

Christian Röthel, BSc

Thin Film Crystal Structure Solution by Combining Molecular Dynamics and X-ray Diffraction

DIPLOMARBEIT

zur Erlangung des akademischen Grades Diplom-Ingenieur der
Studienrichtung Technische Physik



TECHNISCHE UNIVERSITÄT GRAZ

Betreuer:

Ao.Univ.-Prof. Dipl.-Ing. Dr.techn. Roland Resel
Institut für Festkörperphysik
Technische Universität Graz

Graz, Mai 2014

I am among those who think that science has great beauty. A scientist in his laboratory is not only a technician: he is also a child placed before natural phenomena which impress him like a fairy tale.

— Marie Curie

Deutsche Fassung:
Beschluss der Curricula-Kommission für Bachelor-, Master- und Diplomstudien vom 10.11.2008
Genehmigung des Senates am 1.12.2008

EIDESSTÄTTLICHE ERKLÄRUNG

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Graz, am

.....
(Unterschrift)

Englische Fassung:

STATUTORY DECLARATION

I declare that I have authored this thesis independently, that I have not used other than the declared sources / resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

.....
date

.....
(signature)

Abstract

Organic thin films may exhibit crystallographic phases which differ from the bulk phase due to the presence of a surface during the crystallization process. Such thin film phases are crucial for the charge transport in organic semiconducting devices. Therefore a complete structure solution is necessary to understand the complex properties of organic electronic devices. The investigation of such films is challenging due to the small scattering volume and low scattered intensities. Hence standard single crystal structure solution methods are not applicable. Our approach is a combination of experimental and theoretical methods based on grazing incidence X-ray diffraction (GID) and molecular dynamics simulations. Compared to rigid body refinement, molecular dynamics is capable of handling flexible molecules as well, which is especially important for molecules with flexible functional groups. A simulation scheme involving a shrinking simulation box based on the experimental unit cell was developed. In order to avoid simulations getting trapped in local minima of the potential energy surface, multiple simulation runs with different initial molecular packings were carried out. The evaluation of the results was based on the comparison of X-ray structure factors and corresponding energies. The procedure was tested for molecules with increasing complexity beginning with pentacene and ternaphthalene. In both cases a herringbone motif was favored by energy and matched the experimental X-ray structure factors. Also the planarization of the twisted ternaphthalene molecule in a crystallographic unit cell was successfully reproduced. However the herringbone angles were approximately 5° too low compared to literature values. In order to improve the results, a plane-wave DFT optimization on the most promising structures was carried out and lead to excellent agreement with experimental data. DFT optimizations proved to be an excellent tool to improve the the molecular dynamics results since they do not rely on the molecular mechanics force field approximations. For dioctyl-terthiophene we observed a herringbone motif for the conjugated backbone. The orientation of the terminal chains matched the energy minimum of the dihedral potential as obtained by DFT calculations and fulfilled the close-packing principle. The largest investigated system DBDCS, a cyano distyrylbenzene derivative, contained four molecules per unit cell with alkoxy chains attached to the backbone. Despite the extreme flexibility of the side chains, we still managed to identify a parallel stacking motif which features most likely bent terminal chains.

Kurzfassung

Organische dünne Schichten können aufgrund des Einflusses der Substratoberfläche kristallografische Phasen aufweisen, welche sich von der Bulk-Struktur Phase unterscheiden. Solche Dünnschichtphasen spielen eine entscheidende Rolle für den Ladungstransport in der organischen Elektronik und machen eine vollständige Kristallstrukturklärung notwendig. Durch das geringe Streuvermögen gestaltet sich die Untersuchung solcher dünnen Filme sehr schwierig und herkömmliche Einkristallmethoden sind nicht anwendbar. In dieser Arbeit wird eine Methode vorgestellt, welche sich aus der experimentellen Messung mittels Röntgenbeugung unter streifendem Einfall und Molekulardynamik Simulationen zusammensetzt. Im Gegensatz zu einem Rigid Body Refinement Ansatz, welcher Moleküle als starre Gebilde annimmt, ist diese Einschränkung in der Molekulardynamik nicht gegeben. Um zu verhindern, dass Simulationen in lokalen Energieminima stecken bleiben, wurde zum einen eine vergrößerte Einheitszelle verwendet, welche während des Simulationsverlauf auf die experimentelle Größe schrumpft und des weiteren unterschiedliche Startkonfigurationen. Die Auswertung der Ergebnisse basiert auf den gemessenen Röntgenstrukturfaktoren und den jeweiligen Energien der molekularen Packungen. Die Methode wurde auf unterschiedliche Moleküle mit zunehmender Komplexität angewandt. Für Pentacene und Ternaphthalene wurde von der Energie und den Strukturfaktoren ein Herringbone Motiv eindeutig bevorzugt. Des Weiteren wurde bei Ternaphthalene die typische planarisierung von verdrehten Molekülen in einer Einheitszelle beobachtet. Jedoch wich in beiden Fällen der Herringbone Winkel um 5° von den Literaturwerten ab. Eine anschließende Optimierung mittels DFT war in der Lage diese Abweichungen zu beseitigen und ist somit geeignet die Ergebnisse von klassischen Molekulardynamik Simulationen zu verbessern. Für Dioctyl-Terthiophene konnte eine Herringbone Struktur für das Molekül-Backbone identifiziert werden. Die Orientierung der Seitenketten war in guter Übereinstimmung mit dem Energieminimum des Torsionspotential und entsprach einer dichten Kugelpackung. Das komplexeste System DBDCS ist ein cyano Distyrylbenzene Derivat mit Alkoxy Seitenketten und vier Molekülen pro Einheitszelle. Trotz der extrem flexiblen Seitenketten war es möglich ein paralleles Stappeln der Molekül-Backbones und eine bevorzugte Ausrichtung der Seitenketten normal zur Backbone Ebene zu identifizieren.

Acknowledgment

First of all, I would like to thank my supervisor Roland Resel who has supported me throughout my thesis with his patience and knowledge whilst allowing me the room to work in my own way. He made it possible for me to take part in a synchrotron experiment at the ESRF in Grenoble which has been an amazing experience. It was a pleasure to meet and work with Ingo Salzmänn and his group from the Humboldt University of Berlin. Special thanks go to Dominik Kriegner for his 24/7 support for his software package *xrayutilities*.

I am most grateful to Otello M. Roscioni from the University of Bologna who shared his knowledge and experience in MD simulations and provided lots of valuable ideas and advice. Furthermore, Egbert Zojer and Iris Hehn for introducing me to MD simulations and the dCluster at TU Graz and Johannes Gierschner of IMDEA Nanoscience in Madrid for providing us with the DBDCS samples.

I enjoyed sharing an office with Reinhold Hetzel, Michael Zawodzki, Stefan Pachmajer, Katrin Unger and Andrew Jones. We quickly discovered that science is not the only passion we share. It was truly a delight spending so much time with you in the *k-Room*.

The end of my master's thesis is also a good time to look back and to appreciate all the help of my colleagues and friends with whom I had the pleasure of studying physics. A big thank you to Manuel Zingl, Georg Urstöger, Paul Christian and Gernot Kraberger. I would never have been able to get this far without you.

Last but not least, I thank my parents for all the moral support and the amazing chances they have given me over the years.

Contents

1. Introduction	1
2. Fundamentals	3
2.1. Molecular dynamics simulations	3
2.1.1. Numerical integration	3
2.2. Force field methods	5
2.2.1. Force field energy	6
2.2.2. Stretching energy	6
2.2.3. Bending energy	7
2.2.4. Dihedral energy	7
2.2.5. Van der Waals energy	9
2.2.6. The electrostatic energy	9
2.2.7. Parametrization	11
2.3. Computational chemistry	12
2.3.1. Density functional theory	13
2.3.2. Møller-Plesset perturbation theory	16
2.3.3. Basis sets	16
2.4. Crystal lattice	17
2.5. X-ray scattering	20
2.5.1. Structure factors	24
2.5.2. Phase problem	25
2.5.3. Intensity correction	26
3. Methods	29
3.1. Grazing incidence X-ray diffraction	29
3.2. MD simulations with LAMMPS and CGenFF	30
3.2.1. Molecule file formats	31
3.2.2. CHARMM topology and parameter file	32
3.2.3. Protein data bank format	33
3.2.4. Protein structure file (psf)	34

3.2.5. charmm2lammmps	35
3.2.6. LAMMPS data file	36
3.2.7. LAMMPS input file	36
3.2.8. LAMMPS output	39
3.3. Parametrization by analogy	40
3.3.1. Generation of CHARMM topology files	41
3.4. Shrinking cell simulation	41
3.5. DFT optimization	44
3.6. Potential energy surface scans	45
3.6.1. Gaussian09	45
3.6.2. LAMMPS	47
3.6.3. Dihedral parameter fitting	48
3.7. StructFactViewer	49
3.8. Computational equipment	49
4. Results and Discussion	51
4.1. Pentacene	51
4.1.1. Parametrization	51
4.1.2. Energy	52
4.1.3. Structure factors	52
4.2. Ternaphthalene	58
4.2.1. Isolated molecule	58
4.2.2. Parametrization	58
4.2.3. Energy	59
4.2.4. Structure factors	60
4.3. Dioctyl-Terthiophene	63
4.3.1. Isolated molecule	63
4.3.2. Parametrization	69
4.3.3. Energy	70
4.3.4. Average structures	70
4.3.5. DFT optimization	71
4.4. DBDCS	75
4.4.1. Isolated molecule	75
4.4.2. Parametrization	76
4.4.3. Energy	76
4.4.4. Structure factors	81
5. Conclusion	85
A. Appendix	87
A.1. Perylene	87
A.1.1. Intensity correction	87

1. Introduction

Organic π -conjugated molecules are widely used in organic electronic devices such as OFETs [1] and OLEDs [2]. The charge transport in such devices depends on the overlap of π -orbitals and therefore on the orientation of adjacent molecules [3, 4]. Organic molecules such pentacene and polythiophenes are usually applied as thin films in electronic devices. It has been shown that organic molecules in thin films can exhibit a different crystallographic phases compared to the bulk structure [5, 6]. These thin film phases are induced by the substrate surface. Hence the thin film phase is crucial for the charge transport [7], a complete structure solution is necessary to understand the complex behavior of organic electronic devices.

X-ray diffraction techniques are perfect tools for crystal structure solution and have an impressive history. Probably the most famous results of X-ray diffraction has been the discovery of the double-helix structure of DNA [8]. First, due to ground breaking scientific contributions to the understanding of living organism but later more due to the dubious way of how contributions of other researchers were concealed [9].

The work flow for solving the crystal structure of a thin film phase is shown in Figure 1.1. The experimental data is collected by means of *grazing incidence X-ray diffraction (GID)*. Such an experiment can be used for the determination the crystallographic lattice constants. Furthermore, the intensity of diffraction peaks is used to calculate the X-ray structure factors. The difficult part in crystal structure solution is the determination of the positions of atoms or molecules inside the unit cell. This is also known as the *phase problem*. For single crystal techniques many sophisticated methods have been developed to reconstruct the missing phase information, for example the *Direct Methods* or *Patterson Methods* [10]. However they rely on the availability of several hundred or thousand individual structure factors which is not possible for organic thin films. In the case of organic thin film phases, the number of molecules in an unit cell can be estimated from the unit cell volume and molecular weight. This information can be used for a *rigid body refinement*. This method operates in direct space and is basically an optimization of the molecule positions. The aim of the optimization is to minimize the difference between the measured and calculated structure factors. A drawback of this method is

1. Introduction

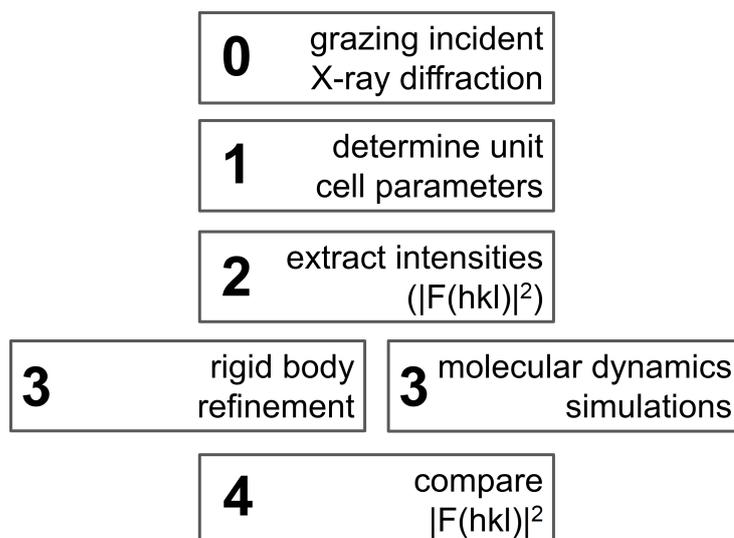


Figure 1.1.: Illustration of thin film crystal structure solution work flow. The rigid body refinement is replaced with molecular dynamics simulations to incorporate the flexibility of molecules.

that molecules are assumed to be rigid. Hence molecular conformational changes are not possible during the crystal structure solution procedure. Nevertheless, this approach has been used successfully for thin film crystal structure solution involving small rigid molecules such as pentacene and ternaphthalene [11, 12].

In order to incorporate flexibility of molecules the rigid body refinement was exchanged for a *molecular dynamics* (MD) approach. MD simulations are widely used in the study of macromolecules like proteins and DNA but also provide an opportunity for small organic molecule simulations [13]. In this work, MD simulations for single unit cells have been carried out and the results have been evaluated by comparing the experimental and simulated structure factors $|F(hkl)|$. Another criteria is the energy of the molecular packings, since an optimal packing should represent a low energy arrangement. Since MD simulations rely on several approximations, the best MD results have been optimized by means of *density functional theory* (DFT) in order to improve the packing.

2. Fundamentals

2.1. Molecular dynamics simulations

In molecular dynamics, Newton's laws of motion [14] and numerical integration is used to simulate the evolution of a quantum mechanical many-body system over time. The result is a trajectory that describes the positions and velocities of the particles.

$$\mathbf{F}(\mathbf{r}) = -\nabla V(\mathbf{r}) = m\mathbf{a} \quad (2.1)$$

Due to the large mass difference between nuclei and electrons, it is possible to treat them separately. This is known as the *Born-Oppenheimer* approximation. Once the problem for the electrons is solved, which usually involves the solution of the Schrödinger equation, a potential $V(\mathbf{r})$ can be defined for the calculations of the classical trajectories of the nuclei. The nuclei are assumed to behave like charged point particles which move due to the force \mathbf{F} that is equal to the negative of the potential energy gradient, see Eq. (2.1).

2.1.1. Numerical integration

Since no analytical solution for a many-body problem with $N > 2$ can be obtained, numerical integration is necessary to find a solution. For numerical integration a large number of different schemes are available. A very popular and easy to implement method is the *Verlet* algorithm [15]. It is based on the Taylor expansion of the position \mathbf{r}_i at time step i .

$$\mathbf{r}_{i+1} = \mathbf{r}_i + \frac{\partial \mathbf{r}_i}{\partial t} \Delta t + \frac{1}{2} \frac{\partial^2 \mathbf{r}_i}{\partial t^2} \Delta t^2 + \frac{1}{6} \frac{\partial^3 \mathbf{r}_i}{\partial t^3} \Delta t^3 + \dots \quad (2.2)$$

The first derivative of \mathbf{r}_i with respect to time is the velocity \mathbf{v}_i at time step i and the second derivative, the acceleration \mathbf{a}_i . Since classical mechanics (neglecting friction)

2. Fundamentals

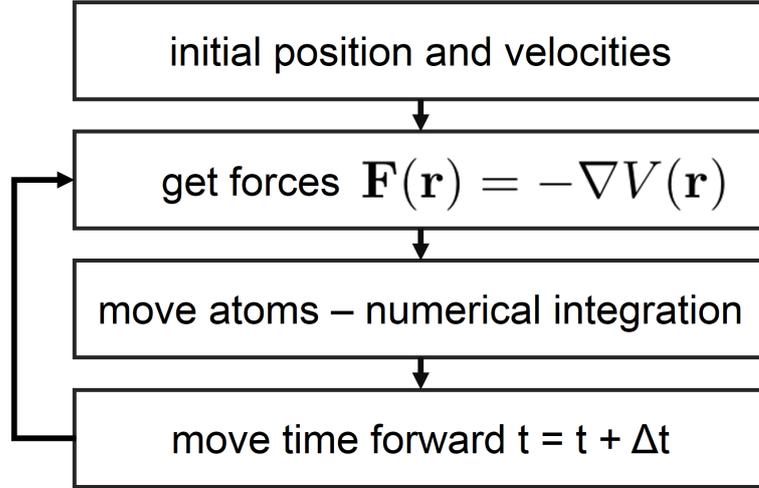


Figure 2.1.: Highly simplified description of a molecular dynamics simulation.

is time-reversal invariant, the position of a particle a small time step Δt earlier can be obtained by substituting Δt for $-\Delta t$.

$$\mathbf{r}_{i-1} = \mathbf{r}_i - \mathbf{v}_i \Delta t + \frac{1}{2} \mathbf{a}_i \Delta t^2 - \frac{1}{6} \frac{\partial^3 \mathbf{r}_i}{\partial t^3} \Delta t^3 + \dots \quad (2.3)$$

Combining Eq. (2.2) and (2.3) and solving for \mathbf{r}_{i+1} yields

$$\mathbf{r}_{i+1} = (2\mathbf{r}_i - \mathbf{r}_{i-1}) + \frac{1}{2} \mathbf{a}_i \Delta t^2 + \dots \quad (2.4)$$

$$\mathbf{a}_i = \frac{\mathbf{F}(\mathbf{r}_i)}{m} = -\frac{1}{m} \nabla V(\mathbf{r}_i). \quad (2.5)$$

Eq. (2.4) can be used for solving Newton's equation numerically with a sufficiently small time step Δt . Hence the term including the third derivative disappears, the approximation is correct to the third order in Δt . For the initial time step, the previous position is naturally not available, but can be estimated from Eqn. (2.2).

$$\mathbf{r}_{-1} = \mathbf{r}_0 - \mathbf{v}_0 \Delta t \quad (2.6)$$

The Verlet algorithm has the drawback that velocities do not appear explicitly. This poses a problem for controlling the temperature of an ensemble, which is usually achieved by

adjusting the velocities of the particles. An alternative version that contains velocities explicitly is the so called *leap-frog* algorithm [16, p. 452]:

$$\mathbf{r}_{i+1} = \mathbf{r}_i + \mathbf{v}_{i+\frac{1}{2}} \Delta t \quad (2.7)$$

$$\mathbf{v}_{i+\frac{1}{2}} = \mathbf{v}_{i-\frac{1}{2}} + \mathbf{a}_i \Delta t. \quad (2.8)$$

Eqs. (2.7) and (2.8) also provide a third-order integration scheme, as the Verlet algorithm, however the position and velocities updates are shifted by half a time step. The major disadvantage of the leap-frog algorithm is obviously that position and velocity are never known at the same time. The Verlet and leap-frog algorithm are usually preferred over the the famous Runge-Kutta integration scheme because the latter is not time-reversible. It can be shown that time-reversibility tends to improve the conservation of energy over long simulation times [17].

2.2. Force field methods

The potential energy function $V(\mathbf{r})$ contains all the information about a system and dictates the forces. Hence it determines ultimately the result of the simulation. Therefore it is crucial to find an appropriate description of the potential energy of an atomic system.

Molecules are approximated by a ball and spring model. Atoms have different masses and electrostatic charges. Bonds are modeled by springs with different stiffness and lengths. Molecules are mostly composed of smaller units that are similar in their structure. For example, benzene rings are part of many different organic molecules. However the benzene ring itself is not heavily influenced by its chemical environment and can usually be described by the same set of parameters, such as bond lengths and angles. Therefore it stands to reason to think of molecules as larger structures which are built up by smaller independent units. Such a set of small structural units, and its corresponding parameters, is called a force field. The smallest units are usually atoms, or more precisely atom types. Atom types account for the different behavior of for example carbon atoms in an aromatic ring or in an aliphatic chain. The force field approach is also known as *molecular mechanics* (MM).

2. Fundamentals

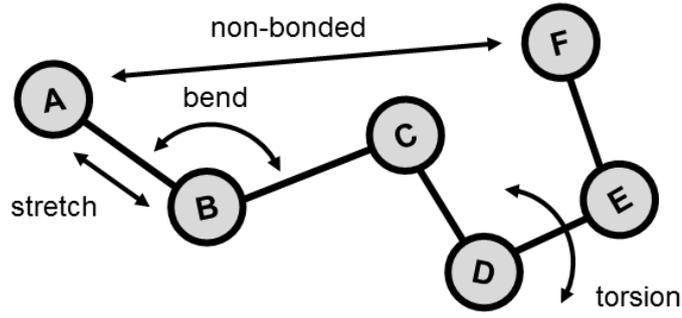


Figure 2.2.: Illustration of force field energy contributions.

2.2.1. Force field energy

The potential or force field energy E_{FF} is split up into two parts, the short range *intra*-molecular and long range *inter*-molecular forces:

$$E_{FF} = E_b + E_{nb} \quad (2.9)$$

$$E_b = E_{str} + E_{bend} + E_{dihed} \quad (2.10)$$

$$E_{nb} = E_{el} + E_{vdW} \quad (2.11)$$

The intra-molecular energy E_b (bonded) considers all the energies and forces that are connected to bond properties. These include stretching (E_{str}), bending (E_{bend}) and the torsion (E_{dihed}) of a bond. The latter is also referred to as the dihedral angle in the context of molecular mechanics.

The inter-molecular energy E_{nb} (non-bonded) takes all long range forces into account. Namely the electrostatic force E_{el} and the van der Waals force E_{vdW} . The bonded and non-bonded contributions to the force field are illustrated in Figure 2.2

2.2.2. Stretching energy

E_{str} is the energy required for stretching a bond between two atoms of type A and B . The simplest approximation for a ball and spring model is a harmonic oscillator which is obtained by a second-order Taylor expansion around the equilibrium bond length l_0 and l being the deflection.

$$E_{str}(l - l_0) = k_{AB}(l - l_0)^2 = k_{AB}\Delta l^2 \quad (2.12)$$

Here $\Delta l = l - l_0$ is the stretching of the bond and k_{AB} is the force constant associated with the A–B bond. Bond breaking or forming is not possible. An alternative to a harmonic potential is the more sophisticated Morse potential [18]. However, the most important region of the potential energy function is near the energy minimum and the agreement between the harmonic potential and quantum mechanical potential is sufficient for a large majority of systems.

2.2.3. Bending energy

E_{bend} is the energy function for bending a bond angle defined by three atom types A , B and C . There are bonds between A and B , and B and C . Also the bending potential is considered to be a harmonic potential, which is sufficient for most applications.

$$E_{bend}(\theta - \theta_0) = k_{ABC}(\theta - \theta_0)^2 = k_{ABC}\Delta\theta \quad (2.13)$$

2.2.4. Dihedral energy

The torsional energy or often called dihedral energy E_{dihed} considers the energy change which is linked to rotations around a bond. Lets consider the arrangement of atoms C , D , E , F in Figure 2.2. The torsion or dihedral angle χ is defined by the rotation around the D–E bond. The dihedral potential is naturally a periodic potential and therefore a Taylor expansion in χ is not applicable. However, a Fourier series can be used to account for the periodicity of E_{dihed} .

$$E_{dihed} = \sum_n k_n(1 + \cos(n\chi - \delta)) \quad (2.14)$$

Here, n allows to define rotations that are periodic by 360° for $n = 1$, $n = 2$ corresponds to a periodicity of 180° and so on. The constant k_n is used to set the rotation barriers around the D–E bond. Although a Fourier series allows one to represent an arbitrary periodic function, the chemical properties of a bond limits the choice of physically reasonable n .

Dihedrals are of special interest for molecular packing in solids since the cost in energy for rotating a bond is relatively low considering the large change in geometry. Hence molecules obey the close-packing principle, dihedral angles are strongly influenced by inter-molecular forces. In contrast, bond lengths and bending angles play only a minor role in the packing of molecules, since these intra-molecular properties are not heavily

2. Fundamentals

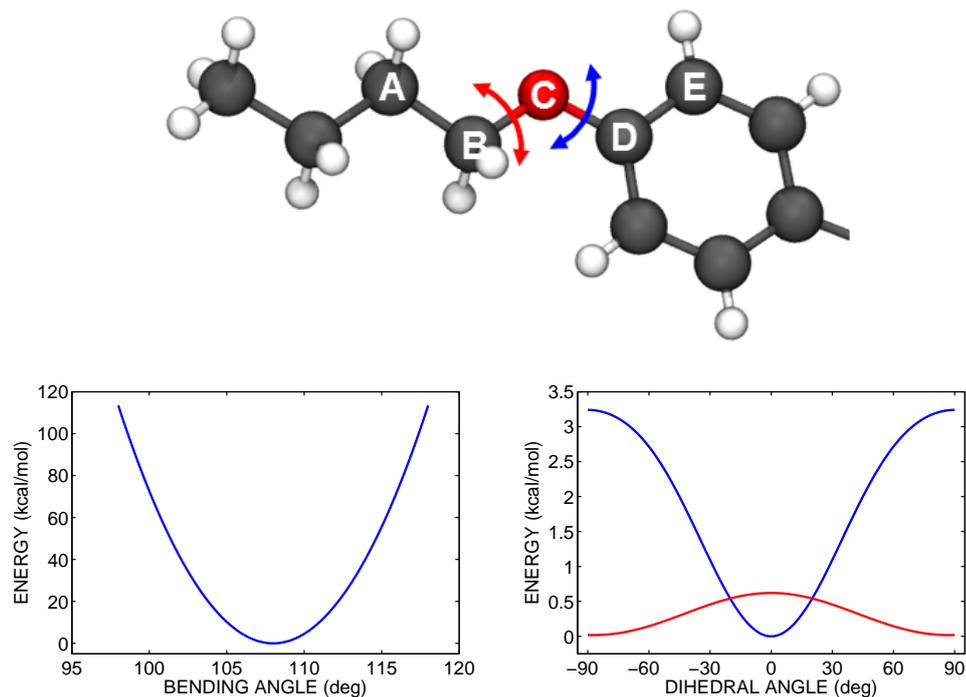


Figure 2.3.: Comparison of the bending and dihedral potentials of a butoxyphenyl group based on the *CGenFF* force field. Left: Bending potential (B – C – E) of the center oxygen atom. Right: Dihedral potentials A – B – C – D (red) and B – C – D – E (blue).

influenced by adjacent molecules. Furthermore, bond lengths and angles have only a small impact on the overall conformation of the molecule.

Figure 2.3 illustrates nicely the huge difference in energy costs for modifying bending angles and dihedrals. Bending the (B – C – E) angle by just 4° , which is a rather small change in the molecular conformation, causes a rise in energy of roughly 20 kcal/mol. In contrast, the rotation barrier for the B – C – D – E (blue) dihedral is just slightly over 3 kcal/mol. Considering that the average kinetic energy at 300 K is about 0.9 kcal/mol [16, p. 26], the importance of the dihedrals for the molecular conformation becomes clear. In the case of shallow potentials even low temperatures are sufficient to overcome the rotation barrier for certain parts of a molecule such as alkyl side chains. Hence those chains tend to move more or less unrestricted due to thermal energy. This behavior makes it rather tricky to find an optimal molecular packing.

