University of Southampton

Faculty of Engineering and Physical Science
School of Chemistry

Computational Screening of Small-Molecule Organic Semiconductors

Chi Yang Cheng

A thesis for the degree of
Doctor of Philosophy

October 2021

University of Southampton

Abstract

Faculty of Engineering and Physical Science
School of Chemistry

Doctor of Philosophy

Computational Screening of Small-Molecule Organic Semiconductors

Chi Yang Cheng

A computational screening workflow for small-molecule organic semiconductors which starts from a defined search space of molecules and ends with a set of proposed molecules was made. The MolBuilder program was developed to use an evolutionary algorithm to optimise the molecular structures of a population of molecules constrained to a search space defined by MolBuilder using a set of molecular fragments. We successfully applied the MolBuilder program to a search space of nitrogen substituted polycyclic aromatic hydrocarbons, and indenofluorenedione derivatives to obtain high-performance n-type organic semiconductors by using a fitness function that optimises for low reorganisation energies and specific electron affinities. In both cases, the computational screening workflow was made to take the best molecules proposed by the evolutionary algorithm through a crystal-structure prediction and electron mobility calculation stage for further evaluations. Based on the results of the computational screening workflows: for the search space of nitrogen substituted polycyclic aromatic hydrocarbons, suggested molecules were found to be competitive to a set of azapentacenes designed from computational considerations; and for the search space of indenofluorenedione derivatives, suggested molecules were found to be superior to a set of previously synthesised indenofluorenedione molecules used in organic electronic applications.

Contents

List of Figures

List of Tables

Additional Material

Additional Computational Data to support the University of Southampton Doctoral thesis, Computational Screening of Small-Molecule Organic Semiconductors

https://doi.org/10.5258/SOTON/D2169

Declaration of Authorship

I declare that this thesis and the work presented in it is my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. Parts of this work have been published as:

   C. Y. Cheng, J. E. Campbell and G. M. Day, Chem. Sci. **11**, 4922 (2020)

   C. Y. Cheng and G. M. Day, Manuscript In Preparation

Signed:..... Date:.....

Acknowledgements

I would like to thank Prof. Graeme M Day for providing useful advice and assistance throughout the years and for being an approachable supervisor who creates an environment of learning and exploration within the research group. I extend my thanks to the Day group members for their useful opinions and I am especially thankful to those who had worked on the group's crystal-structure prediction code as this work would not have been possible without it. I thank Prof. Jochen Blumberger and Dr. Orestis Ziogos for sharing and supporting their pyAOMlite code which became an important part of this work and I also thank them for their useful discussions on carrier mobility calculations. Personal acknowledgements go to my family who has always supported my decision to pursue a career in the sciences. I am grateful for support from the EPSRC Centre for Doctoral Training in Theory and Modelling in Chemical Sciences (grant EP/L015722/1); acknowledge the use of the IRIDIS High-Performance Computing Facility and associated support services at the University of Southampton in the completion of this work; and I am grateful to the UK Materials and Molecular Modelling Hub for computational resources, which is partially funded by EPSRC (EP/P020194/1 and EP/T022213/1).

Abbreviations

AOM	Analytic Overlap Method
BFGS	Broyden-Fletcher-Goldfarb-Shanno (Algorithm)
CCDC	Cambridge Crystallographic Data Centre
CCSD	Coupled-Cluster Singles and Doubles
CISD	Configuration Interaction Singles and Doubles
CSD	Cambridge Structural Database
CSP	Crystal-Structure Prediction
DFT	Density-Functional Theory
DFT-D	Dispersion-Corrected Density-Functional Theory
exp-6	Buckingham (Force Field)
EA	Evolutionary Algorithm
FO-DFT	Fragment Molecular Orbital Density-Functional Theory
FOB-SH	Fragment Orbital-Based Surface Hopping
gTLM	Gated Transfer Length Method
GCA	Gradual Channel Approximation
GDMA	Gaussian Distributed Multipole Analysis
GMH	Generalised Mulliken-Hush
GGA	Generalised Gradient Approximation
HF	Hartree-Fock
KMC	Kinetic Monte Carlo
KS-DFT	Kohn-Sham Density-Functional Theory
LDA	Local-Density Approximation
LEFET	Organic Light-Emitting Field-Effect Transistors
LEIPS	Low-Energy Inverse Photoemission Spectroscopy

OFET	Organic Field-Effect Transistors
OLED	Organic Light-Emitting Diodes
OSC	Organic Solar Cells
PAH	Polycyclic Aromatic Hydrocarbon
QR	Quasi-Random
RHF	Restricted Closed-Shell Hartree-Fock
RKS	Restricted Closed-Shell Kohn-Sham Density-Functional Theory
RTA	Relaxation Time Approximation
SAPT	Symmetry Adapted Perturbation Theory
SAT	Separating Axis Theorem
SC-FET	Single Crystal Field-Effect Transistor
TLT	Transient Localisation Theory
TVP	Target Volume Parameter
UFF	Universal Force Field

Chapter 1

Introduction

This thesis describes the work carried out to create a computational screening workflow for small-molecule organic semiconductors and provides an overview of the relevant background, theory, and programs used. The computational screening of small-molecule organic semiconductors, which begins with a search space of molecules and ends with a specific set of molecules with predicted performances, forms a complex structure of computational methods since a given molecule can be sent through a series of calculations using a large number of different methods with results from one method used as the input for another.

In the development of our computational screening approach, we have primarily focused on: the development of the MolBuilder program used for chemical structure optimisations; benchmark work on the molecular conformer generation, crystal-structure prediction (CSP), and carrier mobility calculation methods; the development of an efficient reduced sampling scheme; and the application and evaluation of the computational screening workflows. We have therefore worked to ensure a reasonable compromise between computational cost and accuracy given the methods and resources currently available for the most important components of the computational screening workflow. In the following paragraphs of this introductory chapter, we will provide a brief outline of some of the subsequent chapters and sections within this thesis.

In Chapter 2 we discuss a number of background topics related to the computational screening of small-molecule organic semiconductors. In Section 2.1

we give an overview of organic semiconductors and the experimental methods used to calculate the carrier mobility of an organic field-effect transistor. In Section 2.2 we provide an overview of the CSP methods used in Chapter 6 of this thesis and some methods used in the sixth blind test of organic crystal-structure prediction. In Section 2.3 we compare and review some of the latest work on the computational screening of organic semiconductors and discuss the different search spaces that are chosen and the methods used to rank a molecule's ability to form high-performance organic semiconductors.

In Chapter 3 we work through the theory behind the computational methods used in our computational screening workflow. In Sections 3.1 and 3.2 we provide an overview of the electronic-structure and density-functional theories respectively. In Section 3.3 we provide the theoretical justifications for the force fields used in the lattice energy minimisation methods of our CSP workflows. In Section 3.4 we derive the spin-boson model, outline methods used to calculate parameters in the spin-boson model, and derive the Marcus theory transition rate equation. In Section 3.5 we describe a number of methods used to calculate carrier mobilities.

In Chapter 4 we provide an overview of CSPy and some related programs that are used to automate parts of the CSP workflow. In Section 4.1 we describe the methods used in the programs GDMA and Mulfit to generate distributed multipoles parameters for our force fields. In Section 4.2 we describe and compare two different versions of the quasi-random structure generation algorithms implemented in CSPy. In Section 4.3 we describe the force fields and methods used by DMACRYS to energy minimise crystal structures. In Chapter 5 we will provide an overview of the MolBuilder program which is used to optimise the molecular structures of a population of molecules using an evolutionary algorithm for specific properties.

In Chapter 6 we describe the work on the application of the computational screening workflow on a search space of nitrogen substituted polycyclic aromatic hydrocarbons to obtain high-performance n-type organic semiconductors. In Chapter 7 we benchmark a number of different components of our computational screening workflow including the molecular conformer generation, the CSP force fields and sampling schemes, and the carrier mobility calculation methods.

By using the CSP benchmark results, we outlined an efficient and reliable reduced sampling scheme to be used for the computational screening of rigid aromatic molecules. In Chapter 8 we describe the work on the application of the computational screening workflow on a search space of indenofluorenedione derivatives using an improved computational screening workflow with an increased number of evaluations in each stage.

Chapter 2

Background

2.1 Organic Semiconductors

Organic semiconductors have the potential to become a useful alternative to their inorganic counterparts despite their typically lower carrier mobilities as they provide a few advantages including lower fabrication costs, flexibility, and transparency. Experimental work in the field has for example looked to apply organic semiconductor materials to the development of a number of organic electronic devices including organic field-effect transistors (OFET), organic solar cells (OSC), and organic light-emitting diodes (OLED) [1].

2.1.1 Categories

The chemical structures of organic molecules can vary considerably leading to a large number of different types of molecules. A few example molecules used in organic semiconductor research are shown in Figure 2.1. Usually, these molecules are grouped into two main categories: small-molecule and conjugated polymer. The small-molecule organic semiconductors usually contain planar aromatic substructures as seen for pentacene, rubrene, and perylenetetracarboxylic dianhydride (PTCDA), although there are some exceptions, for example, [6,6]-phenyl-C_{61}-butyric acid methyl ester (PCBM) which contains a fullerene substructure. The conjugated polymers are formed from a large number of aromatic monomer units as seen in the poly(3-hexylthiophene-2,5-diyl) (PH3T)

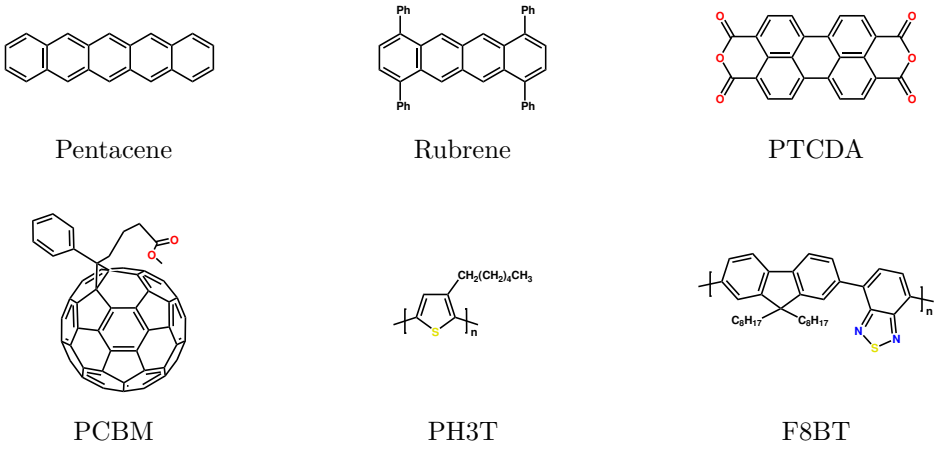


Figure 2.1: Molecular structures of: four small-molecule organic semiconductors pentacene, rubrene, perylenetetracarboxylic dianhydride (PTCDA), and [6,6]-phenyl-C_{61}-butyric acid methyl ester (PCBM); and two conjugated polymer organic semiconductors poly(3-hexylthiophene-2,5-diyl) (PH3T) and poly(9,9-di-n-octyl-fluorene-alt-benzothiadiazole (F8BT).

and poly(9,9-di-n-octyl-fluorene-alt-benzothiadiazole) (F8BT) polymers.

Out of the small-molecule organic semiconductors, rubrene has become one of the more widely researched molecules in the field due to its high hole mobilities and has been commonly used as a model system in both experimental and theoretical work. Some interesting examples include the work by Ren *et al.* which showed a 13% reduction in the hole mobilities when fully substituted ^{13}C-rubrene molecules are used [2]. While a number of experimental and theoretical work had used rubrene to examine the effects of mechanical strain on its hole mobilities [3–8]. For example in the computational work by Elsner *et al.* increases to the hole mobility along the **a**-axis with compressive strain along the **a** or **b** axes were obtained with larger increases shown in the latter case [7].

The conjugated polymers also show some interesting properties, for example, the PH3T polymer can adopt either the regioregular or regioirregular forms which result in different electronic properties. Formed from head-to-tail couplings between its hexane substituted thiophene monomer units, the regioregular form results in reduced steric hindrances between alkyl side groups which allow the polymer to adopt a more planar configuration leading to greater conjugations across the molecule. The resulting regioregular polymers can then self-assemble

to form a solid-state structure containing π-stacked substructures leading to a material with superior conductivities due to its molecular and solid-state structures [9, 10].

2.1.2 Charge Carriers

Organic semiconductors can also be divided into three different types depending on the predominant charge carrier present in the material when used in a typical organic electronic device. Semiconductors are p-type when charge carriers are predominately holes, n-type when charge carriers are predominately electrons, and ambipolar when charge carriers are neither predominately holes or electrons. For example, the molecules: pentacene, rubrene, and PH3T, are p-type semiconductors; PTCDA and PCBM, are n-type semiconductors; and F8BT is an ambipolar semiconductor.

The p-type organic semiconductor materials are formed from electron-rich molecules[1] with small ionisation energies which provide lower barriers for hole injection. For example, both pentacene and rubrene due to their systems of delocalised π-electrons, and PH3T due to its electron-rich thiophene monomer units, are all p-type semiconductors. The n-type organic semiconductor materials are formed from electron-deficient molecules with large electron affinities which provide lower barriers for electron injection. For example, PTCDA due to its two electron-withdrawing anhydride groups and PCBM due to its electron-deficient fullerene, are both n-type semiconductors. Both p-type and n-type semiconductors are important since for example in OSC devices the photoactive layers can be formed from a blend of p-type and n-type semiconductors [1].

F8BT is an example of a donor-acceptor copolymer as it is formed from electron-rich donor and electron-deficient acceptor monomer units; substituted fluorenes act as its electron-rich unit and benzothiadiazole as its electron-deficient

[1] We use the standard organic chemistry meaning for the phrase electron-rich and electron-deficient when referring to heterocyclic compounds. For example, PH3T is described as electron-rich as the sulfur heteroatom of the 3-hexylthiophene-2,5-diyl monomer feeds electron density into the π-system while PTCDA is described as electron-deficient as its anhydride groups withdraw electron density from the π-system. PBCM is a less obvious example but is an electron-deficient molecule as the fullerene group behaves chemically like an electron-deficient alkene or arene [11].

unit which overall produces an ambipolar semiconductor. When coupled with its electroluminescence properties F8BT has been usefully applied to the fabrication of organic light-emitting field-effect transistors (LEFET), a device with an architecture similar to an OFET but emits light like an OLED. Although LEFETs can be formed from unipolar semiconductors inefficiencies can occur as the electron-hole recombinations take place near the electrode since a majority of only one type of charge carrier is present within the semiconductor. By using an ambipolar semiconductor like F8BT which allows for both electron and hole transport, the recombination processes can be moved into the transistor channel leading to improved efficiencies [12, 13].

2.1.3 OFET

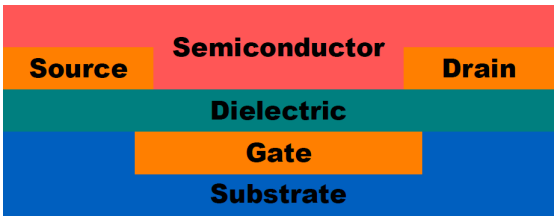


Figure 2.2: A field-effect transistor using the bottom-contact, bottom-gate architecture.

OFETs are devices that control the current flow between its source and drain electrodes by applying a voltage to its drain and gate electrodes. Figure 2.2 shows an OFET using the bottom-contact, bottom-gate architecture and is one of a number of different architectures that can be created each with its own set of advantages and disadvantages [14]. These OFET devices are most commonly described using the gradual channel approximation (GCA) [15], for a p-type OFET the equations for the drain current are

$$I_{\mathrm{DS,lin}} = \frac{W}{2L}\mu C_{\mathrm{diel}}\left[2(V_{\mathrm{GS}} - V_{\mathrm{T}})V_{\mathrm{DS}} - V_{\mathrm{DS}}^2\right] \qquad V_{\mathrm{DS}} < V_{\mathrm{GS}} - V_{\mathrm{T}} \qquad (2.1)$$

$$I_{\mathrm{DS,sat}} = \frac{W}{2L}\mu C_{\mathrm{diel}}\left(V_{\mathrm{GS}} - V_{\mathrm{T}}\right)^2 \qquad V_{\mathrm{DS}} > V_{\mathrm{GS}} - V_{\mathrm{T}} \qquad (2.2)$$

where $V_{GS} > V_T$ for both equations, $I_{DS,lin}$ and $I_{DS,sat}$ are the drain-source currents in the linear and saturation regimes, W and L are the channel width and lengths, μ is the charge carrier mobility, C_{diel} is the gate dielectric capacitance per unit area, V_{GS} is the gate-source voltage, V_{DS} is the drain-source voltage, and V_T is the threshold voltage.

The current between the source and drain electrodes is small when $V_{GS} < V_T$ and large when $V_{GS} > V_T$ so that the OFET has the ability to act as an on/off switch. The ratios between the currents of the OFET in the on and off states is known as the on/off current ratio and needs to be large around 10^6 or above so the two states can be easily distinguished. It is also important for V_T to be close to 0 V for applications in low power devices. The deviation of V_T from zero can occur when a number of trap states are present which can for example occur due to the defects or impurities within the semiconductor [15].

The carrier mobility is another important parameter that needs to be large to give high-performances which can also lead to an improvement to the on/off ratio. The Eqs. (2.1) and (2.2) give two equations for determining the charge carrier mobilities of the OFET experimentally

$$\mu_{lin} = \frac{L}{WC_{diel}V_{DS}} \frac{\partial I_{DS}}{\partial V_{GS}} \tag{2.3}$$

$$\mu_{sat} = \frac{2L}{WC_{diel}} \left(\frac{\partial \sqrt{I_{DS}}}{\partial V_{GS}} \right)^2 \tag{2.4}$$

however a great deal of care must be taken when applying these equations. In a recent commentary article, Choe *et al.* described a number of the common pitfalls which could lead to erroneous results [16]. In brief, some of the difficulties arise due to the assumptions made in the GCA [17] which for example requires that the contact resistances R_c are much smaller than the channel resistance R_{ch}.

The R_c are related to the barrier for the injection of a charge carrier from the electrode into the semiconductor, so to form a high-performance OFET we will want the R_c to be as small as possible. There are a number of experimental methods used to determine R_c and have their own set of advantages and disadvantages, one of these is known as the gated transfer length method

(gTLM) [14,15,18]. In gTLM an equation for R_{ch} is obtained by applying Ohm's law to Eq. (2.1) and taking the limit that $V_{\text{DS}} \ll V_{\text{GS}} - V_{\text{T}}$. The device resistance is the sum of R_{ch} and R_{c} so that

$$R_{\text{device}} = \frac{L}{W\mu C_{\text{diel}}(V_{\text{GS}} - V_{\text{T}})} + R_{\text{c}} \tag{2.5}$$

therefore R_{c} can then be obtained by plotting L against R_{device} and extrapolating to $L = 0$ so that gTLM requires carrying out measurements for a number of different channel lengths.

Another method is known as the gated linear four-point probes [14,15,18,19] which add two additional electrodes onto the OFET placed at the 1/3 and 2/3 positions between the source and drain contacts. Two measurements are carried out at these two electrodes to obtain voltages V_1 and V_2 at positions L_1 and L_2. By assuming a linear potential across the channel we can determine the potential drop on the source and drain due to the contact resistance by extrapolating the potential of V_1 and V_2 to the contacts. By applying Ohm's law we can obtain the contact resistance due to the source or drain electrodes

$$R_{\text{c,s}} = \frac{\Delta V_{\text{drop,s}}}{I_{\text{SD}}} \qquad R_{\text{c,d}} = \frac{\Delta V_{\text{drop,d}}}{I_{\text{SD}}} \tag{2.6}$$

where $R_{\text{c,s}}$ and $R_{\text{c,d}}$ are the contact resistances due to the source and drain electrodes, $\Delta V_{\text{drop,s}}$ and $\Delta V_{\text{drop,s}}$ is the difference in the potential between the extrapolated and expected values. The total contact resistance is the sum of the resistances of the two contacts $R_{\text{c}} = R_{\text{c,s}} + R_{\text{c,d}}$. Using the gated linear four-point probes we can also calculate the mobilities using

$$\mu_{\text{lin,4p}} = \frac{(L_2 - L_1)}{WC_{\text{diel}}} \frac{\partial(I_{\text{DS}}/(V_2 - V_1))}{\partial V_{\text{GS}}} \tag{2.7}$$

so that the impact of R_{c} on the calculated carrier mobilities can be reduced.

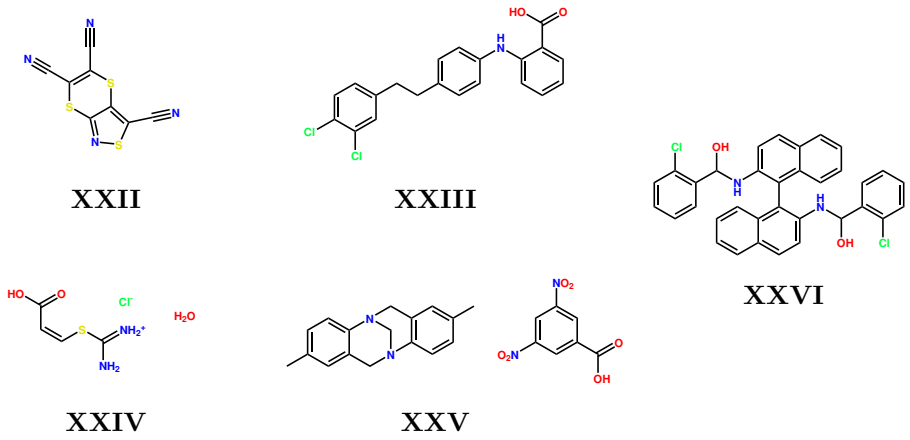


Figure 2.3: The five test systems used in the sixth blind test of organic crystal-structure prediction methods.

2.2 Crystal-Structure Prediction

This project works towards the development of a computational screening workflow for small-molecule organic semiconductors to be used in an organic electronic device such as the OFETs. For the OFET example, the semiconducting layer will be formed from a single crystal or a thin-film of the molecule. The computational screening workflow will therefore require the solid-state structure to be known so that other properties can be calculated such as the carrier mobility and therefore its predicted performances can be determined. Crystal-structure prediction (CSP) methods can be used to predict the solid-state structures of an organic molecule so will therefore play an important role in our computational screening workflow.

Since 1999 there have been a total of six blind tests of organic CSP that have been organised with the sixth blind test ending in 2015 [20–25] which collectively provide an excellent overview of the progress made in the field throughout the years. In each blind test computational groups attempt to predict the crystal structures of a number of molecules from their molecular structures and since the fourth blind test its crystallisation conditions as well. To ensure unbiased testing, unpublished crystal structures are obtained from experimental crystallographers and are then sent to an external referee who selects one crystal structure for each

category. The categories are selected to give a variety of molecules at different levels of difficulty. The molecular structure and related information for each category are sent to computational groups participating in the blind test and a deadline is set for the submission of predicted crystal structures. At some time after the submission deadline, experimental structures are circulated to allow for post-analysis and a workshop is then held to discuss the results.

The sixth blind test was the largest that had been undertaken with a total of 25 participants which therefore showcases a large range of different approaches that can be taken. The five test systems used in the sixth blind test are shown in Figure 2.3. According to the authors of the sixth blind test [25], CSP can be split into three individual stages which they had defined to be

1. Exploration of the conformational preferences of the target molecules.

2. Generating plausible crystal-packing arrangements of the target molecules.

3. Ranking the likelihood of resulting crystal structures forming using some form of scoring or fitness function.

so that for example in Chapter 6 we carried out CSP using the following steps

1. Using RDKit generate a single conformer, optimise geometries using universal force field (UFF), and reoptimise using density-functional theory (DFT).

2. Treat the optimised geometry as a rigid body and use a quasi-random sampling method to generate crystal structures and optimise them using a force field with atomic multipoles.

3. Rank the likelihood of the crystal structures using the energy evaluated with the force field with atomic multipoles.

a more technical overview of the procedure we are using can be seen in Chapter 4. The methods chosen for each step of the CSP procedure can depend on a number of factors such as the flexibility and functional groups of the molecules, and the computational resources available to the researcher.

2.2.1 Conformational Exploration

In order to carry out step 1 efficiently, it can be helpful to consider the flexibility of the molecule. In Chapter 6 we carried out CSP on polycyclic aromatic molecules which have the chance to form π-stacked crystal structures which are required to obtain large carrier mobilities. These molecules are relatively small, similar to the organic semiconductor pentacene and tetracene. The combination of its small sizes and extended π-bonded structures mean that these molecules will be relatively rigid and exist in a single conformer. This allowed for the rather simple approach to step 1 in Chapter 6 which only required the generation of a single conformer followed by UFF and DFT optimisations.

However, in most cases, further provisions must be made to steps 1 and/or 2 to include the flexibility and any conformational isomerism of the molecule. One approach taken in the sixth blind test was to perform conformer searches of the isolated molecule, for example, Day *et al.* used the OPLS2005 [26–28] force field to generate a large number of conformers which were followed up by an optimisation using dispersion-corrected density-functional theory (DFT-D) [29,30]. In another approach, Cole *et al.* generated conformers using the CSD conformer generator which were then optimised using the CSD knowledge-based force fields [31]. While Facelli *et al.* instead allowed certain torsional angles to be included in the variables of the crystal structure in a genetic algorithm [32–35].

2.2.2 Crystal-Packing

In Chapter 6 we obtain *plausible crystal-packing arrangements* required in step 2 by using a quasi-random crystal structures generation algorithm to return a set of crystal structures with a distribution of different cell lengths and angles, and molecule positions and orientations within the unit cell, which are then followed up with a series of optimisations using an atom-atom potential energy model. For a more technical overview see Chapter 4. There are therefore a number of different choices that need to be made in step 2 such as the number of space groups to sample, and the structure generation and optimisation procedures used.

The space groups that are sampled can be made based on the frequency

of their occurrences in the Cambridge Structure Database (CSD). Since for example, only the 10 most common space groups are required to cover 90% of all organic crystals with a single molecule in its asymmetric unit. So one method to choose the space groups that are sampled can be made by sampling only the most frequently observed. We used this method in Chapter 6 and was also frequently seen in the sixth blind test, for example, Day *et al.* sampled the 25 most common space groups for the systems **XXIII**, **XXIV**, **XXV**, and **XXVI**; while Hofmann & Kuleshova sampled the 10 most common space groups for all tests systems. Although there were groups who did not take this route for example Lv, Wang & Ma choose to sample all space groups that have less than or equal to four molecules in their unit cells while Chadha & Singh used all 230 space groups in their searches. Others used this strategy only for certain systems for example Neumann, Kendrick & Leusen sampled all 230 space groups for the system **XXII** while for the other systems they had used only the most common.

Once the number of molecules in the asymmetric unit and space groups are chosen, a process of structure generation and optimisation is carried out. The structure generation and optimisation stages can be tied together in some way and to different extents. In Chapter 6 we keep the two procedures mostly separate so that it simply takes a structure from the quasi-random structure generation algorithm and optimises them, although there are multiple stages to the optimisation procedure we use. The two procedures can be tied together more closely, for example, Day *et al.* generated crystal structures using the molecular conformers with quasi-random lattice parameters, molecular positions, and molecular orientations [36] which were optimised with DMACRYS [37] which fixes molecular geometries so that optimisations are run over rigid molecules and the intermolecular interactions were modelled with an exp-6 force field with multipoles up to hexadecapoles. This optimisation stage with rigid molecules is therefore the structure generation to the next optimisation stage which takes the lowest energy structures for reoptimisation using CrystalOptimizer [38] which allow for both the internal molecular and crystals structures to be optimised together.

Facelli *et al.* tie the two processes even more closely together in their genetic algorithm which randomly generates an initial population of crystal structures

which is then used to create a new generation of crystal structures. In each generation, they create a new population using genetic algorithm type operations with the previous generation and optimises the crystal structures using the MGAC_CHARMM [32–35] or MGAC2_QE [39] procedures. This process of structure generation and optimisation is then repeated continually until 150 generations are created. The best structures from the genetic algorithm are then reoptimised using QUANTUM ESPRESSO [40] if the optimisations in the genetic algorithm had used the MGAC_CHARMM procedure.

2.2.3 Ranking

In Chapter 6 for the final step of the CSP workflow, we rank crystal structures by the energies obtained from the final optimisation in step 2 so that further calculations are not required. This method had seen usages in the sixth blind test particularly for cases when the optimisation method used in step 2 were deemed to be sufficiently accurate so that further energy revaluations were not necessary. When this is not the case attempts to improve ranking can be done by recalculating energies with a higher level of theory for example Obata & Goto used single point PBE+TS [41] energies of crystal structures that had been optimised using the MMFF94 force field [42]. While in another example Marom *et al.* used single point PBE0+MBD [43,44] energies of crystal structures that had been optimised using PBE+TS.

Other approaches try to include thermodynamic contributions to the ranking, for example, Day *et al.* calculated rigid molecule phonon frequencies to calculate free energies [45,46] which were used to rank crystal structures for molecules **XXII** and **XXV**. Pickard *et al.* carry out a similar procedure but instead included full atom harmonic and anharmonic corrections [47] to the free energy. While Tuckerman, Szalewicz *et al.* ranked their crystal structures based on thermally averaged energies obtained by running molecular dynamics simulations [48]. While in one case kinetic contributions to the ranking were considered instead, van den Ende, Cuppen *et al.* used kinetic Monte Carlo simulations [49, 50] to rank crystal structures based on their lowest critical nucleus size.

2.2.4 Summary

In this section, we have given a brief overview of the methods used in each of the three stages of CSP in Chapter 6 and some methods in the sixth blind test of organic crystal-structure prediction. Although CSP methods are used in Chapter 6 of this work and the sixth blind test, there are however some significant differences between the aims of this work and the aims of a blind test which result in some differences between how CSP is applied. In this work, we use CSP as a part of a computational screening workflow where the aim will be to apply CSP to a large number of molecules. It will therefore be preferable to use more efficient methods so that a large number of molecules can be screened. This makes it quite different to a blind test where it will be more favourable to use CSP methods which are in general more accurate and computationally demanding as the aim of the blind tests are to obtain matches with the experimental structures.

Since we use CSP for a large number of molecules we require a black-box approach to CSP with nearly no human intervention required during any stage of the CSP workflow. This means that certain methods may be more difficult to apply such as the ones which include thermodynamic effects by calculating phonon frequencies as issues with imaginary frequencies for structures that are not on an energy minimum may occur. This requirement also places restrictions on the search space of molecules in our computational screening workflow since efficient black-box flexible CSP methods are still in development. We are therefore restricted to rigid molecules which do not change drastically between the gas and solid states. Restrictions will also occur on the choice of energy models since we require a computational method that can reproduce experimental structures within an expected energy range but must also be efficient enough to be able to screen hundreds of molecules.

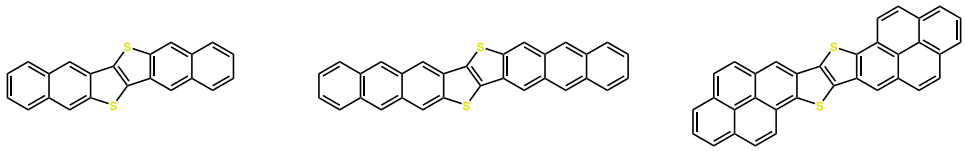
These restrictions force us to use a more streamlined CSP process compared to those in the sixth blind test. Extensive conformational explorations are not required, crystal-packing searches are reduced, optimisation methods are restricted to the more computationally efficient force field models, and ranking will be simply the final energies calculated. Although inevitably there will be shortcomings to the CSP methodology used in the project we show in Chapter 7 that it is still

able to provide a method that can be used to obtain structures in agreement with experiment. However, as CSP and computational screening workflows develop in the future it may be possible to see some methods used in the sixth blind test in a computational screening workflow as more computationally efficient variants are developed or more sophisticated multistaged computational screening workflows with the further crystal structure optimisations are developed.

2.3 Computational Screening

The computational screening for organic semiconductors aims to search for promising molecules that could be used as a part of an organic electronic device. Computational screening can therefore be a useful initial stage of a materials discovery program for an organic semiconductor with the final aim of device fabrication and commercialisation. Computational screening workflows begin by first defining and evaluating molecules within a specified search space and are ranked on their ability to form a material for the required application. Usually, further evaluation stages are carried out on continually smaller subsets using increasingly sophisticated computational methods. At the endpoint, a small number of molecules are evaluated using the most sophisticated computational methods available giving in theory suggestions for the best molecular targets within the search space. As we shall see this methodology can be particularly flexible, with different search spaces, methods, and stages that can be used. In this section, we will review a selection of the latest work carried out on the computational screening of organic semiconductors and highlight some of the different strategies that have been employed.

2.3.1 Dinaphthothienothiophene Derivatives



1 **2** **7**

Figure 2.4: The dinaphthothienothiophene (**1**) molecule that the search space was based on in a study by Sokolov *et al.* and two molecules **2** and **7** that passed the first evaluation stage. Molecule **2** passed a second evaluation stage so was chosen for synthesis and used for the fabrication of a single crystal field-effect transistor (SC-FET) leading to reported hole mobilities of 12.3 and 16 $\mathrm{cm}^2(\mathrm{Vs})^{-1}$ in the saturation and linear regimes respectively [51].

In a study by Sokolov *et al.* a computational screening workflow was applied to a search space containing 7 different dinaphthothienothiophene (**1**) derivatives for organic semiconductors with high hole mobilities [51]. This search space was chosen based on past experimental work on **1** which had been used to create an OFET device with a hole mobility of 3.1 $\mathrm{cm}^2(\mathrm{Vs})^{-1}$ [52, 53]. A two-staged evaluation process was carried out where in the first stage four different molecular properties were calculated. The first three properties were the HOMO and LUMO energies and the HOMO-LUMO gap and were used to determine their relative stabilities. The final property calculated was the hole reorganisation energy which was used to provide some ideas of the relative hole mobilities of the semiconductors.

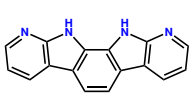
Two molecules **2** and **7** (Figure 2.4) were taken into a second evaluation stage mainly due to their low reorganisation energies of 0.084 and 0.075 eV respectively. In the second stage, they obtained predicted crystal structures of **2** and **7** by assuming that the molecular packings of both molecules were similar to the experimental structure of **1** and then optimising them using the Dreiding force field [54] with B3LYP/6-311+G** derived point charges. Isotropic hole mobilities of both crystal structures were calculated using the Einstein relation and the

diffusion equations [55–57]

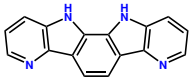
$$\mu = \frac{e_0}{k_B T} D \qquad D = \frac{1}{6} \sum_i R_i^2 P_i k_i \qquad P_i = \frac{k_i}{\sum_i k_i} \tag{2.8}$$

where k_i is the transition rate calculated using Marcus theory, R_i is the hopping distance and P_i is the hopping probability, to the neighbouring site i. Using the Eqs. (2.8) hole mobilities of the predicted crystal structures of **2** and **7** were calculated to be 3.34 and 1.45 $\mathrm{cm}^2(\mathrm{Vs})^{-1}$. Molecule **2** was synthesised and used in the fabrication of a single crystal field-effect transistor (SC-FET) with gold contacts using the bottom-contact, bottom-gate architecture due to its larger calculated hole mobilities and greater predicted air stabilities. The best performing SC-FET device was determined to have hole mobilities of 12.3 and 16 $\mathrm{cm}^2(\mathrm{Vs})^{-1}$ when operating in the saturation and linear regimes respectively.

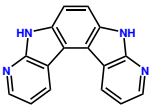
2.3.2 Pyrrole-Based Azaphenacene Isomers



1



4



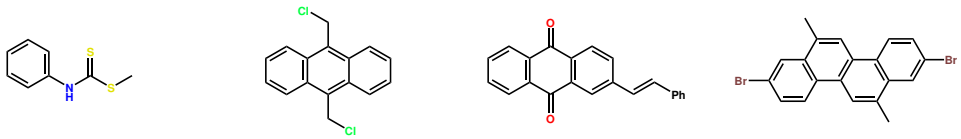
5

Figure 2.5: The pyrido[2,3-b]pyrido[3',2':4,5]pyrrolo[3,2-g]indole (**1**) molecule that the search space was based on in a study by Yang *et al.* and 2 selected molecules **4** and **5** with predicted crystal structure landscapes which were found to have the most favourable properties to form high-performance organic semiconductors [58].

In a study by Yang *et al.* a computational screening workflow was applied to a search space containing 27 different pyrrole-based azaphenacene isomers for organic semiconductors with high electron mobilities [58]. This search space was based around the molecule pyrido[2,3-b]pyrido[3',2':4,5]pyrrolo[3,2-g]indole (**1**) whose crystal structures were found to have π-stacked columns [59] which can increase the electronic coupling between molecular units. A single evaluation stage was carried out which took all 27 molecules through a CSP workflow with the quasi-random generation of crystal structures followed by minimisation.

Isotropic electron mobilities were calculated for each unique crystal structure within a $7\,\mathrm{kJ\,mol^{-1}}$ range from the global lattice energy minimum structure for each molecule using Marcus theory rates and the Eqs. (2.8). The authors report three measures to determine the molecules ability to produce an organic semiconductor with high electron mobilities: the electron mobilities of the global minimum structure, and the maximum and weighted average of the electron mobilities of the crystal structure landscape. From the results, they conclude the molecules **4** and **5** (Figure 2.5) were found to have the most promising properties to form high-performance organic semiconductors.

2.3.3 CSD Screening



S1 **S2** **S3** **S4**

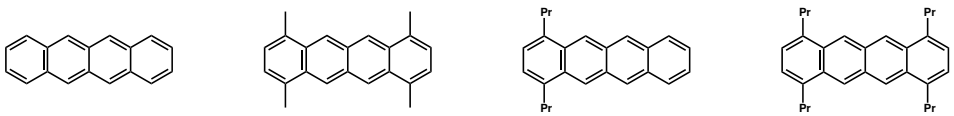
Figure 2.6: Four molecules that had not previously been considered for organic electronic applications that were proposed by a computational screening workflow over the Cambridge Structural Database (CSD) for crystal structures with the most favourable properties to form high-performance organic semiconductors [60].

In a study by Schober *et al.* a computational screening workflow was applied to the Cambridge Structural Database (CSD) [61] for organic semiconductors with high hole mobilities [60]. The CSD was first reduced to a total of 95,445 targets by excluding crystal structures that for example contained: organometallic molecules, more than one type of molecule, and structural disorder. A three-staged evaluation process was applied; in the first two stages, electronic coupling calculations were carried out with a more accurate method used in the second. In both stages, structures were carried forward only when one or more of its electronic couplings were calculated to have an absolute value greater than 50 meV. After the first two stages, the search space had been reduced to a total of 10,214 structures. In the final stage, a more in-depth evaluation was carried out which included the structures' reorganisation energies and percolation

pathways. From this, the authors determine four promising candidates shown in Figure 2.6 that have not previously been considered for organic electronic applications.

A few years later the CSD had been screened by Nematiaram *et al.* again for organic semiconductors with high hole mobilities [62]. In their approach, the CSD was first reduced to a search space of around 40,000 targets by using the data set from Ref. [63]. To obtain this set of targets, Padula *et al.* had first excluded crystal structures that had for example more than one type of molecule and molecules which contained more than 100 atoms. Further reductions are then made by selecting molecules with specific HOMO-LUMO gaps where for the final step they had selected molecules whose gaps were in a range between 2 and 4 eV to finally produce the set of around 40,000 target crystal structures. In the first screening stage, the number of crystal structures was reduced to a total of 4,801 by including only those with electronic couplings with absolute values of 0.1 eV or more. In the next stage, all 4,801 structures were evaluated using transient localisation theory [64, 65] and were grouped into three sets with low, intermediate, and high mobilities. A total of 40 molecules were contained in the high mobility group with 15 structures previously considered for carrier transport and another 10 for optoelectronic applications.

2.3.4 Tetracene Derivatives



TET TMT DPrT TPrT

Figure 2.7: The tetracene (TET) molecule which the search space was based on in a study by Ziogos *et al.* and the molecules of 3 candidate organic semiconductors, TMT, DPrT, and TPrT which showed high hole mobilities and/or allowed for charge carriers to perform *intercolumnar polaron diffusive jumps* [66].

In a study by Ziogos *et al.* a computational screening workflow was applied to 9 different tetracene (TET) derivatives for organic semiconductors with high hole

mobilities [66]. Here the search space was chosen due to the high experimental hole mobilities of the crystal structures of related molecules TET and rubrene with values of 2.4 and 19.0 $cm^2(Vs)^{-1}$ in the saturation regimes [67]. Additionally, the experimental crystal structures of the molecules in the chosen search space were also available in the CSD so that CSP searches were not required. Ziogos *et al.* used a single evaluation stage that applied the fragment orbital-based surface hopping (FOB-SH) method to calculate hole mobilities on the experimental crystal structures.

From all 9 molecules they highlight the molecules TMT, DPrT, and TPrT (Figure 2.7): TMT and TPrT because of their high hole mobilities which were calculated to be 20.8 and 15.8 $cm^2(Vs)^{-1}$, and TMT and DPrT because they had structures that allowed for hole carriers to perform *intercolumnar polaron diffusive jumps* [66]. TMT and DPrT have structures with π-stacked molecular columns with non-vanishing electronic couplings between molecules from neighbouring columns. This allows for charge carrier transport to occur through the molecular columns but also across columns through the *intercolumnar polaron diffusive jump* mechanism which they suggest will mean that the carrier mobilities will be more resilient to defects, such as molecular vacancies, within a given molecular column.

2.3.5 Summary

In this section, we have given a brief overview of a few recent examples of computational screening workflows for high-performance small-molecule organic semiconductors. Even between these five reports, there's quite a range of different strategies used. The search space for example varied widely from a space containing 7 targets to an entire database. Excluding the CSD screening studies, there is one common feature across all choices of the search space, which is that they all contain extended aromatic structures. This choice is based on experience as it is known that these aromatic structures are necessary to obtain large electronic couplings between molecular sites. With this fact, it is therefore unsurprising that the best molecules from both CSD searches typically contained these aromatic structures. Although in the CSD search by Schober *et*

al. the selected molecule **S1** from Figure 2.6 goes against this expectation. Unfortunately, the authors do not give any particular reason why other than the fact that its crystal structures contained a particularly high maximum electron coupling of 190 meV.

If we compare the CSD searches carried out by Schober *et al.* and Nematiaram *et al.* one interesting thing to note is that all four molecules proposed by Schober *et al.* had not shown up in the high mobilities list generated by Nematiaram *et al.* If we also compare both CSD searches with the work on the tetracene derivatives it is interesting to note that neither TMT nor TPrT was mentioned in either CSD screening studies despite the high hole mobilities obtained from FOB-SH calculations. Both structures were in the CSD as early as 2009 [68], several years before both CSD screening papers and should have been within their search spaces. It is not clear exactly why TMT and TPrT do not turn up in either CSD searches but it does suggest that perhaps there are some disagreements between the methods used to determine a crystal structures ability to form a high-performance organic semiconductor.

We can broadly divide the computational screening of organic semiconductors into two categories, those that select a search space containing molecules with reported experimental structures (*exclude-csp*) and those that do not so will require some sort of CSP stage (*include-csp*). This places the CSD searches and the tetracene derivatives studies in the first category *exclude-csp* and the others in the second *include-csp*. There are some significant advantages to the *exclude-csp* approach as synthetic routes for these molecules and their crystal structures are known. This allows the more costly CSP stage to be avoided and therefore more computational resources to be used on either a larger search space or on more accurate mobility evaluation methods.

The *include-csp* approach is however necessary for hypothetical molecules that have not been synthesised or for known molecules whose experimental crystal structures are not available. This approach can lead to a more high-risk scenario due to the lack of known experimental crystal structures leading to greater uncertainties in any methods used to determine the molecules performances. Significant challenges will therefore exist to the development of *include-csp* computational screening workflows due to higher computational costs, uncer-

tainties in the CSP results, and issues with the synthesizability and stability of hypothetical molecules.

Finally, one interesting point to take note of is the usage of the Eqs. (2.8) to obtain carrier mobilities using Marcus theory. The earliest appearance of these equations seems to have occurred in a paper by Deng *et al.* [55] who do not provide a derivation for them. Eqs. (2.8) therefore appears to have been constructed by an educated guess and the validity of these equations have been questioned previously [69, 70]. Additionally, the hopping probability term is particularly unusual given that Marcus theory describes transition probabilities or equivalently decay probabilities which are related to the mean lifetime of a given state since Marcus theory rates are themselves an approximation of Fermi's golden rule. In Section 3.5.1 we derive a diffusion equation Eq. (3.173) from the Kubo equation which unlike Eqs. (2.8) gives an orientational dependence of the mobilities and follows more closely the equations used in the kinetic Monte Carlo simulations.

Chapter 3

Theoretical Foundations

3.1 Electronic-Structure Theory

Electronic-structure theory [71–73] involves the study of the mathematical methods and techniques used to solve the quantum mechanical equations of a many-body system of electrons and nuclei. A number of approximations are taken usually to reduce the computational complexity of the equations while preserving accuracy. In this section, we will outline some electronic-structure theory methods used to obtain an approximate solution to the electronic Hamiltonian

$$\begin{aligned} \hat{H}_{\mathrm{el}} &= \hat{H}_{\mathrm{e}} + \hat{V}_{\mathrm{nn}} \\ \hat{H}_{\mathrm{e}} &= \hat{T}_{\mathrm{e}} + \hat{V}_{\mathrm{ne}} + \hat{V}_{\mathrm{ee}} \end{aligned} \tag{3.1}$$

with

$$\begin{aligned} \hat{T}_{\mathrm{e}} &= -\frac{1}{2}\sum_{i} \hat{\boldsymbol{\nabla}}_i^2 & \hat{V}_{\mathrm{ne}} &= -\sum_{iK} \frac{Z_K}{|\mathbf{r}_i - \mathbf{r}_K|} \\ \hat{V}_{\mathrm{ee}} &= \sum_{i,j>i} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} & \hat{V}_{\mathrm{nn}} &= \sum_{K,L>K} \frac{Z_K Z_L}{|\mathbf{r}_K - \mathbf{r}_L|} \end{aligned} \tag{3.2}$$

where $\hat{T}_{\mathrm{e}}$, $\hat{V}_{\mathrm{ne}}$, $\hat{V}_{\mathrm{ee}}$, and $\hat{V}_{\mathrm{nn}}$ are the electron kinetic, nuclear-electron potential, electron-electron potential, and nuclear-nuclear potential energy operators in

atomic units with the nuclear positions fixed in space. From the variational principle

$$E_0 = \min_{\psi} E[\psi] \qquad \psi_0 = \arg\min_{\psi} E[\psi] \tag{3.3}$$

under the constraint that $\langle\psi|\psi\rangle = 1$, where E_0 and ψ_0 are the ground-state energy and wavefunction and $E[\psi]$ is energy functional

$$E[\psi] = \langle\psi|\hat{H}_{\mathrm{e}}|\psi\rangle + V_{\mathrm{nn}} \tag{3.4}$$

since we are working with fixed nuclei the nuclear-nuclear potential energy V_{nn} will be a constant. So to obtain an approximate solution for the ground-state energy and wavefunction we must find the wavefunction that minimises the energy functional Eq. (3.4).

3.1.1 Hartree-Fock

The Hartree-Fock (HF) method uses a single Slater determinant to approximate the wavefunction for a system of electrons

$$\psi_{\mathrm{HF}} = |\phi_i \cdots \phi_N\rangle = \frac{1}{\sqrt{N!}} \det|\phi_i \cdots \phi_N| \tag{3.5}$$

where N is the number of electrons and the functions ϕ form an orthonormal set of molecular orbitals which have both a spatial and spin component. The method of Lagrange multipliers is applied to the minimisation of the energy functional with this wavefunction

$$E[\{\phi\}] = \langle\phi_i \cdots \phi_N|\hat{H}_{\mathrm{e}}|\phi_i \cdots \phi_N\rangle + V_{\mathrm{nn}} \tag{3.6}$$

subject to the constraint that $\langle\phi_i|\phi_j\rangle = \delta_{ij}$ leading to the HF equations [71]

$$\hat{f}(\mathbf{x})\phi_p(\mathbf{x}) = \varepsilon_p\phi_p(\mathbf{x}) \qquad p = i, \cdots, N \tag{3.7}$$

where ε_p are the orbital energies and $\hat{f}(\mathbf{x})$ is the Fock operator

$$\hat{f}(\mathbf{x}) = \hat{h}(\mathbf{r}) + \sum_{i}^{N} \hat{j}_i(\mathbf{x}) - \hat{k}_i(\mathbf{x}) \tag{3.8}$$

where

$$\hat{h}(\mathbf{r}_1) = -\frac{1}{2}\hat{\boldsymbol{\nabla}}_1^2 - \sum_{K} \frac{Z_K}{|\mathbf{r}_1 - \mathbf{r}_K|} \tag{3.9}$$

$$\hat{j}_i(\mathbf{x}_1) = \int \mathrm{d}^4\mathbf{x}_2 \, \phi_i^*(\mathbf{x}_2) \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} \phi_i(\mathbf{x}_2) \tag{3.10}$$

$$\hat{k}_i(\mathbf{x}_1) = \int \mathrm{d}^4\mathbf{x}_2 \, \phi_i^*(\mathbf{x}_2) \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} \hat{\mathcal{P}}_{12}\phi_i(\mathbf{x}_2) \tag{3.11}$$

so that $\hat{h}(\mathbf{x})$ is the orbital kinetic and nuclear-electron potential energy operator, $\hat{j}_i(\mathbf{x})$ and $\hat{k}_i(\mathbf{x})$ are the orbital coulomb and exchange operators and $\hat{\mathcal{P}}_{12}$ is an operator which exchanges the coordinates 1 and 2.

For closed-shell systems whose orbitals are occupied by a pair of electrons with opposite spins, we can simplify Eq. (3.7) to Eq. (3.11) by integrating away the spin coordinates and changing the summation to run over the spatial orbitals to obtain the restricted closed-shell Hartree-Fock (RHF) equations

$$\hat{f}(\mathbf{r})\varphi_p(\mathbf{r}) = \varepsilon_p\varphi_p(\mathbf{r}) \qquad\qquad p = i', \cdots, N/2 \tag{3.12}$$

$$\hat{f}(\mathbf{r}) = \hat{h}(\mathbf{r}) + \sum_{i'}^{N/2} 2\hat{j}_{i'}(\mathbf{r}) - \hat{k}_{i'}(\mathbf{r}) \tag{3.13}$$

where

$$\hat{h}(\mathbf{r}_1) = -\frac{1}{2}\hat{\boldsymbol{\nabla}}_1^2 - \sum_K \frac{Z_K}{|\mathbf{r}_1 - \mathbf{r}_K|} \tag{3.14}$$

$$\hat{j}_{i'}(\mathbf{r}_1) = \int \mathrm{d}^3\mathbf{r}_2 \, \varphi_{i'}^*(\mathbf{r}_2) \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} \varphi_{i'}(\mathbf{r}_2) \tag{3.15}$$

$$\hat{k}_{i'}(\mathbf{r}_1) = \int \mathrm{d}^3\mathbf{r}_2 \, \varphi_{i'}^*(\mathbf{r}_2) \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} \hat{\mathcal{P}}_{12} \varphi_{i'}(\mathbf{r}_2) \tag{3.16}$$

and $\varphi_{i'}(\mathbf{r})$ is the spatial component of the spin-orbitals $\phi_i(\mathbf{x}) = \varphi_{i'}(\mathbf{r})\alpha(\omega)$ and $\phi_j(\mathbf{x}) = \varphi_{i'}(\mathbf{r})\beta(\omega)$. The RHF equations can be solved by constructing the orbitals from a linear combination of functions

$$\varphi_{i'}(\mathbf{r}) = \sum_\mu C_{\mu i'} g_\mu(\mathbf{r}) \tag{3.17}$$

where for molecular calculations $g_\mu(\mathbf{r})$ are usually formed with a sum of Gaussian type functions. The coefficients are obtained by solving the Roothaan equations

$$\mathbf{FC} = \mathbf{SC}\boldsymbol{\varepsilon} \tag{3.18}$$

$$F_{\mu\nu} = \langle g_\mu | \hat{f}(\mathbf{r}) | g_\nu \rangle \qquad\qquad S_{\mu\nu} = \langle g_\mu | g_\nu \rangle \tag{3.19}$$

iteratively since the Coulomb and exchange operators depend on the orbitals and therefore the coefficients $C_{\mu i'}$ themselves.

3.1.2 Post-Hartree-Fock

Post-Hartree-Fock methods approximate the wavefunction with a linear combination of HF Slater determinants at different excitations. Since the total number of HF excitations that can be formed can be extremely large, Post-Hartree-Fock methods take different approaches to truncate this expansion. We will briefly describe three Post-Hartree-Fock methods: configuration interaction, coupled-

cluster theory, and Møller-Plesset perturbation theory. In the configuration interaction method when all HF excitations are summed over the configuration interaction wavefunction can be written as

$$\psi_{\text{CI}} = c\psi_{\text{HF}} + \sum_{a}\sum_{i} c_i^a \psi_i^a + \sum_{a,b>a}\sum_{i,j>i} c_{ij}^{ab}\psi_{ij}^{ab} + \cdots \tag{3.20}$$

where ψ_i^a are HF wavefunctions with the orbital ϕ_i replaced with excited state orbital ϕ_a and the terms c_i^a are a set of coefficients. A truncation to this expansion to a given excitation level is then made, for example in configuration interaction singles and doubles (CISD) a truncation to only include single and doubly excited HF wavefunctions.

In the coupled-cluster theory methods, the wavefunction is made by applying an exponential operator onto the HF wavefunction

$$\psi_{\text{CC}} = e^{\hat{T}}\psi_{\text{HF}} = \left[1 + \hat{T} + \frac{1}{2!}\hat{T}^2 + \frac{1}{3!}\hat{T}^3 + \cdots\right]\psi_{\text{HF}} \tag{3.21}$$

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \cdots \tag{3.22}$$

$$\hat{T}_1\psi_{\text{HF}} = \sum_{a}\sum_{i} t_i^a \psi_i^a \qquad \hat{T}_2\psi_{\text{HF}} = \sum_{a,b>a}\sum_{i,j>i} t_{ij}^{ab}\psi_{ij}^{ab} \tag{3.23}$$

where $\hat{T}_1$ is the single excitation operator, $\hat{T}_2$ is the double excitation operator and t_i^a and t_{ij}^{ab} are their corresponding cluster amplitudes. The coupled-cluster method now applies the truncation onto the cluster operator Eq. (3.22) where for example in the coupled-cluster singles and doubles (CCSD) [74] method the expansion is truncated to double excitations. The truncation on the cluster operator means that CCSD for example partially includes higher excitation terms due to the exponential operator Eq. (3.21).

Møller-Plesset perturbation theory (MPPT) applies the Rayleigh-Schrödinger perturbation theory onto the HF reference by using a sum of Fock operators for the unperturbed Hamiltonian while using the difference between the full and

unperturbed Hamiltonian for the perturbation so that.

$$\hat{H}_{\mathrm{e}} = \hat{F} + \hat{\mathcal{V}} \tag{3.24}$$

$$\hat{F} = \sum_i \hat{f}(\mathbf{x}_i) \qquad \hat{\mathcal{V}} = \hat{H}_{\mathrm{e}} - \hat{F} = \hat{V}_{\mathrm{ee}} - \sum_i \hat{v}_{\mathrm{HF}}(\mathbf{x}_i) \tag{3.25}$$

$$\hat{v}_{\mathrm{HF}}(\mathbf{x}_i) = \sum_j \left[\hat{j}_j(\mathbf{x}_i) - \hat{k}_j(\mathbf{x}_i)\right] \tag{3.26}$$

The ground-state wavefunction can be written as a series of HF wavefunctions at different excitations

$$\psi_{\mathrm{MP}} = \psi_{\mathrm{HF}} + \sum_{a,b>a} \sum_{i,j>i} \frac{\langle \psi_{ij}^{ab} | \hat{V}_{\mathrm{ee}} | \psi_{\mathrm{HF}} \rangle}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b} \psi_{ij}^{ab} + \cdots \tag{3.27}$$

where ε_i are the orbital energies. Standard techniques from perturbation theory can then be used to obtain corrections to the energy or wavefunction to the required order of the perturbation.

3.2 Density-Functional Theory

Density-functional theory (DFT) provides an alternative approach to the electronic-structure methods described in the previous section. Of particular importance is its Kohn-Sham density-functional theory (KS-DFT) [75–79] variant which offers results with a quality better than the HF methods but at a computational cost smaller than or similar to the HF depending on the functional used. In this research project, we will be using KS-DFT to obtain ground-state geometries and electron densities for a number of organic molecules so that a set of atom centred multipoles can be created to model the interactions between molecules in the solid-state, and obtain electronic coupling parameters and reorganisation energies used to calculate carrier mobilities of a crystal structure.

In the foundations of DFT, the existence of a mapping between the external

potentials, ground-state wavefunctions, and ground-state electron densities is given by the first Hohenberg-Kohn theorem [78, 80]. In particular, for a non-degenerate ground-state, a one-to-one mapping exists between a set of external potentials (which differ from each other by no more than the addition of a constant) and the ground-state wavefunction and electron density. The existence of the one-to-one mapping implies that the ground-state wavefunction is itself a functional of the ground-state electron density. From this, we can rewrite Eq. (3.4) to state the functional dependence on the electron density by the wavefunction so that

$$E[n] = \langle \psi[n] | \hat{H}_{\mathrm{e}} | \psi[n] \rangle + V_{\mathrm{nn}} \tag{3.28}$$

where $\psi[n]$ is the ground-state wavefuntion functional and n is the electron density.

Eq. (3.28) is a density-functional which we can use to solve for the ground-state density and energy using the second Hohenberg-Kohn theorem [78, 80] which proved the existence of a variational principle between the ground-state density and energy

$$E_0 = \min_n E[n] \qquad\qquad n_0 = \arg\min_n E[n] \tag{3.29}$$

with the constraint that

$$\int \mathrm{d}^3\mathbf{r}\, n(\mathbf{r}) = N \tag{3.30}$$

where n_0 is the ground-state density and N is the number of electrons of the system. Although an energy functional of the electron density exists, only the nuclear-electron potential energy functional

$$V_{\mathrm{ne}}[n] = -\sum_K \int \mathrm{d}^3\mathbf{r}\, \frac{n(\mathbf{r}) Z_K}{|\mathbf{r} - \mathbf{r}_K|} \tag{3.31}$$

is known exactly. The remaining terms must be approximated. In most cases, a

classical electrostatic potential energy functional

$$V_{\text{elst}}[n] = \frac{1}{2}\int \mathrm{d}^3\mathbf{r}_1 \int \mathrm{d}^3\mathbf{r}_2 \, \frac{n(\mathbf{r}_1)n(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} \tag{3.32}$$

is introduced so that the remaining terms to be approximated are the kinetic energy and the correction to the electron-electron potential energy from the classical electrostatic potential energy.

3.2.1 Kohn-Sham DFT

In the KS-DFT approach, the kinetic energy functional for a non-interacting system is introduced

$$T_{\text{KS}}[n] = \langle \psi_{\text{KS}} | \hat{T}_{\text{e}} | \psi_{\text{KS}} \rangle \tag{3.33}$$

where ψ_{KS} is the wavefunction of a non-interacting system whose electron density is fixed so that it is equivalent to the electron density of the fully interacting system. All the remaining correction terms are then collected into the exchange and correlation energy functionals so that the energy functional can be written as.

$$E[n] = T_{\text{KS}}[n] + V_{\text{ne}}[n] + V_{\text{elst}}[n] + E_{\text{x}}[n] + E_{\text{c}}[n] + V_{\text{nn}} \tag{3.34}$$

The exchange and correlation energy functionals are defined by the equations [77]

$$E_{\text{x}}[n] = \langle \psi_{\text{KS}} | \hat{V}_{\text{ee}} | \psi_{\text{KS}} \rangle - V_{\text{elst}}[n] \tag{3.35}$$

$$E_{\text{c}}[n] = T[n] - T_{\text{KS}}[n] + V_{\text{ee}}[n] - V_{\text{elst}}[n] - E_{\text{x}}[n] \tag{3.36}$$

where $T[n]$ and $V_{\text{ee}}[n]$ are the exact kinetic and electron-electron potential energy functionals.

Similarly to the RHF equations, the restricted closed-shell Kohn-Sham density-functional theory (RKS) equations can be derived using the method of Lagrange multipliers for the minimisation of the energy functional under the constraint

that the KS spin-orbitals are orthonormal $\langle\phi_i|\phi_j\rangle = \delta_{ij}$ and by integrating away the spin coordinates and changing the summation to run over the spatial orbitals

$$\left\{\hat{h}(\mathbf{r}) + \hat{v}_{\text{elst}}(\mathbf{r}) + \hat{v}_{\text{x}}(\mathbf{r}) + \hat{v}_{\text{c}}(\mathbf{r})\right\}\varphi_p(\mathbf{r}) = \varepsilon_p\varphi_p(\mathbf{r}) \tag{3.37}$$

where $p = i, \cdots, N/2$ and $\hat{v}_{\text{elst}}(\mathbf{r})$, $\hat{v}_{\text{x}}(\mathbf{r})$, and $\hat{v}_{\text{c}}(\mathbf{r})$ are the electrostatic, exchange, and correlations potentials which are obtained by a functional derivative with respect to the density. So the electrostatic potential is given by a functional derivative of the classical electrostatic potential energy functional

$$\hat{v}_{\text{elst}}(\mathbf{r}_1) = \frac{\delta V_{\text{elst}}[n]}{\delta n} = \int \mathrm{d}^3\mathbf{r}_2 \, \frac{n(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} \tag{3.38}$$

while the exchange and correlation potentials are determined by a functional derivative of the approximate exchange or correlation functionals.

