al.[67] for various benzene dimer configurations defined in Figure 5.2. The dashed lines show the AFM-CG forces, and the S-function fitting is shown by solid lines.

parametrized using isotropic force matching.[110] Comparison of the accuracy between the published CG models and the AFM-CG model developed in this work will help validate the performance of the method compared with other CG methods.

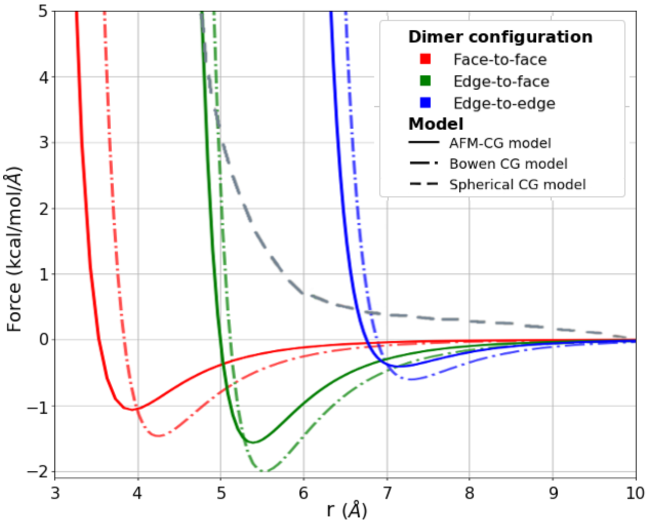


Figure 5.4: Pair force functions of AFM-CG (solid curves) and Bowen model (dot-dash curves) for three CG benzene dimer configurations: face-to-face (red), edge-to-face (green), and edge-to-edge (blue), together with isotropic pair force function (dashed curves) adopted from Pranami et al.[110]

Average densities of benzene in NPT simulations of the FG models, the AFM-CG and the Bowen models, together with experimental data are shown in Table 5.1. The average densities calculated in FG model simulation reasonably agrees with experimental values,[111] deviating by less than 5%. The AFM-CG model yields densities close to those of the FG model with less than 4% difference for both temperatures. In comparison, the Bowen model overestimated the density, with at most 15% deviation from those for the FG model, and nearly 10% from the experimental density at 350 K. This could be due to overestimation of

Table 5.1: Experimental and simulated densities of benzene at 1 atm for 300 and 350 K

	Density ($\mathrm{g\,cm^{-3}}$)			
Temperature	exp[a]	FG	AFM-CG	Bowen[67]
300 K	0.872	0.842	0.875	0.934
350 K	0.818	0.780	0.800	0.898

[a] Experimental data from ref. [111], FG, AFM-CG, and Bowen denote values from simulations using the respective models.

the force well depths for this model as shown in Figure 5.4. Even though the average densities of the spherical model published by Pranami et al.[110] are not calculated, due to their force function not being fitted into an analytical form for MD simulations, one would not expect the density to be close to the FG or the experimental densities. This is because CG models produced by isotropic force matching without pressure correction have been shown to not capture the density correctly in NPT simulation.[64] Therefore, it can be concluded that the AFM-CG model perform the best in matching the density to the FG model.

In addition to matching the density, it is important to show that the CG force function in a CG simulation can reproduce relevant structural properties of the FG system, expressed by a set of distribution functions. The simplest and most commonly used for homogeneous systems is the radial distribution function (RDF) $g(r)$. This function gives the probability of finding a pair of particles a distance r apart within the whole simulation system:

$$g(r) = \frac{N-1}{4\pi\rho r^2}\langle\delta(r-r_{ij})\rangle_{ij} \tag{5.9}$$

where N is the total number of particles, ρ is the number density of the system, and r_{ij} is the distance between the centers of mass of particles i and j.

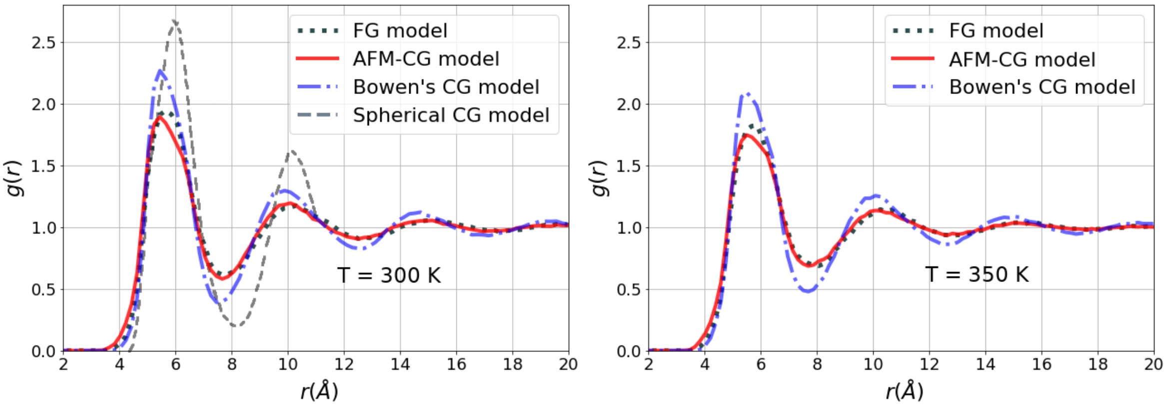


Figure 5.5: Comparison between RDFs of the CG models to the FG model of benzene at 300 and 350 K. (For the spherical model,[110] the RDF data is not available for distances greater than 11 Å or for $T = 350$ K.)

Figure 5.5 compares the RDFs of the CG models to the FG model at 300 and 350 K. It can be seen that the AFM-CG model performs the best out of the three CG models, evidenced by outstanding matching of the RDF to that for the FG model. The RDF of the spherical model has the largest deviation to the FG data, resulting from the dimer pair force not parametrized to describe anisotropy. The peak position and height are shifted to the right due to positive force at long dimer separation distance. Also, this RDF for the spherical model is from simulation in the NVT ensemble without pressure control. One would not expect the same result in an NPT simulation, as parametrization using force-matching method for spherical models without pressure correction normally fails to correctly describe the density and therefore the structural properties compared to the atomistic level. On the other hand, the model parametrized using the AFM-CG method allows direct parametrization of the forces for all directions, and simulations in NPT are in good agreement with the FG model. Compared to the anisotropic model of Bowen, the AFM-CG model also shows improvement in matching to the RDF of the FG model. The higher peaks at distances of 5.5 Å and 10 Å for the Bowen model are possibly due to overestimation of the force function well depths in comparison to the AFM-CG model. This could be a result of the unsystematic method used to parametrize this model, in which molecular fluctuations and many-body interactions may not be fully considered. Although parametrization was completed at 300 K, the RDFs at 350 K of the FG and AFM-CG model still match well, indicating good transferability of the model over temperatures in this range.

The orientational order parameter S_2 is also an important structural quantity that characterizes possible liquid-crystalline phases of anisotropic molecules. In general, it measures the average alignment of molecules within a system along a preferred direction. Here, a local order parameter $S_2(r)$ that describes the average alignment of molecules separated by a distance r is investigated:

$$S_2(r) = \langle P_2(\cos\theta(r))\rangle = \langle \frac{3\cos^2\theta(r) - 1}{2} \rangle \tag{5.10}$$

where $P_2(...)$ is the second Legendre polynomial, $\theta(r)$ is value of the angle between the normal axes of two benzene molecules that are a distance r apart. The angle brackets denote averaging over all molecules and time.

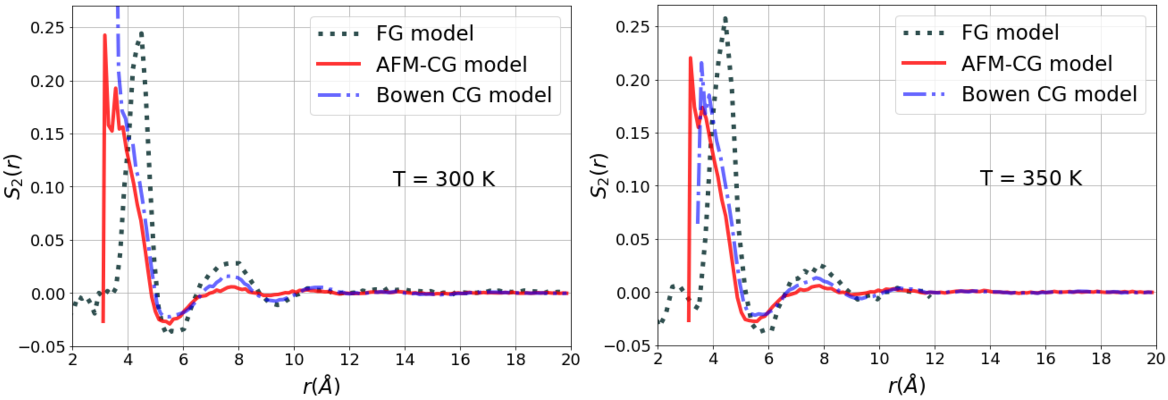


Figure 5.6: Pair order parameter versus separation distance for simulations using FG and CG models of benzene at 300 K and 350 K

Figure 5.6 shows the order parameter as a function of separation distance between pairs of molecules for the FG and CG models at 300 and 350 K. The order parameters for the AFM-CG and Bowen models have quite similar features for both temperatures, and are both comparable to the FG model. The most obvious difference between $S_2(r)$ of the FG and CG models is in the position and height of the first peaks which is at about 4 Å in the CG model compared with 4.5 Å for the FG model. This peak is due to the face-to-face π-stacking interactions between benzene molecules, which can be affected by the rigid shape of CG particles compared to with the flexible bonded atoms in the FG model. For the AFM-CG model, the peak is shifted to slightly shorter distance than the Bowen CG model. This may be because for the force become repulsive at shorter distances in the S-function fitted AFM-CG compared with the Bowen model. Overall, CG models of benzene using disk-shaped particles effectively capture the molecular orientational correlations of the FG model, which is not possible a single-site spherical model. Therefore, this demonstrates that these models allow accurate and simple interpretation of the orientations of molecules in relevant applications.