2.2.5. Van der Waals energy

Van der Waals interaction (E_{vdW}) is a purely quantum mechanical effect and occurs due to instantaneous dipoles. Thus, it is also observed between non-polar molecules. The van der Waals energy is positive at small distances, features a negative minimum near the touching distance of the particles and approaches zero as the distance becomes larger. Probably the most famous model for the van der Waals interaction is the *Lennard-Jones* (LJ) potential [19]. A LJ potential, as in Eq. (2.15), is defined by two parameters. r_{AB}^{min} defines the distance at which the energy minimum occurs and ϵ_{AB} defines the depth of the potential well. Figure 2.4 illustrates the typical behavior of an LJ potential for a hydrogen-oxygen pair.

$$E_{vdW} = E_{LJ} = \epsilon_{AB} \left[\left(\frac{r_{AB}^{min}}{r_{AB}} \right)^{12} - 2 \left(\frac{r_{AB}^{min}}{r_{AB}} \right)^6 \right] \quad (2.15)$$

Since van der Waals interaction is calculated between all non-bonded atoms, the number of possible combinations can become very large and each combination requires its own LJ parameters. In order to circumvent this problem, a technique known as *parameter mixing* is applied. The LJ parameters for the interaction of two atom types A and B are calculated by Eq. (2.16) and (2.17).

$$\epsilon_{AB} = \sqrt{\epsilon_A \epsilon_B} \quad (2.16)$$

$$r_{AB}^{min} = \frac{r_A^{min} + r_B^{min}}{2} \quad (2.17)$$

2.2.6. The electrostatic energy

The second non-bonded interaction is due to electrostatic forces and modeled by the *Coulomb* potential as given in Eq.(2.18). The electron charge is not evenly distributed over the whole molecule. Thus, certain parts of a molecule are positively or negatively charged.

$$E_{el} = \frac{Q_A Q_B}{\epsilon r_{AB}} \quad (2.18)$$

Quantum mechanics uses orbitals to describe electron, thus the charge density is a continuous function. In order to approximate the charge distribution for a molecule, each

2. Fundamentals

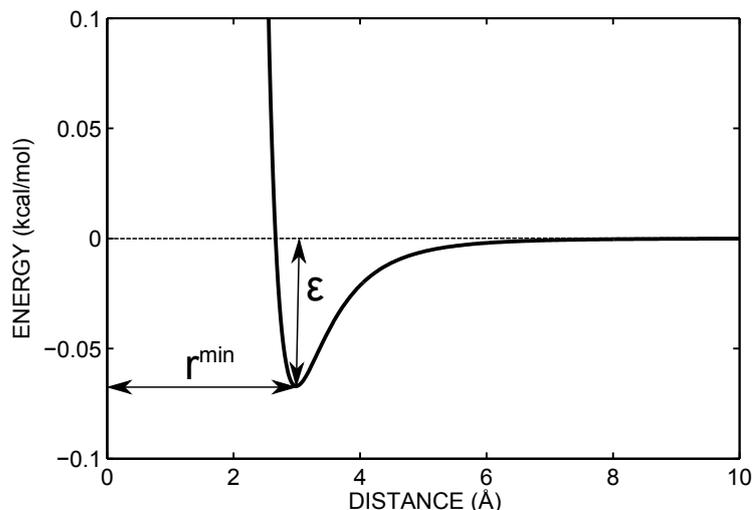


Figure 2.4.: Lennard-Jones potential for a hydrogen-oxygen pair.

atom is assigned a partial charge. The assignment of these charges can be done by using empirical rules or more commonly by fitting the electrostatic potential to theoretical data, for example the *molecular electrostatic potential* as obtained by computational chemistry methods.

This very basic approximation has naturally several drawbacks. The fitting of the atomic charges is usually performed on the equilibrium conformation of the molecule. Any change of the charge distribution due to different molecular conformations are neglected. As a consequence, the conformational energy for polar molecules is less accurate as for non-polar systems.

The computational cost of including long range intermolecular forces scales like $\mathcal{O}(N^2)$. A first step to get on top of this undesirable scaling is to introduce a cutoff distance, beyond which the long range interaction potential is set to zero. A typical cutoff radius would be 10 Å. To avoid unphysical large forces at the cutoff distance, due to the discontinuity of the energy, two cutoff radii are defined. The region between these two distances is used to reduce the energy function smoothly to zero. For this purpose a so called *switching function* such as $S(r)$ is multiplied onto the potential.

$$S(r) = \frac{(r_2^2 - r^2)^2(r_2^2 + 2r^2 - 3r_1^2)}{(r_2^2 - r_1^2)^3} \quad r_1 \leq r \leq r_2 \quad (2.19)$$

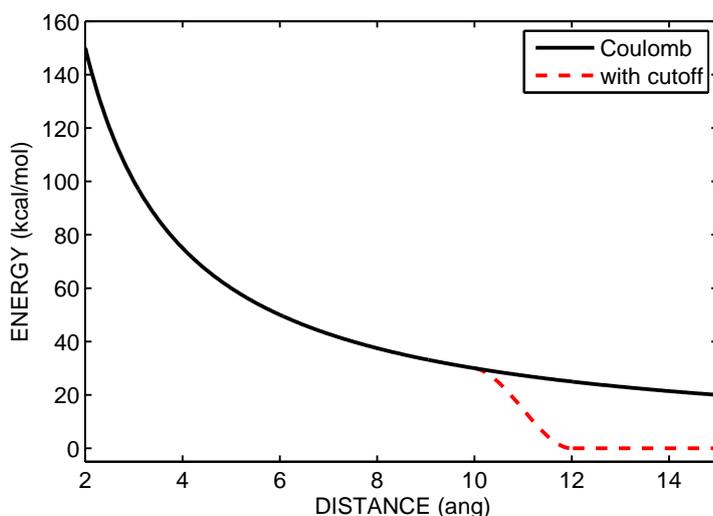


Figure 2.5.: Illustration of a Coulomb potential with a smooth cutoff between 10 Å and 12 Å.

Figure 2.5 demonstrates the effect of switching function $S(r)$ on the Coulomb potential. The cutoff region was set to be between 10 and 12 Å.

This works quite well for E_{vdW} , because the energy at a typical cutoff radius of 10 Å is usually below 0.01 kcal/mol due to r^{-6} decay. However, Coulomb interaction varies only as r^{-1} and has still a noticeable contribution. Therefore it is advisable to check the influence of the cutoff radius on the results. Nevertheless the computational complexity is still $\mathcal{O}(N^2)$ inside of the cutoff sphere. Many efficient methods for the summation of long range forces are based on *Ewald summation* [20]. Such summation schemes, rely on the fact that certain sums converge rapidly in reciprocal space. If modern numerical Fourier transform methods, like *Fast Fourier Transform* (FFT) in combination with *particle mesh* methods such as the *particle-particle-particle mesh* (PPPM) are applied, the complexity of the summation can be reduced to $\mathcal{O}(N \log N)$ [21].

2.2.7. Parametrization

An essential part of a molecular mechanics force field, besides the functional description, is the parameter set which is used to describe the interaction between different atom types. Each combination of atom types requires its own parameter set. Unfortunately the number of parameters becomes absolutely huge even for a simple force field which contains for example 50 atom types. Each of them requires two van der Waals

2. Fundamentals

parameters, so 100 parameters. Not all atom types are able to form bonds, even if just half of them form bonds, one needs to find 576 stretching parameters, for the 288 possible combinations. It gets even worse for bond angles (13800 parameters) and dihedrals (500000 parameters).

The intra-molecular parameters are mostly determined by quantum mechanical calculations and do not rely so much on experimental data as the non-bonded parameters. Anyhow, calculating many hundred thousands of dihedral parameters is still not feasible. So usually only parameters for very common chemical compounds are available. Hence it is quite common that one is in need for a replacement for missing parameters. A straightforward approach is to look for chemically similar compounds, for example, dihedral parameters for A – B – B – A are missing, but parameters for A – B – B – C exist. The results of such an approach, referred to as *Parametrization by Analogy*, depends of course on the similarity of atom types. If no fitting replacement is available in the force field one needs to fit the energy function of the corresponding parameter to some reference data in order to obtain new force field parameters. Such a procedure is described in more detail in Section 3.3.

One should keep in mind, that new parameters should always be in balance with the rest of the force field. Therefore mixing of parameters of different force fields is not advisable and will most likely produce useless results. Also re-optimization of parameters should always be carried out by following the guidelines of the particular force field.

2.3. Computational chemistry

For accurately describing atoms, a complete quantum mechanical approach is not avoidable. Such methods are called *ab initio*, due to the fact that they all focus on the solution of the time-independent Schrödinger equation as given in Eq. (2.20) and do not rely on empirical data.

$$\hat{H} |\psi\rangle = E |\psi\rangle \quad (2.20)$$

Again, the *Born-Oppenheimer* approximation is applied, which allows one to solve the Schrödinger equation with fixed nuclei coordinates. This leads to the following many-body Hamiltonian:

$$\begin{aligned} \hat{H} |\psi\rangle &= [\hat{T} + \hat{V} + \hat{U}] |\psi\rangle = \\ &= \left[\sum_i^N \left(-\frac{\hbar^2}{2m_i} \nabla_i^2 \right) + \sum_i^N V(\mathbf{r}_i) + \sum_{i=1}^N \sum_{j<i}^N U(\mathbf{r}_i, \mathbf{r}_j) \right] |\psi\rangle = E |\psi\rangle \end{aligned} \quad (2.21)$$

where E is the total energy. \hat{T} is the kinetic energy, \hat{V} is the potential due to the nuclei and \hat{U} covers the problematic electron-electron interaction, which makes it impossible to separate the many-body equation into simpler single-particle equations. Even an exact numerical solution of a coupled many-body problem is computationally not feasible and some further simplifications are required.

2.3.1. Density functional theory

The foundation for many modern computational chemistry methods for solving the many-body Schrödinger equation is the *Hartree-Fock* (HF) formalism. Since a detailed treatment of the HF method and density function theory (DFT) would go easily beyond the scope of this work, only the most important aspects will be discussed. A practical introduction to DFT can be found in [22].

The exact wave function for a many-body systems consists of a linear combination of Slater determinants. Since this wave function would be too huge for any real application, it is approximated by a single Slater determinant. Using this ansatz yields the Hartree-Fock equation. It becomes clear that the electron-electron interaction is approximated by a single-particle potential, which depends solely on the electron density $n(\mathbf{r})$. The single-particle potential can further be split up into two parts. The so called Hartree potential and the problematic exchange potential (Fock potential). The former can be expressed exactly, however the exchange potential can only be evaluated by using the rather intuitive Slater approximation.

A highly successful method, which is based on insights gained by HF is *density functional theory* (DFT). The idea behind DFT is to find a single-particle potential for the exchange potential \hat{U} , which depends solely on the $n(\mathbf{r})$ like the Hartree potential. The connection

2. Fundamentals

between the complicated antisymmetric wave function Ψ and electron density $n(\mathbf{r})$ is given by:

$$n(\mathbf{r}) = N \int d^3r_2 \int d^3r_3 \dots \int d^3r_N \Psi^*(\mathbf{r}, \mathbf{r}_2, \dots, \mathbf{r}_N) \Psi(\mathbf{r}, \mathbf{r}_2, \dots, \mathbf{r}_N) \quad (2.22)$$

The important part is, that the relation in Eq. (2.22) can be reversed. Hence it is possible to calculate the ground-state wavefunction Ψ_0 for a given ground-state density $n_0(\mathbf{r})$. This means, Ψ_0 is a *unique* functional of $n_0(\mathbf{r})$ [23]:

$$\Psi_0 = \Psi[n_0]. \quad (2.23)$$

Furthermore the ground-state electron density n_0 determines also the ground-state energy E_0

$$E_0 = E[n_0] = \langle \Psi[n_0] | \hat{T} + \hat{V} + \hat{U} | \Psi[n_0] \rangle. \quad (2.24)$$

The external potential \hat{V} is defined by the investigated system. In contrast, \hat{T} and \hat{U} are called universal potentials, because they do not depend on the investigated system. If the functional $T[n]$ and $U[n]$ are known, one has to minimize

$$E[n] = T[n] + U[n] + \int V(\mathbf{r})n(\mathbf{r})d^3r \quad (2.25)$$

with respect to $n(\mathbf{r})$ in order to find the ground-state density n_0 . First, the electron-electron interaction $U[n]$ is neglected:

$$E_s[n] = \langle \Psi_s[n] | \hat{T} + \hat{V}_s | \Psi_s[n] \rangle \quad (2.26)$$

where \hat{V}_s is an external effective potential in which the non-interacting electrons move. It was shown by *Hohenberg* and *Kohn* that the electron density that minimizes the energy of the overall functional is the true electron density corresponding to the solution of the Schrödinger equation [24]. However, the true exchange-correlation functional has not been discovered yet, therefore approximations, as discussed in in the next section, are applied. Nevertheless, the variational problem of minimizing the energy is solved by using Lagrangian multipliers. This leads to the famous *Kohn-Sham* equations:

$$\left[-\frac{\hbar^2}{2m} \nabla^2 + V_s(\mathbf{r}) \right] \phi_i(\mathbf{r}) = \varepsilon_i \phi_i(\mathbf{r}) \quad (2.27)$$

The difference to the original problem in Eq. (2.21) is that the *Kohn-Sham* equation

is missing the summations. This is due to the fact that the *Kohn-Sham* equations are single-electron equations. The solutions $\phi_i(\mathbf{r})$ allow one to calculate the sought electron density $n_s(\mathbf{r})$.

$$n_s(\mathbf{r}) = \sum_i^N |\phi_i(\mathbf{r})|^2. \quad (2.28)$$

A more detailed expression of the effective single-particle potential is given by:

$$V_s(\mathbf{r}) = V(\mathbf{r}) + \int \frac{e^2 n_s(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d^3 \mathbf{r}' + V_{XC}[n_s(\mathbf{r})] \quad (2.29)$$

The second term is the already known Hartree potential and the third term V_{XC} is the still unknown exchange-correlation potential, which includes all the many-body interactions. Due to the fact, that the Hartree and exchange-correlation potential depend on $n_s(\mathbf{r})$, the Kohn-Shams equations can only be solved iteratively. One starts with a reasonable guess for $n_s(\mathbf{r})$, subsequently calculates V_s and solves the Kohn-Sham equations. The obtained solutions $\phi_i(\mathbf{r})$ are used to calculate a new density $n_s(\mathbf{r})$ for the next iterative step. This steps are repeated until the density $n_s(\mathbf{r})$ is converged or below a certain tolerance level.

Exchange-correlation potential

A rather simple approximation for V_{XC} is the *Local Density Approximation* (LDA) [25]. It is assumed that exchange-correlation just depends on the position where the functional is evaluated:

$$E_{XC}^{LDA} = \int \varepsilon_{XC}(n) n(\mathbf{r}) d^3 r \quad (2.30)$$

Since exchange-correlation potentials are often obtained by Monte Carlo methods, an appropriate analytical representation is desired. Such fitting formulas where, for example created by *Perdew and Wang* (PW) [26]. An expansion of LDA is the *generalized gradient approximation* (GGA). It is still local but considers also the gradient of $n(\mathbf{r})$:

$$E_{XC}^{GGA} = \int \varepsilon_{XC}(n, \nabla n) n(\mathbf{r}) d^3 r \quad (2.31)$$

Two widely used GGA functionals are *PBE*, named after *Perdew, Burke and Ernzerhof* [27] and *PW91* [28].

2. Fundamentals

Another kind of functional are the so called *hybrid functionals*. They consist of a mix of HF exact exchange energy, based on the Kohn-Sham orbitals $\phi_i(\mathbf{r})$ rather than on the density $n(\mathbf{r})$, and various other exchange and correlation density functionals. Probably the most common hybrid functional is *B3LYP* [29, 30].

The correct or true exchange-correlation functional is still unknown, hence all solutions obtained by means of DFT are not exact solutions of the Schrödinger equation. Probably the most well-known inaccuracies of DFT are the underestimation of band gaps in semiconductors and the lack of proper van der Waals interaction.

2.3.2. Møller-Plesset perturbation theory

Møller-Plesset perturbation theory (MP) is a post-Hartree-Fock method which uses perturbation theory to add electron correlation effects. The Møller-Plesset theorem states that the correlation potential does not contribute in first-order to the exact electronic energy [31]. Therefore second order perturbation theory (MP2) is the lowest order which changes the unperturbed Hartree-Fock result. MP2 and even higher order calculations are only used for small systems due to the relatively large computational cost compared to DFT.

2.3.3. Basis sets

So far only the description of the particle interaction was discussed. However, for an actual calculation a linear combination of functions is used to approximate the unknown molecular orbitals. Such a set of functions is called a basis set. For solids with periodic boundary conditions, usually a set of plane waves in combination with pseudo potentials is used. Plane waves up to a certain cutoff energy are included in the calculations. For isolated molecules a linear combination of atomic orbitals (LCAO) is used to approximate the molecular orbitals. Since many different basis sets are available, the most popular sets will be discussed briefly.

The simplest basis sets are called *minimal basis sets*. Only a single function (called an atomic orbital) is used for each electron of the atoms. Thus, for hydrogen and helium a single $1s$ -orbital and for second row elements two s -functions ($1s$, $2s$) and three p -orbitals ($2p_x$, $2p_y$, $2p_z$) are sufficient. An expansion of the minimal basis set is the *Double Zeta* (DZ) or *Triple Zeta* (TZ) basis set. Double and triple refers to the doubling or tripling of the atomic orbitals for each electron. Hence ten atomic orbitals (two $1s$, two $2s$, six $2p_{x,y,z}$) would be necessary for second row elements like carbon for a DZ basis

set. Such a basis set is more flexible and allows for a better description of the electron distribution in different directions. This is especially important to accurately describe σ and π -bonds. The drawback is a larger basis set which increases the computational cost. Keeping in the mind that only valence electrons contribute to bonds, it stands to reason to split the electrons into core and valence electrons. Such basis sets are known as *split valence basis*.

Further improvements are to include polarization and diffuse functions. An s -orbital can be polarized by mixing with a p -orbital and d -orbitals are used to polarize p -orbitals. Diffuse functions use smaller exponents to account for the larger region of influence of anions and elements like sulfur, nitrogen, oxygen and fluorine in organic molecules.

Names of basis sets can be quite cryptic and confusing. The most popular notations for basis sets are based on the work of the group of Pople [32] and Dunning [33]. A short list of commonly used basis sets is given in Table 2.1.

Table 2.1.: List of basis sets with a short description. Top four are basis sets in Pople notation and the bottom three are correlation-consistent basis sets in Dunning notation.

basis set	description
6-31G	split valence double-zeta
6-31G*	* adds d-type polarization on non-hydrogen atoms
6-311G*	split valence triple-zeta with d-type polarization on non-hydrogen atoms
6-31+G*	+ adds one set of diffuse s- and p-functions for non-hydrogen atoms
cc-pVDZ	correlation-consistent, polarized valence double-zeta
cc-pVTZ	correlation-consistent, polarized valence triple-zeta
aug-cc-pVTZ	aug includes a set of diffuse s- and p-functions

2.4. Crystal lattice

Many properties of solids can be explained by their periodic structure. Therefore it is possible to reduce the large macroscopic bulk structure to a single microscopic repetitive unit cell. Generally, the unit cell is defined by six parameters, which are the lengths of the edges (a , b and c) and the angles between them (α , β and γ). The positions of the atoms \mathbf{r}_i (also known as basis, however not to be confused with the lattice basis vectors) are defined relative to the lattice point. The positions \mathbf{R} of the lattice points (or unit

2. Fundamentals

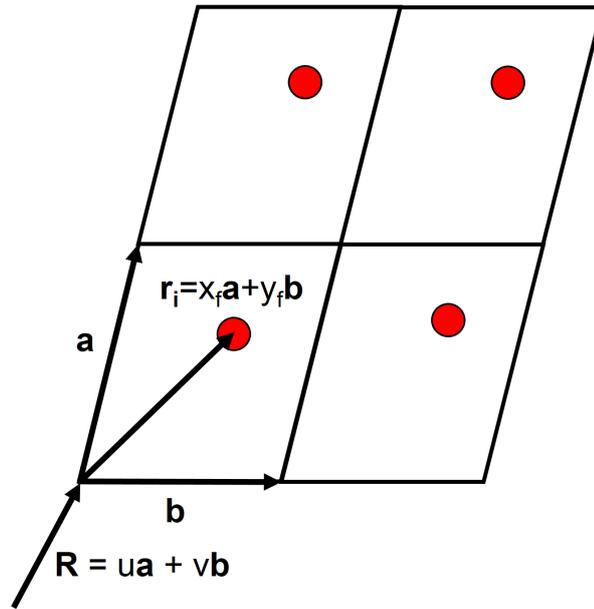


Figure 2.6.: Example of a 2-dimensional lattice with lattice vectors \mathbf{a} and \mathbf{b} . \mathbf{r}_i is the the position of atom i in fractional coordinates.

cells) create crystal lattice. It is obtained by translation of the lattice basis vectors \mathbf{a} , \mathbf{b} and \mathbf{c} .

$$\mathbf{R} = u\mathbf{a} + v\mathbf{b} + w\mathbf{c} \quad u, v, w \in \mathcal{Z} \quad (2.32)$$

The combination \mathbf{R} and \mathbf{r}_i allows one to reconstruct a crystalline solid. This circumstance is sometimes expressed rather casually as:

lattice + basis = crystal structure

For a orthorhombic lattice ($\alpha = \beta = \gamma = 90^\circ$) the basis vectors for the lattice are expressed by Eq. (2.33) and are equal to a Cartesian basis.

$$\mathbf{a} = (a, 0, 0) \quad \mathbf{b} = (0, b, 0) \quad \mathbf{c} = (0, 0, c) \quad (2.33)$$

Usually the positions of the atoms \mathbf{r}_i inside the unit cell are not represented in Cartesian coordinates \mathbf{r}_i^c but in fractional coordinates \mathbf{r}_i^f (unit cell coordinates). For a general

2.4. Crystal lattice

triclinic system defined by the lattice parameters $a, b, c, \alpha, \beta, \gamma$ and V , such a basis transformation is performed by Eq. (2.34).

$$\mathbf{B} = \begin{bmatrix} a & b \cos \gamma & c \cos \beta \\ 0 & b \sin \gamma & \frac{c(\cos \alpha - \cos \beta \cos \gamma)}{\sin \gamma} \\ 0 & 0 & \frac{V}{ab \sin \gamma} \end{bmatrix}$$

$$\mathbf{r}_i^c = \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = \mathbf{B} \begin{bmatrix} x_f \\ y_f \\ z_f \end{bmatrix} \quad (2.34)$$

The volume V as used above is given by $V = \mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})$. For the transformation of vector in Cartesian coordinates into fractional coordinates, one simply needs to invert \mathbf{B} :

$$\mathbf{r}_i^f = \begin{bmatrix} x_f \\ y_f \\ z_f \end{bmatrix} = \mathbf{B}^{-1} \begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} \quad (2.35)$$

Since X-ray diffraction is usually analyzed in the reciprocal space, the following relations are used to calculate the reciprocal lattice vectors \mathbf{a}^* , \mathbf{b}^* and \mathbf{c}^* :

$$\mathbf{a}^* = \frac{\mathbf{b} \times \mathbf{c}}{\mathbf{a} \cdot \mathbf{b} \times \mathbf{c}} = \frac{\mathbf{b} \times \mathbf{c}}{V} \quad (2.36)$$

$$\mathbf{b}^* = \frac{\mathbf{c} \times \mathbf{a}}{\mathbf{a} \cdot \mathbf{b} \times \mathbf{c}} = \frac{\mathbf{c} \times \mathbf{a}}{V} \quad (2.37)$$

$$\mathbf{c}^* = \frac{\mathbf{a} \times \mathbf{b}}{\mathbf{a} \cdot \mathbf{b} \times \mathbf{c}} = \frac{\mathbf{a} \times \mathbf{b}}{V}. \quad (2.38)$$

An extremely useful relation between the direct and reciprocal lattice vectors is given in Eq. (2.39) and (2.40). Depending on the definition of the reciprocal lattice vectors the scalar product of the direct and reciprocal lattice vector is either 1 or 2π .

$$\mathbf{a} \cdot \mathbf{a}^* = \mathbf{b} \cdot \mathbf{b}^* = \mathbf{c} \cdot \mathbf{c}^* = 1 \quad (2.39)$$

$$\mathbf{a} \cdot \mathbf{b}^* = \mathbf{a} \cdot \mathbf{c}^* = \mathbf{b} \cdot \mathbf{c}^* = 0 \quad (2.40)$$

An arbitrary vector in reciprocal space is a linear combination of the reciprocal lattice vectors. If the coefficients are integers such as h, k and l they are called *Miller indices* and represent a reciprocal vector \mathbf{h}_{hkl} to a reciprocal lattice point (see Eq. (2.41)).

2. Fundamentals

Furthermore, it is possible to derive a handy expression for the corresponding netplane distance d_{hkl} as shown in Eq. (2.42).

$$\mathbf{h}_{hkl} = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^* \quad h, k, l \in \mathcal{Z} \quad (2.41)$$

$$d_{hkl} = \frac{1}{h_{hkl}} \quad (2.42)$$

One should note that the vector \mathbf{h}_{hkl} is perpendicular to the corresponding (hkl) -netplane. Since \mathbf{h}_{hkl} points to a reciprocal lattice point, it allows one to construct the reciprocal lattice by using all integer combinations of hkl .