3.2.2 LDA Functionals

In the local-density approximation (LDA) the exchange-correlation functional is derived for a homogeneous electron gas. One example is the SVWN5 exchange-correlation functional which is formed from the Slater exchange [76, 78] and VWN5 correlation [78, 79, 81] which are given as

$$E_{\text{x}}^{\text{S}}[n] = \int \mathrm{d}^3\mathbf{r} \; n(\mathbf{r})\varepsilon_{\text{x}}^{\text{S}}(n(\mathbf{r})) \tag{3.39}$$

$$E_{\text{c}}^{\text{VWN}}[n] = \int \mathrm{d}^3\mathbf{r} \; n(\mathbf{r})\varepsilon_{\text{c}}^{\text{VWN}}(n(\mathbf{r})) \tag{3.40}$$

where $\varepsilon_{\mathrm{x}}^{\mathrm{S}}(n)$ and $\varepsilon_{\mathrm{c}}^{\mathrm{VWN}}(n)$ are the Slater exchange and VWN5 correlation energies per electron

$$\varepsilon_{\mathrm{x}}^{\mathrm{S}}(n) = -\alpha\frac{9}{8}\left(\frac{3}{\pi}n\right)^{\frac{1}{3}} \tag{3.41}$$

$$\varepsilon_{\mathrm{c}}^{\mathrm{VWN}}(n) = \frac{1-\ln 2}{\pi^2}\Bigg\{\ln\frac{x(n)^2}{X(x(n))} + \frac{2b}{Q}\tan^{-1}\frac{Q}{2x(n)+b} - \frac{bx_0}{X(x_0)}\Bigg[\ln\frac{[x(n)-x_0]^2}{X(x(n))} + \frac{2(b+2x_0)}{Q}\tan^{-1}\frac{Q}{2x(n)+b}\Bigg]\Bigg\} \tag{3.42}$$

with

$$X(x) = x^2 + bx + c \qquad Q = \sqrt{4c-b^2} \qquad x(n) = \left[\frac{3}{4\pi n}\right]^{\frac{1}{3}} \tag{3.43}$$

and $\alpha = 2/3$, $x_0 = -0.10498$, $b = 3.72744$ and $c = 12.9352$. Where the form of the Slater exchange was determined analytically while the VWN5 correlation functional was constructed to interpolate between the analytic high and low-density limits of the correlation energy for the homogeneous electron gas and fitted to Monte Carlo calculations.

3.2.3 GGA Functionals

Attempts to correct the LDA by including some description of an inhomogeneous system of electrons are made in the generalised gradient approximation (GGA) by including a dependence on both the electron density and its gradient. One example is the PBE exchange-correlation functional [78, 82]

$$E_{\mathrm{x}}^{\mathrm{PBE}}[n] = \int \mathrm{d}^3\mathbf{r}\, n(\mathbf{r})\varepsilon_{\mathrm{x}}^{\mathrm{LDA}}(n(\mathbf{r}))f(\xi(n(\mathbf{r}), \boldsymbol{\nabla} n(\mathbf{r}))) \tag{3.44}$$

$$E_{\mathrm{c}}^{\mathrm{PBE}}[n] = \int \mathrm{d}^3\mathbf{r}\, n(\mathbf{r})\left[\varepsilon_{\mathrm{c}}^{\mathrm{LDA}}(n(\mathbf{r})) + h(n(\mathbf{r}), t(n(\mathbf{r}), \boldsymbol{\nabla} n(\mathbf{r})))\right] \tag{3.45}$$

with

$$f(\xi) = 1 + \frac{a\xi}{1 + b\xi} \qquad \xi(n, \boldsymbol{\nabla} n) = \left[\frac{|\boldsymbol{\nabla} n|}{2k_{\mathrm{F}}(n)n}\right]^2 \tag{3.46}$$

and

$$h(n,t) = \gamma \ln\left(1 + \frac{\beta}{\gamma}\frac{t^2 + A(n)t^4}{1 + A(n)t^2 + A(n)^2t^4}\right) \tag{3.47}$$

$$t(n, \boldsymbol{\nabla} n) = \frac{|\boldsymbol{\nabla} n|}{2k_{\mathrm{TF}}(n)n} \qquad A(n) = \frac{\beta}{\gamma}\left[\exp\left(-\frac{\varepsilon_{\mathrm{c}}^{\mathrm{LDA}}(n)}{\gamma}\right) - 1\right]^{-1} \tag{3.48}$$

where $a = 0.21951$, $b = 0.2730$, $\beta = 3a/\pi^2$, $\gamma = (1 - \ln 2)/\pi^2$, $k_{\mathrm{F}}(n) = (3\pi^2 n)^{1/3}$ and $k_{\mathrm{TF}}(n) = (4k_{\mathrm{F}}/\pi)^{1/2}$, and $\varepsilon_{\mathrm{x}}^{\mathrm{LDA}}(n)$ and $\varepsilon_{\mathrm{c}}^{\mathrm{LDA}}(n)$ are LDA exchange and correlation energies per electron.

3.2.4 Hybrid Functionals

In the adiabatic connection, an interelectronic coupling-strength parameter λ is introduced which scales the electron-electron interaction, resulting in the Hamiltonian

$$\hat{H}_{\mathrm{e}}^{\lambda} = \hat{T}_{\mathrm{e}} + \hat{V}_{\mathrm{ne}} + \lambda\hat{V}_{\mathrm{ee}} \tag{3.49}$$

a λ-dependent wavefunction is then defined to be

$$\psi_n^{\lambda} = \arg\min_{\psi} \langle\psi|\hat{H}_{\mathrm{e}}^{\lambda}|\psi\rangle \tag{3.50}$$

with the constraint that

$$\langle\psi_n^{\lambda}|\hat{n}(\mathbf{r}_{\mathrm{e}}, \mathbf{r})|\psi_n^{\lambda}\rangle = n(\mathbf{r}) \qquad \hat{n}(\mathbf{r}_{\mathrm{e}}, \mathbf{r}) = \sum_{i=1}^{N}\delta(\mathbf{r}_i - \mathbf{r}) \tag{3.51}$$

where $\hat{n}(\mathbf{r}_{\mathrm{e}}, \mathbf{r})$ is the density operator and $\mathbf{r}_{\mathrm{e}}$ stands for the coordinates of the electrons. Therefore the electron density for the wavefunction ψ_n^{λ} has the electron

density $n(\mathbf{r})$ for all values of λ. From this, an expression for the exchange-correlation energy can be derived to give.

$$E_{\mathrm{xc}}[n] = \int_0^1 \mathrm{d}\lambda\, W^\lambda[n] \qquad\qquad W^\lambda[n] = \langle\psi_n^\lambda|\hat{V}_{\mathrm{ee}}|\psi_n^\lambda\rangle - V_{\mathrm{elst}}[n] \tag{3.52}$$

The integral in Eq. (3.52) was approximated by Becke [83] with a single trapezoid between the two endpoints so that

$$E_{\mathrm{xc}}[n] \approx \frac{1}{2}W^0[n] + \frac{1}{2}W^1[n] \tag{3.53}$$

where the $W^0[n]$ can be identified as the HF exchange energy over the KS orbitals.

$$W^0[n] = E_x^{\mathrm{HF}}[n] - \frac{1}{2}\sum_{ij}^{N}\int \mathrm{d}^3\mathbf{r}_1 \int \mathrm{d}^3\mathbf{r}_2\, \frac{\phi_i^*(\mathbf{r}_1)\phi_j^*(\mathbf{r}_2)\phi_j(\mathbf{r}_1)\phi_i(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} \tag{3.54}$$

Becke then approximated $W^1[n]$ with an approximate exchange-correlation functional leading to the half-and-half hybrid functionals.

$$E_{\mathrm{xc}}^{\mathrm{HH}}[n] = \frac{1}{2}E_{\mathrm{x}}^{\mathrm{HF}}[n] + \frac{1}{2}E_{\mathrm{xc}}^{\mathrm{DFT}}[n] \tag{3.55}$$

An empirical generalisation of the half-and-half hybrid functionals includes different fractions of the HF exchange and the DFT exchange and correlations energies, for example, the B3LYP functional [84, 85]

$$\begin{aligned} E_{\mathrm{xc}}^{\mathrm{B3LYP}}[n] =& E_{\mathrm{xc}}^{\mathrm{LDA}}[n] + a_0\left(E_{\mathrm{x}}^{\mathrm{HF}}[n] - E_{\mathrm{x}}^{\mathrm{LDA}}[n]\right) + a_{\mathrm{x}}\left(E_{\mathrm{x}}^{\mathrm{B88}}[n] - E_{\mathrm{x}}^{\mathrm{LDA}}[n]\right) \\ &+ a_{\mathrm{c}}\left(E_{\mathrm{c}}^{\mathrm{LYP}}[n] - E_{\mathrm{c}}^{\mathrm{LDA}}[n]\right) \end{aligned} \tag{3.56}$$

where $E_{\mathrm{xc}}^{\mathrm{LDA}}[n]$ is the LDA functional SVWN3 or SVWN5 depending on the implementation, $E_{\mathrm{x}}^{\mathrm{B88}}[n]$ is the Becke 1988 functional [86], $E_{\mathrm{c}}^{\mathrm{LYP}}[n]$ is the Lee, Yang and Parr functional [87], and the empirical parameters are $a_0 = 0.2$, $a_{\mathrm{x}} = 0.72$ and $a_{\mathrm{c}} = 0.81$.

3.3 Force Fields

In CSP we approximate the lattice energy of a crystal structure by modelling the intermolecular interactions between rigid molecules using a series of two-body atom-atom potentials. In this project, we use the Buckingham (exp-6) force field [88] to model the atom-atom interactions which decompose them into the exchange-repulsion, dispersion, and electrostatic energy terms. Using this model the interaction between the atoms K and R for example takes the form

$$V_{KR} = A_{KR}\exp(-B_{KR}r_{KR}) - C_{KR}r_{KR}^{-6} + E_{KR}^{\mathrm{DMA}} \tag{3.57}$$

where A_{KR}, B_{KR}, and C_{KR} are fitting parameters for a specific pair of atom types and $r_{KR} = |\mathbf{r}_K - \mathbf{r}_R|$. The term $A_{KR}\exp(-B_{KR}r_{KR})$ models the exchange repulsion, $-C_{KR}r_{KR}^{-6}$ models the dispersion interaction, and E_{KR}^{DMA} models electrostatic interaction by using a series of multipoles centred at the atoms K and R.

The lattice energy can be approximated as a sum of the exchange-repulsion, dispersion, and electrostatic energies

$$U = \sum_{\mathcal{A},\mathcal{B}>\mathcal{A}} E_{\mathcal{A}\mathcal{B}}^{\mathrm{exch}} + E_{\mathcal{A}\mathcal{B}}^{\mathrm{disp}} + E_{\mathcal{A}\mathcal{B}}^{\mathrm{elst}} \tag{3.58}$$

where $E_{\mathcal{A}\mathcal{B}}^{\mathrm{exch}}$, $E_{\mathcal{A}\mathcal{B}}^{\mathrm{disp}}$, and $E_{\mathcal{A}\mathcal{B}}^{\mathrm{elst}}$ are the exchange-repulsion, dispersion, and electrostatic energies between the molecules $\mathcal{A}$ and $\mathcal{B}$. The summation of $\mathcal{A}$ runs over molecules of a single unit cell while $\mathcal{B}$ runs over all molecules in the crystal. The indexes that $\mathcal{A}$ run over are smaller than the indexes of the molecules in all other unit cells so inequality $\mathcal{B} > \mathcal{A}$ ensures that we do not double count the interaction between two molecules in the same unit cell. The approximation in Eq. 3.58 with the atom-atom potential model Eq. 3.57 excludes a number of other interaction terms such as the induction and higher-order mixed exchange, induction, and dispersion interactions, although these terms are partially included through the various fitting parameters used in the force field.

The decomposition of the interaction energy to obtain the dispersion and electrostatic energy terms are made by analysing terms in a perturbation expansion

for large molecule-molecule separations known as the polarisation approximation. The exchange-repulsion is a short-range effect that is not included in the polarisation approximation and requires more sophisticated energy decomposition methods such as the symmetry-adapted perturbation theory (SAPT) [89]. A basic expression for the exchange-repulsion between two atoms can be obtained by considering the differences in the energy of a Hartree product wavefunction and the antisymmetrised Slater determinant. In the following subsections, we will outline some of the theoretical justifications for a force field with the form Eq. (3.57).

3.3.1 Polarisation Approximation

In the polarisation approximation [90] we assume that the supermolecular wavefunction can be formed from a product of isolated molecule wavefunctions

$$\psi^{(0)}_{000\cdots} = \prod_{\mathcal{A}} \psi^{(0)}_{\mathcal{A}0} \tag{3.59}$$

where the superscript denotes the order of the expansion in perturbation theory and the subscripts are the state of the molecular wavefunctions. So that $\psi^{(0)}_{000\cdots}$ is the zeroth-order supermolecular wavefunction with $000\cdots$ denoting that the wavefunction is formed from the molecular wavefunctions $\mathcal{ABC}\cdots$ which are all in the ground state and $\psi^{(0)}_{\mathcal{A}0}$ is the wavefunction of the isolated molecule $\mathcal{A}$ in its ground state.

For the intermolecular interactions between the molecules of a crystal structure, the zeroth-order Hamiltonian will be simply the sum of the Hamiltonians of each isolated molecule while the first-order correction will be all the remaining interacting terms between each pair of molecules.

$$\hat{H}^{(0)} = \sum_{\mathcal{A}} \hat{H}^{(0)}_{\mathcal{A}} \qquad\qquad \hat{H}^{(1)} = \sum_{\mathcal{A},\mathcal{B}>\mathcal{A}} \hat{H}^{(1)}_{\mathcal{AB}} \tag{3.60}$$

$$\hat{H}^{(0)}_{\mathcal{A}} = -\frac{1}{2}\sum_i \hat{\boldsymbol{\nabla}}^2_{\mathcal{A}i} + \sum_{K,L>K} \frac{Z_{\mathcal{A}K} Z_{\mathcal{A}L}}{|\mathbf{r}_{\mathcal{A}K} - \mathbf{r}_{\mathcal{A}L}|} - \sum_{iK} \frac{Z_{\mathcal{A}K}}{|\mathbf{r}_{\mathcal{A}i} - \mathbf{r}_{\mathcal{A}I}|} + \sum_{i,j>i} \frac{1}{|\mathbf{r}_{\mathcal{A}i} - \mathbf{r}_{\mathcal{A}j}|} \tag{3.61}$$

$$\hat{H}^{(1)}_{\mathcal{A}\mathcal{B}} = \sum_{KR} \frac{Z_{\mathcal{A}K} Z_{\mathcal{B}R}}{|\mathbf{r}_{\mathcal{A}K} - \mathbf{r}_{\mathcal{B}R}|} - \sum_{pK} \frac{Z_{\mathcal{A}K}}{|\mathbf{r}_{\mathcal{B}p} - \mathbf{r}_{\mathcal{A}K}|} - \sum_{iR} \frac{Z_{\mathcal{B}R}}{|\mathbf{r}_{\mathcal{A}i} - \mathbf{r}_{\mathcal{B}R}|} + \sum_{ip} \frac{1}{|\mathbf{r}_{\mathcal{A}i} - \mathbf{r}_{\mathcal{B}p}|} \tag{3.62}$$

The zeroth-order energy from perturbation theory is

$$E^{(0)}_{000\cdots} = \langle \psi^{(0)}_{000\cdots} | \hat{H}^{(0)} | \psi^{(0)}_{000\cdots} \rangle = \sum_{\mathcal{A}} E^{(0)}_{\mathcal{A}0} \tag{3.63}$$

$$E^{(0)}_{\mathcal{A}0} = \langle \psi^{(0)}_{\mathcal{A}0} | \hat{H}^{(0)}_{\mathcal{A}} | \psi^{(0)}_{\mathcal{A}0} \rangle \tag{3.64}$$

so that it is simply a sum of all isolated molecule energies. The first-order correction to the energy is

$$E^{(1)}_{000\cdots} = \langle \psi^{(0)}_{000\cdots} | \hat{H}^{(1)} | \psi^{(0)}_{000\cdots} \rangle = \sum_{\mathcal{A},\mathcal{B}>\mathcal{A}} E^{\mathrm{elst}}_{\mathcal{A}\mathcal{B}} \tag{3.65}$$

$$E^{\mathrm{elst}}_{\mathcal{A}\mathcal{B}} = \langle \psi^{(0)}_{\mathcal{A}0} \psi^{(0)}_{\mathcal{B}0} | \hat{H}^{(1)}_{\mathcal{A}\mathcal{B}} | \psi^{(0)}_{\mathcal{A}0} \psi^{(0)}_{\mathcal{B}0} \rangle \tag{3.66}$$

which is simply a sum of all two-body classical electrostatic interactions between the nuclei point charges and electron density of one molecule with the nuclei point charges and electron density of another.

The second-order correction to the energy is

$$\begin{aligned} E^{(2)}_{000\cdots} = &\sum_{m\neq 0} \frac{|\langle\psi^{(0)}_{000\cdots}|\hat{H}^{(1)}|\psi^{(0)}_{m00\cdots}\rangle|^2}{E^{(0)}_{000\cdots} - E^{(0)}_{m00\cdots}} + \sum_{n\neq 0} \frac{|\langle\psi^{(0)}_{000\cdots}|\hat{H}^{(1)}|\psi^{(0)}_{0n0\cdots}\rangle|^2}{E^{(0)}_{000\cdots} - E^{(0)}_{0n0\cdots}} + \cdots \\ &+ \sum_{\substack{m\neq 0 \\ n\neq 0}} \frac{|\langle\psi^{(0)}_{000\cdots}|\hat{H}^{(1)}|\psi^{(0)}_{mn0\cdots}\rangle|^2}{E^{(0)}_{000\cdots} - E^{(0)}_{mn0\cdots}} + \sum_{\substack{m\neq 0 \\ o\neq 0}} \frac{|\langle\psi^{(0)}_{000\cdots}|\hat{H}^{(1)}|\psi^{(0)}_{m0o\cdots}\rangle|^2}{E^{(0)}_{000\cdots} - E^{(0)}_{m0o\cdots}} + \cdots \end{aligned} \tag{3.67}$$

the terms involving excited states of only one molecular wavefunction are the induction interactions between that molecule and the ground state electron densities and nuclear point charges of all other molecules. Terms involving excited states of two molecular wavefunctions are the dispersion interaction between those two molecules. Terms with three or more excited state molecular wavefunctions are zero so that to second-order there are no many-body interactions. The second-order correction to the energy can be rewritten as

$$E^{(2)}_{000\cdots} = \sum_{\mathcal{A},\mathcal{B}>\mathcal{A}} E^{\text{ind}}_{\mathcal{AB}} + E^{\text{disp}}_{\mathcal{AB}} \tag{3.68}$$

$$\begin{aligned} E^{\text{ind}}_{\mathcal{AB}} = &- \sum_{m\neq 0} \frac{|\langle\psi^{(0)}_{\mathcal{A}0}\psi^{(0)}_{\mathcal{B}0}|\hat{H}^{(1)}_{\mathcal{AB}}|\psi^{(0)}_{\mathcal{A}m}\psi^{(0)}_{\mathcal{B}0}\rangle|^2}{E^{(0)}_{\mathcal{A}m} - E^{(0)}_{\mathcal{A}0}} \\ &- \sum_{n\neq 0} \frac{|\langle\psi^{(0)}_{\mathcal{A}0}\psi^{(0)}_{\mathcal{B}0}|\hat{H}^{(1)}_{\mathcal{AB}}|\psi^{(0)}_{\mathcal{A}0}\psi^{(0)}_{\mathcal{B}n}\rangle|^2}{E^{(0)}_{\mathcal{B}n} - E^{(0)}_{\mathcal{B}0}} \end{aligned} \tag{3.69}$$

$$E^{\text{disp}}_{\mathcal{AB}} = - \sum_{\substack{m\neq 0 \\ n\neq 0}} \frac{|\langle\psi^{(0)}_{\mathcal{A}0}\psi^{(0)}_{\mathcal{B}0}|\hat{H}^{(1)}_{\mathcal{AB}}|\psi^{(0)}_{\mathcal{A}m}\psi^{(0)}_{\mathcal{B}n}\rangle|^2}{E^{(0)}_{\mathcal{A}m} - E^{(0)}_{\mathcal{A}0} + E^{(0)}_{\mathcal{B}n} - E^{(0)}_{\mathcal{B}0}} \tag{3.70}$$

where we have defined $E^{\text{ind}}_{\mathcal{AB}}$ as the induction energy of $\mathcal{A}$ due to the electric potential of $\mathcal{B}$ plus the induction energy of $\mathcal{B}$ due to the electric potential of $\mathcal{A}$.

3.3.2 Distributed Multipoles

Assuming we have the electron densities of each isolated molecule in our molecular crystal we can obtain the electrostatic potential in principle by carrying out the integration Eq. (3.66). However, we instead take a more computationally efficient approach by approximating it using a distributed multipole expansion. First we Taylor expand the electron coordinates of molecule $\mathcal{A}$ around an origin $\mathbf{r}_{\mathcal{O}}$ and the electron coordinates of molecule $\mathcal{B}$ around an origin $\mathbf{r}_{\mathcal{O}'}$

$$\begin{aligned}\hat{H}^{(1)}_{\mathcal{AB}}(\mathbf{r}_{\mathcal{O}},\mathbf{r}_{\mathcal{O}'}) = \sum_{KR}\sum_{\alpha\beta\cdots} & T_{KR}Z_KZ_R + T_{K\mathcal{O}'}Z_K\hat{q}_{\mathcal{B}} + T_{\mathcal{O}R}\hat{q}_{\mathcal{A}}Z_R + T_{\mathcal{OO}'}\hat{q}_{\mathcal{A}}\hat{q}_{\mathcal{B}} \\ & + T_{K\mathcal{O}'\alpha}Z_K\hat{\mu}_{\mathcal{B}\alpha}(\mathbf{r}_{\mathcal{O}'}) - T_{\mathcal{O}R\alpha}\hat{\mu}_{\mathcal{A}\alpha}(\mathbf{r}_{\mathcal{O}})Z_R + T_{\mathcal{OO}'\alpha}\big[\hat{q}_{\mathcal{A}}\hat{\mu}_{\mathcal{B}\alpha}(\mathbf{r}_{\mathcal{O}'}) \\ & - \hat{\mu}_{\mathcal{A}\alpha}(\mathbf{r}_{\mathcal{O}})\hat{q}_{\mathcal{B}}\big] + \tfrac{1}{3}T_{K\mathcal{O}'\alpha\beta}Z_K\hat{\theta}_{\mathcal{B}\alpha\beta}(\mathbf{r}_{\mathcal{O}'}) + \tfrac{1}{3}T_{\mathcal{O}R\alpha\beta}\hat{\theta}_{\mathcal{A}\alpha\beta}(\mathbf{r}_{\mathcal{O}})Z_R \\ & + T_{\mathcal{OO}'\alpha\beta}\big[\tfrac{1}{3}\hat{q}_{\mathcal{A}}\hat{\theta}_{\mathcal{B}\alpha\beta}(\mathbf{r}_{\mathcal{O}'}) - \hat{\mu}_{\mathcal{B}\alpha}(\mathbf{r}_{\mathcal{O}})\hat{\mu}_{\mathcal{B}\beta}(\mathbf{r}_{\mathcal{O}'}) + \tfrac{1}{3}\hat{\theta}_{\mathcal{A}\alpha\beta}(\mathbf{r}_{\mathcal{O}})\hat{q}_{\mathcal{B}}\big] + \cdots\end{aligned} \tag{3.71}$$

where the charge monopole, dipole, and quadrupole operators of molecule $\mathcal{A}$ are

$$\begin{aligned}\hat{q}_{\mathcal{A}} &= -\hat{n}_{\mathcal{A}} \qquad\qquad \hat{\mu}_{\mathcal{A}\alpha}(\mathbf{r}_{\mathcal{O}}) = -\sum_i (\mathbf{r}_i - \mathbf{r}_{\mathcal{O}})_\alpha \\ \hat{\theta}_{\mathcal{A}\alpha\beta}(\mathbf{r}_{\mathcal{O}}) &= -\sum_i \frac{3}{2}(\mathbf{r}_i - \mathbf{r}_{\mathcal{O}})_\alpha(\mathbf{r}_i - \mathbf{r}_{\mathcal{O}})_\beta - \frac{1}{2}|\mathbf{r}_i - \mathbf{r}_{\mathcal{O}}|^2\delta_{\alpha\beta}\end{aligned} \tag{3.72}$$

where $\hat{n}_{\mathcal{A}}$ is the number operator, and $\mathbf{r}_i$ are the coordinates of the electrons of molecule $\mathcal{A}$ and the tensors are.

$$\begin{aligned}T_{\mathcal{OO}'} &= \frac{1}{|\mathbf{r}_{\mathcal{O}'} - \mathbf{r}_{\mathcal{O}}|} \qquad\qquad T_{\mathcal{OO}'\alpha} = -\frac{(\mathbf{r}_{\mathcal{O}'} - \mathbf{r}_{\mathcal{O}})_\alpha}{|\mathbf{r}_{\mathcal{O}'} - \mathbf{r}_{\mathcal{O}}|^3} \\ T_{\mathcal{OO}'\alpha\beta} &= \frac{3(\mathbf{r}_{\mathcal{O}'} - \mathbf{r}_{\mathcal{O}})_\alpha(\mathbf{r}_{\mathcal{O}'} - \mathbf{r}_{\mathcal{O}})_\beta - |\mathbf{r}_{\mathcal{O}'} - \mathbf{r}_{\mathcal{O}}|^2\delta_{\alpha\beta}}{|\mathbf{r}_{\mathcal{O}'} - \mathbf{r}_{\mathcal{O}}|^5}\end{aligned} \tag{3.73}$$

We can partition the integration for the electrostatic energy so that

$$E^{\text{elst}}_{\mathcal{AB}} \approx E^{\text{DMA}}_{\mathcal{AB}} = \sum_{\mathcal{OO}'}\langle\psi^{(0)}_{\mathcal{A}0}\psi^{(0)}_{\mathcal{B}0}|\hat{H}^{(1)}_{\mathcal{AB}}(\mathbf{r}_{\mathcal{O}},\mathbf{r}_{\mathcal{O}'})\hat{w}_{\mathcal{AO}}\hat{w}_{\mathcal{BO}'}|\psi^{(0)}_{\mathcal{A}0}\psi^{(0)}_{\mathcal{B}0}\rangle \tag{3.74}$$

where $\hat{w}_{\mathcal{A}\mathcal{O}} = \prod_i \hat{w}_{\mathcal{O}i}(\mathbf{r}_i)$ is a weight function that partitions the electron coordinates of molecule $\mathcal{A}$. The origins for the Taylor expansions of Eq. (3.71) are then chosen depending on these partitions. Since the electron densities of a molecule are highly peaked around its nuclei we will consider the case where each partition is centred around a given nucleus and set the associated origin of the Taylor series expansions at that nuclei.

$$E^{\mathrm{DMA}}_{\mathcal{A}\mathcal{B}} = \sum_{KR} \langle \psi^{(0)}_{\mathcal{A}0} \psi^{(0)}_{\mathcal{B}0} | \hat{H}^{(1)}_{\mathcal{A}\mathcal{B}}(\mathbf{r}_K, \mathbf{r}_R) \hat{w}_{\mathcal{A}K} \hat{w}_{\mathcal{B}R} | \psi^{(0)}_{\mathcal{A}0} \psi^{(0)}_{\mathcal{B}0} \rangle \tag{3.75}$$

Inserting Eq. (3.71) into Eq. (3.75) we obtain an approximation of the electrostatic energy as a summation of the interactions between the atom centred multipoles

$$\begin{aligned} E^{\mathrm{DMA}}_{\mathcal{A}\mathcal{B}} = \sum_{KR} \sum_{\alpha\beta\cdots} & T_{KR} z_K z_R + T_{KR\alpha} \left[z_K \mu_{R\alpha} - \mu_{K\alpha} z_R \right] \\ & + T_{KR\alpha\beta} \left[\tfrac{1}{3} z_K \theta_{R\alpha\beta} - \mu_{K\alpha} \mu_{R\beta} + \tfrac{1}{3} \theta_{K\alpha\beta} z_R \right] + \cdots \end{aligned} \tag{3.76}$$

which is usually truncated to only include terms that are for example of order r^{-n} with $n \leq 5$. The monopole, dipole, and quadrupole moments are obtained from the integrals.

$$\begin{aligned} z_K &= Z_K + q_K & q_K &= \langle \psi^{(0)}_{\mathcal{A}0} | \hat{q}_{\mathcal{A}} \hat{w}_{\mathcal{A}K} | \psi^{(0)}_{\mathcal{A}0} \rangle \\ \mu_{K\alpha} &= \langle \psi^{(0)}_{\mathcal{A}0} | \hat{\mu}_{\mathcal{A}\alpha}(\mathbf{r}_K) \hat{w}_{\mathcal{A}K} | \psi^{(0)}_{\mathcal{A}0} \rangle & \theta_{K\alpha\beta} &= \langle \psi^{(0)}_{\mathcal{A}0} | \hat{\theta}_{\mathcal{A}\alpha\beta}(\mathbf{r}_K) \hat{w}_{\mathcal{A}K} | \psi^{(0)}_{\mathcal{A}0} \rangle \end{aligned} \tag{3.77}$$

Alternatively, the electrostatic interaction Eq. (3.76) and the multipole moments Eq. (3.77) can be rewritten into a spherical-tensor form [90]. In Section 4.1 we describe a method implemented in the program GDMA that can be used to partition the electron density and obtain atom centred spherical-tensor multipoles from a single Slater determinant wavefunction with orbitals formed from a linear combination of Gaussian type functions. While in Section 4.3 we describe the calculation of the electrostatic interaction between spherical-tensor multipoles of two molecules by using a system of global and local molecular axis frames in the program DMACRYS.

3.3.3 Distributed Polarisabilities

The induction energies can be evaluated using Eq. (3.69) which we will approximate by carrying out a multipole expansion of the interaction and partition the integration around nuclei centres similar to the methodology described in Subsection 3.3.2 for the electrostatic interaction so that

$$E_{\mathcal{AB}}^{\text{ind}} \approx -\sum_{m\neq 0}(E_{\mathcal{A}m}^{(0)} - E_{\mathcal{A}0}^{(0)})^{-1} \sum_{KRLS} \Big\{ \langle \psi_{\mathcal{A}0}^{(0)} \psi_{\mathcal{B}0}^{(0)} | \hat{H}_{\mathcal{AB}}^{(1)}(\mathbf{r}_K, \mathbf{r}_R) \hat{w}_{\mathcal{A}K} \hat{w}_{\mathcal{B}R} | \psi_{\mathcal{A}m}^{(0)} \psi_{\mathcal{B}0}^{(0)} \rangle \times \langle \psi_{\mathcal{A}m}^{(0)} \psi_{\mathcal{B}0}^{(0)} | \hat{H}_{\mathcal{AB}}^{(1)}(\mathbf{r}_L, \mathbf{r}_S) \hat{w}_{\mathcal{A}L} \hat{w}_{\mathcal{B}S} | \psi_{\mathcal{A}0}^{(0)} \psi_{\mathcal{B}0}^{(0)} \rangle \Big\} + \cdots \tag{3.78}$$

where the first term of the induction interaction between the molecules $\mathcal{A}$ and $\mathcal{B}$ describes the many-body interaction between two atomic multipoles at the atoms R and S on the molecule $\mathcal{B}$ and the non-local polarisability of atoms K and L on the molecule $\mathcal{A}$. To obtain an expression that is only formed from two-body interactions we must ignore all terms where $K \neq L$ and $R \neq S$. The first significant term to the induction energy is

$$E_{\mathcal{AB}}^{\text{ind}} \approx -\sum_{KR}\sum_{\alpha\beta} T_{KR\alpha} T_{KR\beta} \left[q_{\mathcal{A}K} q_{\mathcal{A}K} \alpha_{\mathcal{B}RR\alpha\beta} + \alpha_{\mathcal{A}KK\alpha\beta} q_{\mathcal{B}R} q_{\mathcal{B}R} \right] \tag{3.79}$$

$$\alpha_{\mathcal{A}KK\alpha\beta} = \sum_{m\neq 0} \frac{\langle \psi_{\mathcal{A}0}^{(0)} | \hat{\mu}_{\mathcal{A}\alpha}(\mathbf{r}_K) \hat{w}_{\mathcal{A}K} | \psi_{\mathcal{A}m}^{(0)} \rangle \langle \psi_{\mathcal{A}m}^{(0)} | \hat{\mu}_{\mathcal{A}\beta}(\mathbf{r}_K) \hat{w}_{\mathcal{A}K} | \psi_{\mathcal{A}0}^{(0)} \rangle}{E_{\mathcal{A}m}^{(0)} - E_{\mathcal{A}0}^{(0)}} \tag{3.80}$$

where $\alpha_{\mathcal{A}KK\alpha\beta}$ is the dipole-dipole polarisability of the atom K.

Similarly, we can approximate the dispersion interaction with a multipole type expansion of the interacting Hamiltonian and partition the integrations around nuclei centres. By ignoring all terms where $K \neq L$ and $R \neq S$ we obtain an expression involving only the local polarisabilities where the first significant

term to the dispersion energy is

$$\begin{aligned} E_{\mathcal{AB}}^{\text{disp}} \approx -\sum_{\substack{m\neq 0\\ n\neq 0}} (E_{\mathcal{A}m}^{(0)} - E_{\mathcal{A}0}^{(0)} + E_{\mathcal{B}n}^{(0)} - E_{\mathcal{B}0}^{(0)})^{-1} \sum_{KR} \sum_{\alpha\beta\gamma\delta} \Big\{ & T_{KR\alpha\beta} T_{KR\gamma\delta} \\ & \times \langle \psi_{\mathcal{A}0}^{(0)} | \hat{\mu}_{\mathcal{A}\alpha}(\mathbf{r}_K) \hat{w}_{\mathcal{A}K} | \psi_{\mathcal{A}m}^{(0)} \rangle \langle \psi_{\mathcal{A}m}^{(0)} | \hat{\mu}_{\mathcal{A}\beta}(\mathbf{r}_K) \hat{w}_{\mathcal{A}R} | \psi_{\mathcal{A}0}^{(0)} \rangle \\ & \times \langle \psi_{\mathcal{B}0}^{(0)} | \hat{\mu}_{\mathcal{B}\gamma}(\mathbf{r}_R) \hat{w}_{\mathcal{B}K} | \psi_{\mathcal{B}n}^{(0)} \rangle \langle \psi_{\mathcal{B}n}^{(0)} | \hat{\mu}_{\mathcal{B}\delta}(\mathbf{r}_R) \hat{w}_{\mathcal{B}R} | \psi_{\mathcal{B}0}^{(0)} \rangle \Big\} \end{aligned} \tag{3.81}$$

and can rewrite this expression using the identity

$$\frac{1}{x+y} = \frac{2}{\pi} \int_0^\infty \mathrm{d}z \, \frac{x}{x^2+z^2} \, \frac{y}{y^2+z^2} \tag{3.82}$$

so that

$$E_{\mathcal{AB}}^{\text{disp}} \approx -\frac{1}{2\pi} \sum_{KR} \sum_{\alpha\beta\gamma\delta} T_{KR\alpha\beta} T_{KR\gamma\delta} \int_0^\infty \mathrm{d}\omega \; \alpha_{\mathcal{A}KK\alpha\beta}(-i\omega; i\omega) \alpha_{\mathcal{B}RR\gamma\delta}(-i\omega; i\omega) \tag{3.83}$$

where $\alpha_{\mathcal{A}KK\alpha\beta}(-i\omega; i\omega)$ is a component of the frequency-dependent polarisability of the atom K on molecule $\mathcal{A}$ at imaginary frequencies.

$$\begin{aligned} \alpha_{\mathcal{A}KK\alpha\beta}(-i\omega; i\omega) = & 2 \sum_{m\neq 0} \frac{E_{\mathcal{A}m}^{(0)} - E_{\mathcal{A},0}^{(0)}}{(E_{\mathcal{A}m}^{(0)} - E_{\mathcal{A}0}^{(0)})^2 + \omega^2} \langle \psi_{\mathcal{A}0}^{(0)} | \hat{\mu}_{\mathcal{A}\alpha}(\mathbf{r}_K) \hat{w}_{\mathcal{A}K} | \psi_{\mathcal{A}m}^{(0)} \rangle \\ & \times \langle \psi_{\mathcal{A}m}^{(0)} | \hat{\mu}_{\mathcal{A}\beta}(\mathbf{r}_K) \hat{w}_{\mathcal{A}R} | \psi_{\mathcal{A}0}^{(0)} \rangle \end{aligned} \tag{3.84}$$

To obtain the isotropic part of the dispersion energy we use an averaged frequency-dependent polarisability

$$\bar{\alpha}_{\mathcal{A}KK} = \frac{1}{3}(\alpha_{\mathcal{A}KK\alpha\alpha} + \alpha_{\mathcal{A}KK\beta\beta} + \alpha_{\mathcal{A}KK\gamma\gamma}) \tag{3.85}$$

for each atom in both molecules, the dispersion interaction now becomes.

$$\begin{aligned} E_{\mathcal{AB}}^{\text{disp}} &\approx -\frac{1}{2\pi}\sum_{KR}\sum_{\alpha\beta} T_{KR\alpha\beta}T_{KR\alpha\beta}\int_0^\infty \mathrm{d}\omega\ \bar{\alpha}_{\mathcal{A}KK}(-i\omega;i\omega)\bar{\alpha}_{\mathcal{B}RR}(-i\omega;i\omega) \\ &\approx -\sum_{KR}\frac{C_{6,KR}}{|\mathbf{r}_R-\mathbf{r}_K|^6} \end{aligned} \tag{3.86}$$

Finally obtaining the $-r^{-6}$ expression for the dispersion energy where the $C_{6,KR}$ coefficient is obtained from the integral.

$$C_{6,KR} = \frac{3}{\pi}\int_0^\infty \mathrm{d}\omega\ \bar{\alpha}_{\mathcal{A}KK}(-i\omega;i\omega)\bar{\alpha}_{\mathcal{B}RR}(-i\omega;i\omega) \tag{3.87}$$

3.3.4 Exchange-Repulsion

To obtain the purely repulsive part of the exchange-repulsion we will work in the limit of zero electron-electron interactions so that exchange stabilisation effects are not included. The contribution of the repulsive part of the exchange-repulsion internally to a molecule for this system could be defined as

$$E_{\text{exch}} = \langle\psi|\hat{T}_{\text{e}} + \hat{V}_{\text{ne}}|\psi\rangle - \langle\psi_{\text{H}}|\hat{T}_{\text{e}} + \hat{V}_{\text{ne}}|\psi_{\text{H}}\rangle \tag{3.88}$$

where ψ is a Slater determinant and ψ_{H} is the Hartree product. We will want to find some function of E_{exch} between two molecules to build a force field. We can get a basic idea of what this could be by considering two hydrogen atoms with no electron-electron interactions. The ground state energy for the Hartree product wavefunction of this hydrogen molecule will simply be twice the energy of the ground state molecular orbital.

$$\langle\psi_{\text{H}}|\hat{T}_{\text{e}} + \hat{V}_{\text{ne}}|\psi_{\text{H}}\rangle = 2E_0 \tag{3.89}$$

When creating ψ we will need to force both electrons to be spin-parallel so that we obtain an expression of the exchange-repulsion that models the interaction between two closed-shell systems rather than the bonding case when

they are anti-parallel. For the hydrogen molecule of a system of non-interacting spin-parallel electrons, the wavefunction will be a Slater determinate made from the ground state and the next excited state molecular orbitals. The total energy will be the sum of the energies of these two orbitals

$$\langle\psi|\hat{T}_{\mathrm{e}} + \hat{V}_{\mathrm{ne}}|\psi\rangle = E_0 + E_1 \tag{3.90}$$

the intermolecular exchange-repulsion energy is therefore

$$E_{\mathrm{exch}} = E_1 - E_0 = \Delta E \tag{3.91}$$

where ΔE is the difference between the ground state and excited state molecular orbitals and has been determined analytically at long ranges to be [91].

$$\Delta E = 4e^{-r-1}\left[r + \frac{1}{2} + O\left(r^{-1}\right)\right] \tag{3.92}$$

The exponential is the dominant term at long ranges and is one of the theoretical justifications for the Born-Mayer model for the repulsive interaction which takes the form

$$V(r) = Ae^{-Br} \tag{3.93}$$

where A and B are fitting parameters. Since Eq. (3.92) was obtained for energy between two atoms, a reasonable approximation to the repulsive energy between two molecules could be formed from a summation of two-body atomic potentials in the form of Eq. (3.93).

3.4 The Spin-Boson Model

In this project, we calculated the charge mobility using a hopping model which assumes that the charge carriers will be localised to a given molecule and travels across the crystal structure by a series of hops from one molecule to another. We can derive a model with this type of action by starting from the spin-boson

model [92]

$$\begin{aligned}\hat{H}'_{\text{SB}} = &\sum_n E'_n|n\rangle\langle n| + \sum_{n,n'\neq n} V_{nn'}|n\rangle\langle n'| \\ &+ \sum_{n\alpha} \hbar\omega_\alpha V_{n\alpha}(\hat{a}^\dagger_\alpha + \hat{a}_\alpha)|n\rangle\langle n| + \hat{H}_{\text{B}} \\ \hat{H}_{\text{B}} = &\sum_\alpha \hbar\omega_\alpha \left(\hat{a}^\dagger_\alpha \hat{a}_\alpha + \frac{1}{2}\right)\end{aligned} \tag{3.94}$$

where $|n\rangle$ is a diabatic charge localised wavefunction with its charge density centred on the molecule at a site n, E'_n is the energy for the occupation of the state $|n\rangle$, $V_{nn'}$ is the electronic coupling between states $|n\rangle$ and $|n'\rangle$, ω_α is the nuclear vibrational frequency of the mode α, $V_{n\alpha}$ is the vibronic coupling between the states $|n,\nu_\alpha\rangle$ and $|n,\nu'_\alpha\rangle$ where $|n,\nu_\alpha\rangle$ is a product of the diabatic charge localised wavefunction $|n\rangle$ and the nuclear vibrational wavefunction $|\nu_\alpha\rangle$ of vibrational mode α, and $\hat{a}^\dagger_\alpha$ and $\hat{a}_\alpha$ are the creation and annihilation operators in the vibrational mode α.

We can carry out a transformation onto Eqs. (3.94) to remove the coupling between the site occupation of the charge carrier and the nuclear vibrational modes to obtain

$$\hat{H}_{\text{SB}} = \sum_n E_n|n\rangle\langle n| + \sum_{n,n'\neq n} V_{nn'} e^{\hat{\Omega}_{nn'}}|n\rangle\langle n'| + \hat{H}_{\text{B}} \tag{3.95}$$

where

$$\begin{aligned}&\hat{H}_{\text{SB}} = \hat{H}_0 + \hat{V} && \hat{H}_0 = \hat{H}_{\text{D}} + \hat{H}_{\text{B}} \\ &\hat{H}_{\text{D}} = \sum_n E_n|n\rangle\langle n| && E_n = E'_n - \sum_\alpha \hbar\omega_\alpha V^2_{n\alpha} \\ &\hat{V} = \sum_{n,n'\neq n} \hat{V}_{nn'} && \hat{V}_{nn'} = V_{nn'} e^{\hat{\Omega}_{nn'}}|n\rangle\langle n'| \\ &\hat{\Omega}_{nn'} = \hat{\Omega}_n - \hat{\Omega}_{n'} && \hat{\Omega}_n = \sum_\alpha V_{n\alpha}(\hat{a}^\dagger_\alpha - \hat{a}_\alpha)\end{aligned} \tag{3.96}$$

this transformed Hamiltonian now describes a dressed quantum state with a stabilised site occupation energy E_n and a transformed coupling operator $\hat{V}_{nn'}$ which now describes the coupling between two dressed quantum states $|n, \boldsymbol{\nu}\rangle$ and $|n', \boldsymbol{\nu}'\rangle$ where $|n, \boldsymbol{\nu}\rangle$ is a product of the diabatic charge localised wavefunction $|n\rangle$ and the nuclear vibrational wavefunction $|\boldsymbol{\nu}\rangle = \prod_\alpha |\nu_\alpha\rangle$. In the next sections, we will derive the spin-boson model from the Schrödinger equation of nuclei and electrons and give exact definitions of all the parameters used and highlight some of the approximations that are made to obtain the spin-boson model.

3.4.1 Derivation

First, we must define the parameters in the spin-boson model and the nuclear vibrational and electronic wavefunctions that will be used. We start from the Hamiltonian for a system of nuclei and electrons for the neutral system

$$\begin{aligned} \hat{H}(\mathbf{r}_\mathrm{e}, \mathbf{r}_\mathrm{n}) &= \hat{T}_\mathrm{n}(\mathbf{r}_\mathrm{n}) + \hat{H}_\mathrm{el}(\mathbf{r}_\mathrm{el}, \mathbf{r}_\mathrm{n}) \\ \hat{H}_\mathrm{el}(\mathbf{r}_\mathrm{e}, \mathbf{r}_\mathrm{n}) &= \hat{H}_\mathrm{e}(\mathbf{r}_\mathrm{e}, \mathbf{r}_\mathrm{n}) + \hat{V}_\mathrm{nn}(\mathbf{r}_\mathrm{n}) \\ \hat{H}_\mathrm{e}(\mathbf{r}_\mathrm{e}, \mathbf{r}_\mathrm{n}) &= \hat{T}_\mathrm{e}(\mathbf{r}_\mathrm{e}) + \hat{V}_\mathrm{ne}(\mathbf{r}_\mathrm{e}, \mathbf{r}_\mathrm{n}) + \hat{V}_\mathrm{ee}(\mathbf{r}_\mathrm{e}) \end{aligned} \tag{3.97}$$

with

$$\begin{aligned} \hat{T}_\mathrm{n}(\mathbf{r}_\mathrm{n}) &= -\sum_K \frac{\hbar^2}{2m_K} \hat{\boldsymbol{\nabla}}_K^2 & \hat{T}_\mathrm{e}(\mathbf{r}_\mathrm{e}) &= -\frac{\hbar^2}{2m_\mathrm{e}} \sum_i \hat{\boldsymbol{\nabla}}_i^2 \\ \hat{V}_\mathrm{ne}(\mathbf{r}_\mathrm{e}, \mathbf{r}_\mathrm{n}) &= -\frac{e^2}{4\pi\epsilon_0} \sum_{iK} \frac{Z_K}{|\mathbf{r}_i - \mathbf{r}_K|} & \hat{V}_\mathrm{ee}(\mathbf{r}_\mathrm{e}) &= \frac{e^2}{4\pi\epsilon_0} \sum_{i,j>i} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} \\ \hat{V}_\mathrm{nn}(\mathbf{r}_\mathrm{n}) &= \frac{e^2}{4\pi\epsilon_0} \sum_{K,L>K} \frac{Z_K Z_L}{|\mathbf{r}_K - \mathbf{r}_L|} \end{aligned} \tag{3.98}$$

where $\mathbf{r}_\mathrm{e}$ and $\mathbf{r}_\mathrm{n}$ is used to signify that the operators are dependent on the coordinates of the electrons and/or nuclei. The nuclear vibrational wavefunctions are then obtained from an effective nuclear Hamiltonian in the Born-Oppenheimer

approximation [93] with the electrons in its ground state

$$\begin{aligned}\hat{H}(\mathbf{r}_\mathrm{n}) &= \hat{T}_\mathrm{n}(\mathbf{r}_\mathrm{n}) + \hat{V}_\mathrm{n}(\mathbf{r}_\mathrm{n}) \\ \hat{V}_\mathrm{n}(\mathbf{r}_\mathrm{n}) &= \langle\psi_0|\hat{H}_\mathrm{e}(\mathbf{r}_\mathrm{e},\mathbf{r}_\mathrm{n})|\psi_0\rangle + \hat{V}_\mathrm{nn}(\mathbf{r}_\mathrm{n})\end{aligned} \tag{3.99}$$

where $|\psi_0\rangle$ is the adiabatic ground state electronic wavefunction. Using standard methods we can approximate Eq. (3.99) with the vibrational Hamiltonian [93,94]

$$\begin{aligned}\hat{H}_\mathrm{vib}(\mathbf{x}_\mathrm{n}) &= \hat{H}_\mathrm{B}(\mathbf{x}_\mathrm{n}) + \hat{V}_\mathrm{n}(\mathbf{0}) \\ \hat{H}_\mathrm{B}(\mathbf{x}_\mathrm{n}) &= -\sum_\alpha \frac{\hbar^2}{2\mu_\alpha}\frac{\partial^2}{\partial x_\alpha^2} + \frac{1}{2}\sum_\alpha \mu_\alpha\omega_\alpha^2\hat{x}_\alpha^2\end{aligned} \tag{3.100}$$

where $\mathbf{x}_\mathrm{n}$ is used to signify that the operators depend on the coordinates of the nuclei in a normal mode coordinate system, the origin $\mathbf{0}$ is set to the energy minimum of the harmonic potential, and μ_α is the reduced mass and ω_α is the vibrational frequency of the vibrational mode α. So $\hat{H}_\mathrm{B}$ will be equivalent to a Hamiltonian of an N-dimensional harmonic oscillator

$$\hat{H}_\mathrm{B}|\boldsymbol{\nu}\rangle = E_{\boldsymbol{\nu}}|\boldsymbol{\nu}\rangle \tag{3.101}$$

with eigenfunctions $|\boldsymbol{\nu}\rangle = \prod_\alpha |\nu_\alpha\rangle$ which is a product of harmonic oscillator wavefunctions for each vibrational mode α.

The electronic wavefunctions are obtained by solving the Schrödinger equation for the charged system as a function of the nuclear coordinates

$$\hat{H}_\mathrm{e}^\pm(\mathbf{r}_\mathrm{e},\mathbf{r}_\mathrm{n})|\xi_i\rangle = E_i(\mathbf{r}_\mathrm{n})|\xi_i\rangle \tag{3.102}$$

where $\hat{H}_\mathrm{e}^\pm(\mathbf{r}_\mathrm{e},\mathbf{r}_\mathrm{n})$ is a charged Hamiltonian and $|\xi_i\rangle$ are its eigenfunctions which are a set of adiabatic states which have a parametric dependence on the nuclear coordinates so that for an example nuclear-electronic wavefunction

$$\langle\xi_i,\boldsymbol{\nu}|\hat{T}_\mathrm{n}(\mathbf{r}_\mathrm{n})|\xi_j,\boldsymbol{\nu}'\rangle = \langle\boldsymbol{\nu}|\hat{T}_\mathrm{n}(\mathbf{r}_\mathrm{n}) + \hat{\Lambda}_{ij}(\mathbf{r}_\mathrm{n})|\boldsymbol{\nu}'\rangle \tag{3.103}$$

where [95].

$$\hat{\Lambda}_{ij}(\mathbf{r}_\mathrm{n}) = -\sum_K \frac{\hbar^2}{2m_K}\left[2\langle\xi_i|\hat{\boldsymbol{\nabla}}_K\xi_j\rangle\cdot\hat{\boldsymbol{\nabla}}_K + \langle\xi_i|\hat{\boldsymbol{\nabla}}_K^2\xi_j\rangle\right]$$
$$\langle\xi_i|\hat{\boldsymbol{\nabla}}_K^2\xi_j\rangle = \sum_k \langle\xi_i|\hat{\boldsymbol{\nabla}}_K\xi_k\rangle\cdot\langle\xi_k|\hat{\boldsymbol{\nabla}}_K\xi_j\rangle + \hat{\boldsymbol{\nabla}}_K\cdot\langle\xi_i|\hat{\boldsymbol{\nabla}}_K\xi_j\rangle \tag{3.104}$$

An orthogonal transformation is then carried out on the states $|\xi_i\rangle$ to form a set of diabatic states $|n\rangle$

$$|n\rangle = \sum_i U_{in}(\mathbf{r}_\mathrm{n})|\xi_i\rangle \tag{3.105}$$

where the elements of the matrix $\mathbf{U}(\mathbf{r}_\mathrm{n})$ are selected so that $\langle n|\hat{\boldsymbol{\nabla}}_K n'\rangle = \mathbf{0}$ for all electronic states and nuclear configurations.

A generic wavefunction can be formed from a linear combination of products of the electronic and nuclear wavefunctions using the basis sets from Eq. (3.101) for the nuclear and Eq. (3.105) for the electronic coordinates

$$|\Psi\rangle = \sum_{n\boldsymbol{\nu}} c_{n\boldsymbol{\nu}}|n,\boldsymbol{\nu}\rangle \tag{3.106}$$

which can be used to solve the charged nuclear-electronic Hamiltonian.

$$\hat{H}^\pm(\mathbf{r}_\mathrm{e},\mathbf{r}_\mathrm{n}) = \hat{T}_\mathrm{n}(\mathbf{r}_\mathrm{n}) + \hat{H}_\mathrm{e}^\pm(\mathbf{r}_\mathrm{e},\mathbf{r}_\mathrm{n}) + \hat{V}_\mathrm{nn}(\mathbf{r}_\mathrm{n}) \tag{3.107}$$

We can simplify this Hamiltonian by first adding and subtracting the harmonic potential of the neutral system Eq. (3.99) so that

$$\begin{aligned}\hat{H}^\pm(\mathbf{r}_\mathrm{e},\mathbf{r}_\mathrm{n}) = &-\sum_\alpha \frac{\hbar^2}{2m_\alpha}\frac{\partial^2}{\partial r_\alpha^2} + \frac{1}{2}\sum_{\alpha\beta} H_{\alpha\beta}\hat{r}_\alpha\hat{r}_\beta + \hat{V}_\mathrm{nn}(\mathbf{r}_\mathrm{n}) \\ &- \frac{1}{2}\sum_{\alpha\beta} H_{\alpha\beta}\hat{r}_\alpha\hat{r}_\beta + \hat{H}_\mathrm{e}^\pm(\mathbf{r}_\mathrm{e},\mathbf{r}_\mathrm{n})\end{aligned} \tag{3.108}$$

and restrict the nuclear coordinates to use a normal mode coordinate system so that translational and rotational motions are removed and the harmonic

potentials are diagonal.

$$\hat{H}^{\pm}(\mathbf{r}_{\mathrm{e}}, \mathbf{x}_{\mathrm{n}}) = \hat{H}_{\mathrm{B}}(\mathbf{x}_{\mathrm{n}}) + \hat{V}_{\mathrm{nn}}(\mathbf{x}_{\mathrm{n}}) - \frac{1}{2}\sum_{\alpha} \mu_{\alpha}\omega_{\alpha}^{2}\hat{x}_{\alpha}^{2} + \hat{H}_{\mathrm{e}}^{\pm}(\mathbf{r}_{\mathrm{e}}, \mathbf{x}_{\mathrm{n}}) \tag{3.109}$$

To solve this Hamiltonian using the wavefunction in Eq. (3.106) would require evaluating the following types of integrals

$$\langle n, \boldsymbol{\nu}| \left[\hat{H}_{\mathrm{B}}(\mathbf{x}_{\mathrm{n}}) + \hat{V}_{\mathrm{nn}}(\mathbf{x}_{\mathrm{n}}) - \frac{1}{2}\sum_{\alpha} \mu_{\alpha}\omega_{\alpha}^{2}\hat{x}_{\alpha}^{2} + \hat{H}_{\mathrm{e}}^{\pm}(\mathbf{r}_{\mathrm{e}}, \mathbf{x}_{\mathrm{n}}) \right] |n', \boldsymbol{\nu}'\rangle \tag{3.110}$$

so that by carrying out the integrals over the electronic coordinates we obtain

$$\langle \boldsymbol{\nu}| S_{nn'}\hat{H}_{\mathrm{B}}(\mathbf{x}_{\mathrm{n}}) + \hat{E}_{nn'}(\mathbf{x}_{\mathrm{n}})|\boldsymbol{\nu}'\rangle \tag{3.111}$$

where

$$\hat{E}_{nn'}(\mathbf{x}_{\mathrm{n}}) = S_{nn'} \left[\hat{V}_{\mathrm{nn}}(\mathbf{x}_{\mathrm{n}}) - \frac{1}{2}\sum_{\alpha} \mu_{\alpha}\omega_{\alpha}^{2}\hat{x}_{\alpha}^{2} \right] + \langle n|\hat{H}_{\mathrm{e}}^{\pm}(\mathbf{r}_{\mathrm{e}}, \mathbf{x}_{\mathrm{n}})|n'\rangle \tag{3.112}$$

with $S_{nn'} = \langle n|n'\rangle$.

We can simplify this equation by Taylor expanding the correction to $\hat{H}_{\mathrm{B}}$ around the nuclear coordinates at $\mathbf{0}$

$$\hat{E}_{nn'}(\mathbf{x}_{\mathrm{n}}) = \hat{E}_{nn'}(\mathbf{0}) + \sum_{\alpha} \left.\frac{\partial \hat{E}_{nn'}}{\partial x_{\alpha}}\right|_{\mathbf{0}} \hat{x}_{\alpha} + O(\hat{x}^{2}) \tag{3.113}$$

truncating to first-order and rewriting the position operator in terms of creation and annihilation operators to obtain

$$\hat{E}_{nn'}(\mathbf{x}_{\mathrm{n}}) \approx E_{nn'} + \sum_{\alpha} \hbar\omega_{\alpha} V_{nn'\alpha}(\hat{a}_{\alpha}^{\dagger} + \hat{a}_{\alpha}) \tag{3.114}$$

where we have defined the zeroth-order term as an occupation energy/electronic coupling parameter and the first-order derivative evaluated at the origin as a

vibronic coupling parameter.

$$E_{nn'} = \hat{E}_{nn'}(\mathbf{0}) \qquad \hbar\omega_\alpha V_{nn'\alpha} = \sqrt{\frac{\hbar}{2\mu_\alpha\omega_\alpha}}\frac{\partial \hat{E}_{nn'}}{\partial x_\alpha}\bigg|_{\mathbf{0}} \tag{3.115}$$

Once we have obtained these parameters we can rewrite an element of the nuclear-electronic Hamiltonian as

$$\hat{H}^{\pm}_{nn'} \approx E_{nn'} + \sum_\alpha \hbar\omega_\alpha V_{nn'\alpha}(\hat{a}^\dagger_\alpha + \hat{a}_\alpha) + S_{nn'}\hat{H}_{\text{B}} \tag{3.116}$$

and if we assume that $S_{nn'} = \delta_{nn'}$ we can rewrite this to obtain.

$$\hat{H}^{\pm} \approx \sum_{nn'} E_{nn'}|n\rangle\langle n'| + \sum_{nn'\alpha} \hbar\omega_\alpha V_{nn'\alpha}(\hat{a}^\dagger_\alpha + \hat{a}_\alpha)|n\rangle\langle n'| + \hat{H}_{\text{B}} \tag{3.117}$$

If we take the Condon approximation we assume that the electronic couplings are independent of the nuclear coordinates so that we can disregard the non-local vibronic coupling terms and obtain the spin-boson model

$$\begin{aligned}\hat{H}'_{\text{SB}} = \sum_n E'_n|n\rangle\langle n| + \sum_{n,n'\neq n} V_{nn'}|n\rangle\langle n'| \\ + \sum_{n\alpha} \hbar\omega_\alpha V_{n\alpha}(\hat{a}^\dagger_\alpha + \hat{a}_\alpha)|n\rangle\langle n| + \hat{H}_{\text{B}}\end{aligned} \tag{3.118}$$

where we have split the $E_{nn'}$ into a diagonal $E_{nn} = E'_n$ occupation energy and off-diagonal $E_{nn'} = V_{nn'}$ electronic coupling and dropped an index for the diagonal vibronic coupling term so that $V_{n\alpha} = V_{nn\alpha}$. The parameters for the spin-boson model can be obtained from electronic-structure calculations using the equations

$$E'_n = \langle n|\hat{H}^{\pm}_{\text{e}}(\mathbf{r}_{\text{e}}, \mathbf{0})|n\rangle + \hat{V}_{\text{nn}}(\mathbf{0}) \tag{3.119}$$

$$V_{nn'} = \langle n|\hat{H}^{\pm}_{\text{e}}(\mathbf{r}_{\text{e}}, \mathbf{0})|n'\rangle \tag{3.120}$$

$$V_{n\alpha} = \frac{1}{\hbar\omega_\alpha}\sqrt{\frac{\hbar}{2\mu_\alpha\omega_\alpha}}\left\{\frac{\partial\langle n|\hat{H}^{\pm}_{\text{e}}(\mathbf{r}_{\text{e}}, \mathbf{x}_{\text{n}})|n\rangle}{\partial x_\alpha}\bigg|_{\mathbf{0}} + \frac{\partial \hat{V}_{\text{nn}}(\mathbf{x}_{\text{n}})}{\partial x_\alpha}\bigg|_{\mathbf{0}}\right\} \tag{3.121}$$

where we have assumed that $S_{nn'} = \delta_{nn'}$.