5.5 Conclusions and outlook

In summary, the AFM-CG method, which uses force-matching to parametrize simpler anisotropic CG models from FG models, was used to parametrize an anisotropic model of liquid benzene. The accuracy of the method was successfully validated by good agreement between simulation results of the CG model and the underlying FG model for various structural and thermodynamic properties. The AFM-CG model also shows improvements over a spherical model and a similar disk-shaped model of benzene that were parametrized using other methods.[67,110] The method could be applied more generally to MD simulations of any system of organic materials in which the effects of molecular anisotropy on mesoscale structural or dynamical properties are important.

6

Conclusions and future directions

"Learn, learn more, learn forever"

– Vladimir I. Lenin

6.1 Conclusions

In this work, several molecular simulation studies relating to molecular anisotropy have been presented. The first study relates to the structures at interfaces between organic semiconductor materials that are known to be affected by molecular anisotropy, which were investigated using atomistic/fine-grained (FG) molecular dynamics (MD) simulations. The second study relates to the development and validation of a method for systematically parametrizing simplified coarse-grained (CG) models from detailed FG models, which are able to describe molecular anisotropic properties and speed up MD simulations.

Chapter 3 presented an investigation into the structures at the interfaces between an organic photovoltaic (OPV) bulk-heterojunction system, consisting of benzodithiophene quaterthiophene rhodanine (BQR) as the electron donor, and [6,6]-phenyl-C71-butyric acid methyl ester ($PC_{71}BM$) as the electron acceptor. Parametrization of FG models for BQR and $PC_{71}BM$ followed by MD simulations were carried out for bulk systems of the pure materials and interface systems of mixtures. In particular for the interface systems, two initial configurations with different interfacial orientation arrangements were simulated: a face-on system in which the aromatic rings in BQR molecules faced $PC_{71}BM$ molecules, and a end-on system with one end of the BQR molecules pointing towards a $PC_{71}BM$ layer. Simulations were completed at temperatures near the experimental melting point of BQR (460–500 K),[47] allowing the molecules to be mobile enough to adopt preferred structural features at the interfaces. During simulations of the face-on interface system, BQR molecules maintained the initial orientations with respect to the interface, indicating that the face-on configuration is thermodynamically stable around the melting point of BQR. On the other hand, it was shown that the end-on configuration is not stable even below the melting points of both materials, with BQR molecules rotating to be parallel to the interface. These results agree with previously reported experimental X-ray scattering data[78,112–114] on the high ratio of face-on orientation between different donor materials with respect to fullerene-based acceptors at material interfaces in high-performing devices. Moreover, analysis of pair orientations between BQR molecules showed that BQR molecules close to the interface were less ordered with respect to one another than those far from the interface, which could assist charge separation by increasing energetic disorder at the interface. Together, these factors may enhance charge generation and transport, contributing to the outstanding performance of BQR:$PC_{71}BM$ system compared to other materials with similar chemical moieties but different molecular shapes that have been tested experimentally.[47]

In Chapter 4, a systematic CG parametrization method, termed the anisotropic force-matching coarse-grained (AFM-CG) method, was developed for CG models that take into account molecular anisotropy. Using a rigorous statistical mechanics basis, several mapping rules from a FG model to its anisotropic CG description were derived, together with a force-matching condition for thermodynamically consistent CG models to be developed. In particular, matching of the mass and inertia tensor of each CG site to those of the corresponding groups of FG atoms is necessary for consistency between the probability distribution functions for momenta in the two models. In addition, matching of the configuration distribution function provides an expression for the force exert on a CG site, as an average of the sum of the forces on the FG atoms that make up a CG site in equivalent configurations of the FG and CG systems.

In implementation of the method, a variational principle was used to make the force matching equation a regression problem, which can be solved in a least-squares manner. In particular, the CG force vector functions were written as a linear combination of basis force vectors multiplied scalar coefficients, with the solution being the set of coefficients that minimize the difference between the FG and CG forces. The basis force vectors were written in terms of scalar radial basis functions (RBFs), which are commonly used in approximation of multivariate functions, as in this case. In applications to CG parametrization of anisotropic molecules, the AFM-CG method may be used to construct pairwise non-bonded potentials as these potentials normally have complicated forms that depend on the relative position and orientation of interacting pairs. Other simpler potentials such as those describing bonded interactions can be more conveniently parametrized using less complicated CG methods. For applications to pairwise non-bonded interactions, the RBFs are functions of many variables that describe the relative position and orientation between pairs of interacting CG sites. An efficient method to solve for the basis vector coefficients was also developed, using sequential Householder triangularization,[101] to reduce the size of the least-squares problem and the computational cost for solving it.

The AFM-CG algorithm was validated in Chapter 5 by developing an anisotropic model of benzene in which each molecule is coarse-grained into a single disk-shaped site. Matching of the position, mass, shape and orientation of each CG site from a atomistic benzene molecule was done according to the mapping rules developed in Chapter 4. A set of basis vector coefficients was obtained using the AFM-CG method, which matched pairwise non-bonded CG force functions between pairs of benzene molecules to those extracted and calculated from a MD simulation using the FG model. In order to use the AFM-CG results in MD simulations, the force functions were fitted to analytical forms that are derived from the S-function expansion potential. This potential has been implemented for MD simulations previously by Bowen et al.[67] to study the liquid structure of organic materials that exhibit molecular anisotropy. Results from MD simulations using the AFM-CG model of benzene were compared to corresponding data obtained for simulations of various models. These include the FG model, a previously reported single-site spherical CG model,[110] and an anisotropic model developed by Bowen et al.[67] for liquid benzene under the same simulation conditions. Overall, the AFM-CG model showed the best agreement with the FG model for the density and various structural properties, which are comparable to experimental results. While the Bowen model shows good agreement with the FG model for orientational correlations between pairs of molecules, the model overestimates the density in constant-pressure simulations, and has larger deviations in the pair distribution function compared to the AFM-CG model. This demonstrates improved accuracy of the AFM-CG method over other unsystematic CG methods previously used to simulate molecules with anisotropy. Transferability of the AFM-CG model was also shown with good matching to the FG model at a different temperature to the temperature at which parametrization was completed.

6.2 Future directions

This thesis has investigated several aspects of the role of molecular anisotropy in the structure and thermodynamics of condensed phase systems. However, further study is required to comprehensively understand the phenomena and methods that have been examined in this work. Several directions for future work are outlined below.

6.2.1 Structures at material interfaces in organic photovoltaics and effects on device efficiency

While the work presented in Chapter 3 supports the conclusion that parallel alignment and face-on orientation of BQR at the interface with $PC_{71}BM$ is preferred, more simulations of a larger system over longer time scales are required in order to obtain structures that are representative of those formed under realistic device fabrication conditions. This is only feasible using simulations of CG models of the system, in which molecular anisotropy must be described for a meaningful description of the interfacial orientation of the materials. Once CG sites are suitably defined, the AFM-CG method can be effectively applied to construct the pairwise non-bonded potentials between these sites. Bonded interactions between these sites can be readily obtained using other CG method such as the iterative Boltzmann inversion (IBI) method. Previous use of the IBI method for bonded interactions is available for defining parametrization procedure.[65] With the use of CG models with anisotropy for BQR and $PC_{71}BM$, MD simulations of a system that is several times larger than the FG system can be completed for microsecond timescales. The FG models and simulation results obtained in Chapter 3 can be used as a reference to validate the accuracy of the CG models. Moreover, the CG models can also be used in simulations of other systems with similar chemical structures for comparison of the interface structure and the resulting effects on charge dynamics for OPV devices. In particular, a system consisting of BMR and $PC_{71}BM$ could potentially exhibit the most variations to the structure of the BQR:$PC_{71}BM$ system, due to it having the lowest efficiency in the corresponding OPV devices recorded by Geraghty et al.[47] Study of this system may provide further knowledge about how changes in molecular shape affect interface structures between materials and lead to molecular design rules for how to control OPV performance.

Once detailed descriptions of the material interfaces between these systems are obtained using CG simulations, calculations of the electronic structure and charge dynamic modeling at the interface could be performed. In order to use the information obtained from the molecular dynamic simulations to study the consequences of the interface structure on charge separation and recombination, development of electronic structure and transport calculations using a tight-binding model that include the effect of meso-scale electrostatic forces is required. This is because charge carrier and charge-transfer states can significantly change under the effects of nearby molecular dipoles and quadrupoles that can vary depending on the distance from them to the interface and the ordering at the interface[115–117] . In this site-based model, electronic states and couplings will be described at the molecular level and so will explicitly represent the structural order or disorder obtained in the molecular dynamic simulations. This together with a realistic structural description of the interface, will bring direct benefits toward studying energy and charge transfer phenomena happening at the donor/acceptor interfaces in organic photovoltaic devices. Similar approaches to develop models to calculate electronic properties specifically for organic semiconductor materials are available in the literature.[33,118] Comparison between results for electronic processes at the interfaces of these systems in relations to the interface structures obtained using MD simulations can provide insight about the underlying rules of controlling electronic processes at the donor–acceptor interface in OPVs.

6.2.2 Further development and optimization of the AFM-CG method

The AFM-CG method presented in Chapter 4 was successfully validated by parametrization of a single-site disk-shaped CG model of benzene. Thermodynamically consistent matching of MD simulation results to those of corresponding the FG model guaranteed the accuracy of the method for MD simulations under the same conditions. MD simulations at a higher temperature than the parametrization conditions also showed good agreement with a FG simulation. Further investigation for a wider range of temperature could be completed to assess the transferability of the CG model and the AFM-CG method. Overall, the CG model was shown to be in good agreement with the FG model, with only small deviations in the structural and thermodynamic properties observed. These deviations could be because the AFM-CG force functions were not used for MD simulations directly but via fitting to an analytical form of potential, which is not able to represent all of the features in the AFM-CG results. This can be solved if the MD simulation package could read and output force values directly using the basis vector expansion from the AFM-CG results, which in principle can be implemented. If this is available, a better assessment on the accuracy and transferability of the AFM-CG method could also be made.

Regarding technical aspects of the AFM-CG method, increasing the speed for constructing the force matrix is necessary for more efficient applications of CG molecular models. This can be optimized using several methods. Using different types of RBFs instead of the currently implemented compactly supported functions is one solution. Using compactly supported RBFs means a large number of these functions are needed to construct a complicated multivariate function such as pairwise non-bonded functions. Instead, a smaller number of global RBFs such as functions of Gaussian form can be used. It has been shown that these global functions can in general provide a better convergence rate and better accuracy in approximation of multivariate functions.[102] This means less FG reference simulation data is required, reducing the computational cost of FG simulations and CG parametrization. Another solution is the construction of RBFs on scattered points instead of uniformly distributed points, which could provide a more efficient approximation of the function of interest for where data is available. Once optimized, a CG parametrization package for the AFM-CG method will be published for general users, allowing parametrization of CG models with anisotropy to be more easily used by researchers in the field.