2.5. X-ray scattering

A thorough derivation of X-ray scattering involves many technical and rather cumbersome mathematical rearrangements. Therefore the following section will only focus on important key elements in order to introduce some principal concepts of X-ray diffraction. A rigorous derivation can be found in [10] or [34]. The basic concept of elastic

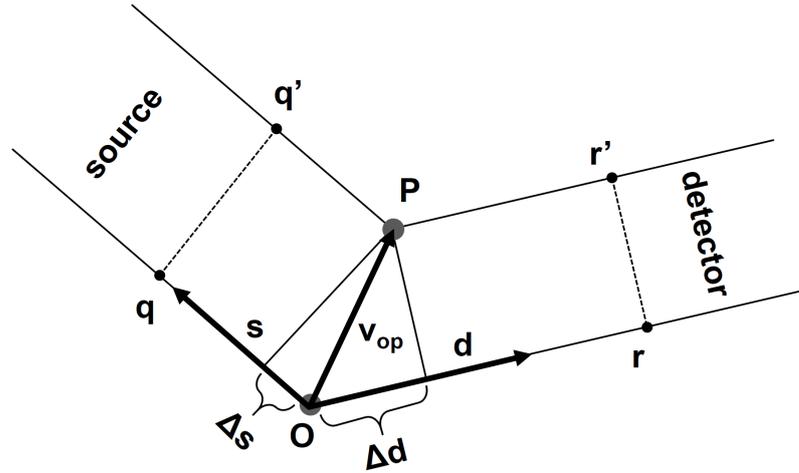


Figure 2.7.: Illustration of a scattering event. Unit vectors \mathbf{s} and \mathbf{d} point in the direction of the source and detector, respectively. An incoming wave is scattered at O and P. The interference is determined by the path difference Δs and Δd .

scattering will be explained on basis of Figure 2.7. The at O scattered electromagnetic (EM) wave is observed at position r . The amplitude of the EM wave is denoted as E_s . At some time t the scattered wave at r is $E_r(t) = E_s \cos \omega t$. However at same time t , the

2.5. X-ray scattering

situation for a wave scattered at P at r' looks different due to the shorter path. The difference is $L = \Delta s + \Delta d$. This path difference can also be expressed as a phase difference $\varphi_{OP} = \omega t - \omega(t + L/c) = -\omega L/c$, where c is the speed of light. Introducing direction vectors \mathbf{s} and \mathbf{d} allows one to define the phase difference by using the projection of the vector \mathbf{v}_{OP} onto \mathbf{s} and \mathbf{d} , respectively:

$$\varphi_{OP} = -\frac{\omega L}{c} = -\frac{2\pi\nu L}{c} = -\frac{2\pi}{\lambda} (\mathbf{v}_{OP} \cdot (\mathbf{s} + \mathbf{d})) \quad (2.43)$$

Switching to the complex representation of the electric field allows for a more compact notation. Using $\omega = 2\pi\nu$ yields:

$$E_O(t) = E_s e^{i2\pi\nu t} \quad E_P(t) = E_s e^{i(2\pi\nu t + \varphi_{OP})} \quad (2.44)$$

The electric field which hits the detector is naturally the sum of both scattered waves E_O and E_P .

$$E_{OP} = E_s \left(e^{i2\pi\nu t} + e^{i(2\pi\nu t + \varphi_{OP})} \right) \quad (2.45)$$

During an experiment the detected intensity will be proportional to the square of the EM field, which will be independent of time. Therefore one can set $t = 0$ in Eq. (2.45) which yields

$$E_{OP} = E_s + E_s e^{i\varphi_{OP}} \quad (2.46)$$

In order to explain the scattering which occurs in a solid, many single scatters have to be included. The scatterer or electron at the origin is removed, however a now hypothetical scattered wave E_s as a reference is still assumed. It stands to reason to use following approach for the diffracted electric field E_d for a n electron system:

$$E_d = E_s \sum_{j=1}^n e^{i\varphi_j} \quad \text{with} \quad \varphi_j = -\frac{2\pi}{\lambda} (\mathbf{v}_{OP_j} \cdot (\mathbf{s} + \mathbf{d})) \quad (2.47)$$

where \mathbf{v}_{OP_j} is the vector pointing to electron j . However, *real* electrons are not located at a point. Due to their quantum mechanical nature the electron charge density or the probability of finding an electron $\rho(\mathbf{r})$ is spatially distributed. Hence the electron density depends on the position \mathbf{r} . The integral over space has to be unity or equal to the number of electron for a multi-electron system. Considering now infinitesimal small

2. Fundamentals

volumes allows one to transform the sum in Eq. (2.47) into an integral over the volume V of the crystal.

$$E_d = E_s \int_V \rho_j e^{i\varphi_j} dV = E_s \int_V \rho_j e^{-i\frac{2\pi}{\lambda}(\mathbf{v}_{OP_j} \cdot (\mathbf{s} + \mathbf{d}))} dV \quad (2.48)$$

In order to connect the geometry with an actual crystal lattice, the position O is set to a crystal lattice point. Thus, the positions of the atoms \mathbf{v}_{OP_j} can be described by lattice vectors and fractional coordinates.

$$\mathbf{v}_{OP_j} = \mathbf{R}_j + \mathbf{r}_j^f = (u_j \mathbf{a} + v_j \mathbf{b} + w_j \mathbf{c}) + (x_j^f \mathbf{a} + y_j^f \mathbf{b} + z_j^f \mathbf{c}) \quad (2.49)$$

By plugging Eq. (2.49) into Eq. (2.48) one gets the rather unhandy expression for the scattered wave of the whole crystal:

$$E_x = E_s \int_V \rho_j e^{-i\frac{2\pi}{\lambda}(u_j \mathbf{a} + v_j \mathbf{b} + w_j \mathbf{c} + x_j^f \mathbf{a} + y_j^f \mathbf{b} + z_j^f \mathbf{c}) \cdot (\mathbf{s} + \mathbf{d})} dV \quad (2.50)$$

Since a crystal consists of a periodic arrangement of unit cells, the integration can be limited to the volume of one single unit cell V_C . The triple sum takes care of the summation over all unit cells in the crystal volume. Also index j in the fractional coordinates can be neglected since the atom positions are the same in each unit cell.

$$E_x = E_s \sum_U \sum_V \sum_W e^{-i\frac{2\pi}{\lambda}(U\mathbf{a} + V\mathbf{b} + W\mathbf{c}) \cdot (\mathbf{s} + \mathbf{d})} \int_{V_C} \rho(\mathbf{r}_j^f) e^{-i\frac{2\pi}{\lambda}(x^f \mathbf{a} + y^f \mathbf{b} + z^f \mathbf{c}) \cdot (\mathbf{s} + \mathbf{d})} dV \quad (2.51)$$

The first part of Eq. (2.51) depends only on lattice constants and defines a necessary condition for diffraction. However, the second part depends on the content of the unit cell and determines the intensity of the peak. The last part which needs modification is the phase difference expression, since it is still not connected to the crystal lattice. Vector \mathbf{s} and \mathbf{d} are pointing in the direction of the source and detector and are divided by the wavelength. Hence they can be expressed as a reciprocal vector:

$$\frac{\mathbf{s} + \mathbf{d}}{\lambda} = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^* = \mathbf{h}_{hkl} = \mathbf{k}_i - \mathbf{k}_f \quad (2.52)$$

Due to the fact that peaks only occur at constructive interference it is no coincidence that the coefficients h, k, l in Eq. (2.52) turn out to be the infamous Miller indices. Eq. (2.52) is known as the *Laue condition* [35] and defines a requirement for observing diffraction peaks. The detector and source must be moved in a position where the vector

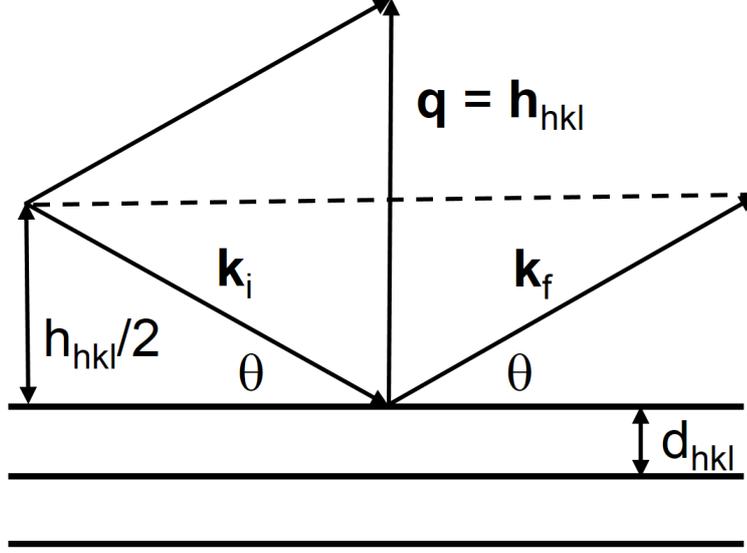


Figure 2.8.: Orientation of incoming and outgoing wave vector \mathbf{k}_i and \mathbf{k}_f , respectively. Scattering vector \mathbf{q} and corresponding net plane distance is indicated as d_{hkl} .

sum of s and d is parallel to \mathbf{h}_{hkl} . Instead of using s and d , it is convenient to use the incoming and diffracted wave vector, \mathbf{k}_i and \mathbf{k}_f , to denote these directions as illustrated in Figure 2.8. Thus, in the case of constructive interference, the scattering vector \mathbf{q} is equal to the reciprocal lattice vector \mathbf{h}_{hkl} . Since we are only considering elastic scattering, \mathbf{h}_{hkl} bisects the angle defined by \mathbf{k}_i and \mathbf{k}_f and furthermore is perpendicular to the corresponding (hkl) net plane. Bragg's law [36] can be obtained by using Eq. (2.42) and $|k| = 1/\lambda$:

$$\sin \theta = \frac{h_{hkl}/2}{k_f} = \frac{h_{hkl}}{2k_f}$$

$$\lambda = 2d_{hkl} \sin \theta \quad (2.53)$$

Combining Eq. (2.51), Eq. (2.52) and considering the helpful relation between the direct and reciprocal vectors in Eq. (2.36) one obtains:

$$E_x = E_s \sum_U \sum_V \sum_W e^{-i\frac{2\pi}{\lambda}(hU+kV+lW)} \int_{V_C} \rho(\mathbf{r}^f)_j e^{-i\frac{2\pi}{\lambda}(x^f h+y^f k+z^f l)} dV \quad (2.54)$$

The first part of Eq. (2.54) also explains that the volume of the crystal or number of unit cells, which take part in the scattering is proportional to the scattered intensity.

2. Fundamentals

However for a given crystal size and incoming intensity, the sums just pose a constant factor D . This allows one to simplify the expression for the scattered EM wave and find its final form:

$$E_{hkl} = D \int_{V_C} \rho(\mathbf{r}_f) e^{-2\pi i(hx_f + ky_f + lz_f)} dV = D \int_{V_C} \rho(\mathbf{r}_f) e^{-2\pi i(\mathbf{h}_{hkl} \cdot \mathbf{r}_f)} dV \quad (2.55)$$

2.5.1. Structure factors

The structure factors F_{hkl} play a important role in crystal structure determination, since they are intimately connected to the the electron density ρ and therefore to the atom positions. The structure factors should be independent of the incident beam intensity and crystal size. This is achieved by dividing Eq. (2.55) by D , which acts as a constant multiplier depending on the crystal size and beam intensity:

$$F_{hkl} = \frac{E_{hkl}}{D} = \int_{V_C} \rho(\mathbf{r}_f) e^{-2\pi i(hx_f + ky_f + lz_f)} dV \quad (2.56)$$

Since usually no analytical expression for the periodic electron density ρ is available, a Fourier series expansion seems to be the obvious solution.

$$\rho(\mathbf{r}_f) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \sum_{o=-\infty}^{\infty} C_{mno} e^{-2\pi i(mx_f + ny_f + oz_f)} \quad (2.57)$$

Plugging the Fourier expression for ρ into Eq. (2.56) yields after a lengthy and cumbersome evaluation that the Fourier coefficients C_{mno} can be expressed by the structure factors $F_{\mathbf{h}}$:

$$\rho(\mathbf{r}_f) = \frac{1}{V_c} \sum_{\mathbf{h}} F_{\mathbf{h}} e^{-2\pi i(\mathbf{h} \cdot \mathbf{r}_f)} \quad (2.58)$$

This means that the electron density ρ and the structure factors $F_{\mathbf{h}}$ are connected by a Fourier transform. Since the electrons are localized around their corresponding atoms, the *spherical atom approximation* is applied. The electron density is modeled as a sphere centered around the atom positions. Using this model, an expression for the structure factor can be obtained, which contains the *atomic form factor* f_j . It defines the scattering amplitude of an isolated atom. Since no convenient analytical expression for f_j is available, it is usually modeled by Eq. (2.59). The coefficients a_i , b_i and c are known as

Cromer–Mann coefficients and can be found in the *International Tables for Crystallography* [37].

$$f_j(q) = \sum_{i=1}^4 a_i e^{-b_i \left(\frac{q}{4\pi}\right)^2} + c \quad (2.59)$$

Figure 2.9 illustrates f_j for elements commonly found in organic molecules and its dependence on the reciprocal scattering vector magnitude q . For $q = 0$ it equates the atomic number of the element. Since f_j increases with the atomic number, it becomes clear that X-rays are not very sensitive to hydrogen and other light elements. Furthermore the contrast between neighboring atoms in the periodic table is rather small which makes it difficult to separate them.

Using the atomic form factor allows one to write the structure factor F_{hkl} in following convenient form which just contains the sum over all atoms in the unit cell:

$$F_{hkl} = \sum_{j=1}^n f_j(hkl) e^{-2\pi i(hx_j^f + ky_j^f + lz_j^f)} = \sum_{j=1}^n f_j(hkl) e^{-2\pi i(\mathbf{h} \cdot \mathbf{r}_j^f)} \quad (2.60)$$

The argument hkl in $f_j(hkl)$ refers to the scattering vector as defined in Eq. (2.41) and \mathbf{r}_j^f is the position of atom j in fractional coordinates. The phase of the complex structure factor F_{hkl} depends solely on the fractional coordinates of the atoms, exactly the information which is required for a crystal structure solution. If sufficient enough structure factor are known, the electron density ρ and therefore the atom position can be reconstructed by means of Fourier synthesis.

2.5.2. Phase problem

The intensity of an EM wave is proportional to E^2 . In the case of an X-ray experiment this means that the measured intensity is proportional to the incoming beam intensity and furthermore depends for the main part on the structure factor:

$$I \propto |F(hkl)|^2 \quad (2.61)$$

This may look convenient at first glance but is a huge problem for crystal structure solution, since only the magnitude of the complex structure factor F_{hkl} is accessible with an experiment. However, for a successful Fourier synthesis of the electron density ρ , the magnitude *and* phase are necessary. This inconvenient circumstance is known as the *phase problem*.

2. Fundamentals

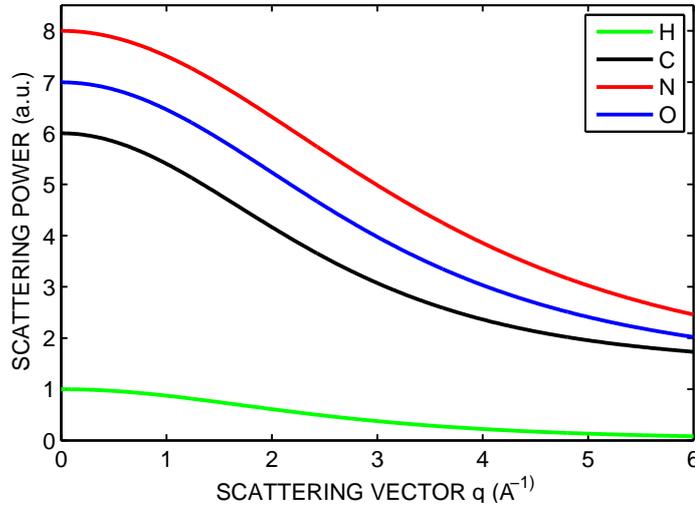


Figure 2.9.: Atomic form factor for hydrogen, carbon, nitrogen and oxygen.

2.5.3. Intensity correction

The recorded intensity for a reflection (hkl) in an experiment is only proportional to $|F(hkl)|^2$. For a successful crystal structure solution it is absolutely crucial to correct the measured intensities to obtain reliable structure factors. Properties of X-rays and the experimental setup have a strong influence on the intensities. Since the direction of polarization of an X-ray beam changes during scattering, the polarization of the incident beam influences the scattered intensity. X-rays with a polarization perpendicular to scattering plane are unaffected and the intensity does not change. However, the electric field of X-rays with polarization in the plane of scattering are reduced by the cosine of the scattering angle 2θ . Hence, the intensity, which is proportional to the electric field squared, is reduced by $\cos^2 2\theta$. In contrast to linear polarized synchrotron radiation, X-rays produced by a typical X-ray tube are a homogenous mixture of all polarization directions and the correction factor is the mean of the two extreme cases and is given in Eq. (2.62).

$$P = \frac{1 + \cos^2 2\theta}{2} \quad (2.62)$$

Another important factor is the Lorentz correction which takes the finite size of reciprocal lattice points into account. The form of the Lorentz factor depends on experimental

2.5. X-ray scattering

arrangement and sample. The relation in Eq. (2.63) holds for powder samples measured in a *Bragg-Brentano* geometry.

$$L = \frac{1}{\sin \theta \sin 2\theta} \quad (2.63)$$

Correction factors for more sophisticated setups such as grazing incidence diffraction are discussed in [11, 38].

3. Methods

3.1. Grazing incidence X-ray diffraction

Grazing incidence X-ray diffraction (GID) is a surface sensitive technique due to a primary beam incidence angle α_i near the critical angle ($\approx 0.2^\circ$ for Cu K_α). A typical experimental setup is given in Figure 3.1 and described in [39]. In contrast to a specular setup, GID probes netplanes perpendicular to the sample surface and is used to study the in-plane order of thin films. The scattering vector \mathbf{q} is split up in the in-plane (\mathbf{q}_p) and out-of-plane (\mathbf{q}_z) component and plotted in a 2D reciprocal space map (RSM).

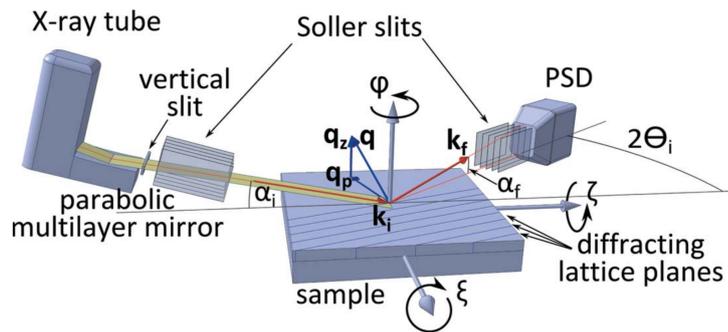


Figure 3.1.: Schematic of a GID laboratory setup using an X-ray tube and 1D detector showing the incidence angle α_i , wave vector \mathbf{k}_i and \mathbf{k}_f of the incoming and scattered beam. Corresponding scattering vector \mathbf{q} and its in-plane (\mathbf{q}_p) and out-of-plane (\mathbf{q}_z) components are shown. Reprinted from [39].

Since the peak positions solely depend on the lattice parameters, the unit cell can be determined by adjusting the lattice parameters until experimental and calculated peak positions match. The peak intensities are extracted by integrating along one direction in the RSM. This yields a 1D intensity profile which is usually fitted by Gaussian, Lorentzian or Pseudo-Voigt functions in order to obtain the peak intensity.

3. Methods

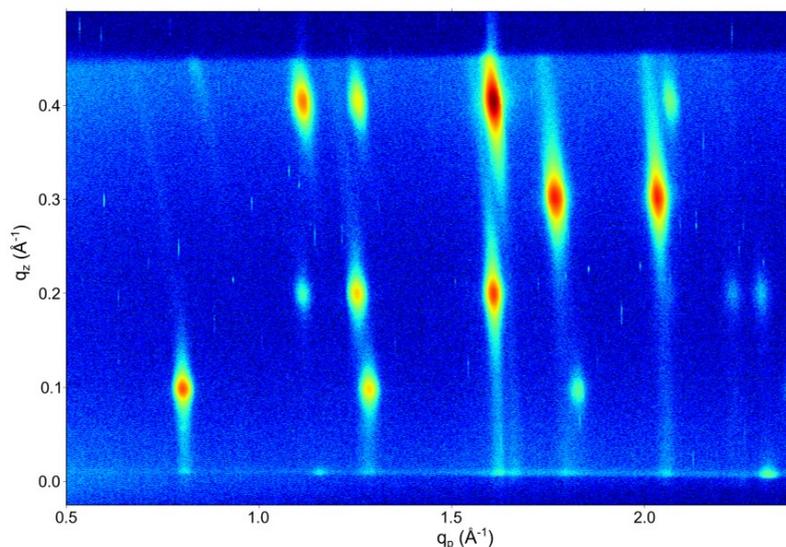


Figure 3.2.: Reciprocal space map of a thin film perylene sample.

3.2. MD simulations with LAMMPS and CGenFF

This part will cover the steps which are necessary to set up MD simulations with *LAMMPS* using the *CHARMM General Force Field (CGenFF)*. *LAMMPS* is an open source classical molecular dynamics simulator that can be used with several force fields. It is available for Windows and Linux and can be used on normal desktop PCs and large scale computer clusters. The *CHARMM* program suite (Chemistry at HARvard Macromolecular Mechanics) was developed to study molecules of biological interest including macromolecules such as proteins, lipids and DNA. It offers among other tools, a MD simulator and force fields. The *CHARMM* simulator itself is not free, however the *CHARMM* force fields can be used free of charge. The *CHARMM* suite offers a variety of force fields which are suited for different applications. The *CHARMM22* and *CHARMM27* force fields are optimized for protein-DNA binding. The most versatile force field is the *CHARMM General Force Field (CGenFF)* [40]. It was designed for small drug-like molecules and covers many typical subunits of small organic molecules.

In order to convert the *CHARMM* force field data to *LAMMPS* compatible files many different file formats and programs are required. Nevertheless, Figure 3.3 tries to provide an overview of all the file formats and programs to avoid too much confusion.

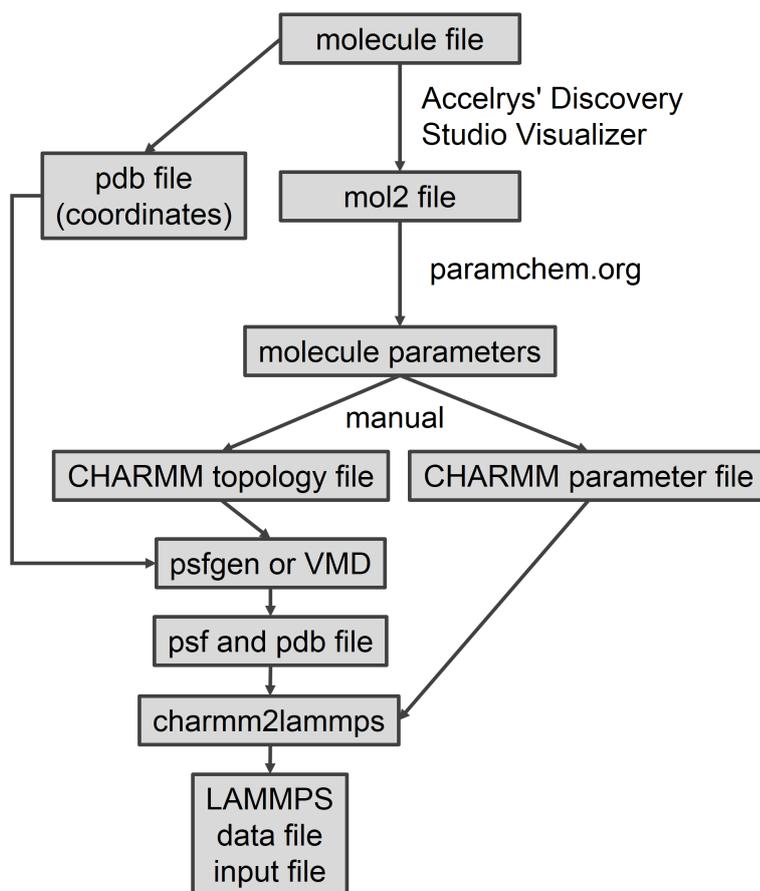


Figure 3.3.: Overview of files and programs included in the conversion of CHARMM to LAMMPS compatible data.

3.2.1. Molecule file formats

Molecules are available in a large variety of data formats. The simplest format is the XYZ format. It just contains the Cartesian coordinates of the atoms and the element. Due to its simplicity it is also quite limited. In crystallography the *Crystallographic Information File* (cif) is very popular. It is used to save complete crystal structures including lattice constants and symmetries. Another very common file format is the *Protein Data Bank* (pdb) format which was used extensively during this work because most programs are compatible with this file format. The format is discussed in more detail in Section 3.2.3.

3. Methods

3.2.2. CHARMM topology and parameter file

The CHARMM topology file (.rtf or .tpc) contains only chemical information (atom types and bonds) about a molecule and no coordinates. An excerpt of a topology file for acetic acid is shown in Listing 3.1. The first entry in the file tells *psfgen* to generate automatically the angles and dihedrals for the molecule. The first part of the file lists all the masses and corresponding atom types. The second part begins with the unique name of the molecule is indicated by the keyword RESI. The expression residue is used to describe a single molecule, although it has a different meaning in the field of macromolecules. The next part of the topology file assigns an unique name to each atom, the type and the partial charge. The last part defines the bonds between the atoms. There is no need to define a bond type, since this information is already provided by the atom type. The actual assignment of atom types, charges and penalty scores is covered in Section 3.3.

Listing 3.1: Excerpt of a CHARMM topology file for acetic acid.

```
1
2 AUTOgenerate ANGLES DIHEDRAL
3
4 MASS 258 HGA3 1.00800 ! alphatic proton, CH3
5 MASS 320 CG331 12.01100 ! aliphatic C for methyl group (-CH3)
6 MASS 293 CG202 12.01100 ! carbonyl C: esters, [neutral] carboxylic acids
7 MASS 359 OG2D1 15.99940 ! carbonyl O: amides, esters, carboxylic acids
8 MASS 266 HGP1 1.00800 ! polar H
9 MASS 372 OG311 15.99940 ! hydroxyl oxygen
10
11 RESI ACE 0.000 ! param penalty= 0.000 ; charge penalty= 0.000
12 GROUP ! CHARGE CH_PENALTY
13 ATOM O1 OG311 -0.600 ! 0.000
14 ATOM H1 HGP1 0.430 ! 0.000
15 ATOM O2 OG2D1 -0.549 ! 0.000
16 ATOM C1 CG202 0.749 ! 0.000
17 ATOM C2 CG331 -0.300 ! 0.000
18 ATOM H2 HGA3 0.090 ! 0.000
19 ATOM H3 HGA3 0.090 ! 0.000
20 ATOM H4 HGA3 0.090 ! 0.000
21
22 BOND O1 H1
23 BOND O1 C1
24 BOND O2 C1
25 BOND C1 C2
26 BOND C2 H2
27 BOND C2 H3
28 BOND C2 H4
29
30 END
```

3.2. MD simulations with LAMMPS and CGenFF

The CHARMM parameter file (.prm) contains the parameters for bonds, angles and dihedrals, all of which are defined by atom types. As one can imagine, there are many thousand possible combinations of atom types and each corresponds to a unique chemical setup with different properties. Listing 3.2 shows only a very small fraction of all available parameters as a demonstration. As one can see, for example, bonds are defined by two atom types and the parameters refer to the force constant and the equilibrium bond length of a harmonic potential. All CGenFF parameters are combined in one single file named *par_all36_cgenff.prm*. The generation of a parameter file and necessary modifications for new molecules are explained in Section 3.3.

Listing 3.2: Excerpt of the CGenFF parameter file.

```
1
2 BONDS
3 CG202 CG331 200.00 1.5220 ! acetic acid pure solvent
4 CG202 OG2D1 750.00 1.2200 ! acetic acid pure solvent;
5 CG202 OG311 230.00 1.4000 ! acetic acid pure solvent
6 CG331 HGA3 322.00 1.1110 ! alkane update
7 OG311 HGP1 545.00 0.9600 ! methanol vib fit;
8
9 ANGLES
10 CG331 CG202 OG2D1 70.00 25.00 20.00 2.44200 ! acetic acid pure solvent;
11 CG331 CG202 OG311 55.00 110.50 ! acetic acid vibrations
12 OG2D1 CG202 OG311 50.00 123.00 210.00 2.26200 ! acetic acid
13 CG202 CG331 HGA3 33.00 109.50 30.00 2.16300 ! acetic acid
14 HGA3 CG331 HGA3 35.50 108.40 5.40 1.80200 ! alkane update
15 CG202 OG311 HGP1 55.00 115.00 ! acetic acid pure solvent
16
17 DIHEDRALS
18 OG2D1 CG202 CG331 HGA3 0.0000 6 180.00 ! from lipid methyl acetate
19 OG311 CG202 CG331 HGA3 0.0000 6 180.00 ! from lipid methyl acetate
20 CG331 CG202 OG311 HGP1 2.0500 2 180.00 ! acetic Acid C-0h rot barrier
21 OG2D1 CG202 OG311 HGP1 2.0500 2 180.00 ! acetic Acid C-0h rot barrier
```

3.2.3. Protein data bank format

This section will explain just the basic structure of a pdb file and how they are used for MD simulations. A full reference of the pdb format can be found online. This file format was originally designed to work with macromolecules. Many features are therefore not used or adapted to work for this kind of MD simulations. Since the pdb format is widely spread most of the programs are able to handle pdb files. Hence they are essential for the communication between various tools and programs.

pdb files contain the coordinates of the atoms but also allow to distinguish between multiple molecules. Listing 3.3 shows a short example of a pdb file for acetic acid. The

3. Methods

first line is an optional header. The second line contains the crystallographic unit cell parameters. Each following line contains information about a single atom. It starts with the *ATOM* keyword and is followed by a unique atom index. The next column is the unique name of the atom. The atom name must be the exact same name as in the CHARMM topology file. The name is the only link between the pdb file which contains the positions of atoms and the topology file which stores the information about the atom types, charges and bonds. The next column contains the *residue name* and is only used for creating the psf file. The fifth column contains the very important *residue id*. This id is used to separate single molecules. The next three columns are the Cartesian coordinates of the atom. The example in Listing 3.3 contains two acetic acid molecules and each molecule consists of eight atoms. The atom index is increasing for all atoms, however the atoms names repeat for each single molecule. Since the pdb file format dates back to time of punched cards the format is very strict and any violation of the column formatting is not allowed. MATLAB provides helpful functions for reading and writing of pdb files (*pdbwrite* and *pdbread*).