3.4.2 Electronic Coupling

To evaluate Eq. (3.119) to Eq. (3.121), a set of diabatic charged localised states are required which can be obtained by an orthogonal transformation of the adiabatic states Eq. (3.105). However, we will want to simplify the problem first by truncating the number of states $|\xi_i\rangle$ so that for example there is only one state per molecule.

$$|n\rangle = \sum_{i}^{N_{\text{mol}}} U_{in}(\mathbf{r}_{\text{n}})|\xi_i\rangle \tag{3.122}$$

This truncation will, unfortunately, mean that a transformation $\mathbf{U}(\mathbf{r}_{\text{n}})$ cannot be found which gives a set of diabatic states that have the property that $\langle n|\hat{\boldsymbol{\nabla}}_K n'\rangle = \mathbf{0}$ for all electronic states and nuclear configurations [96]. Instead, we must try to obtain a set of diabatic (quasi-diabatic) states where the derivative couplings $\langle n|\hat{\boldsymbol{\nabla}}_K n'\rangle$ are small.

One example is the Boys localisation procedure which uses a transformation matrix that maximises the localisation sum

$$f(\mathbf{U}) = \sum_{nn'}^{N_{\text{mol}}} |\langle n|\hat{\boldsymbol{\mu}}|n\rangle - \langle n'|\hat{\boldsymbol{\mu}}|n'\rangle|^2 \tag{3.123}$$

where $\hat{\boldsymbol{\mu}}$ is the dipole operator. The Boys localisation procedure therefore will find a set of diabatic states $|n\rangle$ so that their charge centres will be as far away from each other as possible. Using this method Subotnik *et al.* describes a simple case for the diabatisation of the four lowest energy adiabatic states obtained from a CASSCF(7,8) calculation using the 6-31G* basis set for a supermolecular cation of four Helium atoms (He_4^+) and showed that the Boys localisation procedure produced a set of four diabatic states with the same energies but with each state having its charge centres close to one of the He nucleus [97].

A good question to ask is whether the states obtained from the Boys localisation procedure are actually diabatic. Fatehi *et al.* showed that in an example case

for a *p*-benzoquinone molecule, with distorted geometries so that its CIS/6-31G** adiabatic excited states are closely spaced, the derivative coupling between the adiabatic states were found to be extremely large as expected. By applying the Boys localisation procedure to obtain a set of diabatic states they then showed that the derivative coupling between the diabatic states to be near zero proving that at least for this one example the Boys localisation procedure does indeed result in a set of diabatic states [98].

A method closely related to Boys localisation [97] known as the generalised Mulliken-Hush (GMH) method [99, 100] an equation for the electronic coupling for a two-state system is given in terms of only adiabatic parameters

$$|V_{nn'}| \approx \frac{|\boldsymbol{\mu}_{ij} \cdot \mathbf{v}_0||E_i - E_j|}{\sqrt{|\boldsymbol{\mu}_{ii} - \boldsymbol{\mu}_{jj}|^2 + 4(\boldsymbol{\mu}_{ij} \cdot \mathbf{v}_0)^2}} \tag{3.124}$$

where

$$\boldsymbol{\mu}_{ij} = \langle \xi_i | \hat{\boldsymbol{\mu}} | \xi_j \rangle \qquad \mathbf{v}_0 = \frac{\boldsymbol{\mu}_{ii} - \boldsymbol{\mu}_{jj}}{|\boldsymbol{\mu}_{ii} - \boldsymbol{\mu}_{jj}|} \tag{3.125}$$

which assumes that there exists a transformation that generates a set of diabatic states so that $\boldsymbol{\mu}_{ab} \cdot \mathbf{v}_0 = 0$. Although the GMH can be extended for multiple states it is, however, difficult to extend this to systems with more than two sites as it is not clear what the vector $\mathbf{v}_0$ should be for non-linear systems [97]. Therefore Eq. (3.124) will need to be carried out by going through isolated dimers to approximate the electronic couplings for a system with multiple sites.

Another method known as the fragment molecular orbital density-functional theory (FO-DFT) [101] takes a different approach and generates diabatic states from isolated molecule DFT orbitals. In the $2N \pm 1$ variant of FO-DFT diabatic states are created by orthogonalising the orbitals of the two isolated molecules

and using them to generate a Slater determinant of $2N \pm 1$ spin-orbitals.

$$
\begin{aligned}
|n\rangle &= \frac{1}{\sqrt{(2N \pm 1)!}} \det|\tilde{\phi}_i \tilde{\phi}_j \cdots \tilde{\phi}_{N\pm1} \tilde{\phi}_{i'} \tilde{\phi}_{j'} \cdots \tilde{\phi}_{N'}| \\
|n'\rangle &= \frac{1}{\sqrt{(2N \pm 1)!}} \det|\tilde{\phi}_i \tilde{\phi}_j \cdots \tilde{\phi}_{N} \tilde{\phi}_{i'} \tilde{\phi}_{j'} \cdots \tilde{\phi}_{N'\pm1}|
\end{aligned}
\tag{3.126}
$$

Two charged Hamiltonians are formed from a sum of KS Hamiltonians one for each of the two diabatic states

$$
\hat{H}_n^{\pm} = \sum_i^{2N\pm1} \hat{h}_{\mathrm{KS},ni}^{\pm} \qquad\qquad \hat{H}_{n'}^{\pm} = \sum_i^{2N\pm1} \hat{h}_{\mathrm{KS},n'i}^{\pm} \tag{3.127}
$$

which are then used to approximate the electronic Hamiltonian

$$
\hat{H}_{\mathrm{e}}^{\pm} \approx \frac{1}{2}\left(\hat{H}_{\mathrm{KS},n}^{\pm} + \hat{H}_{\mathrm{KS},n'}^{\pm}\right) \tag{3.128}
$$

where we have taken the average of the two Hamiltonians to ensure that $V_{nn'} = V_{n'n}$ for cases where the two molecules are not the same [102]. The electronic coupling is then approximated with Eq. (3.126) and Eq. (3.128) so for hole and electron transport.

$$
V_{nn'} \approx \langle n|\hat{H}_{\mathrm{e}}^{+}|n'\rangle = \frac{1}{2}\langle\tilde{\phi}_N|\hat{h}_{\mathrm{KS},n}^{+} + \hat{h}_{\mathrm{KS},n'}^{+}|\tilde{\phi}_{N'}\rangle \tag{3.129}
$$

$$
V_{nn'} \approx \langle n|\hat{H}_{\mathrm{e}}^{-}|n'\rangle = \frac{1}{2}\langle\tilde{\phi}_{N+1}|\hat{h}_{\mathrm{KS},n}^{-} + \hat{h}_{\mathrm{KS},n'}^{-}|\tilde{\phi}_{N'+1}\rangle \tag{3.130}
$$

In the $2N$ variant of FO-DFT [57] the electronic couplings are approximated by the equation

$$
V_{nn'} \approx \frac{H_{nn'} - S_{nn'}(H_{nn} - H_{n'n'})}{1 - S_{nn'}^2} \tag{3.131}
$$

where

$$
H_{nn'} = \frac{1}{2}\langle\phi_n^{\mathrm{H/L}}|\hat{h}_{\mathrm{KS},n} + \hat{h}_{\mathrm{KS},n'}|\phi_{n'}^{\mathrm{H/L}}\rangle \qquad\qquad S_{nn'} = \langle\phi_n^{\mathrm{H/L}}|\phi_{n'}^{\mathrm{H/L}}\rangle \tag{3.132}
$$

and $\phi_n^{\mathrm{H/L}}$ is the HOMO for hole transport and LUMO of electron transport of the molecule n, so that the Eq. (3.131) transforms the coupling to the orthogonal basis set. Therefore the $2N$ variant of FO-DFT further approximates one-electron Hamiltonians in Eq. (3.129) and Eq. (3.130) with the neutral Hamiltonians so that $\hat{H}^{\pm}_{\mathrm{KS},n} \approx \hat{H}_{\mathrm{KS},n}$ and $\hat{H}^{\pm}_{\mathrm{KS},n'} \approx \hat{H}_{\mathrm{KS},n'}$.

3.4.3 Vibronic Coupling

Although local vibronic coupling parameters can be obtained from the derivatives of occupation energies with respect to the nuclear coordinates it is more commonly obtained from the differences between the optimised geometries for the different electronic states. The nuclear Hamiltonian for the spin-boson model where the electronic wavefunction occupies the state $|n\rangle$ is

$$\langle n|\hat{H}'_{\mathrm{SB}}|n\rangle = E'_n + \sum_{\alpha} \hat{H}_{n\alpha} \tag{3.133}$$

$$\hat{H}_{n\alpha} = \hbar\omega_\alpha(\hat{a}^\dagger_\alpha \hat{a}_\alpha + 1/2) + \hbar\omega_\alpha V_{n\alpha}(\hat{a}^\dagger_\alpha + \hat{a}_\alpha) \tag{3.134}$$

where the Hamiltonian $\hat{H}_{n\alpha}$ can be rewritten as

$$\hat{H}_{n\alpha} = \frac{\hbar^2}{2\mu_\alpha}\frac{\partial^2}{\partial x_\alpha^2} + \frac{1}{2}\mu_\alpha\omega_\alpha^2\hat{x}_\alpha^2 + \hbar\omega_\alpha V_{n\alpha}\sqrt{\frac{2\mu_\alpha\omega_\alpha}{\hbar}}\hat{x}_\alpha \tag{3.135}$$

and by completing the square we can rewrite the potential on the normal mode coordinate in a shifted form.

$$\hat{H}_{n\alpha} = \frac{\hbar^2}{2\mu_\alpha}\frac{\partial^2}{\partial x_\alpha^2} - \hbar\omega_\alpha V_{n\alpha}^2 + \frac{1}{2}\mu_\alpha\omega_\alpha^2\left[\hat{x}_\alpha + V_{n\alpha}\sqrt{\frac{2\hbar}{\mu_\alpha\omega_\alpha}}\right]^2 \tag{3.136}$$

This potential energy surface of the nuclei will have a minimum situated at $x_{n\alpha} = -V_{n\alpha}\sqrt{2\hbar/\mu_\alpha\omega_\alpha}$. We can therefore approximate the coupling $V_{n\alpha}$ as the difference in the normal mode coordinates from the optimised nuclei positions of a charged system in the state $|n\rangle$ to the optimised nuclei positions of the neutral

system so that

$$V_{n\alpha} = -x_{n\alpha}\sqrt{\frac{\mu_\alpha \omega_\alpha}{2\hbar}} \tag{3.137}$$

since the nuclear coordinates of the neutral system was defined to be at the origin $\mathbf{0}$. It is useful to define a shift parameter $\lambda_{nn'\alpha} = V_{n\alpha} - V_{n'\alpha}$ as the difference between the local vibronic couplings of two different states where its square is the Huang-Rhys factor $S_{nn'\alpha} = \lambda^2_{nn'\alpha}$. The shift parameters can therefore be calculated from the distances between the two different nuclear configurations in the normal mode coordinate system.

3.4.4 Reorganisation Energy

In the Marcus theory rate equation, a parameter called the reorganisation energy is used which is related to the shift parameters by the equation

$$\Lambda_{nn'} = \sum_\alpha \hbar\omega_\alpha \lambda^2_{nn'\alpha} \tag{3.138}$$

and is commonly obtained by calculating energy differences for the system at different geometries. From Eq. (3.136) the potential energy for the α component of the system when the electronic wavefunction is in state $|n\rangle$ is for example.

$$\hat{V}_{n\alpha}(x_\alpha) = \frac{1}{2}\mu_\alpha\omega_\alpha^2\left[\hat{x}_\alpha + V_{n\alpha}\sqrt{\frac{2\hbar}{\mu_\alpha\omega_\alpha}}\right]^2 - \hbar\omega_\alpha V_{n\alpha}^2 \tag{3.139}$$

The difference of this potential at the optimised geometry for state $|n'\rangle$ and state $|n\rangle$ is

$$\hat{V}_{n\alpha}(-V_{n'\alpha}\sqrt{2\hbar/\mu_\alpha\omega_\alpha}) - \hat{V}_{n\alpha}(-V_{n\alpha}\sqrt{2\hbar/\mu_\alpha\omega_\alpha}) = \hbar\omega_\alpha(V_{n\alpha} - V_{n'\alpha})^2 \tag{3.140}$$

so the reorganisation energy $\Lambda_{nn'}$ is the sum of the differences in the nuclear potential energies for all modes α with an electronic state $|n\rangle$ at the optimised

geometries of $|n'\rangle$ and $|n\rangle$ to give

$$\Lambda_{nn'} = \hat{V}_n(\mathbf{x}_{\mathrm{opt},n'}) - \hat{V}_n(\mathbf{x}_{\mathrm{opt},n}) \tag{3.141}$$

which is a generalisation of the four-point scheme.

The four-point scheme [103] assumes that when the electronic wavefunction is in the state $|n\rangle$ the excess charge will be completely localised on a single molecule at the site labelled n and affects only that molecule. The optimised coordinates $\mathbf{x}_{\mathrm{opt},n}$ will therefore be the coordinates of the charged species at the site n with the neutral geometries for all remaining molecules in the crystal. Similarly $\mathbf{x}_{\mathrm{opt},n'}$ will be the optimised coordinates of the charged species at the site n' with the neutral geometries for all remaining molecules. It also assumes that we can split in the nuclear potential energies as a sum of isolated molecule energies so that we can rewrite potential energy as

$$\hat{V}_n(\mathbf{x}_{\mathrm{opt},n'}) = E_n(\mathbf{x}_{n0}) + E_{n'}(\mathbf{x}_{n'\pm}) + \sum_{m\neq n,n'} E_m(\mathbf{x}_{m0}) \tag{3.142}$$

$$\hat{V}_n(\mathbf{x}_{\mathrm{opt},n}) = E_n(\mathbf{x}_{n\pm}) + E_{n'}(\mathbf{x}_{n'0}) + \sum_{m\neq n,n'} E_m(\mathbf{x}_{m0}) \tag{3.143}$$

where $\mathbf{x}_{\mathrm{opt},n'} = (\mathbf{x}_{n0}, \mathbf{x}_{n'\pm}, \cdots)$ and $\mathbf{x}_{\mathrm{opt},n} = (\mathbf{x}_{n\pm}, \mathbf{x}_{n'0}, \cdots)$ so that for example $\mathbf{x}_{n0}$ represents the nuclei of the neutral molecule at the site n and is a subset of the coordinates of $\mathbf{x}_{\mathrm{opt},n'}$. For crystals composed of a single type of molecule, we can make the replacements $E_n(\mathbf{x}) \to E_\pm(\mathbf{x})$ and $E_{n'}(\mathbf{x}) \to E_0(\mathbf{x})$. As the isolated molecule energies are invariant to translations or rotations we can also make the replacement $\mathbf{x}_{n0} \to \mathbf{x}_0$ and $\mathbf{x}_{n\pm} \to \mathbf{x}_\pm$. By applying the substitutions we finally arrive at the four-point scheme for the reorganisation energy at the isolated molecule approximation.

$$\Lambda_\pm = \Big[E_\pm(\mathbf{x}_0) - E_0(\mathbf{x}_0)\Big] + \Big[E_0(\mathbf{x}_\pm) - E_\pm(\mathbf{x}_\pm)\Big] \tag{3.144}$$

3.4.5 Marcus Theory

With the transformed Hamiltonian Eq. (3.95) we can use Fermi's golden rule to determine the transition rate from one state to another by assuming that $\hat{V}$ acts as a small perturbation onto the Hamiltonian $\hat{H}_0$ which has unperturbed eigenstates $|n, \boldsymbol{\nu}\rangle$ which is a product of the state $|n\rangle$ with the charge localised at the site n and the vibrational modes $|\boldsymbol{\nu}\rangle$ of the crystal. Therefore if we prepare a wavefunction in state $|2, \boldsymbol{\nu}\rangle$ the transition rate to a continuum of vibrational states with the charge carrier occupying the state $|1\rangle$ is.

$$k_{1\leftarrow 2,\boldsymbol{\nu}} = \frac{2\pi}{\hbar} \sum_{\boldsymbol{\nu}'} |\langle 2, \boldsymbol{\nu}|\hat{V}_{2,1}|1, \boldsymbol{\nu}'\rangle|^2 \, \delta(E_{2,\boldsymbol{\nu}} - E_{1,\boldsymbol{\nu}'}) \tag{3.145}$$

If we assume that the thermal relaxation in the vibrational modes is fast relative to the transition rate an initially prepared state $|2, \boldsymbol{\nu}\rangle$ will interact with the environment and thermally equilibrate. So after some amount of time much shorter than the time required for the transition to occur.

$$|2, \boldsymbol{\nu}\rangle \rightarrow \sum_{\boldsymbol{\nu}} P_{\boldsymbol{\nu}} |2, \boldsymbol{\nu}\rangle \qquad\qquad P_{\boldsymbol{\nu}} = \frac{e^{-\beta E_{\boldsymbol{\nu}}}}{\sum_{\boldsymbol{\nu}} e^{-\beta E_{\boldsymbol{\nu}}}} \tag{3.146}$$

The thermally averaged rate from a state with the charge localised to site 2 to a state where the charge is localised to site 1 is therefore.

$$k_{1\leftarrow 2} = \frac{2\pi}{\hbar} \sum_{\boldsymbol{\nu}} P_{\boldsymbol{\nu}} \sum_{\boldsymbol{\nu}'} |\langle 2, \boldsymbol{\nu}|\hat{V}_{2,1}|1, \boldsymbol{\nu}'\rangle|^2 \, \delta(E_{2,\boldsymbol{\nu}} - E_{1,\boldsymbol{\nu}'}) \tag{3.147}$$

The transition rates from states $|2\rangle$ to $|1\rangle$ are independent of the electronic couplings between any other states so that the transitions are uncorrelated, which was due to the truncation of the interacting term $\hat{V}$ to first-order in perturbation expansion used in the derivation of Fermi's golden rule.

Eq. (3.147) can be written in terms of a time-correlation function

$$k_{1\leftarrow 2} = \frac{1}{\hbar^2} \int_{-\infty}^{\infty} \mathrm{d}t \, e^{i\omega_{2,1}t} \, \langle e^{i\hat{H}_{\mathrm{B}}t/\hbar} e^{\hat{\Omega}_{2,1}} e^{-i\hat{H}_{\mathrm{B}}t/\hbar} e^{\hat{\Omega}_{1,2}} \rangle_{\hat{H}_{\mathrm{B}}} \tag{3.148}$$

with $\omega_{2,1} = E_{2,1}/\hbar = (E_2 - E_1)/\hbar$ and

$$E_{2,\boldsymbol{\nu}} - E_{1,\boldsymbol{\nu}'} = E_2 - E_1 + E_{\mathrm{B},\boldsymbol{\nu}} - E_{\mathrm{B},\boldsymbol{\nu}'} \tag{3.149}$$

where we have used the identity.

$$\delta(x) = \frac{1}{2\pi\hbar} \int_{-\infty}^{\infty} \mathrm{d}t \; e^{ixt/\hbar} \tag{3.150}$$

We can evaluate the thermal average over the vibrational states using the Weyl and Bloch identities [92, 104] so that the rate equation then becomes

$$k_{1\leftarrow 2} = \frac{|V_{2,1}|^2}{\hbar^2} e^{-\sum_\alpha \lambda_\alpha^2 (2n_\alpha + 1)} \int_{-\infty}^{\infty} \mathrm{d}t \; e^{i\omega_{2,1}t + \sum_\alpha \lambda_\alpha^2 [n_\alpha e^{i\omega_\alpha t} + (n_\alpha + 1)e^{-i\omega_\alpha t}]} \tag{3.151}$$

where $\lambda_\alpha = V_{2\alpha} - V_{1\alpha}$ is the shift parameter and we have dropped the site indices for clarity and $n_\alpha = 1/(e^{\beta\hbar\omega_\alpha} - 1)$. With high temperatures and large shift parameters, the time-correlation function becomes uncorrelated quickly so that for large times from $t = 0$ the integrand does not contribute significantly to the integral Eq. (3.151) so that we can take the short-time approximation by expanding the exponent of the integrand to second order in $\omega_\alpha t$

$$\begin{aligned} k_{1\leftarrow 2} &= \frac{|V_{2,1}|^2}{\hbar^2} \int_{-\infty}^{\infty} \mathrm{d}t \; e^{i\omega_{2,1}t - i\sum_\alpha \lambda_\alpha^2 \omega_\alpha t - (1/2)\sum_\alpha \lambda_\alpha^2 \omega_\alpha^2 (2n_\alpha + 1)t^2} \\ &= \frac{|V_{2,1}|^2}{\hbar^2} \sqrt{\frac{\pi}{a}} e^{-(\Lambda/\hbar - \omega_{2,1})^2/4a} \qquad a = \frac{1}{2}\sum_\alpha (2n_\alpha + 1)\lambda_\alpha^2 \omega_\alpha^2 \end{aligned} \tag{3.152}$$

where $\Lambda = \sum_\alpha \hbar\omega_\alpha \lambda_\alpha^2$ is the reorganisation energy. We then take the classical limit for the probability of a vibrational mode in state α so that $n_\alpha = k_\mathrm{B}T/\hbar\omega_\alpha - 1/2$ which is the analytic part of the Laurent series of n_α to obtain the equation

$$k_{\mathrm{MT},1\leftarrow 2} = \frac{|V_{2,1}|^2}{\hbar} \sqrt{\frac{\pi}{\Lambda k_\mathrm{B} T}} \exp\left[-\frac{(\Lambda - E_{2,1})^2}{4\Lambda k_\mathrm{B} T} \right] \tag{3.153}$$

which is Marcus theory transition rate from one state to another.

3.5 Carrier Mobilities

For an OFET device, we are interested in the induced electric currents that flow through the organic semiconductor due to an external electric field generated when a voltage is applied to the drain and gate electrodes. Theoretically, we will want an equation that describes the current density of the semiconductor as a function of an external electric field. We can write this current density as a Taylor expansion [105]

$$j_\alpha(\boldsymbol{\Xi}) = j_\alpha(\mathbf{0}) + \sum_\beta \sigma_{\alpha\beta}\Xi_\beta + \cdots \tag{3.154}$$

$$\sigma_{\alpha\beta} = \left.\frac{\partial j_\alpha}{\partial \Xi_\beta}\right|_{\mathbf{0}} \qquad \sigma_{\alpha\beta} = N_\mathrm{c} q \mu_{\alpha\beta} \tag{3.155}$$

where $\boldsymbol{\Xi}$ is the external electric field, $\sigma_{\alpha\beta}$ is the conductivity, $\mu_{\alpha\beta}$ is the carrier mobility, N_c is the number of charge carriers, and q is the carriers charge. We can truncate Eq. (3.154) to first-order so we are only interested in the linear response of the current density due to the external electric field similar to the GCA in Eqs. (2.1) and (2.2). In this section, we will describe a few methods that can be used to approximate carrier mobilities.

3.5.1 Kubo Formula

In the Schrödinger representation, the time-dependent charge density of a many-particle system of electrons in a state $\psi(\mathbf{r}_\mathrm{e}, t)$ can be obtained using the charge density operator

$$n_\mathrm{el}(\mathbf{r}, t) = -e\langle\psi(\mathbf{r}_\mathrm{e}, t)|\hat{n}(\mathbf{r}_\mathrm{e}, \mathbf{r})|\psi(\mathbf{r}_\mathrm{e}, t)\rangle \tag{3.156}$$

$$\hat{n}(\mathbf{r}_\mathrm{e}, \mathbf{r}) = \sum_{i=1}^{N} \delta(\mathbf{r}_i - \mathbf{r}) \tag{3.157}$$

a current density operator can be derived from the continuity equation [106]

$$\frac{\partial n_{\mathrm{el}}(\mathbf{r},t)}{\partial t} = -\boldsymbol{\nabla}\cdot\mathbf{j}_{\mathrm{el}}(\mathbf{r},t) \tag{3.158}$$

$$\mathbf{j}_{\mathrm{el}}(\mathbf{r},t) = \langle\psi(\mathbf{r}_{\mathrm{e}},t)|\hat{\mathbf{j}}_{\mathrm{el}}(\mathbf{r}_{\mathrm{e}},\mathbf{r})|\psi(\mathbf{r}_{\mathrm{e}},t)\rangle \tag{3.159}$$

where $\mathbf{j}_{\mathrm{el}}(\mathbf{r},t)$ is the current density and the current density operator is defined as.

$$\hat{\mathbf{j}}_{\mathrm{el}}(\mathbf{r},\mathbf{r}_{\mathrm{e}}) = \frac{e_0}{2m_e}\sum_i \hat{\mathbf{p}}_i\delta(\mathbf{r}_i-\mathbf{r}) + \delta(\mathbf{r}_i-\mathbf{r})\hat{\mathbf{p}}_i \tag{3.160}$$

For carrier mobility calculations it is more useful to use the sum of the current densities which can be obtained by taking the time derivative of the charge polarisation [105]

$$\mathbf{j}_{\mathrm{el}}(t) = \frac{\partial}{\partial t}\langle\psi(\mathbf{r}_{\mathrm{e}},t)|\hat{\boldsymbol{\mu}}(\mathbf{r}_{\mathrm{e}})|\psi(\mathbf{r}_{\mathrm{e}},t)\rangle = \int \mathrm{d}^3\mathbf{r}\,\mathbf{j}_{\mathrm{el}}(\mathbf{r},t) \tag{3.161}$$

where $\hat{\boldsymbol{\mu}}(\mathbf{r}_{\mathrm{e}}) = -e_0\sum_i \hat{\mathbf{r}}_i$ so that the time derivative is equivalent to carrying out an integration over space of the current density, a full derivation is given in Appendix A. So that this current density operator can be obtained from the expression.

$$\hat{\mathbf{j}}_{\mathrm{el}} = \frac{1}{i\hbar}\left[\hat{\boldsymbol{\mu}},\hat{H}\right] \tag{3.162}$$

The Kubo formula [105] gives an expression for the carrier mobility to be

$$\begin{aligned}\mu_{\alpha\beta} &= \frac{1}{k_{\mathrm{B}}T}\frac{\pi\hbar}{e_0 N_{\mathrm{c}}}\sum_{ij} P_i\,\langle\Psi_i|\hat{j}_\alpha|\Psi_j\rangle\langle\Psi_j|\hat{j}_\beta|\Psi_i\rangle\,\delta(E_i-E_j)\\ &= \frac{1}{2k_{\mathrm{B}}T}\frac{1}{e_0 N_{\mathrm{c}}}\int_{-\infty}^{\infty}\mathrm{d}t\,\langle e^{i\hat{H}t/\hbar}\hat{j}_\alpha e^{-i\hat{H}t/\hbar}\hat{j}_\beta\rangle_{\hat{H}}\end{aligned} \tag{3.163}$$

where $\mu_{\alpha\beta}$ is the mobility tensor and for this section $\hat{H}$ will be the Hamiltonian

for the spin-boson model with $N_{\mathrm{c}} = 1$ since our diabatic wavefunctions are formed from charged localised states with only a single excess electron/hole localised to a given site. The polarisation operator in the spin-boson model is

$$\hat{\mu}_a = -e_0 \sum_n R_{n\alpha} |n\rangle\langle n| \qquad\qquad R_{n\alpha} = \langle n|\hat{r}_\alpha|n\rangle \tag{3.164}$$

where $\hat{r}_\alpha = \sum_i \hat{r}_{i\alpha}$ is the position operator so by using Eq. (3.162) we can obtain the current operator

$$\hat{j}_\alpha = -\frac{e_0}{i\hbar} \sum_{n,n' \neq n} R_{nn'\alpha} \hat{V}_{nn'} \tag{3.165}$$

with $R_{nn'\alpha} = R_{n\alpha} - R_{n'\alpha}$, inserting Eq. (3.165) into (3.163) we now obtain the equation

$$\mu_{\alpha\beta} = \frac{e_0}{2k_{\mathrm{B}}T} \sum_{\substack{n,n' \neq n \\ m,m' \neq m}} R_{nn'\alpha} R_{m'm\beta} \frac{1}{\hbar^2} \int_{-\infty}^{\infty} \mathrm{d}t \, \langle e^{i\hat{H}_{\mathrm{SB}}t/\hbar} \hat{V}_{nn'} e^{-i\hat{H}_{\mathrm{SB}}t/\hbar} \hat{V}_{mm'} \rangle_{\hat{H}_{\mathrm{SB}}} \tag{3.166}$$

where thermal average can be rewritten as

$$\langle e^{i\hat{H}_{\mathrm{SB}}t/\hbar} \hat{V}_{nn'} e^{-i\hat{H}_{\mathrm{SB}}t/\hbar} \hat{V}_{mm'} \rangle_{\hat{H}_{\mathrm{SB}}} = \frac{\mathrm{Tr}\left[e^{-\beta\hat{H}_0} \hat{U}(-i\hbar\beta) \hat{U}^{-1}(t) \hat{V}_{nn'}(t) \hat{U}(t) \hat{V}_{mm'} \right]}{\mathrm{Tr}\left[e^{-\beta\hat{H}_0} \hat{U}(-i\hbar\beta) \right]} \tag{3.167}$$

with

$$\hat{U}(t) = 1 + \frac{-i}{\hbar} \int_0^t \mathrm{d}t_1 \, \hat{V}(t_1) + \left(\frac{-i}{\hbar}\right)^2 \int_0^t \mathrm{d}t_1 \int_0^{t_1} \mathrm{d}t_2 \, \hat{V}(t_1)\hat{V}(t_2) + \cdots \tag{3.168}$$

where the time-dependent electronic coupling operators are in the interaction representation so that $\hat{V}_{nn'}(t) = e^{i\hat{H}_0 t/\hbar} \hat{V}_{nn'} e^{-i\hat{H}_0 t/\hbar}$. Next, we truncate Eq. (3.168) to the zeroth-order in the coupling to approximate the thermal average

$$\langle e^{i\hat{H}_{\mathrm{SB}}t/\hbar} \hat{V}_{nn'} e^{-i\hat{H}_{\mathrm{SB}}t/\hbar} \hat{V}_{mm'} \rangle_{\hat{H}_{\mathrm{SB}}} \approx \langle \hat{V}_{nn'}(t) \hat{V}_{mm'} \rangle_{\hat{H}_0} \tag{3.169}$$

which can then be rewritten as

$$\begin{aligned}\langle \hat{V}_{nn'}(t)\hat{V}_{mm'}(0)\rangle_{\hat{H}_0} =& V_{nn'}V_{mm'}\langle e^{i\hat{H}_\mathrm{D}t/\hbar}|n\rangle\langle n'|e^{-i\hat{H}_\mathrm{D}t/\hbar}|m\rangle\langle m'|\rangle_{\hat{H}_\mathrm{D}} \\ &\times \langle e^{i\hat{H}_\mathrm{B}t/\hbar}e^{\hat{\Omega}_{nn'}}e^{-i\hat{H}_\mathrm{B}t/\hbar}e^{\hat{\Omega}_{mm'}}\rangle_{\hat{H}_\mathrm{B}}\end{aligned} \tag{3.170}$$

which is only non-zero when $n = m'$ and $n' = m$

$$\langle e^{i\hat{H}_\mathrm{D}t/\hbar}|n\rangle\langle n'|e^{-i\hat{H}_\mathrm{D}t/\hbar}|n'\rangle\langle n|\rangle_{\hat{H}_\mathrm{D}} = P_n e^{i\omega_{nn'}t} \qquad P_n = \frac{e^{-\beta E_n}}{\sum_n e^{-\beta E_n}} \tag{3.171}$$

so that Eq. (3.166) is approximated with

$$\begin{aligned}\mu_{\alpha\beta} = \frac{e_0}{2k_\mathrm{B}T}\sum_{n,n'\neq n}\Bigg\{ & P_n R_{nn'\alpha}R_{nn'\beta} \\ & \times \frac{|V_{nn'}|^2}{\hbar^2}\int_{-\infty}^{\infty}\mathrm{d}t\, e^{i\omega_{nn'}t}\langle e^{i\hat{H}_\mathrm{B}t/\hbar}e^{\hat{\Omega}_{nn'}}e^{-i\hat{H}_\mathrm{B}t/\hbar}e^{\hat{\Omega}_{n'n}}\rangle_{\hat{H}_\mathrm{B}}\Bigg\}\end{aligned} \tag{3.172}$$

where we can now replace the integral with the Marcus theory equation by taking the high-temperature limit as outlined in Subsection 3.4.5.

$$\mu_{\alpha\beta} = \frac{e_0}{2k_\mathrm{B}T}\sum_{n,n'\neq n} P_n R_{nn'\alpha}R_{nn'\beta}k_{\mathrm{MT},n'\leftarrow n} \tag{3.173}$$

For crystal structures there are many repeating terms, the summation over n can be reduced by simply running the summation over unique sites in the unit cell and multiplying by the number of unit cells of the crystal. The same reduction can be made for the summation in the denominator of P_n so there will be a cancellation of the two prefactors. The summation over n' must be over all sites in the crystal structure but since k_MT decreases rapidly with distance, the summation n' can be reduced to only run over molecules in close contact.

3.5.2 Kinetic Monte Carlo

The Einstein relation gives an expression that relates the diffusion of a charge carrier and its carrier mobility.

$$\mu_{\alpha\beta} = \frac{e_0}{k_\mathrm{B}T} D_{\alpha\beta} \tag{3.174}$$

From Subsection 3.5.1 we arrived at an expression for the carrier mobility in Eq. (3.173) and from the Einstein relation we can see that the diffusion of a carrier can be approximated to be

$$D_{\alpha\beta} = \frac{1}{2} \sum_{n,n' \neq n} P_n R_{nn'\alpha} R_{nn'\beta} k_{\mathrm{MT},n' \leftarrow n} \tag{3.175}$$

which is only a sum of first-order decay rates from a state $|n\rangle$ to $|n'\rangle$ with no intermediate states. To include higher-order terms would require terms beyond the zeroth-order in the expansion of Eq. (3.168) in the thermal average of Eq. (3.167).

We could instead try to obtain the diffusion coefficients directly by running a kinetic Monte Carlo (KMC) [107] simulation which models the charge carrier diffusion across the crystal structure through a series of hops. KMC evolves the system using a Markov chain which requires that the transition rates from one state to another are independent of the trajectory that the charge carrier had taken. This requires that the charge carrier occupies a state for a long enough time for it to be sufficiently randomised by the environment so the transition rates become independent of the previous transitions that had occurred. The vibrational modes must be fast relative to the rate of hopping the same assumptions used to obtain the Marcus theory rate equation in Subsection 3.4.5. We can therefore calculate the diffusion coefficient of a particle with the expression

$$D_{\alpha\beta} = \lim_{t \to \infty} \frac{1}{2t} \langle r_\alpha(t) r_\beta(t) \rangle \tag{3.176}$$

where the particle is situated at the origin at $t = 0$ and $\langle \cdots \rangle$ denotes an average over a large number of KMC trajectories.

To run a KMC simulation we must define an initial state and a set of transition

rates from one state to another which will be calculated using Marcus theory. We first prepare a charge carrier in state $|n\rangle$ at the origin at $t = 0$ with a probability.

$$P_n = \frac{e^{-\beta E_n}}{\sum_n e^{-\beta E_n}} \tag{3.177}$$

The total decay rate from the state $|n\rangle$ to any other state is

$$k_{\mathrm{MT},\{n'\}\leftarrow n} = \sum_{n'} k_{\mathrm{MT},n'\leftarrow n} \tag{3.178}$$

and the probability that the charge carrier remains in that state is

$$p_n(t) = \exp\left(-k_{\mathrm{MT},\{n'\}\leftarrow n} t\right) \tag{3.179}$$

where the total decay rate was obtained by extending the golden rule rate equations Eq. (3.145) to include all coupling operators from a state $|n\rangle$ to all other states and taking the high-temperature limit. We then carry out a hopping move from the state $|n\rangle$ to another state with the probability.

$$P_{n'\leftarrow n} = k_{\mathrm{MT},n'\leftarrow n} / k_{\mathrm{MT},\{n'\}\leftarrow n} \tag{3.180}$$

The hopping move takes an amount of time Δt with the equation

$$\Delta t = -\ln(r) / k_{\mathrm{MT},\{n'\}\leftarrow n} \tag{3.181}$$

where r is a randomly generated number in the range $0 < r \leq 1$. Another hop occurs again from the new state and these hops are carried out continually until a large number of hops are carried out or a specified simulation time is reached. A large number of these KMC trajectories are carried out to evaluate Eq. (3.176) and obtain the carrier mobility with Eq. (3.174).

3.5.3 Transient Localisation Theory

Transient localisation theory (TLT) [64, 65] takes another approach to evaluate the Kubo formula Eq. (3.163) which we will rewrite here in an alternate but

equivalent form.

$$\mu_{\alpha\beta} = \frac{1}{k_{\mathrm{B}}T}\frac{1}{e_0 N_{\mathrm{c}}}\,\mathrm{Re}\int_0^\infty \mathrm{d}t\,\langle e^{i\hat{\mathcal{H}}t/\hbar}\hat{\mathcal{J}}_\alpha e^{-i\hat{\mathcal{H}}t/\hbar}\hat{\mathcal{J}}_\beta\rangle_{\hat{\mathcal{H}}} \tag{3.182}$$

Using the relaxation time approximation (RTA) the current-current correlation function of the real system is re-expressed in terms of the current-current correlation function of a reference system which decays over time. The Kubo formula in the RTA is given as

$$\mu_{\alpha\beta} = \frac{1}{k_{\mathrm{B}}T}\frac{1}{e_0 N_{\mathrm{c}}}\,\mathrm{Re}\int_0^\infty \mathrm{d}t\,\langle e^{i\hat{H}t/\hbar}\hat{j}_\alpha e^{-i\hat{H}t/\hbar}\hat{j}_\beta\rangle_{\hat{H}}\, e^{-(1/\tau)t} \tag{3.183}$$

or equivalently

$$\mu_{\alpha\beta} = \frac{1}{k_{\mathrm{B}}T}\frac{L^2_{\alpha\beta}(\tau)}{2\tau} \tag{3.184}$$

the term $L^2_{\alpha\beta}(\tau)$ is the squared localisation length

$$L^2_{\alpha\beta}(\tau) = \frac{\hbar^2}{e_0 N_{\mathrm{c}}}\sum_{ij} P_i\,\langle\Psi_i|\hat{j}_\alpha|\Psi_j\rangle\langle\Psi_j|\hat{j}_\beta|\Psi_i\rangle\frac{2}{(\hbar/\tau)^2 + (E_i - E_j)^2} \tag{3.185}$$

where we have used the relation

$$\mathrm{Re}\int_0^\infty \mathrm{d}t\, e^{ixt}e^{-st} = \frac{s}{s^2 + x^2} \tag{3.186}$$

for real values of x and s. In this subsection, we use script lettering to denote the Hamiltonian and operators of the real system and normal lettering for the reference system.

For example, we can model the real system in TLT using the tight-binding

Hamiltonian

$$\begin{aligned}\hat{\mathcal{H}} = &\sum_n E'_n|n\rangle\langle n| + \sum_{n,n'\neq n} V_{nn'}|n\rangle\langle n'| + \sum_{n\alpha} \hbar\omega_\alpha V_{n\alpha}(\hat{a}^\dagger_\alpha + \hat{a}_\alpha)|n\rangle\langle n| \\ &+ \sum_{n,n'\neq n,\alpha} \hbar\omega_\alpha V_{nn'\alpha}(\hat{a}^\dagger_\alpha + \hat{a}_\alpha)|n\rangle\langle n'| + \hat{H}_{\mathrm{B}}\end{aligned} \tag{3.187}$$

where all parameters are analogous to those derived for the spin-boson model in Section 3.4. Usually, the reference system is defined by a tight-binding Hamiltonian

$$\hat{H} = \sum_n \tilde{E}_n|n\rangle\langle n| + \sum_{n,n'\neq n} \tilde{V}_{nn'}|n\rangle\langle n'| \tag{3.188}$$

where the site energies $\tilde{E}_n$ and electronic couplings $\tilde{V}_{nn'}$ are sampled from a normal distribution. The mean and deviation of the normal distributions are determined from the Hamiltonian of the real system, some example equations to calculate them are derived in Appendix B. This reference Hamiltonian represents a single realisation of the disorder caused by the slow nuclear vibrations of the system. To calculate carrier mobilities an average of the squared localisation lengths of a number of different realisation of the disorder is used so that

$$\mu_{\alpha\beta} = \frac{1}{k_{\mathrm{B}}T}\frac{\bar{L}^2_{\alpha\beta}(\tau)}{2\tau} \tag{3.189}$$

where $\bar{L}^2_{\alpha\beta}(\tau)$ is the average squared localisation length. A practical implementation of the procedure to carry out mobility calculation using this reference Hamiltonian can be found in Ref. [65].

Chapter 4

CSPy

CSPy is a program developed within the research group that allows for the automation of some parts of the CSP workflow but requires molecular geometries and atom centred multipoles to be obtained from other computational methods. Typically multipoles up to hexadecapoles are obtained using the Gaussian Distributed Multipole Analysis (GDMA) program from a calculated electron density and a set of monopoles are generated by fitting them to match the electrostatic potential of the multipoles using MULFIT. CSPy can then be set up to run a variety of different methods with multiple optimisation stages. In the current version of CSPy with the default settings, optimisations are carried out in three separate stages. A high-level flow diagram of the overall workflow is shown in Figure 4.1.

So when the default settings are used and the molecular geometries and multipoles have been obtained the CSP procedure begins by generating quasi-random crystal structures in a select number of space groups which are then minimised with an electrostatic potential model using: monopoles with the PMIN program followed by DMACRYS with the crystal structure under a small external pressure and finally with multipoles using DMACRYS. In all stages, the remaining interactions are modelled using the exp-6 force field. The structure generations and minimisations are carried out continually until a specified number of successfully minimised structures have been obtained for a given space group. At the end duplicated crystal structures are removed by generating simulated X-

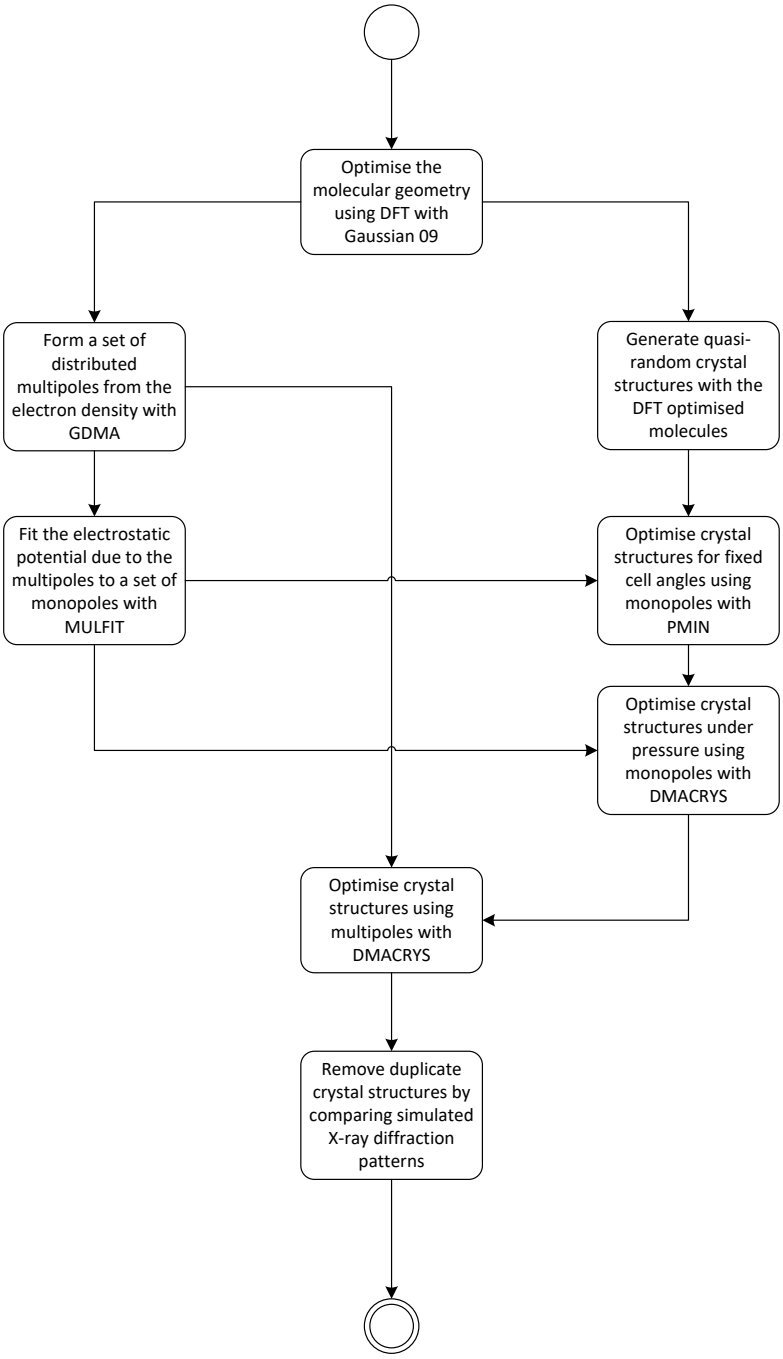


Figure 4.1: High-level flow diagram of the CSP workflow used in this project.

ray diffraction patterns using PLATON [108] and calculating similarities between crystal structures (within a $1\,\mathrm{kJ\,mol^{-1}}$ and $0.05\,\mathrm{g\,cm^{-3}}$ window) with the cosine similarity and constrained dynamic time warping methods and retaining only the lowest energy structure when similarities are greater than 0.8 and less than 10.0 respectively.

CSPy automates parts or all of stages 2 and 3 of the CSP process defined by the authors of the sixth blind test which are [25].

1. Exploration of the conformational preferences of the target molecules.
2. Generating plausible crystal-packing arrangements of the target molecules.
3. Ranking the likelihood of resulting crystal structures forming using some form of scoring or fitness function.

Stage 1 is therefore mostly absent and requires other programs to generate molecular conformers such as CREST [109] or as done in this project RDKit [110] followed up with UFF and DFT optimisations. Stages 2 and 3 can be fully or partially completed by CSPy depending on the CSP workflow being used. For example, a CSP workflow may include optimisation in stage 2 with periodic DFT optimisation using VASP [111–114] on unique structures obtained from CSPy. Others may decide that structures from CSPy are sufficiently accurate so no further actions are taken in stage 2 and may wish to only carry out further energy evaluations using more accurate methods to improve ranking adding an additional step to stage 3. Due to the requirement for an efficient CSP procedure in our computational screening process no further additional calculations are carried out after CSPy so that stages 2 and 3 will be completed entirely by CSPy.

4.1 GDMA

We use the GDMA program [115] to partition the electron density and perform integrations that are analogous to the Eqs. (3.77) to obtain a set of distributed multipoles. In closed-shell single Slater determinant methods such as RHF or

RKS the electron density can be obtained from

$$n(\mathbf{r}) = \langle\psi|\hat{n}(\mathbf{r}_\mathrm{e}, \mathbf{r})|\psi\rangle = 2\sum_{i}^{N/2} \varphi_i^*(\mathbf{r})\varphi_i(\mathbf{r}) \tag{4.1}$$

where $\varphi_i(\mathbf{r})$ form an orthonormal set of orbitals and are typically made from a linear combination of functions

$$\varphi_i(\mathbf{r}) = \sum_{\mu} c_{\mu i} g_\mu(\mathbf{r}) \tag{4.2}$$

the coefficients $c_{\mu i}$ are determined by running the RHF or RKS equations until self-consistency. Typically the functions $g_\mu(\mathbf{r})$ are a sum of Gaussian type functions centred on an atomic nucleus. For simplicity, we assume for all cases $g_\mu(\mathbf{r})$ is a single Gaussian type function

$$g_\mu(\mathbf{r}) = N_\mu (x - r_{Kx})^{l_\mu} (y - r_{Ky})^{m_\mu} (z - r_{Kz})^{n_\mu} e^{-\alpha_\mu |\mathbf{r} - \mathbf{r}_K|^2} \tag{4.3}$$

where N_μ is a normalisation factor and in this case, $g_\mu(\mathbf{r})$ is centred on the atomic nucleus K. The electron density can be rewritten as a sum of these functions

$$n(\mathbf{r}) = 2\sum_{\mu\nu}\sum_{i} c_{\mu i}^* c_{\nu i} \varphi_\mu^*(\mathbf{r})\varphi_\nu(\mathbf{r}) = \sum_{\mu\nu} D_{\mu\nu}\varphi_{\mu\nu}(\mathbf{r}) \tag{4.4}$$

where the product of the Gaussian type functions is another Gaussian type function

$$\varphi_{\mu\nu}(\mathbf{r}) = N_{\mu\nu} (x - r_{KLx})^{l_{\mu\nu}} (y - r_{KLy})^{m_{\mu\nu}} (z - r_{KLz})^{n_{\mu\nu}} e^{-\alpha_{\mu\nu} |\mathbf{r} - \mathbf{r}_{KL}|^2} \tag{4.5}$$

$$\alpha_{\mu\nu} = \alpha_\mu + \alpha_\nu \qquad\qquad \mathbf{r}_{KL} = \frac{\alpha_\mu \mathbf{r}_K + \alpha_\nu \mathbf{r}_L}{\alpha_{\mu\nu}} \tag{4.6}$$

where $N_{\mu\nu} = N_\nu N_\mu$, $l_{\mu\nu} = l_\mu + l_\nu$, $m_{\mu\nu} = m_\mu + m_\nu$, and $n_{\mu\nu} = n_\mu + n_\nu$.

For the atom centred multipoles GDMA partitions the electron density into

individual parts centred around each nucleus of the molecule

$$n(\mathbf{r}) = \sum_M n_M(\mathbf{r}) = \sum_{\mu\nu} D_{\mu\nu}\varphi_{\mu\nu}(\mathbf{r}) \tag{4.7}$$

the summation on the right is separated so that

$$\sum_M n_M(\mathbf{r}) = \sum_{\mu'\nu'} D_{\mu'\nu'}\varphi_{\mu'\nu'}(\mathbf{r}) + n'(\mathbf{r}) \tag{4.8}$$

and the summations over μ' and ν' are for the case where $\alpha_{\mu\nu} \geq Z$ and $n'(\mathbf{r})$ is the remaining electron density, GDMA uses a default setting of $Z = 4.0$. GDMA now assigns terms on the left of the equality to those on the right first by assigning $D_{\mu'\nu'}\varphi_{\mu'\nu'}(\mathbf{r})$ to the closest nucleus from its centre and assigns partitions of $n'(\mathbf{r})$ using Beckes fuzzy cell [116] so that

$$n_M(\mathbf{r}) = \sum_{\mu'\nu'} D_{M\mu'\nu'}\varphi_{M\mu'\nu'}(\mathbf{r}) + w_M(\mathbf{r})n'(\mathbf{r}) \tag{4.9}$$

and the functions $D_{M\mu'\nu'}\varphi_{M\mu'\nu'}(\mathbf{r})$ have been assigned to nucleus M and $w_M(\mathbf{r})$ is a weight function that partitions the electron density into a Becke fuzzy cell for the space around nuclei M. The atomic centred multipole moments can be obtained by integrating

$$Q_{Ml\kappa} = -\int \mathrm{d}^3\mathbf{r}\, R_{Ml\kappa}(\mathbf{r})n_M(\mathbf{r}) \tag{4.10}$$

where $R_{Ml\kappa}(\mathbf{r})$ are real regular solid harmonics centred at $\mathbf{r}_M$ and $Q_{Ml\kappa}$ are atomic spherical-tensor multipole moments centred at $\mathbf{r}_M$. Equations to convert between spherical-tensor multipole moments and Cartesian-tensor multipole moments can be found in the GDMA manual Ref. [117].

In CSPy we use a point charge model to initially minimise quasi-random crystal structures as the interactions between point charges can be evaluated more rapidly. However in general it is not recommended to use point charges from GDMA as higher-order multipole moments are necessary to model the electrostatic potential of a molecule due to the distortions of the electron density

caused by chemical bonding [117]. However, if point charges are required Stone recommends using the program MULFIT [118–121] which can fit multipoles from GDMA to another set of lower-ranked multipoles. To fit multipoles to monopoles MULFIT takes the multipoles situated on a given atom and fits this to a set of distributed monopoles centred on that atom and each neighbouring atom within a distance $r_{\text{vdw}} + r_{\text{incl}}$. This results in a series of monopoles on each atom due to the multipoles on that atom and each neighbouring atom which are summed over leaving a single monopole per atom. The molecule with a set of distributed monopoles on each atom collectively reproduces an electrostatic potential which matches more closely the potential due to the reference set of multipoles than the set of monopoles from the reference.

4.2 Quasi-Random Structures

In CSPy initial structures are generated by mapping an N-dimensional vector of quasi-random (QR) numbers $x_i \in [0, 1)$ generated using a Sobol sequence [122, 123] to the N structural parameters of a crystal structure of rigid molecules. The number of structural parameters depends on the chosen number of molecules in the asymmetric unit and the space group of the crystal. In CSPy there are two different versions of the QR structure generation methods which are used in this project. In both versions 6 parameters are used for each molecule in the asymmetric unit, 3 QR numbers are mapped to a unit quaternion and are used to rotate the molecule, and another 3 QR numbers are used to translate the molecule from the origin to a position in fractional coordinates. Another 0-3 parameters depending on the space group are mapped to the unit cell angles. For each cell angle that is not fixed by the space group symmetries, a QR number is mapped to the cell angle with the equation.

$$\theta_i = \left[\frac{1}{2}\arccos(1 - 2x_i)\right] + \frac{\pi}{4} \tag{4.11}$$

Another 1-3 parameters depending on the space group are used for the generation of the unit cell lengths. The next two subsections explain the two different methods that are used to set and adjust cell lengths before accepting or rejecting

the QR structure for minimisation.

4.2.1 Version I

In version I [36] the cell lengths are chosen so that the mean of the cell volumes generated is equal to the sum of the molecular volumes of all molecules in the unit cell multiplied by the target volume parameter (TVP). The molecular volumes are calculated from the volume of the smallest box required to contain the molecule with van der Waals radii around each atom with edges of this box oriented so that they are parallel to the molecules axes of inertia[1]. A minimum $s_{\min,i}$ and maximum $s_{\max,i}$ cell lengths are defined to be the smallest and largest lengths of the projection of the molecule and van der Waals radii onto the cell axes. For cases when there is only one molecule in the unit cell $s_{\max,i}$ is increased by 50%. The lengths of all but one of the cell axes that are not fixed by the space group symmetries are obtained using the QR number and the equation.

$$l_i = \frac{3}{4}\left[s_{\min,i} + x_i(N_{\text{mols}}\, s_{\max,i} - s_{\min,i})\right] \tag{4.12}$$

The final QR number is used so that the last cell length is sampled from a normal distribution so that cell volumes have a mean equal to the target volume and a standard deviation of $0.15 s_{\min,i}$. Throughout the structure generation procedure, the cell axes orderings are permuted so that each axis will be defined as the last axis an equal number of times.

The SAT-expand method is then used as an initial minimisation stage which is analogous to the minimisation of a crystal structures unit cell volume using a hard-sphere potential model. SAT-expand starts by first generating a crystal structure using a TVP of 1.0, the structure is then used in the next minimisation stages if there are no molecular collisions detected using the separating axis theorem (SAT) with molecular convex hull extended with atomic van der Waals radii. If collisions are detected the cell lengths increased iteratively. For each

[1] In the actual implementation of version I the moment of inertia tensor is calculated with all masses set to unity, in version II the moment of inertia tensor is calculated in the usual way using the atomic masses.

iteration unit cell lengths are increased by

$$\Delta l_i = l_i \left| \frac{v_{\text{over},i}}{v_{\text{cent},i}} \right| + 0.001 \tag{4.13}$$

in fractional coordinates where $v_{\text{over},i}$ is the overlap vector between the molecules extended convex hulls and $v_{\text{cent},i}$ is the vector between the centroid of the two molecules. The unit cell expansions are carried out up until a maximum cell volume of 2.5 multiplied by the sum of molecular volumes. If collisions are still detected at this point the structure will be rejected.

4.2.2 Version II

Convex hulls cannot be generated for molecules with less than four atoms or molecules that do not extend out in all three dimensions e.g. exactly linear or planar molecules. Version I, therefore, cannot be applied to these types of molecules although there are some methods that can be used to get around this issue, for example, a tetrahedron is used to model single atoms. This issue and others have led to the development of version II which uses a slightly different method for unit cell generation, collision detection, and initial minimisation. Unit cell lengths are mapped using a similar method to version I

$$l_i = l_{\text{min},i} + x_i(l_{\text{max},i} - l_{\text{min},i}) \tag{4.14}$$

where $l_{\text{min},i}$ and $l_{\text{max},i}$ are the minimum and maximum unit cell lengths and depend on the space group used. For the cases when there are three independent unit cell lengths the $l_{\text{min},i}$ terms are calculated by taking the maximum projection of the atomic positions onto the unit cell axes plus twice the maximum van der Waals radii of the atoms of the molecule. The $l_{\text{max},i}$ are then calculated with the

following equations

$$
\begin{aligned}
l_{\max,i} &= V_{\max}(V_{\text{ang}} l_{\min,j} l_{\min,k})^{-1} \\
l_{\max,j} &= V_{\max}(V_{\text{ang}} l_i l_{\min,k})^{-1} \\
l_{\max,k} &= V_{\max}(V_{\text{ang}} l_i l_j)^{-1}
\end{aligned}
\tag{4.15}
$$

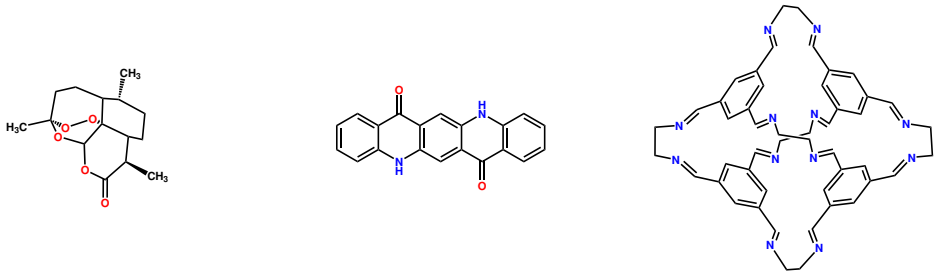
where

$$
V_{\text{ang}} = \left[1 - \cos^2(\alpha) - \cos^2(\beta) - \cos^2(\gamma) + 2\cos(\alpha)\cos(\beta)\cos(\gamma)\right]^{1/2} \tag{4.16}
$$

so that V_{ang} is the angular component to the unit cell volume and $V_{\max}$ is the sum of molecular box volumes in the unit cell multiplied by a TVP of 2.5. The unit cells must therefore be calculated in a specific order and similarly to version I the orderings are permuted throughout the structure generation procedure.

Once the crystal structure has been generated a minimisation is carried out which removes collisions and reduces the unit cell volumes. First, a collision detection is run between molecules of the crystal structure; molecules are colliding if the distance between any two atoms is less than a sum of their covalent radii plus 0.5 Å. If a collision has been detected the shortest unit cell length is increased by 1.0 Å. This is carried out iteratively until there are no collisions or when the unit cell volume becomes greater than $V_{\max}$ causing the structure to be rejected. If there are no collisions the minimisation moves onto the second stage which contract cell length by 1.0 Å for each iteration. This contraction stage is carried out for each independent cell length starting with the longest. If this contraction results in a collision the previous structure is used and the second-largest cell length undergoes the contraction iterations. At the end when the contraction for the shortest cell length results in a collision the last crystal structure without a collision is accepted and sent to the minimisation stage of the CSP process.

4.2.3 Comparisons



Artemisinin Quinacridone CC1

Figure 4.2: Molecular structures and names of three molecules used to compare performances of versions I and II of the QR structure generation methods.

To compare the two versions of the QR structure generation methods we carry out CSP on three different molecules (Figure 4.2): artemisinin, quinacridone, and CC1. We run CSP searches using each version of the QR structure generation methods with artemisinin in Z'=1 | $P2_12_12_1$ and Z'=4 | $P1$ until a total of 10,000 and 50,000 successfully minimised structures are obtained, quinacridone in Z'=1 | $P2_1/c$, $P\bar{1}$ and Z'=2 | $P\bar{1}$ until a total of 10,000 and 50,000 successfully minimised structures are obtained, and CC1 in Z'=1 | $P2_1/c$, $P1$ until a total of 10,000 successfully minimised structures are obtained. Final stage lattice energy minimisations were carried out using the FIT force field [124] with atom centred multipoles up to hexadecapoles obtained from GDMA with B3LYP/6-311G** GAUSSIAN09 electron densities. Ewald summations were carried out for charge-charge, charge-dipole, and dipole-dipole interactions, while all other higher-order electrostatics up to r^{-5} and repulsion-dispersion interactions were calculated for atom-atom distances within a 30 Å cut-off radius. Duplicated crystal structures were identified and removed by generating simulated X-ray diffraction patterns using PLATON [108] and calculating powder pattern similarities as described in Chapter 4.