Appendices

Appendix A. All-atom force field for simulations of BQR and PC$_{71}$BM

Appendix A.1. All-Atom force field for BQR

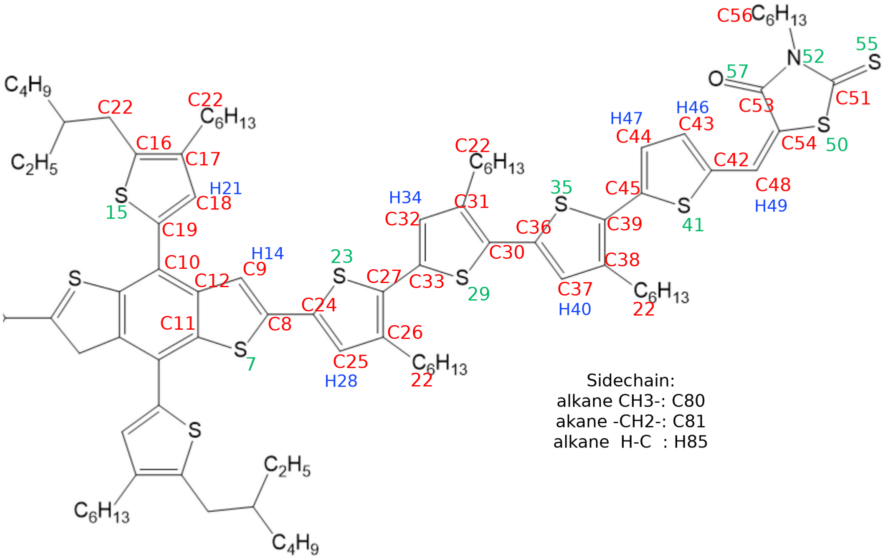


Figure A.1: Atoms types assigned for BQR. Atom types are shown in colors, not to be confused with chemical structures in black. Carbon atoms bonded to aromatic rings are assigned to different atom types from other from the sidechains. Alkane C and H atom types are adapted from the OPLS-AA (2005) force field[70]

Table A.1: Atomic charges calculated for atom types of BQR using "Charges from the Electrostatic Potential on a Grid" (ChelpG) method for electrostatic potential fitting with B3LYP/6-31+G** level of theory. Atomic charges for C and H atoms on side chains are taken from the OPLS-AA (2005) force field[70]

Name	Charge (*e*)	Name	Charge (*e*)	Name	Charge (*e*)	Name	Charge (*e*)
S7	-0.0863	C22	-0.0600	S35	-0.1042	C48	-0.1883
C8	0.0948	S23	-0.0470	C36	0.2048	H49	0.1649
C9	-0.2686	C24	0.1261	C37	-0.4083	S50	-0.0537
C0	0.0328	C25	-0.4381	C38	0.1796	C51	0.0554
C11	0.0430	C26	0.2853	C39	-0.0707	N52	0.1114
C12	-0.0177	C27	-0.1663	H40	0.2292	C53	0.3823
H14	0.1676	H28	0.2080	S41	-0.0193	C54	-0.0430
S15	-0.0622	S29	-0.0973	C42	0.1783	S55	-0.2597
C16	-0.0180	C30	-0.1160	C43	-0.2978	C56	-0.2290
C17	0.1153	C31	0.1967	C44	-0.1411	O57	-0.4597
C18	-0.3241	C32	-0.3953	C45	0.0626	C80	-0.1800
C19	0.0815	C33	0.2033	H46	0.1855	C81	-0.1200
H21	0.1833	H34	0.2235	H47	0.1759	H85	0.0600

Table A.2: Bond parameters for pairs of atoms in BQR. Bond energy is calculated as $U_{\text{bond}} = K_b(b - b_0)^2$ for spring constant K_b and equilibrium bond length b_0. All bond energy parameters are adapted from the OPLS-AA (2005) force field,[70] while equilibrium bond length was measured from an optimized geometry of BQR B3LYP/6-31+G** level of theory

Atom 1	Atom 2	K_b (kcal/mol/Å^2)	b_0 (Å)
S7	C8	250	1.77
S7	C11	250	1.762
C8	C9	546	1.371
H14	C9	367	1.082
C9	C12	469	1.431
C11	C12	447	1.427
C10	C12	469	1.418
C10	C11	469	1.398
C10	C19	385	1.472
S15	C16	250	1.75
S15	C19	250	1.756
C16	C17	546	1.376
C17	C18	469	1.433
C18	C19	546	1.371
C22	C16	317	1.501
C22	C17	317	1.506
H21	C18	367	1.085
C8	C24	378	1.446
C24	C25	546	1.379
C25	C26	469	1.419
H28	C25	367	1.084
C22	C26	317	1.506
C26	C27	546	1.393
C24	S23	250	1.749
C27	S23	250	1.762
C27	C33	378	1.443
C33	C32	546	1.384
C32	C31	469	1.416
H34	C32	367	1.085
C22	C31	317	1.505
C31	C30	546	1.393
C33	S29	250	1.761
C30	S29	250	1.756
C30	C36	378	1.441
C36	C37	546	1.385
C37	C38	469	1.414
H40	C37	367	1.084
C22	C38	317	1.505
C38	C39	546	1.395
C36	S35	250	1.755
C39	S35	250	1.762
C39	C45	378	1.438
C45	C44	546	1.394
C44	C43	469	1.401
H47	C44	367	1.084
H46	C43	367	1.084
C43	C42	546	1.391
C45	S41	250	1.758
C42	S41	250	1.758
C42	C48	385	1.424
H85	C48	340	1.089
C48	C54	549	1.361
C54	S50	250	1.767
S50	C51	250	1.781
C51	S55	300	1.649
C51	N52	490	1.369
N52	C53	490	1.414
N52	C56	337	1.46
H85	C56	340	1.09
C81	C56	268	1.529
C53	O57	570	1.22
C53	C54	410	1.474

Table A.3: Angle parameters for BQR, calculated as $U_{\text{angle}} = K_\theta(\theta - \theta_0)^2$ for θ_0, the angle made by atoms 1–3, and spring constant K_θ. Atom types are simplified for some common types: CT as alkane C, HC as alkane H, and HA as aromatic H.

Atom 1	Atom 2	Atom 3	K_θ (kcal/mol/rad^2)	θ (°)
C8	S7	C11	74	91.3
S7	C8	C9	70	111.7
S7	C11	C12	70	110.8
S7	C8	C16	70	119.7
C8	C9	C12	85	114.1
C9	C12	C11	85	112
C9	C8	C16	70	120
HA	C9	C8	35	123.1
HA	C9	C12	35	122.7
HA	C8	C9	35	128.4
HA	C8	S7	35	119.9
S7	C11	C10	70	125.6
C9	C12	C10	85	127.7
C10	C12	C11	85	120.2
C10	C11	C12	85	123.4
C11	C10	C12	63	116.3
C11	C10	C19	63	122.8
C12	C10	C19	63	120.8
C16	S15	C19	74	92.3
S15	C16	C17	70	111.1
C16	C17	C18	70	112.2
C17	C18	C19	70	114.7
S15	C19	C18	70	109.7
CT	C16	S15	70	119.1
CT	C16	C17	70	129.8
HC	CT	C16	35	109.5
CT	C17	C18	70	121.9
CT	C17	C16	70	125.9
HC	CT	C17	35	109.5
CT	CT	C17	63	114
CT	CT	C16	63	114
C10	C19	S15	70	122.4
C10	C19	C18	70	127.9
HA	C18	C17	35	123.1
HA	C18	C19	35	122.2
C8	C16	C17	70	120
C8	C16	S15	70	120
C24	S23	C27	74	92.2
S23	C27	C26	70	110.5
C27	C26	C25	70	112.3
C26	C25	C24	70	114.9
C25	C24	S23	70	110
C8	C24	C25	70	127.6
C8	C24	S23	70	122.3
HA	C25	C24	35	122.4
HA	C25	C26	35	122.7
CT	C26	C25	70	121.9
CT	C26	C27	70	125.7
HC	CT	C26	35	109.5
CT	CT	C26	63	114
C29	C30	C36	70	118.2
C31	C30	S36	70	131.3
CT	C31	C30	70	125.7
CT	C31	C32	70	122
HC	C13	C31	35	109.5
CT	C13	C31	63	114
HA	C32	C31	35	122.1
HA	C32	C33	35	122.5
C27	C33	C29	70	123.1
S27	C33	C32	70	127.6
C36	S35	C39	74	92.4
S35	C39	C38	70	110.5
C39	C38	C37	70	112.3
C38	C37	C36	70	115.5
C37	C36	S35	70	109.3
C30	C36	C37	70	127.4
C30	C36	S35	70	123.2
HA	C37	C36	35	122.4
HA	C37	C38	35	122.1
CT	C38	C37	70	122
CT	C38	C39	70	125.7
HC	CT	C38	35	109.5
CT	CT	C38	63	114
C38	C39	C45	70	131.1
S35	C39	C45	70	118.4
C45	S41	C42	74	92.3
S41	C42	C43	70	109.8
C42	C43	C44	70	114.1
C43	C44	C45	70	114
C44	C45	S41	70	109.8
S41	C42	C48	70	124.8
C43	C42	C48	70	125.3
HA	C43	C42	35	122
HA	C43	C44	35	123.9
HA	C44	C43	35	123.4
HA	C44	C45	35	122.6
C39	C45	S41	70	122.9
C39	C45	C44	70	127.4
C42	C48	HC	35.5	115.2
C42	C48	C54	85	131.1
HC	C48	C54	35	113.8
C54	S50	C51	74	92.2
S50	C51	N52	70	110
C51	N52	C53	70	117.5
N52	C53	C54	70	110
C53	C54	S50	70	110.3
S50	C51	S55	70	123
N52	C51	S55	80	127
C51	N52	C56	50	122.2
C53	N52	C56	50	120.3

Continued on next page

Table A.3 – *Continued from previous page*

Atom 1	Atom 2	Atom 3	K_θ (kcal/mol/rad^2)	θ (°)
C26	C27	C33	70	131.3
S23	C27	C33	70	118.5
C33	S29	C30	74	92.5
S29	C30	C31	70	110.5
C30	C31	C32	70	112.3
C31	C32	C33	70	115.5
C32	C33	S29	70	109.3
HC	C56	N52	35	109.5
CT	C56	N52	80	109.7
N52	C53	O57	80	123.2
C54	C53	O57	80	126.8
C48	C54	C53	85	120.2
C48	C54	S50	85	129.5

Dihedral parameters were adopted mostly from the OPLS-AA force field, except ones between thiophene rings. These dihedral were parametrized following the procedure from ref. 72 and are presented in Table A.4. Lennard-Jones parameters are adopted from the OPLS-AA force field of similar atom types.