Listing 3.3: Example of a pdb file for acetic acid.

```
1  HEADER      CSD ENTRY ACETAC07
2  CRYST1    13.1510   3.9230   5.7620  90.00  90.00  90.00      Pna21
3  ATOM      1  O1  UNK      1      4.912   3.486   0.000   1.00  0.00      O
4  ATOM      2  H1  UNK      1      4.248   3.692  -0.455   1.00  0.00      H
5  ATOM      3  O2  UNK      1      3.248   2.397   0.981   1.00  0.00      O
6  ATOM      4  C1  UNK      1      4.409   2.754   0.978   1.00  0.00      C
7  ATOM      5  C2  UNK      1      5.401   2.417   2.038   1.00  0.00      C
8  ATOM      6  H2  UNK      1      4.958   1.973   2.777   1.00  0.00      H
9  ATOM      7  H3  UNK      1      6.115   1.969   1.717   1.00  0.00      H
10 ATOM      8  H4  UNK      1      5.813   3.229   2.426   1.00  0.00      H
11 ATOM      9  O1  UNK      2      8.239   0.437   2.881   1.00  0.00      O
12 ATOM     10  H1  UNK      2      8.903   0.231   2.426   1.00  0.00      H
13 ATOM     11  O2  UNK      2      9.903   1.526   3.862   1.00  0.00      O
14 ATOM     12  C1  UNK      2      8.742   1.169   3.859   1.00  0.00      C
15 ATOM     13  C2  UNK      2      7.750   1.506   4.919   1.00  0.00      C
16 ATOM     14  H2  UNK      2      8.193   1.950   5.658   1.00  0.00      H
17 ATOM     15  H3  UNK      2      7.036   1.954   4.598   1.00  0.00      H
18 ATOM     16  H4  UNK      2      7.338   0.694   5.307   1.00  0.00      H
```

3.2.4. Protein structure file (psf)

The *Protein Structure Files* (psf) are used in an intermediate stage for the data conversion with *charmm2lammps*. As one can guess from its name, it contains only structural information about a system and no coordinates. Usually manual editing of psf files is not required and therefore a more detailed explanation is not necessary. However they play an important role in data visualization as explained in Section 3.2.8.

3.2. MD simulations with LAMMPS and CGenFF

For the generation of psf files, a CHARMM topology file and a pdb file with the coordinates is required. The conversion is possible with VMD (Visual molecular dynamics), which provides a graphical user interface and therefore requires manual user input. Another more efficient alternative is *psfgen*, a simple console program which can be part of an automated script for structure generation. *psfgen* is part of another MD simulator called *NAMD* and is available free of charge.

psfgen needs a short configuration script to work properly (Listing 3.4). In line 2 the name of the CHARMM topology file is defined. Line 4 is used to change (if necessary) the residue name of UNK to ACE (for acetic acid). The residue name must match the name in the topology file. The segment name is not used in the simulations but needs to be defined. In line 9 the name of the pdb file, which contains the coordinates, is given. The last two lines just define the output name of the psf and pdb file.

Listing 3.4: Input script for psfgen.

```
1 package require psfgen
2 topology top_acetic.rtf
3
4 pdbalias residue UNK ACE
5 segment DBD1 {
6   pdb unit_cell.pdb
7 }
8
9 coordpdb unit_cell.pdb DBD1
10
11 writepsf unit_cell.psf
12 writepdb unit_cell.pdb
```

3.2.5. charmm2lammps

Finally *charmm2lammps* is used to create the input files for LAMMPS and is part of the LAMMPS package. However *charmm2lammps* is a Perl script which relies on the use of many Linux command line tools. Therefore it is not possible to use *charmm2lammps* on Windows without *Cygwin*. *Cygwin* provides a minimal Unix-like environment for Windows. It includes many of the useful standard tools which are part of every Linux distribution. More information about *Cygwin* is available online.

charmm2lammps needs a psf and pdb file (structural information + coordinates) and the CHARMM parameter file to create the LAMMPS data file which includes all the information about the system for the simulation. The generated input file (.in) for LAMMPS can be ignored.

3. Methods

3.2.6. LAMMPS data file

The LAMMPS data file contains the structural information and all the force field parameters in LAMMPS format. The first part of the file is a summary of the parameters of the system. The important part begins at line 12. The next 4 lines define the dimensions and geometry of the simulation box. Line 12 to 14 define the x , y , z dimensions of the box for a rectangular box. Line 15 is used to define a non-orthogonal (triclinic) box. It is important to note, that this formalism differs from the usual crystallographic unit cell definition. For non-orthogonal unit cells the conversion as described in [41] needs to be applied in order to convert unit cell parameters into the correct LAMMPS box parameters. This part of the data file is modified to fit the experimental unit cell of the organic thin film structure for the MD simulation.

The remaining part of the data file lists all atom types, initial atom positions and charges, intra-molecular pairs and corresponding coefficients, as well as the inter-molecular pair coefficients. Usually there is no need to modify this part of the data file. Again, a comprehensive description of the data file is available online in the LAMMPS documentation.

3.2.7. LAMMPS input file

The actual simulation is set up in the input file (.in). Listing 3.5 shows a simple LAMMPS input script that performs a simulation run of 50 ps. The initial temperature is set to 100 K and decreases during the simulation to 50 K.

The most important commands and settings will be explained in more detail. In line 3 the system of units is defined. The keyword *real* sets the measure of length to Ångströms, time to femtoseconds and energy to kcal/mol. This setting was used for all further simulations. Line 10 defines the type of long range inter-molecular interaction and the corresponding cutoff radii. The most important setting is the *kpace_style* setting. The so called particle-particle-particle mesh (PPPM) is a very efficient method to deal with the long range particle interaction. However for LAMMPS (version 1 Jul 2013) it was only implemented for orthogonal simulation boxes. In order to use triclinic boxes it was necessary to switch to the less efficient Ewald summation schema. The loss in performance presents not a large problem due to the relatively small particle numbers. It seems that for more recent versions of LAMMPS, PPPM is also supported for triclinic systems.

3.2. MD simulations with LAMMPS and CGenFF

The information about the particles and force field is defined in line 14 and is read from the previously generated data file (see Section 3.2.6). Line 16 is used to control the data dumping. It tells LAMMPS to save all atoms positions every 1000 time steps in a dcd file. Line 17 is used to unwrap the coordinates of the atoms. Due to the periodic boundary conditions, molecules which stick out of the simulation box are broken up into multiple pieces. In order to conserve the complete molecular structure in the output files, unwrapping of the coordinates is applied. This means that the periodic boundary conditions are ignored for the coordinates in the output file.

thermo and *thermo_style* defines the periodic output of thermodynamic data as such potential energy, kinetic energy, temperature etc. *timestep* obviously defines the time step for the simulation and is was set to 1 fs. *variable* allows one the define symbolic variables. *minimize* is used to perform a energy minimization before the simulation start to minimize the potential energy of the system due to a suboptimal starting geometry, mostly due to too close adjacent atoms. Without such an energy minimization it is very likely to have an unusual high potential energy in the system which renders the simulation useless.

In line 30, initial velocities are assigned to the atoms to represent a temperature of 100 K. *fixes* are used to control the simulation. In line 31, the *momentum fix* is used to zero the linear momentum of the system, to avoid the so called *flying ice cube* effect [42]. The next *fix* in line 32 sets up an NVE (constant number of particles, volume and energy) simulation. Such an simulation represents a microcanonical ensemble. In order to control the temperature of the system a Berendsen thermostat *fix* was used. The temperature is decreasing during the simulation from 100 K to 50 K. Finally, the *run* command in line 34 tells the simulator to calculate 50000 time steps. As usual, a more detailed explanation of all commands and *fixes* is available online in the LAMMPS documentation [43].

Listing 3.5: Input script for LAMMPS.

```
1 # Minimize the molecules in an expanded unit cell
2
3 units          real
4 neigh_modify  delay 2 every 1
5
6 atom_style    full
7 bond_style    harmonic
8 angle_style   charmm
9 dihedral_style charmm
10 pair_style    lj/charmm/coul/long 8. 10.
11 pair_modify   mix arithmetic
12 kspace_style  ewald/disp 1e-4
13 special_bonds charmm
```

3. Methods

```
14 read_data          cell.data
15
16 dump               1 all dcd 1000 minimization.dcd
17 dump_modify       1 unwrap yes
18
19 thermo_style      custom step cpu etotal temp press evdwl ecout eangle
    edihed pe ke lx ly lz
20 thermo            1000
21 #thermo_style     multi
22 timestep          1
23
24 variable var_temp equal 100.0
25 variable var_temp_end equal 50.0
26
27
28 minimize 1e-6 1e-9 1000 100000
29 # Run the MD simulation.
30 velocity all create ${var_temp} 5782341 dist uniform
31 fix 1 all momentum 1 linear 1 1 1
32 fix 2 all nve
33 fix 3 all temp/berendsen ${var_temp} ${var_temp_end} 100.
34 run               50000
```

Listing 3.6: LAMMPS data file.

```
1 Created by charmm2lammps v1.8.1 on Wed, Nov 20, 2013 3:52:38 PM
2      50 atoms
3      53 bonds
4      86 angles
5      118 dihedrals
6
7      6 atom types
8      7 bond types
9      10 angle types
10     15 dihedral types
11
12     -4.178      12.321 xlo xhi
13     -12.561     12.561 ylo yhi
14     0.425      40.934 zlo zhi
15     0.027 0.937 2.371 xy xz yz
16
17 Masses
18     1      1.008 # HGA3
19     2      1.008 # HGR61
20     3     12.011 # CG2R61
21     4     12.011 # CG2R67
22     5     12.011 # CG331
23     6    15.9994 # OG301
24 Pair Coeffs
25     1      0.024 2.387609 0.024 2.387609 # HGA3
26     2      0.03 2.420037 0.03 2.420037 # HGR61
27     3      0.07 3.550053 0.07 3.550053 # CG2R61
28     4      0.07 3.550053 0.07 3.550053 # CG2R67
29 Atoms
30     1      1      6    -0.391 3.828 3.651 11.665 # OG301
31     2      1      3     0.219 3.756 3.185 12.954 # CG2R61
```

3.2. MD simulations with LAMMPS and CGenFF

```
32      3      1      3    -0.115 3.085 3.805 13.98 # CG2R61
33      4      1      2     0.115 2.627 4.6   13.827 # HGR61
34      5      1      3    -0.114 3.097 3.239 15.243 # CG2R61
35 Bond Coeffs
36      1      305     1.375 # CG2R61  CG2R61
37      2      305     1.375 # CG2R61  CG2R67
38      3      340     1.08  # CG2R61  HGR61
39      4      230     1.382 # CG2R61  OG301
40 Bonds
41      1      7      1     22 # CG331  OG301
42      2      4      1     2  # CG2R61  OG301
43      3      1      2     10 # CG2R61  CG2R61
44      4      1      2     3  # CG2R61  CG2R61
```

3.2.8. LAMMPS output

LAMMPS allows one to use several file formats and methods to save data during a simulation run. Usually the positions of the particles are periodically saved in a file. Such a file, commonly called a dump file, contains therefore the trajectories of the particles and allows one to visualize the system's behavior over time. Since the amount of data is quite large, even for small system many hundred megabyte of data are not uncommon, a simple format like the xyz format would not be efficient. Therefore the compressed dcd file format is used by most MD simulators to store trajectories and additional data.

However dcd files only hold information about the coordinates. The structural information about the system (bonds) is only available through the psf file, which was generated during the preparation of the simulation. VMD [44] provides the easiest way to visualize dcd files, only in combination with the correct psf file, and allows one to analyze many different aspects of a simulation. It is also capable of rendering high quality images by using the Tachyon ray tracing engine [45].

In order to use dcd files in MATLAB, a third-party library can be found here [46].

3. Methods

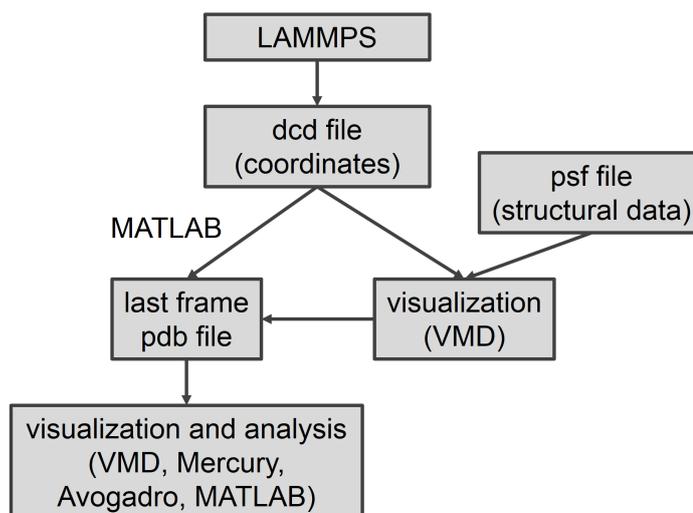


Figure 3.4.: Overview of LAMMPS data output.

3.3. Parametrization by analogy

The last sections covered the conversion of the CHARMM data in LAMMPS format. For this purpose it was assumed that the CHARMM topology and CHARMM parameter file already available. This section will explain the generation of the topology and parameter file.

In order to perform a successful MD simulation the correct atom types need to be assigned to each atom of a molecule. Since manual parametrization or even re-optimization for new molecules is quite time consuming and also prone to human error, the assignment of atom types new parameters is therefore performed by a method called *Parametrization by analogy*. This method tries to compute new charges and use already available parameters of *CGenFF* to find those missing parameters. *Analogy* refers to the fact that parameters of chemically similar compounds are used during the parametrization. For each new parameter a penalty score is calculated, which tries to quantify the dissimilarity between the parameters for the new molecule and the reference material. A more detailed description of this process and penalty scores can be found in [47].

In order to use this kind parametrization several steps are necessary. The only file format which works properly is *mol2*. Unfortunately, there are only a few editors which are capable of editing and saving a *mol2* files which are conform to the *mol2* standard. The only easy available editor so far is *Accelrys' Discovery Studio Visualizer*. It is a proprietary but free of charge software. For a successful parametrization all bond orders must be

chosen correctly. The created *mol2* file is subsequently uploaded to *paramchem.org*. This website implements the current *Parametrization by analogy* algorithm and allows the user to download a simple text file which contains the results (atom types and new intra-molecular parameters) of the parametrization.

3.3.1. Generation of CHARMM topology files

The part of the parametrization results concerning the the topology (atom types and charges) are moved to new topology file. If already all parameters are available in *CGenFF*, this means for every combination of atom types present in the new molecules, exists an entry the CHARMM parameter file, there is no need to modify or expand the parameter file. If this is not the case, the best match (lowest penalty score) is used to replace the missing parameters. However this new combination of atom types is of course not the standard CHARMM parameter file. Therefore it is necessary to add these new parameters. Another possibility is to copy all parameters for the molecule, including the new one, in a separate file. If parametrization by analogy fails to find a suitable replacement for the missing parameter, re-optimization of the best parameter set is necessary. This quite time consuming task is described in Section 3.6.

3.4. Shrinking cell simulation

Usually MD simulations are performed with large super cells which consists of several unit cells in order to achieve a proper statistical ensemble and to avoid that molecules interact with its own image in a neighboring cell. Since we are not interested in statistical properties, such as thermodynamic quantities, the number of unit cells does not pose a problem. However the size of the simulation box could lead to unphysical structures but since the results will be compared to X-ray data, such structures can be easily rejected. Furthermore, a super cell approach would bring along different challenges, such as increased calculation times, averaging over multiple unit cells etc.

The work flow of the shrinking cell MD simulation is illustrated in Figure 3.5. The simulations are based on the experimental unit cell parameters either obtained by GID experiments or literature. At first, the initial packings were created by randomly distributed molecules. Such a straightforward approach was not successful since a unit cell is too small and molecules are locked in their initial configuration. Therefore it was necessary

3. Methods

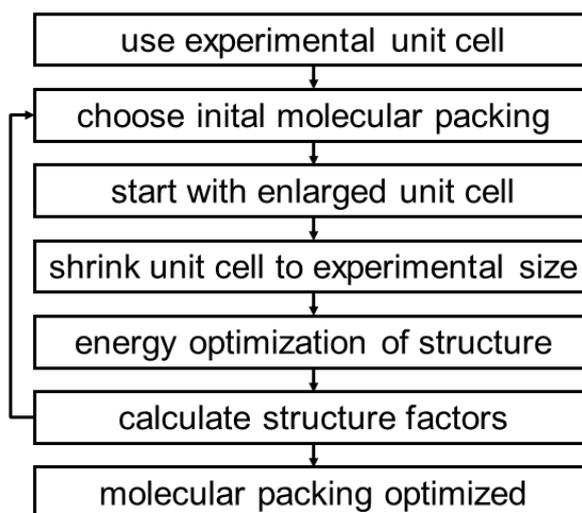


Figure 3.5.: Work flow of the shrinking cell MD simulation.

to use a slightly larger simulation box. Subsequently, the box was shrunk to its experimental size during the simulation. The initial size and shrinking speed is essential for a successful realization of this approach. If the initial cell was too large, all information of the initial packing was lost. On the other hand, if the cell shrunk too fast, molecules did not have enough time to rearrange properly and the system got stuck in a high energy configuration. An expansion by a factor of 1.1 to 1.3 and shrinking over 50 ps produced satisfying results. Nevertheless, several hundred independent simulations with different initial conditions were performed in order to avoid local minima.

Depending on the complexity of the simulated system the choice of initial molecular packings can be crucial. The ideal case is, of course, a random generation of initial packings to assure unbiased simulations in order to fully explore the parameter space. However, a complete random generation of initial packings produces also many useless simulations. In order to increase the efficiency, constraints in form of crystallographic motifs, such as herringbone or π -stacking, were introduced and modified randomly by translations and rotations within reasonable limits depending on the system under investigation.

An excerpt of the LAMMPS input file with comments is shown in Listing. 3.7. The simulations was carried out using Ewald summation since PPPM was not supported for triclinic systems in this version of LAMMPS with a time step of 1 fs. The radii for the smooth cutoff region for long range forces was set to 8 and 10 Å. A Berendsen thermostat [48] was applied for controlling the temperature. The shrinking was performed

3.4. Shrinking cell simulation

over 50 ps at 150 K and the subsequent cool down from 150 K to 20 K took 20 ps. The last step was an energy optimization based on a conjugate-gradient search in the force field.

Since several hundred or even thousand individual simulation runs were required a fully automated script was developed in MATLAB to set up the simulations. A prototype structure (pdb file), containing the content and size of the unit cell was used to generate the initial structure and saved as a pdb file. This file was used for generating the psf file and subsequently the LAMMPS data file by utilizing the *charmm2lammmps* script. The LAMMPS data file was modified to use the expanded unit cell size and the correct triclinic tilt parameters if necessary. After the simulation, the energy of the final results was extracted from the LAMMPS log file and the dcd file containing the coordinates was converted to pdb.

Listing 3.7: Input script for shrinking cell simulations in LAMMPS.

```
1 # shrinking cell simulation
2 timestep 1
3
4 # factor for shrinking, box dimension set to 1.1 exp. size
5 variable var_scale equal 1./1.1
6
7 # FF energy optimization
8 minimize 1e-4 1e-6 200 1000
9
10 velocity all create 150. 5782341 dist uniform
11 fix 1 all momentum 1 linear 1 1 1
12
13 # use constant energy and volume (NVE) simulation
14 fix 2 all nve
15 # set temperature to 150 K during shrinking
16 fix 3 all temp/berendsen 150. 150. 100.
17
18 # shrinking to experimental cell size in 50000 steps (50 ps)
19 fix 4 all deform 1 x scale
20     ${var_scale} y scale ${var_scale} z scale ${var_scale} units box
21 run 50000
22 unfix 3
23 unfix 4
24
25 # cool down to 20 K in experimental unit cell
26 fix 5 all temp/berendsen 150. 20. 100.
27 run 20000
28
29 # FF energy optimization
30 minimize 1e-4 1e-6 200 1000
31
32 #save last frame
33 dump 1 all dcd 1 shrinking.dcd
34 dump_modify 1 unwrap yes
35 run 0
```

3.5. DFT optimization

For DFT optimizations the *Vienna Ab initio Simulation Package* (VASP) [49–52] was utilized. VASP is based on pseudo potentials and plane waves and uses periodic boundary conditions. For the optimization *projector augmented wave* (PAW) potentials [53, 54] implementing GGA functionals with an energy cutoff of 345 eV were used. The k-mesh was generated automatically by using a Monkhorst-Pack scheme [55]. A VASP calculation requires four input files.

- POTCAR: contains the pseudo potentials
- POSCAR: unit cell and atom positions
- KPOINTS: defines the k-mesh
- INCAR: what to do and how to do it (Listing 3.8)

The complete documentation explaining various files and parameters is available online [56]. POSCAR is a very simple format that contains more or less just coordinates. However the ordering of the elements must be consistent with the POTCAR file. Therefore the atoms are usually ordered by elements, which in most cases messes up the accustomed atom order of the MD simulation results (pdb file). The easiest way to get on top of this annoying peculiarity of the POSCAR format is to write a special pdb to POSCAR converter. This allows one to restore the original ordering of atoms during the conversion of the VASP results back into the pdb format. VASP calculation were carried out on the *dcluster* at TU Graz.

Listing 3.8: Input script for geometry relaxation in VASP.

```
1 SYSTEM = Pentacene
2
3 ISTART = 0
4 ICHARG = 2
5 ISMEAR = -5
6 EDIFF = 1.0E-4
7 EDIFFG = -0.01
8 LWAVE = .TRUE.
9 LCHARG = .TRUE.
10 LVTOT = .FALSE.
11 ENCUT = 345.00 eV # energy cut-off for the calculation
12 ISIF = 2 # forces and stress are optimized
13 NPAR = 4
14
15 IBRION = 2 # CG algorithm
16 NSW = 100 # 100 ionic steps
```

3.6. Potential energy surface scans

The first step for adding new parameters to a force field is the creation of quantum mechanical target data, usually obtained by ab initio calculations. The aim is to find force field parameters which are able to reproduce the target data. An optimization procedure was developed for dihedral parameters, since they are most important for the molecular packing. A dihedral potential (or more generally a 1D potential energy surface), as described in more detail in Section 2.2.4, is the energy in dependence of the dihedral angle. Thus, the dihedral in question, is rotated to a certain angle and the total energy of the molecular conformation is calculated. Potential energy surface (PES) scans were performed in *Gaussian09* and LAMMPS. The next section will explain the generation of the QM target data and how to perform PES scans with LAMMPS.

3.6.1. Gaussian09

During this work Gaussian09, a widely used quantum chemistry program, was used for geometry optimizations and for dihedral potential calculations. The results were used as target data for the FF optimization. Gaussian offers various levels of theory, including HF and post HF methods, like Møller-Plesset perturbation theory (MP2) and of course several DFT based methods like B3LYP, PBE and PW91.

Geometry optimization

An input file for a geometry optimization is given in Listing 3.9. The first two lines are used to define the available memory and processors for the calculation. The geometry optimization is selected by the keyword *OPT* and is carried out for example at the B3LYP/cc-pVDZ level of theory. *pop=esp* produces charges fit to the electrostatic potential. It is used for calculating the partial charges of atoms with a method called *Restrained Electrostatic Potential* (RESP). Unfortunately, such an approach is not eligible for CHARMM force fields. The initial geometry is given in Cartesian coordinates. The other keywords are used for output formatting and are not important at the moment. However, the interested reader may be referred to the online documentation of Gaussian [57].

3. Methods

Listing 3.9: Gaussian input file for geometry optimization. Geometry is given in Cartesian coordinates.

```
1 %mem=100MB
2 %nprocshared=8
3 #p B3LYP/cc-pVDZ ginput IOP(6/7=3) OPT pop=esp
4
5 2T
6
7 0 1
8 C      20.17400      -5.76700      -0.00100
9 C      19.67000      -4.52000      -0.00100
10 S      18.91300      -6.87100       0.00000
11 C      18.13900      -4.52800       0.00000
12 C      17.64400      -5.77100       0.00000
```

Potential energy surface scan

If a geometry optimization is carried out for each step of the potential energy surface scan the obtained PES is called a *relaxed* PES. The geometry optimization is necessary because the rotated molecule conformation might have a different relaxed geometry compared to the unrotated conformation. Since input files for Gaussian calculations are quite tricky an example for a relaxed PES is presented in Listing 3.10. The important settings are in line 3. *opt=modredundant* is used to set up an geometry optimization using redundant coordinates, which are provided at the end of the file. *B3LYP/6-311G** defines the level of theory.

For PES scans the use of internal coordinates is required. Gaussian uses the *Z-matrix* notation to represent to molecular structure. The geometry of the molecule is defined by relative distances and angles (bond length and angle, dihedrals). The conversion from Cartesian coordinates into *Z-matrix* representation can be done with several molecular file viewer, for example *Avogadro*, which is also a convenient tool for viewing Gaussian log files. The most important part are the last two lines. These lines are used to define a new dihedral angle between atoms (18, 11, 12, 9). The numbers refer to the ordering in the *Z-matrix*. (*, 11, 12, *) tells Gaussian to include all dihedrals which are connected to the center atoms 11 and 12. Otherwise the rotated structure would be useless. *S 36 5.0* is used to perform a scan over the dihedral with a step size of 5.0° for 36 times. The result of the scan can be extracted from the Gaussian log file. Since Gaussian log files are absolutely huge and confusing a handy script was developed for efficient extraction of the required data.

3.6. Potential energy surface scans

Listing 3.10: Gaussian relaxed potential energy scan input file (not all atoms and variables shown). Geometry is given in Z-matrix notation.

```
1 %mem=100MB
2 %nprocshared=8
3 #p opt=modredundant B3LYP/6-311G* nosymm ginput IOP(6/7=3)
4
5 relaxed pes scan
6
7 0 1
8 C
9 C 1 B1
10 H 2 B2 1 A2
11 C 1 B3 2 A3 3 D3
12 H 4 B4 1 A4 2 D4
13 C 2 B5 1 A5 3 D5
14 H 6 B6 2 A6 1 D6
15 C 6 B7 2 A7 1 D7
16 C 4 B8 1 A8 2 D8
17 C 9 B9 4 A9 1 D9
18 ...
19 Variables:
20 B1      1.39796
21 B2      1.08617
22 A2      120.13426
23 B3      1.39702
24 A3      119.43305
25 D3      178.51562
26 B4      1.08603
27 A4      120.16348
28 ...
29
30 * 11 12 *
31 18 11 12 9 S 36 5.0
```

3.6.2. LAMMPS

LAMMPS does not offer directly an option for PES scans. However it is possible to constraint certain bonded parameters during a simulation run. Hence, a dihedral parameter was set to a fixed value and a geometry optimization of the molecule was carried out in a large simulation box. The large box avoids interaction over the periodic boundary conditions. A high temperature at the beginning is necessary to escape local energy minima. Subsequently, the temperature was lowered during the simulation run in order to reach the global energy minimum. To obtain the PES scan, a separate run for each dihedral angle was performed. A MATLAB script was used to start the simulations and subsequently extract the energies of the LAMMPS log files. Listing 3.11 shows the LAMMPS input file for the calculation for on point of the PES scan. By setting the dihedral constant to 10 000 kcal/mol in line 5 and 9, the dihedral defined by atoms (22, 10,

3. Methods

9, 24) is fixed to value of *angle1*. This constant, the initial temperature and cooling rate may have to be adjusted to work properly with the system under investigation.