Figure 4.3 shows the energy-density plots of the CSPs for both versions of the QR structure generation methods for all three molecules. Versions I and II produce energy-density plots that are visually quite similar, although on a closer inspection some differences between the CSPs of the artemisinin and

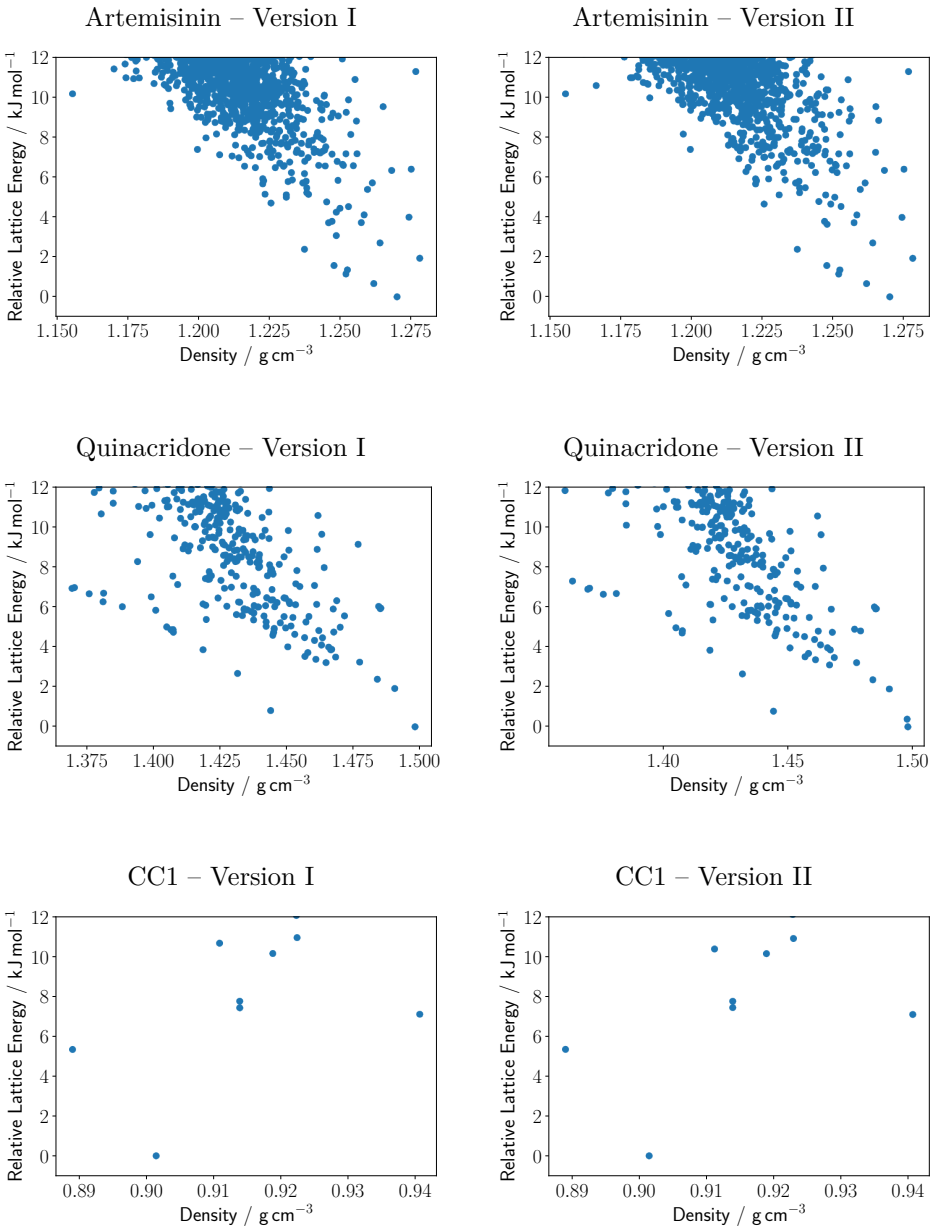


Figure 4.3: Energy-density plots of the molecules used to compare versions I and II of the QR structure generation methods.

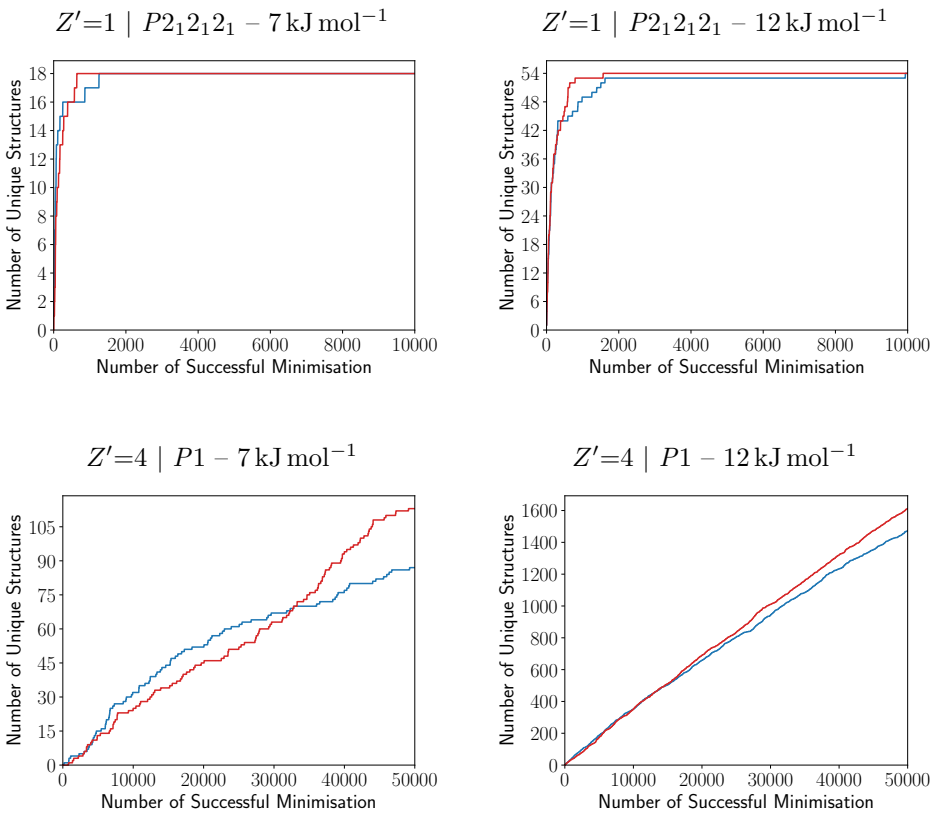


Figure 4.4: The number of unique structures found within 7 and 12 kJ mol^{-1} from the global minimum of that Z' | SG search as a function of the number of successful minimisations. Each line plot shows the progress of the CSP for the artemisinin molecule using version I (blue) or version II (red) of the QR structure generation methods.

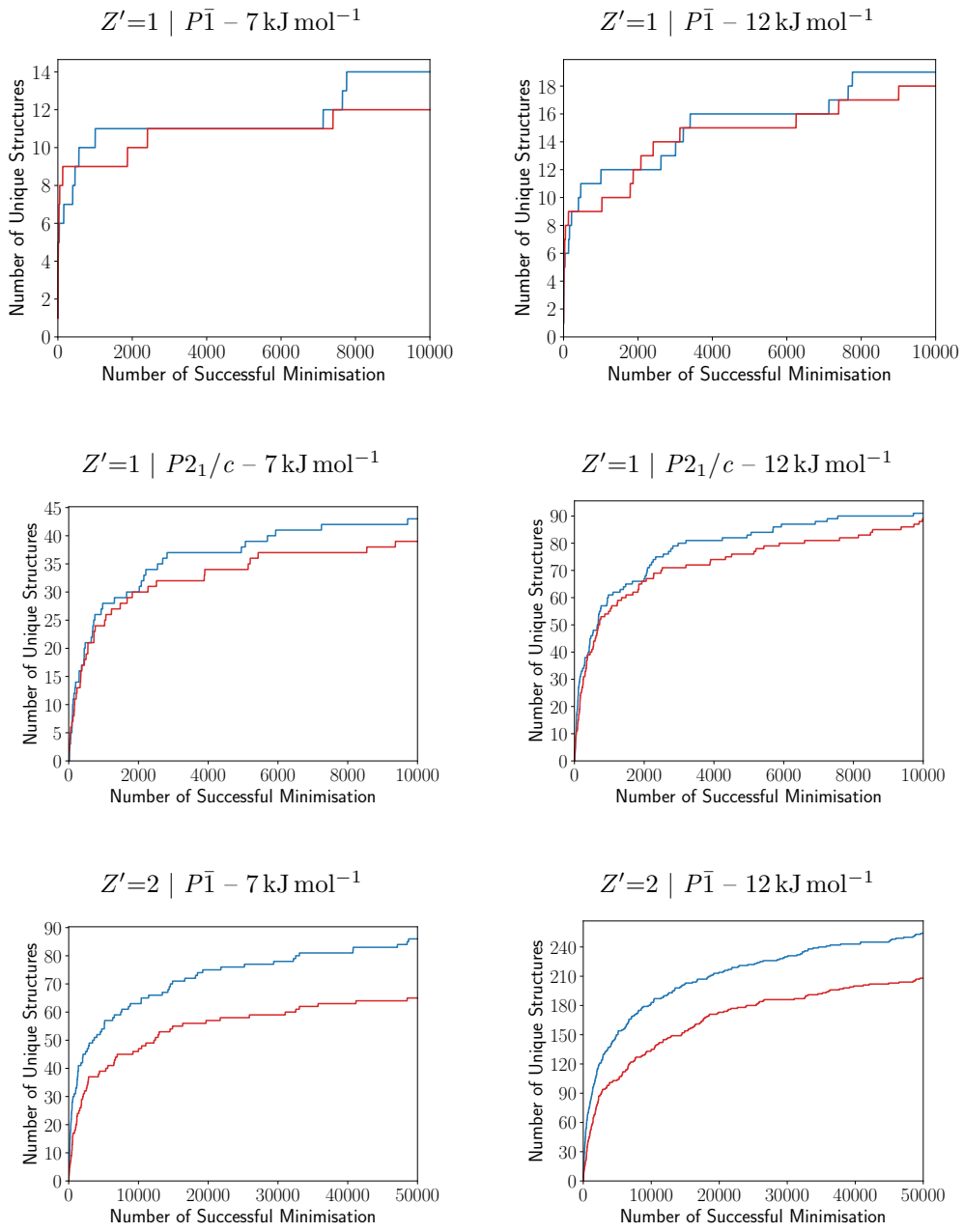


Figure 4.5: The number of unique structures found within 7 and 12 kJ mol^{-1} from the global minimum of that Z' | SG search as a function of the number of successful minimisations. Each line plot shows the progress of the CSP for the quinacridone molecule using version I (blue) or version II (red) of the QR structure generation methods.

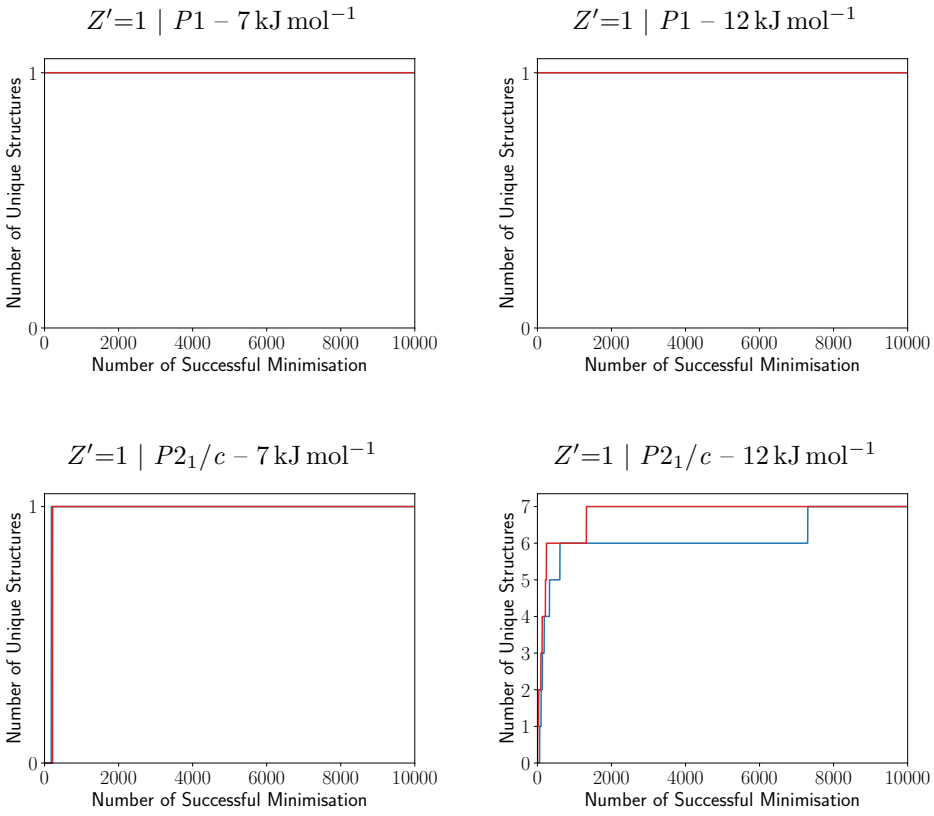


Figure 4.6: The number of unique structures found within 7 and 12 kJ mol^{-1} from the global minimum of that Z' | SG search as a function of the number of successful minimisations. Each line plot shows the progress of the CSP for the CC1 molecule using version I (blue) or version II (red) of the QR structure generation methods.

quinacridone are seen while for CC1 both energy-density plots are the same. Figures 4.4 to 4.6 shows the number of unique structures found within 7 and 12 kJ mol^{-1} from the global minimum for that particular Z' | SG search as a function of the number of successful minimisation. For the CSP of artemisinin in Z'=1 | $P2_12_12_1$ we see that both versions find unique structures at a similar rate but it is clear that version II is slightly faster at certain points; for Z'=4 | $P1$ a clear difference between the two versions is seen with version II finding more structures at a faster rate than version I. The CSPs of quinacridone in Z'=1 show a slightly faster rate of finding unique structures when version I is used and has more unique structures at the end of its search; for Z'=2 | $P\bar{1}$ a clear difference between the two versions is seen with version I showing a much faster rate of finding unique structures and produces a much larger number of unique structures at the end of its search. In the CSPs of CC1 there are only a few structures in the low energy region so the rate of finding unique structures and the number of unique structures found at the end of the CSPs are almost the same for both versions.

For this small test set of molecules, we can see that overall the results do not show either version being particularly better than the other. Version I appears to be better for sampling the crystal structure landscape of quinacridone, version II appears to be better for artemisinin and both versions perform similarly for CC1. The main differences in the rates and the final number of unique structures for artemisinin and quinacridone are likely to be down to the different mappings of the QR number to the unit cell parameters as well as the different minimisation algorithms used to remove collisions. Since both methods are based on QR numbers it is likely that on average across a large number of molecules a similar performance will be seen for both versions. Further results including the final Sobol seed used, number of rejected QR structures, number of failed minimisations, and the time taken to complete each CSP is given in Appendix C.

4.3 DMACRYS

CSPy optimises crystal structures generated from the QR sampling procedure using PMIN and DMACRYS. In both programs, crystal structures are minimised

by optimising lattice parameters, and rigid molecule positions and orientations. We first minimise QR structures using PMIN [125] as it allows us to fix the cell angles during the minimisation to avoid structures with small cell angles. For the PMIN minimisations, we use the exp-6 force field with point charges from MULFIT fitted to multipoles up to hexadecapoles from GMDA. PMIN minimised structures are followed up by further reoptimisations using DMACRYS [37]. The DMACRYS minimisations are carried out in two stages the first with the exp-6 force field with point charges from MULFIT under 0.1 GPa of pressure and finally with the exp-6 force field with multipoles up to hexadecapoles from GDMA.

4.3.1 Parameters

DMACRYS supports a number of different atom-atom potentials which can be used to optimise the crystal structure. In this project, we use the exp-6 force field with distributed multipoles for the electrostatic interactions.

$$V_{KR} = A_{KR}\exp(-B_{KR}r_{KR}) - C_{KR}r_{KR}^{-6} + E_{KR}^{\mathrm{DMA}} \tag{4.17}$$

This atom-atom potential model requires three parameters for each pair of atom types. There are two sets of parameters used in this project which creates a set of force fields known as the W99 and FIT force fields. In Table 4.1 we show a few of the different ways that the atom types are defined and their associated NEIGHCRYS labels [126].

The W99 force fields were originally developed by Williams for the C-H group hydrogen and the three and four coordinated carbon atom types by fitting parameters to 134 hydrocarbon crystal structures with 8 heats of sublimation [127]. The force field was then changed to include one and two coordinated oxygen and the alcohol and carboxyl group hydrogen atom types with revisions made to the hydrocarbon parameters by fitting to 124 oxohydrocarbon crystal structures and 7 heats of sublimation [128]. The force field was then extended to include two coordinated carbon, N-H group hydrogen, four different nitrogen atom types by fitting to 76 azahydrocarbon crystal structures and 11 heats of sublimation [124]. In W99 only parameters for the interactions between the same atom types were defined independently with the remaining terms derived using the following

Element	Details	W99	FIT
H	C-H group	H_W1	H_F1
H	Alcohol group	H_W2	H_F2
H	Carboxyl group	H_W3	H_F2
H	N-H group	H_W4	H_F2
C	Bonded to two atoms	C_W2	C_F1
C	Bonded to three atoms	C_W3	C_F1
C	Bonded to four atoms	C_W4	C_F1
N	Bonded to two or more hydrogens	N_W1	N_F1
N	Bonded to one hydrogens	N_W2	N_F1
N	Not bonded to any hydrogens and not in a triple bond	N_W3	N_F1
N	In a triple bond	N_W4	N_F1
O	Bonded to one atoms	O_W1	O_F1
O	Bonded to two atoms	O_W2	O_F1
F	Fluorines	-	F_F1

Table 4.1: A selection of atom types defined by the W99 and FIT force fields in NEIGHCRYS.

combining rules

$$A_{KR} = (A_{KK}A_{RR})^{1/2}$$

$$B_{KR} = \frac{1}{2}(B_{KK} + B_{RR}) \tag{4.18}$$

$$C_{KR} = (C_{KK}C_{RR})^{1/2}$$

so that the overall number of parameters to be determined has been reduced. Additionally, the centre of all hydrogen atom potentials are moved onto a foreshortened hydrogen position and the E_{KR}^{DMA} contributions were modelled using distributed point charges fitted to reproduce the electrostatic potentials of an HF/6-31G** electron density.

The FIT force fields [129] are based on an older set of parameters derived by Williams *et al.* [130–132] which similarly to the W99 force field, fixed parameters between different atom types with the combining rules Eq. (4.18), and used foreshortened hydrogen positions and point charges to model the electrostatic

potential. The main differences are that the FIT force field uses a distributed multipoles series to model its electrostatic potential. In particular, Coombes *et al.* [129] outlined a FIT+0.9DMA method which derives the multipoles from an HF/6-31G** electron density that was scaled to 0.9 of its calculated values. Additionally FIT leaves hydrogen atom potentials centred onto the hydrogen nuclei and not their foreshortened positions and parametrises a further hydrogen atom type for polar hydrogens using 13 hydrogen-bonded structures and 6 heats of sublimation.

In this project, we use both the W99 and FIT force fields except that in both cases the electrostatic potential models are replaced with atom centred distributed multipoles calculated using GDMA with B3LYP/6-311G** electron densities. We have therefore assumed that the electrostatic potential models between the ones derived for the W99 and FIT+0.9DMA force fields should be relatively similar to one derived using GDMA with the B3LYP/6-311G** electron density so that reparameterising the force field with this electrostatic model would not likely lead to major differences. Comparisons of the FIT and W99 force fields using B3LYP/6-311G** derived multipoles as used here with other force field and electronic-structure methods for lattice energy calculations using the X23 set of crystal structures are given by Nyman *et al.* [133].

4.3.2 Ewald Summation

As previously mentioned in Section 3.3 the lattice energy can be calculated by summing over all unique interacting terms between the molecules in one unit cell and all other molecules in the crystal structure.

$$U = \sum_{\mathcal{A},\mathcal{B}>\mathcal{A}} E_{\mathcal{A}\mathcal{B}}^{\text{exch}} + E_{\mathcal{A}\mathcal{B}}^{\text{disp}} + E_{\mathcal{A}\mathcal{B}}^{\text{elst}} \tag{4.19}$$

However due to the long-range terms up to r^{-3} within the electrostatic interaction and the fact that the number of interactions increases by approximately r^2 the summations in Eq. (4.19) may be conditionally convergent. DMACRYS carries out Ewald summations [134, 135] on the charge-charge, charge-dipole, and dipole-dipole interactions so that the long-range conditionally convergent contributions

to the energies are removed. The remaining exchange, dispersion, and short-ranged electrostatic interaction terms are simply summed to a specified cut-off.

The electrostatic energy for a periodic structure of point multipoles consisting of a charge and dipole can be written as [135]

$$\mathcal{E}_{q\mu}^{\text{elst}} = \frac{1}{2}\sum_{KL\mathbf{n}} \hat{M}_K(\mathbf{r}')\hat{M}_L(\mathbf{r})\frac{1}{|\mathbf{r}'-\mathbf{r}+\mathbf{n}|} \tag{4.20}$$

and the Ewald summation of this is

$$\begin{aligned}\mathcal{E}_{q\mu}^{\text{ES}} =& \frac{1}{2}\sum_{KL\mathbf{n}} \hat{M}_K(\mathbf{r}')\hat{M}_L(\mathbf{r})\frac{\operatorname{erfc}(\alpha|\mathbf{r}'-\mathbf{r}+\mathbf{n}|)}{|\mathbf{r}'-\mathbf{r}+\mathbf{n}|} \\ &+ \frac{1}{2\pi V}\sum_{\mathbf{k}\neq\mathbf{0}} \frac{\exp(-\pi^2k^2/\alpha^2)}{k^2}|S(\mathbf{k})|^2 + \mathcal{E}_{1,3}^{\text{self}}\end{aligned} \tag{4.21}$$

where

$$\hat{M}_L(\mathbf{r})V(\mathbf{r}) = z_L V(\mathbf{r}) + \sum_\alpha \mu_{L\alpha}\frac{\partial V(\mathbf{r})}{\partial r_\alpha}\bigg|_{\mathbf{r}_L} \tag{4.22}$$

$$S(\mathbf{k}) = \sum_L [z_L + 2\pi i\boldsymbol{\mu}_L\cdot\mathbf{k}]\exp(2\pi i\mathbf{k}\cdot\mathbf{r}_L) \tag{4.23}$$

$$\mathcal{E}_{q\mu}^{\text{self}} = -\frac{\alpha}{\sqrt{\pi}}\sum_L\left(z_L^2 + \frac{2\alpha^2}{3}\mu_L^2\right) \tag{4.24}$$

and the summations over K and L run over nuclei in the unit cell and the summation over K, L and $\mathbf{n}$ in Eq. (4.20) and Eq. (4.21) does not include the terms with $K = L$ and $\mathbf{n} = \mathbf{0}$. We can calculate the electrostatic interactions between the rigid body molecules by removing the direct summation of intramolecular electrostatic charge and dipole energies from the Ewald summation.

$$E_{q\mu}^{\text{ES}} = \mathcal{E}_{q\mu}^{\text{ES}} - \sum_{\mathcal{A}}\sum_{\substack{K\in\mathcal{A},L\in\mathcal{A}\\L>K}} \hat{M}_K(\mathbf{r}')\hat{M}_L(\mathbf{r})\frac{1}{|\mathbf{r}'-\mathbf{r}|} \tag{4.25}$$

We then replace the electrostatic energy between charges and dipoles in Eq. (4.19) with Eq. (4.25)

$$\sum_{\mathcal{A},\mathcal{B}>\mathcal{A}} E^{\text{elst}}_{\mathcal{A}\mathcal{B}} \rightarrow E^{\text{ES}}_{q\mu} + \sum_{\mathcal{A},\mathcal{B}>\mathcal{A}} E^{\text{elst}}_{5,\mathcal{A}\mathcal{B}} \tag{4.26}$$

where $E^{\text{elst}}_{5,\mathcal{A}\mathcal{B}}$ is the direct summation of the electrostatic energy for the remaining higher-order interaction terms up to r^{-5} between multipoles up to hexadecapoles of molecules $\mathcal{A}$ and $\mathcal{B}$. The lattice energy calculated by DMACRYS is therefore

$$U_{\text{intr}} = E^{\text{ES}}_{q\mu} + \sum_{\mathcal{A},\mathcal{B}>\mathcal{A}} E^{\text{exch}}_{\mathcal{A}\mathcal{B}} + E^{\text{disp}}_{\mathcal{A}\mathcal{B}} + E^{\text{elst}}_{5,\mathcal{A}\mathcal{B}} \tag{4.27}$$

which is the intrinsic contribution to the energy and is absolutely convergent. It is important to note that the direct summation is not equivalent to the Ewald summation so that in other words $\mathcal{E}^{\text{elst}}_{q\mu} \neq \mathcal{E}^{\text{ES}}_{q\mu}$. The Ewald summation removes the conditionally convergent contribution by specifically disregarding the problematic $\mathbf{k} = \mathbf{0}$ point which for a charge-neutral system leads to an undefined 0/0 result.

4.3.3 Axis Frames

The electrostatic interactions can be calculated between Cartesian-tensor or the spherical-tensor multipoles. DMACRYS uses spherical-tensors multipoles so that a system of global and local molecular axis frames can be used to determine the orientational dependence of the electrostatic interaction. With this method, the multipole moments can be fixed to their values in the local axis frame and the orientational dependence of the interaction can be moved into a spherical-tensor term $T_{KRl_1\kappa_1 l_2\kappa_2}$. So the electrostatic energy using real spherical-tensor multipole moments can be written as [90]

$$E^{\text{DMA}}_{\mathcal{A}\mathcal{B}} = \sum_{KR} \sum_{l_1 l_2 \kappa_1 \kappa_2} {}^{\mathcal{A}}Q_{Kl_1\kappa_1} {}^{\mathcal{B}}Q_{Rl_2\kappa_2} T_{KRl_1 l_2 \kappa_1 \kappa_2} \tag{4.28}$$

where

$$T_{KRl_1\kappa_1 l_2\kappa_2} = \begin{pmatrix} l_1+l_2 \\ l_1 \end{pmatrix} \bar{S}_{KRl_1\kappa_1 l_2\kappa_2(l_1+l_2)} |\mathbf{r}_R - \mathbf{r}_K|^{-l_1-l_2-1} \tag{4.29}$$

$$\begin{aligned} \bar{S}_{KRl_1\kappa_1 l_2\kappa_2 j} =& i^{l_1-l_2-j} \begin{pmatrix} l_1 & l_2 & j \\ 0 & 0 & 0 \end{pmatrix}^{-1} \sum_{k_1 k_2} \sum_{m_1 m_2 m} X_{\kappa_1 k_1} [D^{l_1}_{m_1 k_1}(\mathbf{\Omega}_{\mathcal{A}})]^* \\ & \times X_{\kappa_2 k_2} [D^{l_2}_{m_2 k_2}(\mathbf{\Omega}_{\mathcal{B}})]^* C_{jm}(\theta,\phi) \begin{pmatrix} l_1 & l_2 & j \\ m_1 & m_2 & m \end{pmatrix} \end{aligned} \tag{4.30}$$

and the atoms K are associated with the molecule $\mathcal{A}$ and the atoms R are with the molecule $\mathcal{B}$ so for example ${}^{\mathcal{A}}Q_{Kl\kappa}$ are the real spherical-tensor multipole in a local axis frame of the molecule $\mathcal{A}$. The 2 by 3 array of numbers are Wigner 3-j symbols, $C_{jm}(\theta,\phi)$ are the renormalised spherical harmonics, $X_{\kappa k}$ are transformation coefficients, $D^l_{mk}(\mathbf{\Omega})$ are Wigner rotation matrix elements, and $\mathbf{\Omega} = (\alpha,\beta,\gamma)$ are Euler angle rotations. So in Eq. (4.28) the real spherical-tensor multipole moments are transformed and rotated since

$${}^{\mathrm{G}}Q_{Klm} = \sum_{\kappa k} {}^{\mathcal{A}}Q_{Kl\kappa} X_{\kappa k} [D^l_{mk}(\mathbf{\Omega}_{\mathcal{A}})]^* \tag{4.31}$$

so that ${}^{\mathrm{G}}Q_{Klm}$ are the spherical-tensor multipole moments in the global axis frame and $\mathbf{\Omega}_{\mathcal{A}}$ are the Euler angle rotations which rotate the global axis frame to the local axis frame of molecule $\mathcal{A}$. The complex conjugation on the Wigner rotation matrices reverses the rotation.

A few of the spherical-tensors are shown below

$$
\begin{aligned}
T_{KR0000} &= \frac{1}{r_{RK}} & T_{KR1\alpha00} &= \frac{r_{\mathcal{A}\alpha}}{r_{RK}} \\
T_{KR2000} &= \frac{3}{2}\frac{r_{\mathcal{A}z}^2 - 1}{r_{RK}^3} & T_{KR21c00} &= \frac{\sqrt{3} r_{\mathcal{A}x} r_{\mathcal{A}z}}{r_{RK}^3} \\
T_{KR21s00} &= \frac{\sqrt{3} r_{\mathcal{A}y} r_{\mathcal{A}z}}{r_{RK}^3} & T_{KR22c00} &= \frac{\sqrt{3}}{2}\frac{(r_{\mathcal{A}x}^2 - r_{\mathcal{A}y}^2)}{r_{RK}^3} \\
T_{KR22s00} &= \frac{\sqrt{3} r_{\mathcal{A}x} r_{\mathcal{A}y}}{r_{RK}^3} & T_{KR1\alpha1\beta} &= \frac{3 r_{\mathcal{A}\alpha} r_{\mathcal{B}\beta} + c_{\alpha\beta}}{r_{RK}^3}
\end{aligned}
\tag{4.32}
$$

with $r_{RK}^2 = (\mathbf{r}_R - \mathbf{r}_K) \cdot (\mathbf{r}_R - \mathbf{r}_K)$, $c_{\alpha\beta} = \mathbf{e}_{\mathcal{A}\alpha} \cdot \mathbf{e}_{\mathcal{B}\beta}$, $r_{\mathcal{A}\alpha} = \mathbf{e}_{\mathcal{A}\alpha} \cdot \mathbf{e}_{RK}$, and $r_{\mathcal{B}\beta} = -\mathbf{e}_{\mathcal{B}\beta} \cdot \mathbf{e}_{RK}$ where $\mathbf{e}_{\mathcal{A}\alpha}$ is a unit vector of the local axis frame of molecule $\mathcal{A}$ and $\mathbf{e}_{RK} = (\mathbf{r}_R - \mathbf{r}_K)/|\mathbf{r}_R - \mathbf{r}_K|$. A full list of terms up to r_{RK}^{-6} are given by Stone [115]. Since in this project both the exchange-repulsion and the dispersion interactions are isotropic they will only depend on r_{RK}. The total interaction energies can therefore be written as a function of 16 dot products which are all the different components of $r_{\mathcal{A}\alpha}$, $r_{\mathcal{B}\beta}$ and $c_{\alpha\beta}$ and r_{RK}.

4.3.4 Quasi-Newton Method

Crystal structures are optimised in DMACRYS using the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm [37, 136–142], a quasi-newton method that provides an equation for the next trial structure based on the first-order derivatives and an approximation of the second-order derivatives

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \delta\mathbf{x}_k \tag{4.33}$$

$$\delta\mathbf{x}_k = -\mathbf{B}_k^{-1}\mathbf{F}_k \tag{4.34}$$

where the subscript k denotes the current iteration of the optimisation. The elements of $\mathbf{x}_k$ are the lattice parameters and molecular coordinates of the crystal

structure and $\delta\mathbf{x}_k$ is a displacement of these coordinates which are multiplied by a scaling factor α_k which depends on specific conditions that are met during the line search stage of the algorithm [140]. $\mathbf{F}_k$ is a vector of first-order derivatives and $\mathbf{B}_k$ is an approximation of the Hessian matrix.

The approximation of the inverse of $\mathbf{B}_k$ is updated with each iteration of the optimisation procedure using the equations.

$$\mathbf{B}_{k+1}^{-1} = \left(\mathbf{I} - \rho_k \mathbf{s}_k \mathbf{y}_k^{\mathrm{T}}\right) \mathbf{B}_k^{-1} \left(\mathbf{I} - \rho_k \mathbf{y}_k \mathbf{s}_k^{\mathrm{T}}\right) + \rho_k \mathbf{s}_k \mathbf{s}_k^{\mathrm{T}} \tag{4.35}$$

$$\mathbf{s}_k = \alpha_k \delta\mathbf{x}_k \qquad \mathbf{y}_k = \mathbf{F}_{k+1} - \mathbf{F}_k \qquad \rho_k = \frac{1}{\mathbf{y}_k^{\mathrm{T}} \mathbf{s}_k} \tag{4.36}$$

BFGS gives equations for the inverse Hessian matrix for the next iterations but requires an initial guess for $\mathbf{B}_0^{-1}$. In DMACRYS $\mathbf{B}_0$ is calculated using

$$\mathbf{B}_0 = \frac{1}{2}\left(\mathbf{H}_0 + \mathbf{H}_0^{\mathrm{T}}\right) \tag{4.37}$$

where $\mathbf{H}_0$ is an approximation of the Hessian and contains analytical second derivatives for all terms except the cross terms between the translation, rotation, and strain coordinates which are set to zero. A symmetrised Hessian is used in Eq. (4.37) as the orientation-orientation derivatives for orientational coordinates of the same molecule lead to an asymmetric Hessian since the infinitesimal rotation operators (angular momentum operators) do not commute and lead to the commutator relation

$$[\hat{J}_\alpha, \hat{J}_\beta] = i\epsilon_{\alpha\beta\gamma}\hat{J}_\gamma \tag{4.38}$$

and therefore

$$\left(\frac{\partial}{\partial\theta_A}\frac{\partial}{\partial\phi_A} - \frac{\partial}{\partial\phi_A}\frac{\partial}{\partial\theta_A}\right) U = -\frac{\partial}{\partial\psi_A} U \tag{4.39}$$

so that the Hessian is only symmetric at a stationary point since the first-order derivatives will be zero. The derivatives of the lattice energy with the $6N + 6$ variables, 3 positions, and 3 orientations for the N molecules in the unit cell and

6 variables for the unit cell and are obtained through a chain rule [141] with the variables of the atom-atom interactions in Eq. (4.28).

The BFGS algorithm, therefore, starts by first calculating an approximation of the initial symmetric Hessian Eq. (4.37) which is used to determine the search direction with Eq. (4.34). A scaling factor is determined by running a line search along the search direction until specific conditions are met and the move Eq. (4.33) is then made. The inverse symmetric Hessian of the new position is determined using Eq. (4.35) and another line search and move is made, this is then repeated until specific convergence conditions are reached. In DMACRYS further iterations of the BFGS algorithm are stopped when the step sizes or energy changes are small. For the cases when the minimisations are required under specific space group symmetries, DMACRYS uses projection operators to form symmetry adapted Hessian matrices and derivatives to determine symmetry adapted displacements [37].

Chapter 5

MolBuilder

MolBuilder is a molecule fragment-based chemical structure optimisation program that applies an evolutionary algorithm (EA) onto a population of molecules. Some aspects of the work in this chapter have been described in the published article Ref. [143] and was itself a generalisation of some earlier work carried out by Josh E. Campbell who had applied a genetic algorithm on a population of nitrogen substituted polycyclic aromatic hydrocarbons [144]. MolBuilder, therefore, includes some major changes from Josh E. Campbells work including support for a larger number of molecules but maintains some design choices such as the fitness functions.

5.1 Search Space

To reduce the time required for the optimisation of a molecule's chemical structure by an EA it is necessary to restrict the target search space that can be explored. This is carried out in MolBuilder using a selection of molecules, fragments, and mutations. MolBuilder has specific operations that are applied onto these molecules, fragments, or mutations to create other molecules. In MolBuilder the search space of molecules that can be accessed by the EA is defined in terms of five variables, an example of which is given in Listing 1. The size of the search space can be found by finding all unique molecules that can be made by the random molecule generator as described in Section 5.3.

```
molecules = ['c1ccccc1', 'c1ccsc1']
fragments = ['c1c**cc1', 'c1**sc1', '*1*csc1', 'c1c***1',
↪  'c1cc2cccc3c2c(c1)***3', 'c1c****1', '*1**s*1']
mutations_1 = ['[#6R1]-F', '[#6R1&H]', '[#7R1&H0]']
mutations_2 = ['[#8R1r5]', '[#16R1r5]', '[#7R1r5&H]',
↪  '[#6R1r5]=O']
molsize = [4, 6]
```

Listing 1: Example settings for the five variables used in MolBuilder to define the search space that the EA can explore.

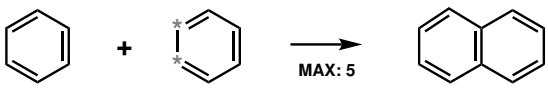
In Listing 1 the variable `molecules` is a list of molecules represented in the SMILES [145, 146] string format, this list is used by MolBuilder as a starting point to build larger molecules. The variable `fragments` is a list of fragments that are used by MolBuilder to attach onto molecules to create larger molecules. These fragments are also represented in the SMILES string format but unlike the molecule SMILES strings they have asterisks in place of atoms which denote the fragments attachment type and position. The variables `mutations_1` and `mutations_2` are lists of SMARTS [147, 148] strings and are used to modify a molecule. The SMARTS strings are used to match and replace specific groups or atoms. The list `molsize` gives the minimum and maximum size of the molecules defined by the number of rings present.

5.2 Molecular Operators

There are four types of molecular operators in MolBuilder and are named addition, crossover, recombination, and mutation. These operators have been developed for the exploration of planar aromatic molecules for organic electronic applications and so assumes that the input molecules and fragments are a planar aromatic of some sort. All four molecular operators break or create bonds of the molecules and fragments and are carried out using RDKits molecule objects [110].

5.2.1 Addition

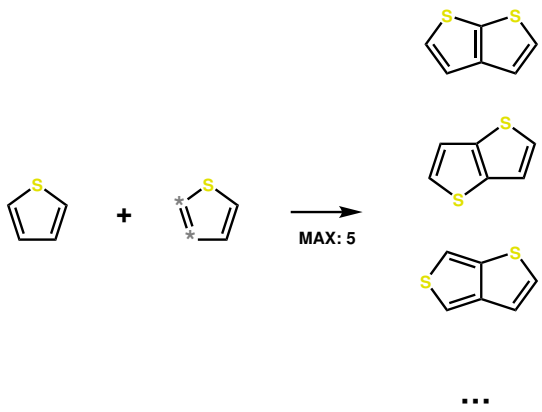
The addition operator takes a molecule, a list of fragments, and a maximum molecule size and returns a molecule for example



$$A(\texttt{'c1ccccc1'},\ [\texttt{'c1c**cc1'}],\ 5) = \texttt{'c1ccc2ccccc2c1'} \tag{5.1}$$

which has added a two-point benzene fragment `'c1c**cc1'` to a benzene molecule `'c1ccccc1'`. The exact position of the benzene molecule that the fragment attaches onto and the orientation that the fragment takes are randomly chosen. Due to the symmetries of both the molecule and fragment, in all cases this addition operation results in a naphthalene molecule. We can use the addition operator again onto this molecule with another attachable fragment creating an even larger molecule.

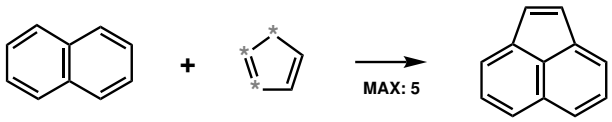
In certain cases, the molecule or fragments have an asymmetry of some sort so the position of the molecule that the fragment attaches onto and the orientation of the fragment can result in a range different of molecules



$$A(\texttt{'c1ccsc1'},\ [\texttt{'c1**sc1'}],\ 5) = \begin{cases} \texttt{'c1csc2c1ccs2'} & 0 \leq x < 1 \\ \texttt{'c1csc2c1scc2'} & 1 \leq x < 2 \\ \texttt{'c1csc2c1csc2'} & 2 \leq x < 3 \\ \cdots \end{cases} \tag{5.2}$$

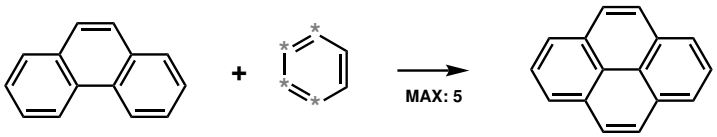
where x is a random number in the range $[0, N)$ and N is the total number of possibilities for the addition of the thiophene molecule `'c1ccsc1'` and a two-point thiophene fragment `'c1**sc1'`. Since two or more attachment positions may result in the same molecule, duplicate entries in the right side of Eq. (5.2) can occur and will increase the probability of this molecule forming after the addition operation.

There are other attachable fragments with more than two attachment points such as the three-point cyclopentadiene fragment `'c1c***1'` and four-point benzene fragment `'c1c****1'`. These types of fragments are chosen to attach to specific molecules that have positions available for the position operator to be applied. For example, the naphthalene molecule is a candidate for `'c1c***1'` to be attached to forming an acenaphthylene molecule



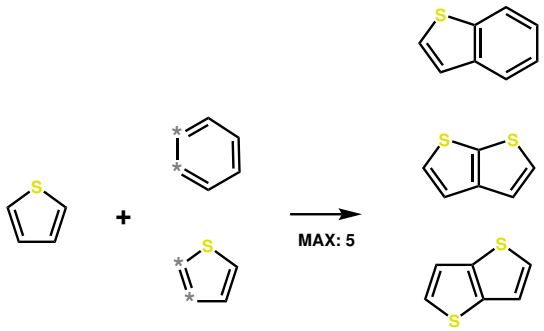
$$A(\texttt{'c1ccc2ccccc2c1'},\ [\texttt{'c1c***1'}],\ 5) = \texttt{'c1cc2cccc3c2c(c1)C=C3'} \tag{5.3}$$

while phenanthrene is a candidate for `'c1c****1'` to be attached to forming a pyrene molecule.



$$A(\texttt{'c1ccc2c(c1)ccc3c2cccc3'},\ [\texttt{'c1c****1'}],\ 5) = \texttt{'c1cc2ccc3cccc4c3c2c(c1)cc4'} \tag{5.4}$$

When more than one fragment is present in the list of fragments then the possibility of all molecules that can be formed is the sum of the possibilities with each fragment



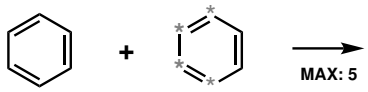
$$A(\texttt{'c1ccsc1'},\ \texttt{fragments},\ 5) = \begin{cases} \texttt{'c1ccc2c(c1)ccs2'} & 0 \leq x < 1 \\ \texttt{'c1csc2c1ccs2'} & 1 \leq x < 2 \\ \texttt{'c1csc2c1scc2'} & 2 \leq x < 3 \\ \ldots & \end{cases} \tag{5.5}$$

$$\texttt{fragments} = [\texttt{'c1c**cc1'},\ \texttt{'c1**sc1'}]$$

where x is a random number in the range $[0, N)$ and N is the total number of

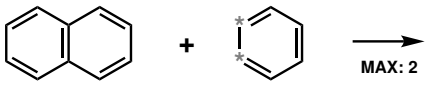
possibilities for the addition of the thiophene molecule and a two-point thiophene fragment or a two-point benzene fragment.

In the examples above we have shown cases where the addition operator was able to find and create at least one molecule from the inputs. There are a number of cases where there is no chemically reasonable ways that a new molecule could be created or no new molecule that can be created within the specified molecule sizes. For these cases, the addition operator simply returns `None`. One example of the addition operator returning a `None` result is the addition of a benzene molecule and a four-point benzene fragment



$$A(\texttt{'c1ccccc1'},\ \texttt{['c1c****1']},\ 5) = \texttt{None} \tag{5.6}$$

since there is simply no position on the benzene molecule where the four-points can attach onto which retains the molecules and fragments hexagonal structure. Another example occurs for the addition operation between a naphthalene molecule and a two-point benzene fragment with a maximum molecule size of 2



$$A(\texttt{'c1ccc2ccccc2c1'},\ \texttt{['c1c**cc1']},\ 2) = \texttt{None} \tag{5.7}$$

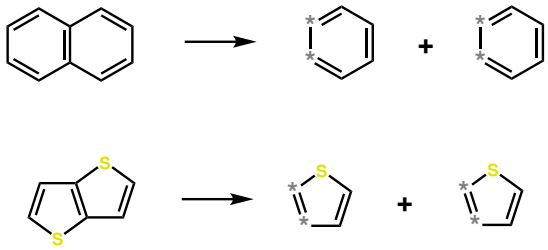
since it will always result in a molecule with a molecule size of 3.

5.2.2 Crossover

The crossover operator takes in three variables, two molecules, and a molecule size list and returns two molecules for example

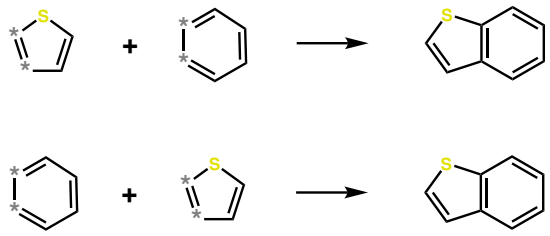
$$C(\texttt{'c1ccc2ccccc2c1'}, \texttt{'c1csc2c1scc2'}, [1, 5]) \\ = \texttt{'c1ccc2c(c1)ccs2'} + \texttt{'c1ccc2c(c1)ccs2'} \tag{5.8}$$

which is a simple example of the crossover between the naphthalene and thieno[3,2-b]thiophene molecules producing two 1-benzothiophene molecules. The crossover operator can be broken down into two individual steps or suboperations. Firstly crossover fragments each molecule into two parts along an intersecting bond. For naphthalene, this will result in two two-point benzene molecules and for thieno[3,2-b]thiophene, two two-point thiophene molecules.



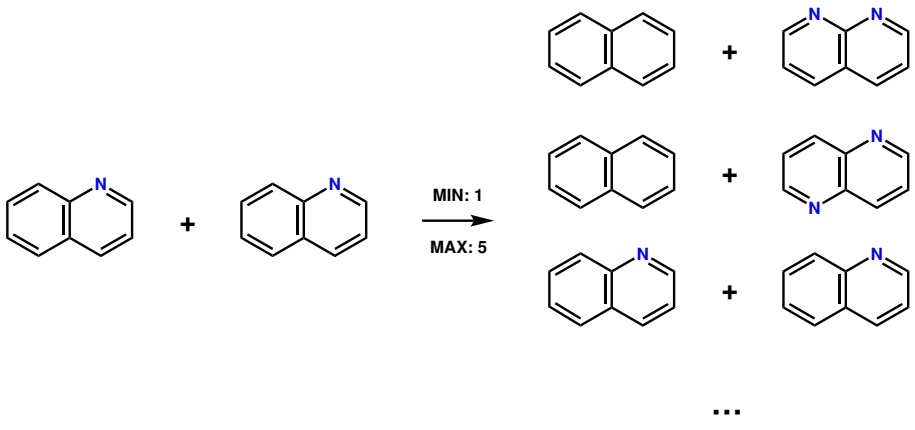
$$\tilde{F}(\texttt{'c1ccc2ccccc2c1'}) = \texttt{'c1c**cc1'} + \texttt{'c1c**cc1'} \\ \tilde{F}(\texttt{'c1csc2c1scc2'}) = \texttt{'c1**sc1'} + \texttt{'c1**sc1'} \tag{5.9}$$

Two addition suboperations are carried out using one fragment from each of the above fragmentation suboperations resulting in the crossover molecules Eq. (5.8)



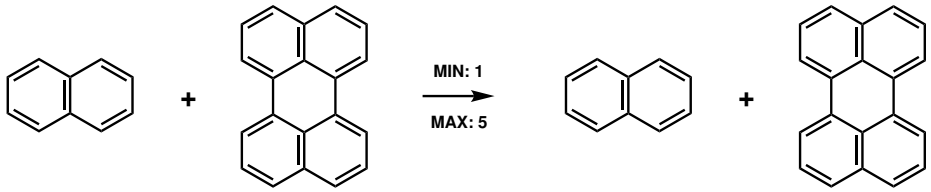
$$\tilde{A}(\texttt{'c1**sc1'},\ \texttt{'c1c**cc1'}) = \texttt{'c1ccc2c(c1)ccs2'}$$
$$\tilde{A}(\texttt{'c1c**cc1'},\ \texttt{'c1**sc1'}) = \texttt{'c1ccc2c(c1)ccs2'} \tag{5.10}$$

the addition suboperator is similar to the addition operator described in Section 5.2.1 except this suboperator works on two fragments and attaches them together at their two attachment points with a random orientation. For more complex molecules the crossover operator can result in one of a number of different pairs of molecules for example the crossover between two quinoline molecules.



$$C(\texttt{'c1ccc2c(c1)cccn2'}, \texttt{'c1ccc2c(c1)cccn2'}, [1, 5])$$
$$= \begin{cases} \texttt{'c1ccc2ccccc2c1'} + \texttt{'c1cc2cccnc2nc1'} & 0 \leq x < 1 \\ \texttt{'c1ccc2ccccc2c1'} + \texttt{'c1cc2c(cccn2)nc1'} & 1 \leq x < 2 \\ \texttt{'c1ccc2c(c1)cccn2'} + \texttt{'c1ccc2c(c1)cccn2'} & 2 \leq x < 3 \\ \ldots \end{cases} \tag{5.11}$$

Some additional rules within the crossover operator ensure that the resulting molecules are within the molecule size ranges and that the suboperations in crossover can be made to always produce valid fragments or molecules. In some situations there are no valid structures that can be made, the crossover operator will then return the input molecules. For example, the attempted crossover between the perylene molecule and naphthalene will return the input molecules.



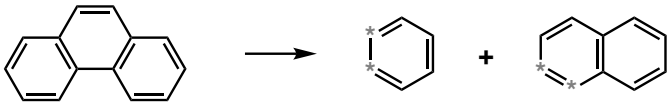
$$C(\texttt{'c1ccc2ccccc2c1'}, \texttt{'c1cc2cccc3c2c(c1)c4cccc5c4c3ccc5'}, [1, 5])$$
$$= \texttt{'c1ccc2ccccc2c1'} + \texttt{'c1cc2cccc3c2c(c1)c4cccc5c4c3ccc5'} \tag{5.12}$$

5.2.3 Recombination

The recombination operator takes in one molecule and returns an isomer of the original molecule, for example

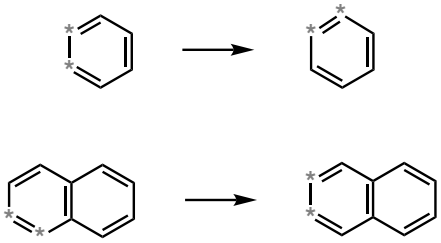
$$R(\texttt{'c1ccc2c(c1)ccc3c2cccc3'}) = \texttt{'c1ccc2cc3ccccc3cc2c1'} \tag{5.13}$$

which is the recombination of phenanthrene into an anthracene molecule. The recombination operator can be broken down into three individual suboperations, fragmentation, switch, and addition. The recombination operator first runs by fragmenting the molecule into two fragments



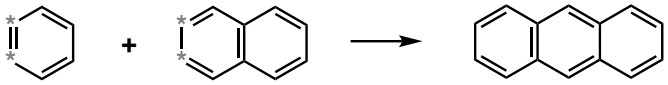
$$\tilde{F}(\texttt{'c1ccc2c(c1)ccc3c2cccc3'}) = \texttt{'c1c**cc1'} + \texttt{'*1ccc2ccccc2*1'} \tag{5.14}$$

where in this example the phenanthrene molecule has been fragmented to a two-point benzene fragment and a two-point naphthalene fragment. The switch suboperator is applied to each fragment which moves their attachment points around



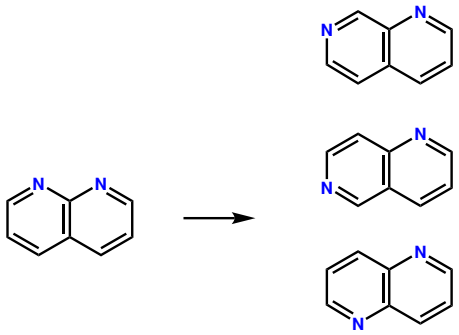
$$\tilde{S}(\texttt{'c1c**cc1'}) = \texttt{'c1cc**c1'}$$
$$\tilde{S}(\texttt{'*1ccc2ccccc2*1'}) = \texttt{'*1*cc2ccccc2c1'} \tag{5.15}$$

and an addition suboperation is then carried out onto each fragment forming the anthracene molecule.



$$\tilde{A}(\texttt{'c1cc**c1'},\ \texttt{'*1*cc2ccccc2c1'}) = \texttt{'c1ccc2cc3ccccc3cc2c1'} \tag{5.16}$$

Similarly to crossover, the recombination operator has additional rules which ensure that all suboperations result in valid fragments or molecules. Since the position and orientations of the suboperators are random there can be more than one possibility for a single molecule.

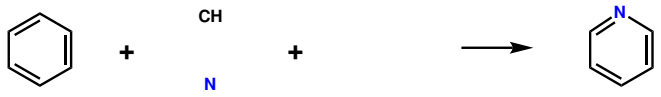


...

$$R(\texttt{'c1cc2cccnc2nc1'}) = \begin{cases} \texttt{'c1cc2ccncc2nc1'} & 0 \leq x < 1 \\ \texttt{'c1cc2cnccc2nc1'} & 1 \leq x < 2 \\ \texttt{'c1cc2c(cccn2)nc1'} & 2 \leq x < 3 \\ \ldots \end{cases} \tag{5.17}$$

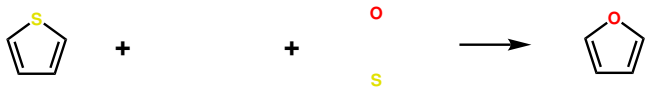
5.2.4 Mutation

The mutation operator takes in one molecule and two lists of mutations and returns a molecule for example



$$M(\texttt{'c1ccccc1'},\ \texttt{['[\#6R1\&H]',\ '[\#7R1\&H0]']},\ \texttt{[]}) = \texttt{'c1ccncc1'} \tag{5.18}$$

where we have left the `mutations_2` list variable empty. The mutation operator requires two separate lists of mutations as the different mutations require a different set of operations on the RDKit molecule objects for a valid molecule to be created. The first list `mutations_1` should contain SMARTS expressions that match positions of atoms that are only within one ring except for the positions defined for the second list. The second list `mutations_2` should contain SMARTS expressions that match the heteroatom position of an aromatic five-membered ring such as the sulfur atom in the thiophene molecule or the nitrogen atom in the pyrrole molecule for example



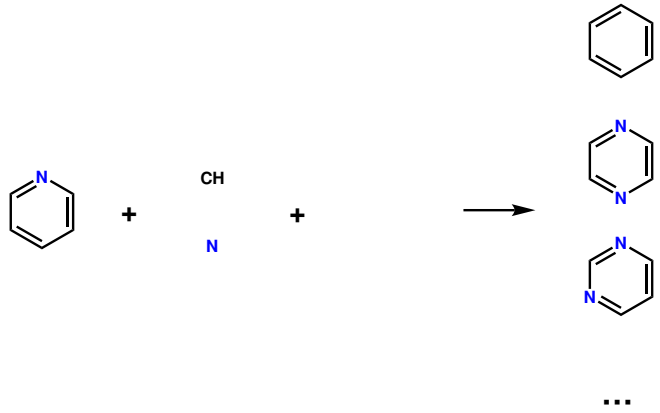
$$M(\texttt{'c1ccsc1'},\ \texttt{[]},\ \texttt{['[\#8R1r5]',\ '[\#16R1r5]']}) = \texttt{'c1ccoc1'} \tag{5.19}$$

which describes the mutation of the thiophene into a pyrrole molecule where `mutations_1` is empty and `mutations_2` contains SMARTS expressions.

The mutation operator works by first determining the mutable positions 1 and 2 by finding all atom positions that match the expressions in the list `mutations_1` or `mutations_2`. Next, a random position is selected out of all the positions and a mutation is carried out by replacing the atoms at that position with a different substructure from `mutations_1` for mutable positions 1 or `mutations_2` for mutable positions 2. For cases where there are no matches for either list, the operator returns the input molecule

$$M(\texttt{'c1ccccc1'},\ \texttt{[]},\ \texttt{['[\#8R1r5]',\ '[\#16R1r5]']}) = \texttt{'c1ccccc1'} \tag{5.20}$$

and as with all other operators, a number of different molecules can be formed.



$$M(\texttt{'c1ccncc1'},\ \texttt{mutations\_1},\ \texttt{[]}) = \begin{cases} \texttt{'c1ccccc1'} & 0 \leq x < 1 \\ \texttt{'n1ccncc1'} & 1 \leq x < 2 \\ \texttt{'c1ncncc1'} & 2 \leq x < 3 \\ \ldots & \end{cases} \tag{5.21}$$

$$\texttt{mutations\_1} = \texttt{['[\#6R1\&H]',\ '[\#7R1\&H0]']}$$

5.3 Random Molecule Generation

In the evolutionary algorithm, an initial population is required which we obtain by randomly generating a set of molecules using the addition and mutation operators described in Section 5.2. First, a molecule is randomly selected from the list `molecules` and a random number in the range `molsize` is chosen. The addition operator is carried out continually onto the molecule using fragments from the `fragments` list until a molecule is generated with a size equal to the random number. If at any point the addition operator returns `None` the molecule is discarded and the molecule generation is restarted. If a molecule has been generated at the required size, a specified number of mutation operators are applied onto this molecule using mutations from `mutations_1` and `mutations_2` which results in a randomly generated molecule.

The size of the search space that can be accessed in the initial population is found by running the random molecule generator in an infinite loop and converting each molecule generated into its canonical InChi [149] string which is then added to a Python set. The size of the search space is determined to be the number of elements in this Python set after no changes have occurred over several days. The total size of the target search space (defined as the total number of unique molecules that can be accessed by MolBuilder throughout the EA) is in most cases equivalent to the size of the search space that can be accessed in the initial population. Therefore in most cases, the size of the target search space can be determined using the random molecule generation method explained here.

5.4 Evolutionary Algorithm

```
population_size = 100
initial_number_mutations = 500
elitism_population_size = 10
mutation_rate = 0.05
recombination_rate = 0.05
tournament_win_rate = 0.75
total_generations = 30
maximize = False
```

Listing 2: Example settings used in MolBuilder which controls certain parts of the EA.

With an initial population of randomised molecules, a method of selection is made to favour molecules with specific properties and generate another population of molecules that are likely to have similar characteristics and therefore increase the likelihood of finding high performing molecules for the target property with each generation. With the configuration Listing 2 the EA starts by first generating an initial population of 100 randomised molecules using the molecule generation algorithm with 500 mutations. The fitness of each molecule in this initial population (generation 1) is carried out. We then carry out a selection and modification process to generate the next population. We use an elitist type selection so that the 10 best performing molecules are selected and placed into the next population.

The remaining molecules are generated with the following procedure: two molecules are selected by running two 2-way tournament selections, a single 2-way tournament selection runs by randomly selecting two molecules from the current population (generation 1) and then chooses the fittest molecule out of the two with a probability of 75%. A crossover operation is then carried out on the two molecules from the two 2-way tournament selections to create two child molecules, each child molecule undergoes a mutation with a probability of 5% and recombination with a probability of 5%. These tournament selections are carried out continually until the new population reaches the specified population size of 100. The new generation (generation 2) is now set as the current generation, the

fitness of each molecule is evaluated and a new generation is again created from this one, this is done continuously until generation 30 is reached.

5.5 Fitness Functions

In the evolutionary algorithms a fitness function is defined which determines the likelihood that one individual undergoes some selection and mutation process. For the computational screening of organic semiconductors, we are looking to search and target organic molecules that form crystal structures that could be used to produce high-performance OFET devices. Since in general, such functions do not exist we must therefore design one that selects for important properties that a molecule must have to form a high-performance OFET. A simple fitness function could be formed which seeks to maximise the average carrier mobility of the crystal structure landscape of a molecule

$$F = \langle \mu \rangle = \sum_i \mu_i P_i \qquad P_i = \frac{e^{-\beta \Delta E_i}}{\sum_i e^{-\beta \Delta E_i}} \tag{5.22}$$

where ΔE_i is the energy difference from the global minimum, μ_i is the carrier mobility for the crystal structures i, $\beta^{-1} = 2.70\,\mathrm{kJ\,mol^{-1}}$ is a constant obtained by fitting energy differences between experimentally known polymorphs [46, 150] and all summations run through all crystal structures within an energy range from the global minimum structure for that CSP search.

Even for the relatively simple fitness function Eq. (5.22) some major challenges need to be overcome before such a function could be used. The computational cost involved per molecule would be extremely large as the evaluation of this fitness function would require a CSP search per molecule followed by a carrier mobility evaluation per crystal structure. Additionally, this fitness function does not take into account the resulting contact resistances that an OFET device with this semiconductor may have and the impact that it will have on the overall OFET performance. Alternatively, we can instead use the EA to optimise the molecules chemical structure based only on its molecular properties so that CSP and carrier mobility evaluations can be carried out in later reevaluations stages

on a smaller subset containing the best performing molecules.