Table A.4: Dihedral parameters for BQR, calculated as $U_{\text{dihedral}} = 1/2K_1(1+\cos\phi) + 1/2K_2(1-\cos(2\phi)) + 1/2K_3(1+\cos(3\phi)) + 1/2K_4(1-\cos(4\phi))$, for dihedral angle ϕ between atoms 1–4

Atom 1 1	Atom 2 2	Atom 3 3	Atom 4 4	K_1 (kcal/mol)	K_2 (kcal/mol)	K_3 (kcal/mol)	K_4 (kcal/mol)
C9	C8	C24	C25	1.016426	5.528661	0.087027	1.013261
C9	C8	C24	S23	0.000000	0.000000	0.000000	0.000000
S7	C8	C24	S23	0.000000	0.000000	0.000000	0.000000
S7	C8	C24	C25	0.000000	0.000000	0.000000	0.000000
26	C27	C33	S29	0.000000	0.000000	0.000000	0.000000
26	C27	C33	C32	2.786189	4.826969	0.173521	0.061187
S23	C27	C33	S29	0.000000	0.000000	0.000000	0.000000
S23	C27	C33	C32	0.000000	0.000000	0.000000	0.000000
S29	C30	C36	C37	0.000000	0.000000	0.000000	0.000000
S29	C30	C36	S35	0.000000	0.000000	0.000000	0.000000
31	C30	C36	S35	0.000000	0.000000	0.000000	0.000000
31	C30	C36	C37	2.281305	4.754088	0.213218	0.073487
38	C39	C45	S41	0.000000	0.000000	0.000000	0.000000
38	C39	C45	C44	2.198861	5.037005	0.782737	0.496626
S35	C39	C45	S41	0.000000	0.000000	0.000000	0.000000
S35	C39	C45	C44	0.000000	0.000000	0.000000	0.000000

Appendix A.2. All-Atom force field for PC$_{71}$BM

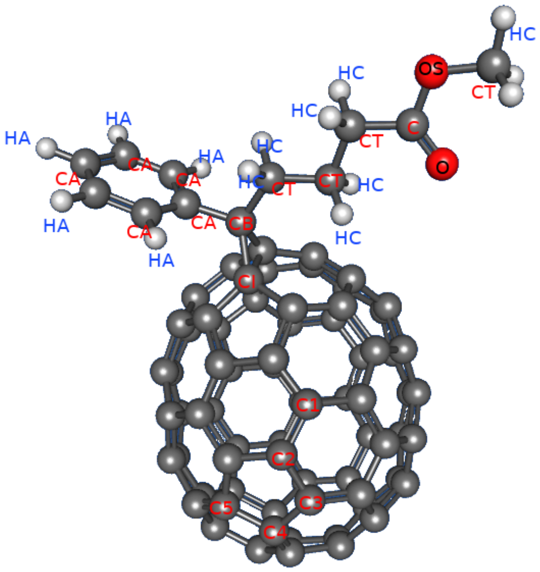


Figure A.2: Atoms types assigned for PCBM. Atom types of sidechain adopted from ref. [119]

Most force field parameters for PC$_{71}$BM were obtained from Cheung & Troisi.[119] However, it was found that using only one atom type for C$_{70}$ ball was not sufficient due to the anisotropic shape and sidechain of the molecule. In this work several C atom types (Figure A.2) have to be used to ensure the structure and interaction of the molecule is accurate.

Table A.5: Atomic charges for PCBM

Name	Charge	Definition
CA	0.23593	Aromatic C bonded to aliphatic C
CS	0.85678	Carbonyl (ester) C
CT	0.05946	Aliphatic C
CT	0.02139	Aliphatic C (methyl)
CT	0.56284	Aliphatic C bonded to Fullerene
C	0.56445	Carbony (ester) C=O
O	0.40815	Ester O
HA	0.11628	Aromatic H
HC	0.0482	Aliphatic H
C1	0.0426	Fullerene C1
CI	0.3774	Fullerene C bonded to aliphatic C
C2	0.0103	Fullerene C2
C3	0.0066	Fullerene C3
C4	0.0201	Fullerene C4
C5	0.0367	Fullerene C5

Table A.6: Bond parameters for PCBM. Bond energy is calculated as $U_{\text{bond}} = K_b(b - b_0)^2$ for spring constant K_b and equilibrium bond length b_0. Bond constants are adopted from the OPLS-AA force field and ref. [119]

Atom 1	Atom 2	K_b (kcal/mol/Å^2)	b_0 (Å)
CA	CA	469.0	1.401
CA	CT	317.0	1.510
CA	HA	367.0	1.075
CT	CT	268.0	1.530
CT	C	317.0	1.510
CT	OS	320.0	1.412
CT	HC	340.0	1.08
C	O	520.0	1.221
C	OS	214.0	1.334
C1	C1	469.0	1.453
C1	C2	469.0	1.410
C1	C3	469.0	1.410
C2	C2	469.0	1.444
C2	C3	469.0	1.380
C3	C3	469.0	1.444
C4	C2	469.0	1.388
C4	C3	469.0	1.388
C4	C1	469.0	1.496
C4	C4	469.0	1.444
C4	C5	469.0	1.388
C5	C2	469.0	1.380
C5	C3	469.0	1.380
C5	C5	469.0	1.428
C5	C1	469.0	1.410

Other parameters for angle bending and dihedral torsions are taken from the OPLS-AA force field for similar molecules without modification.

Appendix B. The molecular rigid body and classical approximation

Appendix B.1. The molecular rigid body and classical approximation

The development of the anisotropic CG force matching method makes the assumption that CG sites behave as "rigid" bodies. Using the rigid body approximation, we assume that the CG inter-site potentials only depend on the positions and orientations the component sites. This approximation ignores internal rotation and vibration of the atoms within the "rigid" CG sites. Thus for a CG system containing N sites, the potential energy operator $U(\mathbf{R}^N, \boldsymbol{\omega}^N)$ can be described using only the coordinates $\mathbf{R}^N = \mathbf{R}_1 ... \mathbf{R}_N$ and orientations $\boldsymbol{\omega}^N = \boldsymbol{\omega}_1 ... \boldsymbol{\omega}_N$ of the sites.[76] In the classical approximation, quantum effects are ignored and translational and rotational motions of the CG sites are treated classically.

Appendix B.2. The Dirac delta function

Definitions

The Dirac delta can be thought of as a function on the real line which is zero everywhere except at the origin, where it is infinite,

$$\delta(x) = \begin{cases} +\infty, & x = 0 \\ 0, & x \neq 0 \end{cases} \tag{B.1}$$

which results in the identity

$$\int_{-\infty}^{\infty} \delta(x) dx = 1. \tag{B.2}$$

Appendix C. Euler angles between two coordinate frames

Appendix C.1. Transformation matrix from space-fixed to body-fixed reference frame

In the CG algorithm, interactions between each CG site and all other neighbouring sites contribute to the RBFs that construct the force acting on that CG site.

It is easier to define these interactions in the body-fixed reference frame. Consider two CG sites I and J whose orientations are specified by two orthogonal rotation matrices $\hat{\mathbf{i}}$, and $\hat{\mathbf{j}}$:

$$\hat{\mathbf{i}} = \begin{pmatrix} \hat{\mathbf{i}}_1^T \\ \hat{\mathbf{i}}_2^T \\ \hat{\mathbf{i}}_3^T \end{pmatrix} = \begin{bmatrix} \hat{i}_{1x} & \hat{i}_{1y} & \hat{i}_{1z} \\ \hat{i}_{2x} & \hat{i}_{2y} & \hat{i}_{2z} \\ \hat{i}_{3x} & \hat{i}_{3y} & \hat{i}_{3z} \end{bmatrix}, \tag{C.1}$$

$$\hat{\mathbf{j}} = \begin{pmatrix} \hat{\mathbf{j}}_1^T \\ \hat{\mathbf{j}}_2^T \\ \hat{\mathbf{j}}_3^T \end{pmatrix} = \begin{bmatrix} \hat{j}_{1x} & \hat{j}_{1y} & \hat{j}_{1z} \\ \hat{j}_{2x} & \hat{j}_{2y} & \hat{j}_{2z} \\ \hat{j}_{3x} & \hat{j}_{3y} & \hat{j}_{3z} \end{bmatrix}, \tag{C.2}$$

where rows of $\hat{\mathbf{i}}$ and $\hat{\mathbf{j}}$ are the principal axes in the space-fixed frame of CG site I and J, respectively.

To transform the principal axes of CG site J from the space-fixed reference frame to the frame fixed to the principal axes of CG site I, one simply pre-multiplies the axis vectors of J by the rotation matrix of I:

$$\begin{pmatrix} \hat{\mathbf{j}}_{1I} & \hat{\mathbf{j}}_{2I} & \hat{\mathbf{j}}_{3I} \end{pmatrix} = \hat{\mathbf{i}} \cdot \begin{pmatrix} \hat{\mathbf{j}}_1 & \hat{\mathbf{j}}_2 & \hat{\mathbf{j}}_3 \end{pmatrix} \tag{C.3}$$

Appendix C.2. Calculating Euler angles from rotation matrix

Geometric description of Euler angles

One can also produce a set of three variables called Euler angles to describe the relative orientation between two coordinate frames. Considering the same CG sites I and J when site J is in reference frame fixed to site I as the result of the transformation in C.1, successive rotations as defined in Fig. C.3 transform the principle axes of I to those of J. The Euler angles (α,β,γ) are defined as the three angles of these rotations. It is assumed that the two systems share a common origin.