Listing 3.11: Input script for PES scan with LAMMPS based on [58].

```
1 # minimize molecule energy with restraints
2 velocity all create 600.0 8675309 mom yes rot yes dist gaussian
3 fix NVE all nve
4 fix TFIX all langevin 600.0 0.0 100 24601
5 fix REST all restrain dihedral 22 10 9 24 0.0 10000.0 ${angle1}
6 fix_modify REST energy yes
7 run 10000
8 fix TFIX all langevin 0.0 0.0 100 24601
9 fix REST all restrain dihedral 22 10 9 24 10000.0 10000.0 ${angle1}
10 fix_modify REST energy yes
11 run 10000
12 # sanity check for convergence
13 minimize 1e-6 1e-9 1000 100000
14 # report unrestrained energies
15 unfix REST
16 run 0
```

3.6.3. Dihedral parameter fitting

The fitting procedure for missing dihedral parameters is based on the description in [59]. The idea is to fit the energy difference between the QM and force field data. At first, parameters for each dihedral were assigned during the parametrization by analogy as described in Section 3.3. To obtain the energies without the contribution of the dihedral which is to be optimized, the force constant for this dihedral is set to zero and a PES in LAMMPS is recorded. The easiest way to manipulate the dihedral parameters is simply by locating the corresponding line in the LAMMPS data file and changing the dihedral constant. Subsequently, the difference of the relative energies (by subtraction of the energy minimum of the respective scan) of the QM and LAMMPS PES is calculated and fitted by using the function of the dihedral potential as defined in Eq. (2.14).

The new optimized dihedral parameters are used for another PES scan in LAMMPS to check if the optimization worked. If this is the case, the force field should be able to reproduce the QM energies within reasonable limits. The same procedure can be used to check a suspicious dihedral (high penalty score) if further optimization is necessary. This is simply done by comparing the relative energies of the QM and LAMMPS result.

3.7. StructFactViewer

In order to work efficiently with the amount of data produced by the simulations, a MATLAB tool called *StructFactView* was developed. It is capable of batch processing the MD results and calculating the structure factors based on Eq. (2.60). The structures are sorted according to the difference of all or a selection of structure factors based on experimental data. The graphical user interface allows one to navigate easily through to results and visualize the structure factor spectrum and molecular packing. Even the influence of the position and orientation of molecules on the structure factor spectrum can be studied by rotating and translating molecules in the unit cell. Another feature is the capability to load .dcd files containing several frames of a MD run and visualize the structure factors for each time step and in doing so, the impact of thermal motions or other conformational changes on structure factors can be investigated.

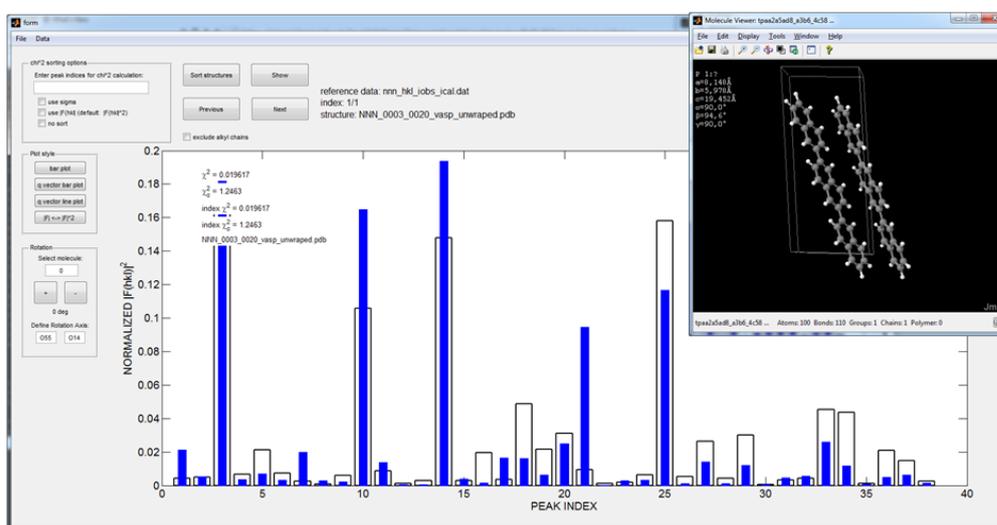


Figure 3.6.: Screenshot of StructFactViewer.

3.8. Computational equipment

Gaussian09, VASP and partly LAMMPS calculations were carried out on the *dcluster* at TU Graz. It consists of 120 computing nodes, each of which is built up of two octo-core Intel XEON processors and 64 GB RAM. The total performance of the dCluster is

3. *Methods*

approximately 30 Teraflops. VASP was executed on 16 CPUs and for Gaussian09 usually eight CPUs were used.

The less computationally intensive shrinking cell simulations with LAMMPS were carried out on a desktop PC consisting of an Intel Core i7-3770 quad-core processor and 8 GB RAM. Due to hyper threading, it was possible to use eight parallel threads for the calculations. Roughly a thousand shrinking cell simulations were possible in approximately 12 hours, for unit cells containing about 100 atoms.

4. Results and Discussion

This chapter will present and discuss the results obtained by MD simulations and DFT calculations. The unit cell parameters for each molecule are summarized in Table 4.1

Table 4.1.: Experimental lattice constants used for the shrinking cell simulations.

molecule	a (Å)	b (Å)	c (Å)	α (deg)	β (deg)	γ (deg)	ref.
pentacene	5.958	7.596	15.61	81.25	86.56	89.80	[6]
NNN	8.15	5.98	19.45	90.00	94.58	90.00	[12]
s-DOTT	5.43	7.71	33.2	90.00	90.00	90.00	[60]
b-DOTT	5.59	7.56	32.4	90.00	90.00	90.00	[60]
perylene	11.277	10.826	10.263	90.00	100.55	90.00	[61]
B-DBDCS	9.851	10.094	27.455	85.42	89.67	85.91	[12]

4.1. Pentacene

Pentacene consists of five linearly-fused benzene rings and is one of the most famous representatives of acenes. Due to its high electron mobility it is widely used in organic electronics and was investigated on many occasions. The combination of its rigid geometry and the availability of several reference solutions of the thin film phase makes pentacene an ideal candidate to check the capabilities of MD simulations and especially the *CGenFF* force field. The experimental data was provided by Armin Moser, who used rigid body refinement to solve the crystallographic structure of the film phase.

4.1.1. Parametrization

The parameters for pentacene were obtained by analogy as described in Section 3.3. All force constants (bonds, angles and dihedrals) and partial charges are available in

4. Results and Discussion

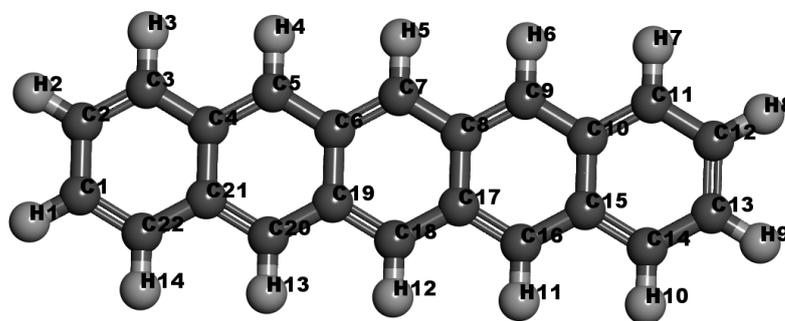


Figure 4.1.: Pentacene molecule with atom labels as used for MD simulations.

CGenFF without any penalties. The obtained atom types and charges are shown in Table 4.3).

4.1.2. Energy

For a first evaluation only the energy of the final pentacene structures were considered. As expected there is a significant difference in energy of approximately 11 kcal/mol or 0.48 eV between the π -stack and herringbone motif (see Figure 4.2). Since most of the herringbone structures are too close in energy it is not possible to select an unambiguous solution. Therefore it was necessary to compare the results with complementary X-ray data. The final selection of the correct structure is based on the experimental structure factors $|F(hkl)|^2$.

4.1.3. Structure factors

The difference $\chi^2 = \sum_{hkl} (|F_{calc}(hkl)|^2 - |F_{exp}(hkl)|^2)^2$ between the experimental and calculated structure factors was used to evaluate the results. Figure 4.3 shows the comparison of the best match based on structure factors. The calculated structure factors are in excellent agreement with the experimental data. As already illustrated in the energy plot, the packing of this structure features the typical herringbone motif. Since the same data set was used for a rigid body refinement of the thin film pentacene phase,

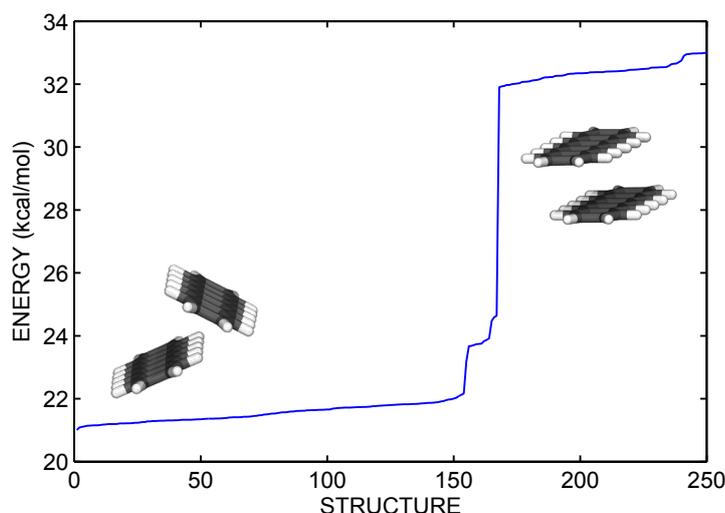


Figure 4.2.: First 250 MD results for pentacene in order of increasing energy.

the first five structure factors were set to zero as an additional constraint during the optimization. However the herringbone angle θ of 50.5° is about 4.5° too low compared to published results (see Table 4.2).

There is clearly a correlation between χ^2 and energy since a low χ^2 indicates a good agreement with the experimental data and therefore the actual molecular packing, which is of course a low energy packing. In Figure 4.4 the correlation of χ^2 and the energy is illustrated. The difference between lowest energy packing and the best X-ray data match is only 0.28 kcal/mol. Due to the large number of degrees of freedom and thermal motion, which is always present in MD simulations, a perfect match in energy is not very likely. Even the geometry optimization at the end of the simulation run gets stuck easily in one of the many local energy minima. This gives rise to many different molecular packings, which are very close in energy and geometry that can be considered to be equal though.

All structures with energies above 22 kcal/mol are a result of failed random initial packings. Rotating and translating of molecules without constraints can lead to interlocking molecules. Such unphysical packings are of no use and are neglected. Only solutions on the left bottom of Figure 4.4 (low χ^2 and low energy) are of interest.

In order to compare the simulation results with solved pentacene thin film phases, the orientation angles of the molecules were calculated (see Table 4.2). One can see that the tilt angle χ (5.5°) is significantly lower than in the bulk phase (22°) and agrees pass-

4. Results and Discussion

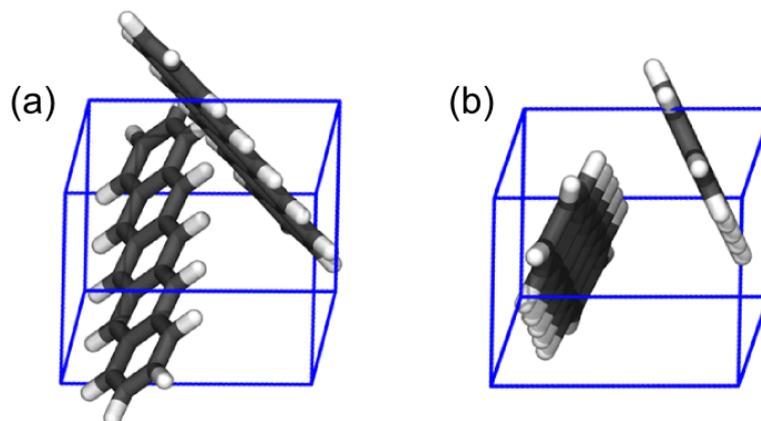


Figure 4.3.: (a) Initial random molecular packing for MD simulation. (b) Final packing after MD simulation and conjugate gradient geometry optimization.

ably with the reference solutions. The upright orientation is typical for surface induced phases. The herringbone angle θ of 50.5° is roughly 4.5° to small compared to the reference solutions. Since many initial packings produced no useful results, in consequence of the unrestricted random initial packings, constraints were introduced to focus on more likely packings. Just applying information which was gained from the simulations so far, it is quite obvious that only a herringbone motif is able to match the experimental data and low energy criterion. Hence the initial packings were limited to slightly modified herringbone packings. The best MD result was used as a template and rotations and translations within reasonable limits were applied to randomize this structure and a second batch of MD simulations was performed. However the herringbone angle just improved by 1.1° which is still too less.

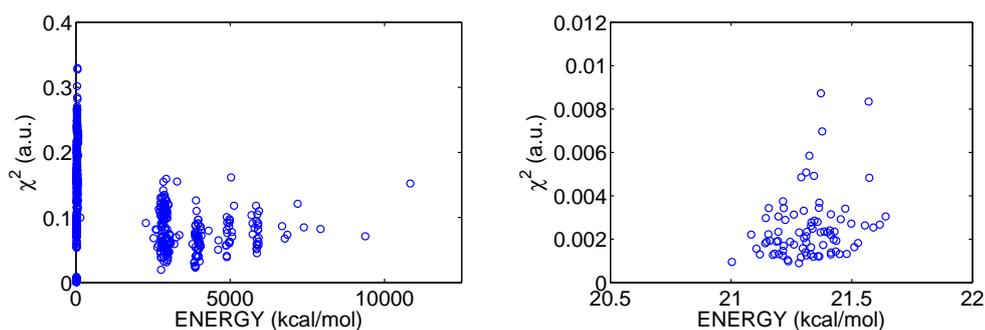


Figure 4.4.: Correlation of energy and χ^2 for pentacene results. Right: all results. Left: Region with reasonable results.

Since pentacene is a remarkable rigid molecule the intra-molecular forces only play a minor role and the herringbone angle is mostly affected by inter-molecular forces. The cause of the mismatch is most likely due to the distribution of the partial charge of the molecule. However the force constants and partial charges for pentacene are perfectly covered (no penalties) by *CGenFF*. In order to improve the result one could start using a different partitioning schema for the charges but such an optimization would be necessary for each molecule and without an reference solution it is quite hard to estimate the success of such an approach. A rather different approach is to avoid completely the problem of charge partition. This can be achieved by using ab initio DFT methods. Again, the best structure based on the X-ray data was used as an input for the geometry optimization with VASP. The optimization was able to improve the agreement in the structure factors and most important, the herringbone angle θ of 53.7° matches the experimental data.

Although pentacene is a rigid molecule the fused benzene rings are not absolutely flat. Due to thermal motion, dihedral angles up to 1° are observed in the benzene units. Those deviations of the ideal flat structure are reduced during the DFT optimization to about 0.1° . Nevertheless the non planar geometry introduces a small uncertainty in the determination of the herringbone angle.

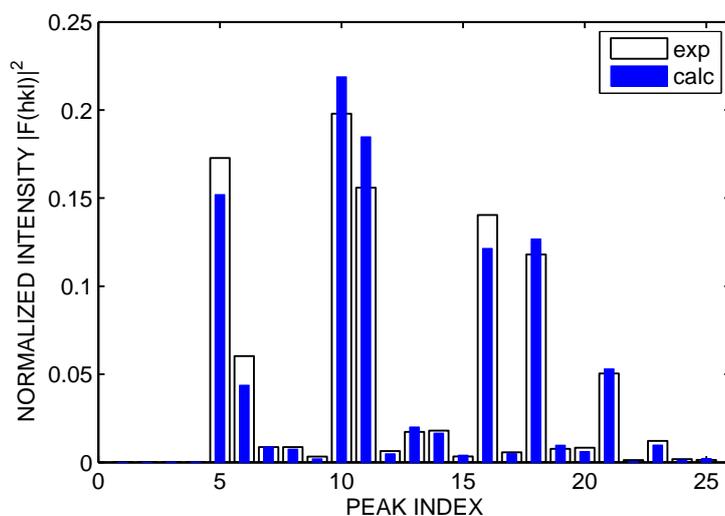


Figure 4.5.: Comparison of first MD run (calc) with random initial packings and X-ray (exp) structure factors $|F(hkl)|^2$.

4. Results and Discussion

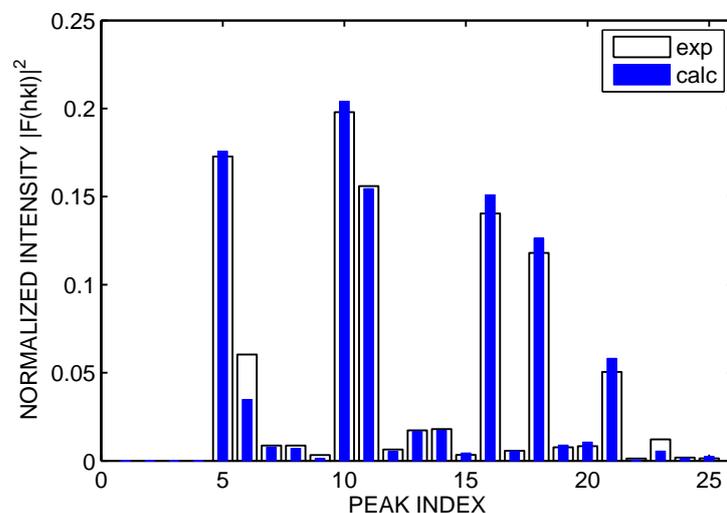


Figure 4.6.: Comparison of DFT optimized packing (calc) and measured (exp) structure factors $|F(hkl)|^2$ for pentacene.

Table 4.2.: Orientation of pentacene molecules in the bulk and thin film phase compared to the MD simulation results obtained with random initial packings (MD), the lowest energy packing (energy), randomized herringbone packing (MD HB) and DFT optimized results. χ measures the angle between of the long molecular axis (LMA) and the normal vector on the *ab*-plane of the unit cell. δ is the angle between the LMA of the molecules.

	θ (deg)	χ_1 (deg)	χ_2 (deg)	δ (deg)
bulk [62]	52.5	22.4	20.5	2.2
Nabok [63]	54.96	3.01	3.00	0.02
Moser [11]	54.13	4.79	4.83	0.14
energy	50.5	6.1	5.3	0.93
MD	50.5	5.3	6.2	0.92
MD HB	51.6	3.99	4.27	0.85
DFT	53.7	4.38	3.87	0.52

Table 4.3.: CHARMM CGenFF parameters for pentacene obtained by analogy. All penalty scores are zero. Atom labels see Figure 4.1.

atom	type	charge	atom	type	charge
C1	CG2R61	-0.116	C19	CG2R61	0.002
C2	CG2R61	-0.116	C20	CG2R61	-0.117
C3	CG2R61	-0.116	C21	CG2R61	0.003
C4	CG2R61	0.003	C22	CG2R61	-0.116
C5	CG2R61	-0.117	H1	HGR61	0.115
C6	CG2R61	0.002	H2	HGR61	0.115
C7	CG2R61	-0.117	H3	HGR61	0.115
C8	CG2R61	0.002	H4	HGR61	0.115
C9	CG2R61	-0.117	H5	HGR61	0.115
C10	CG2R61	0.003	H6	HGR61	0.115
C11	CG2R61	-0.116	H7	HGR61	0.115
C12	CG2R61	-0.116	H8	HGR61	0.115
C13	CG2R61	-0.116	H9	HGR61	0.115
C14	CG2R61	-0.116	H10	HGR61	0.115
C15	CG2R61	0.003	H11	HGR61	0.115
C16	CG2R61	-0.117	H12	HGR61	0.115
C17	CG2R61	0.002	H13	HGR61	0.115
C18	CG2R61	-0.117	H14	HGR61	0.115

4.2. Ternaphthalene

4.2.1. Isolated molecule

Ternaphthalene consists of three connected rigid naphthalene units. However the whole molecule itself is flexible. The equilibrium geometry in vacuum for an isolated molecule shows a distinct rotation of the center naphthalene unit as shown in Figure 4.8. The calculation for the MD simulation was performed in a sufficiently large box to avoid interaction between the adjacent molecules due to periodic boundary conditions. The exact rotation angle of the center naphthalene unit depends clearly on the applied method. The agreement of the MD result and MP2 was exceeded since the *CGenFF* force field is based on MP2 calculations. The difference between B3LYP and MP2 results is a well known fact and illustrates that especially dihedral angles are subject to large systematic uncertainties.

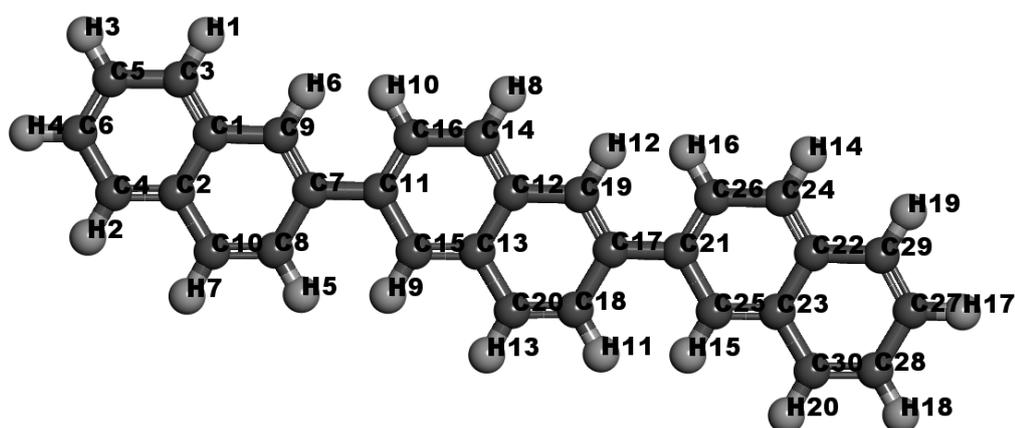


Figure 4.7.: Ternaphthalene molecule with atom labels as used for MD simulations.

4.2.2. Parametrization

Pentacene and naphthalene, both are composed of laterally fused benzene rings, are in the broader sense chemically very similar. Therefore it is not surprising that the parametrization by analogy found the same atom types for the naphthalene units. Only atoms which are involved in the bond between the naphthalene units (e.g. C7 or C11) were assigned a different atom type.

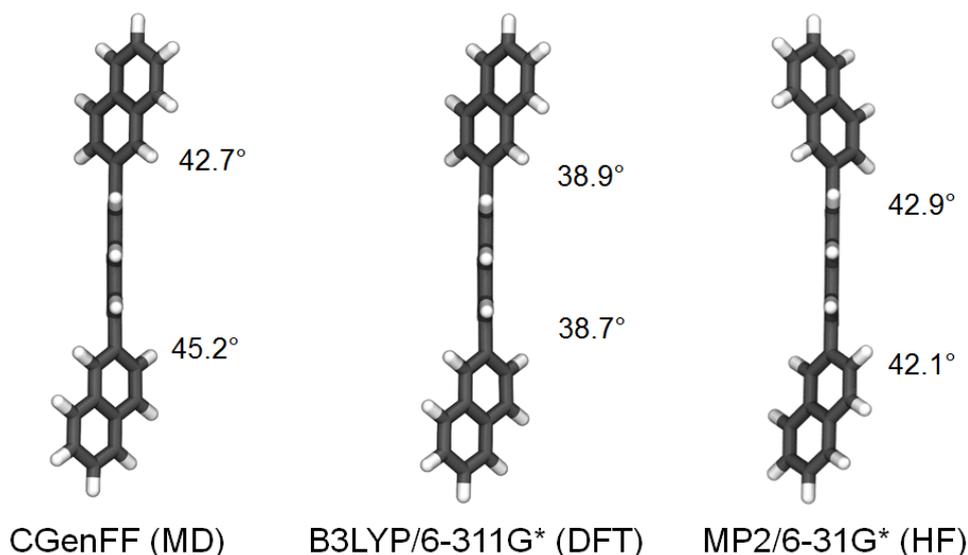


Figure 4.8.: Isolated NNN molecule in vacuum. Comparison of different geometry optimizations based on molecular dynamics (left), DFT (center) and MP2 (right) calculations.

4.2.3. Energy

A first evaluation of the results is solely based on energy. The energy of the first 100 MD results is shown in Figure 4.9. Since a possible solution should have a low energy, only structures up to the first energy step are reasonable. The lowest energy results features a herringbone arrangement of the molecules and the naphthalene units of each molecule are in the same plane. Such a planarization of a twisted molecules in a periodic arrangement is due to the minimization of the total energy. Although the energy of a single flat molecule is higher than the twisted geometry, the total energy of the flat herringbone packing is lower (see Figure 4.11).

The first step in energy is connected to an increase of the orientation angle χ . Other packing motifs with higher energies are illustrated in Figure 4.11a and 4.11b. Parallel packing and even a mix of herringbone and parallel stacking can be excluded as an likely structure due to the high energies. Hence the flexible bond between the naphthalene units allows for more rotational freedom the energy difference between different packing motifs is not as pronounced as with pentacene.

4. Results and Discussion

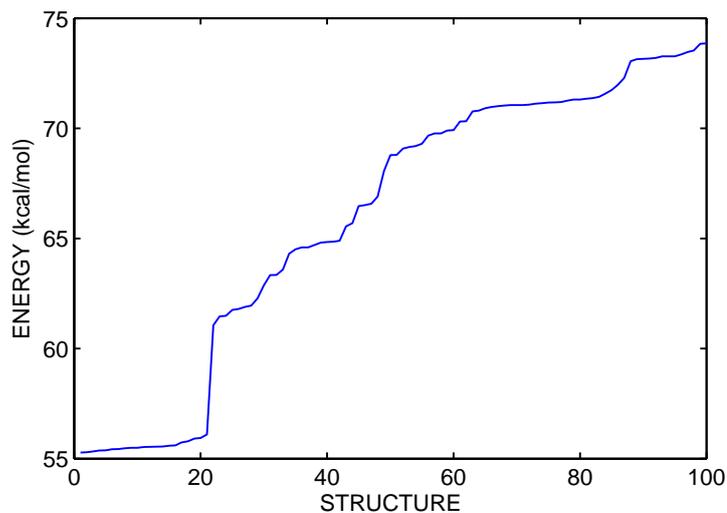


Figure 4.9.: First 100 MD results for ternaphthalene in order of increasing energy.

4.2.4. Structure factors

The GID measurement of NNN and subsequent intensity extraction for the structure factors was performed by Alexander Pichler during his master thesis [12]. The thin film phase was solved successfully by means of rigid body refinement and is used for verification of the MD results.

The best results of the MD simulation were obtained by comparison with the experimental structure factors. The energy of this herringbone packing is only slightly higher than the lowest energy results. Considering thermal motion, the difference between the lowest energy and the best X-ray results is negligible ($\Delta E = 0.26$ kcal/mol). The orientation angles χ match the solved thin film phase, however the herringbone angle θ is again 4.5° too low. Nearly the same mismatch was observed for pentacene. This systematic underestimation of the herringbone angle seems to be a characteristic of *CGenFF*. Thus a subsequent DFT optimization of the packing was necessary to improve the herringbone angle. The overall agreement with the structure factors was increased and the herringbone angle matches the experimental value.

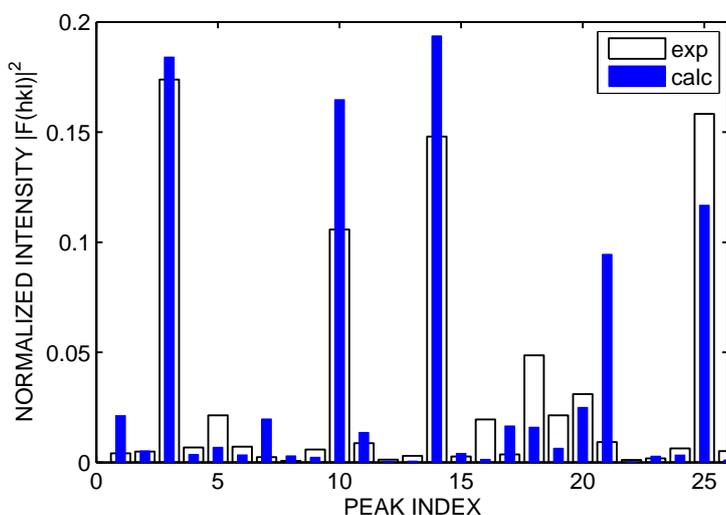


Figure 4.10.: Comparison of DFT optimized packing (calc) and X-ray intensity (exp) structure factors $|F(hkl)|^2$.