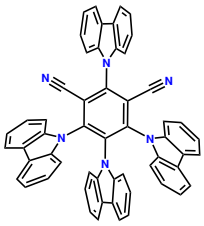
Since carrier mobilities can be approximated as a sum of Marcus theory rates a reasonable first approach would be to design a molecular fitness function that reduces the reorganisation energy at the isolated molecule approximation Eq. (3.144). Since the gas-phase adiabatic electron affinity or ionisation energies can be obtained from the reorganisation energy calculations without any further calculations, we can also use this to reduce the contact resistances that the OFET device might have. Using these two parameters we have designed two molecule-based fitness functions to maximise electron mobilities whilst minimising the OFET contact resistances

$$F_{1,W}(\Lambda_-, \mathrm{A_s}) = \begin{cases} \Lambda_- + (W - \mathrm{A_s}) & \mathrm{A_s} < W \\ \Lambda_- & \mathrm{A_s} \geq W \end{cases} \tag{5.23}$$

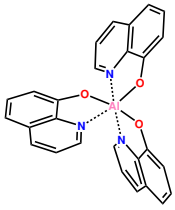
$$F_{2,W}(\Lambda_-, \mathrm{A_s}) = \sqrt{\Lambda_-^2 + (W - \mathrm{A_s})^2} \tag{5.24}$$

where Λ_- is the electron reorganisation energy, W is the work function of the contact, and $\mathrm{A_s}$ is the solid-state electron affinity of the semiconductor. So that we have included a contribution due to the Schottky barrier $\Theta = W - \mathrm{A_s}$ in our fitness functions to ensure that the barrier for the injection of an electron into the organic semiconductor is reduced and overall OFET device performances are increased. So the Schottky barrier for the injection of an electron is a property that is dependent on the work function of the metal and the electron affinity of the thin-film or crystal of the molecule.

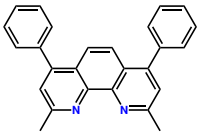
Instead of carrying out solid-state electron affinity calculations we can obtain an approximation from gas-phase adiabatic electron affinities calculations and correct them using a linear regression model fitted to mean values of experimental solid-state electron affinities determined from low-energy inverse photoemission spectroscopy (LEIPS) [151]. We use mean values as in some cases thin-films of molecular semiconductor stack in different orientations resulting in different experimentally determined electron affinities. For example, pentacene thin-films stack in the standing orientation with a SiO_2 substrate and the lying orientation with a highly oriented pyrolytic graphite substrate and have experimental LEIPS



4CzIPN

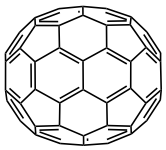


Alq_3

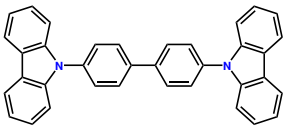


BCP

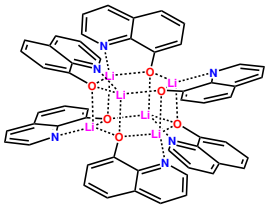
C60



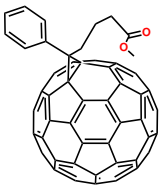
C70



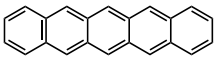
CBP



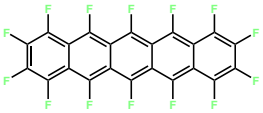
Liq



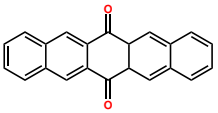
PBCM



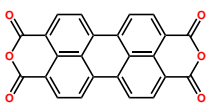
PEN



PFP



PNQ



PTCDA

Figure 5.1: Molecular structures and names of the 12 molecules used to fit calculated gas-phase electron affinities to LEIPS experimental solid-state electron affinities. Reproduced from Ref. [143] with permission from the Royal Society of Chemistry.

electron affinities of 2.35 and 3.14 eV [152] respectively. In total, the reference set contains mean values of solid-state electron affinities for twelve different small organic and organometallic molecules whose molecular structures are shown in Figure 5.1.

The gas-phase adiabatic electron affinities were calculated with GAUSSIAN09 [153] by carrying out geometry optimisations using B3LYP/6-311+G** at the neutral state followed by another optimisation for the charged state and taking differences in the final energies between the two states. Initial geometries for the molecules 4CzIPN, Alq3, BCP, CBP, Liq, and PCBM were obtained by extracting them from the crystal structures YUGDOV [154, 155], QATMON [156, 157], TICBUD [158, 159], KANYUU [160, 161], ADATOP [162, 163], and PESJII01 [164, 165] to ensure the correct conformers were used. Initial geometries for the remaining molecules C60, C70, PEN, PFP, PNQ, and PTCDA were obtained using RDKit initial coordinates generation and UFF optimisation functions as these molecules are unlikely to form more than one conformer.

Results of the DFT gas-phase calculations and the corresponding solid-state experimental electron affinities for the twelve molecules are given in Table 5.1. A plot of the DFT gas-phase calculations and mean experimental electron affinities are shown with a linear regression model ($m = 1.00$ and $c = 1.11$ with $R^2 = 0.97$) in Figure 5.2. Therefore an approximation for the solid-state electron affinities can be obtained from B3LYP/6-311+G** gas-phase adiabatic electron affinities by the addition of a polarisation correction term of 1.1 eV

$$\mathrm{A_s} \approx \Big[E_0(R_0) - E_-(R_-)\Big] + 1.1 \tag{5.25}$$

where $E_0(R_0)$ is the ground state energy with the minimum energy nuclear coordinates R_0 for the neutral state and $E_-(R_-)$ is the ground state energy with the minimum energy nuclear coordinates R_- for the charged state.

Molecule	DFT A_g / eV	LEIPS $\bar{A}_s$ / eV	LEIPS A_s / eV	References
4CzIPN	2.03	2.81	2.81	[166]
Alq_3	1.02	2.06	2.06	[166]
BCP	0.616	1.89	1.89	[166]
C60	2.64	3.98	3.98	[167]
C70	2.69	4.00	4.00	[167]
CBP	0.786	1.75	1.75	[166]
Liq	0.749	1.85	1.85	[166]
PCBM	2.53	3.75	3.64, 3.76, 3.84	[167,168]
PEN	1.54	2.73	2.35, 2.70, 3.14	[152,169]
PFP	2.85	3.85	3.58, 4.12	[152]
PNQ	1.65	2.83	2.34, 2.83, 3.32	[170]
PTCDA	3.17	4.11	4.11	[151]

Table 5.1: Calculated B3LYP/6-311+G** gas-phase adiabatic electron affinities, average experimental LEIPS solid-state electron affinities, and experimental LEIPS solid-state electron affinities for the different molecular orientations. Reproduced from Ref. [143] with permission from the Royal Society of Chemistry.

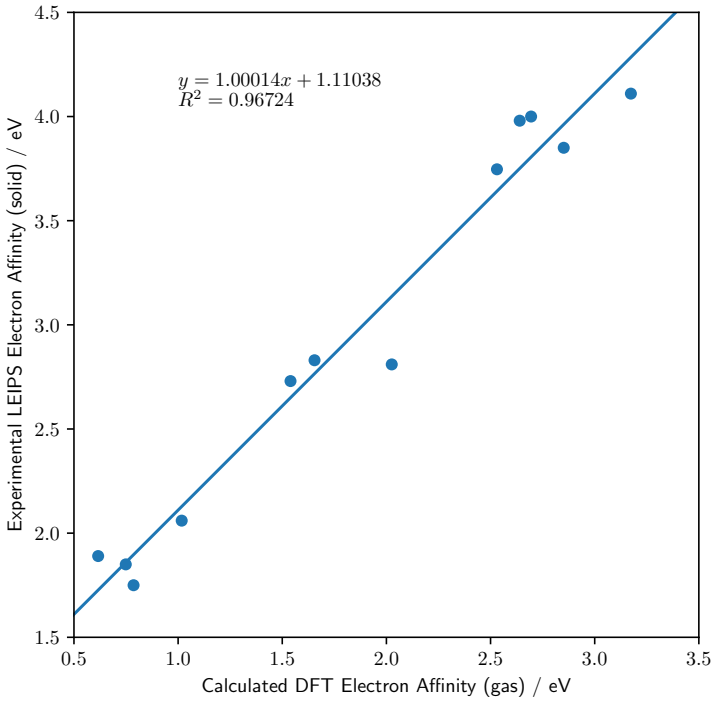


Figure 5.2: A linear fit between the average experimental LEIPS solid-state electron affinities and the calculated B3LYP/6-311+G** gas-phase adiabatic electron affinities. Reproduced from Ref. [143] with permission from the Royal Society of Chemistry.

Chapter 6

Nitrogen Substituted PAHs

The work in this chapter has been published in Ref. [143] and was based on some previous work by Josh E. Campbell [144] who had applied a genetic algorithm on a population of nitrogen substituted polycyclic aromatic hydrocarbons (PAHs). The work here follows closely the published article but contains major changes to Josh E. Campbell original work such as the CSP and mobility evaluation methods. Most importantly we used the MolBuilder program to carry out the molecular structure optimisations and use a larger search space of nitrogen substituted PAHs, example molecules are shown in Fig 6.1.

The work by Josh E. Campbell on nitrogen substituted PAHs was itself inspired by earlier computational studies on nitrogen substituted pentacene molecules for organic semiconductor applications by Chen and Chao [171], and Winkler and Houk [172]. Both Chen and Chao, and Winkler and Houk had studied the effect of nitrogen substitution on the pentacene molecule and found it to be an effective measure to modify the electron affinity of pentacene while

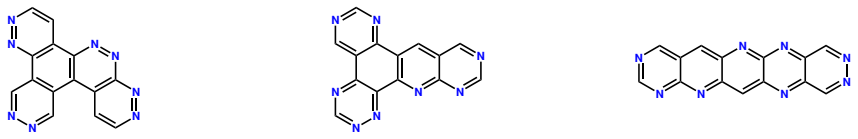


Figure 6.1: Molecular structures of four randomly generated molecules using MolBuilder with the molecules, fragments, mutations, and molecule sizes defined in Listing 3. Reproduced from Ref. [143] with permission from the Royal Society of Chemistry.

maintaining its low reorganisation energy. Although no CSP had been carried out by either Chen and Chao or Winkler and Houk they had both suggested that the increased nitrogen substitution could lead to more π-stacked sheet type structures due to the introduction of intermolecular C-H$\cdots$N interactions.

The suggestion of obtaining more favourable packing configurations for azapentacene molecules provided a good opportunity for the application of CSP methods. This led to the work by Campbell *et al.* [150] who had applied CSP to a total of six azapentacene molecules with four molecules from the work by Winkler and Houk and an extra two that were generated to examine the effect of more asymmetric nitrogen substitution patterns. The results of the CSP of six hypothetical azapentacene molecules showed how the nitrogen substitution could have a large effect on the preferred packing type of their predicted crystal structures. However, the relationship between a specific crystal-packing type and calculated carrier mobilities were not clear as variations in the mobility for a given packing type were found to be quite large.

The work in this chapter can be seen as a generalisation and continuation of the previous work on azapentacenes. Here we will be running an EA over a search space of nitrogen substituted PAHs for low reorganisation energies and high electron affinities. This will therefore be going beyond the earlier work by Chen and Chao, and Winkler and Houk by evaluating a far larger number of molecules with a greater diversity of the molecular structures. While additionally building upon the CSP work on azapentacenes by Campbell *et al.* but instead look to apply CSP to nitrogen substituted PAHs proposed by an EA.

6.1 Evolutionary Algorithm

```
molecules = ['c1ccccc1']
fragments = ['c1c**cc1']
mutations_1 = ['[#6R1&H]', '[#7R1&H0]']
mutations_2 = []
molsize = [5, 5]
```

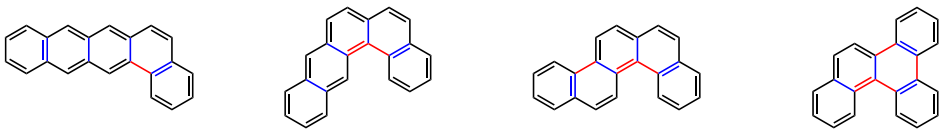
Listing 3: The five variables used to define a search space of nitrogen substituted PAHs that the EA will explore and consists of a benzene molecule, a two-point benzene fragment, C-H and N mutations, and a min and max molecule size of 5. The search space contains all aromatic structures with five six-membered rings and any number of nitrogen substitutions excluding the perylene and pyrene derivatives.

The search space is defined with the input variables Listing 3 and contains a total of 68,064 molecules which includes all aromatic structures containing five six-membered rings with any number of nitrogen substitutions excluding the perylene and pyrene derivatives. In this chapter we define the types of structures sampled by the EA with the amount of non-linearity which is defined by the number of bonds that connect two rings but not counting the intersecting bonds, examples are given in Figure 6.2. Using this system there are a total of five different groups of structures.

The settings for the EA such as the elitism populations sizes are the same as those described in Section 5.4. To evaluate a molecules ability to produce crystal structure of high electron mobilities we use the fitness function

$$F_{1,W}(\Lambda_-, \mathrm{A_s}) = \begin{cases} \Lambda_- + (W - \mathrm{A_s}) & \mathrm{A_s} < W \\ \Lambda_- & \mathrm{A_s} \geq W \end{cases} \tag{6.1}$$

where both the reorganisation energy Eq. (3.144) and electron affinities Eq. (5.25) were calculated using B3LYP/6-311+G** with GAUSSIAN09. We carry out 10 EA searches for two different values of the contacts work function. The first we target a work function of 0.0 eV this effectively reduces the fitness function to the electron reorganisation energy $F_{1,0} = \Lambda_-$ and in the second we target a work function of 4.1 eV.



1 2 3 4

Figure 6.2: Example chemical structures with their non-linearity values below, the amount of non-linearity is defined by the number of bonds (red) that connect two rings but not the intersecting bonds (blue). Reproduced from Ref. [143] with permission from the Royal Society of Chemistry.

6.2 Reorganisation Energy

The pentacene molecule is the optimum structure for the electron reorganisation energy minimisation for a search space defined by Listing 3 since pentacene is the most symmetric structure that can be formed so that a charge on this molecule can be delocalised effectively across it. We, therefore, use the fitness function $F_{1,0}$ to study the performance of the EA we have developed and run a total of 10 individual searches. Table 6.1 show some basic statistics for each EA search for its ability to find the global minimum structure pentacene. In general, we see that the EA performs well for all searches finding the global minimum after around 10 generations with most searches only sampling $\sim$1% of the search space.

In Figure 6.3 we show how the population of molecules changes as the EA progresses. Figure 6.3a shows the mean reorganisation energy of the populations which quickly decreases with each generation. In general, we see that the EA generates an initial population with mean reorganisation energy of around 0.28 eV which drops to below 0.15 eV for nearly all 10 searches at the 20^{th} generation. In Figure 6.3b we plot the minimum reorganisation energy found in the populations for each generation which shows more variability for the initial population but the minimum reorganisation energy quickly drops to the same minimum for all 10 searches. Figures 6.3a and 6.3b with Table 6.1, therefore, proves that the EA with the options set with Listing 2 works sufficiently well when optimising molecular structures with the fitness function $F_{1,0}$.

Figures 6.3c and 6.3d show a more detailed analysis of the different types

Run	Number of Generations	Molecules Sampled	Percentage of Search Space Sampled
1	9	642	0.94%
2	11	745	1.09%
3	9	672	0.99%
4	11	778	1.14%
5	15	1035	1.52%
6	8	572	0.84%
7	17	1110	1.63%
8	6	420	0.62%
9	7	513	0.75%
10	9	631	0.93%

Table 6.1: The number of EA generations required, the number of unique molecules sampled, and the percentage of the search space sampled before locating pentacene, the global minimum for the electron reorganisation energy (fitness function $F_{1,0}$) for a search space of nitrogen substituted PAHs. Reproduced from Ref. [143] with permission from the Royal Society of Chemistry.

of molecules found in the population as the EA progresses. Figure 6.3c shows the mean number of nitrogen atoms of the population and as expected the initial population for all 10 searches were found to have a mean number of nitrogen atoms of around 7. As the EA progresses we see the mean number of nitrogen atoms of the population drops steadily similarly to the decrease in the reorganisation energy. Figure 6.3d shows how linear molecules become dominant as the EA progresses however there appears to be more variability across each EA search than for example the mean number of nitrogen atoms showing that linear molecules such as pentacene may not be the only type of molecules that have low electron reorganisations energies.

From all 10 runs of the EA, we form a list of the 10 fittest molecules, molecular structures are shown in Figure 6.4. We see that as expected the pentacene molecule was found to have the smallest reorganisation energy. A majority of the next set of molecules were found to have a planar non-linear type structure with a nitrogen atom within its fjord region. The range of the reorganisation energies between these 10 molecules are extremely small ranging from 0.1346 to 0.1399 eV so that between this set of molecules the electronic coupling and

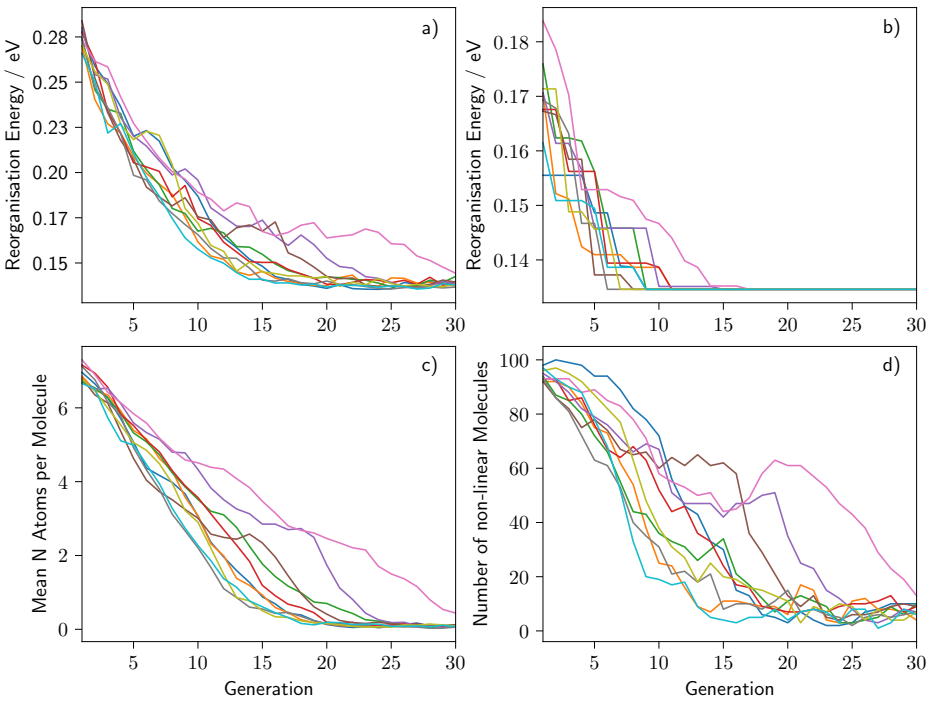
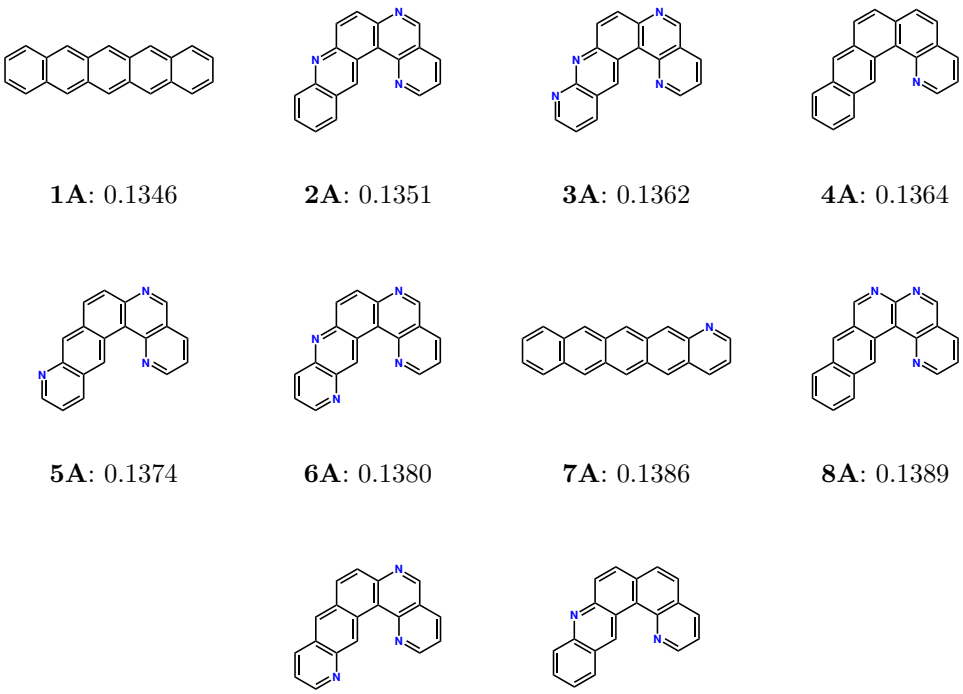


Figure 6.3: Plots showing the changes in — a) the mean reorganisation energy of the population, b) the minimum reorganisation energy, c) the mean number of nitrogen atoms per molecule, and d) the number of non-linear molecules in the population — for ten different EA searches using the fitness function $F_{1,0}$. Reproduced from Ref. [143] with permission from the Royal Society of Chemistry.



1A: 0.1346 **2A**: 0.1351 **3A**: 0.1362 **4A**: 0.1364

5A: 0.1374 **6A**: 0.1380 **7A**: 0.1386 **8A**: 0.1389

9A: 0.1394 **10A**: 0.1399

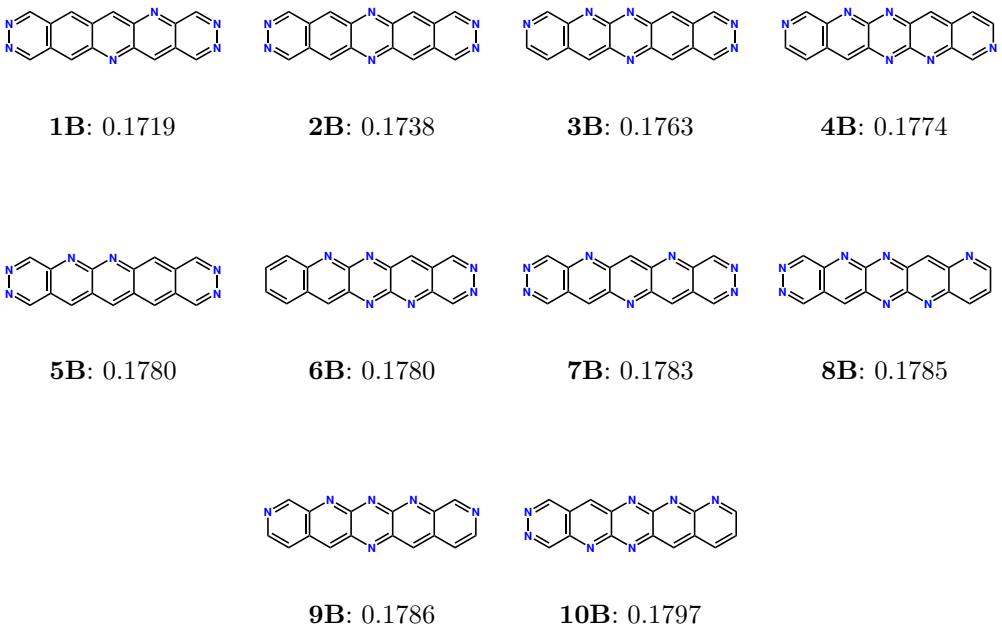
Figure 6.4: Molecular structures of the top 10 best-performing molecules from 10 runs of the EA for the minimisation of the fitness function $F_{1,0}$ (electron reorganisation energy) with molecule labels and $F_{1,0}$ fitness values in eV. Reproduced from Ref. [143] with permission from the Royal Society of Chemistry.

therefore its crystal structures will be the main factor that differentiates these molecules from each other when comparing electron mobilities between their predicted crystal structures.

6.3 Electron Affinity

We perform another 10 runs of the EA but for the minimisation of the fitness function $F_{1,4.1}$ which is set to favour molecules with solid-state electron affinities of 4.1 eV or larger which should match with higher work function contacts. The top 10 best performing molecules are shown in Figure 6.5 and are all linear azapentacenes with 6 or 7 nitrogen atoms in each molecule. Additionally, all molecules were found to have solid-state electron affinities greater than 4.1 eV showing that the fitness function had strongly disfavoured low electron affinities molecules. Table 6.2 shows the reorganisation energies and electron affinities of the top 10 molecules for the fitness functions of $F_{1,0}$ and $F_{1,4.1}$. Slightly larger reorganisation energies are obtained for the $F_{1,4.1}$ set of molecules but their larger electron affinities may lead to lower electron injection barriers and better performance organic electronic devices.

We compare the molecules suggested by MolBuilder against four reference molecules by Winkler and Houk [172], molecular structures and labels are shown in Figure 6.6. These four molecules were designed based on computational calculations and aimed to target molecules with low reorganisation energies and gas-phase electron affinities of around 3.0 eV or equivalently a solid-state electron affinity of 4.1 eV. Table 6.2 shows reorganisation energies and solid-state electron affinities of the top 10 molecules proposed by MolBuilder with fitness functions $F_{1,0}$ and $F_{1,4.1}$ and the molecules proposed by Winkler and Houk. In general, the Winkler and Houk molecules are similar to those from MolBuilder with the fitness function $F_{1,4.1}$. However, **WH5A** and **WH5B** have smaller electron affinities than the 4.1 eV target which may result in larger injection barriers for high work function contacts.

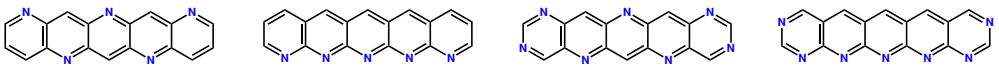


1B: 0.1719 **2B**: 0.1738 **3B**: 0.1763 **4B**: 0.1774

5B: 0.1780 **6B**: 0.1780 **7B**: 0.1783 **8B**: 0.1785

9B: 0.1786 **10B**: 0.1797

Figure 6.5: Molecular structures of the top 10 best-performing molecules from 10 runs of the EA for the minimisation of the fitness function $F_{1,4.1}$ with molecule labels and $F_{1,4.1}$ fitness values in eV. As the solid-state electron affinities of all 10 molecules are greater than 4.1 eV, fitness values are also the molecules electron reorganisation energies. Reproduced from Ref. [143] with permission from the Royal Society of Chemistry.



WH5A **WH5B** **WH7A** **WH7B**

Figure 6.6: Four azapentacenes proposed by Winkler and Houk [172] due to their low reorganisation energies and high electron affinities. Reproduced from Ref. [143] with permission from the Royal Society of Chemistry.

6.4 Property Maps

We generate property maps using the calculated electron affinities and reorganisation energies from the molecules sampled by the EA for all searches using $F_{1,0}$ and $F_{1,4.1}$. A total of 15,870 molecules were sampled corresponding to around 23.3% of the search space. In Figure 6.7 we colour the plot by the fitness functions the molecules were sampled by: $F_{1,0}$ not $F_{1,4.1}$, $F_{1,4.1}$ not $F_{1,0}$, or $F_{1,4.1}$ and $F_{1,0}$. We can see that the fitness function $F_{1,0}$ predominately sampled lower electron affinity molecules while $F_{1,4.1}$ predominately sampled higher electron affinity molecules. In Figure 6.8a we colour the plot points by the number of nitrogen atoms found in the molecule while in Figure 6.8b we colour the plot points by the amount of non-linearity of the molecule. In Figure 6.9 we mark the positions of the four molecules proposed by Winkler and Houk in the property map of nitrogen substituted PAHs sampled by the EA.

We can see that in general, an increase in the number of nitrogen atoms in the molecule resulted in an increase in the electron affinity. While an increase in the non-linearity of the molecule resulted in an increase in the reorganisation energy. In Figure 6.8b we see a shift towards higher electron affinities and lower reorganisation energies for all linear structures. Comparing figures 6.8a and 6.8b we see that a smaller number of nitrogen substitutions are required to achieve higher electron affinities in linear molecules when compared to the non-linear molecules. This information suggests two possible design rules for the n-type OFETs with nitrogen substituted PAHs based on molecular properties

- For linear nitrogen substituted PAHs, 5-7 nitrogen atoms, and for nonlinear nitrogen substituted PAHs, 8-10 nitrogen atoms, are required to obtain a solid-state electron affinity of around 4.0 eV.
- Linear nitrogen substituted PAHs are preferred over the nonlinear counterparts due to their lower reorganisations energies which will result in higher electron mobilities assuming the electronic coupling networks are similar.

where we have chosen an electron affinity of around 4.0 eV to lower electron injection barriers with higher work function contacts while avoiding highly electrophilic molecules which may not be stable.

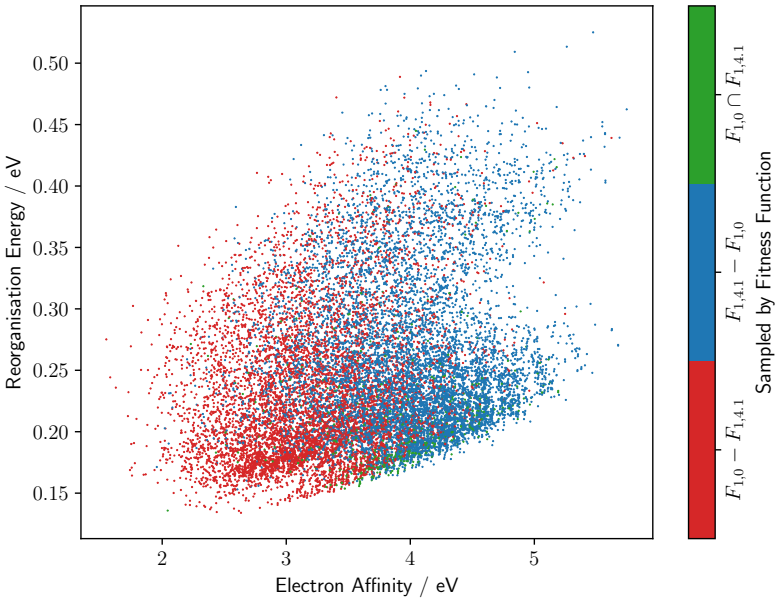


Figure 6.7: Property map containing 15,870 molecules which were sampled by the EA with 10 searches for the fitness functions $F_{1,0}$ and $F_{1,4.1}$ for a search space of nitrogen substituted PAHs defined by Listing 3. Molecules are plotted with their reorganisation energies against solid-state electron affinities and are coloured by the fitness functions that they were sampled by. Reproduced from Ref. [143] with permission from the Royal Society of Chemistry.

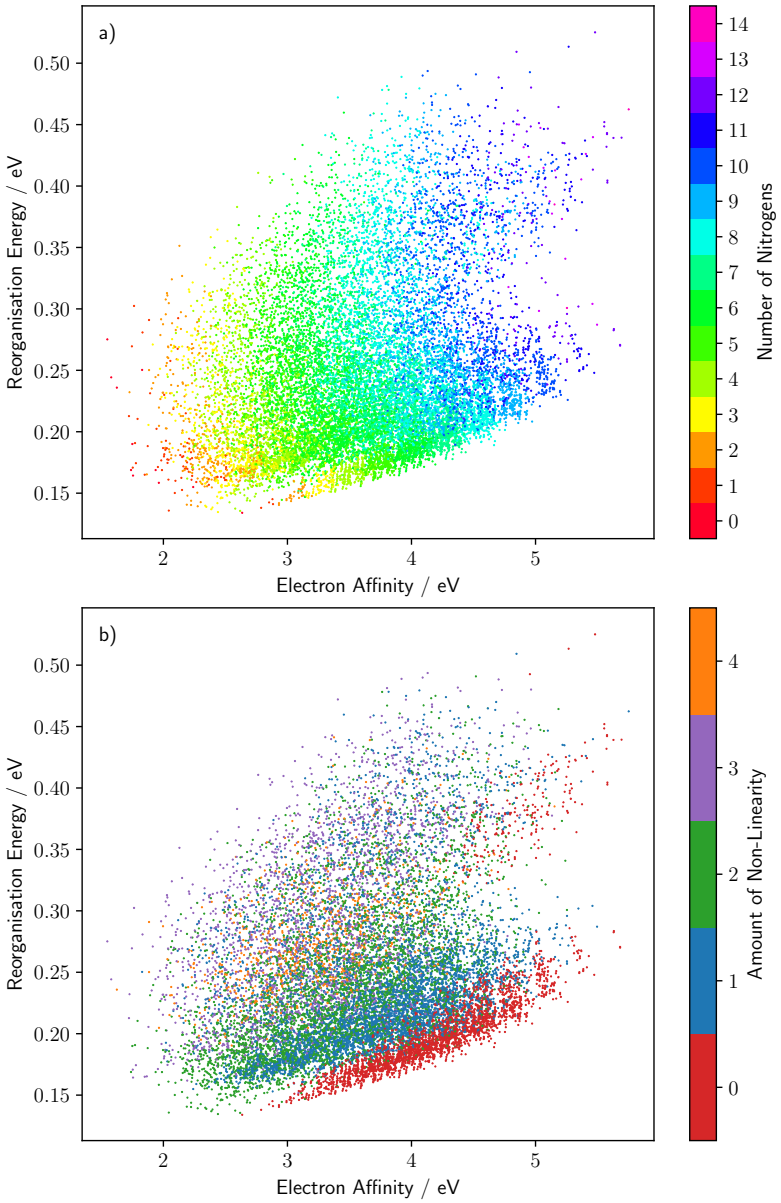


Figure 6.8: Property map containing 15,870 molecules which were sampled by the EA with 10 searches for the fitness functions $F_{1,0}$ and $F_{1,4.1}$ for a search space of nitrogen substituted PAHs defined by Listing 3. Molecules are plotted with their reorganisation energies against solid-state electron affinities and are coloured by a) the number of nitrogen atoms and b) the amount of non-linearity. Reproduced from Ref. [143] with permission from the Royal Society of Chemistry.

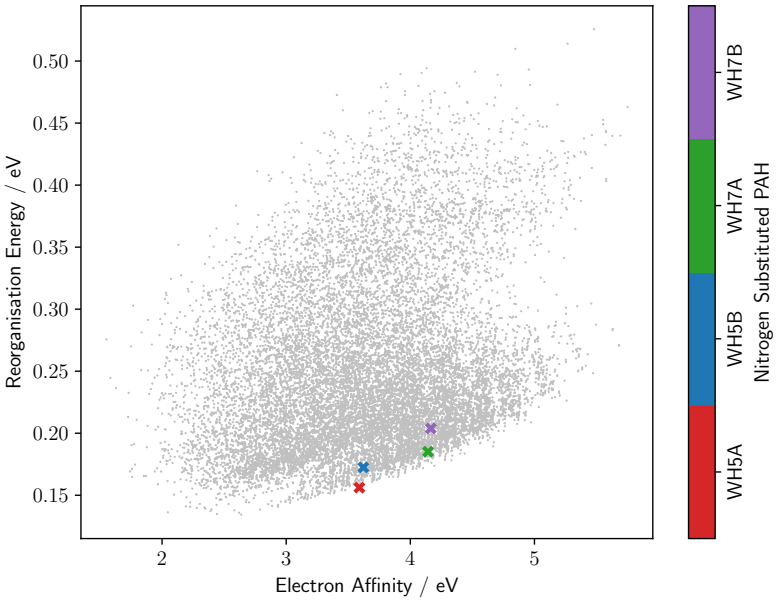


Figure 6.9: The positions of four azapentacenes proposed by Winkler and Houk [172] in the property map of molecules sampled by the EA with 10 searches for the fitness functions $F_{1,0}$ and $F_{1,4.1}$ for a search space of nitrogen substituted PAHs defined by Listing 3. Out of the four azapentacenes the molecules **WH5A**, **WH5B**, and **WH7A** had been sampled by the EA.

6.5 Electron Mobilities

We evaluate the molecules **1A-10A**, **1B-10B**, and **WH5A-WH7B** ability to produce crystal structures that result in high electron mobilities by carrying out CSP on each molecule and evaluating electron mobilities using KMC simulations with Marcus theory transition rates for each predicted crystal structure with an energy less than 7 kJ mol^{-1} from its global minimum. Each CSP search was carried out using version I of the QR structure generation method with Z'=1 until a total of 4,000 successfully minimised structures are obtained for each of the 6 most common space groups ($P2_1/c$, $P2_12_12_1$, $P\bar{1}$, $P2_1$, $Pbca$, $C2/c$) and 2,000 for each of the next 5 most common ($Pna2_1$, Cc, $Pca2_1$, $C2$, $P1$) for organic crystals with Z'=1. Final stage lattice energy minimisations were carried out using the W99 force field [124] with atom centred multipoles up to hexadecapoles obtained using GDMA with B3LYP/6-311G** GAUSSIAN09 electron densities. Ewald summations were carried out for charge-charge, charge-dipole, and dipole-dipole interactions, while all other higher-order electrostatics up to r^{-5} and repulsion-dispersion interactions were calculated for atom-atom distances within a 35 Å cut-off radius. Duplicated crystal structures were identified and removed by generating simulated X-ray diffraction patterns using PLATON [108] and calculating powder pattern similarities as described in Chapter 4.

Marcus theory transition rates between two states were calculated with reorganisation energies obtained from the EA fitness evaluations and electronic couplings which were approximated using the FO-DFT PBE/TZ2P approach implemented in the ADF program [173]. Electronic couplings were scaled by 1.325 to bring FO-DFT values in line with ab-initio calculations [174, 175] and the total number of electron coupling evaluations were reduced by finding equivalent dimers in the crystal structure using the Kabsch algorithm [176] and only evaluating them once. The components of the distance between the average position of the two states $R_{nn'\alpha}$ were approximated with the centroid-centroid distances between the two molecules. With the transition rates and distances, KMC simulations were carried out to determine electron mobilities by averaging 100,000 calculated diffusion tensors with 1,000 KMC iterations per trajectory. The isotropic electron mobilities were obtained from the mobility tensor using

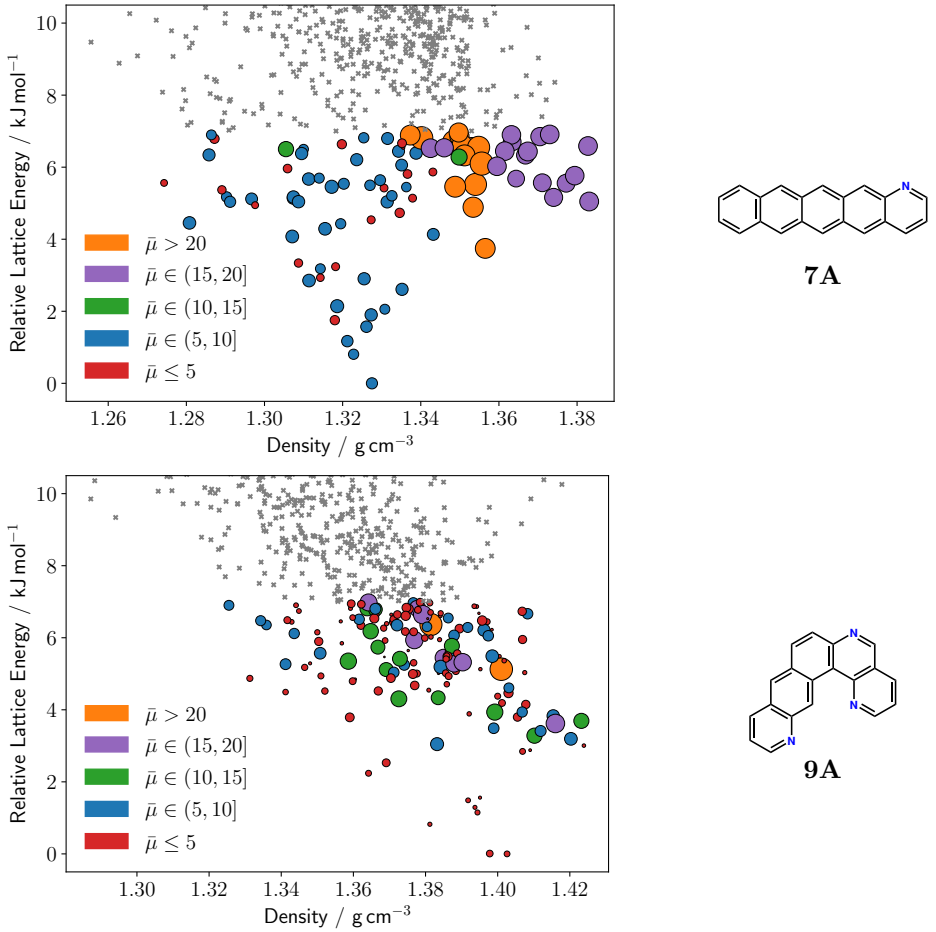


Figure 6.10: ESF maps of the molecules **7A** and **9A** which showed larger electron mobilities for their high-density crystal structures. Electron mobility calculations were carried out for all structures within 7 kJ mol^{-1} from the global lattice energy minimum, plot points are scaled and coloured by their calculated mobilities. Ranges of the mobilities defined in the ESF map legends are given in units of cm^2(Vs)$^{-1}$. Reproduced from Ref. [143] with permission from the Royal Society of Chemistry.

$\bar{\mu} = \mathrm{tr}(\boldsymbol{\mu})/3$ and for all equations, the temperature was set to 300 K.

To visualise the results we plot ESF maps for each crystal structure landscape with plot points coloured and scaled proportionally to its isotropic electron mobility, ESF maps for all 24 molecules are shown in Appendix D. The number of crystal structures in the low energy region varies between molecules. The distributions of the electron mobilities of the crystal structure in the ESF map for a given molecule appear more or less uniformly distributed. However, the ESF maps of **7A** and **9A** shown in Figures 6.10 do suggest that the higher density structures have formed to give larger carrier mobilities possibly more sheet type packing which could result in larger electronic couplings due to more favourable π-π stacking.

6.6 Evaluation

To compare each molecules ability to form high mobility organic semiconductors quantitatively we evaluate the crystal structures using three different performance metrics with results given in Table 6.2. The first is $\bar{\mu}_{\mathrm{GM}}$ which is used to evaluate the molecules based on the most likely crystal structure they will form based on the force field energies. Using $\bar{\mu}_{\mathrm{GM}}$ we can see that there is a wide range of mobilities with molecules **4A** and **WH5A** giving particularly large electron mobilities of 17.00 and 20.27 $\mathrm{cm}^2(\mathrm{Vs})^{-1}$. However, $\bar{\mu}_{\mathrm{GM}}$ is the least useful metric as there can be several structures close to the global minimum. Additionally, experimental structures may correspond to a higher energy structure in our ESF map due to inaccuracies in the force field model as well as the thermodynamic or kinetic effects we have not accounted for.

A more useful performance metric to compare molecules is the average mobility of its crystal structure landscape. Since lower energy structures have a higher probability to be formed we weight the average by this probability. Essentially we obtain an ensemble average of the crystal structure landscape of the molecule with $T = 325\,\mathrm{K}$ or $k_{\mathrm{B}}T = 2.70\,\mathrm{kJ\,mol^{-1}}$ which was obtained by comparing observed experimental polymorph pairs against calculated energies

[46, 150].

$$\langle \bar{\mu} \rangle = \sum_i \bar{\mu}_i P_i \qquad\qquad P_i = \frac{e^{-\beta \Delta E_i}}{\sum_i e^{-\beta \Delta E_i}} \tag{6.2}$$

Due to the computational costs of the mobility evaluations we cut off the summation so that it only runs over the crystal structures which are $7\,\mathrm{kJ\,mol^{-1}}$ from the global minimum which will include around 95% of all experimentally observed polymorph pairs [46]. The average mobilities across all molecules are much more similar to each other than the global minimum mobilities. The molecule with the largest average is **WH5A** followed by **2B** with 14.60 and $10.87\,\mathrm{cm^2(Vs)^{-1}}$ respectively. The molecule **2B** is an unusual case with a global minimum equal to its average due to it containing only a single structure with a relatively large electron mobility within the $7\,\mathrm{kJ\,mol^{-1}}$ cut off.

The third performance metric we use is a deviation in the mobilities of the crystal structures

$$\langle \Delta\mu^2 \rangle^{1/2} = \left[\langle \mu^2 \rangle - \langle \mu \rangle^2 \right]^{1/2} \tag{6.3}$$

as with Eq. (6.2) the averages are weighted by the probability for a given crystal to be observed experimentally with the summations cut off to include only the crystal structures which are $7\,\mathrm{kJ\,mol^{-1}}$ from the global minimum. This deviation describes the spread of the mobilities of the crystal structures landscape of the molecule. It is preferable for the deviation to be as small as possible to avoid obtaining polymorphs with lower mobilities. The deviation allows us to assess the overall risk for forming low mobility crystal structures for a given molecule. The molecule with the smallest value of the deviation is **2B** with $0.000\,\mathrm{cm^2(Vs)^{-1}}$ due to the single structure in the low energy region, **2B**, therefore, represents a particularly low-risk option compared to all other molecules.

Although not directly related to the carrier mobility the solid-state electron affinity is another important quantity to consider when fabricating an OFET device. Since we are searching for high-performance n-type OFETs we require the Schottky barrier for the injection of electrons from the contact to be as small as possible. Typically less reactive higher work function contacts are used

such as gold ($W = 5.1\,\mathrm{eV}$) we will therefore need the electron affinity of the semiconductor to match the work function of contacts like gold. The molecules **1B-10B**, **WH7A**, and **WH7B** are therefore preferable over the others since they have higher electron affinities.

Using all performance metrics we have developed it appears to us that the most favourable molecule to take into the next stages of a computational screening process such as more complete CSP searches with improved energy models and more sophisticated carrier mobility evaluations would be the molecule **2B**. The single crystal in the low energy region of **2B** would mean that this molecule would be a particularly low-risk option to take into further computational or experimental stages. Its high electron affinity would allow high work function metal contacts to be used giving lower barriers for the injection of an electron and together with its high electron mobilities **2B** has good possibilities for forming a high-performance n-type organic semiconductor.

Molecule	Number of Structures	$\bar{\mu}_{\mathrm{GM}}$ / $\mathrm{cm}^2(\mathrm{Vs})^{-1}$	$\langle\bar{\mu}\rangle$ / $\mathrm{cm}^2(\mathrm{Vs})^{-1}$	$\langle\Delta\bar{\mu}^2\rangle^{1/2}$ / $\mathrm{cm}^2(\mathrm{Vs})^{-1}$	Λ / eV	A_s / eV
1A	30	3.546	6.908	3.236	0.1346	2.640
2A	126	0.969	3.525	4.022	0.1351	2.444
3A	22	13.33	5.654	4.849	0.1362	2.623
4A	274	17.00	7.886	6.079	0.1364	2.049
5A	40	5.407	6.870	5.452	0.1374	2.391
6A	61	8.161	7.353	5.136	0.1380	2.583
7A	93	7.026	9.715	6.493	0.1386	2.791
8A	42	14.17	6.686	5.183	0.1389	2.401
9A	162	1.887	4.591	4.884	0.1394	2.351
10A	133	3.271	5.894	4.873	0.1399	2.282
1B	44	10.34	7.316	4.079	0.1719	4.101
2B	1	10.87	10.87	0.000	0.1738	4.191
3B	9	9.695	8.658	2.874	0.1763	4.168
4B	11	2.406	1.974	1.204	0.1775	4.140
5B	13	13.38	5.457	5.005	0.1780	4.112
6B	54	1.477	6.344	5.246	0.1780	4.129
7B	8	6.250	4.376	2.853	0.1783	4.329
8B	52	9.832	7.600	3.901	0.1785	4.275
9B	43	5.118	7.000	2.574	0.1786	4.141
10B	35	4.478	5.960	2.785	0.1797	4.278
WH5A	139	20.27	14.60	6.584	0.1562	3.591
WH5B	54	15.22	8.543	4.318	0.1724	3.623
WH7A	85	8.446	9.210	3.049	0.1849	4.138
WH7B	49	0.528	5.408	3.493	0.2036	4.160

Table 6.2: Electron mobility statistics of the crystal structure landscapes and electronic properties of the 10 best molecules from the EA searches with fitness functions $F_{1,0}$ and $F_{1,4.1}$ each and the 4 molecules proposed by Winkler and Houk [172]. Reproduced from Ref. [143] with permission from the Royal Society of Chemistry.

6.7 Conclusion

The work in this chapter outlines the foundations of a computational screening workflow that attempts to tackle the CSP and mobility evaluations relatively high computational cost by first using an EA to suggest the best molecules for further evaluations. There are however significant challenges that must be overcome to improve this computational screening workflow so that it can be used to confidently suggest molecules that will produce crystal structures with high electron mobilities. One difficulty of this work is knowing whether the EA stage with fitness functions $F_{1,0}$ and $F_{1,4.1}$ had assisted the computational screening in the later CSP and mobility evaluation stage.

For the search space of nitrogen substituted PAHs there are a large number of molecules with similarly small reorganisation energies. The differences in the electron mobilities of the crystal structure between these molecules will therefore be due to the electronic coupling. This issue can be seen in Table 6.2 where the molecules **1A-10A** do not show particularly higher mobilities despite lower reorganisation energies. By carrying out CSP for only 20 molecules will mean that we are at risk of missing out on other higher reorganisation energy molecules that may have crystal structures with high electron mobilities due to more favourable molecular packing. This could be easily overcome by running CSP for a large number of molecules perhaps more than several hundred. This will however increase computational costs significantly which could be alleviated with more efficient CSP and mobility evaluation methods.

Another important difficulty is the reliability of the mobility evaluation methods used which were derived through a large number of approximations, it is uncertain what the effects on the results will be quantitatively. However, we continued to use this Marcus theory based method due to its simplicity and lower computational cost as it is not the aim of this chapter to develop, evaluate, or improve models used for carrier mobilities calculations. We, therefore, took a more practical approach and focused on the development of a computational screening procedure but we acknowledge that there may be inadequacies in the carrier mobility model. Replacement with more sophisticated carrier mobility models such as TLT [64, 65] and benchmark work of these methods will be

necessary.

Despite the difficulties mentioned some interesting results were still obtained such as the existence of low risk **2B** type molecules. These molecules have only a small number of crystal structures in their low energy regions which means that applying more sophisticated carrier mobility methods such as TLT which are only affordable for a smaller number of crystal structures will be possible. Molecules like **2B** offer an opportunity to use more computationally expensive methods so that we can be more confident with our results before they are taken to the experimental stages. However the **2B** result depends on the CSP sampling scheme used, it is important to note that the CSPs carried out in this chapter did not contain any $Z' > 1$ searches. It is likely that a greater number of low energy predicted structures would be found with alternative sampling schemes since, for example, we know that we would require a Z'=2 CSP searches to obtain all the experimental polymorphs of pentacene. It is therefore clear that the space groups and sampling used which included only Z'=1 searches may not be the best choice for the CSP of these types of molecules.

Chapter 7

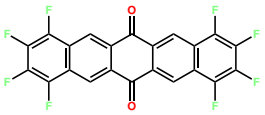
Computational Screening Benchmark

In this chapter, we will benchmark the DFT geometry optimisation, the FIT force field, the CSP sampling methods, and the Marcus theory and TLT mobility evaluation methods. For the DFT geometry optimisation, FIT force field, and CSP sampling benchmark we use a test set containing 40 polyaromatic molecules and their experimental crystal structures. Molecular structures and molecule names — which we have derived from the CSD refcodes of their associated experimental crystal structures — are shown in Figure 7.1 and the reference codes, polymorph forms, Z', and space groups of the associated crystal structures are given in Table 7.1. This set of molecules and their crystal structures were chosen to cover molecules that are

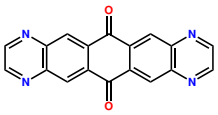
- Polyaromatics which include C-F, aromatic N, and C=O groups.
- Preferably formed from a total of five rings.
- Formed from only five-membered or six-membered rings.
- Rigid molecules without flexible functional groups.

so that the test set molecules are structurally and chemically similar to those used in our computational screening workflows such as the Nitrogen substituted PAHs in Chapter 6 and the indenofluorenedione derivatives in Chapter 8.

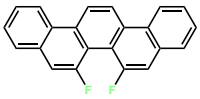
The molecular and crystal structures generated in our CSP workflow will be benchmarked against experimental references from the CSD [177–280]. Using the CSP benchmark results we will then devise an efficient sampling scheme to be used for the computational screening of small molecule organic semiconductors. For the mobility evaluation benchmark test set we use the experimental crystal structures of 10 different tetracene derivatives [68, 217, 281–292], molecular structures and CSD refcodes are shown in Figure 7.2. We choose this set of 10 tetracene derivatives as high-quality hole mobilities from FOB-SH simulations have been reported [66].



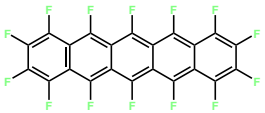
AROBUFMOL



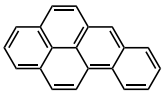
AROCAMMOL



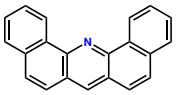
BEGKIKMOL†



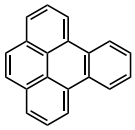
BEZLUOMOL



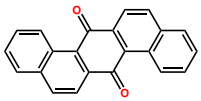
BNPYREMOL



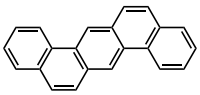
BNZACRMOL



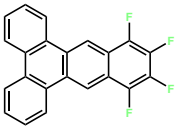
CEQGELMOL



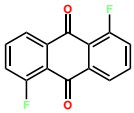
DBANQUMOL



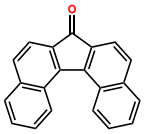
DBNTHRMOL



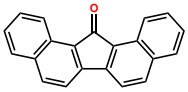
DUXSASMOL



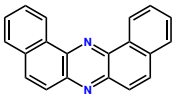
FANTRQMOL



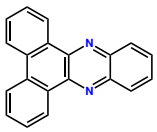
HAMDUWMOL†



KONBAQMOL

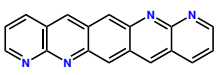


LOHVAHMOL

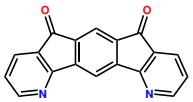


MIVRUEMOL

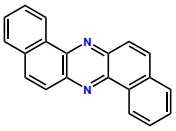
Figure 7.1: Molecular structures and names (†chiral molecules) of the 40 molecules used to benchmark the DFT geometry optimisation, the FIT force field, and the CSP sampling methods used in the computational screening workflow and to derive an efficient sampling scheme from the benchmark results.



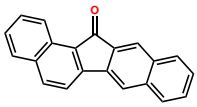
MIVSESMOL



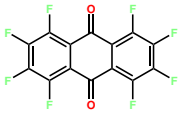
MORRODMOL



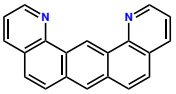
NAAZASMOL



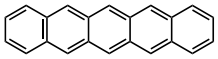
NIJCEQMOL



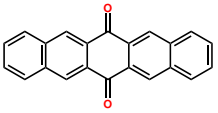
OFANTQMOL



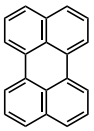
PAGMERMOL



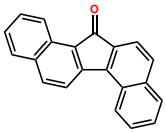
PENCENMOL



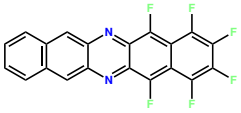
PENTQUMOL



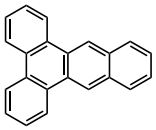
PERLENMOL



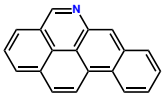
POVLUHMOL



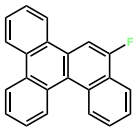
RUNLAPMOL



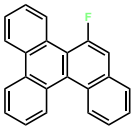
SANQIIMOL



SEVYEAMOL

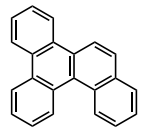


UPIRUIMOL†

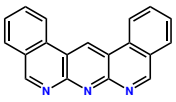


UPIXAUMOL†

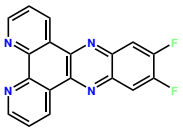
Figure 7.1: *continued*



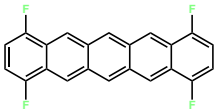
UPIXEYMOL†



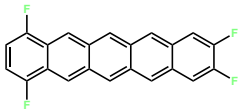
VIBMAVMOL



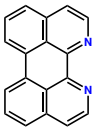
VIQTATMOL



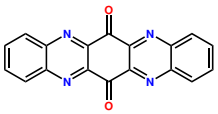
VUFKALMOL



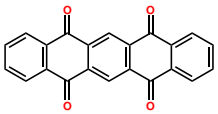
VUFNIWMOL



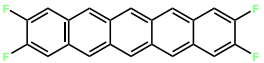
WOTZUDMOL



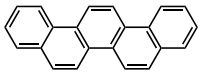
WUPYIQMOL



YOFROBMOL



ZETQUNMOL



ZZZOYCMOL

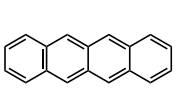
Figure 7.1: *continued*

Molecule	CSD Refcode	Polymorph	Z'	SG
AROBUFMOL	AROBUF	-	0.5	$P\bar{1}$
AROCAMMOL	AROCAM	-	0.3	$C2/m$
BEGKIKMOL	BEGKIK	-	1	$Pbca$
BEZLUOMOL	BEZLUO	-	0.5	$P2_1/c$
BNPYREMOL	BNPYRE10	monoclinic	1	$P2_1/c$
BNPYREMOL	BNPYRE11	monoclinic	1	$P2_1/c$
BNPYREMOL	BNPYRE12	monoclinic	1	$P2_1/c$
BNZACRMOL	BNZACR	-	1	$Pna2_1$
CEQGELMOL	CEQGEL	-	2	$P\bar{1}$
CEQGELMOL	CEQGEL01	-	2	$P\bar{1}$
DBANQUMOL	DBANQU	-	1	$Pca2_1$
DBNTHRMOL	DBNTHR01	orthorhombic	0.5	$Pcab$
DBNTHRMOL	DBNTHR02	orthorhombic	0.5	$Pcab$
DBNTHRMOL	DBNTHR10	monoclinic	1	$P2_1$
DUXSASMOL	DUXSAS	-	1	$P\bar{1}$
FANTRQMOL	FANTRQ10	-	$0.5 + 0.5$	$P2_1/c$
HAMDUWMOL	HAMDUW	-	1	$P2_1/n$
HAMDUWMOL	HAMDUW01	-	1	$P2_1/n$
HAMDUWMOL	HAMDUW02	-	1	$P2_1/n$
KONBAQMOL	KONBAQ	-	0.5	$Pnma$
LOHVAHMOL	LOHVAH	-	1	$P2_12_12_1$
MIVRUEMOL	MIVRUE	-	1	$P2_12_12_1$
MIVSESMOL	MIVSES	-	0.5	$C2/c$
MORRODMOL	MORROD	-	1	$P2_1/n$
NAAZASMOL	NAAZAS	-	0.5	$P2_1/c$
NIJCEQMOL	NIJCEQ	-	1	$P2_1/c$
OFANTQMOL	OFANTQ	-	1	$P112_1/b$
OFANTQMOL	OFANTQ01	-	1	$P2_1/c$
PAGMERMOL	PAGMER	-	1	$P2_1/n$
PENCENMOL	PENCEN	I	$0.5 + 0.5$	$P\bar{1}$
PENCENMOL	PENCEN01	II	$0.5 + 0.5$	$P\bar{1}$
PENCENMOL	PENCEN02	II	$0.5 + 0.5$	$P\bar{1}$
PENCENMOL	PENCEN03	II	$0.5 + 0.5$	$P\bar{1}$
PENCENMOL	PENCEN04	II	$0.5 + 0.5$	$P\bar{1}$
PENCENMOL	PENCEN05	I	$0.5 + 0.5$	$P\bar{1}$
PENCENMOL	PENCEN06	II	$0.5 + 0.5$	$P\bar{1}$

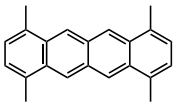
Table 7.1: The benchmark molecules and the CSD refcodes, polymorph forms, Z', and space groups of their associated experimental crystal structures.