The process starts with rotation of the system ($\hat{\mathbf{i}}_1$,$\hat{\mathbf{i}}_2$,$\hat{\mathbf{i}}_3$) counterclockwise an angle α about $\hat{\mathbf{i}}_3$. This will transform $\hat{\mathbf{i}}_1$ and $\hat{\mathbf{i}}_2$ to $\hat{\mathbf{i}}_1^{\prime}$ and $\hat{\mathbf{i}}_2^{\prime}$, respectively, while keeping $\hat{\mathbf{i}}_3$ unchanged. The next step is a counterclockwise rotation of an angle β, this time around the newly rotated $\hat{\mathbf{i}}_2^{\prime}$. The process is completed with a counterclockwise rotation around $\hat{\mathbf{i}}_3^{\prime\prime}$ by an angle γ, rotating the system to the final state, which matches the principal axes of CG site J ($\hat{\mathbf{j}}_1$,$\hat{\mathbf{j}}_2$,$\hat{\mathbf{j}}_3$).

Calculating Euler angles

In the reference frame fixed to CG site I, the rotation matrix of CG site J has rows that are its three principal axes. The Euler angles relating orientation of I and J, as described above, can be calculated from the components of the rotation matrix of site J:[107]

$$\hat{\mathbf{j}}_I^T = \begin{bmatrix} \cos\alpha\cos\gamma - \sin\alpha\cos\beta\sin\gamma & -\cos\alpha\sin\gamma - \sin\alpha\cos\beta\cos\gamma & \sin\alpha\sin\beta \\ \sin\alpha\cos\gamma + \cos\alpha\cos\beta\sin\gamma & -\sin\alpha\sin\gamma + \cos\alpha\cos\beta\cos\gamma & -\cos\alpha\sin\beta \\ \sin\beta\sin\gamma & \sin\beta\cos\gamma & \cos\beta \end{bmatrix} \tag{C.4}$$

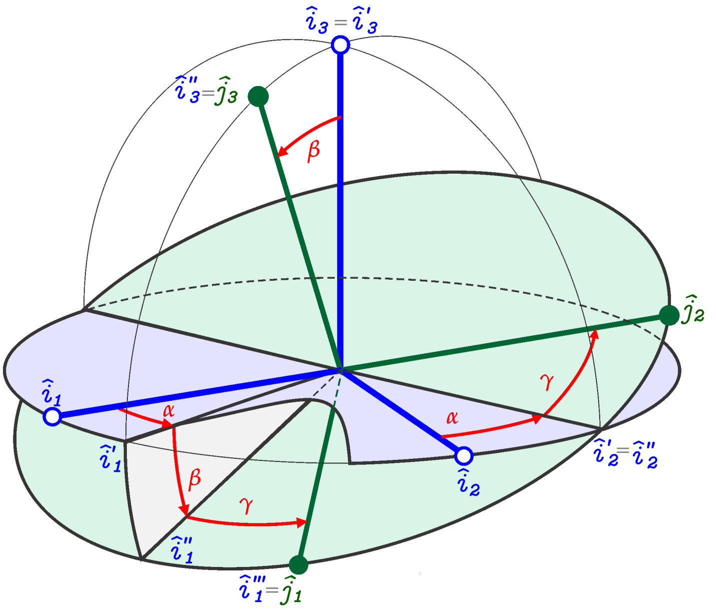


Figure C.3: Geometrical description of Euler angles. The reference frame fixed to I is shown in blue, while the rotated system corresponding to principle axes of J is shown in green. The Euler angles (α,β,γ) are shown in red. A prime is added to the axis labels every time a rotation is applied.

Appendix D. Radial basis functions

Appendix D.1. Radial basis functions on grid points

Fig. D.4 shows a grid of points in two dimensions. A radial distribution function built on point (2,3) evaluated at the sample point (x,y) will be dependent on the distance d between the two points.

An example of a set of uniformly distributed basis functions built on a 2D grid are shown in Fig. D.5a. If we call this set of basis functions $\mathcal{V}$, a scalar function represented as a surface can be built on it (Fig. D.5 (b)). The surface can be expressed mathematically from the basis functions as:

$$\mathcal{S}(x,y) = \sum \lambda_i \mathbf{v}_i(x,y) \quad \text{for} \quad \mathbf{v}_i(x,y) \in \mathcal{V} \tag{D.1}$$

The magnitude of $\mathcal{S}(x,y)$ is shown in vertical direction in Fig. D.5 (b). The vertical fluctuation of the surface depends on the value of the coefficients λ_i for each basis vectors $\mathbf{v}_i(x,y)$.

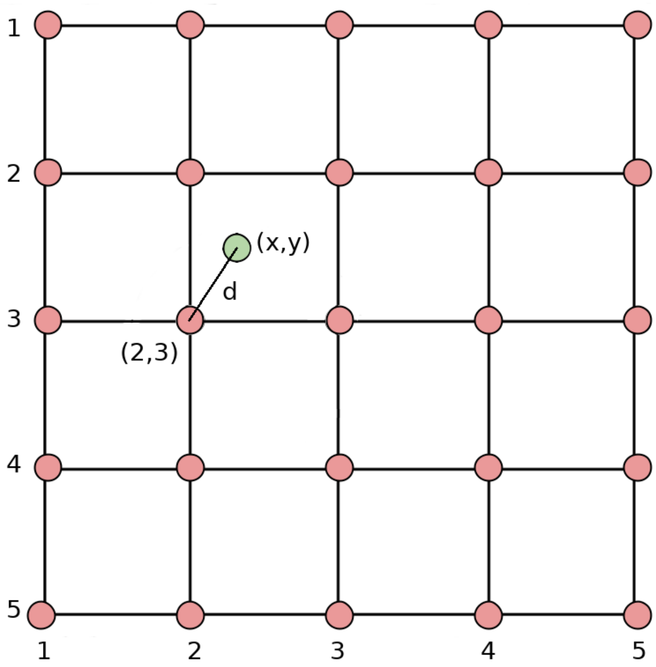


Figure D.4: A grid in two dimensions. Grid points are shown in pink, and a sample point is shown in green.

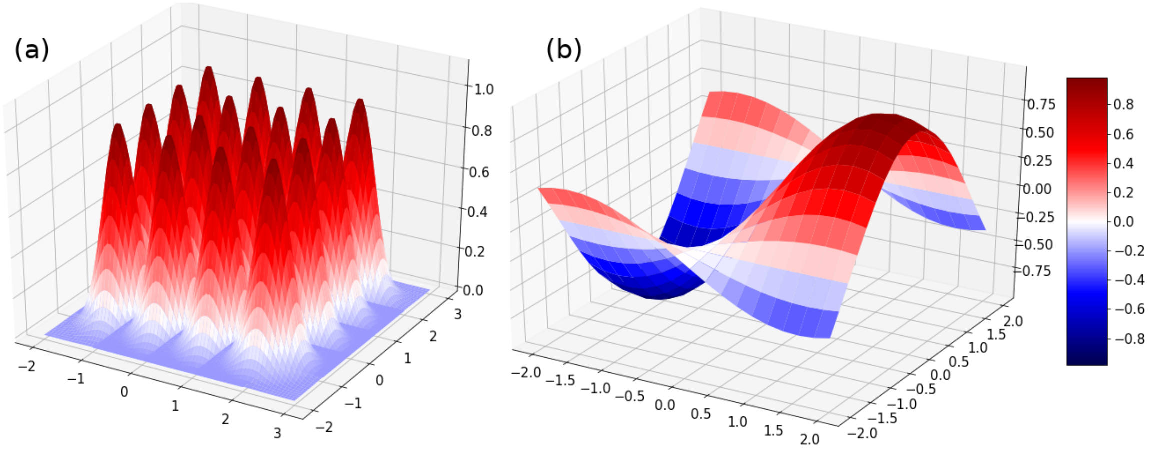


Figure D.5: (a) RBFs that are built on a 2D grid. (b) A surface that can be constructed from the 2D RBFs.

Appendix D.2. Choice of RBFs

One can use globally supported RBFs or compactly supported RBFs. The figure below shows some of them.

Globally supported RBFs converge more readily (i.e. give smaller errors) and have better stability (ie. changing number of grids, amount of input data will not affect the results significantly) but matrix built using these RBFs are completely dense so saving and solving it is limited by computer memory.[103]

Using compactly supported can solve the problem of computer memory, but at the same time reduces the convergence and stability of the results. Typically one has to change the radius from a centre at which the RBF are non-zeros to cover at least 50 other centres, to ensure

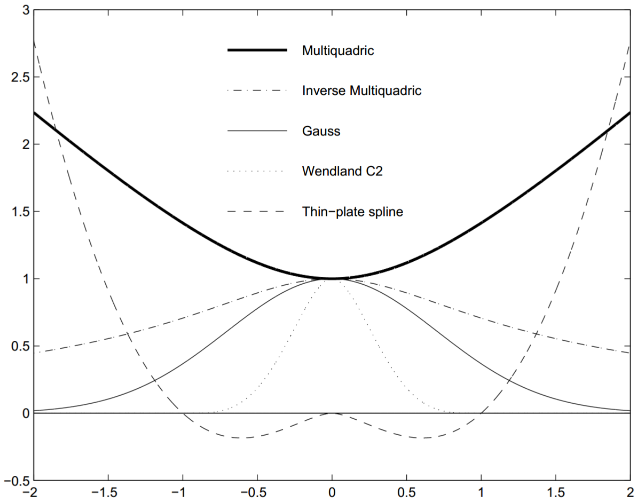


Figure D.6: Some commonly used RBFs. Note that here only Wendland RBFs which were implemented in the AFM-CG method is compactly supported. Image taken from ref. [120]

that the result is reliable.[120]

Appendix E. Fitting of AFM-CG benzene model to S-function potential

Appendix E.1. Fitting of AFM-CG benzene model to S-function potential

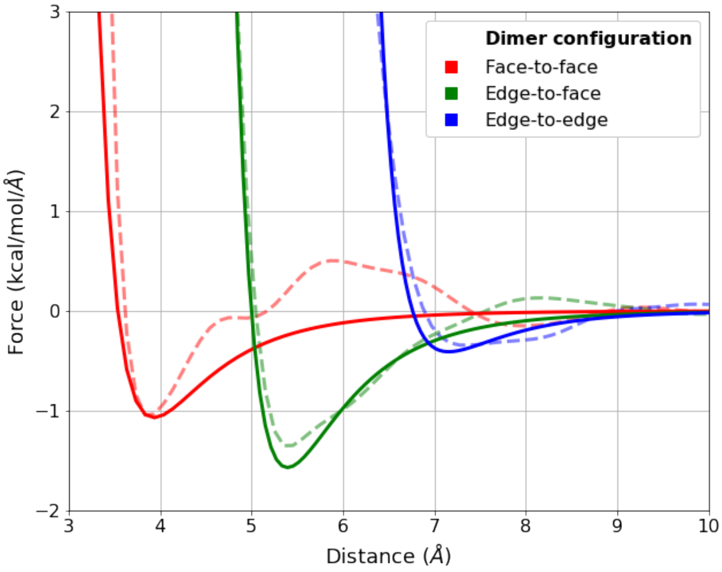


Figure E.7: Force curves versus separation distance at face-to-face, edge-to-face, and edge-to-edge configurations of CG benzene dimers