Table 4.4.: CHARMM CGenFF parameters for ternaphthalene obtained by analogy. All penalty scores are zero. Atom labels see Figure 4.7.

name	type	charge	name	type	charge	name	type	charge
C1	CG2R61	0.006	C18	CG2R61	-0.119	H5	HGR61	0.115
C2	CG2R61	0.004	C19	CG2R61	-0.122	H6	HGR61	0.115
C3	CG2R61	-0.116	C20	CG2R61	-0.115	H7	HGR61	0.115
C4	CG2R61	-0.116	C21	CG2R67	0.005	H8	HGR61	0.115
C5	CG2R61	-0.116	C22	CG2R61	0.004	H9	HGR61	0.115
C6	CG2R61	-0.116	C23	CG2R61	0.006	H10	HGR61	0.115
C7	CG2R67	0.005	C24	CG2R61	-0.115	H11	HGR61	0.115
C8	CG2R61	-0.119	C25	CG2R61	-0.122	H12	HGR61	0.115
C9	CG2R61	-0.122	C26	CG2R61	-0.119	H13	HGR61	0.115
C10	CG2R61	-0.115	C27	CG2R61	-0.116	H14	HGR61	0.115
C11	CG2R67	0.005	C28	CG2R61	-0.116	H15	HGR61	0.115
C12	CG2R61	0.006	C29	CG2R61	-0.116	H16	HGR61	0.115
C13	CG2R61	0.006	C30	CG2R61	-0.116	H17	HGR61	0.115
C14	CG2R61	-0.115	H1	HGR61	0.115	H18	HGR61	0.115
C15	CG2R61	-0.122	H2	HGR61	0.115	H19	HGR61	0.115
C16	CG2R61	-0.119	H3	HGR61	0.115	H20	HGR61	0.115
C17	CG2R67	0.005	H4	HGR61	0.115			

4. Results and Discussion

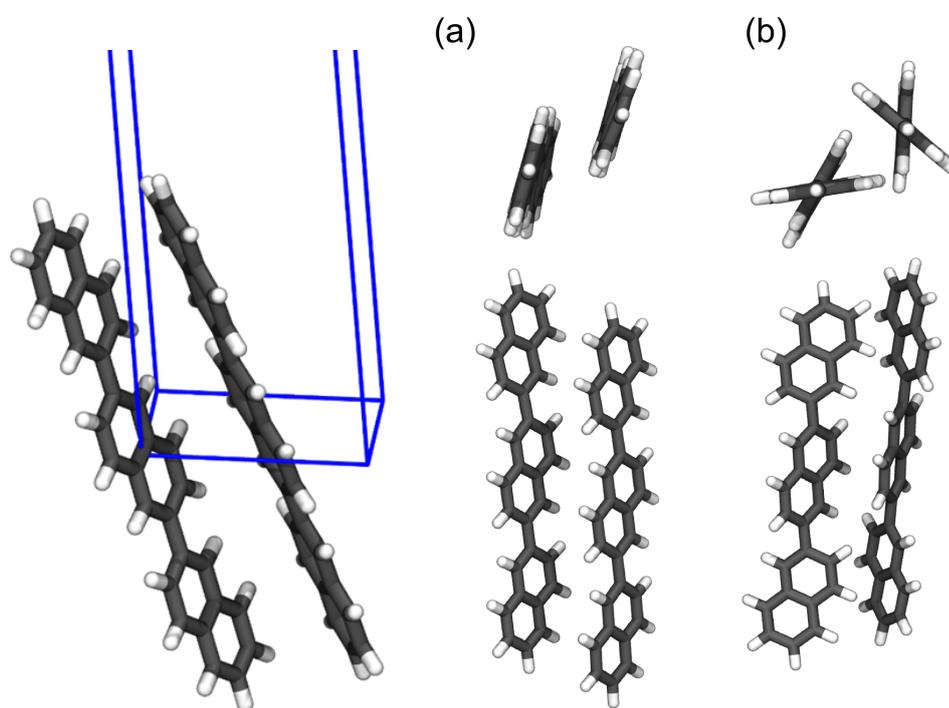


Figure 4.11.: Left: Result of best MD simulation according to X-ray data. The planarization of the twisted isolated molecules in a crystallographic unit cells is reproduced by the simulation. Right: Molecular packings with higher energy. a) Parallel stacking motif. b) Mix of parallel and herringbone packing for single naphthalene units.

Table 4.5.: Orientation of NNN molecules for rigid body refinement (rbr), lowest energy, MD and DFT results. χ measures the angle between of the long molecular axis (LMA) and the normal vector on the ab-plane of the unit cell. δ is the angle between the LMA of the molecules.

	θ (deg)	χ_1 (deg)	χ_2 (deg)	δ (deg)
rbr	52.4	22.9	22.9	2.0
energy	48.0	22.2	22.1	0.5
MD	47.9	22.1	21.9	0.2
DFT	52.3	21.9	21.9	0.4

4.3. Dioctyl-Terthiophene

Dioctyl-Terthiophene (DOTT) consists of a backbone containing three thiophene rings (3T) connected by a carbon-carbon bond and octyl chains at each end of the backbone. DOTT is used typically for the active layer of organic semiconducting devices [64]. The octyl chains are able to rotate quite easily, which increases the overall flexibility of the molecule and different orientation are therefore observed.

DOTT was investigated extensively by Christoph Lercher during his master's thesis [60]. Among other things, he was able to determine the unit cell parameters for the b- and s-phase. Those unit cell parameters were used for MD simulations. However, due to the weak scattered intensities of the GID experiment, it was not possible to use the measured intensity data for further structure refinement. The evaluation of the MD simulations is therefore based solely on the energy of the simulation results.

4.3.1. Isolated molecule

In order to investigate the important thiophene inter-ring dihedral and the octyl-thiophene dihedral potential, the molecule was split in smaller units. Figure 4.13 illustrates a bithiophene unit (2T) which was used to study the inter-ring dihedral ϕ . The octyl-thiophene (OT) dihedral ψ was investigated in a system as shown in Figure 4.15. The popular quantum chemistry methods, MP2 and DFT on basis of the B3LYP hybrid functional were used and the influence of different basis sets were studied and compared. The calculations were carried out in *Gaussian09*. For 2T, a step size of 5° and 10° was chosen for MP2 and B3LYP, respectively. For OT, a step size of 5° was used for all calculations. At each step a geometry optimization was performed. At detailed description of the procedure and corresponding input files are explained in Section 3.6.1.

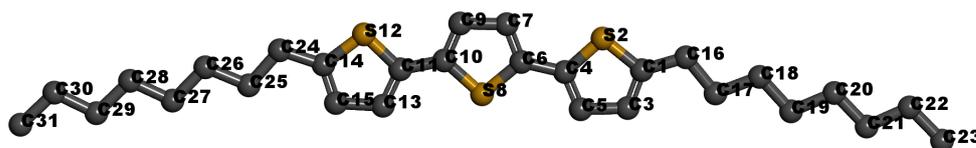


Figure 4.12.: Dioctyl-Terthiophene with atom labels. For the sake of clarity, hydrogen atoms are not shown. H1 to H6 are attached to the thiophene rings, H7 to H23 and H24 to H40 are part of the octyl chains.

4. Results and Discussion

Bithiophene (2T)

The results for 2T are summarized in Figure 4.14. The dihedral potential features two local minima, the position of which depends on the applied method. A difference of approximately 15° for the minima positions was observed for the MP2 and B3LYP calculation. Also the height of the rotation barriers is influenced by the applied method. The energy difference for the anti-syn barrier, can get up to 1.2 kcal/mol, depending on the choice of basis set. Nevertheless, the position of the anti-syn barrier at 90° seems to be independent of the level of theory. The potential around $\phi = 180^\circ$ is very shallow and the anti-anti barrier is in the case of MP2 0.4 kcal/mol and even lower for B3LYP (0.1 kcal/mol). The barrier seems to vanish completely for the B3LYP/aug-cc-pVDZ calculation. However using a even larger aug-cc-pVTZ basis set did bring the rotation barrier back but the computational effort increased roughly by a factor of 12. Nevertheless, the influence of the basis set is rather small and only minor deviations are observed. The use of a mixed basis set consisting of aug-cc-pVDZ for sulfur and cc-pVDZ for all other atoms produced the lowest rotation barrier. Usually such a mixed basis set is used to save time since diffuse function increase drastically the computational effort. All in all, the calculations are in good agreement with the results as reported in [65, 66].

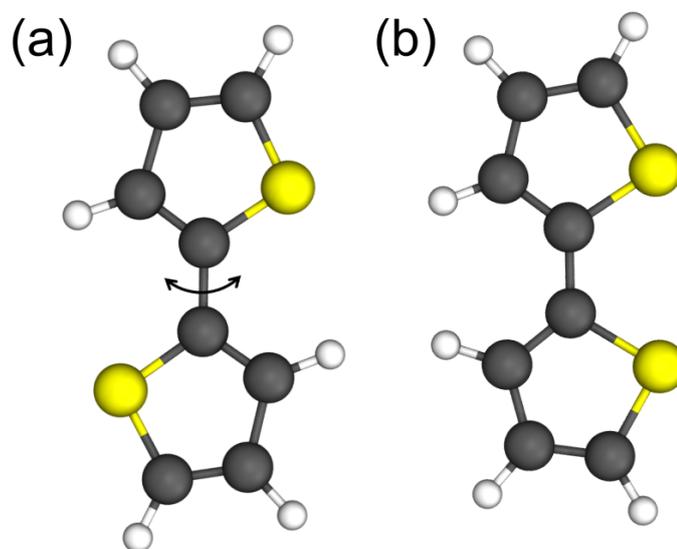


Figure 4.13.: Bithiophene (2T) model as used for the study of the inter-ring dihedral potential ϕ . a) 2T in anti conformation with $\phi = 180^\circ$. b) 2T in syn conformation with $\phi = 0^\circ$.

Additionally, unconstrained geometry optimization at various levels of theory were performed and are summarized in Table 4.6. In order to find a global minimum the flat anti geometry was rotated by 10° . Without those modifications the geometry optimization yielded a perfectly flat ($\phi = 180.0^\circ$) molecule, which does not correspond to the lowest energy conformation with dihedral angles $\phi_{of} \approx 140$ to 160° . The variation of the optimized dihedrals may seem quite large, however the global minimum is relatively flat. Merely a change of $0.05 \text{ kcal/mol} \approx 2 \text{ meV}$ in energy is enough to shift the dihedral angle roughly by 15° but there is still a significant difference between MP2 and B3LYP.

Table 4.6.: Unconstrained geometry optimization at various levels of theory for bithiophene. Dihedral angle ϕ (in degree) corresponds to the global energy minimum as illustrated in Figure 4.14. Bottom half lists the relative computing time based on the time for MP2/cc-pVTZ of 18.00 h.

	6-31G*	6-311G*	cc-pVDZ	cc-pVTZ	aug-cc-pVDZ
MP2	141.5	136.7	146	150.8	148.1
B3LYP	157.5	149.4	163.3	157.9	158.4
MP2	0.044	0.092	0.097	1.000	0.529
B3LYP	0.019	0.023	0.032	0.278	0.194

Octyl-Thiophene (OT)

In Figure 4.16 the PES for the octyl-thiophene dihedral ψ is illustrated. In contrast to the inter-ring potential ϕ , the level of theory has a much smaller impact on the chain-ring potential ψ . The global minimum is located at $\psi \approx 70^\circ$ and a local yet very shallow minimum is at $\psi = 180^\circ$, which allows the octyl chain to move freely in a range of $\pm 30^\circ$. In contrast to ϕ potential of 2T, MP2 produces now larger rotation barriers and the B3LYP potential especially around $\psi = 180^\circ$ is extremely flat and features only a small barrier of $\Delta E \approx 0.15 \text{ kcal/mol}$. The barrier at $\psi = 0^\circ$ has a height of 2 kcal/mol for MP2 and 1.5 kcal/mol for B3LYP. In order to escape to local minimum during a geometry optimization it was necessary to set $\psi = 120^\circ$. The results for the unconstrained geometry optimization is given in Table 4.7.

DOTT geometry optimization

The knowledge gained by studying the 2T and OT dihedrals was applied in a geometry optimization of the terthiophene (3T) backbone and the whole DOTT molecule. Starting with a flat geometry for 3T yielded a flat optimized 3T molecule with all inter-ring

4. Results and Discussion

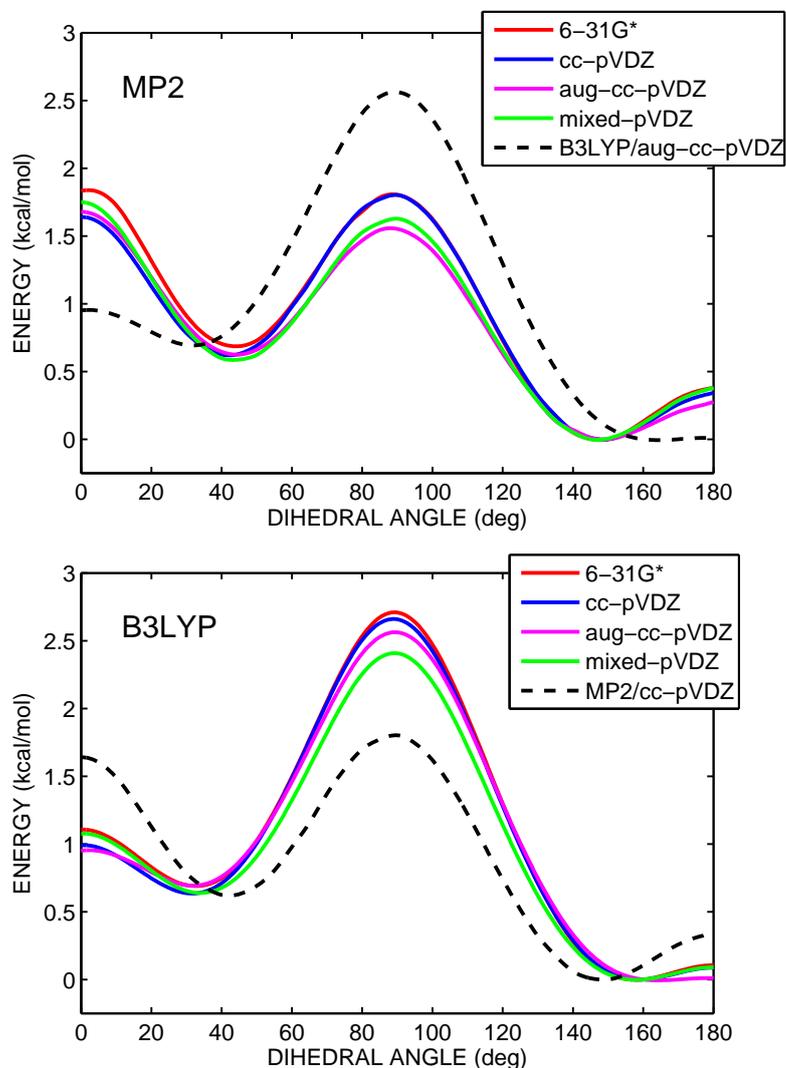


Figure 4.14.: Inter-ring dihedral potential for 2T calculated with MP2 and B3LYP for different basis sets. The green mixed basis set consists of aug-cc-pVDZ for sulfur and cc-pVDZ for all other atoms. The black dashed line illustrates the significant difference in rotation barrier heights and minima positions between MP2 and B3LYP. During the calculations, a step size of 5° and 10° was used for MP2 and B3LYP, respectively. Cubic spline interpolation was used to obtain a continuous potential.

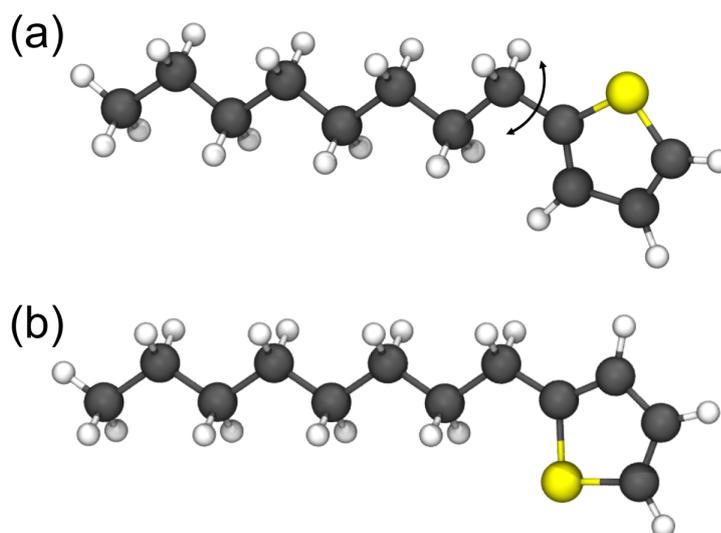


Figure 4.15.: Octyl-Thiophene (OT) as used for the investigation of the chain-ring dihedral ψ . a) OT in anti conformation with $\psi = 180^\circ$. b) OT in syn conformation with $\psi = 0^\circ$.

Table 4.7.: Unconstrained geometry optimization at various levels of theory for octhyl-thiophene (OT). Dihedral angle ψ (in degree) corresponds to the global energy minimum as illustrated in Figure 4.16. Bottom half lists the relative computing time based on the time for MP2/aug-cc-pVDZ of 38.47 h.

	6-31G*	6-311G*	cc-pVDZ	aug-cc-pVDZ
MP2	70.9	74.4	73.1	72.7
B3LYP	68.5	69.6	69.9	69.1
MP2	0.078	0.165	0.153	1.000
B3LYP	0.025	0.041	0.046	0.758

dihedral equal to 180° . A geometry optimization with an initial conformation with inter-ring dihedral set to 175° yielded a twisted structure with both inter-ring dihedral at 153.4° . The calculations were carried out the the B3LYP/6-311G* level of theory and are in good agreement with the 2T relaxed geometry, which yielded a dihedral angle of 149.4° . The energy difference between the flat and the twisted (lower energy) conformation is 0.35 kcal/mol.

DOTT geometry optimizations were performed at the B3LYP/6-31G* level of theory. The octyl-thiophene (ψ) and inter-ring (ϕ) dihedral were set near their equilibrium values in order the avoid a flat optimized structure. The optimized twisted structure is 1.29 kcal/mol lower in energy than the flat conformation and the dihedral angles ψ

4. Results and Discussion

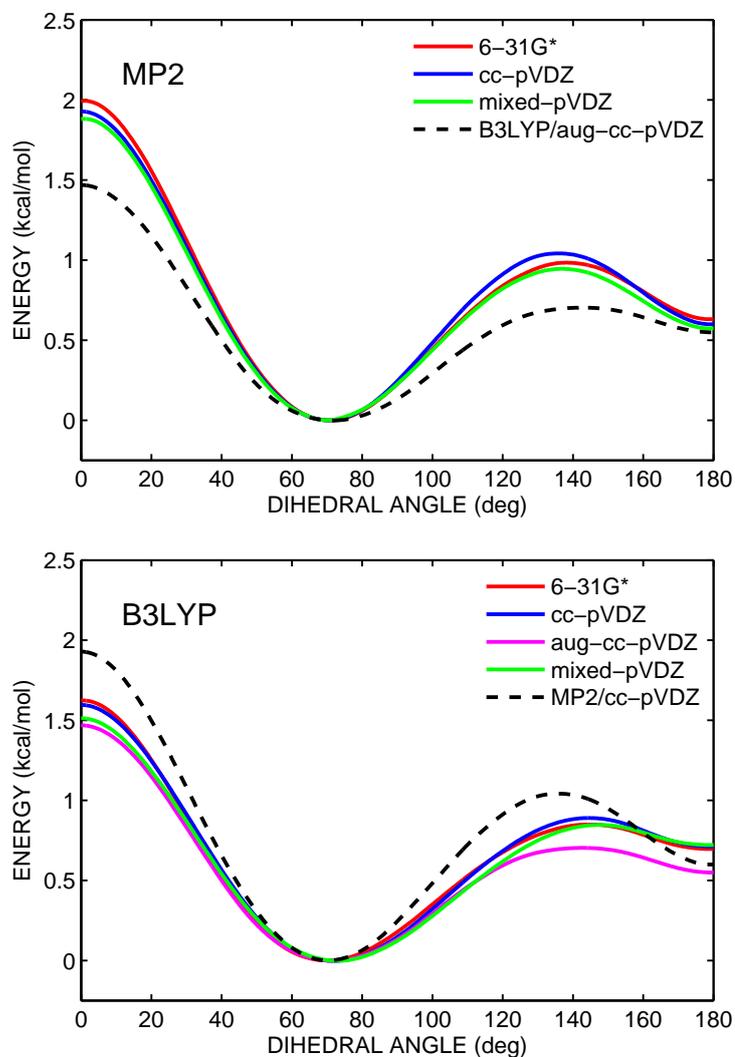


Figure 4.16.: Octyl-thiophene dihedral potential calculated with MP2 and B3LYP for different basis sets. For details see caption of Figure 4.14. A step size of 5° was used for MP2 and B3LYP calculations. It was not possible to carry out the MP2/aug-cc-pVDZ in a reasonable time frame.

are 68.4 and 72.3° which matches the results of the isolated octyl-thiophene system. The inter-ring dihedrals ϕ are 170.8 and 162.5°, which differs from the value of the 2T optimization (157.5°). However the dihedral potential is quite flat in this region and could be very sensitive to the changes introduced by the octyl chains. This might explain the presence of a flatter backbone.

4.3.2. Parametrization

The results for the DOTT parameters are summarized in Table 4.8. Unfortunately not all parameters are available in *CGenFF*. The largest penalty scores are as expected for the link between the backbone and the side chains and the between the single thiophene units in the backbone. For so called hairy rods molecules like DOTT, the primary focus lies on the dihedral parameters for the link between the side chains and backbone. The backbone of DOTT is rather rigid, therefore a flat backbone is expected and is also indicated by single crystal results. In order to check the force field parameters the data presented in Section 4.3.1 was utilized.

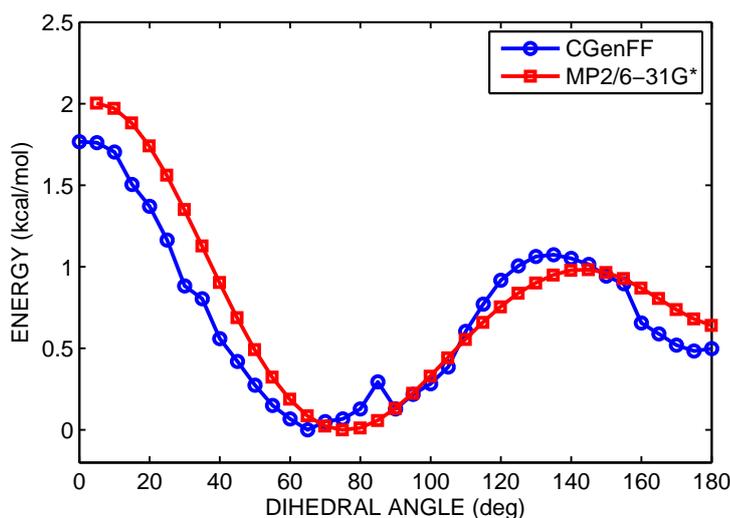


Figure 4.17.: Comparison of the PES recorded with LAMMPS (*CGenFF*) and MP2 for the octyl-thiophene dihedral potential (see Figure 4.15).

4. Results and Discussion

4.3.3. Energy

In Figure 4.18 and 4.19 the results for the s- and b-phase are presented. As expected both phases differ in their herringbone angle. Values for the herringbone angles are presented in the next section. It is interesting to note, that in both cases, a distinct increase in energy is linked to a certain orientation of the side chains. This orientation matches the dihedral potential minimum of approximately 70° (see Figure 4.16). Hence two molecules are present in an unit cell, four important dihedral angles (two per molecule) are shown in the figures. In order to increase the packing density, the side chains rotate to $\pm 70^\circ$. For higher energies it seems that the side chains are in an energetically less favorable orientation.

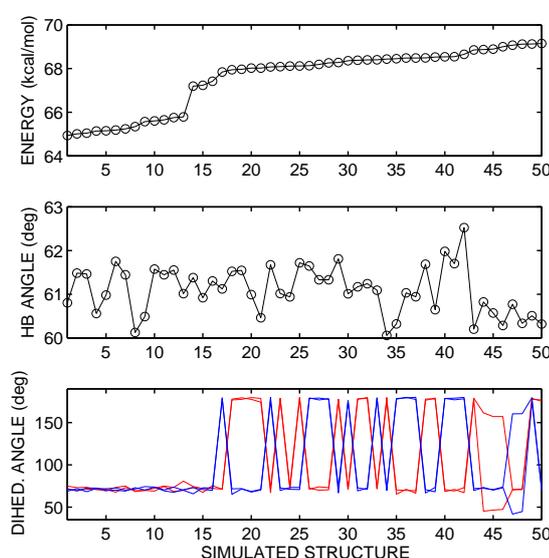


Figure 4.18.: First 50 DOTT s-phase results in increasing order of energy. The herringbone angle of the backbone and the dihedral angles ϕ : S2-C1-C16-C17 and ψ : S12-C14-C24-C25 of the side chains are indicated in red and blue, respectively. One can see clearly the influence of the side chain orientation on the energy.

4.3.4. Average structures

Due to the thermal motion of the terminal chains it was necessary to average the position of the atoms over 1000 time frames to obtain stable results. For this purpose the lowest energy packing was used as an input for another MD simulations with the correct

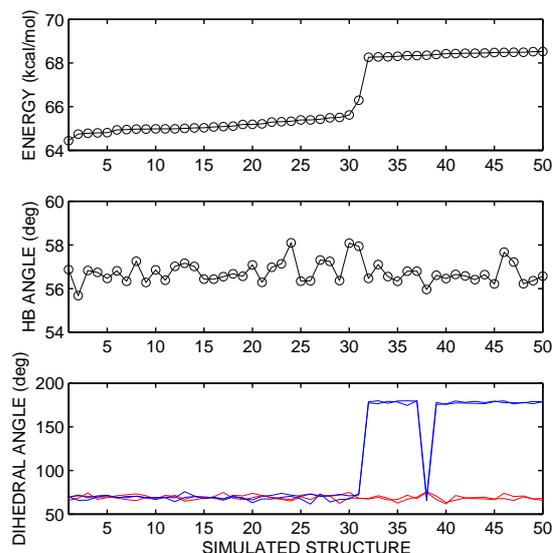


Figure 4.19.: First 50 DOTT b-phase results in increasing order of energy. For a detailed description see caption of Figure 4.18.

experimental unit cell. The simulation was executed at a temperature of 20 K using a Berendsen thermostat. Subsequently the herringbone angle θ and the side chain dihedral angles ϕ and ψ were calculated using the average structure. An illustration of the b- and s-phase is provided in Figure 4.20. The orientation of the long molecular axis (LMA) of the backbone is for both phases normal to the ab -plane of the unit cell.

Furthermore, the influence of temperature on mean values and standard deviations was investigated. For this purpose the same procedure as for averaging was performed at various temperatures (see Figure 4.22). Up to 350 K the average dihedral angle increased by approximately 1° , in contrast the standard deviation σ_D increased fourfold. Nevertheless the average values are not strongly affected by the temperature but uncertainties in the range of a few degrees should be expected.