Molecule	CSD Refcode	Polymorph	Z'	SG
PENCENMOL	PENCEN07	II	$0.5 + 0.5$	$P\bar{1}$
PENCENMOL	PENCEN08	II	$0.5 + 0.5$	$P\bar{1}$
PENCENMOL	PENCEN09	I	$0.5 + 0.5$	$P\bar{1}$
PENCENMOL	PENCEN10	III	$0.5 + 0.5$	$P\bar{1}$
PENCENMOL	PENCEN11	III	$0.5 + 0.5$	$P\bar{1}$
PENCENMOL	PENCEN12	III	$0.5 + 0.5$	$P\bar{1}$
PENTQUMOL	PENTQU	-	0.5	$P112_1/b$
PENTQUMOL	PENTQU01	-	0.5	$P\bar{1}$
PENTQUMOL	PENTQU02	-	0.5	$P2_1/c$
PERLENMOL	PERLEN01	alpha	1	$P2_1/a$
PERLENMOL	PERLEN03	alpha	1	$P2_1/a$
PERLENMOL	PERLEN04	alpha	1	$P2_1/c$
PERLENMOL	PERLEN05	alpha	1	$P2_1/c$
PERLENMOL	PERLEN06	beta	0.5	$P2_1/c$
PERLENMOL	PERLEN07	beta	0.5	$P2_1/c$
PERLENMOL	PERLEN08	alpha	1	$P2_1/c$
POVLUHMOL	POVLUH	-	1	$C2/c$
RUNLAPMOL	RUNLAP	-	1	$P2_1/c$
SANQIIMOL	SANQII	-	2	$P2_1$
SEVYEAMOL	SEVYEA	-	1	$P2_12_12_1$
UPIRUIMOL	UPIRUI	-	1	$Pbca$
UPIXAUMOL	UPIXAU	-	1	$P2_1$
UPIXEYMOL	UPIXEY	-	1	$Pca2_1$
VIBMAVMOL	VIBMAV	-	1	$P2_1/c$
VIQTATMOL	VIQTAT	-	1	$P2_12_12_1$
VUFKALMOL	VUFKAL	-	1	$P2_1$
VUFNIWMOL	VUFNIW	-	2	$P2_1/n$
WOTZUDMOL	WOTZUD	-	1	$P2_1/c$
WUPYIQMOL	WUPYIQ	-	0.5	$P2_1/n$
YOFROBMOL	YOFROB	-	0.5	$P\bar{1}$
YOFROBMOL	YOFROB01	-	0.5	$P\bar{1}$
ZETQUNMOL	ZETQUN	-	$0.5 + 0.5$	$P\bar{1}$
ZZZOYCMOL	ZZZOYC01	monoclinic	1	$P2_1$
ZZZOYCMOL	ZZZOYC02	monoclinic	1	$P2_1$
ZZZOYCMOL	ZZZOYC03	monoclinic	1	$P2_1$
ZZZOYCMOL	ZZZOYC04	monoclinic	1	$P2_1$

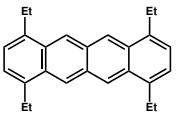
Table 7.1: *continued*



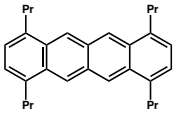
TETCEN



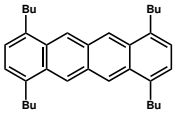
GUMZIY



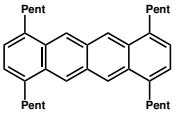
GUMZOE



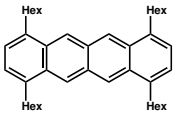
GUMZUK



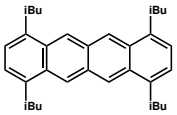
HIGNIV



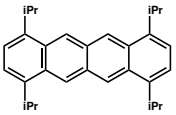
GUNBAT



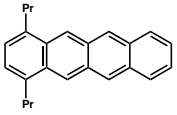
GUNBEX



DUWHEJ



DUWHOT



DUWPER

Figure 7.2: Molecular structures and CSD refcodes of the crystal structures of 10 tetracene derivatives used to benchmark carrier mobilities calculated with Marcus theory and TLT against FOB-SH references values [66].

7.1 Molecular Structures

In our computational screening workflow, we will need to convert molecules from MolBuilder which are outputted as InChi strings to a 3D structure so that it can be used in CSPy. For the CSP workflow to be successful it requires that the methods we use to generate conformers from InChi strings produce molecular structures which approximate the experimental molecular structures in the solid-state. In this section, we evaluate the conformer generation methods that are used in our computational screening workflow which takes an InChi string and generates a single conformer using RDKit which is followed up with a UFF optimisation and a B3LYP/6-311G** reoptimisation using GAUSSIAN09.

This will also test the assumption that for the molecules in our test set and the molecules we will use in our computational screening workflow, only a single conformer will be required. The full set of results of the calculated molecular geometries against the experimental are shown in Appendix E.1. Comparisons between the calculated molecular geometry of PERLENMOL and the experimental molecular structure obtained from PERLEN [293] were not included as the perylene molecules in PERLEN appeared to be quite distorted. A histogram of the single-molecule RMSDs in atomic positions averaged over the experimental molecular structures from each of the molecules associated crystal structures with hydrogen positions ignored is shown in Figure 7.3. In general, results show good agreement with the experimental with a test set average RMSD of 0.060 Å which is an average of the RMSDs averaged over all associated entries in the CSD for a given molecule with hydrogen positions ignored.

Molecules that resulted in RMSDs greater than 0.100 Å are shown in Table 7.2 and overlays of the three comparisons can be seen in Figures 7.4 and 7.5. Two molecules DBANQUMOL and OFATNQMOL have the same dione substructures which have both oxygens in close contact with another atom. For DBANQUMOL this is the nearby hydrogens due to its non-linear structure and for OFANTQMOL it is the nearby fluorines. The RDKit conformer generation with the UFF optimisation followed by DFT reoptimisation gave planar molecules which in both cases disagree with the experimental. Reoptimising the experimental molecular structures with DFT gave lower RMSDs and shows that for molecules

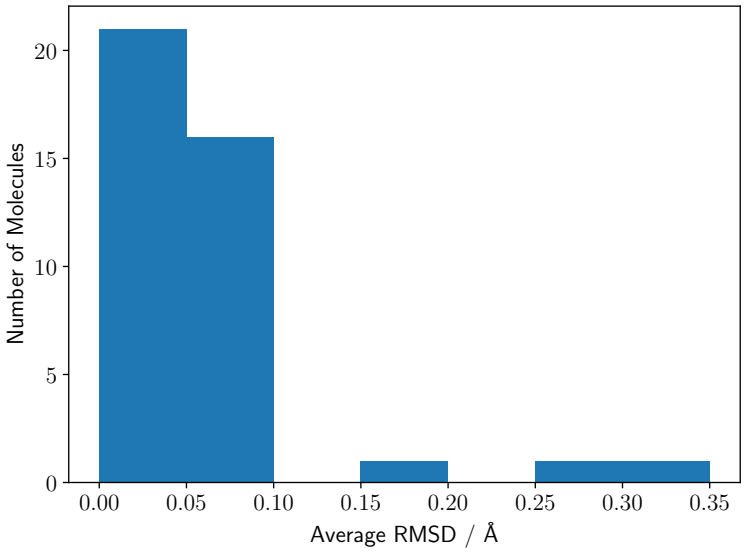


Figure 7.3: Histogram of the single-molecule RMSDs in atomic positions between the molecular structures obtained using the RDKit conformer generation with UFF and B3LYP/6-311G** optimisations against molecular structures from the experimental crystal structures averaged over the experimental molecular structures from each of the molecules associated crystal structures with hydrogen positions ignored.

Molecule	CSD Refcode	RMSD / Å
DBANQUMOL	DBANQU	0.285 (0.184)
OFANTQMOL	OFANTQ	0.325 (0.048)
OFANTQMOL	OFANTQ01	0.331 (0.048)
UPIXAUMOL	UPIXAU	0.197

Table 7.2: Three molecules with large RMSDs for the molecular structure obtained using the RDKit conformer generation with UFF and B3LYP/6-311G** optimisations (and RMSDs for the molecular structure obtained from a B3LYP/6-311G** optimisation using the experimental molecular structure as the initial input geometry) against molecular structures from the experimental crystal structures with hydrogen positions ignored.

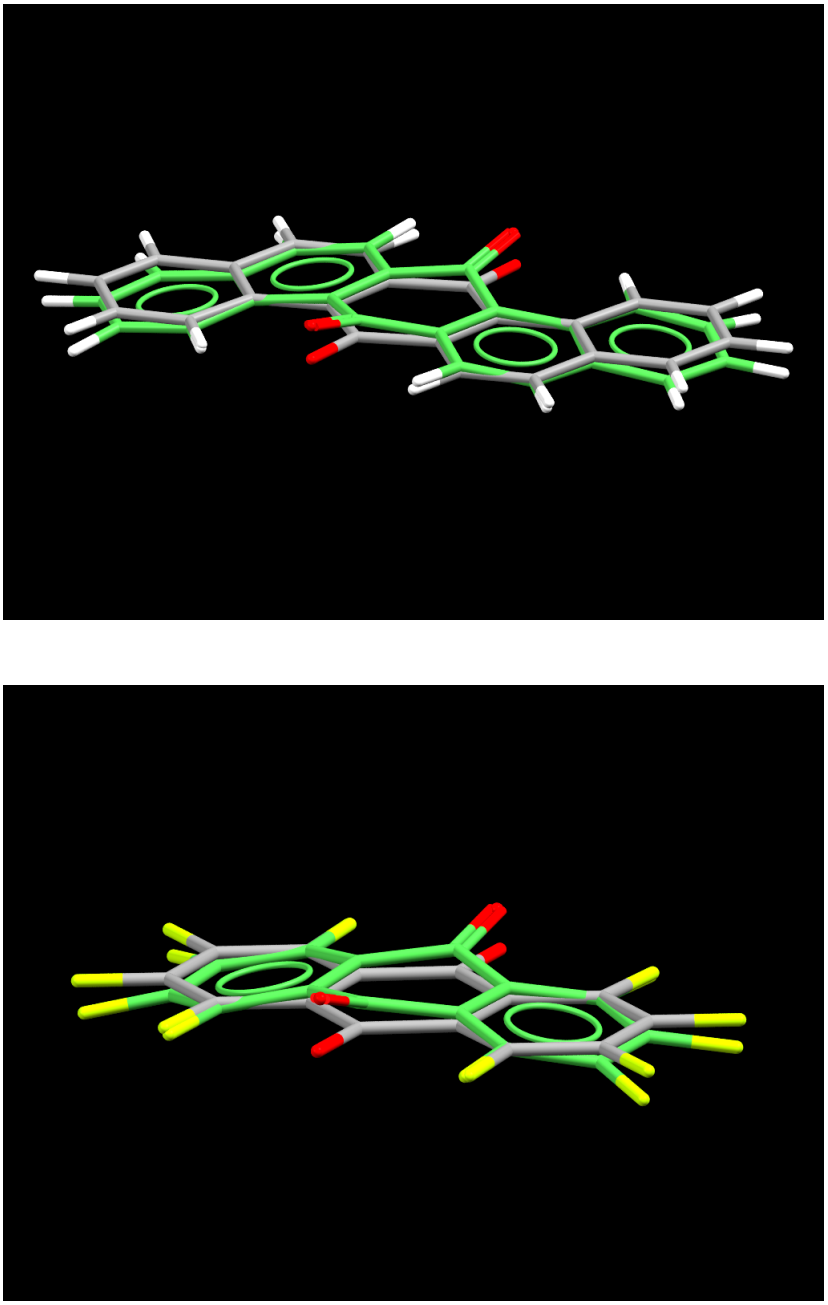


Figure 7.4: Overlays of the molecular structures obtained using the RDKit conformer generation with UFF and B3LYP/6-311G** optimisations with the molecular structures (green) taken from the experimental crystal structures for the molecules DBANQUMOL (top) and OFANTQMOL (bottom).

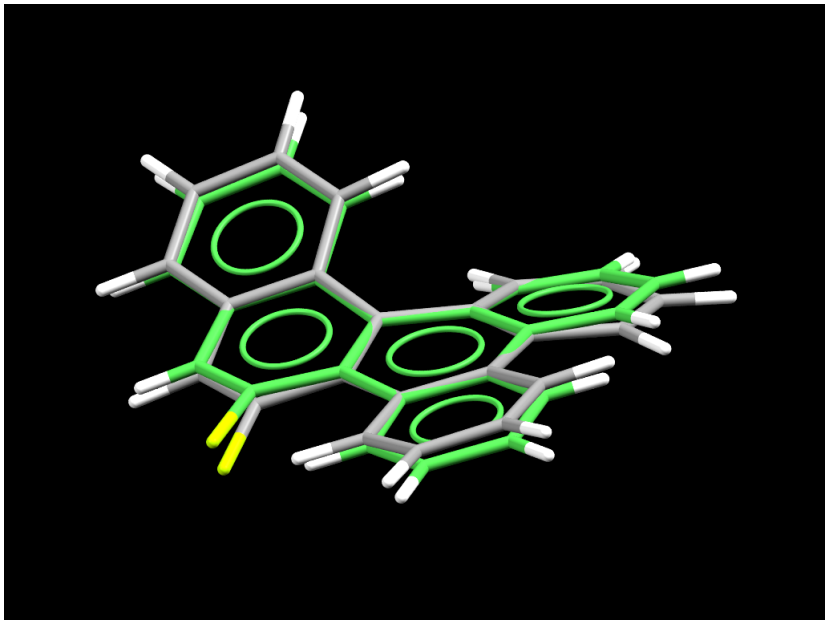


Figure 7.5: Overlays of the molecular structures obtained using the RDKit conformer generation with UFF and B3LYP/6-311G** optimisations with the molecular structures (green) taken from the experimental crystal structures for the molecule UPIXAUMOL.

with this dione substructure the single conformer generation will not be valid in general. This dione substructure should therefore be avoided or a multi-conformer generation stage would need to be included in the computational screening workflow.

Even after the geometry optimisations using the experimental molecular structures poor matches with the experimental were still obtained for the molecule DBANQUMOL. While large RMSDs were still obtained for UPIXAUMOL which showed no differences when experimental structures were used for the initial input geometry. These two examples show that there could be some dependence of the molecular structure on its crystal structure or that DFT may have some difficulties for these non-planar aromatic molecules. In either case improvements to the molecular structure would require more computationally expensive methods such as flexible CSP or electronic-structure methods that go beyond DFT such as MP2. These methods will not be explored here since flexible CSP workflows are still in active development and higher computational costs and greater difficulties

converging MP2 calculations make it more difficult to apply to high-throughput computational screening workflows.

In the CSP sampling benchmark sections, we take all optimised structures for rigid body CSP to benchmark the FIT force field and analyse different CSP sampling schemes. For the molecules DBANQUMOL and OFANTQMOL, we take the DFT optimised structures that used the experimental structure for its initial input geometry. This is done to isolate the errors due to the force fields from the conformer generation procedure when molecules with these dione substructures are used. Since we will not be including molecules with this dione substructure in the search spaces in our computational screening workflow we will therefore be unaffected by these errors.

7.2 CSP Sampling I

CSP searches were carried out using version II of the QR structure generation method with Z'=1 until a total of 5,000 successfully minimised structures are obtained for each of the 10 most common Z'=1 space groups ($P2_1/c$, $P2_12_12_1$, $P\bar{1}$, $P2_1$, $Pbca$, $C2/c$, $Pna2_1$, Cc, $Pca2_1$, $C2$) and Z'=2 until a total of 10,000 successfully minimised structures are obtained for each of the 5 most common Z'=2 space groups ($P\bar{1}$, $P2_1/c$, $P2_1$, $P2_12_12_1$, $P1$) for all 40 DFT optimised test set molecules. Final stage lattice energy minimisations were carried out using the FIT force field [124] with atom centred multipoles up to hexadecapoles obtained from GDMA with B3LYP/6-311G** GAUSSIAN09 electron densities. Ewald summations were carried out for charge-charge, charge-dipole, and dipole-dipole interactions, while all other higher-order electrostatics up to r^{-5} and repulsion-dispersion interactions were calculated for atom-atom distances within a 30 Å cut-off radius. Duplicated crystal structures were identified and removed by generating simulated X-ray diffraction patterns using PLATON [108] and calculating powder pattern similarities as described in Chapter 4. Full results of the CSPs and comparisons against experimental structures can be found in Appendix E.2. Comparisons with the experimental crystal structure PERLEN was not included as its perylene molecules appeared quite distorted.

7.2.1 Performance

Overall results are good with a total of 45 experimental polymorphs found in our search out of a total of 46, only the experimental structure MIVSES was unable to be found. Poor matches were obtained with DUXSAS, PENTQU01, and VUFNIW as they did not result in 30 out of 30 COMPACK matches with distance and angle tolerances of 30% and 30° or smaller and hydrogen positions ignored. The best results found was a match of 29 out of 30 for DUXSAS, 28 out of 30 for PENTQU01, and 28 out of 30 for VUFNIW. A total of six matches were found with RMSDs greater than 0.6 Å for the 30 out of 30 match. Matches with NIJCEQ and UPIXAU had particularly large RMSDs with values greater than 1.0 Å.

A histogram of the COMPACK RMSDs averaged over all entries in the CSD for a given polymorph with hydrogen positions ignored for a 30 out of 30 and 60 out of 60 molecule cluster is shown in Figure 7.6. For all results that were able to obtain complete COMPACK matches a test set average RMSDs of 0.370 and 0.459 Å was obtained which is an average of the RMSDs averaged over all associated entries in the CSD for a given polymorph with hydrogen positions ignored for cluster sizes of 30 and 60 molecules. COMPACK overlays for comparisons between the predicted and experimental crystal structures of BEZLUO, DUXSAS, MORROD, NIJCEQ, PENCEN, PENCEN01, PENCEN10, PENTQU01, UPIXAU, and VUFNIW are shown in Appendix E.2.

We use the CSD Python API [61] to calculate X-ray powder pattern similarities for all matches found with COMPACK. The lowest similarity measure calculated was 0.901 and the test set average of 0.966 was obtained which is an average of the X-ray similarities averaged over all associated entries in the CSD for a given polymorph and shows good agreements between our predicted structures and the experimental. Despite being unable to obtain a 30 out of 30 match in COMPACK, the calculated X-ray powder pattern similarities for DUXSAS, PENTQU01, and VUFNIW were relatively high with values of 0.985, 0.913, and 0.944 respectively. Similarly, NIJCEQ and UPIXAU were also shown to have relatively high X-ray powder pattern similarities of 0.951 and 0.910 despite having larger calculated RMSDs with COMPACK. Of all 45 matches a

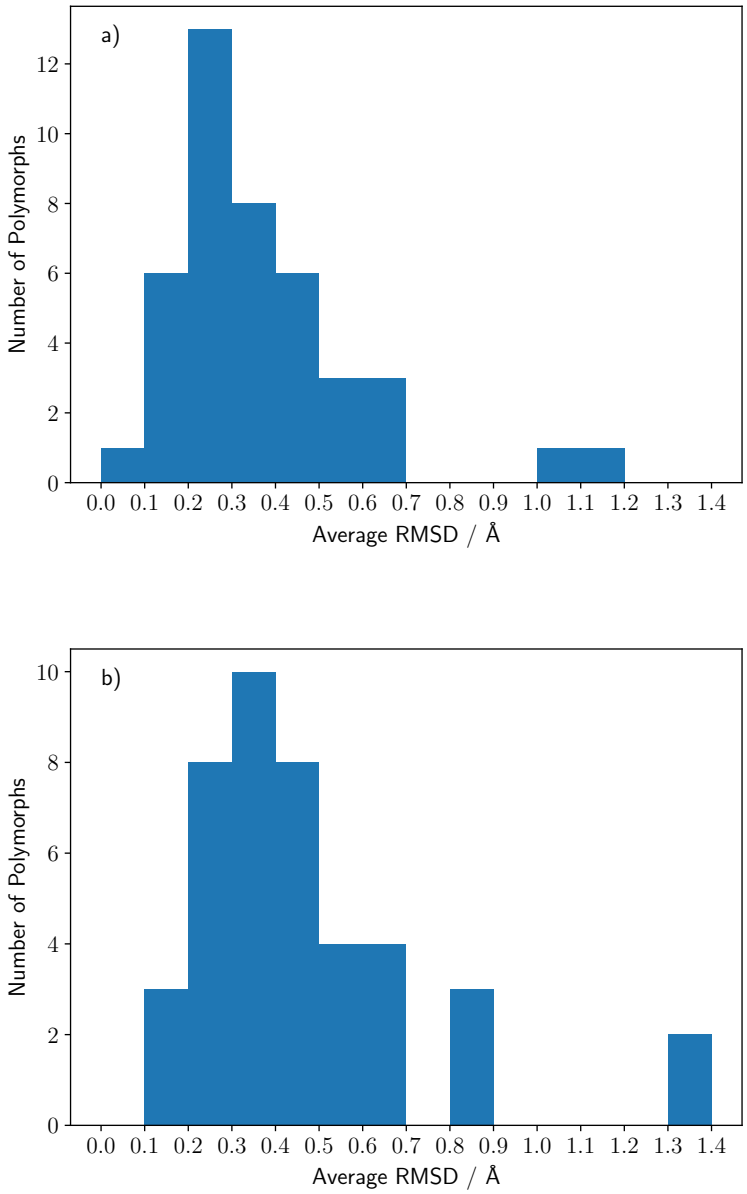


Figure 7.6: Histograms of the COMPACK RMSDs for the matched predicted crystal structure against the experimental averaged over all entries in the CSD for a given polymorph with distance and angles tolerances of 30% and 30° or smaller and hydrogen positions ignored for a) a 30 out of 30 match and b) a 60 out of 60 match.

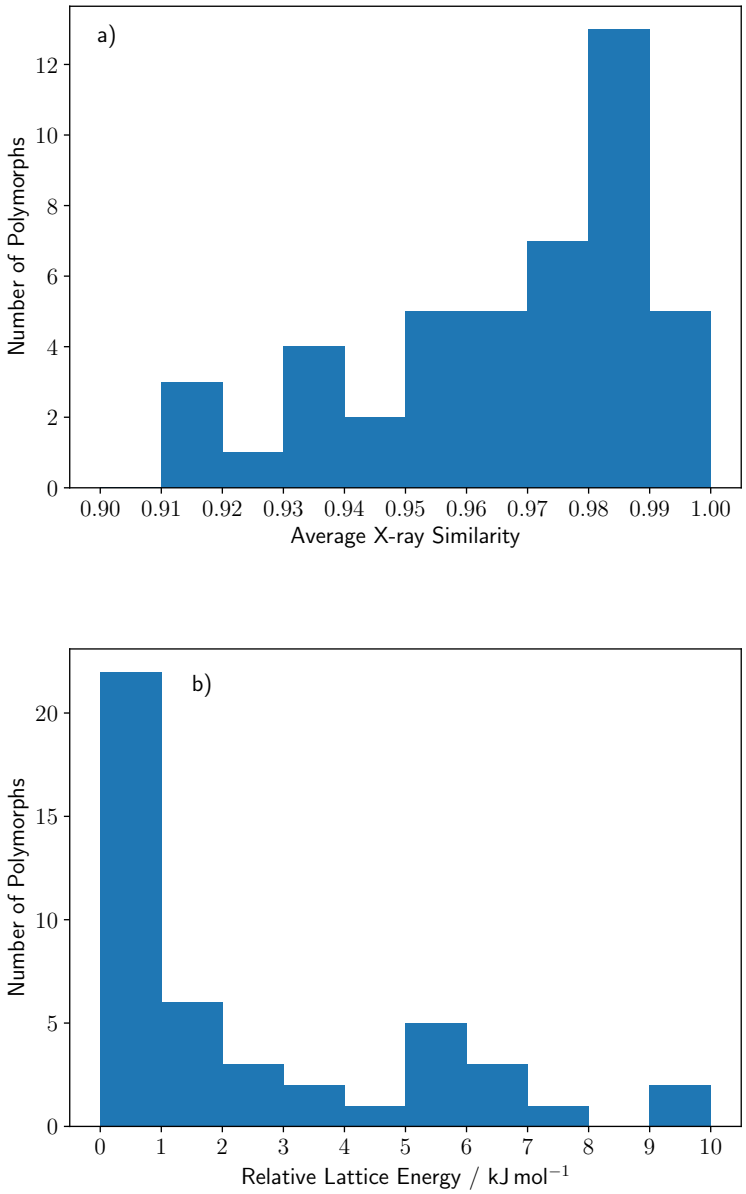


Figure 7.7: Histograms of the a) X-ray similarities between the matched predicted crystal structure and the experimental averaged over all entries in the CSD for a given polymorph and b) relative lattice energy of the matched predicted crystal structure.

total of 17 were found to be the global minimum of the CSP landscape and the average lattice energy relative to the global minimum was found to be around 2.30 kJ mol^{-1}. A histogram of the X-ray powder pattern similarities averaged over all entries in the CSD for a given polymorph and the relative lattice energies are shown in Figure 7.7. Simulated X-ray powder patterns of the predicted and experimental crystal structures of BEZLUO, DUXSAS, MORROD, NIJCEQ, PENCEN, PENCEN01, PENCEN10, PENTQU01, UPIXAU, and VUFNIW are shown in Appendix E.2.

The FIT force field was shown to produce good matches to experimental structures and the sampling scheme used was able to find all experimental polymorphs in our test set, excluding MIVSES. Given these results, the CSP procedure described in this section should act as an excellent first CSP stage in a computational screening workflow for small molecule organic semiconductors. However, the computational expense of the chosen number of space groups and successfully minimised structures described above for each molecule may cause some difficulties when scaling up the CSP stage to include over 100 molecules, at the very least we will want to reduce the computational cost as much as possible. On average the Z'=1 searches had taken around 8 h while for the Z'=2 searches around 17 h when 4 nodes were used on the Iridis 5 compute cluster with dual Intel® Xeon® Gold 6138 CPUs @ 2.00GHz on each node.

7.2.2 MIVSES

MIVSES [207, 208] was found to be particularly problematic with no matches to the experimental structure found. We carry out another test on our force field by running a minimisation on the experimental structure using the DFT optimised isolated molecules geometries pasted into the experimental crystal structure. We use this crystal structure with DFT optimised molecular geometries but the experimental molecular packing so that we are consistent with our CSP procedure. Since our isolated molecule DFT geometries were similar to the experimental molecule geometries the pasted and experimental structures are also similar. Running a COMPACK comparison using for a cluster of 60 molecules with hydrogen positions ignored gave a 60 out of 60 match with an RMSD of 0.062 Å.

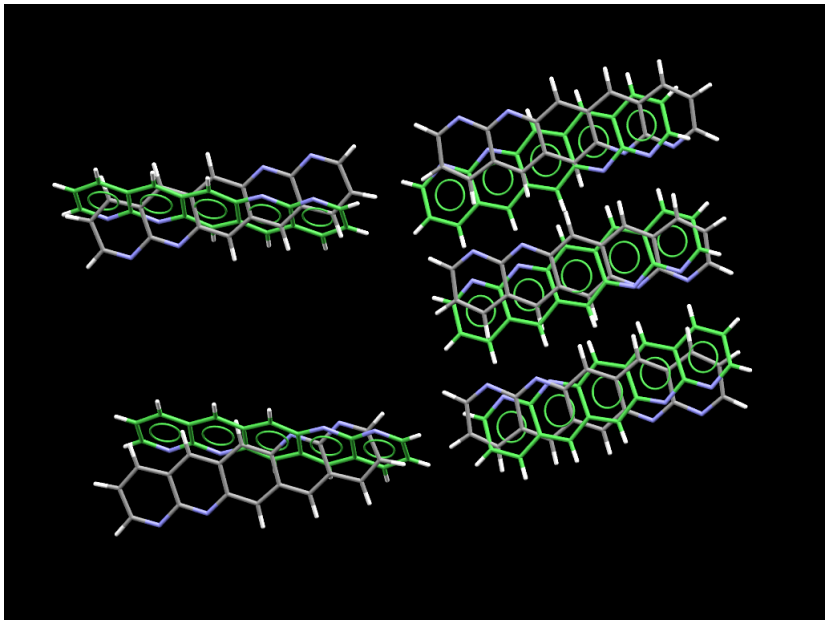


Figure 7.8: The 5 out of 30 COMPACK overlay of the experimental crystal structure reoptimised with DMACRYS using the FIT force field and the original experimental crystal structure (green).

We energy minimise the pasted crystal structure with DMACRYS using the FIT force field and distributed multipoles and the same methods as the final optimisation stage of the CSP described in this section but convert the structure to $P1$ before the minimisation to remove symmetry constraints. Using the minimised structure we carry out COMPACK comparisons with the original experimental structure for a cluster of 30 molecules with distance and angle tolerances of 20% and 20° and hydrogen positions ignored. Results were extremely poor, COMPACK matched only 5 out of 30 molecules showing the DMACRYS had made large moves away from the experimental geometry, Figure 7.8 shows the COMPACK overlay.

The results from the DMACRYS optimisation of the experimental crystal structure led to something quite different from the experimental. The absence of the experimental crystal structure in the CSP suggests that either: the force field itself had some difficulty with this molecule or that the experimental structure had been determined incorrectly. To check whether the failure was due to our

force field we also performed periodic DFT calculations on the experimental structure. If the periodic DFT calculations contained an energy minimum that matched with the experimental it would be clear that higher accuracy methods would have been necessary for these types of molecules.

We ran an energy minimisation with VASP [111–114] using the projector augmented wave method [294, 295], the PBE functional [82, 296] with Grimme's D3 dispersion correction and Becke-Johnson damping [29,30], a planewave energy cut off of 500 eV, and a maximum k-point spacing of 0.05 $Å^{-1}$. We then use COMPACK to compare the experimental structure of MIVSES and the VASP optimised structure for a 30 and 60 molecule clusters with distance and angle tolerances of 30% and 30° and hydrogen positions ignored. For both 30 and 60 molecule clusters, a full match was obtained with RMSDs of 1.117 and 1.467 Å. We check the COMPACK results by calculating X-ray powder pattern similarities using the CSD Python API which resulted in a similarity of 0.744. The powder pattern similarity is particularly poor, much lower than all other comparisons for our test set which had similarities of 0.9 or higher. The poor COMPACK RMSDs and powder pattern similarities suggest overall the VASP optimised structure is not a match. COMPACK overlays for comparisons between the VASP optimised crystal structure and experimental crystal structure of MIVSES are shown in Figure 7.9 and the simulated X-ray powder patterns of the two structures are shown in Figure 7.10.

The disagreement between the computational and experimental results appear to be due to the particularly short distance of 3.005 Å between the nitrogen atoms in the experimental crystal structure. In both DMACRYS and VASP optimisations the N$\cdots$N distances increased as expected since from electrostatic arguments we would expect the nitrogen atoms to be $\delta-$ so that the interaction would be disfavoured. In Figure 7.11 we can see how the VASP optimisation had distorted the MIVSES molecule away from the experimental molecular structure which appears to have occurred so that the N$\cdots$N distances could be increased. In Figure 7.12 we show a histogram generated using IsoStar [298] of the experimental distances in van der Waals corrected distances between aromatic N atoms in 6-rings and aromatic or sp^2 N atoms. A distance of zero in van der Waals corrected distances for two N atoms will be a distance of 3.10 Å.

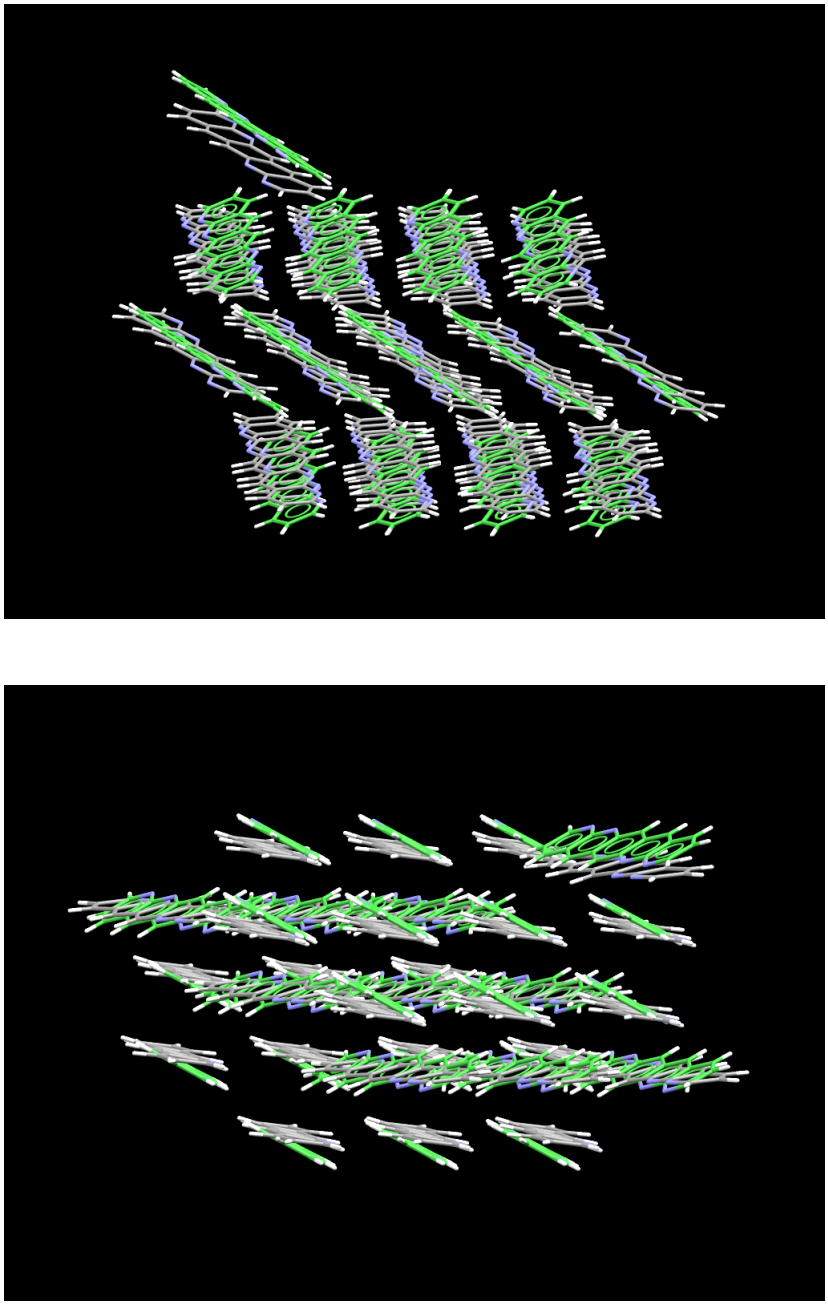
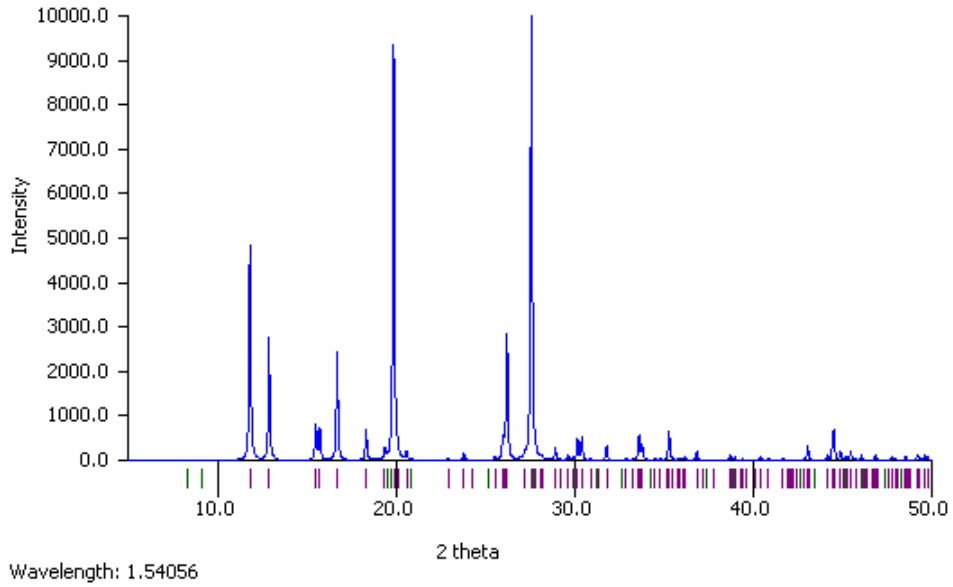


Figure 7.9: Two views of the 30 out of 30 COMPACK overlay with an RMSD of 1.117 Å of MIVSES for the experimental crystal structure reoptimised with VASP and the original experimental crystal structure (green).



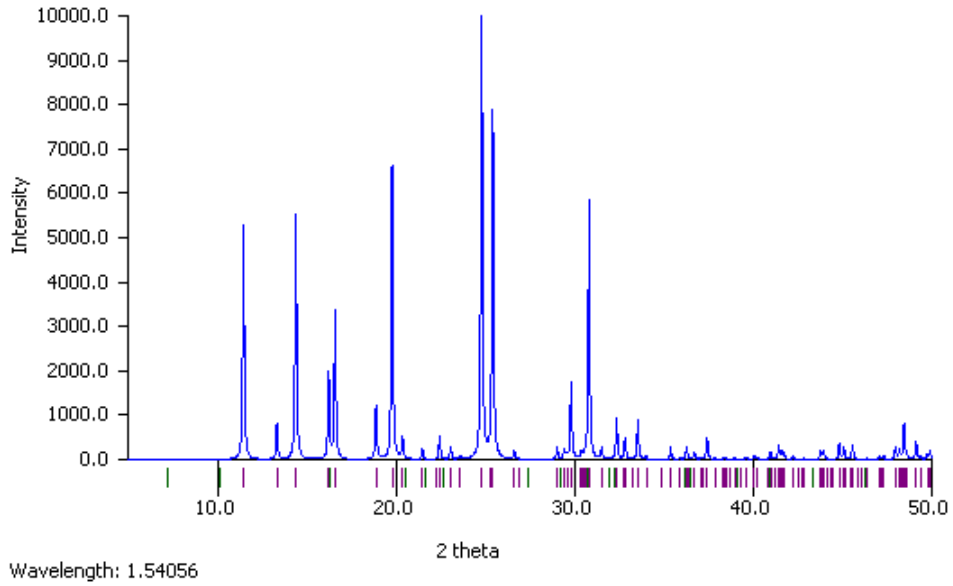


Figure 7.10: Simulated PXRD patterns of MIVSES generated in Mercury [297] for the experimental structure optimised using VASP (top) and the original experimental structure (bottom). X-ray powder patterns similarities between these two structures using the CSD Python API resulted in a similarity of 0.744.

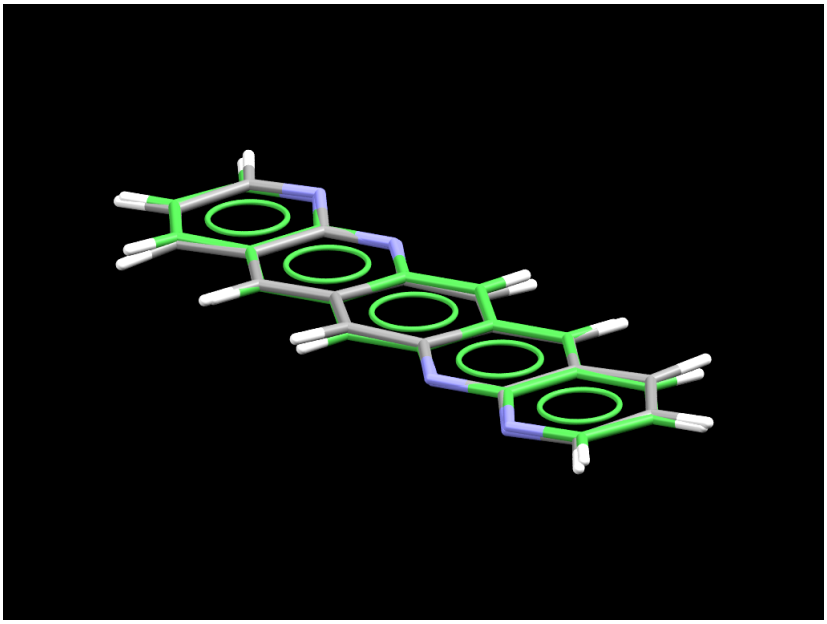


Figure 7.11: Overlays of the molecular geometries of MIVSESMOL obtained from the experimental crystal structure reoptimised with VASP using the PBE functional with Grimme's D3 dispersion correction and Becke-Johnson damping and the molecular geometries (green) from the original experimental crystal structure.

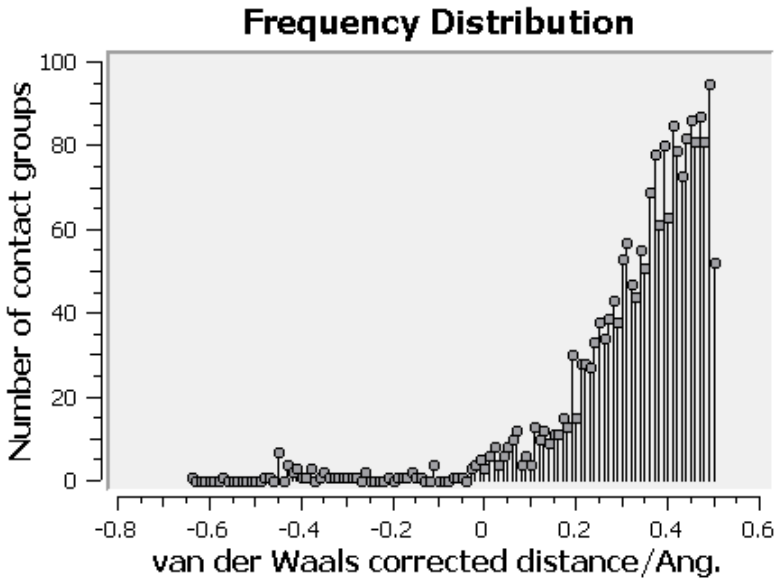


Figure 7.12: Frequency distribution of 2,041 experimental distances in van der Waals corrected distances between aromatic N atoms in 6-rings and aromatic or sp^2 N atoms from IsoStar.

From this graph, we can see that this -0.095 Å corrected distance would be particularly rare. Out of 2,041 contact groups from this IsoStar data set, there are only 49 intermolecular N$\cdots$N distances shorter than the MIVSES example. Taking a look at the crystal structure with N$\cdots$N distances around 3.005 Å we can see that there are usually other interactions that force the N atoms to this short distance. For example, in the crystal structure of AHOHAI [299, 300] there is an intermolecular N$\cdots$N distance of 2.987 Å which appears to be due to a number of N$\cdots$H interactions which has forced the N atoms to come into close contact, see Figure 7.13.

Unlike the AHOHAI example, the MIVSES crystal structure does not appear to have any interactions that would force the nitrogen atoms to come into close contact. What is even more unusual is that there are four pairs of small N$\cdots$N distances per molecule, see Figure 7.14. The disagreement with two different computational methods, arguments based on the electrostatics of the N atom, and past experimental crystal structure data, suggests that the MIVSES crystal structure would be unlikely and further work to re-evaluate the experimental

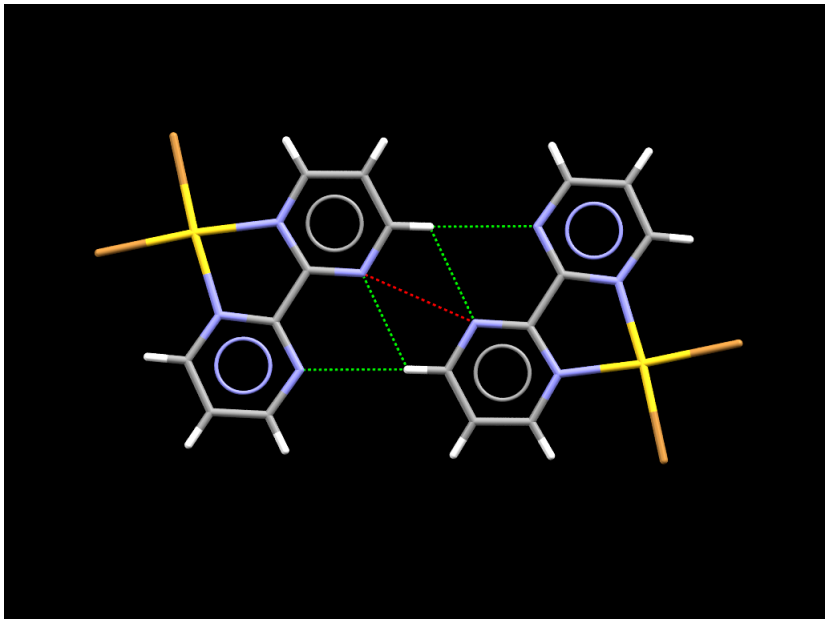


Figure 7.13: Molecular dimer from the crystal structure AHOHAI with a short intermolecular N···N distance (red) of 2.987 Å but a number of favourable N···H interactions (green).

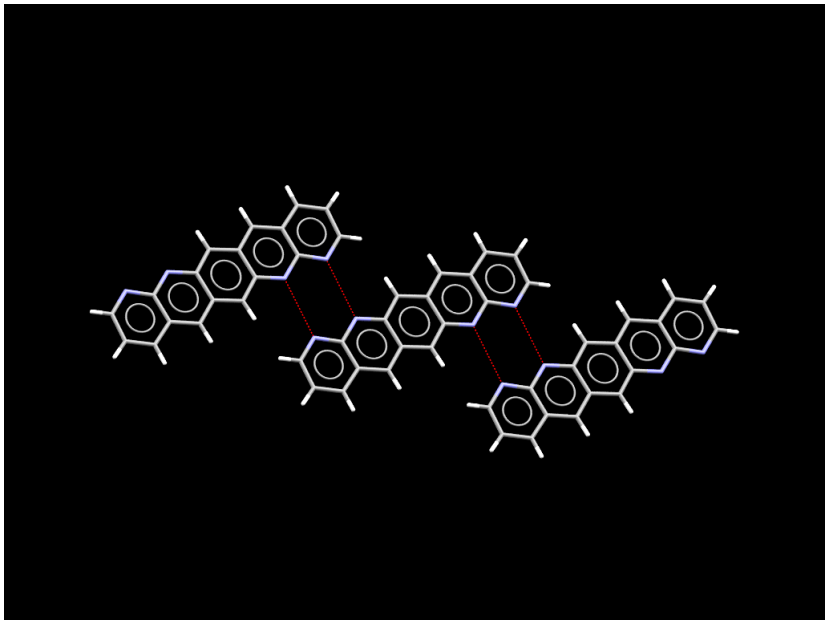


Figure 7.14: Molecules from the MIVSES crystal structure showing a number of short intermolecular N···N distances (red) occurring for each pair of molecules.

structure of MIVSES might need to be carried out. The inability to obtain MIVSES with the FIT force field and sampling scheme should therefore not be of any concern unless future experimental work were to result in the same crystal structure.

7.2.3 Unique Structures

In most cases, we want to carry out CSP in a way to obtain as many minima as possible so that we do not risk missing out on structures that might be discovered in experimental stages due to the sampling scheme that was used. One obvious method that could be used would be to carry out CSP until new unique structures below a given relative lattice energy are unlikely to be found. In Figure 7.15 we show the number of unique structures found (within $12\,\mathrm{kJ\,mol^{-1}}$ from the global lattice energy minimum of the completed search) as a function of the number of successful minimisation of the CSP process for the most common Z'=1 and Z'=2 space groups. For both examples, we see that the numbers of unique structures increase logarithmically with the number of successful minimisations. So that an increase in the unique number of structures is initially large but decreases as the CSP progresses.

Figure 7.15 also shows that for a small proportion of molecules the CSP for that Z' | SG search could be regarded as complete for the $12\,\mathrm{kJ\,mol^{-1}}$ energy window as the number of unique structures becomes a constant with increasing successful minimisations. For these molecules, we could use a smaller number of successful minimisations and obtain similar results. However, for a large proportion of molecules in the more difficult searches such as Z'=1 | $P2_1/c$ the sampling has not been sufficient and more predicted crystal structures could have been found by increasing the number of successfully minimised structures required. It appears that by analysing the number of unique structures as a function of the number of successful minimisations in most cases the CSP searches will need to be expanded in contrast to our aim to devise a reduced sampling scheme.

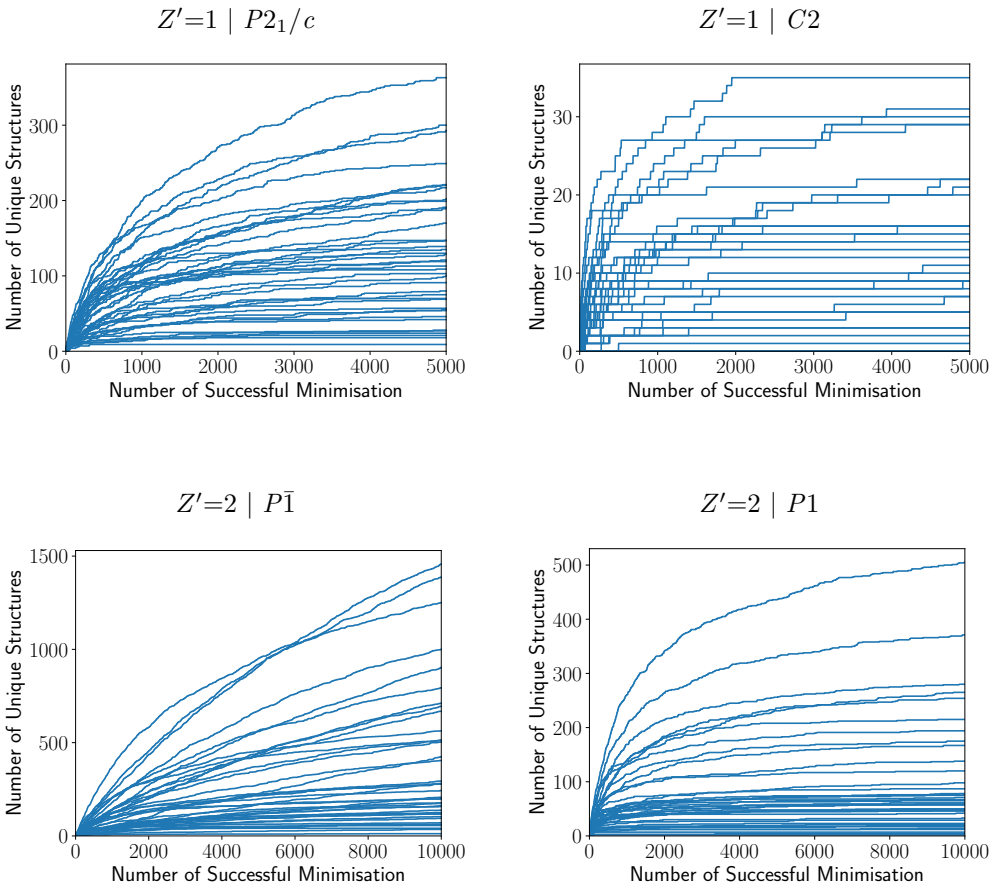


Figure 7.15: The number of unique structures found within 12 kJ mol^{-1} from the global lattice energy minimum of the full search as a function of the number of successful minimisations for the most and least common Z'=1 and Z'=2 space groups used. Each line plot shows the progress of the CSP for a given molecule from the test set.

7.2.4 Reduced Sampling

Although Figure 7.15 suggests that CSP searches need to be expanded, all experimental structures were obtained (ignoring MIVSES). We also know that some oversampling may have occurred since for example the Z'=1 | $P2_1/c$ and Z'=2 | $P\bar{1}$ searches can lead to the same structures. Therefore if we make the assumption that for the aromatic molecules we are interested in that almost all experimental and global minimum structures are found early in the CSP search, our CSP searches do not need to be complete and we can reduce the number of successful minimisations required. This reduced search may be sufficient as the only or first CSP stage in a computational screening procedure.

There are a few possibilities why we might expect to find experimental or global minimum structures early in our CSP searches. Note that the CSP searching methods we use are not completely random and are targeted in some sense. For example, we focus the search in specific space groups that are the most commonly found experimentally and are the space groups that produce the most closely packed structures which typically result in lower lattice energies. Additionally, the minimisations of the crystal structures are similarly not a random process. If for example, experimental structures were to inhabit wide and deep energy wells the minimisation procedure used will likely lead to these energy minima even if other minima exist that are closer to its initial QR structure.

To form a reduced sampling scheme we look to find the minimum number of space groups and number of successfully minimised structures required to obtain all experimental polymorphs in our test set and the lowest energy structures from our original CSPs. Tables E.4 and E.5 show the number of times the global minimum and experimental structures were found in a given Z' | SG search. Only one entry from the CSD per polymorph is used in the reference set so for example only one experimental structure for the polymorph II form of the crystal structure of pentacene (PENCEN01) is used. Tables E.6 and E.7 show the minimum number of successful minimisation required to obtain the global minimum and experimental structures for each Z' | SG search. From these tables we can see that there are a number of structures that can only be obtained in a

single Z' | SG search, a summary is shown in Table 7.3.

The results in Table 7.3 forces us to choose specific space groups and the number of minimisations in our reduced sampling scheme. It is possible to devise a minimum set of space groups that obtains all global minima and experimental structures with only Z'=1 | $P2_1/c$, $Pbca$, $Pna2_1$, $Pca2_1$ and Z'=2 | $P\bar{1}$, $P2_1$, $P1$ searches which are just the space groups required to satisfy the molecules in Table 7.3. The choice of the Z' | SG searches are due to the complete or partial coverage of one Z' | SG search with another. Of course, it could have been down to the set of molecules and the sampling scheme we had chosen. The most reasonable thing to do is therefore to include all space groups which are more commonly found experimentally than the least common space group in the minimum set for both Z' searches. This leaves us with the Z'=1 | $P2_1/c$, $P2_12_12_1$, $P\bar{1}$, $P2_1$, $Pbca$, $C2/c$, $Pna2_1$, Cc, $Pca2_1$ and Z'=2 | $P\bar{1}$, $P2_1/c$, $P2_1$, $P2_12_12_1$, $P1$ searches which are similar to what we started with except for the removal of the Z'=1 | $C2$ search.

We can get an idea of what the minimum number of successful minimisations could be by assigning each match to the space groups with the smallest minimisations required to obtain the reference of that match, choosing Z'=1 over Z'=2 and the most common space group if the same number of minimisation were required. For example, we assign the match between the CSP of AROBUFMOL and the reference AROBUF to Z'=1 | $P2_1$ since only 331 successful minimisations were required to obtain the reference in the Z'=1 | $P2_1$ search. We then assign each Z' | SG searches the maximum of all the matches that were assigned to that Z' | SG search. So for example in Z'=1 | $P2_1/c$ this would be 2,287 since the match of the CSP of BNPYREMOL with its GM was assigned to Z'=1 | $P2_1/c$ and was the largest out of all the assignments for this Z'=1 | $P2_1/c$ search.

In Table 7.4 we show the maximum number of minimisations of the set of matches assigned to a given Z' | SG search as described above. For the Z'=1 searches, Table 7.4 suggests a larger number of minimisations of around 5,000 will be required for $P2_1/c$ while the other space groups in the Z'=1 searches a smaller number around 1,000 would be required. We have used a greater number of successful minimisations than the assigned value in Table 7.4 to ensure transferability to molecules outside of the test set. The imbalance between

CSP	Reference	Z'	SG
BNZACRMOL	BNZACR	1	$Pna2_1$
CEQGELMOL	CEQGEL	2	$P\bar{1}$
DBANQUMOL	DBANQU	1	$Pca2_1$
FANTRQMOL	FANTRQ10	2	$P2_1$
HAMDUWMOL	HAMDUW	1	$P2_1/c$
MIVRUEMOL	GM	1	$P2_1/c$
MIVSESMOL	GM	2	$P1$
PENTQUMOL	PENTQU01	2	$P\bar{1}$
SANQIIMOL	GM	2	$P\bar{1}$
SANQIIMOL	SANQII	2	$P2_1$
UPIRUIMOL	UPIRUI	1	$Pbca$
UPIXAUMOL	GM	1	$P2_1/c$
UPIXEYMOL	GM	1	$P2_1/c$
UPIXEYMOL	UPIXEY	1	$Pca2_1$

Table 7.3: Experimental crystal structures or global lattice energy minimum crystal structures that could only be obtained in a single Z' | SG search.

Z'	SG	CSP	Reference	Minimisations
1	$P2_1/c$	BNPYREMOL	GM	2287
1	$P2_12_12_1$	PAGMERMOL	GM	143
1	$P\bar{1}$	YOFROBMOL	YOFROB	3
1	$P2_1$	AROBUFMOL	GM	331
1	$Pbca$	UPIRUIMOL	UPIRUI	243
1	$C2/c$	POVLUHMOL	POVLUH	17
1	$Pna2_1$	BNZACRMOL	BNZACR	529
1	Cc	-	-	-
1	$Pca2_1$	UPIXEYMOL	UPIXEY	87
2	$P\bar{1}$	PENTQUMOL	PENTQU01	9351
2	$P2_1/c$	RUNLAPMOL	RUNLAP	132
2	$P2_1$	FANTRQMOL	FANTRQ	1765
2	$P2_12_12_1$	-	-	-
2	$P1$	MIVSESMOL	GM	5409

Table 7.4: The maximum minimisations of the set of matches assigned to a given Z' | SG search as described in Subsection 7.2.4, no matches were assigned to the Z'=1 | Cc and Z'=2 | $P2_12_12_1$ searches.

Z'=1 | $P2_1/c$ and other Z'=1 searches are likely down to the larger number of lower energy minima found in Z'=1 | $P2_1/c$. Compare for example the number of unique structures in the low energy region between Z'=1 | $P2_1/c$ and Z'=1 | $C2$ in Figure 7.15.

For the Z'=2 searches, Table 7.4 suggests a greater number of minimisation will be required as expected given the greater dimensionality for these searches. In the Z'=2 | $P\bar{1}$, $P2_1$, $P1$ searches the assigned matches with the maximum minimisations were also the only structures that could be found in that search. A larger 12,500 might be necessary for Z'=2 | $P\bar{1}$ since the PENTQU01 was found at a particularly late stage of the CSP search. A smaller number of minimisations of around 7,500 should be adequate to cover the other Z'=2 searches. As with the Z'=1 case we have used a greater number of successful minimisations than the assigned value in Table 7.4 to ensure transferability to molecules outside of the test set.

From our analysis, we form a reduced sampling scheme with searches in Z'=1 | $P2_1/c$ with 5,000, Z'=1 | $P2_12_12_1$, $P\bar{1}$, $P2_1$, $Pbca$, $C2/c$, $Pna2_1$, Cc, $Pca2_1$ with 1,000, Z'=2 | $P\bar{1}$ with 12,500, and Z'=2 | $P2_1/c$, $P2_1$, $P2_12_12_1$, $P1$ with 7,500 successfully minimised structures. This reduced search should produce the same global minimum structures and find all experimental structures in our test set and should be applicable to other molecules that are similar to the ones in the test set. By using timings from the CSP of BNZACRMOL we estimate that this reduced search would take around 2 h for the Z'=1 and 14 h for the Z'=2 searches for a 4 node job on the Iridis 5 compute cluster with dual Intel® Xeon® Gold 6138 CPUs @ 2.00GHz on each node.

7.3 CSP Sampling II

From Section 7.2 the Z'=2 searches were found to be the most difficult requiring the most computational time. One method to make further reductions to the Z'=2 searches could be to remove space group symmetry constraints in the final stage minimisation which could allow the Z'=2 structures to be found via the more efficient Z'=1 searches. We, therefore, attempt to take advantage of the work by Gavezzotti who had found that 83% of experimental Z'=2 crystal

structures showed some form of pseudosymmetry [301]. In this section, we carry out CSPs for all 40 DFT optimised test set molecules using the same methods as those described in Section 7.2 except in the final stage of the lattice energy minimisation we remove all symmetry constraints by converting all structures to $P1$ before the final stage of minimisation.

7.3.1 Performance

The full results of the CSPs and comparisons against experimental structures can be found in Appendix E.3. Comparisons of the experimental structures with the predicted structures obtained in Section 7.2 and those in this section were almost the same with at most only slight differences in the calculated RMSDs. There was however one difference in the CSP of PENCENMOL which produced a match with the experimental structures of PENCEN10, PENCEN11, and PENCEN12 which gave slightly smaller average RMSDs of 0.268 and 0.324 Å vs. 0.362 and 0.460 Å for a cluster of 30 and 60 molecules and slightly lower relative energy of 3.583 vs. 3.613 kJ mol^{-1}. An increase in the computational cost of the CSP due to the removal of the symmetry constraints in the final optimisation stage was seen. On average the Z'=1 searches had taken around 10 h while for Z'=2 around 17 h when 4 nodes were used on the Iridis 5 compute cluster with dual Intel® Xeon® Gold 6138 CPUs @ 2.00GHz on each node. The Z'=1 search cost had increased by around 25% while only small increases were seen for Z'=2.

7.3.2 Reduced Sampling

Similarly to Section 7.2 we look to find a reduced sampling scheme that can obtain all experimental structures and global minima at the lowest computational cost. Table 7.5 shows the number of matches that were only found in a given Z' | SG search. By comparing with Table 7.3 we can see that there are fewer structures that are only found in one Z' | SG search showing a greater overlap between searches in this sampling scheme. We can choose the minimum number of Z' | SG searches required to satisfy the structures in Table 7.5 but like before choose to retain all searches more common than our least common search. This leaves us with Z'=1 | $P2_1/c$, $P2_12_12_1$, $P\bar{1}$, $P2_1$, $Pbca$, $C2/c$, $Pna2_1$, Cc, $Pca2_1$ and

CSP	Reference	Z'	SG
BNZACRMOL	BNZACR	1	$Pna2_1$
CEQGELMOL	CEQGEL	2	$P\bar{1}$
DBANQUMOL	DBANQU	1	$Pca2_1$
HAMDUWMOL	HAMDUW	1	$P2_1/c$
MIVRUEMOL	GM	1	$P2_1/c$
PENTQUMOL	PENTQU01	2	$P\bar{1}$
SANQIIMOL	GM	2	$P\bar{1}$
SANQIIMOL	SANQII	2	$P2_1$
UPIRUIMOL	UPIRUI	1	$Pbca$
UPIXAUMOL	GM	1	$P2_1/c$
UPIXEYMOL	GM	1	$P2_1/c$

Table 7.5: Experimental crystal structures or global lattice energy minimum crystal structures that could only be obtained in a single Z' | SG search.