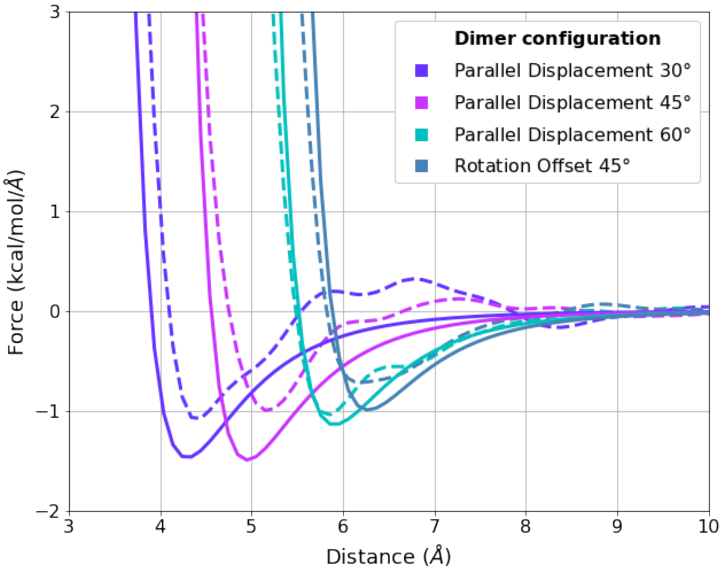


Figure E.8: Force curves versus separation distance at different configurations of CG benzene dimers

Appendix E.2. Coarse-grained model parameters for benzene

Table E.1: CG parameters of fitted AFM-CG results to S-function model, and Bowen CG model[67]

Parameter	AFM-CG model	Bowen model
σ_0	3.23	3.357
σ_{000}	1.63	1.588
σ_{cc2}	-0.795	-0.698
σ_{220}	-0.2	-0.175
σ_{222}	-0.1	-0.268
σ_{224}	0.7	0.841
ε_0	0.9	1.4
ε_{000}	1.5	1.344
ε_{cc2}	1.35	1.090
ε_{220}	0.1	0.293
ε_{222}	-0.5	0.063
ε_{224}	-4.3	-4.152

The set of parameters presented in Table E.1 are required to construct a S-function anisotropic potential. More about the potential can be found in ref. [67].

References

[1] Smil, V. *Energy transitions: global and national perspectives*; ABC-CLIO, 2016.

[2] What is AMOLED - SAMSUNG. `http://www.samsung.com/global/galaxy/what-is/amoled/`, Accessed: 2017-12-30.

[3] OLED Technology - LG company. `http://www.lg.com/us/experience-tvs/oled-tv/technology`, Accessed: 2017-12-30.

[4] Kolesov, V. A.; Fuentes-Hernandez, C.; Chou, W.-F.; Aizawa, N.; Larrain, F. A.; Wang, M.; Perrotta, A.; Choi, S.; Graham, S.; Bazan, G. C.; Nguyen, T.; Marder, S. R.; Kippelen, B. Solution-based electrical doping of semiconducting polymer films over a limited depth. *Nat. Mater.* **2017**, *16*, 474–480.

[5] Khlyabich, P. P.; Burkhart, B.; Thompson, B. C. Efficient ternary blend bulk heterojunction solar cells with tunable open-circuit voltage. *J. Am. Chem. Soc.* **2011**, *133*, 14534–14537.

[6] Arias, A. C.; MacKenzie, J. D.; McCulloch, I.; Rivnay, J.; Salleo, A. Materials and Applications for Large Area Electronics: Solution-Based Approaches. *Chem. Rev.* **2010**, *110*, 3–24.

[7] Sun, K.; Xiao, Z.; Lu, S.; Zajaczkowski, W.; Pisula, W.; Hanssen, E.; White, J. M.; Williamson, R. M.; Subbiah, J.; Ouyang, J.; Holmes, A. B.; Wong, W. W. H.; Jones, D. J. A molecular nematic liquid crystalline material for high-performance organic photovoltaics. *Nat. Commun.* **2015**, *6*, 6013.

[8] Lin, Y.; Li, Y.; Zhan, X. Small molecule semiconductors for high-efficiency organic photovoltaics. *Chem. Soc. Rev.* **2012**, *41*, 4245–4272.

[9] Spanggaard, H.; Krebs, F. C. A brief history of the development of organic and polymeric photovoltaics. *Sol. Energy Mater Sol. Cells* **2004**, *83*, 125–146.

[10] Green, M. A.; Hishikawa, Y.; Warta, W.; Dunlop, E. D.; Levi, D. H.; Hohl-Ebinger, J.; Ho-Baillie, A. W. Solar cell efficiency tables (version 50). *Prog. Photovoltaics* **2017**, *25*, 668–676.

[11] Tang, C. W. Two-layer organic photovoltaic cell. *Appl. Phys. Lett.* **1986**, *48*, 183–185.

[12] Huang, D. M.; Faller, R.; Do, K.; Moulé, A. J. Coarse-Grained Computer Simulations of Polymer/Fullerene Bulk Heterojunctions for Organic Photovoltaic Applications. *J. Chem. Theory Comput.* **2010**, *6*, 526–537.

[13] Knupfer, M. Exciton binding energies in organic semiconductors. *Appl. Phys. A* **2003**, *77*, 623–626.

[14] Coughlin, J. E.; Henson, Z. B.; Welch, G. C.; Bazan, G. C. Design and synthesis of molecular donors for solution-processed high-efficiency organic solar cells. *Acc. Chem. Res* **2013**, *47*, 257–270.

[15] Kan, B.; Zhang, Q.; Li, M.; Wan, X.; Ni, W.; Long, G.; Wang, Y.; Yang, X.; Feng, H.;

Chen, Y. Solution-processed organic solar cells based on dialkylthiol-substituted benzodithiophene unit with efficiency near 10%. *J. Am. Chem. Soc.* **2014**, *136*, 15529–15532.

[16] Zheng, Y.; Huang, J.; Wang, G.; Kong, J.; Huang, D.; Beromi, M. M.; Hazari, N.; Taylor, A.; Yu, J. A highly efficient polymer non-fullerene organic solar cell enhanced by introducing a small molecule as a crystallizing-agent. *Mater. Today* **2018**, *21*, 79–87.

[17] Zhang, G.; Yang, G.; Yan, H.; Kim, J.; Ade, H.; Wu, W.; Xu, X.; Duan, Y.; Peng, Q. Efficient nonfullerene polymer solar cells enabled by a novel wide bandgap small molecular acceptor. *Adv. Mater.* **2017**, *29*, 1606054.

[18] Sheraw, C. D.; Jackson, T. N.; Eaton, D. L.; Anthony, J. E. Functionalized pentacene active layer organic thin-film transistors. *Adv. Mater.* **2003**, *15*, 2009–2011.

[19] Min, J.; Jiao, X.; Ata, I.; Osvet, A.; Ameri, T.; Bäuerle, P.; Ade, H.; Brabec, C. J. Time-Dependent Morphology Evolution of Solution-Processed Small Molecule Solar Cells during Solvent Vapor Annealing. *Adv. Energy Mater.* **2016**, *6*, 1502579.

[20] Sun, Y.; Welch, G. C.; Leong, W. L.; Takacs, C. J.; Bazan, G. C.; Heeger, A. J. Solution-processed small-molecule solar cells with 6.7% efficiency. *Nat. Mater.* **2012**, *11*, 44–48.

[21] Zhou, H.; Yang, L.; Stuart, A. C.; Price, S. C.; Liu, S.; You, W. Development of fluorinated benzothiadiazole as a structural unit for a polymer solar cell of 7% efficiency. *Angew. Chem. Int. Ed.* **2011**, *123*, 3051–3054.

[22] Krebs, F. C.; Tromholt, T.; Jørgensen, M. Upscaling of polymer solar cell fabrication using full roll-to-roll processing. *Nanoscale* **2010**, *2*, 873–886.

[23] Krebs, F. C. Polymer solar cell modules prepared using roll-to-roll methods: knife-over-edge coating, slot-die coating and screen printing. *Sol. Energy Mater Sol. Cells* **2009**, *93*, 465–475.

[24] Hwang, Y.; Li, H.; Courtright, B. A. E.; Subramaniyan, S.; Jenekhe, S. A. Non-fullerene Polymer Solar Cells with 8.5% Efficiency Enabled by a New Highly Twisted Electron Acceptor Dimer. *Adv. Mater.* **2016**, *28*, 124–131.

[25] Coakley, K. M.; McGehee, M. D. Conjugated polymer photovoltaic cells. *Chem. Mater.* **2004**, *16*, 4533–4542.

[26] Stapleton, A.; Vaughan, B.; Xue, B.; Sesa, E.; Burke, K.; Zhou, X.; Bryant, G.; Werzer, O.; Nelson, A.; David Kilcoyne, A. L.; Thomsen, L.; Wanless, E.; Belcher, W.; Dastoor, P. A multilayered approach to polyfluorene water-based organic photovoltaics. *Sol. Energy Mater. Sol. Cells* **2012**, *102*, 114–124.

[27] Poelking, C.; Andrienko, D. Design Rules for Organic Donor–Acceptor Heterojunctions: Pathway for Charge Splitting and Detrapping. *J. Am. Chem. Soc.* **2015**, *137*, 6320–6326.

[28] Lin, Y. L.; Fusella, M. A.; Rand, B. P. The impact of local morphology on organic donor/acceptor charge transfer states. *Adv. Energy Mater.* **2018**, *8*, 1702816.