4.3.5. DFT optimization

Comparing the carbon-carbon bond length between the thiophene rings of the single crystal solution and DFT optimized structures of 2T showed a significant difference. The bond length according to B3LYP/cc-pVTZ is 1.447 \AA . The MD simulations showed a bond lengths of 1.37 \AA . This could be caused by the suboptimal force field parameters. Hence

4. Results and Discussion

a complete DFT geometry optimization was carried out as described in Section 3.5. After the optimization the intra-molecular distances matched the DFT results for 2T and single crystal data. The optimization had also an influence on the inter-molecular arrangement. As seen with previous systems, the herringbone angle decreased roughly about 5° for both phases. The dihedral angle between the octyl chains and backbone was basically unaffected and is still approximately 70° .

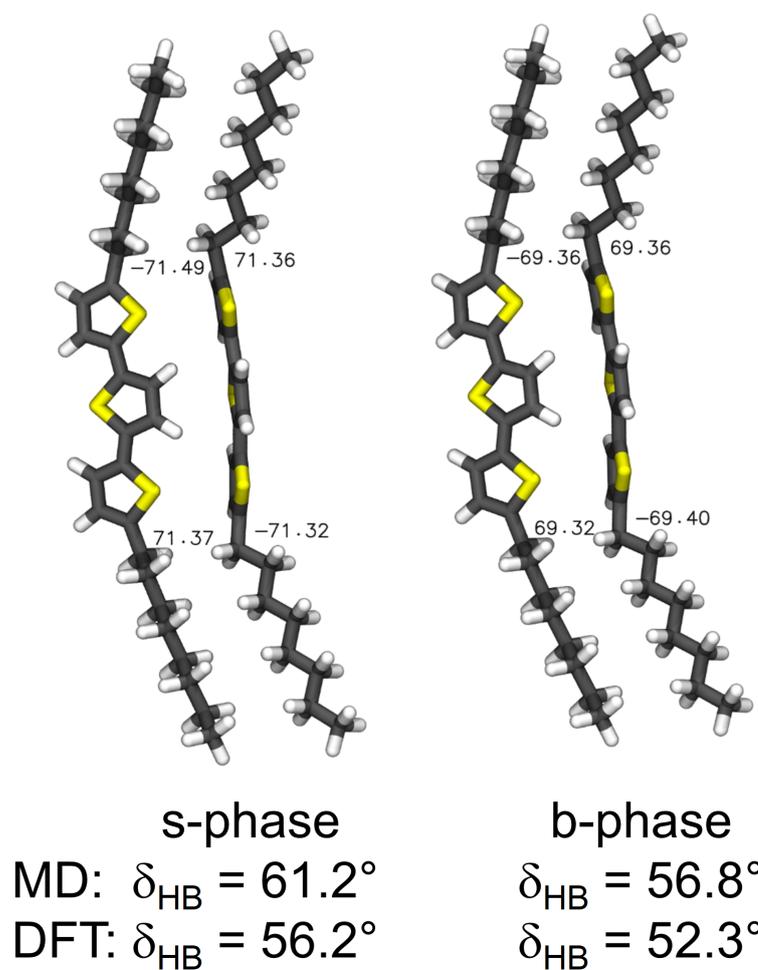


Figure 4.20.: DOTT structures obtained by averaging over 1000 frames and DFT optimization. Herringbone angle δ_{HB} and dihedral angles ϕ , ψ are indicated.

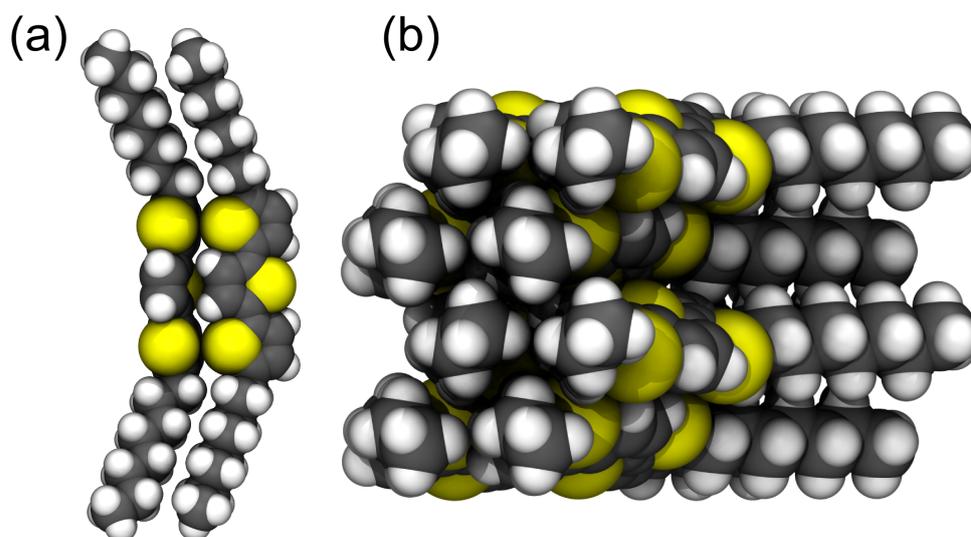


Figure 4.21.: DOTT b-phase. a) Single unit cell with two DOTT molecules. b) Multiple unit cells which demonstrate the close-packing of the octyl side chains.

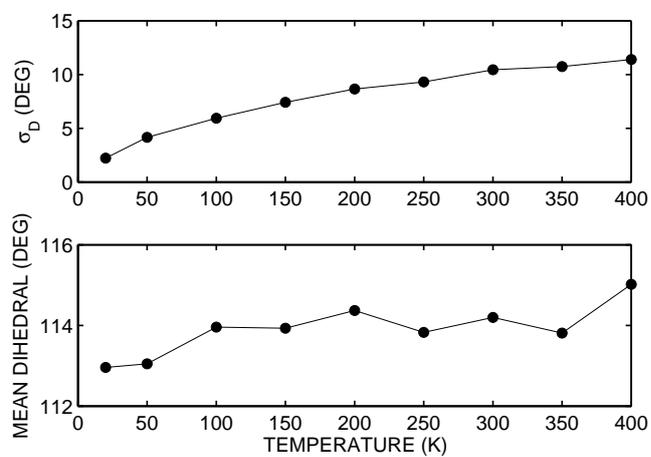


Figure 4.22.: Influence of temperature on the mean value and standard deviation σ_D of dihedral angle ϕ_1 for the DOTT s-phase.

4. Results and Discussion

Table 4.8.: CHARMM CGenFF parameters for dioctyl-terthiophene obtained by analogy. Atom labels see Figure 4.12.

name	type	charge	penalty	name	type	charge	penalty
C1	CG2R51	-0.011	73	H6	HGR51	0.193	0
S2	SG2R50	-0.069	67	H7	HGA2	0.09	3
C3	CG2R51	-0.253	10	H8	HGA2	0.09	3
C4	CG2R51	0.092	3	H9	HGA2	0.09	0
C5	CG2R51	-0.249	2	H10	HGA2	0.09	0
C6	CG2R51	0.091	0	H11	HGA2	0.09	0
C7	CG2R51	-0.249	0	H12	HGA2	0.09	0
S8	SG2R50	-0.098	0	H13	HGA2	0.09	0
C9	CG2R51	-0.249	0	H14	HGA2	0.09	0
C10	CG2R51	0.091	0	H15	HGA2	0.09	0
C11	CG2R51	0.092	3	H16	HGA2	0.09	0
S12	SG2R50	-0.069	67	H17	HGA2	0.09	0
C13	CG2R51	-0.249	2	H18	HGA2	0.09	0
C14	CG2R51	-0.011	73	H19	HGA2	0.09	0
C15	CG2R51	-0.253	10	H20	HGA2	0.09	0
C16	CG321	-0.088	62	H21	HGA3	0.09	0
C17	CG321	-0.181	5	H22	HGA3	0.09	0
C18	CG321	-0.181	3	H23	HGA3	0.09	0
C19	CG321	-0.18	0	H24	HGA2	0.09	3
C20	CG321	-0.18	0	H25	HGA2	0.09	3
C21	CG321	-0.178	0	H26	HGA2	0.09	0
C22	CG321	-0.183	0	H27	HGA2	0.09	0
C23	CG331	-0.269	0	H28	HGA2	0.09	0
C24	CG321	-0.088	62	H29	HGA2	0.09	0
C25	CG321	-0.181	5	H30	HGA2	0.09	0
C26	CG321	-0.181	3	H31	HGA2	0.09	0
C27	CG321	-0.18	0	H32	HGA2	0.09	0
C28	CG321	-0.18	0	H33	HGA2	0.09	0
C29	CG321	-0.178	0	H34	HGA2	0.09	0
C30	CG321	-0.183	0	H35	HGA2	0.09	0
C31	CG331	-0.269	0	H36	HGA2	0.09	0
H1	HGR51	0.193	0	H37	HGA2	0.09	0
H2	HGR51	0.207	0	H38	HGA3	0.09	0
H3	HGR51	0.207	0	H39	HGA3	0.09	0
H4	HGR51	0.207	0	H40	HGA3	0.09	0
H5	HGR51	0.207	0				

4.4. DBDCS

DBDCS is a cyano distyrylbenzene (DBS) derivative with butyloxy groups at the terminal rings (see Figure 4.23). Such conjugated materials are extensively investigated due to their optoelectronic properties. An overview about DBDCS and similar DBS derivatives is provided in [67]. DBDCS exhibits two-color fluorescence switching in response to pressure and temperature. These changes can be explained by two crystal phases, the optical green G-phase and blue B-phase. The structure of the G-phase was solved by means of single crystal diffraction and is reported in [68]. For the B-phase only thin film samples are available. The samples have been investigated by means of GID. The unit cell parameters and several structure factors were determined by Alexander Pichler during his Master's thesis [12]. The unit cell for the B-phase contains four molecules. The backbone itself is expected to be rigid and only small deviations ($\approx 5^\circ$) of a planar conformation are indicated by the G-phase. In contrast, the terminal chains are able to change their orientation quite easily. Therefore the rigid body refinement was not able to produce reasonable results.

4.4.1. Isolated molecule

An isolated DBDCS molecule features a twisted backbone due to the nitrile groups. Results of DFT calculations are reported in [68] and shown in Figure 4.23. The planarization of the single crystal conformation is explained by hydrogen bonds between the nitrogen and oxygen atoms and hydrogen atoms of adjacent molecules. The dihedral potentials for both torsions were calculated using MP2 and B3LYP, basically following the same procedure as demonstrated with DOTT. For the calculations the molecule was

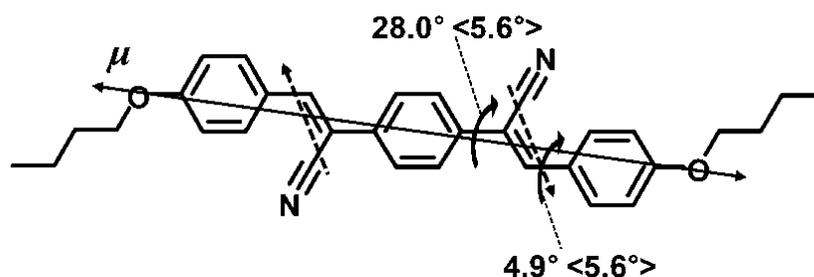


Figure 4.23.: Molecular structure of DBDCS. The local dipoles and transition dipole moment μ are indicated as dashed and solid arrows, respectively. The dihedral angles for the inner bonds are given for an isolated molecule and in single crystal conformation (in brackets). Image adapted from [68].

4. Results and Discussion

reduced to a smaller system consisting of two benzene rings and one nitrile group as illustrated in Figure 4.24. The potentials are shown in Figure 4.25. Again, a significant difference in the height of the rotations barriers between MP2 and B3LYP calculations is observed. Nevertheless, the potential minima are in good agreement with the results as presented in Figure 4.23.

4.4.2. Parametrization

CGenFF contains proper parameters for the alkyl chains and benzene rings. However, the dihedral parameters for parts which are related to the nitrile group, especially dihedral angle ϕ_1 and ϕ_2 , had penalty scores above 100 and were checked against the results of the QM calculations. The recorded data is illustrated in Figure 4.26. The potential for ϕ_1 did not fit at all and was optimized according to the procedure as described in Section 3.6.3. After the optimization, the force field was capable of reproducing the QM potential. Although the penalty score for ϕ_2 was as bad as for ϕ_1 , already the standard *CGenFF* force field was able to provide a satisfying potential.

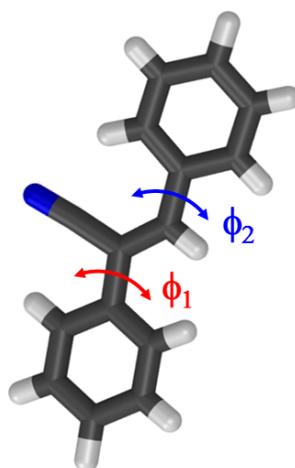


Figure 4.24.: Model system based on the backbone of DBDCS used for the investigation of the dihedral potentials involving the nitrile group.

4.4.3. Energy

The usual approach for generating random initial molecular packings for the shrinking cell MD simulations was not successful in the case of DBDCS due to the number

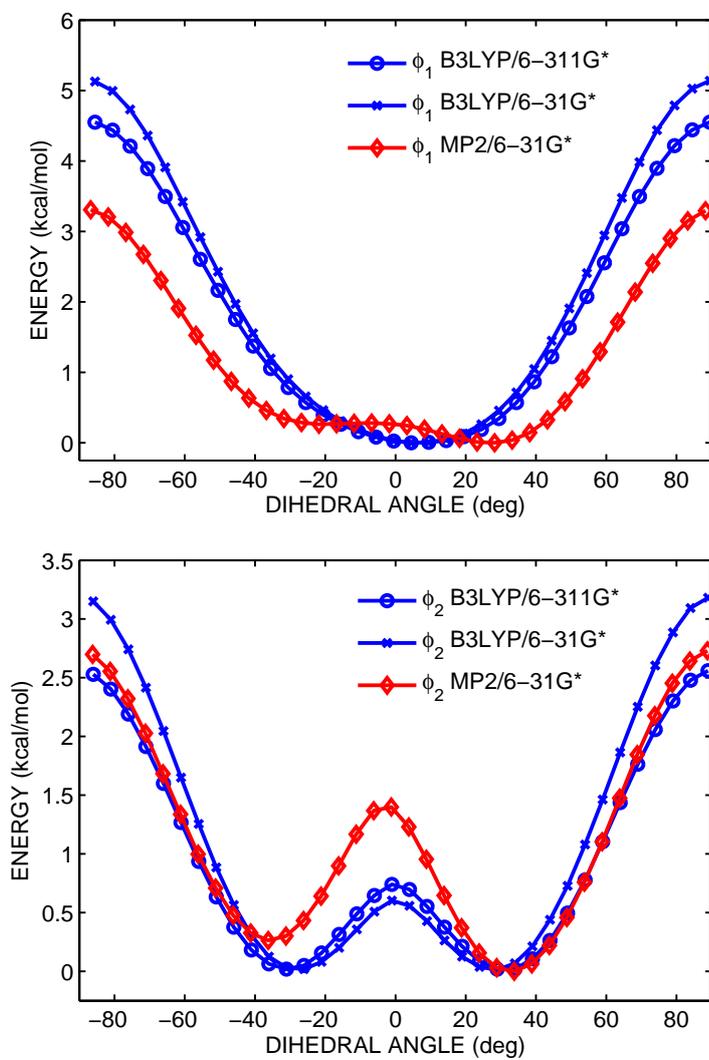


Figure 4.25.: Dihedral potentials ϕ_1 and ϕ_2 obtained by quantum chemical calculations using B3LYP and MP2.

4. Results and Discussion

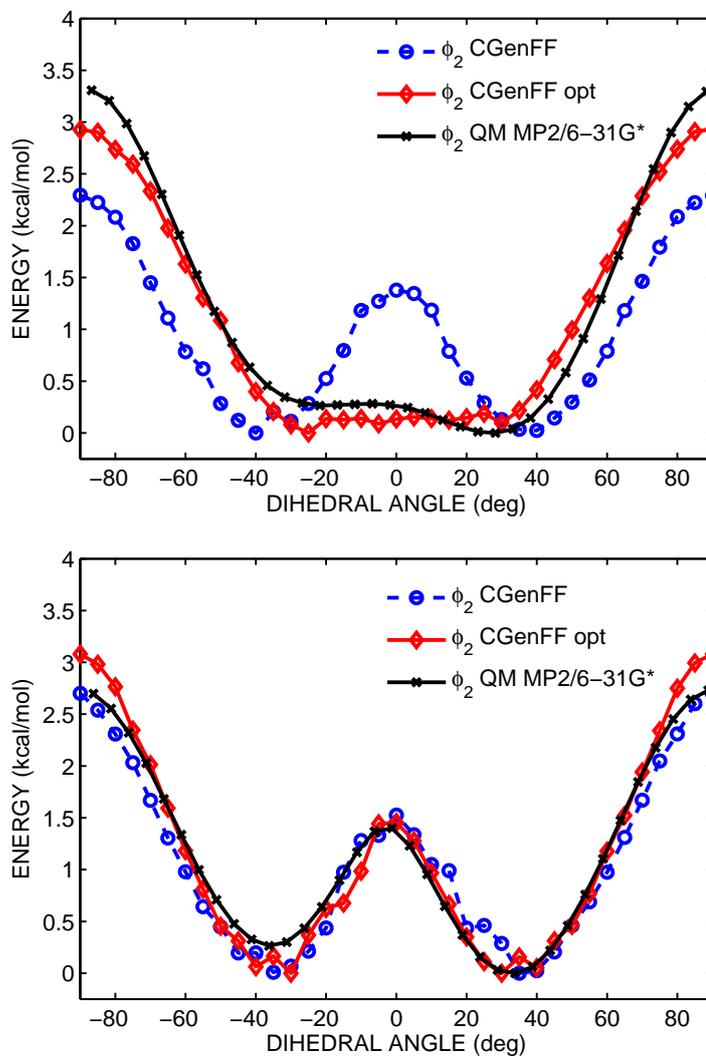


Figure 4.26.: Comparison of *CGenFF* force field and QM data (black) for dihedral potentials ϕ_1 and ϕ_2 . The original force field and optimized data is shown in blue and black, respectively.

of molecules, flexibility and unfavorable height to width ratio. Most simulations got stuck in energetically unfavorable states. Therefore various common packing motifs were chosen and modified randomly to create molecular packings for the simulation. An overview of these packings is given in Figure 4.27. The results of the shrinking cell approach is discussed in the following sections. The results will be identified by their unique serial number (1 to 5000) as used during the MD simulations. In the case of DBDCS the shrinking cell simulations and DFT optimizations were carried out by Otello M. Roscioni (University of Bologna) using a force field which was optimized for DBDCS by Luca Muccioli (University of Bologna).

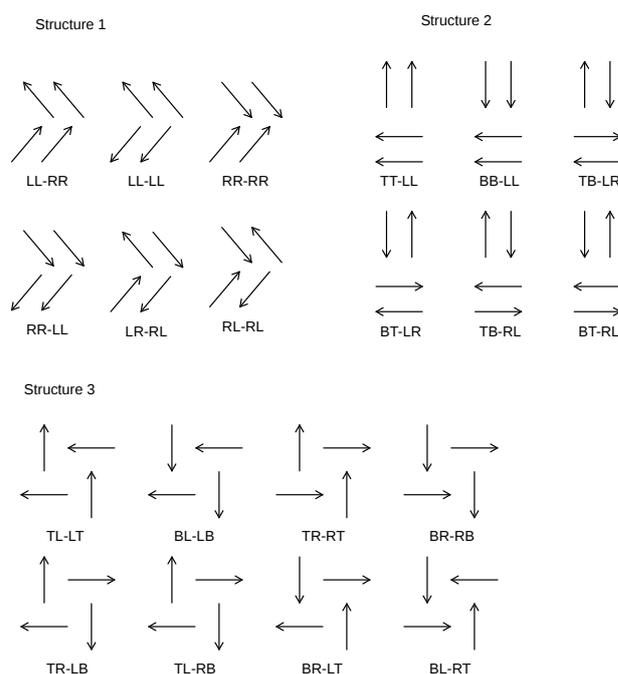


Figure 4.27.: Overview of molecular packing motifs for the DBDCS shrinking cell approach.

The lowest energy packings are shown in Figure 4.28. In contrast to the previous systems, the energy increases more or less continuous and it was not possible to immediately select a motif. However a clear trend to parallel stacking was observed which is also indicated by results as reported in [68]. Furthermore, a slip along the short molecular axis is most likely involved in the structural change from G to B-phase.

4. Results and Discussion

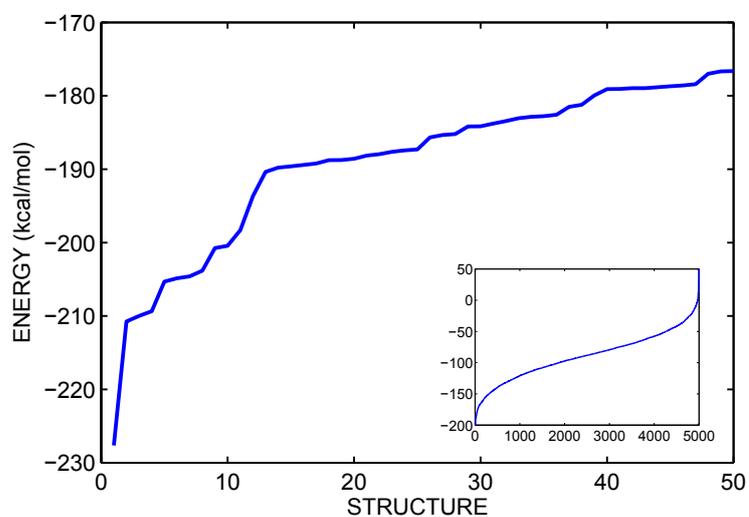


Figure 4.28.: Energies of DBDCS molecular packings in increasing order of energy.

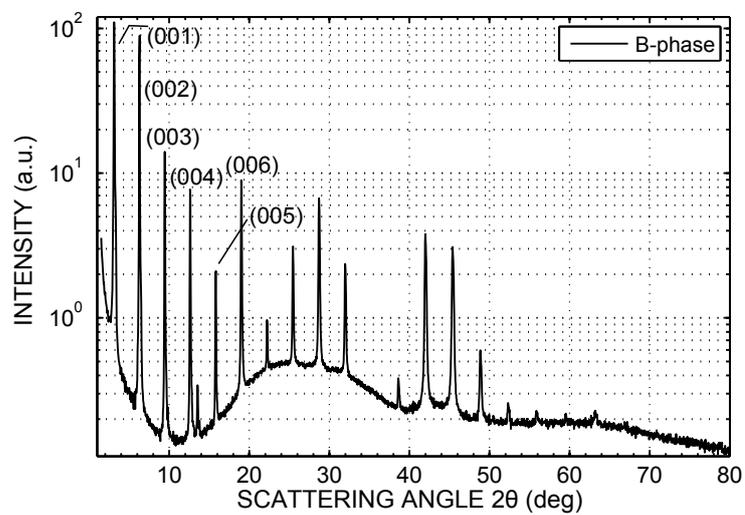


Figure 4.29.: Specular scan of a DBDCS B-phase thin film measured using Cu K_{α} radiation.

4.4.4. Structure factors

Unfortunately the analysis of the X-ray data is not as straightforward as for the other samples. For the evaluation of the simulation results two data set are available. The results of the GID measurement by Alexander Pichler at HASYLAB [12] and specular data recorded with a *Siemens D501* Bragg-Brentano diffractometer with a copper tube at 40 kV and 30 mA. Due to the large lattice constant c , it was possible to detect 15 Bragg peaks. Corrections for polarization and Lorentz factor have been applied to obtain the structure factors.

The lowest energy results (2735, -228 kcal/mol) did not match the specular nor the GID data and was therefore not further investigated. The second lowest packing (0328, -211 kcal/mol) however fits the specular data but the large peak #33 is completely missing. Another promising result is structure 0460 (-57 kcal/mol, third best GID match). Peak #33 is very strong and peak #13 is much weaker compared to 0328.

A similar trend was observed for the specular scans. Motifs such as 0328 tend to have a strong peak #11 and weaker #10 in the specular scan. In general, structures which are in good agreement with the specular data, tend to have only a weak #33 peak in the GID measurement. The best specular match is structure 1945 and is of the same type as 0328. However the side chains are only bent at one end of the backbone. Additional information about the side chain orientation is not available, however bent side chains seem to be favored by energy and X-ray data.

Symmetrization

A reliable comparison of the structure factors is quite difficult because the positions of the side chains are not well defined by the MD simulation. It was not possible to determine proper positions, however it seems that bent alkyl chains are favored by low energy packings and the X-ray data. In order to get a better understanding of the influence of the side chain orientation on the structure factors a symmetrization to space group $P\bar{1}$ was carried out in *PLATON* and subsequently optimized by means of DFT in *CASTEP*. DBDCS is a rather large system, thus DFT optimization was only feasible for a small number of results. The structure factors of the most promising packings are given in Figure 4.32. Based on the fact, that peak #33 is missing in result 0328b it seem unlikely to be correct, although the energy is 6.9 kcal/mol lower. So far none of the simulated structures seem to fulfill all requirements. However, 0460b seems to be the best match so far for following reasons. The dominant peak #33 is present and the

4. Results and Discussion

energy of reasonably low. Furthermore, the packing features a parallel stacking motif with a slip along the short molecular axis, which is in agreement with the properties described in [67].

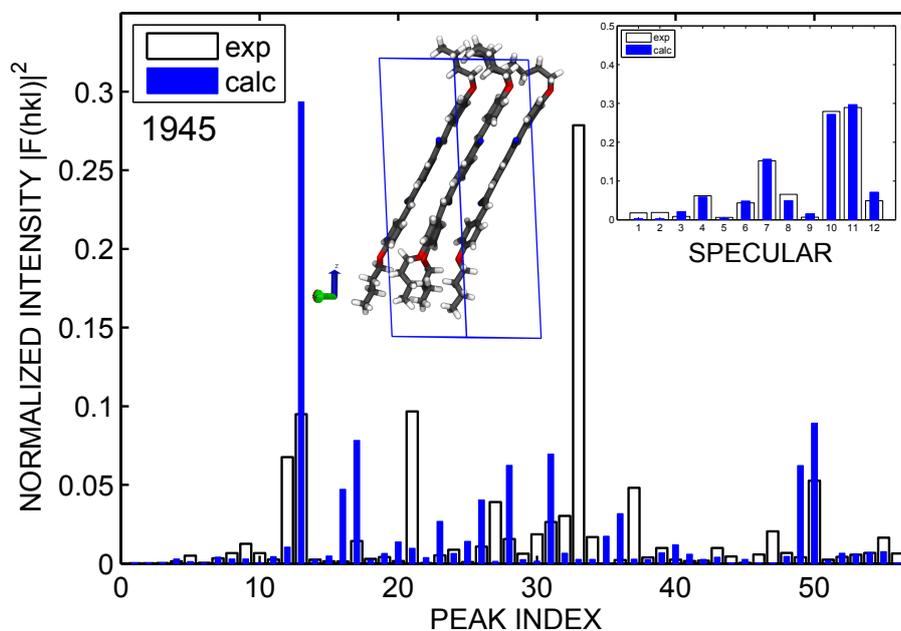


Figure 4.30.: Best structure (1945) based on specular structure factors in comparison with the GID data. 1945 is of the same type as 0328 but the side chains are only bent at one end of the backbone.