Z'	SG	CSP	Reference	Minimisations
1	$P2_1/c$	BNPYREMOL	GM	2286
1	$P2_12_12_1$	PAGMERMOL	GM	143
1	$P\bar{1}$	YOFROBMOL	YOFROB	3
1	$P2_1$	AROBUFMOL	GM	331
1	$Pbca$	UPIRUIMOL	UPIRUI	237
1	$C2/c$	POVLUHMOL	POVLUH	13
1	$Pna2_1$,	BNZACRMOL	BNZACR	522
1	Cc	-	-	-
1	$Pca2_1$	UPIXEYMOL	UPIXEY	87
2	$P\bar{1}$	PENTQUMOL	PENTQU01	9349
2	$P2_1/c$	RUNLAPMOL	RUNLAP	129
2	$P2_1$	PENCENMOL	PENCEN10	3036

Table 7.6: The maximum minimisations of the set of matches assigned to a given Z' | SG search as described in Section 7.3, no matches were assigned to the Z'=1 | Cc search.

Z'=2 | $P\bar{1}$, $P2_1/c$, $P2_1$ searches which is a further removal of the Z'=2 | $P2_12_12_1$ and Z'=2 | $P1$ searches over Section 7.2.

We carry out a similar method to determine the maximum number of minimisations required for each Z' | SG search as described in Section 7.2. Results are shown in Table 7.6 and are similar to the results in Section 7.2. The main differences are the absence of the Z'=2 | $P2_12_12_1$ and Z'=2 | $P1$ searches due to the greater coverage of these searches by the others. Additionally, a change in the assignment of the Z'=2 | $P2_1$ with the polymorph PENCEN10 is made as a better match was found in this section. Overall a similar sampling scheme to the one devised in Section 7.2 can be made. This results in the reduced sampling scheme with searches in Z'=1 | $P2_1/c$ with 5,000, Z'=1 | $P2_12_12_1$, $P\bar{1}$, $P2_1$, $Pbca$, $C2/c$, $Pna2_1$, Cc, $Pca2_1$ with 1,000, Z'=2 | $P\bar{1}$ with 12,500, and Z'=2 | $P2_1/c$, $P2_1$ with 7,500 successfully minimised structures with symmetry constraints removed in the final minimisation. By using timings from the CSP of BNZACRMOL we estimate that this reduced search would take around 3 h for the Z'=1 and 10 h for the Z'=2 searches for a 4 node job on the Iridis 5 compute cluster with dual Intel® Xeon® Gold 6138 CPUs @ 2.00GHz on each node.

7.3.3 Comparisons

Given the shorter timings for the reduced sampling developed in this section, it will therefore be the preferred method over the one described in Section 7.2. We ran CSPs with pentacene on the three different sampling methods to compare their performances and are referred to as Scheme A, Scheme B, and Scheme C. In Scheme A we used the methods and the reduced sampling scheme described in this section. The other two methods used the CSP method described in Section 7.2 which is the same as the one used for Scheme A but the symmetries of the QR crystal structures are maintained throughout the minimisation. Scheme B used the sampling scheme from Yang *et al.* in Ref. [58] which was used for the computational screening of planar pyrrole-based azaphenacene molecules. While in Scheme C we used the sampling scheme described in Chapter 6, a summary of the sampling schemes used by all four methods is given in Table 7.7.

Energy-density plots of three CSPs are shown in Figure 7.16, red plot points

Z'	SG	Scheme A	Scheme B	Scheme C
1	$P2_1/c$	5000	5000	4000
1	$P2_12_12_1$	1000	2000	4000
1	$P\bar{1}$	1000	2000	4000
1	$P2_1$	1000	2000	4000
1	$Pbca$	1000	2000	4000
1	$C2/c$	1000	2000	4000
1	$Pna2_1$	1000	2000	2000
1	Cc	1000	2000	2000
1	$Pca2_1$	1000	2000	2000
1	$C2$	-	2000	2000
1	$P1$	-	2000	2000
1	$Pbcn$	-	2000	-
1	Pc	-	2000	-
1	$P2_12_12$	-	2000	-
1	$P4_32_12$	-	2000	-
1	$P4_1$	-	2000	-
1	$P3_2$	-	2000	-
1	$Fdd2$	-	2000	-
1	$Pccn$	-	2000	-
1	$P2/c$	-	2000	-
1	$P6_1$	-	2000	-
1	$I4_1/a$	-	2000	-
1	$R\bar{3}$	-	2000	-
2	$P\bar{1}$	12500	-	-
2	$P2_1/c$	7500	-	-
2	$P2_1$	7500	-	-

Table 7.7: The number of successful minimisations for each Z' | SG search used in the reduced sampling scheme devised in Subsection 7.3.2 (Scheme A) and the sampling schemes used for the computational screening of small molecule organic semiconductors in Ref. [58] (Scheme B) and Chapter 6 (Scheme C).

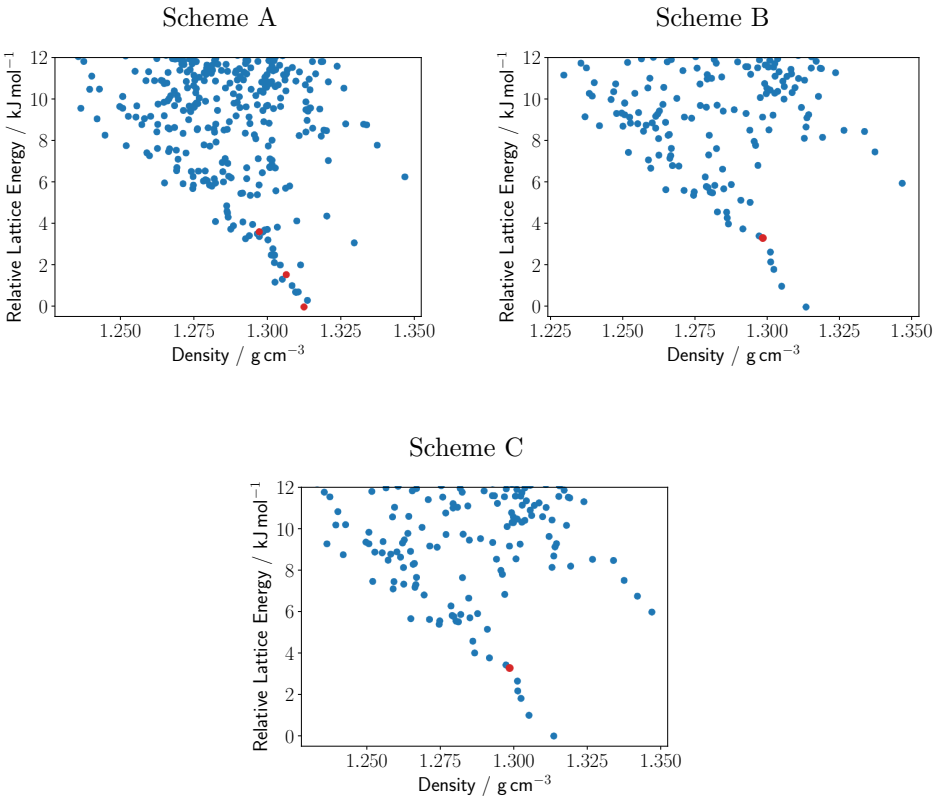


Figure 7.16: Energy-density plots of pentacene using three different sampling schemes. Plot points coloured in red indicate matches with experimental crystal structures from the CSD.

Sampling	Wall Time	Number of Unique Structures	Unique Structures per Hour
Scheme A	13h 32m 5s	12326	910.7
Scheme B	9h 25m 58s	5209	552.2
Scheme C	5h 43m 44s	3098	540.8

Table 7.8: Wall times, the number of unique structures, and the number of unique structures per hour for the CSPs of pentacene using three different sampling schemes. CSP searches were carried out using 4 nodes on the Iridis 5 compute cluster with dual Intel® Xeon® Gold 6138 CPUs @ 2.00GHz on each node.

are the predicted structures that match with the experimental. Scheme A was the only method that found all three pentacene polymorphs and had shown a larger number of unique structures in its crystal structure landscape. The other two methods had only been able to match with one out of three of the known experimental polymorphs of pentacene. In Table 7.8 we show the wall times of the CSPs for each sampling method using 4 nodes on the Iridis 5 compute cluster with dual Intel® Xeon® Gold 6138 CPUs @ 2.00GHz on each node, the total number of unique structures, and the average number of unique structures per hour. Scheme A was the most computational expensive but it is still far more efficient in terms of the number of unique structures per hour out of the three schemes. From these results, Scheme A shows some clear advantages and should therefore be used over the other two sampling schemes for the computational screening of organic semiconductors.

7.4 Carrier Mobilities

Before mobility evaluations are carried on the tetracene derivatives test set we first optimise their molecular geometries with B3LYP/6-311G** using GAUSSIAN09 but with all nuclei fixed except for the hydrogen nuclei and paste this optimised molecule back into their experimental crystal structures. We optimise only the hydrogen positions as experimental hydrogen bond lengths are known to be underestimated and we leave all other atoms fixed so that the correct conformations are maintained. We change all crystal structures with pasted

CSD Refcode	RMSD30 / Å	RMSD60 / Å	PXRD
GUMZOE	0.357	0.459	0.945
DUWHOT	0.170	0.207	0.994
TETCEN	0.360	0.481	0.948
GUMZIY	0.109	0.132	0.995
DUWHEJ	0.196	0.231	0.995
DUWPER	0.604	0.747	0.944
GUNBAT	0.480	0.690	0.985
GUMZUK	0.268	0.368	0.986
GUNBEX	0.136	0.167	0.996
HIGNIV	0.194	0.240	0.992

Table 7.9: COMPACK RMSDs for a 30 out of 30 and a 60 out of 60 cluster with distance and angles tolerances of 30% and 30° or smaller and hydrogen positions ignored and X-ray similarities of the experimental structures reoptimised with DMACRYS using the FIT force field against the original experimental crystal structure.

molecules to the $P1$ subgroup and energy minimise using the same methods as the final minimisation stage used for the CSPs described in Section 7.2. Energy minimising the crystal structure with a force field can lead to differences with the experimental and therefore greater errors in the carrier mobilities. However, this is done so that it is consistent with the computational screening workflow and to allow for phonon calculations to be carried out to obtain nonlocal electron-phonon coupling parameters.

In Table 7.9 we show the RMSDs for the COMPACK comparisons using a 30 and 60 molecule cluster with distance and angle tolerances of 30% and 30° or smaller and hydrogen positions ignored and the simulated X-ray powder pattern similarities using the CSD Python API between the experimental and the force field optimised structures. Overall results are good with low RMSDs for both the 30 and 60 molecule cluster, although slightly larger RMSDs are obtained for GUNBAT and DUWPER. For the X-ray powder pattern similarities, good agreements are obtained for all crystal structures. From these results, we can be fairly confident that the errors due to the differences between the experimental and the optimised structures should be small relative to the errors of the mobility evaluation methods themselves. In the following subsections, we describe the three different methods used to evaluate the carrier mobilities against FOB-SH.

7.4.1 Marcus Theory

In one of the three methods, we determine mobilities with the Marcus theory transition rates, the equation

$$\mu_{\alpha\beta} = \frac{e_0}{2k_{\mathrm{B}}T} \sum_{n,n' \neq n} P_n R_{nn'\alpha} R_{nn'\beta} k_{\mathrm{MT},n' \leftarrow n} \tag{7.1}$$

as described in Subsection 3.5.1, and $\bar{\mu} = \mathrm{tr}(\boldsymbol{\mu})/3$ to obtain the isotropic component, results are shown in Table 7.13. The site occupation probabilities P_n are approximated with the fraction $1/N$ where N is the number of molecules in the unit cell since all crystal structures are formed from only one type of molecule and $\mathbf{R}_{nn'}$ are approximated using centroid-centroid distances between the molecules n and n'. The Marcus theory transition rates $k_{\mathrm{MT},n' \leftarrow n}$ were evaluated with reorganisation energies calculated using the four-point scheme with using B3LYP/6-311+G** GAUSSIAN09 energies, while electronic couplings were calculated using the analytic overlap method (AOM) [175, 302–304] with the pyAOMlite library. Transition rates were calculated at 300 K and were determined from each molecule in the unit cell to all active molecules. Active molecules are all molecules in close contact with any unit cell molecule. Molecules are determined to be in close contact when any of their atom-atom distances are found to be less than the sum of their van der Waals radii plus 1.5 Å.

7.4.2 Transient Localisation Theory

The other two methods will run TLT calculations which will need the crystal structures to be expanded out twice, once for the phonon calculations and once again to create the TLT Hamiltonian. To determine the best expansion directions we calculate the mobility tensor using Marcus theory and obtain the mobilities along each unit cell axis direction. We then expand the crystal structure so that the unit cell lengths are approximately in the same ratios as the square roots of the mobilities along its unit cell axis direction. We use square roots of the mobilities so that they are proportional to the localisation lengths. For the phonon calculations, all optimised crystal structures were expanded to a supercell containing a total of 24 molecules in its unit cell. To build the TLT Hamiltonians

CSD Refcode	Phonon Expansion	TLT Expansion	TLT Molecules
GUMZOE	3, 2, 4	27, 12, 24	7776
DUWHOT	1, 3, 2	3, 36, 18	7776
TETCEN	2, 6, 1	42, 18, 5	7560
GUMZIY	8, 1, 1	144, 6, 3	7776
DUWHEJ	24, 1, 1	144, 9, 6	7776
DUWPER	1, 6, 1	5, 126, 3	7560
GUNBAT	24, 1, 1	144, 9, 6	7776
GUMZUK	12, 2, 1	108, 12, 6	7776
GUNBEX	1, 6, 1	3, 126, 5	7560
HIGNIV	1, 6, 1	3, 126, 5	7560

Table 7.10: Expansion of the crystal structures of the 10 tetracene derivatives used for the phonon and TLT calculations, and the total number of molecules in the unit cell of the crystal structures used for the TLT calculations.

the supercells are expanded again so that the new supercells contain a total of 7,560 or more unit cell molecules. Additionally, the TLT expansion minimum unit cell lengths are set so that they are larger than the graph of electronic couplings. The expansions that are used and the total number of molecules in the TLT system are given in Table 7.10.

We will refer to the two different TLT calculations as TLT1 and TLT2. In TLT1, all local electron-phonon couplings were set to zero while for TLT2 local intramolecular electron-phonon couplings are included. For TLT2 local electron-phonon coupling parameters were approximated by running geometry optimisations and frequency calculations on the neutral molecules. Force calculations are then carried out using the neutral optimised geometries with a positive charge on the molecule. The energy derivatives from the force calculations are transformed into the normal mode coordinate system using the frequency calculation results and local electron-phonon coupling parameters are then calculated with the equation

$$V_{n\alpha} = \frac{1}{\hbar\omega_\alpha}\sqrt{\frac{\hbar}{2\mu_\alpha\omega_\alpha}}\left\{\frac{\partial\langle n|\hat{H}_{\mathrm{e}}^{\pm}(\mathbf{r}_{\mathrm{e}},\mathbf{x}_{\mathrm{n}})|n\rangle}{\partial x_\alpha}\bigg|_{\mathbf{0}} + \frac{\partial\hat{V}_{\mathrm{nn}}(\mathbf{x}_{\mathrm{n}})}{\partial x_\alpha}\bigg|_{\mathbf{0}}\right\} \tag{7.2}$$

as derived in Subsection 3.4.1. Where we used B3LYP/6-311+G** with the `Int=UltraFine` integration grid and for the geometry optimisations the `Opt=Tight`

CSD Refcode	Λ (Four-point) / meV	Λ (Local Couplings) / meV	Λ (Ref. [66]) / meV
GUMZOE	123.5	125.6	136.0
DUWHOT	144.7	140.9	130.2
TETCEN	113.7	116.5	116.5
GUMZIY	120.0	122.3	123.2
DUWHEJ	133.0	132.2	135.0
DUWPER	121.3	123.4	123.7
GUNBAT	133.3	131.0	139.5
GUMZUK	130.9	130.1	138.1
GUNBEX	136.5	134.7	138.9
HIGNIV	136.9	134.9	140.8

Table 7.11: Reorganisation energies of the 10 tetracene derivatives: calculated using B3LYP/6-311+G** with the four-point scheme and Eq. (7.4) and calculated values from Ref. [66].

convergence criteria.

Reorganisation energies with the local electron-phonon coupling parameters can be calculated using the equation

$$\Lambda_{nn'} = \sum_{\alpha} \hbar\omega_{\alpha}(V_{n\alpha} - V_{n'\alpha})^2 \tag{7.3}$$

where summation over α run over all vibrational modes of the crystal structure which reduces to

$$\Lambda_{nn'} = 2\sum_{\alpha'} \hbar\omega_{\alpha'} V^2_{\mathrm{mol},\alpha'} \tag{7.4}$$

as the vibrational modes were approximated as a sum of the isolated molecule vibrational modes α'. To check the correctness of our methodology we compare reorganisation energies using the four-point scheme, local electron-phonon couplings with Eq. (7.4), and values from Ref. [66]. Good agreements are obtained between all three methods, results are shown in Table 7.11.

The nonlocal electron-phonon coupling parameters are calculated using the

equation

$$V_{nn'\alpha} = \frac{1}{\hbar\omega_\alpha}\sqrt{\frac{\hbar}{2\mu_\alpha\omega_\alpha}}\frac{\partial\langle n|\hat{H}_{\mathrm{e}}^{\pm}(\mathbf{r}_{\mathrm{e}},\mathbf{x}_{\mathrm{n}})|n'\rangle}{\partial x_\alpha}\bigg|_{\mathbf{0}} \tag{7.5}$$

where we consider only the Γ-point phonon frequencies and eigenvectors of rigid molecules. We run DMACRYS minimisations followed by the phonon calculations on the supercell. For both calculations the same methods as the final optimisation stage of the CSPs described in Section 7.2 were used but with changes made to tighten the convergence criteria with LIMI set to 1×10^{-8} and changes to increase the accuracy of the Ewald summation with ACCM set to 1×10^{10}. Electronic couplings and derivatives were calculated for each molecule in the unit cell to all active molecules in close contact with that molecule and were approximated with AOM using the pyAOMlite library. Derivatives of the electron couplings were calculated by applying finite displacements for each degree of freedom (three translations and three rotations) of each molecule in the unit cell. We use the symmetric difference quotient and set the maximum absolute value of the displacements to 0.01 Å for any atom of that molecule. This generates a rank-3 tensor of derivatives with respect to one of the degrees of freedom of a molecule in the unit cell. Using the phonon frequencies, eigenvectors, and reduced masses from the phonon calculations the rank-3 tensor of derivatives were transformed to a rank-3 tensor of nonlocal electron-phonon coupling parameters.

The mean values of the site energies for all calculations were set to zero, variances in the site energies, and the means and variances in the electronic couplings were calculated from the local and nonlocal electron-phonon coupling parameters. We use the equations derived in Appendix B but run the summations for the intramolecular vibrational modes over the isolated molecular vibrational modes as explained for the calculation of the reorganisations energies. The equation for the variances of the site energies becomes

$$\sigma_n^2 = \sum_a \hbar^2\omega_a^2 V_{\mathrm{mol},a}^2 \coth\left(\frac{\hbar\omega_a}{2k_{\mathrm{B}}T}\right) \tag{7.6}$$

where the summation over a runs over all low-frequency intramolecular vibrational

modes with $\hbar\omega_a < 2k_\mathrm{B}T$. While the equations for the means and variances of the electronic couplings becomes

$$\mu_{nn'} = V_{nn'} \exp\left[-\sum_b V_{\mathrm{mol},b}^2 \coth\left(\frac{\hbar\omega_b}{2k_\mathrm{B}T}\right)\right] \tag{7.7}$$

$$\begin{aligned}\sigma_{nn'}^2 = \Bigg\{ &\sum_c \hbar^2\omega_c^2 V_{nn'c}^2 \coth\left(\frac{\hbar\omega_c}{2k_\mathrm{B}T}\right) \\ &\times \exp\left[-2\sum_b V_{\mathrm{mol},b}^2 \coth\left(\frac{\hbar\omega_b}{2k_\mathrm{B}T}\right)\right]\Bigg\}\end{aligned} \tag{7.8}$$

where summation over b runs over all high-frequency intramolecular vibrational modes with $\hbar\omega_b \geq 2k_\mathrm{B}T$ and the summation over c runs over all intermolecular vibrational modes from the rigid molecule calculations. With the means and variances of the site energies and electronic couplings, we run TLT calculations following closely the procedure described in Ref. [65]. We use $\hbar/\tau = 5\,\mathrm{meV}$, a temperature of 300 K, and the localisation lengths were averaged over 50 realisations of disorder. Isotropic mobilities were calculated by taking an average over the diagonal elements of the localisation length tensor

$$\bar{\mu}_\mathrm{TLT} = \frac{e}{k_\mathrm{B}T}\frac{\bar{L}^2(\tau)}{2\tau} \qquad \bar{L}^2(\tau) = \frac{1}{3}\left[\bar{L}_{xx}^2(\tau) + \bar{L}_{yy}^2(\tau) + \bar{L}_{zz}^2(\tau)\right] \tag{7.9}$$

results from both TLT methods are shown in Table 7.13.

7.4.3 Comparisons

In Table 7.12 we show the largest absolute value and deviations of the electronic couplings for the different methods. Generally, electronic couplings show fairly good agreement between our methods and the values from FOB-SH. Some differences are expected as FOB-SH runs a molecular dynamics simulation of the crystal structure while our methods use static structures. The deviations in the electronic coupling used in the TLT methods are consistently underestimated as expected since we calculated nonlocal electron-phonon couplings for rigid molecules while FOB-SH uses fully atomistic molecular dynamics simulations.

CSD Refcode	Electronic Coupling (Marcus Theory) / meV	Electronic Coupling (TLT) / meV	Electronic Coupling (FOB-SH) / meV
GUMZOE	2.7	2.6 ± 6.4	18.4 ± 33.6
DUWHOT	23.3	23.4 ± 1.9	76.2 ± 29.4
TETCEN	45.7	45.3 ± 14.6	88.7 ± 32.4
GUMZIY	135.2	135.4 ± 12.0	146.1 ± 21.8
DUWHEJ	117.9	117.8 ± 18.1	124.1 ± 56.3
DUWPER	139.0	139.0 ± 13.9	101.5 ± 27.3
GUNBAT	160.9	161.0 ± 10.9	122.4 ± 25.7
GUMZUK	163.9	163.8 ± 11.5	143.4 ± 23.3
GUNBEX	207.7	207.7 ± 10.7	137.1 ± 28.3
HIGNIV	214.6	214.5 ± 11.8	156.2 ± 25.2

Table 7.12: The largest absolute value of the electronic couplings and their deviations for the crystal structures of the 10 tetracene derivatives from the Marcus Theory, TLT1, and TLT2 calculations, and FOB-SH calculations from Ref. [66].

CSD Refcode	$\bar{\mu}$ (Marcus Theory) / $cm^2(Vs)^{-1}$	$\bar{\mu}$ (TLT1) / $cm^2(Vs)^{-1}$	$\bar{\mu}$ (TLT2) / $cm^2(Vs)^{-1}$	$\bar{\mu}$ (FOB-SH) / $cm^2(Vs)^{-1}$
GUMZOE	0.02	0.39	0.26	0.6
DUWHOT	0.91	20.77	13.29	2.7
TETCEN	1.68	9.58	6.81	3.5
GUMZIY	5.19	20.48	19.00	20.8
DUWHEJ	6.42	8.87	9.87	1.3
DUWPER	10.17	21.04	20.01	8.2
GUNBAT	10.56	30.34	27.97	9.6
GUMZUK	11.40	27.62	25.57	15.8
GUNBEX	14.96	38.67	35.29	11.0
HIGNIV	15.54	35.58	32.96	13.2

Table 7.13: Hole mobilities of the crystal structures of the 10 tetracene derivatives calculated using the Marcus Theory, TLT1, and TLT2 methods, and FOB-SH values from Ref. [66].

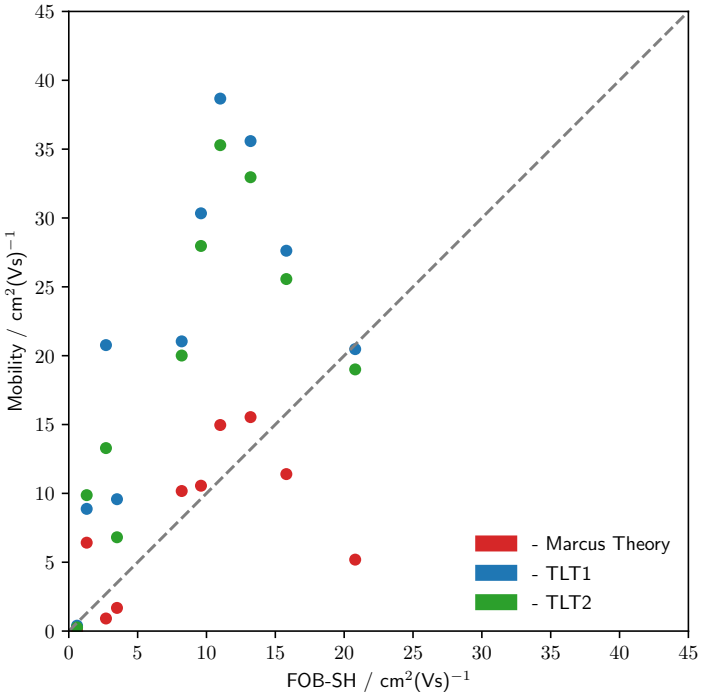


Figure 7.17: Scatter plot of the hole mobilities of the crystal structures of the 10 tetracene derivatives calculated using the Marcus Theory, TLT1, and TLT2 methods against FOB-SH values from Ref. [66].

Results from Marcus theory and both TLT methods and FOB-SH reference values are shown in Table 7.13. Plots against FOB-SH references are shown in Figure 7.17 which show that the TLT methods had overestimated the hole mobilities for nearly all tetracene derivatives as expected given the underestimated nonlocal electron-phonon couplings. We can see that similar to the Marcus theory results, hole mobilities for the crystal structures GUMZIY and GUMZUK were not predicted by the TLT methods to have the largest hole mobilities going against the FOB-SH results. Overall clear improvements are not made with the TLT methods using rigid molecule phonons so that Marcus theory will still be the preferred method for a computational screening procedure given its much lower computational costs.

7.5 Conclusion

In this chapter, we benchmarked the conformer generation methods, the FIT force field, and two different sampling schemes used in the CSP workflow and three different mobility evaluation methods. We determined that for the aromatic molecules similar to those that will be used in the computational screening workflow, the generation of only a single conformer per molecule was sufficient for almost all molecules in the test set. From the results of the CSP benchmark, we were able to devise a reduced sampling scheme that was capable of obtaining all experimental crystal structures and the lowest energy structures from the original CSPs. While from the mobility evaluations benchmark results we determined that TLT methods using rigid molecule phonons did not provide a clear improvement over Marcus theory results and came to the conclusion that for the computational screening of small molecular organic semiconductors, carrier mobilities calculated with Marcus theory still appear to be the best choice given its particularly low computational costs.

We also showed that the reduced sampling scheme we developed had been far more efficient at finding unique structures than the sampling schemes used by Yang *et al.* in Ref. [58] and the one described in Chapter 6 when applied to the CSP of pentacene. This suggests that the sampling schemes used by Yang *et al.* and in Chapter 6 had oversampled some of their space groups. Additionally, all three experimental polymorphs of pentacene had been obtained by the reduced sampling scheme while only one of the three polymorphs had been obtained by the sampling schemes used by Yang *et al.* and in Chapter 6 which suggests that the CSP using these sampling schemes would not be the most reliable at finding experimental structures for planar aromatic molecules like pentacene. The main difference in our reduced sampling scheme is the inclusion of the Z'=2 searches which had led to both increased efficiencies and an improved ability to obtain experimental structures and shows that Z'=1 only sampling schemes may be inadequate in general.

In this benchmark chapter, we used DFT with the B3LYP/6-311G** functional and basis set for certain calculations involved in the CSP searches. The choice of the B3LYP/6-311G** functional and basis set was made to stay in line

with standard practices within the research group. Although it is reasonable to expect that with changes made to the functional and basis sets some improved results could be obtained for our set of benchmark molecules, we however do not attempt any benchmarking of this type since such attempts would undoubtedly require a much larger number of CSP searches to be carried out. Additionally, it may be that to obtain significantly improved agreements with experimental crystal structures an improved description of the force field would be necessary rather than any changes to the functional and basis sets used. Further work to develop and benchmark improved CSP workflows are important but are beyond the scope of this project.

The experimental crystal structures GUMZIY and GUMZUK were not predicted by the Marcus theory and both TLT methods to have the largest hole mobilities against the FOH-SH results. The exact reason for the disagreement will be difficult to determine given the large number of approximations used in the Marcus theory, TLT, and FOB-SH methods. Although one possibility for the disagreement of TLT with FOB-SH could be due to the methods used to determine the deviations in the electronic couplings which had used rigid molecule phonons. Another possibility could be the tight-binding Hamiltonian used to derive the means and deviations of the electronic coupling which had assumed that the site energies and electronic couplings were linear to the normal mode coordinates of the nuclei.

Full atom phonon calculations or molecular dynamics simulations may therefore be needed to obtain the mean and deviations of the site energies and electronic couplings to bring TLT results in line with FOB-SH. If these adjustments show no further improvements it may be that TLT is simply unable to capture the physics that gives the GUMZIY and GUMZUK structures their large hole mobilities, in this case, other methods that go beyond TLT will be required. One possibility for example is the dynamical localisation corrections method which interpolates between the transient localisation and the Bloch-Boltzmann theories [305]. Further work to benchmark and evaluate TLT with further adjustments and alternative methods will be necessary but is beyond the scope of this project.

Chapter 8

IFO Derivatives

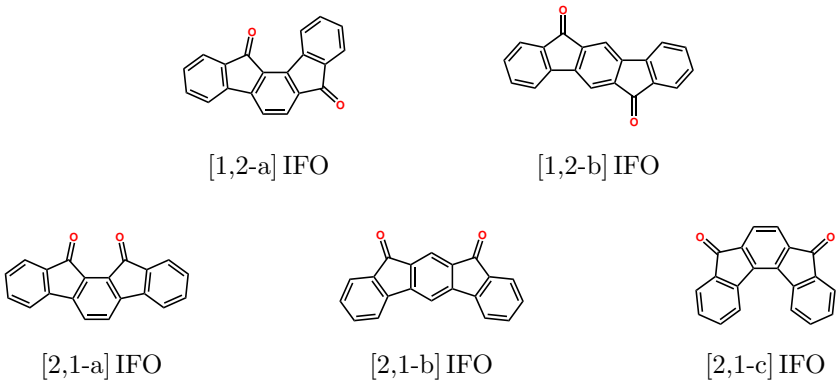


Figure 8.1: The five IFO isomers used for the computational screening of high-performance IFO based organic semiconductors.

In this chapter, we will try to tackle some of the issues raised in the conclusion of Chapter 6 by increasing the number of molecules evaluated in the CSP stage and make use of the reduced sampling scheme developed in Section 7.3. We will carry out the computational screening on a search space of indenofluorenedione (IFO) molecules with nitrogen and/or fluorine substitutions. All five unsubstituted IFO isomers are shown in Figure 8.1. The IFO search space was chosen due to previously reported experimental work which focused only on [1,2-b] IFO derivatives which suggests that this IFO search space including the derivatives of the other IFO isomers could contain promising molecules for organic electronic applications.

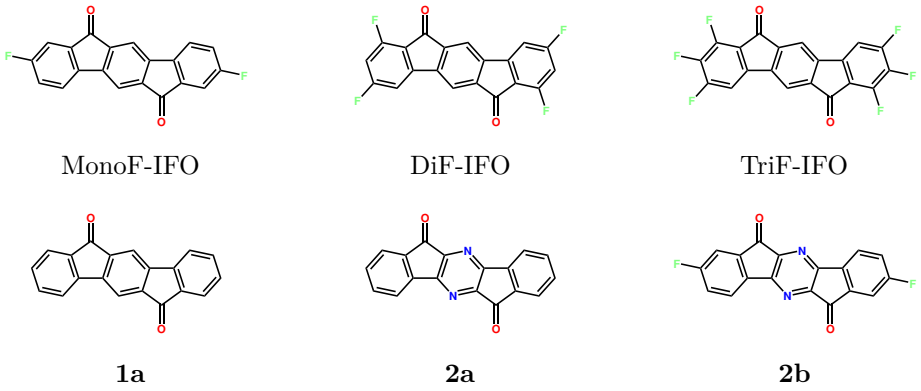


Figure 8.2: Six experimentally synthesised IFO derivatives used in the fabrication of n-type OFETs.

Previous experimental results on the IFO derivatives include the work by Park *et al.* [306] who fabricated OFETs devices with gold contacts using the MonoF-IFO, DiF-IFO, and TriF-IFO semiconductors resulting in electron mobilities of 0.14, 0.07, and 0.16 $\mathrm{cm}^2(\mathrm{Vs})^{-1}$; threshold voltages of 32.1, 22.2, and 9.2 V; and on/off ratios above 10^5. Additionally, they found that the TriF-IFO OFET showed excellent air stability over a three month test period. In another example Nakagawa *et al.* [307] fabricated OFETs using a number of different IFO derivatives including [1,2-b] IFO (**1a**), MonoF-IFO, diindeno[1,2-b;1',2'-e]pyrazine-6,12-dione (**2a**), and a fluorine substituted diindeno[1,2-b;1',2'-e]pyrazine-6,12-dione (**2b**). In particular, they found that an OFET using gold contacts in the top contact geometry and the black solid polymorph of **2b** gave an OFET with electron mobilities of 0.17 $\mathrm{cm}^2(\mathrm{Vs})^{-1}$, a threshold voltage of 17 V, and an on/off ratio of 10^7. Molecular structures for MonoF-IFO, DiF-IFO, TriF-IFO, **1a**, **2a**, and **2b** are drawn in Figure 8.2 and a summary of the experimental properties of the OFETs fabricated with these molecules is given in Table 8.1.

Semiconductor	μ / $cm^2(Vs)^{-1}$	on/off	V_T / V	Geometry	Surface	Contact	Reference
MonoF-IFO	7.0×10^{-2}	6.9×10^5	31.7	top-contact	polystyrene	LiF/Al	[306]
MonoF-IFO	1.4×10^{-1}	1.2×10^5	32.1	top-contact	polystyrene	Au	[306]
MonoF-IFO	1.9×10^{-4}	6.0×10^3	88	bottom-contact	bare	Au	[307]
MonoF-IFO	1.7×10^{-1}	2.0×10^7	69	bottom-contact	HMDS	Au	[307]
MonoF-IFO	6.6×10^{-2}	2.0×10^4	75	top-contact	HMDS	Au	[307]
DiF-IFO	5.0×10^{-2}	4.3×10^5	26.1	top-contact	polystyrene	LiF/Al	[306]
DiF-IFO	7.0×10^{-2}	5.4×10^5	22.2	top-contact	polystyrene	Au	[306]
TriF-IFO	9.0×10^{-2}	1.8×10^5	4.7	top-contact	polystyrene	LiF/Al	[306]
TriF-IFO	1.6×10^{-1}	6.8×10^5	9.2	top-contact	polystyrene	Au	[306]
1a	-	-	-	bottom-contact	HMDS	Au	[307]
2a	2.7×10^{-4}	2.0×10^4	70	bottom-contact	HMDS	Au	[307]
2b (red)	2.7×10^{-4}	2.0×10^4	21	bottom-contact	bare	Au	[307]
2b (red)	8.7×10^{-3}	9.0×10^5	27	bottom-contact	HMDS	Au	[307]
2b (black)	8.8×10^{-4}	2.0×10^5	24	bottom-contact	bare	Au	[307]
2b (black)	1.1×10^{-2}	1.0×10^6	27	bottom-contact	HMDS	Au	[307]
2b (black)	1.7×10^{-1}	1.0×10^7	17	top-contact	HMDS	Au	[307]

Table 8.1: A summary of the electrical properties of a number of OFETs fabricated using the MonoF-IFO, DiF-IFO, TriF-IFO, **1a**, **2a**, and **2b** organic semiconductors with a SiO_2/Si substrate, the OFET fabricated using **1a** showed no gate effect.

8.1 Evolutionary Algorithm

```
molecules = ['O=c1c2ccccc2c2cc3c(=O)c4ccccc4c3cc12',
 ↪  'O=c1c2ccccc2c2c1ccc1c3ccccc3c(=O)c12',
 ↪  'O=c1c2ccccc2c2ccc3c4ccccc4c(=O)c3c12',
 ↪  'O=c1c2ccccc2c2cc3c(cc12)c(=O)c1ccccc13',
 ↪  'O=c1c2ccccc2c2c1ccc1c(=O)c3ccccc3c12']
fragments = []
mutations_1 = ['[#6R1&H]', '[#6R1]-F', '[#7R1&H0]']
mutations_2 = []
molsize = [5, 5]
```

Listing 4: The five variables used to define a search space of IFO derivatives that the EA will explore and consists of: all IFO isomers; C-H, C-F and N mutations; and a min and max molecule size of 5. The search space contains all IFO isomers with any number of nitrogen or fluorine substitutions on the six-membered rings of the IFO structure.

The IFO derivatives search space is defined by the input variables in Listing 4 and contains a total of 177,876 unique molecules. The settings used in the EA are the same as those described in Section 5.4 except for the recombination rate which was set to zero to ensure that the searches are restricted to only the IFO derivatives. To evaluate a molecule's ability to produce crystal structures with high electron mobilities we use the fitness function

$$F_{2,W}(\Lambda_-, \mathrm{A_s}) = \sqrt{\Lambda_-^2 + (W - \mathrm{A_s})^2} \tag{8.1}$$

where both the reorganisation energy Eq. (3.144) and electron affinity Eq. (5.25) were calculated using B3LYP/6-311+G** with GAUSSIAN09. Unlike Chapter 6 the aim here is to generate a large sample of molecules with MolBuilder since from the nitrogen substituted PAHs study we showed that higher reorganisation energies could be compensated with higher electronic couplings. We, therefore, run MolBuilder using nine different target work functions: 3.00, 3.25, 3.50, 3.75, 4.00, 4.25, 4.50, 4.75, and 5.00 eV running 4 EA searches for each target work function which should sample the search space with a range of electron affinities biased towards lower reorganisation energies.

8.2 Property Maps

From all 36 EA searches with the various target work functions of the fitness function $F_{2,W}$ we generate the property map of the sampled IFO derivatives which includes a total of 35,985 molecules or around 20.2% of the search space. In Figures 8.3 and 8.4 we can see that the IFO derivatives, in general, have much higher reorganisation energies and the distributions of reorganisations energies and electron affinities appear less correlated than the nitrogen substituted PAHs. The higher electron affinity nitrogen substituted PAHs tended to have higher reorganisation energies while this does not appear to be the case for the IFOs.

In Figure 8.3a we plot the property map coloured by the IFO structure of the molecule. In general, we see the molecules with the [1,2-a] IFO structures are able to obtain the lowest reorganisation energies out of all the other structures across a range of electron affinities. There appears to be some similarity between the [1,2-a] IFO structures and the low reorganisation energy molecules **2A**-**6A**, **8A**-**10A** found in the nitrogen substituted PAHs searches. In both cases, low reorganisation energies were achieved from a non-linear structure. We also find that the [1,2-b] IFO structure which most experimental work had concentrated on was not the best molecule based on its reorganisations energies, see Figure 8.5. In Figure 8.6 we show histograms of the reorganisations energies for the five different IFO structures of the molecules sampled by the EA which visually shows that the [1,2-a] IFO and [2,1-b] IFO had a greater number of low reorganisations energy molecules so that in general the [1,2-a] IFO and [2,1-b] IFO derivatives could be the most likely to lead to the highest electron mobilities crystal structures.

In Figure 8.3b we plot the property map coloured with the number of nitrogens on each molecule. Similarly to the previous study on nitrogen substituted PAHs we can see that there is a direct relationship between the number of nitrogen atoms and the electron affinity of the molecule. In Figure 8.4a we plot the property map coloured with the number of fluorines for all molecules and in Figure 8.4b we only plot only the molecules that do not contain nitrogen atoms. Comparing Figures 8.3b and 8.4b we can see that there is a much weaker effect on the electron affinity with increasing fluorine atoms. It appears that to reach high electron affinity organic semiconductors it is preferable to have nitrogen

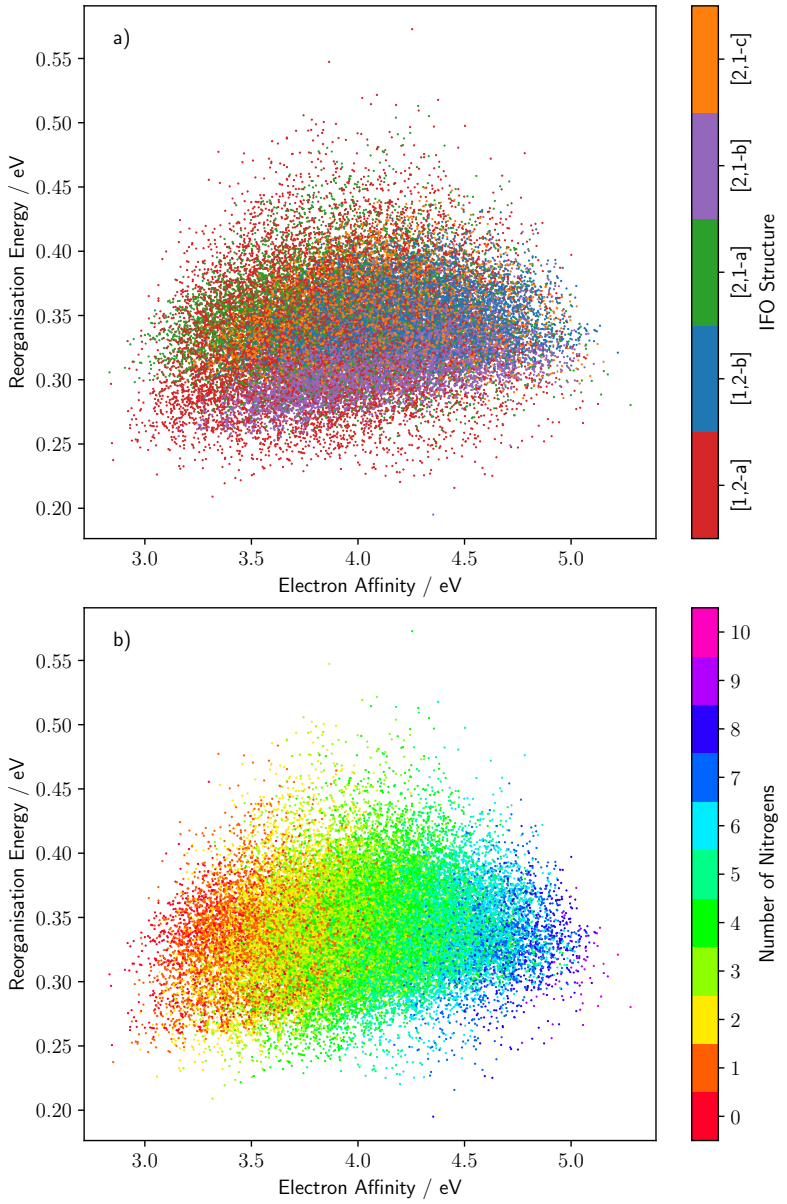


Figure 8.3: Property maps of 35,985 molecules that were sampled by the EA for fitness function $F_{2,W}$ using 9 different values of W with 4 EA searches each for the search space of IFO derivatives defined by Listing 4. Molecules are plotted with their reorganisation energies against solid-state electron affinities and are coloured by a) the IFO structure and b) the number of nitrogen atoms.

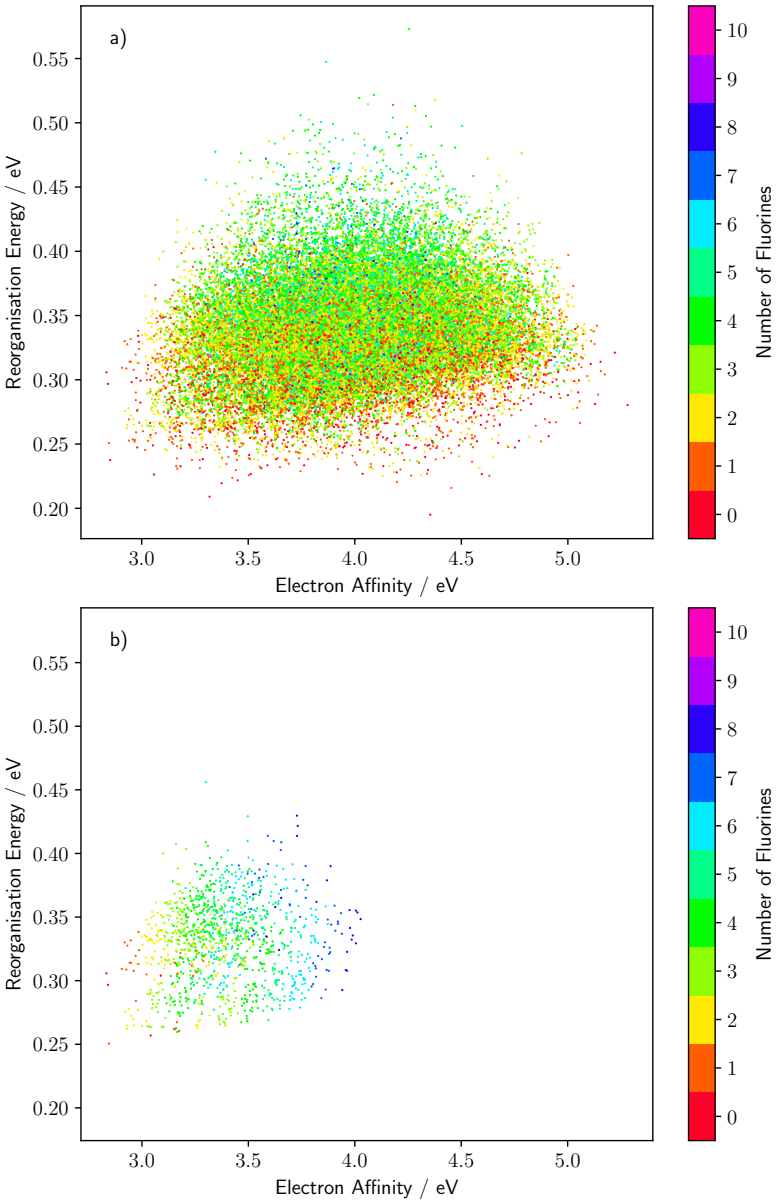


Figure 8.4: Property maps of 35,985 molecules that were sampled by the EA for fitness function $F_{2,W}$ using 9 different values of W with 4 EA searches each for the search space of IFO derivatives defined by Listing 4. Molecules are plotted with their reorganisation energies against solid-state electron affinities and are coloured by the number of fluorine atoms contained in the molecule. In figure a) we plot all molecules while in b) we only plot molecules that do not contain any nitrogen atoms.

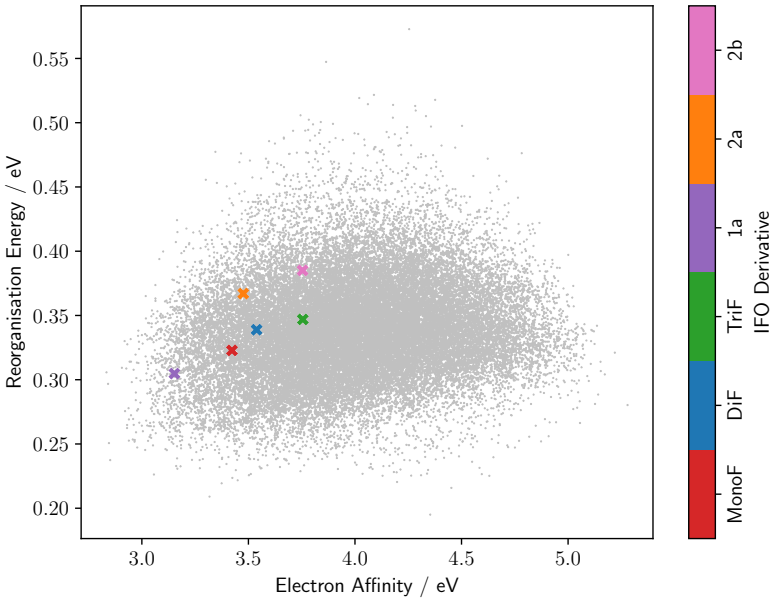


Figure 8.5: The positions of six experimentally synthesised IFO derivatives used in the fabrication of n-type OFETs in the property map of molecules sampled by the EA for fitness function $F_{2,W}$ using 9 different values of W with 4 EA searches each for the search space of IFO derivatives defined by Listing 4. Out of the six IFO derivatives only molecule **2a** had been sampled by the EA.

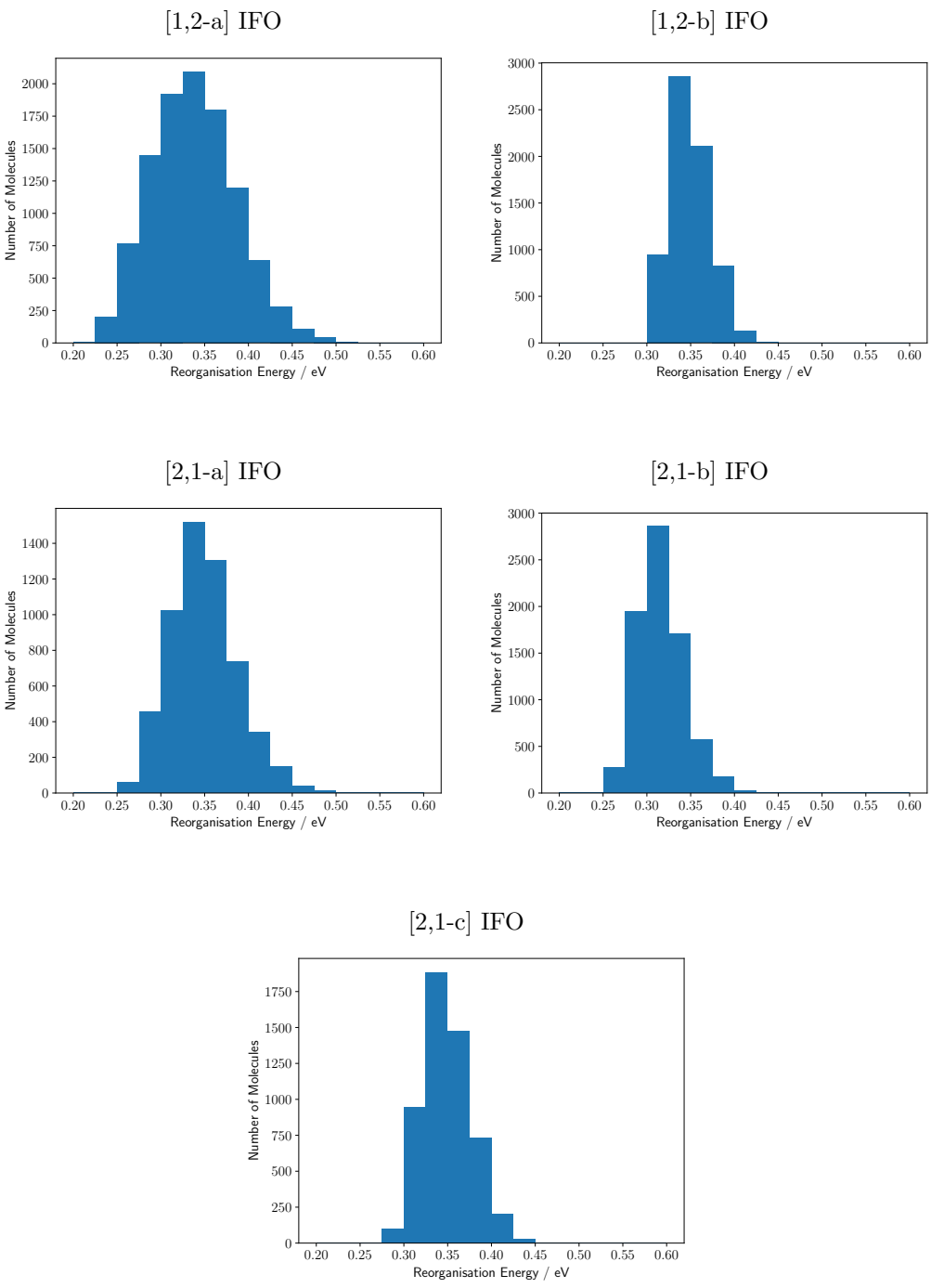


Figure 8.6: Histograms of the reorganisation energies for the five different IFO structures of the molecules sampled by the EA.

substitutions rather than fluorine.

8.3 Electron Mobilities

For the mobility evaluation stage of the computational screening workflow, we divide the evaluated molecules by the solid-state electron affinities into eight groups. Each group contains molecules which have electron affinities in the intervals of $(3.00, 3.25]$, $(3.25, 3.50]$, $(3.50, 3.75]$, $(3.75, 4.00]$, $(4.00, 4.25]$, $(4.25, 4.50]$, $(4.50, 4.75]$, or $(4.75, 5.00]$. This should give a range of molecules that could be matched to contacts with a specific work function and reduce the electron injection barriers. The minimum electron affinities were set to avoid having organic semiconductors which would need to be matched with particularly reactive low work function contacts while the maximum was set to avoid unstable molecules due to their highly electrophilic nature. From each group, we select 40 molecules with the lowest reorganisation energies to a CSP and electron mobility evaluations stage giving a total of 320 molecules consisting of 296 [1,2-a] IFO, 0 [1,2-b] IFO, 12 [2,1-a] IFO, 7 [2,1-b] IFO, and 5 [2,1-c] IFO derivatives.

CSP searches were carried out on all 320 molecules using the methods and space group sampling devised in Section 7.3. We use version II of the QR structure generation method with Z'=1 until a total of 5,000 successfully minimised structures are obtained for the most common Z'=1 space group ($P2_1/c$) and 1,000 in the next eight most common ($P2_12_12_1$, $P\bar{1}$, $P2_1$, $Pbca$, $C2/c$, $Pna2_1$, Cc, $Pca2_1$), and Z'=2 until a total of 12,500 successfully minimised structure are obtained for the most common Z'=2 space group ($P\bar{1}$) and 7,500 for the next two most common ($P2_1/c$, $P2_1$). In the final stage of lattice energy minimisations, all structures were converted to $P1$ to remove symmetry constraints. The FIT force field with atom centred multipoles up to hexadecapoles obtained from GDMA with B3LYP/6-311G** GAUSSIAN09 electron densities was used. Ewald summations were carried out for charge-charge, charge-dipole, and dipole-dipole interactions, while all other higher-order electrostatics up to r^{-5} and repulsion-dispersion interactions were calculated for atom-atom distances within a 30 Å cut-off radius. Duplicated crystal structures were identified and removed by generating simulated X-ray diffraction patterns using PLATON [108]

and calculating powder pattern similarities as described in Chapter 4.

Electron mobilities were calculated for each unique predicted crystal structure within a 12 kJ mol^{-1} range from the global minimum for each CSP using

$$\mu_{\alpha\beta} = \frac{e_0}{2k_\mathrm{B}T} \sum_{n,n' \neq n} P_n R_{nn'\alpha} R_{nn'\beta} k_{\mathrm{MT},n' \leftarrow n} \tag{8.2}$$

from Subsection 3.5.1 and $\bar{\mu} = \mathrm{tr}(\boldsymbol{\mu})/3$ to obtain the isotropic component. The site occupation probabilities P_n were approximated using $1/N$ where N is the number of molecules in the unit cell since all crystal structures are formed from only one type of molecule. The distances $\mathbf{R}_{nn'}$ were approximated using centroid-centroid distances between the molecules n and n'. The Marcus theory transition rates $k_{\mathrm{MT},n' \leftarrow n}$ were evaluated using reorganisation energies taken from the EA results and electronic couplings calculated using the AOM [175, 302, 304] with the pyAOMlite library. Transition rates were calculated at 300 K and were determined from each molecule in the unit cell to all active molecules. Active molecules are all molecules in close contact with any unit cell molecule. Molecules are determined to be in close contact when any of their atom-atom distances are found to be less than the sum of their van der Waals radii plus 1.5 Å.

For each molecule, the mobility of the global minimum $\bar{\mu}_\mathrm{GM}$ and the weighted average $\langle\bar{\mu}\rangle$ of the mobilities of the crystal structure landscape were used to determine its ability to form a crystal structure with high electron mobilities as done in Section 6.6. In Figure 8.7 we generate property maps of the reorganisation energies and calculated electron mobilities of the global minimum crystal structure or the average electron mobilities over the crystal structure landscape and colour plot points by its IFO structure. Both plots show that correlations between the reorganisation energies and the electron mobilities are still apparent even with the strong dependence on the molecular packing. Additionally, we can see that for reorganisation energies below around 0.26 eV the best performing molecules are obtained.

In Figure 8.8 we generate property maps of the electron affinities with the calculated electron mobility of the global minimum crystal structure and the average electron mobility over the crystal structure landscape with plot points coloured by the number of nitrogen atoms in the molecules. Both plots show some

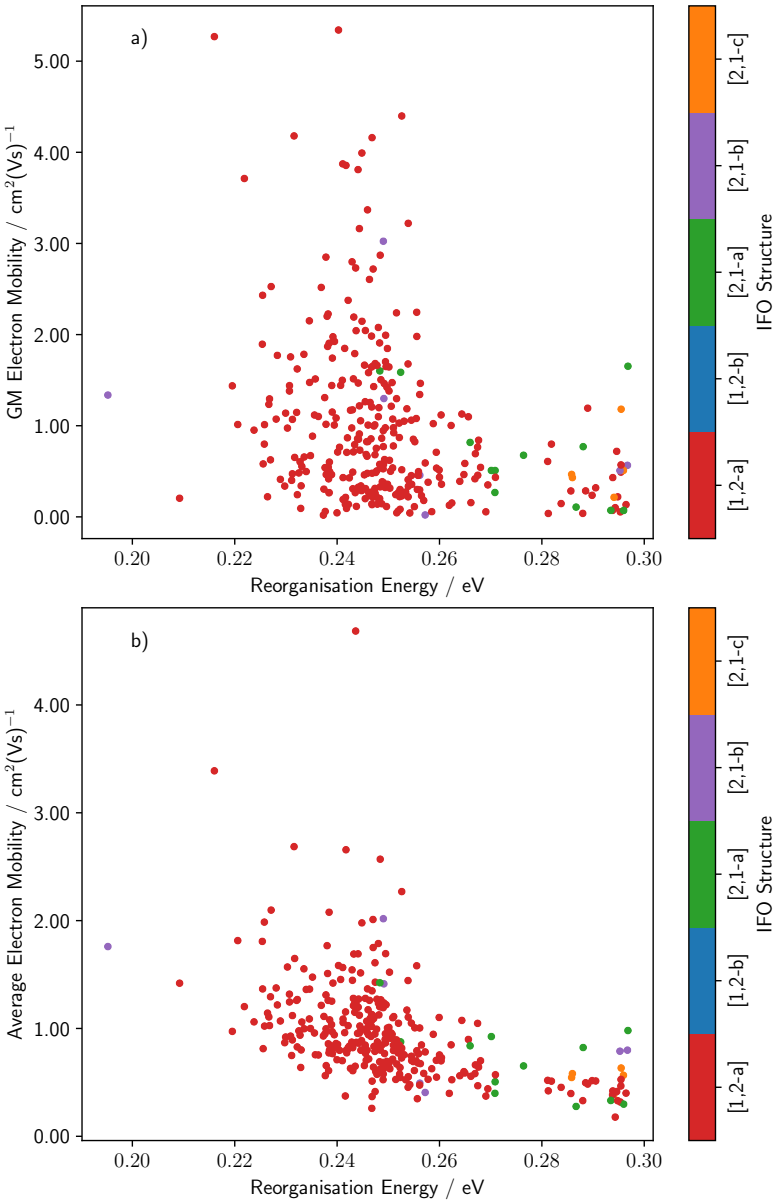


Figure 8.7: Property maps of the 320 molecules that were taken through the mobility evaluation stage of the computational screening workflow. Molecules are plotted with a) its electron mobility of the global minimum crystal structure and b) its average electron mobility against its reorganisations energies. Plot points are coloured by the IFO structure of the molecule.