[29] Jailaubekov, A. E.; Willard, A. P.; Tritsch, J. R.; Chan, W.-L.; Sai, N.; Gearba, R.; Kaake, L. G.; Williams, K. J.; Leung, K.; Rossky, P. J.; Zhu, X. Hot charge-transfer excitons set the time limit for charge separation at donor/acceptor interfaces in organic

photovoltaics. *Nat. Mater.* **2013**, *12*, 66.

[30] van der Hofstad, T. G. J.; Di Nuzzo, D.; van den Berg, M.; Janssen, R. A. J.; Meskers, S. C. J. Influence of Photon Excess Energy on Charge Carrier Dynamics in a Polymer-Fullerene Solar Cell. *Adv. Energy Mater.* **2012**, *2*, 1095–1099.

[31] Vandewal, K. et al. Efficient charge generation by relaxed charge-transfer states at organic interfaces. *Nat. Mater.* **2014**, *13*, 63–68.

[32] Tamura, H.; Burghardt, I. Ultrafast Charge Separation in Organic Photovoltaics Enhanced by Charge Delocalization and Vibronically Hot Exciton Dissociation. *J. Am. Chem. Soc.* **2013**, *135*, 16364–16367.

[33] D'Avino, G.; Muccioli, L.; Olivier, Y.; Beljonne, D. Charge Separation and Recombination at Polymer–Fullerene Heterojunctions: Delocalization and Hybridization Effects. *J. Phys. Chem. Lett.* **2016**, *7*, 536–540.

[34] Verlaak, S.; Beljonne, D.; Cheyns, D.; Rolin, C.; Linares, M.; Castet, F.; Cornil, J.; Heremans, P. Electronic Structure and Geminate Pair Energetics at Organic-Organic Interfaces: The Case of Pentacene/C60 Heterojunctions. *Adv. Funct. Mater.* **2009**, *19*, 3809–3814.

[35] de Gennes, P.; Prost, J. *The Physics of Liquid Crystals*; Oxford University Press, 1995.

[36] Verma, I.; Sidiq, S.; Pal, S. K. Detection of creatinine using surface-driven ordering transitions of liquid crystals. *Liq. Cryst.* **2016**, *43*, 1126–1134.

[37] Mills, S. J.; Care, C. M.; Neal, M. P.; Cleaver, D. J. Computer simulation of an unconfined liquid crystal film. *Phys. Rev. E* **1998**, *58*, 3284–3294.

[38] Wall, G. D.; Cleaver, D. J. Computer simulations of adsorbed liquid crystal films. *Mol. Phys.* **2003**, *101*, 1105–1112.

[39] Antypov, D.; Cleaver, D. J. The role of attractive interactions in rod–sphere mixtures. *J. Chem. Phys.* **2004**, *120*, 10307–10316.

[40] Eremtchenko, M.; Temirov, R.; Bauer, D.; Schaefer, J. A.; Tautz, F. S. Formation of molecular order on a disordered interface layer: Pentacene/Ag (111). *Phys. Rev. B* **2005**, *72*, 115430.

[41] Jakowetz, A. C.; Bohm, M. L.; Sadhanala, A.; Huettner, S.; Rao, A.; Friend, R. H. Visualizing excitations at buried heterojunctions in organic semiconductor blends. *Nat. Mater.* **2017**, *16*, 551–557.

[42] Tumbleston, J. R.; Collins, B. A.; Yang, L.; Stuart, A. C.; Gann, E.; Ma, W.; You, W.; Ade, H. The influence of molecular orientation on organic bulk heterojunction solar cells. *Nat. Photon.* **2014**, *8*, 385–391.

[43] Verlaak, S.; Beljonne, D.; Cheyns, D.; Rolin, C.; Linares, M.; Castet, F.; Cornil, J.; Heremans, P. Electronic Structure and Geminate Pair Energetics at Organic–Organic Interfaces: The Case of Pentacene/C60 Heterojunctions. *Adv. Funct. Mater.* **2009**, *19*, 3809–3814.

[44] Ran, N. A.; Roland, S.; Love, J. A.; Savikhin, V.; Takacs, C. J.; Fu, Y.; Li, H.; Coropceanu, V.; Liu, X.; Brédas, J.; Bazan, G. C.; Toney, M. F.; Neher, D.; Nguyen, T. Impact of interfacial molecular orientation on radiative recombination and charge generation efficiency. *Nat. Commun.* **2017**, *8*, 79.

[45] Kim, Y.; Cook, S.; Tuladhar, S. M.; Choulis, S. A.; Nelson, J.; Durrant, J. R.; Bradley, D. D. C.; Giles, M.; McCulloch, I.; Ha, C.; Ree, M. *Materials For Sustainable Energy: A Collection of Peer-Reviewed Research and Review Articles from Nature Publishing Group*; World Scientific, 2011; pp 63–69.

[46] Hood, S. N.; Kassal, I. Entropy and Disorder Enable Charge Separation in Organic Solar Cells. *J. Phys. Chem. Lett.* **2016**, *7*, 4495–4500.

[47] Geraghty, P. B.; Lee, C.; Subbiah, J.; Wong, W. W. H.; Banal, J. L.; Jameel, M. A.; Smith, T. A.; Jones, D. J. High performance p-type molecular electron donors for OPV applications via alkylthiophene catenation chromophore extension. *Beilstein J. Org. Chem.* **2016**, *12*, 2298–2314.

[48] Lenes, M.; Shelton, S. W.; Sieval, A. B.; Kronholm, D. F.; Hummelen, J. C. K.; Blom, P. W. M. Electron Trapping in Higher Adduct Fullerene-Based Solar Cells. *Adv. Funct. Mater.* **2009**, *19*, 3002–3007.

[49] Eisenmenger, N. D.; Delaney, K. T.; Ganesan, V.; Fredrickson, G. H.; Chabinyc, M. L. Energy Transfer Directly to Bilayer Interfaces to Improve Exciton Collection in Organic Photovoltaics. *J. Phys. Chem. C* **2015**, *119*, 19011–19021.

[50] Scherer, C.; Andrienko, D. Comparison of systematic coarse-graining strategies for soluble conjugated polymers. *Eur. Phys. J. Spec. Top.* **2016**, *225*, 1441–1461.

[51] E., R. S.; Savagatrup, S.; Pais, C. J.; Arya, G.; Lipomi, D. J. Predicting the Mechanical Properties of Organic Semiconductors Using Coarse-Grained Molecular Dynamics Simulations. *Macromolecules* **2016**, *49*, 2886–2894.

[52] Förster, S.; Kohl, E.; Ivanov, M.; Gross, J.; Widdra, W.; Janke, W. Polymer adsorption on reconstructed Au(001): A statistical description of P3HT by scanning tunneling microscopy and coarse-grained Monte Carlo simulations. *J. Chem. Phys.* **2014**, *141*, 164701.

[53] Müller-Plathe, F. Coarse-graining in polymer simulation: from the atomistic to the mesoscopic scale and back. *ChemPhysChem* **2002**, *3*, 754–769.

[54] Reith, D.; Pütz, M.; Müller-Plathe, F. Deriving effective mesoscale potentials from atomistic simulations. *J. Comput. Chem.* **2003**, *24*, 1624–1636.

[55] Karimi-Varzaneh, H. A.; Müller-Plathe, F. *Multiscale molecular methods in applied chemistry*; Springer, 2011; pp 295–321.

[56] Noid, W. G. Perspective: Coarse-grained models for biomolecular systems. *J. Chem. Phys.* **2013**, *139*, 090901.

[57] Fu, C.-C.; Kulkarni, P. M.; Scott Shell, M.; Gary Leal, L. A test of systematic coarse-graining of molecular dynamics simulations: Thermodynamic properties. *J. Chem. Phys.* **2012**, *137*, 164106.

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

[59] Lu, L.; Izvekov, S.; Das, A.; Andersen, H. C.; Voth, G. A. Efficient, regularized, and scalable algorithms for multiscale coarse-graining. *J. Chem. Theory Comput.* **2010**, *6*, 954–965.

[60] Rühle, V.; Junghans, C. Hybrid Approaches to Coarse-Graining using the VOTCA Package: Liquid Hexane. *Macromol. Theory Simul.* **2011**, *20*, 472–477.

[61] Lu, L.; Voth, G. A. Systematic coarse-graining of a multicomponent lipid bilayer. *J. Phys. Chem. B* **2009**, *113*, 1501–1510.

[62] Izvekov, S.; Voth, G. A. Solvent-free lipid bilayer model using multiscale coarse-graining. *J. Phys. Chem. B* **2009**, *113*, 4443–4455.

[63] Chaimovich, A.; Shell, M. S. Coarse-graining errors and numerical optimization using a relative entropy framework. *J. Chem. Phys.* **2011**, *134*, 094112.

[64] Potter, T. D.; Tasche, J.; Wilson, M. R. Assessing the transferability of common top-down and bottom-up coarse-grained molecular models for molecular mixtures. *Phys. Chem. Chem. Phys.* **2019**, *21*, 1912–1927.

[65] Lee, C.; Pao, C. Multiscale Molecular Simulation of Solution Processing of SMDPPEH: PCBM Small-Molecule Organic Solar Cells. *ACS Appl. Mater. Interfaces* **2016**, *8*, 20691–20700.

[66] Srivastava, A.; Ghosh, S. Reduced-order molecular-dynamics model for polystyrene by equivalent-structure coarse graining. *Phys. Rev. E* **2012**, *85*, 026702.

[67] Bowen, A. S.; Jackson, N. E.; Reid, D. R.; de Pablo, J. J. Structural Correlations and Percolation in Twisted Perylene Diimides Using a Simple Anisotropic Coarse-Grained Model. *J. Chem. Theory Comput.* **2018**, *14*, 6495–6504.

[68] Jensen, F. *Introduction to computational chemistry*; John wiley & sons, 2017.

[69] Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.

[70] Banks, J. L. et al. Integrated modeling program, applied chemical theory (IMPACT). *J. Comput. Chem.* **2005**, *26*, 1752–1780.

[71] Moreno, M.; Casalegno, M.; Raos, G.; Meille, S. V.; Po, R. Molecular modeling of crystalline alkylthiophene oligomers and polymers. *J. Phys. Chem. B* **2010**, *114*, 1591–1602.

[72] Jackson, N. E.; Kohlstedt, K. L.; Savoie, B. M.; Olvera de la Cruz, M.; Schatz, G. C.; Chen, L. X.; Ratner, M. A. Conformational order in aggregates of conjugated polymers. *J. Am. Chem. Soc.* **2015**, *137*, 6254–6262.