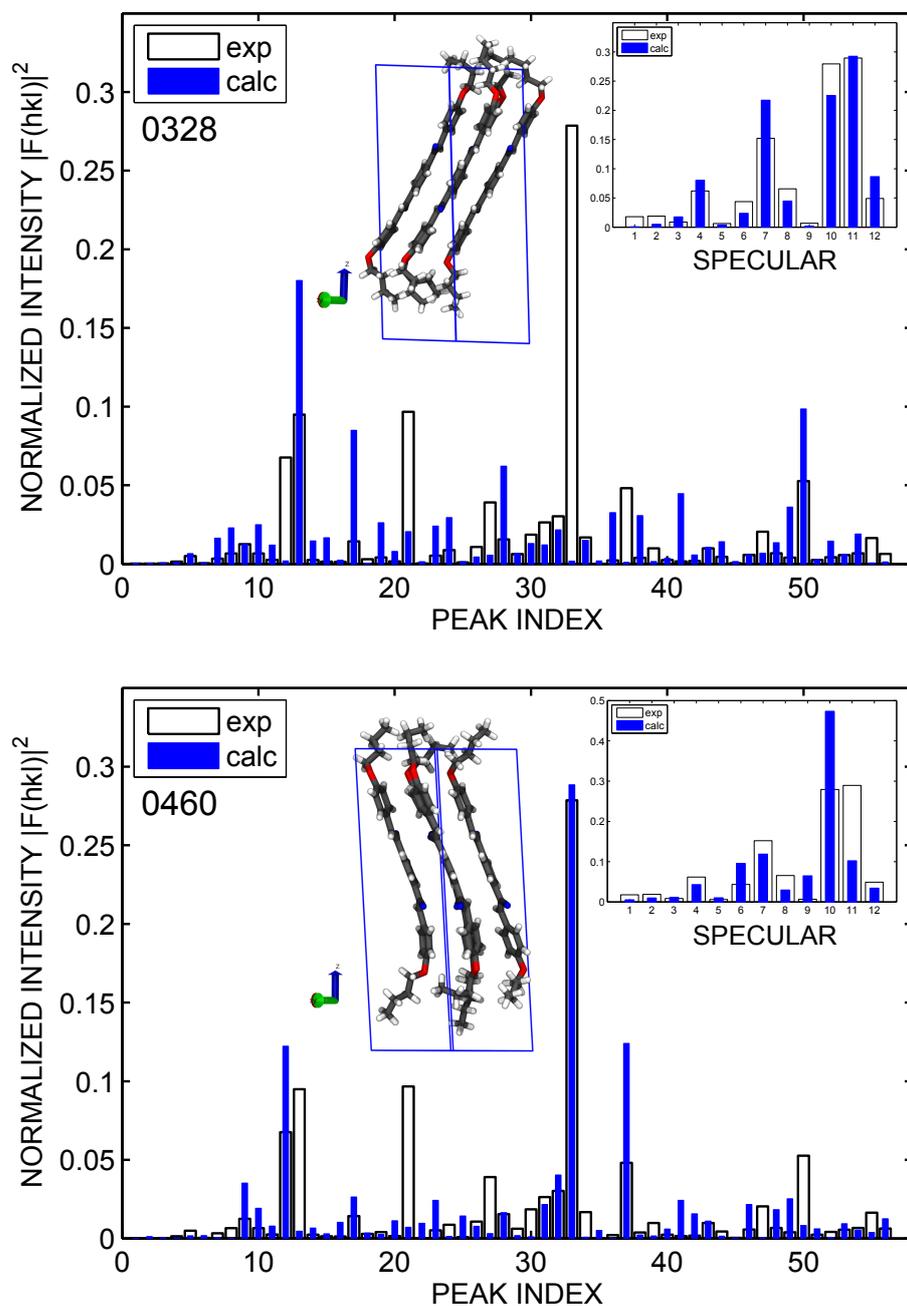


Figure 4.31.: Comparison of the structure factors for structure 0328 and 0460 obtained by MD simulations and X-ray measurements (GID and specular scans). Top: The 0328 (-211 kcal/mol) motif features a strong peak #13 and the absence of peak #33 in the GID measurement. Bottom: In contrast, motifs similar to result 0460 (-57kcal/mol) feature a strong peak #33 for GID and a very weak peak #13.

4. Results and Discussion

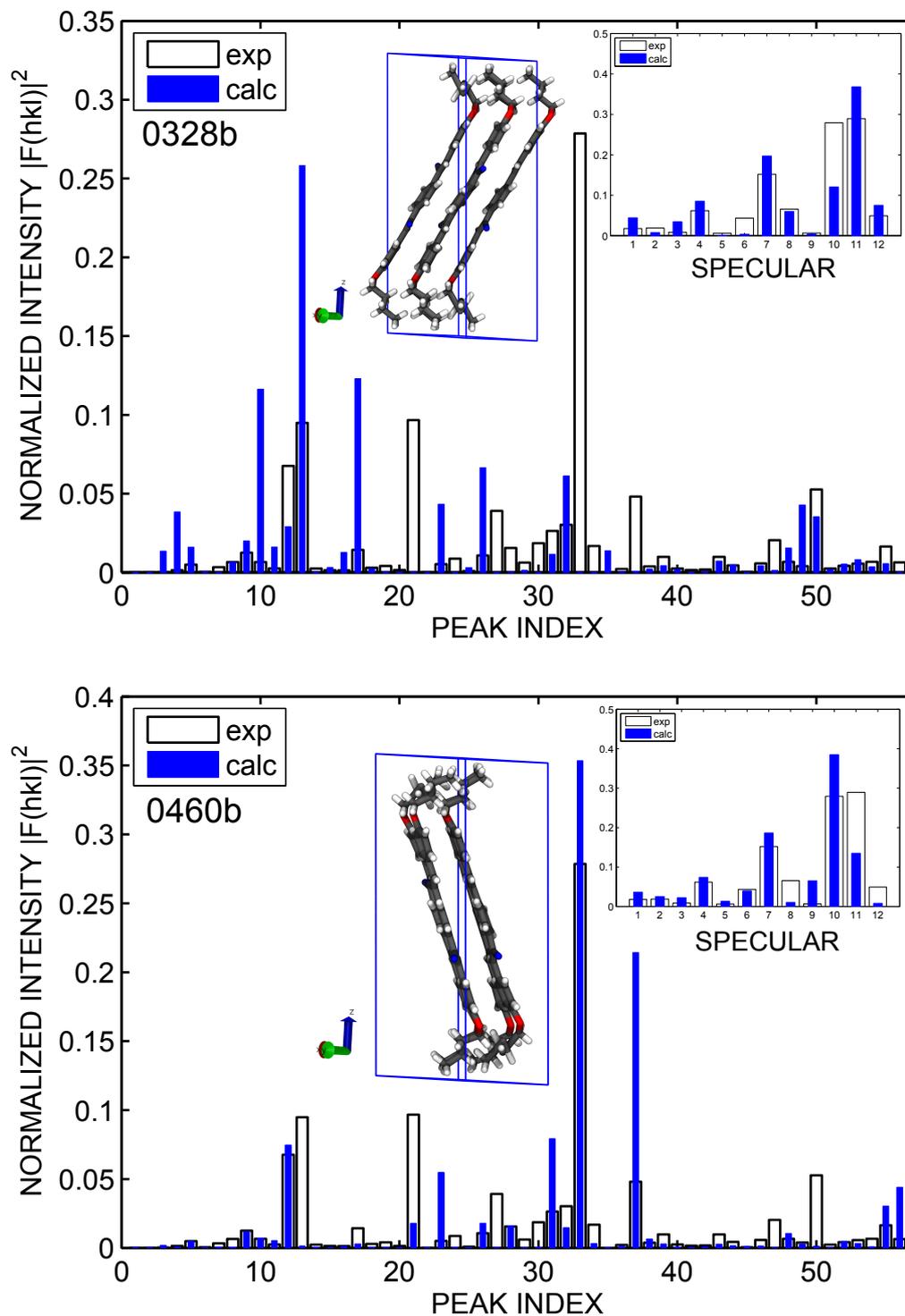


Figure 4.32.: Comparison of the structure factors of structure 0328b and 0460b obtained by symmetrization of 0328 and 0460 to space group $P\bar{1}$. The experimental data was obtained by GID and specular measurements.

5. Conclusion

The capabilities of molecular dynamics simulations for organic thin film crystal structure solution based on grazing incidence X-ray experiments using the CHARMM *CGenFF* force field was investigated. The parametrization by analogy procedure offers in principle a quick and convenient parametrization of new molecules. However, due to the way CHARMM force fields are designed, many parameters which are especially important for the connection of molecular subunits, especially dihedral parameters, are not properly included and manual re-optimization of those parameters is necessary. Even more difficult is the determination of partial charges since a fit to the electro-static potential is not applicable. Charges are based on the reproduction of HF/6-31G* water-model compound interaction energies, which is quite difficult without the complete CHARMM suite.

The *CGenFF* force field was able to model the typical planarization of flexible molecules in a densely packed crystallographic unit cell. Unfortunately the herringbone angles were not reproduced in a satisfying manner, even if all parameters were available in the force field without penalty scores. For pentacene and NNN the herringbone angle was approximately 5° to low compared to literature values. The MD results have been used as starting geometries for plane-wave DFT calculations. A relaxation of the structures was able to reproduce the herringbone angles within a reasonable tolerance.

For DOTT the parametrization did not provide all force field parameters and the best-match alternatives for the thiophene inter-ring dihedral had high penalties. In order to check the usability of these parameters DFT calculations for the octyl-thiophene dihedral have been performed and compared to the PES produced by the MD simulation. The thiophene inter-ring potential was also investigated by means of DFT and the influence of different methods (MP2 and B3LYP) and basis sets was studied. Since a charge and force field optimization was not feasible, the MD simulations have been carried out with the parameters as provided by the analogy parametrization. The orientation of the side chains matched the energy minimum of the dihedral potential and fulfill the close packing principle. The MD results showed clearly a different herringbone angle for the b- and s-phase of DOTT. The 3T backbone which exhibited a twisted conformation in

5. Conclusion

vacuum DFT calculations, was planarized in the shrinking cell simulations. This is in excellent agreement with the single crystal solution. Averaging over 1000 frames lead to more symmetric orientation of the side chains and was therefore able to reduce the influence of thermal motion. However the carbon-carbon bond between the thiophene rings was roughly 0.07 Å too short, which is probably due to the suboptimal force field parameters. A subsequent DFT optimization in VASP was able to reproduce the experimental intra-molecular distances and significantly decreased ($\approx 5^\circ$) the herringbone angle for both phases.

In the case of DBDCS the situation was more difficult for multiple reasons. First of all, the structure factor extraction was difficult due to the awkward peak positions. Furthermore the combination of a large unit cell with four flexible molecules proved to be a challenging system. Furthermore, the MD simulation was not capable of producing well-defined positions for sidechains. However, the results indicated that π -stacking and bent terminal chains are the preferred motif. The former is also expected from experiments, since the activation barrier for the phase transition is very low. Due to the more or less random orientation of the alkyl chains an evaluation based on energy or structure factors was quite difficult and ambiguous. Hence symmetrization was necessary to see an defined impact between certain structural changes on the structure factor spectrum. Subsequent DFT optimization of the symmetrized structures revealed very promising molecular packings.

Structure factors are particularly sensitive to the atomic positions and rather small deviations can have a strong impact. Also the determination of the structure factors, especially for GID measurements, is a quite delicate procedure, due to importance of correction factors and the influence of the experimental setup and sample must be taken into account. In addition, if peaks are too close to each other or even start overlapping a meaningful intensity extraction is not possible.

MD simulations also have their limitations as demonstrated with the problematic DBDCS system, however they still provide an extremely useful alternative to rigid body refinement. Even without an optimized force field, MD simulations can be used to create input structures for further optimizations by means of DFT, which does not rely on the approximations of molecular mechanic force fields. Another important improvement to rigid body refinement is that this approach does not depend on the sensitive X-ray intensity data. It just uses it as an additional source of information to optimize and verify the results.

A. Appendix

A.1. Perylene

Perylene is a flat and rigid aromatic hydrocarbon and consists of two bonded naphthalene units connected by two carbon-carbon bonds. Due to its blue fluorescence, it is used as a blue-emitting dopant material in OLEDs. Perylene crystallizes in a monoclinic system, with four molecules in the unit cell.

A.1.1. Intensity correction

A specular scan of a thin film perylene sample was recorded. The specular scans probes only the out-of-plane order, thus net planes parallel to the sample surface (00 l). The measured intensities have been corrected for the Lorentz and Polarization factor as given in Eq. (2.63) and Eq. (2.62). Figure A.2 illustrates the importance of the correction factors, which are absolutely necessary to obtain meaningful data.

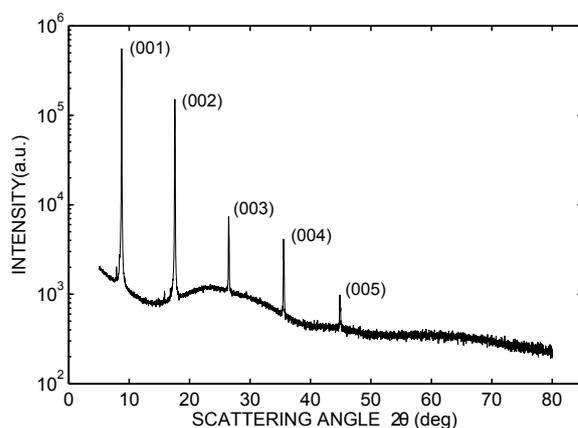


Figure A.1.: Specular scan of a perylene thin film using Cu K_{α} radiation.

A. Appendix

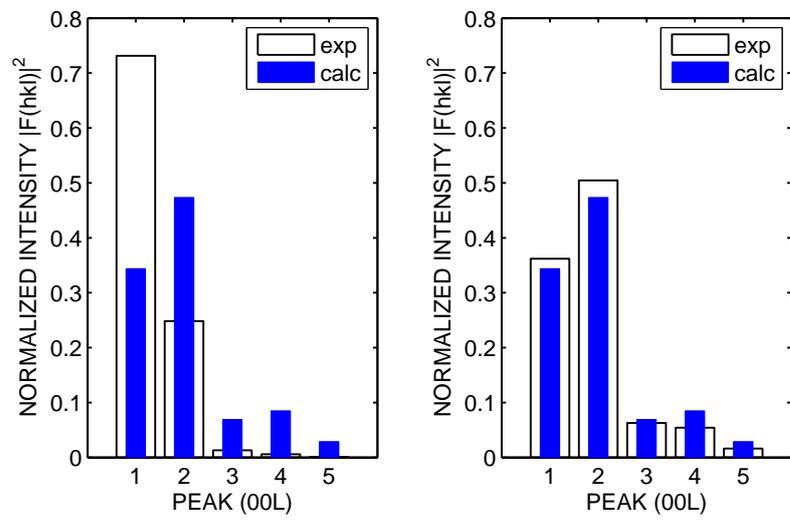


Figure A.2.: Left: Intensities without correction. Right: Intensities corrected for Lorentz- and Polarization factor to obtain meaningful structure factors. Calculated structure factors are based on the structure solution in [61].

Bibliography

- [1] Y. Guo, G. Yu, and Y. Liu: Functional Organic Field-Effect Transistors. *Advanced Materials* **22**(40) (2010), 4427–4447. DOI: 10.1002/adma.201000740.
- [2] M. O’Neill and S. Kelly: Liquid Crystals for Charge Transport, Luminescence, and Photonics. *Advanced Materials* **15**(14) (2003), 1135–1146. DOI: 10.1002/adma.200300009.
- [3] J. E. Anthony: Functionalized Acenes and Heteroacenes for Organic Electronics. *Chemical Reviews* **106**(12) (2006), 5028–5048. DOI: 10.1021/cr050966z.
- [4] I. McCulloch et al.: Influence of molecular design on the field-effect transistor characteristics of terthiophene polymers. *Chemistry of Materials* **17**(6) (2005), 1381–1385.
- [5] C. C. Mattheus et al.: Polymorphism in pentacene. *Acta Crystallographica Section C Crystal Structure Communications* **57**(8) (2001), 939–941. DOI: 10.1107/S010827010100703X.
- [6] S. Schiefer, M. Huth, A. Dobrinevski, and B. Nickel: Determination of the crystal structure of substrate-induced pentacene polymorphs in fiber structured thin films. *Journal of the American Chemical Society* **129**(34) (2007), 10316–10317.
- [7] V. De Cupere et al.: Effect of Interfaces on the Alignment of a Discotic Liquid-Crystalline Phthalocyanine. *Langmuir* **22**(18) (2006), 7798–7806. DOI: 10.1021/la0605182.
- [8] J. D. Watson and F. H. Crick: Molecular structure of nucleic acids. *Nature* **171**(4356) (1953), 737–738.
- [9] A. Klug: Rosalind Franklin and the Discovery of the Structure of DNA. *Nature* **219**(808) (1968).
- [10] D. W. Bennett: *Understanding single-crystal x-ray crystallography*. Weinheim: Wiley-VCH, 2010.
- [11] A. Moser: Crystal Structure Solution Based on Grazing Incidence X-ray Diffraction. PhD thesis. TU Graz, 2012.

Bibliography

- [12] A. Pichler: Crystal Structure Solution from Thin Films. Master's thesis. TU Graz, 2013.
- [13] A. Nemkevich, H.-B. Bürgi, M. A. Spackman, and B. Corry: Molecular dynamics simulations of structure and dynamics of organic molecular crystals. *Physical Chemistry Chemical Physics* **12**(45) (2010), 14916–14929.
- [14] I. Newton: *Philosophiæ naturalis principia mathematica*. Londini: Jussu Societatis Regiæ ac Typis Josephi Streater. Prostat apud plures Bibliopolas, 1687.
- [15] L. Verlet: Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Physical Review* **159**(1) (1967), 98–103. DOI: 10.1103/PhysRev.159.98.
- [16] F. Jensen: *Introduction to computational chemistry*. Chichester, England; Hoboken, NJ: John Wiley & Sons, 2007.
- [17] H. C. Andersen: Molecular dynamics simulations at constant pressure and/or temperature. *Journal of Chemical Physics* **72** (1980), 2384–2393. DOI: 10.1063/1.439486.
- [18] P. M. Morse: Diatomic Molecules According to the Wave Mechanics. II. Vibrational Levels. *Physical Review* **34**(1) (1929), 57–64. DOI: 10.1103/PhysRev.34.57.
- [19] J. E. Jones: On the Determination of Molecular Fields. II. From the Equation of State of a Gas. *Royal Society of London Proceedings Series A* **106** (1924), 463–477. DOI: 10.1098/rspa.1924.0082.
- [20] P. P. Ewald: Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Annalen der Physik* **369**(3) (1921), 253–287. DOI: 10.1002/andp.19213690304.
- [21] A. Y. Toukmaji and J. A. Board Jr.: Ewald summation techniques in perspective: a survey. *Computer Physics Communications* **95**(2) (1996), 73–92. DOI: 10.1016/0010-4655(96)00016-1.
- [22] D. S. Sholl and J. A. Steckel: *Density functional theory: A practical introduction*. Hoboken, N.J.: Wiley, 2009.
- [23] P. Hohenberg and W. Kohn: Inhomogeneous Electron Gas. *Physical Review* **136**(3) (1964), B864–B871. DOI: 10.1103/PhysRev.136.B864.
- [24] W. Kohn and L. J. Sham: Self-Consistent Equations Including Exchange and Correlation Effects. *Physical Review* **140**(4) (1965), A1133–A1138. DOI: 10.1103/PhysRev.140.A1133.
- [25] J. P. Perdew and A. Zunger: Self-interaction correction to density-functional approximations for many-electron systems. *Physical Review B* **23**(10) (1981), 5048.

- [26] J. P. Perdew and Y. Wang: Accurate and simple analytic representation of the electron-gas correlation energy. *Physical Review B* **45**(23) (1992), 13244–13249. DOI: 10.1103/PhysRevB.45.13244.
- [27] J. P. Perdew, K. Burke, and M. Ernzerhof: Generalized Gradient Approximation Made Simple. *Physical Review Letters* **77**(18) (1996), 3865–3868. DOI: 10.1103/PhysRevLett.77.3865.
- [28] J. P. Perdew et al.: Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation. *Physical Review B* **46**(11) (1992), 6671.
- [29] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch: Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *The Journal of Physical Chemistry* **98**(45) (1994), 11623–11627. DOI: 10.1021/j100096a001.
- [30] K. Kim and K. D. Jordan: Comparison of Density Functional and MP2 Calculations on the Water Monomer and Dimer. *The Journal of Physical Chemistry* **98**(40) (1994), 10089–10094. DOI: 10.1021/j100091a024.
- [31] C. Møller and M. S. Plesset: Note on an Approximation Treatment for Many-Electron Systems. *Physical Review* **46**(7) (1934), 618–22. DOI: 10.1103/PhysRev.46.618.
- [32] R. Ditchfield, W. J. Hehre, and J. A. Pople: Self-Consistent Molecular-Orbital Methods. IX. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules. *The Journal of Chemical Physics* **54**(2) (2003), 724–728. DOI: 10.1063/1.1674902.
- [33] T. H. D. Jr: Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *The Journal of Chemical Physics* **90**(2) (1989), 1007–1023. DOI: 10.1063/1.456153.
- [34] B. E. Warren: *X-Ray Diffraction*. Dover Publications, 1969. 381 pp.
- [35] W. Friedrich, P. Knipping, and M. Laue: Interferenzerscheinungen bei Röntgenstrahlen. *Annalen der Physik* **346**(10) (1913), 971–988. DOI: 10.1002/andp.19133461004.
- [36] W. H. Bragg and W. L. Bragg: *X rays and crystal structure*. London : G. Bell, 1915. 260 pp.

Bibliography

- [37] P. J. Brown, A. G. Fox, E. N. Maslen, M. A. O’Keefe, and B. T. M. Willis: Intensity of diffracted intensities. *International Tables for Crystallography*. Ed. by E. Prince. Red. by H. Fuess et al. 1st ed. C. Chester, England: International Union of Crystallography, 2006, 554–595.
- [38] D.-M. Smilgies: Geometry-independent intensity correction factors for grazing-incidence diffraction. *Review of Scientific Instruments* **73**(4) (2002), 1706–1710. DOI: 10.1063/1.1461876.
- [39] M. Neuschitzer et al.: Grazing-incidence in-plane X-ray diffraction on ultra-thin organic films using standard laboratory equipment. *Journal of Applied Crystallography* **45**(2) (2012), 367–370. DOI: 10.1107/S0021889812000908.
- [40] K. Vanommeslaeghe et al.: CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of Computational Chemistry* **31**(4) (2010), 671–690. DOI: 10.1002/jcc.21367.
- [41] *LAMMPS Documentation: Triclinic (non-orthogonal) simulation boxes*. URL: http://lammps.sandia.gov/doc/Section_howto.html#howto_12 (visited on Jan. 30, 2014).
- [42] S. C. Harvey, R. K.-Z. Tan, and T. E. Cheatham: The flying ice cube: Velocity rescaling in molecular dynamics leads to violation of energy equipartition. *Journal of Computational Chemistry* **19**(7) (1998), 726–740. DOI: 10.1002/(SICI)1096-987X(199805)19:7<726::AID-JCC4>3.0.CO;2-S.
- [43] *LAMMPS Documentation*. URL: <http://lammps.sandia.gov/doc/Manual.html> (visited on Jan. 30, 2014).
- [44] W. Humphrey, A. Dalke, and K. Schulten: VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics* **14** (1996), 33–38.
- [45] J. Stone: An Efficient Library for Parallel Ray Tracing and Animation. Master’s thesis. University of Missouri-Rolla, 1998.
- [46] J. Gullingsrud: *MatDCD - Matlab package DCD reading/writing*. URL: <http://www.ks.uiuc.edu/Development/MDTools/matdcd/> (visited on Jan. 30, 2014).
- [47] K. Vanommeslaeghe, E. P. Raman, and A. D. MacKerell: Automation of the CHARMM General Force Field (CGenFF) II: Assignment of Bonded Parameters and Partial Atomic Charges. *Journal of Chemical Information and Modeling* **52**(12) (2012), 3155–3168. DOI: 10.1021/ci3003649.

- [48] H. J. C. Berendsen, J. P. M. Postma, W. F. v. Gunsteren, A. DiNola, and J. R. Haak: Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* **81**(8) (1984), 3684–3690. DOI: 10.1063/1.448118.
- [49] G. Kresse and J. Hafner: Ab initio molecular dynamics for liquid metals. *Physical Review B* **47**(1) (1993), 558.
- [50] G. Kresse and J. Hafner: Ab initio molecular-dynamics simulation of the liquid-metal-amorphous-semiconductor transition in germanium. *Physical Review B* **49**(20) (1994), 14251.
- [51] G. Kresse and J. Furthmüller: Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Computational Materials Science* **6**(1) (1996), 15–50.
- [52] G. Kresse and J. Furthmüller: Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Physical Review B* **54**(16) (1996), 11169.
- [53] P. E. Blöchl: Projector augmented-wave method. *Physical Review B* **50**(24) (1994), 17953.
- [54] G. Kresse and D. Joubert: From ultrasoft pseudopotentials to the projector augmented-wave method. *Physical Review B* **59**(3) (1999), 1758.
- [55] H. J. Monkhorst and J. D. Pack: Special points for Brillouin-zone integrations. *Physical Review B* **13**(12) (1976), 5188–5192.
- [56] *VASP Documentation*. URL: <https://www.vasp.at/index.php/documentation> (visited on Feb. 7, 2014).
- [57] *Gaussian 09 Help*. URL: http://www.gaussian.com/g_tech/g_ur/g09help.htm (visited on Feb. 10, 2014).
- [58] *LAMMPS Documentation: fix restrain command*. URL: http://lammps.sandia.gov/doc/fix_restrain.html (visited on Feb. 10, 2014).
- [59] O. Guvench and A. D. MacKerell Jr: Automated conformational energy fitting for force-field development. *Journal of molecular modeling* **14**(8) (2008), 667–679.
- [60] C. Lercher: Effects of temperature on the polymorphism of dioctylterthiophene in thin films. Master’s thesis. TU Graz, 2013.
- [61] A. Camerman and J. Trotter: The Crystal and Molecular Structure of Perylene. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* **279**(1376) (1964), 129–146. DOI: 10.1098/rspa.1964.0094.
- [62] R. B. Campbell, J. M. Robertson, and J. Trotter: The crystal and molecular structure of pentacene. *Acta Crystallographica* **14**(7) (1961), 705–711. DOI: 10.1107/S0365110X61002163.

Bibliography

- [63] D. Nabok et al.: Crystal and electronic structures of pentacene thin films from grazing-incidence x-ray diffraction and first-principles calculations. *Physical Review B* **76**(23) (2007), 235322. DOI: 10.1103/PhysRevB.76.235322.
- [64] U. Mitschke and P. Bäuerle: The electroluminescence of organic materials. *Journal of Materials Chemistry* **10**(7) (2000), 1471–1507. DOI: 10.1039/A908713C.
- [65] G. Raos, A. Famulari, and V. Marcon: Computational reinvestigation of the bithiophene torsion potential. *Chemical Physics Letters* **379**(3) (2003), 364–372. DOI: 10.1016/j.cplett.2003.08.060.
- [66] V. Marcon, G. Raos, and G. Allegra: Tetrathiophene on Graphite: Molecular Dynamics Simulations. *Macromolecular Theory and Simulations* **13**(6) (2004), 497–505. DOI: 10.1002/mats.200400012.
- [67] J. Gierschner and S. Y. Park: Luminescent distyrylbenzenes: tailoring molecular structure and crystalline morphology. *Journal of Materials Chemistry C* **1**(37) (2013), 5818–5832. DOI: 10.1039/C3TC31062K.
- [68] S.-J. Yoon et al.: Multistimuli Two-Color Luminescence Switching via Different Slip-Stacking of Highly Fluorescent Molecular Sheets. *Journal of the American Chemical Society* **132**(39) (2010), 13675–13683. DOI: 10.1021/ja1044665.