[73] Hockney, R. W.; Eastwood, J. W. *Computer simulation using particles*; crc Press, 1988.

[74] Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **1985**, *31*, 1695.

[75] Izvekov, S.; Voth, G. A. A multiscale coarse-graining method for biomolecular systems. *J. Phys. Chem. B* **2005**, *109*, 2469–2473.

[76] Gray, C. G.; Gubbins, K. E. *Theory of Molecular Fluids: Vol. 1: Fundamentals*; Clarendon Press, 1984.

[77] Noid, W. G.; Liu, P.; Wang, Y.; Chu, J.; Ayton, G. S.; Izvekov, S.; Andersen, H. C.; Voth, G. A. The multiscale coarse-graining method. II. Numerical implementation for

coarse-grained molecular models. *J. Chem. Phys.* **2008**, *128*, 244115.

[78] Tumbleston, J. R.; Collins, B. A.; Yang, L.; Stuart, A. C.; Gann, E.; Ma, W.; You, W.; Ade, H. The influence of molecular orientation on organic bulk heterojunction solar cells. *Nat. Photon.* **2014**, *8*, 385.

[79] Ndjawa, G. O. N.; Graham, K. R.; Mollinger, S.; Wu, D. M.; Hanifi, D.; Prasanna, R.; Rose, B. D.; Dey, S.; Yu, L.; Brédas, J.; McGehee, M. D.; Salleo, A.; Amassian, A. Open-Circuit Voltage in Organic Solar Cells: The Impacts of Donor Semicrystallinity and Coexistence of Multiple Interfacial Charge-Transfer Bands. *Adv. Energy Mater.* **2017**, *7*, 1601995.

[80] Lin, Y. L.; Fusella, M. A.; Kozlov, O. V.; Lin, X.; Kahn, A.; Pshenichnikov, M. S.; Rand, B. P. Morphological Tuning of the Energetics in Singlet Fission Organic Solar Cells. *Adv. Funct. Mater.* **2016**, *26*, 6489–6494.

[81] Marcon, V.; Raos, G. Free energies of molecular crystal surfaces by computer simulation: Application to tetrathiophene. *J. Am. Chem. Soc.* **2006**, *128*, 1408–1409.

[82] Frigerio, F.; Casalegno, M.; Carbonera, C.; Nicolini, T.; Meille, S. V.; Raos, G. Molecular dynamics simulations of the solvent-and thermal history-dependent structure of the PCBM fullerene derivative. *J. Mater. Chem.* **2012**, *22*, 5434–5443.

[83] Girifalco, L. A. Molecular properties of fullerene in the gas and solid phases. *J. Phys. Chem.* **1992**, *96*, 858–861.

[84] Huang, D. M. Computational study of P3HT/C60-fullerene miscibility. *Aust. J. Chem.* **2014**, *67*, 585–591.

[85] Plimpton, S. Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* **1995**, *117*, 1–19.

[86] Jorgensen, W. L.; Maxwell, D. S.; 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.* **1996**, *118*, 11225–11236.

[87] Williams, M.; Tummala, N. R.; Aziz, S. G.; Risko, C.; Brédas, J. Influence of molecular shape on solid-state packing in disordered PC61BM and PC71BM fullerenes. *J. Phys. Chem. Lett.* **2014**, *5*, 3427–3433.

[88] Vandewal, K.; Tvingstedt, K.; Gadisa, A.; Inganäs, O.; Manca, J. V. Relating the open-circuit voltage to interface molecular properties of donor: acceptor bulk heterojunction solar cells. *Phys. Rev. B* **2010**, *81*, 125204.

[89] Schwarz, K. N.; Kee, T. W.; Huang, D. M. Coarse-grained simulations of the solution-phase self-assembly of poly(3-hexylthiophene) nanostructures. *Nanoscale* **2013**, *5*, 2017–2027.

[90] Gay, J. G.; Berne, B. J. Modification of the overlap potential to mimic a linear site–site potential. *J. Chem. Phys.* **1981**, *74*, 3316–3319.

[91] Babadi, M.; Everaers, R.; Ejtehadi, M. R. Coarse-grained interaction potentials for anisotropic molecules. *J. Chem. Phys.* **2006**, *124*, 174708.

[92] Hahn, O.; Site, L. D.; Kremer, K. Simulation of polymer melts: From spherical to ellipsoidal beads. *Macromol. Theory Simul.* **2001**, *10*, 288–303.

[93] Tillack, A. F.; Johnson, L. E.; Eichinger, B. E.; Robinson, B. H. Systematic gener-

ation of anisotropic coarse-grained Lennard-Jones potentials and their application to ordered soft matter. *J. Chem. Theory Comput.* **2016**, *12*, 4362–4374.

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

[95] Ruhle, V.; Junghans, C.; Lukyanov, A.; Kremer, K.; Andrienko, D. Versatile object-oriented toolkit for coarse-graining applications. *J. Chem. Theory Comput.* **2009**, *5*, 3211–3223.

[96] Dunn, N. J. H.; Lebold, K. M.; DeLyser, M. R.; Rudzinski, J. F.; Noid, W. G. Bocs: Bottom-up open-source coarse-graining software. *J. Phys. Chem. B* **2017**, *122*, 3363–3377.

[97] Das, A.; Andersen, H. C. The multiscale coarse-graining method. V. Isothermal-isobaric ensemble. *J. Phys. Chem. Phys.* **2010**, *132*, 164106.

[98] Larini, L.; Lu, L.; Voth, G. A. The multiscale coarse-graining method. VI. Implementation of three-body coarse-grained potentials. *J. Phys. Chem. Phys.* **2010**, *132*, 164107.

[99] Izvekov, S.; Chung, P. W.; Rice, B. M. The multiscale coarse-graining method: Assessing its accuracy and introducing density dependent coarse-grain potentials. *J. Chem. Phys.* **2010**, *133*, 064109.

[100] Tschöp, W.; Kremer, K.; Batoulis, J.; Bürger, T.; Hahn, O. Simulation of polymer melts. I. Coarse-graining procedure for polycarbonates. *Acta Polym.* **1998**, *49*, 61–74.

[101] Lawson, C. L.; Hanson, R. J. Solving least squares problems. 1974.

[102] Buhmann, M. D. *Radial basis functions: theory and implementations*; Cambridge university press, 2003; Vol. 12.

[103] Wendland, H. Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree. *Adv. Comput. Math.* **1995**, *4*, 389–396.

[104] Ganguly, P.; Mukherji, D.; Junghans, C.; van der Vegt, N. F. Kirkwood–Buff coarse-grained force fields for aqueous solutions. *J. Chem. Theory Comput.* **2012**, *8*, 1802–1807.

[105] Lee, C. K.; Hua, C. C.; Chen, S. A. Parametrization of the Gay–Berne potential for conjugated oligomer with a high aspect ratio. *J. Chem. Phys.* **2010**, *133*, 064902.

[106] Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* **1984**, *52*, 255–268.

[107] Goldstein, H. *Classical mechanics*, 2nd ed.; Addison-Wesley series in physics; Addison-Wesley Pub. Co., 1980.

[108] Oliphant, T. E. *A guide to NumPy*; Trelgol Publishing USA, 2006; Vol. 1.

[109] Berardi, R.; Fava, C.; Zannoni, C. A generalized Gay-Berne intermolecular potential for biaxial particles. *Chem. Phys. Lett.* **1995**, *236*, 462–468.

[110] Pranami, G.; Slipchenko, L.; Lamm, M. H.; Gordon, M. S. *Multi-Scale Quantum Models for Biocatalysis*; Springer: Dordrecht, 2009; pp 197–218.

[111] Beg, S. A.; Tukur, N. M.; Al-Harbi, D. K.; Hamad, E. Z. Saturated liquid densities

of benzene, cyclohexane, and hexane from 298.15 to 473.15 K. *J. Chem. Eng. Data* **1993**, *38*, 461–464.

[112] Li, S.; Ye, L.; Zhao, W.; Yan, H.; Yang, B.; Liu, D.; Li, W.; Ade, H.; Hou, J. A wide band gap polymer with a deep highest occupied molecular orbital level enables 14.2% efficiency in polymer solar cells. *J. Am. Chem. Soc.* **2018**, *140*, 7159–7167.

[113] Ide, M.; Saeki, A.; Koizumi, Y.; Koganezawa, T.; Seki, S. Molecular engineering of benzothienoisoindigo copolymers allowing highly preferential face-on orientations. *J. Mater. Chem. A* **2015**, *3*, 21578–21585.

[114] Yang, Y.; Zhang, Z.-G.; Bin, H.; Chen, S.; Gao, L.; Xue, L.; Yang, C.; Li, Y. Side-chain isomerization on an n-type organic semiconductor ITIC acceptor makes 11.77% high efficiency polymer solar cells. *J. Am. Chem. Soc.* **2016**, *138*, 15011–15018.

[115] D'Avino, G.; Muccioli, L.; Castet, F.; Poelking, C.; Andrienko, D.; Soos, Z. G.; Cornil, J.; Beljonne, D. Electrostatic phenomena in organic semiconductors: fundamentals and implications for photovoltaics. *J. Phys. Condens. Matter* **2016**, *28*, 433002.

[116] Poelking, C.; Andrienko, D. Long–range embedding of molecular ions and excitations in a polarizable molecular environment. *J. Chem. Theory Comput.* **2016**, *12*, 4516–4523.

[117] Schwarze, M.; Tress, W.; Beyer, B.; Gao, F.; Scholz, R.; Poelking, C.; Ortstein, K.; Günther, A. A.; Kasemann, D.; Andrienko, D., et al. Band structure engineering in organic semiconductors. *Science* **2016**, *352*, 1446–1449.

[118] Tapping, P. C.; Clafton, S. N.; Schwarz, K. N.; Kee, T. W.; Huang, D. M. Molecular-level details of morphology-dependent exciton migration in poly (3-hexylthiophene) nanostructures. *J. Phys. Chem. C* **2015**, *119*, 7047–7059.

[119] Cheung, D. L.; Troisi, A. Theoretical study of the organic photovoltaic electron acceptor PCBM: Morphology, electronic structure, and charge localization. *J. Phys. Chem. C* **2010**, *114*, 20479–20488.

[120] Chen, C.; Hon, Y.; Schaback, R. Scientific computing with radial basis functions. *Department of Mathematics, University of Southern Mississippi, Hattiesburg, MS* **2005**, *39406*.