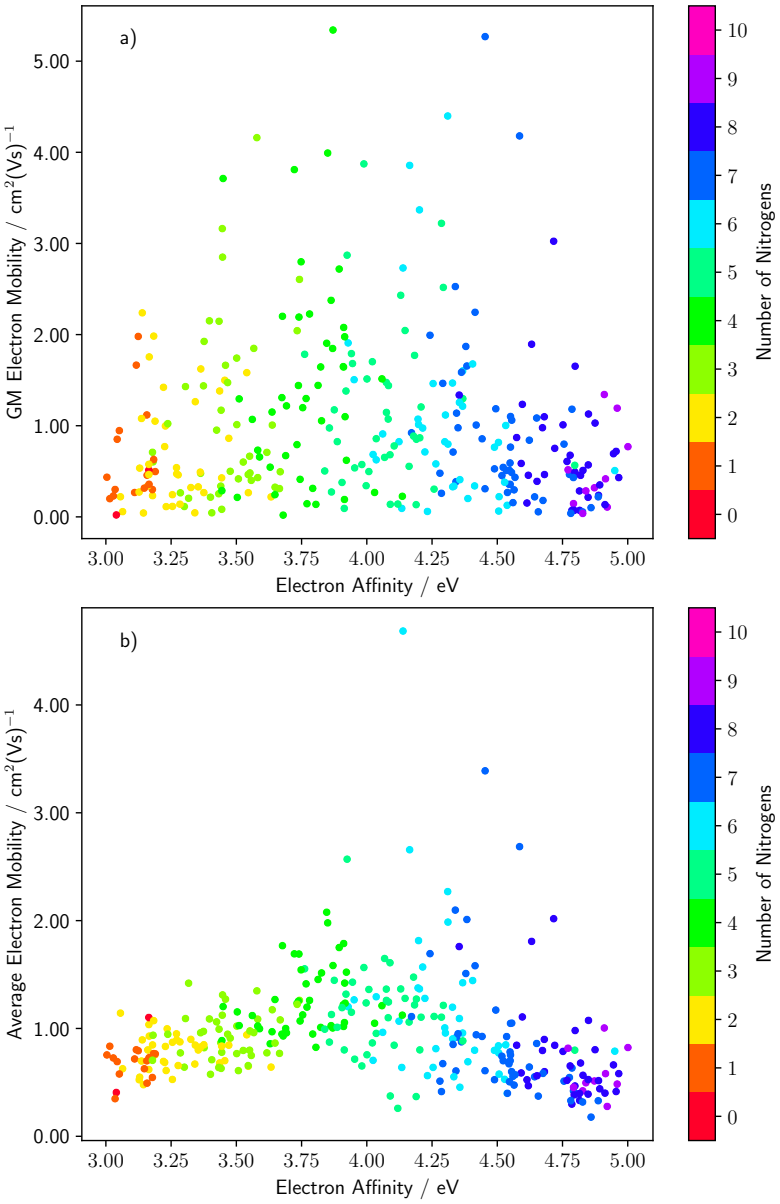


Figure 8.8: Property maps of the 320 molecules that were taken through the mobility evaluation stage of the computational screening workflow. Molecules are plotted with a) its electron mobility of the global minimum crystal structure and b) its average electron mobility against its solid-state electron affinities. Plot points are coloured by the number of nitrogen atoms in the molecule.

structure to the distributions of the mobilities, in particular, the electron affinity values in the range between 3.5 to 4.5 eV or equivalently molecules containing 4-7 nitrogen atoms typically contained the highest values for both the global minimum and average electron mobility values.

8.4 Evaluation

We carry out the same CSP and electron mobility evaluations on the experimentally synthesised molecules MonoF-IFO, DiF-IFO, TriF-IFO, **1a**, **2a**, and **2b** and compare the experimental electron mobilities against our CSP results. A summary is given in Table 8.2 which shows the largest experimentally determined electron mobilities, the electron mobility of the CSP global minimum structure, and the CSP average electron mobilities. We can see that there are poor predictions by the CSP and mobility evaluation for the molecules **1a**, which experimentally showed no gate effect and **2b**, which was experimentally determined to have electron mobility much lower than the electron mobility of the global minimum structure and the calculated CSP average electron mobility.

The disagreement for **1a** can be explained by its rather low solid-state electron affinity which we calculate to be around 3.153 eV which could have lead to high injection barriers for the gold contacts that were used. The disagreement for **2a** is more difficult to explain although looking at the spread of the experimental electron mobilities of the OFETs fabricated with the related molecule **2b** in Table 8.1 suggests that perhaps higher experimental electron mobilities for **2a** could have been obtained. The remaining four molecules show good agreement between the largest experimentally determined electron mobilities and CSP average electron mobilities. Overall the results suggest that CSP average electron mobilities provide a good indication of the typical experimental electron mobilities one would obtain despite the poor agreements for **1a** and **2a** molecules.

We compare the top 18 molecules with the highest crystal structure landscape average mobility from our computational screening workflow against CSP and mobility evaluations of the 6 experimentally synthesised molecules. Molecular structures of the top 18 molecules are shown in Figure 8.9, the mobility evaluation results of all 24 molecules are shown in Table 8.3, and ESF maps are shown in

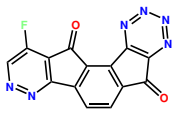
Molecule	Largest Exp Mobility / $cm^2(Vs)^{-1}$	CSP GM Mobility / $cm^2(Vs)^{-1}$	CSP Average Mobility / $cm^2(Vs)^{-1}$
MonoF-IFO	0.17	0.512	0.307
DiF-IFO	0.07	0.122	0.195
TriF-IFO	0.16	0.093	0.201
1a	-	1.341	0.635
2a	2.7×10^{-4}	0.689	0.215
2b	0.17	0.753	0.415

Table 8.2: The largest experimental reported electron mobilities [306,307], the electron mobility of the CSP global minimum structure, and the CSP average electron mobilities calculated in this work for the six IFO derivatives, the experimental OFET fabricated using **1a** showed no gate effect.

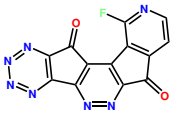
Appendix F. The only performance metric the experimental molecules consistently perform better is their lower deviations which appear to only have occurred because of its consistently lower electron mobilities across the crystal structure landscape.

For the remaining performance metrics, the top EA generated molecules had generally outperformed the experimentally synthesised molecules. Although there were a few examples where the global minimum predicted structure of the experimental molecule **1a** resulted in higher electron mobilities than the equivalent global minimum predicted structures of the EA proposed molecules, overall we see much larger electron mobilities in the global minimum structure of the EA proposed molecules. The crystal structure landscape average mobilities of the EA molecules show around a 3 to 24 times increase over the average mobility of the experimentally synthesised molecules. The electron affinities were also typically higher with EA generated molecules having up to 1.56 eV larger electron affinities. With both these performance metrics and assuming that these molecules can be synthesised and are air-stable, performances of these molecules as an n-type organic semiconductor are expected to be better than the experimentally synthesised molecules since the EA proposed molecules are predicted to have the combination of higher electron mobilities and lower electron injections barriers when combined with higher work function contacts.

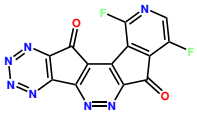
Looking at the 18 best molecules from our computational screening procedure



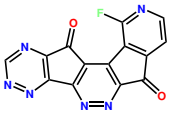
1: 4.676



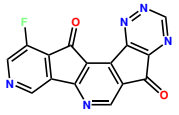
2†: 3.385



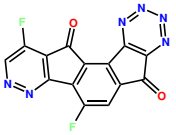
3†: 2.684



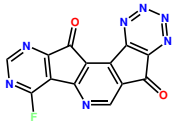
4†: 2.655



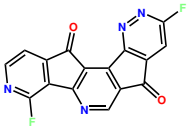
5: 2.568



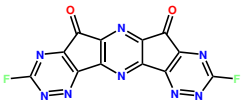
6: 2.269



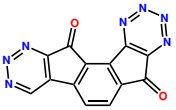
7: 2.098



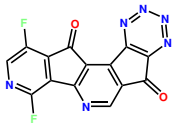
8: 2.078



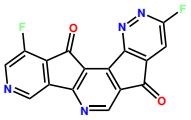
9: 2.018



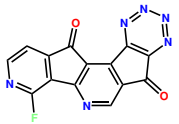
10: 2.011



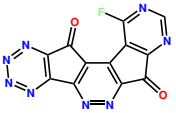
11: 1.987



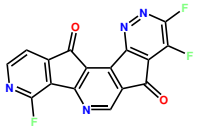
12: 1.980



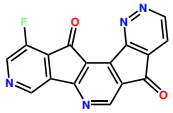
13: 1.816



14†: 1.809

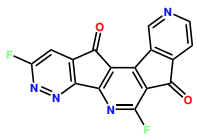


15: 1.789



16: 1.769

17: 1.761



18: 1.752

Figure 8.9: Molecular structures of the top 18 molecules with the largest crystal structure landscape average electron mobilities out of all 320 molecules taken through the mobility evaluations stage of the computational screening workflow, molecular labels (†chiral molecules) and average mobilities in $cm^2(Vs)^{-1}$ are written below.

on a search space of substituted IFOs, we see that the [1,2-a] IFO derivatives were typically the best performing. Although most of this will be attributed to the selection for the lowest reorganisations energy molecules which had resulted in [1,2-a] IFO derivatives being sampled the most. Out of the top 18 molecules, a total of 16 were [1,2-a] IFO derivatives and 2 were [2,1-b] IFO derivatives. Four [1,2-a] IFO derivatives were found to have non-planar helically-shaped chiral structures due to the fluorine substitution in the fjord region which resulted in a collision between the fluorine and oxygens atoms and a breaking of its planar geometry to relieve the collision.

Molecule **1** has the largest average electron mobility with a value of $4.676\,\mathrm{cm}^2(\mathrm{Vs})^{-1}$ and represents a strong candidate for further evaluations. Although it does have the largest deviation $\langle\Delta\bar{\mu}^2\rangle^{1/2}$ in the electron mobility with a value of $3.519\,\mathrm{cm}^2(\mathrm{Vs})^{-1}$ so that the large average is also coupled with quite a large uncertainty. The next three molecules **2**, **3**, and **4** have lower average mobilities but are competitive to **1** due to their lower deviations. For low-risk molecules **10** and **15** are the best candidates with deviations in the electron mobility of around $0.55\,\mathrm{cm}^2(\mathrm{Vs})^{-1}$. Electron affinities are more or less the same across all EA molecules so should all give similarly small electron injection barriers. The only exception is perhaps **16** which has the lowest calculated electron affinity with a value of 3.678 eV so that the other molecules may be preferable if electron injection barriers and therefore contact resistances are predicted to be an issue.

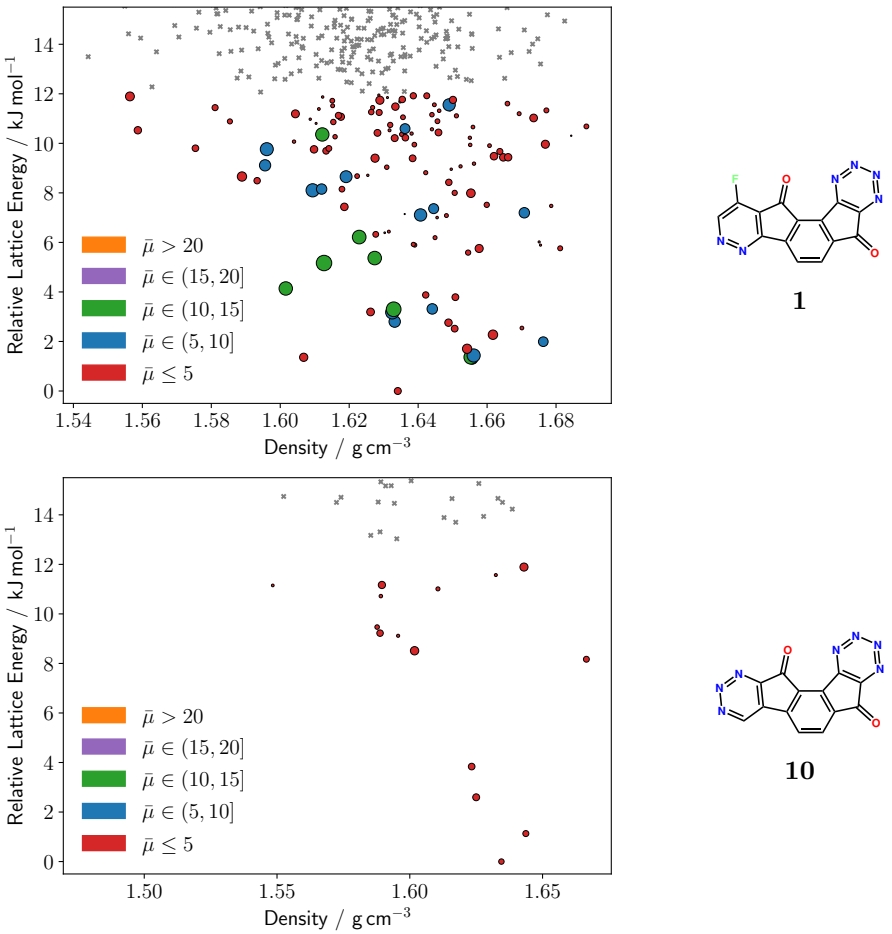


Figure 8.10: ESF map of the molecule **1** which has the largest average electron mobility out of all 320 IFO derivatives evaluated and the ESF map of the molecule **10** which represents a lower risk option with its relatively high average and low deviation due to its sparse crystal structure landscape. Electron mobilities were calculated for all structures within 12 kJ mol^{-1} from the global minimum, plot points are scaled and coloured by their calculated mobilities. Ranges of the mobilities defined in the ESF map legends are given in units of cm^2(Vs)$^{-1}$.

Molecule	Number of Structures	$\bar{\mu}_{GM}$ / $cm^2(Vs)^{-1}$	$\langle\bar{\mu}\rangle$ / $cm^2(Vs)^{-1}$	$\langle\Delta\bar{\mu}^2\rangle^{1/2}$ / $cm^2(Vs)^{-1}$	Λ / eV	A_s / eV
1	130	2.735	4.676	3.519	0.244	4.138
2	35	5.269	3.385	1.753	0.216	4.452
3	26	4.181	2.684	1.469	0.232	4.583
4	19	3.859	2.655	1.367	0.242	4.164
5	139	2.875	2.568	2.908	0.248	3.925
6	174	4.399	2.269	1.655	0.253	4.309
7	80	2.531	2.098	1.070	0.227	4.338
8	105	1.910	2.078	1.111	0.238	3.847
9	41	3.028	2.018	1.130	0.249	4.714
10	15	1.661	2.011	0.554	0.247	4.382
11	76	0.805	1.987	1.651	0.226	4.310
12	132	3.994	1.980	1.282	0.245	3.851
13	403	1.020	1.816	1.699	0.221	4.198
14	86	1.900	1.809	1.408	0.225	4.629
15	72	2.083	1.789	0.559	0.248	3.912
16	385	2.205	1.769	1.188	0.238	3.678
17	148	1.343	1.761	1.268	0.195	4.353
18	79	2.723	1.752	1.06	0.247	3.894
MonoF-IFO	225	0.512	0.307	0.275	0.323	3.422
DiF-IFO	146	0.122	0.195	0.306	0.339	3.537
TriF-IFO	305	0.093	0.201	0.171	0.347	3.753
1a	209	1.341	0.635	0.396	0.305	3.153
2a	415	0.689	0.215	0.176	0.367	3.475
2b	62	0.753	0.415	0.287	0.385	3.752

Table 8.3: Electron mobility statistics of the crystal structure landscapes and electronic properties of the 18 best molecules out of all 320 molecules taken through the mobility evaluations stage of the computational screening workflow and 6 experimental molecules which had previously been used in the fabrication of n-type OFETs.

8.5 Conclusion

In this chapter, we expanded all stages of the computational screening of small molecular organic semiconductors when compared to the work in Chapter 6. Most notable was the expansion of the CSP and mobility evaluations which saw an increase in the number of molecules evaluated at this stage by an order of magnitude. The main advantage we had over the work in Chapter 6 was the application of the AOM for calculating electronic couplings using the pyAOMlite library. Previously electronic coupling calculations had been the rate-determining step in our computational screening workflow while in this chapter the CSP stages had now become the rate-determining step. This change had allowed us to use the reduced sampling scheme which included crystal structures with $Z' > 1$ which demanded a greater number of electronic coupling calculations per structure.

By expanding the CSP and electron mobility evaluations stages we were able to generate the property maps in Figures 8.7 and 8.8 which included crystal structure landscape based properties in addition to property maps in Figures 8.3 and 8.4 which were based only on molecular properties. This is an improvement over the computational screening workflow in Chapter 6 which could only generate property maps and therefore suggest design rules based on molecular properties. Using the property maps generated in this chapter we can therefore suggest a more useful set of design rules for n-type OFETs with IFO derivative semiconductors based on both molecular and crystal structure landscape average properties

- IFO derivatives with 4-5 nitrogen atoms are required to obtain solid-state electron affinities of around 4 eV and the highest electron mobilities.
- To increase electron affinities nitrogen substitutions are preferred over fluorine substitutions.
- The [1,2-a] IFO derivatives are preferred over the other isomers due to their typically lower reorganisations energies and higher crystal structure landscape average electron mobilities.

we have chosen an electron affinity of around 4 eV to lower electron injection barriers with high work function contacts while avoiding highly electrophilic molecules which may not be stable.

The property maps also suggest the existence of a relationship between the molecular structure and its crystal structure landscape average electron mobilities which showed that the highest mobilities were obtained for molecules containing 4-7 nitrogen atoms. This intermediate number of nitrogens makes sense since there would need to be both a number of C-H and N groups to create the C-H$\cdots$N interactions and perhaps more sheet-like packing which could improve electron mobilities. This structure to property relationship suggests that an EA with a fitness function based on a crystal structure landscape average mobility could be a viable strategy for the computational screening of small molecule organic semiconductors. A relatively clear structure to property relationship should allow an EA to optimise the molecular structure space more easily. However, this type of fitness function would be very computational expensive unless a particularly short sampling scheme is used.

In this chapter, we included Z'=2 searches in our CSP and evaluated electron mobilities of structures up to 12 kJ mol^{-1} rather than the 7 kJ mol^{-1} cut-off used in Chapter 6. We used a higher energy cut-off as the results in Chapter 7 showed matches with the experimental for predicted structures up to 9.77 kJ mol^{-1} from the global minimum. It is therefore not surprising that the number of structures evaluated for the top 18 IFO derivatives was generally greater than the nitrogen substituted PAHs from Chapter 6. However, a similar result was found which showed that certain molecules appear to produce sparse crystal structure landscapes for example the IFO derivative **10** with only 15 crystal structures in the low energy region. Like in the conclusion of Chapter 6 it appears that within the search space we had used there are a number of low-risk molecules that are associated with only a small number of crystal structures which could be good candidates for further more computationally expensive evaluations.

Chapter 9

Conclusion

This thesis describes the work we have made in the computational screening of small-molecule organic semiconductors. We have taken a number of steps towards the development of a computational screening workflow with the aim to have as few restrictions on the search space as possible. There is no doubt however that the computational screening carried out in this work describes only a fraction of an ideal workflow due to the computational cost and approximate nature of the computational methods used. Additionally, the lack of efficient black-boxed approaches to flexible CSP methods has placed substantial restrictions on our search space to only rigid molecules, limiting ourselves to a small subset of experimentally realisable organic semiconductors.

Other difficulties with the approach occur due to the problems associated with CSP in general which produce far more predicted structures than what is known experimentally ensuring that there is always a great deal of uncertainty with the results due to a number of computational and experimental factors for the discrepancy [308]. In our case, the number of predicted polymorphs and therefore the uncertainty of the experimental polymorph may have also been inflated by the methods we have used to remove duplicated structures which used quite tight tolerances when comparing simulated X-ray powder patterns. Further work to benchmark a variety of duplicate removal methods and tolerances could be useful for the reduction of the number of unique structures and therefore the uncertainty of the results.

Another problematic issue was the reliability of the carrier mobility evaluations method that was used. We carried out some benchmarking work in Section 7.4 using Marcus theory and TLT which gave results that were somewhat unsatisfactory with poor agreements with FOB-SH for the GUMZIY and GUMZUK crystal structures. Although how precise the agreements should be is not completely clear since we know that carrier mobilities can vary over several orders of magnitude. Further work to determine why both good or poor agreements are obtained between quite different computational methods and further benchmark work on alternative methods are necessary. Once these reasons are known and agreements are obtained for a number of computational methods, benchmark work with experimental results should then be made to improve our confidence with the carrier mobility results.

It is therefore clear that further work will be necessary which goes beyond the work we have described with improved methods and further evaluations stages. Despite the shortcomings of our computational screening workflow, we believe we have made some important advances and provide a solid foundation for future work to be built on. Of particular importance was the development of the reduced sampling scheme in Chapter 7 which allowed for efficient and reliable CSPs to be used in a computational screening workflow which allowed for the application of CSP to a large number of molecules. For example in Chapter 8 we had run CSP searches on a total of 320 EA proposed IFO derivatives and then together with the AOM evaluated the electron mobilities for a huge number of predicted crystal structures.

Another useful result from our computational screening workflow is its ability to form property maps of the search space as shown in Chapters 6 and 8. By interpreting the maps a set of search space specific design rules can be formed for example with the nitrogen substituted PAHs, linear structures are preferred due to their lower reorganisation energies. Given that no comments are made on either the synthesizability or stability of the proposed molecules at the end of our computational screening workflow, these property maps may therefore be more useful to a synthetic chemist as it will provide them with some basic design rules but allow some freedom to design and synthesize the most stable molecules. A clear next step in the development of the computational screening

workflow could therefore be to incorporate some information on the molecules synthesizability and stability into the fitness function of the EA.

Another possible next step in the development of the computational screening workflow could be to include the average mobility of the crystal structure landscape into the fitness function. This would require some CSP to be carried out during EA searches which will be difficult given the computational cost of CSP even with the reduced sampling schemes we have developed. Perhaps a minimal sampling scheme could be made to reproduce the general trends for a test set of molecules for a given average property of the crystal structure landscape. Alternatively, it may be sufficient to have a fitness function using a minimal sampling scheme that outperforms the isolated molecule-based fitness functions at suggesting molecules for the subsequent evaluation stages. A possible computational screening workflow could therefore have an EA with the average mobility using the minimal sampling scheme as its fitness function and a second stage that reevaluates the best molecules using the reduced sampling scheme.

Appendices

Appendix A

Current Density

In Subsection 3.5.1 we stated that the time derivative of the charge polarisation is equivalent to an integration of the current density. So that the time derivative of the charge polarisation is a sum of the current densities

$$j_{\mathrm{el},x}(t) = \frac{\partial}{\partial t}\langle\psi(\mathbf{r}_\mathrm{e},t)|\hat{\mu}_x(\mathbf{r}_\mathrm{e})|\psi(\mathbf{r}_\mathrm{e},t)\rangle = \frac{\partial}{\partial t}\int \mathrm{d}^3\mathbf{r}\, x\, n_\mathrm{el}(\mathbf{r},t) \tag{A.1}$$

where we have written only the x component of the current density to make the notation more clear in the subsequent steps. From the continuity equation, we have an expression for the time derivative of the charge density

$$\frac{\partial n_\mathrm{el}(\mathbf{r},t)}{\partial t} = -\boldsymbol{\nabla}\cdot\mathbf{j}_\mathrm{el}(\mathbf{r},t) \tag{A.2}$$

which we can insert into the rightmost term of Eq. (A.1) to give.

$$j_{\mathrm{el},x}(t) = -\int \mathrm{d}^3\mathbf{r}\, x\boldsymbol{\nabla}\cdot\mathbf{j}_\mathrm{el}(\mathbf{r},t) \tag{A.3}$$

From the product rule for divergence, we have the expression

$$\boldsymbol{\nabla}\cdot[x\,\mathbf{j}_\mathrm{el}(\mathbf{r},t)] = x\boldsymbol{\nabla}\cdot\mathbf{j}_\mathrm{el}(\mathbf{r},t) + (\boldsymbol{\nabla}x)\cdot\mathbf{j}_\mathrm{el}(\mathbf{r},t) \tag{A.4}$$

where

$$(\boldsymbol{\nabla} x) \cdot \mathbf{j}_{\mathrm{el}}(\mathbf{r}, t) = j_{\mathrm{el},x}(\mathbf{r}, t) \tag{A.5}$$

and by integrating all terms over a volume Ω we obtain.

$$\int_{\Omega} \mathrm{d}^3\mathbf{r}\, \boldsymbol{\nabla} \cdot [x\, \mathbf{j}_{\mathrm{el}}(\mathbf{r}, t)] = \int_{\Omega} \mathrm{d}^3\mathbf{r}\, x \boldsymbol{\nabla} \cdot \mathbf{j}_{\mathrm{el}}(\mathbf{r}, t) + \int_{\Omega} \mathrm{d}^3\mathbf{r}\, j_{\mathrm{el},x}(\mathbf{r}, t) \tag{A.6}$$

With the divergence theorem, we can express the leftmost term as a surface integral over the boundary Γ resulting in the equation.

$$\int_{\Gamma} \mathrm{d}\mathbf{s}\, x\, \mathbf{j}_{\mathrm{el}}(\mathbf{r}, t) \cdot \hat{\mathbf{n}} = \int_{\Omega} \mathrm{d}^3\mathbf{r}\, x \boldsymbol{\nabla} \cdot \mathbf{j}_{\mathrm{el}}(\mathbf{r}, t) + \int_{\Omega} \mathrm{d}^3\mathbf{r}\, j_{\mathrm{el},x}(\mathbf{r}, t) \tag{A.7}$$

For crystal structures we want to carry out the volume integrals over the entire structure, similar to the situation with the energy calculations of charged particles in a crystal structure we neglect the surface term to obtain the bulk property so that

$$\int \mathrm{d}^3\mathbf{r}\, j_{\mathrm{el},x}(\mathbf{r}, t) = -\int \mathrm{d}^3\mathbf{r}\, x \boldsymbol{\nabla} \cdot \mathbf{j}_{\mathrm{el}}(\mathbf{r}, t) \tag{A.8}$$

and together with Eq. (A.1) and Eq. (A.3) we obtain the equation

$$j_{\mathrm{el},x}(t) = \frac{\partial}{\partial t} \langle \psi(\mathbf{r}_{\mathrm{e}}, t) | \hat{\mu}_x(\mathbf{r}_{\mathrm{e}}) | \psi(\mathbf{r}_{\mathrm{e}}, t) \rangle = \int \mathrm{d}^3\mathbf{r}\, j_{\mathrm{el},x}(\mathbf{r}, t) \tag{A.9}$$

as stated in Subsection 3.5.1.

Appendix B

TLT Parameters

In this chapter, we will derive a few equations to calculate the TLT parameters for this specific Hamiltonian

$$\begin{aligned}\hat{\mathcal{H}} =& \sum_n E'_n|n\rangle\langle n| + \sum_{n,n'\neq n} V_{nn'}|n\rangle\langle n'| + \sum_{n,\alpha}\hbar\omega_\alpha V_{n\alpha}(\hat{a}^\dagger_\alpha + \hat{a}_\alpha)|n\rangle\langle n| \\ &+ \sum_{n,\beta}\hbar\omega_\beta V_{n\beta}(\hat{a}^\dagger_\beta + \hat{a}_\beta)|n\rangle\langle n| \\ &+ \sum_{n,n'\neq n,\gamma}\hbar\omega_\gamma V_{nn'\gamma}(\hat{a}^\dagger_\gamma + \hat{a}_\gamma)|n\rangle\langle n'| + \hat{H}_{\text{B}}\end{aligned} \tag{B.1}$$

where we have separated the vibrational modes due into the low-frequency intramolecular vibrational modes α, the high-frequency intramolecular vibrational modes β, and the low-frequency rigid molecule intermolecular vibrational modes γ. We assume that there are no vibrational modes that occur in both the local and nonlocal interaction terms. We start by carrying out the polaron transformation on the high-frequency intramolecular vibrational modes using

$$U = \exp\left[\sum_{n\beta} V_{n\beta}(\hat{a}_\beta - \hat{a}^\dagger_\beta)|n\rangle\langle n|\right] \tag{B.2}$$

so that

$$\begin{aligned} U\mathcal{H}U^{-1} = &\sum_{n} E_n|n\rangle\langle n| + \sum_{n,n'\neq n} V_{nn'}e^{\Omega_{nn'}}|n\rangle\langle n'| \\ &+ \sum_{n,\alpha} \hbar\omega_\alpha V_{n\alpha}(\hat{a}^\dagger_\alpha + \hat{a}_\alpha)|n\rangle\langle n| \\ &+ \sum_{n,n'\neq n,\gamma} \hbar\omega_\gamma V_{nn'\beta}(\hat{a}^\dagger_\gamma + \hat{a}_\gamma)e^{\Omega_{nn'}}|n\rangle\langle n'| + \hat{H}_{\text{B}} \end{aligned} \tag{B.3}$$

where.

$$\begin{aligned} &\hat{\Omega}_{nn'} = \hat{\Omega}_n - \hat{\Omega}_{n'} \qquad\qquad \hat{\Omega}_n = \sum_\beta V_{n\beta}(\hat{a}^\dagger_\beta - \hat{a}_\beta) \\ &E_n = E'_n - \sum_\beta \hbar\omega_\beta V^2_{n\beta} \end{aligned} \tag{B.4}$$

Now we thermally averaged over all vibrational modes to obtain the parameter for the TLT calculations so the thermally averaged site energy and the variances are

$$\mu_n = \langle E_n + \sum_\alpha \hbar\omega_\alpha V_{n\alpha}(\hat{a}^\dagger_\alpha + \hat{a}_\alpha)\rangle_{\text{B}} = E_n \tag{B.5}$$

$$\begin{aligned} \sigma^2_n &= \langle (E_n + \sum_\alpha \hbar\omega_\alpha V_{n\alpha}(\hat{a}^\dagger_\alpha + \hat{a}_\alpha) - \mu_n)^2\rangle_{\text{B}} \\ &= \sum_\alpha \hbar^2\omega^2_\alpha V^2_{n\alpha} \coth\left(\frac{\hbar\omega_\alpha}{2k_{\text{B}}T}\right) \end{aligned} \tag{B.6}$$

and the thermally averaged electronic coupling and variances are

$$\begin{aligned}\mu_{nn'} &= \langle V_{nn'} e^{\Omega_{nn'}} + \sum_{\gamma} \hbar\omega_{\gamma} V_{nn'\gamma} (\hat{a}_{\gamma}^{\dagger} + \hat{a}_{\gamma}) e^{\Omega_{nn'}} \rangle_{\mathrm{B}} \\ &= V_{nn'} \exp\left[-\frac{1}{2} \sum_{\beta} (V_{n\beta} - V_{n'\beta})^2 \coth\left(\frac{\hbar\omega_{\beta}}{2k_{\mathrm{B}}T} \right) \right]\end{aligned} \tag{B.7}$$

$$\begin{aligned}\sigma_{nn'}^2 &= \langle (V_{nn'} e^{\Omega_{nn'}} + \sum_{\gamma} \hbar\omega_{\gamma} V_{nn'\gamma} (\hat{a}_{\gamma}^{\dagger} + \hat{a}_{\gamma}) e^{\Omega_{nn'}} - \mu_{nn'})^2 \rangle_{\mathrm{B}} \\ &= \sum_{\gamma} \hbar^2 \omega_{\gamma}^2 V_{nn'\gamma}^2 \langle \hat{a}_{\gamma}^{\dagger} \hat{a}_{\gamma} + \hat{a}_{\gamma} \hat{a}_{\gamma}^{\dagger} \rangle_{\mathrm{B}} \langle e^{\Omega_{nn'}} e^{\Omega_{nn'}} \rangle_{\mathrm{B}} \\ &= \sum_{\gamma} \left\{ \hbar^2 \omega_{\gamma}^2 V_{nn'\gamma}^2 \coth\left(\frac{\hbar\omega_{\gamma}}{2k_{\mathrm{B}}T} \right) \right. \\ &\qquad \left. \times \exp\left[-\sum_{\beta} (V_{n\beta} - V_{n'\beta})^2 \coth\left(\frac{\hbar\omega_{\beta}}{2k_{\mathrm{B}}T} \right) \right] \right\}\end{aligned} \tag{B.8}$$

where we have used the Bloch identity [92, 104] to evaluate thermal averages over the exponential operators. So the site energies $\tilde{E}_n$ and electronic coupling $\tilde{V}_{nn'}$ of the reference Hamiltonian Eq. (3.188) will be sampled from the normal distributions $\mathcal{N}(\mu_n, \sigma_n^2)$ and $\mathcal{N}(\mu_{nn'}, \sigma_{nn'}^2)$ respectively.

Appendix C

QR Sampling Results

Molecule	Z'	SG	Version I	Version II
Artemisinin	1	$P2_12_12_1$	12,180	11,779
Artemisinin	4	$P1$	589,843	334,840
Quinacridone	1	$P\bar{1}$	31,792	24,460
Quinacridone	1	$P2_1/c$	21,878	17,570
Quinacridone	2	$P\bar{1}$	814,229	318,383
CC1	1	$P1$	12,844	10,480
CC1	1	$P2_1/c$	19,718	21,209

Table C.1: Final Sobol seed number of the CSP searches using two different versions of the QR structure generation methods for each Z' | SG search.

Molecule	Z'	SG	Version I	Version II
Artemisinin	1	$P2_12_12_1$	2,061	1,631
Artemisinin	4	$P1$	535,871	280,067
Quinacridone	1	$P\bar{1}$	20,671	13,237
Quinacridone	1	$P2_1/c$	11,438	7,051
Quinacridone	2	$P\bar{1}$	758,923	264,858
CC1	1	$P1$	828	272
CC1	1	$P2_1/c$	3,986	6,552

Table C.2: The number of rejected QR structures for the CSP searches using two different versions of the QR structure generation methods for each Z' | SG search.

Molecule	Z'	SG	Version I	Version II
Artemisinin	1	$P2_12_12_1$	119	148
Artemisinin	4	$P1$	3,972	4,773
Quinacridone	1	$P\bar{1}$	1,121	1,223
Quinacridone	1	$P2_1/c$	440	519
Quinacridone	2	$P\bar{1}$	5306	3,525
CC1	1	$P1$	2,016	208
CC1	1	$P2_1/c$	5,732	4,657

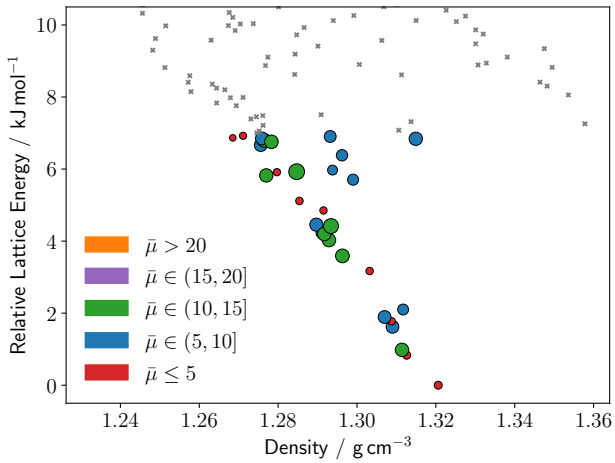
Table C.3: The number of failed minimisations for the CSP searches using two different versions of the QR structure generation methods for each Z' | SG search.

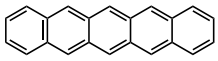
Molecule	Z'	SG	Version I	Version II
Artemisinin	1	$P2_12_12_1$	1h 33m	1h 33m
Artemisinin	4	$P1$	1d 4h 06m	1d 4h 6m
Quinacridone	1	$P\bar{1}$	1h 8m	1h 7m
Quinacridone	1	$P2_1/c$	1h 56m	1h 53m
Quinacridone	2	$P\bar{1}$	18h 07m	17h 54m
CC1	1	$P1$	2h 46m	54m
CC1	1	$P2_1/c$	12h 41m	12h 5m

Table C.4: Approximate total time taken to complete the CSP searches using two different versions of the QR structure generation methods for each Z' | SG search when 4 nodes are used on the Young compute cluster with dual Intel® Xeon® Gold 6248 CPUs @ 2.50GHz on each node.

Appendix D

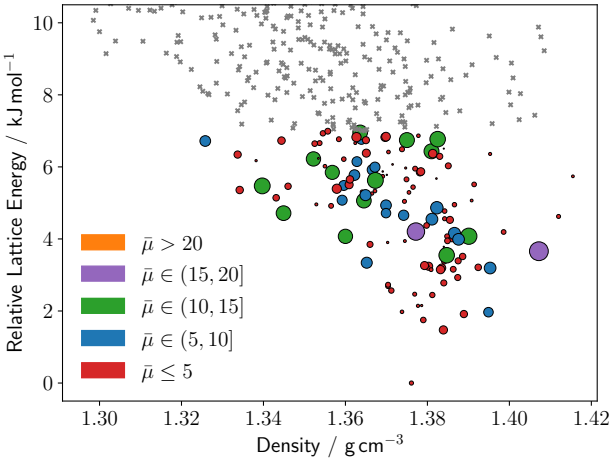
Nitrogen Substituted PAH ESF Maps

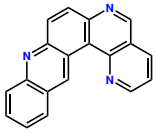




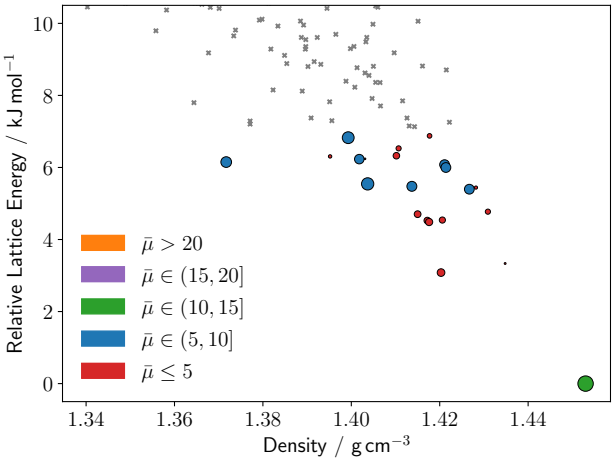
1A

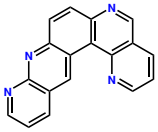
Figure D.1: ESF maps of the 10 best molecules from the EA searches with fitness functions $F_{1,0}$ and $F_{1,4.1}$ as explained in Chapter 6 and the 4 molecules proposed by Winkler and Houk [172]. Electron mobilities were calculated for all structures within 7 kJ mol^{-1} from the global minimum, plot points are scaled and coloured by their calculated electron mobilities. Ranges of the mobilities defined in the ESF map legends are given in units of cm^2(Vs)$^{-1}$. Reproduced from Ref. [143] with permission from the Royal Society of Chemistry.



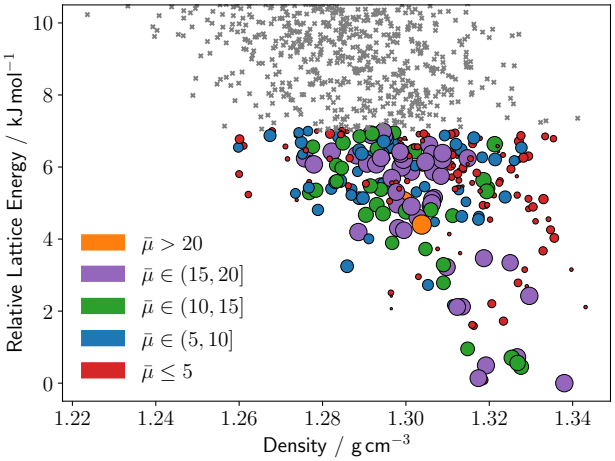


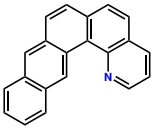
2A





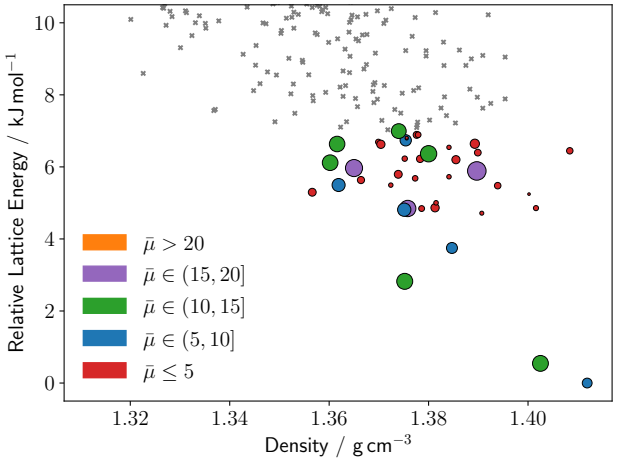
3A

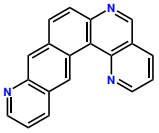




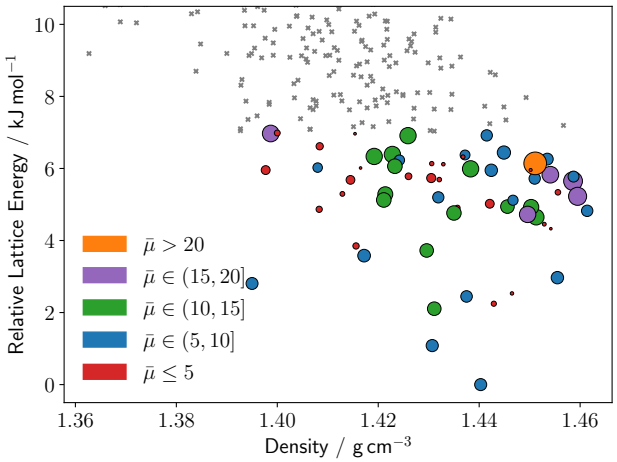
4A

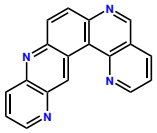
Figure D.1: *continued*



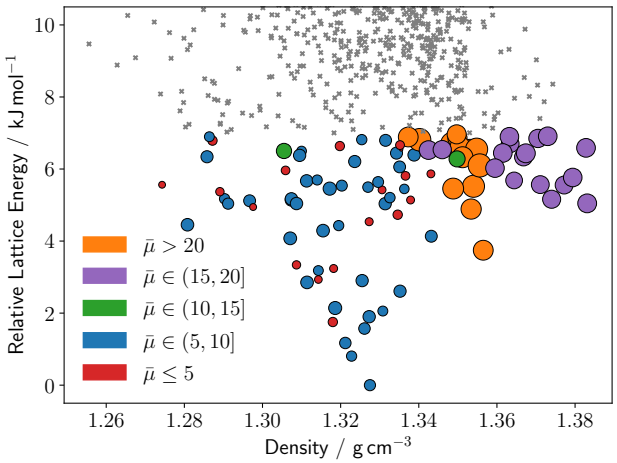


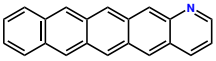
5A





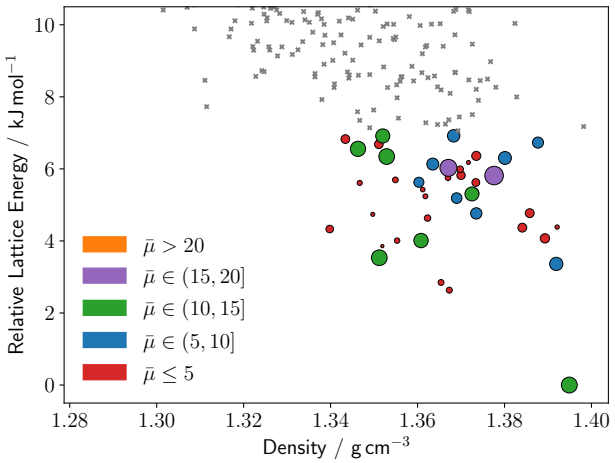
6A

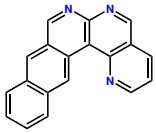




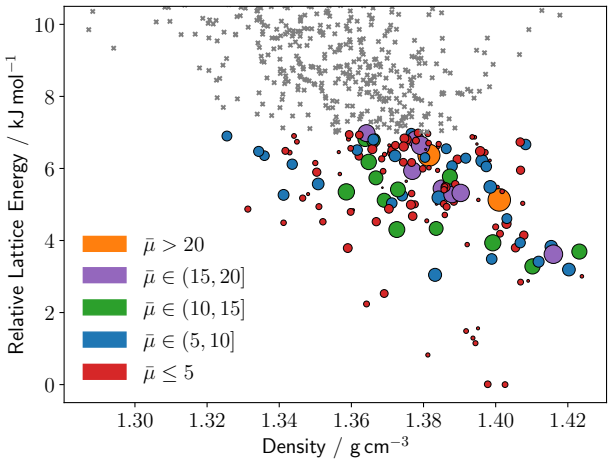
7A

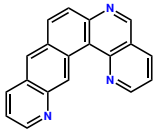
Figure D.1: *continued*



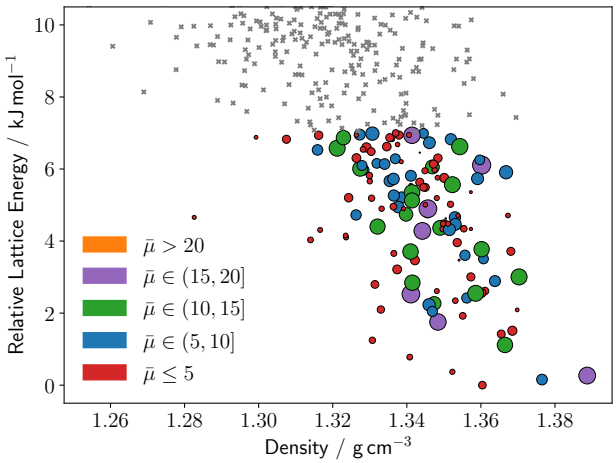


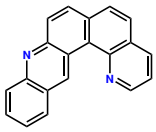
8A





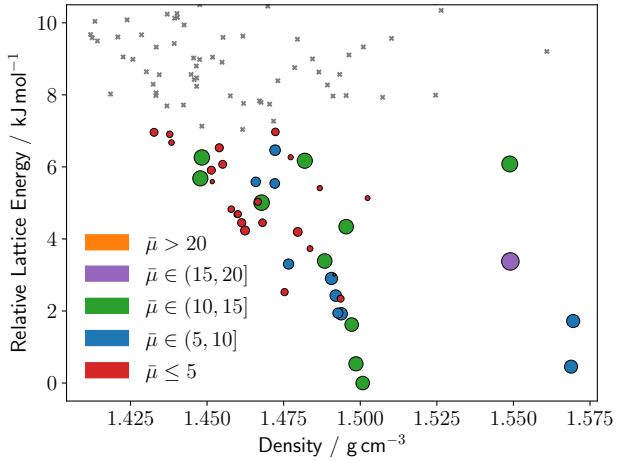
9A

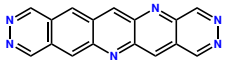




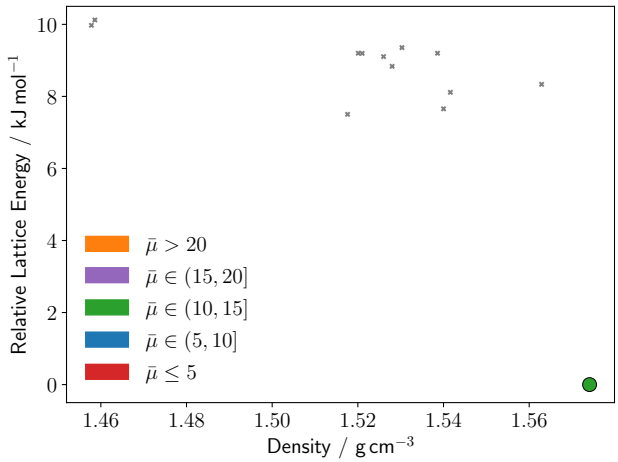
10A

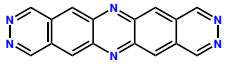
Figure D.1: *continued*



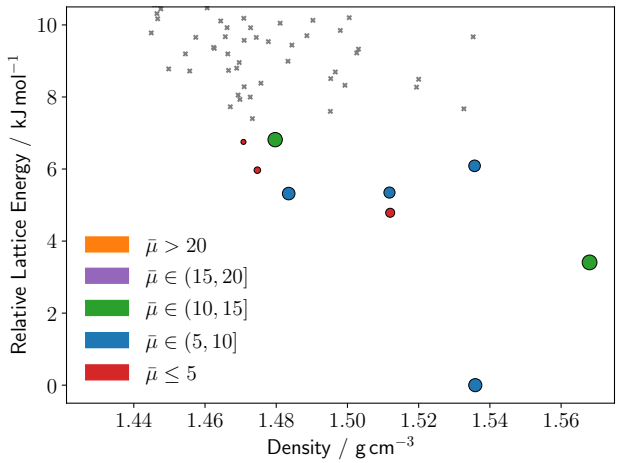


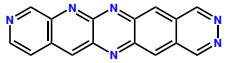
1B





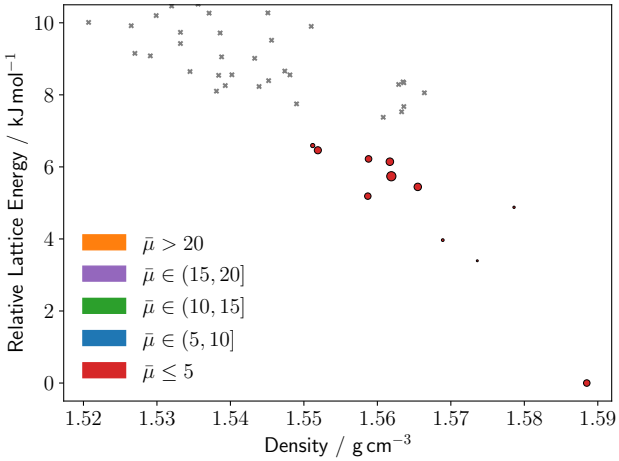
2B

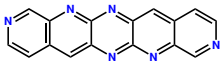




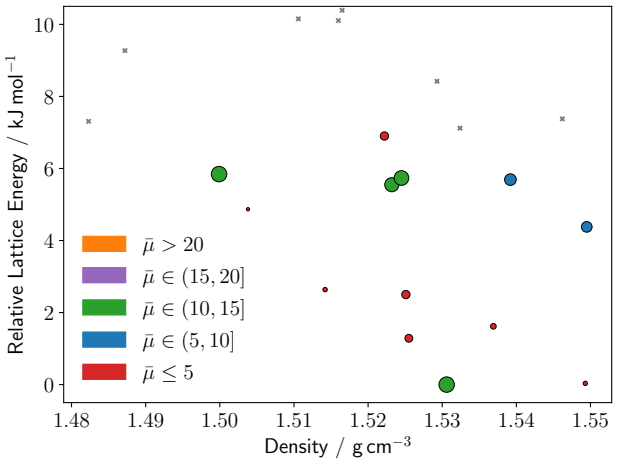
3B

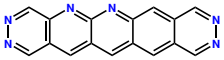
Figure D.1: *continued*



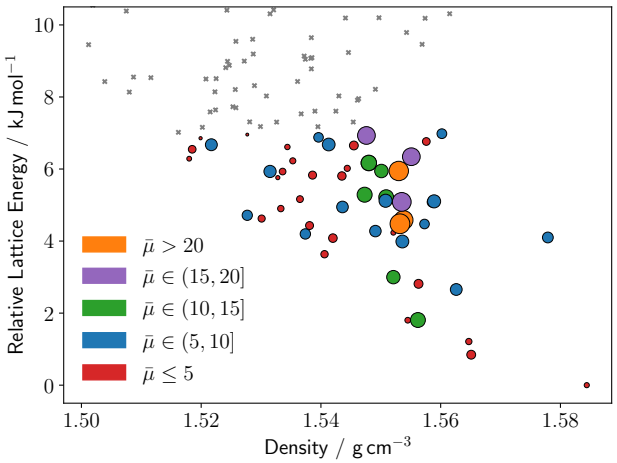


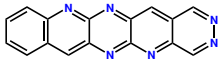
4B





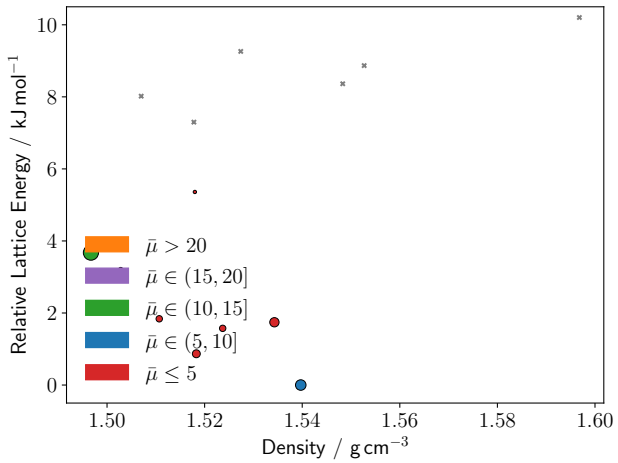
5B

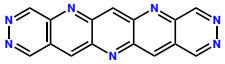




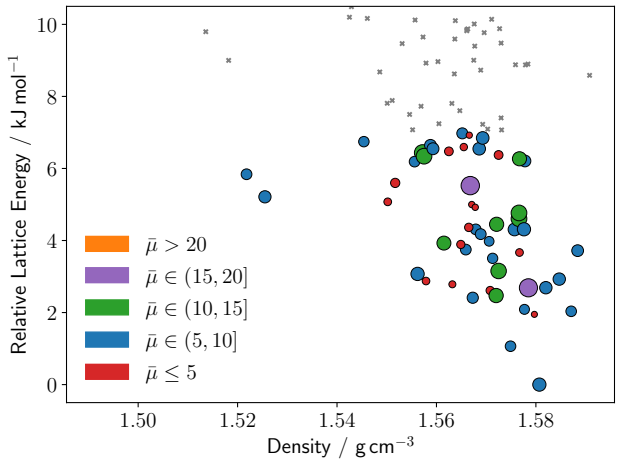
6B

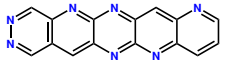
Figure D.1: *continued*



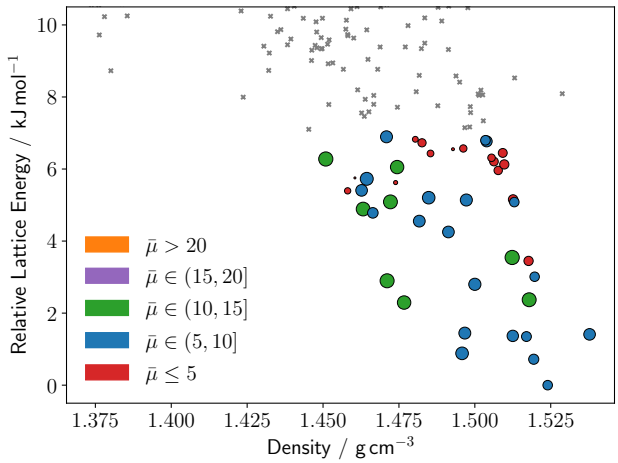


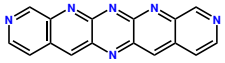
7B





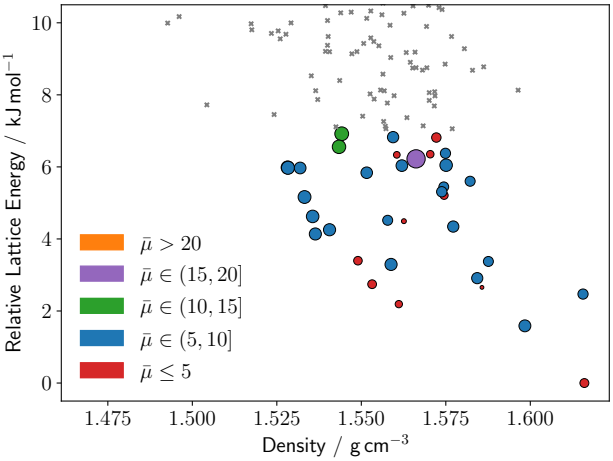
8B

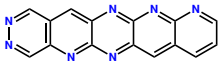




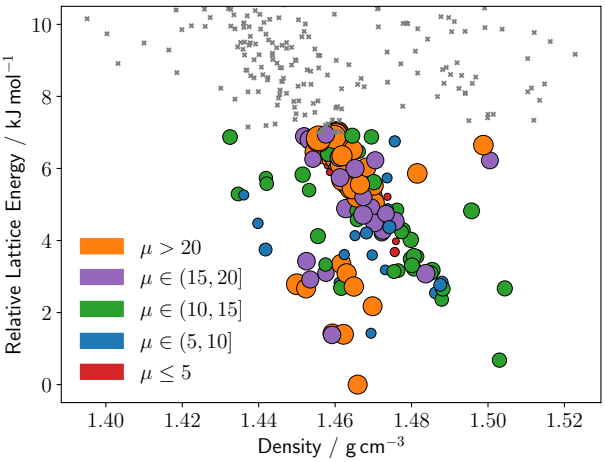
9B

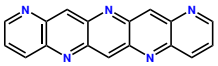
Figure D.1: *continued*



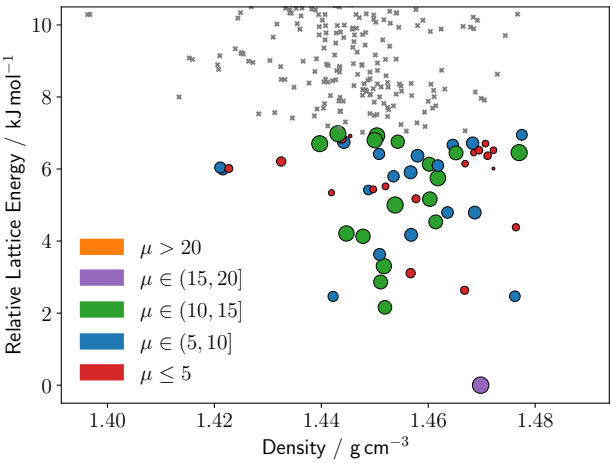


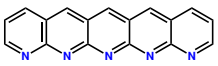
10B





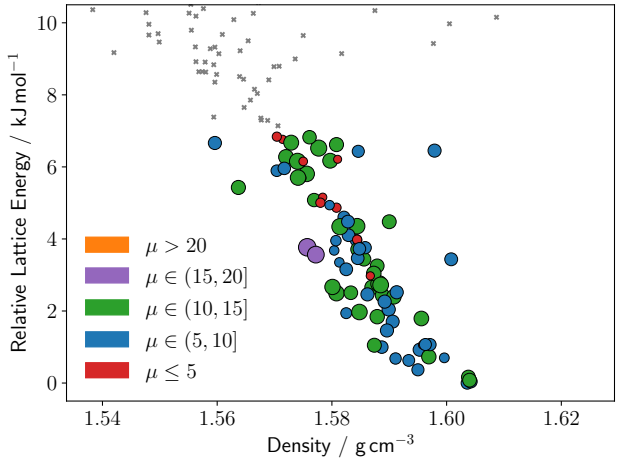
WH5A

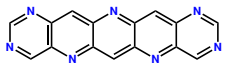




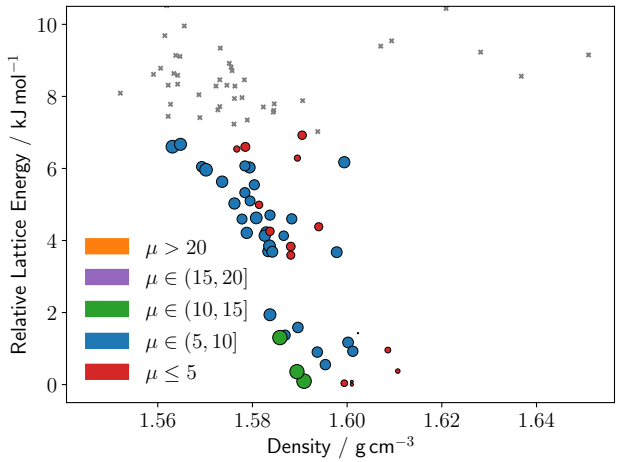
WH5B

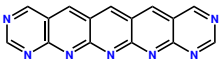
Figure D.1: *continued*





WH7A





WH7B

Figure D.1: *continued*

Appendix E

Computational Screening Benchmark Results

E.1 Molecular Geometries

Molecule	CSD Refcode	RMSD
AROBUFMOL	AROBUF	0.020
AROCAMMOL	AROCAM	0.009
BEGKIKMOL	BEGKIK	0.054
BEZLUOMOL	BEZLUO	0.032
BNPYREMOL	BNPYRE10	0.025
BNPYREMOL	BNPYRE11	0.022
BNPYREMOL	BNPYRE12	0.021
BNZACRMOL	BNZACR	0.030
CEQGELMOL	CEQGEL	0.026
CEQGELMOL	CEQGEL01	0.017
DBANQUMOL	DBANQU	0.285
DBNTHRMOL	DBNTHR01	0.033
DBNTHRMOL	DBNTHR02	0.022
DBNTHRMOL	DBNTHR10	0.061
DUXSASMOL	DUXSAS	0.022

Table E.1: RMSDs in atomic positions with hydrogen positions ignored of the molecular geometries obtained using the RDKit conformer generation followed up with UFF and B3LYP/6-311G** optimisations against the molecular geometries taken from experimental crystal structures.

Molecule	CSD Refcode	RMSD
FANTRQMOL	FANTRQ10	0.030
HAMDUWMOL	HAMDUW	0.057
HAMDUWMOL	HAMDUW01	0.054
HAMDUWMOL	HAMDUW02	0.058
KONBAQMOL	KONBAQ	0.027
LOHVAHMOL	LOHVAH	0.025
MIVRUEMOL	MIVRUE	0.052
MIVSESMOL	MIVSES	0.061
MORRODMOL	MORROD	0.067
NAAZASMOL	NAAZAS	0.023
NIJCEQMOL	NIJCEQ	0.066
OFANTQMOL	OFANTQ	0.325
OFANTQMOL	OFANTQ01	0.331
PAGMERMOL	PAGMER	0.054
PENCENMOL	PENCEN	0.039
PENCENMOL	PENCEN01	0.007
PENCENMOL	PENCEN02	0.015
PENCENMOL	PENCEN03	0.016
PENCENMOL	PENCEN04	0.017
PENCENMOL	PENCEN05	0.022
PENCENMOL	PENCEN06	0.009
PENCENMOL	PENCEN07	0.007
PENCENMOL	PENCEN08	0.014
PENCENMOL	PENCEN09	0.025
PENCENMOL	PENCEN10	0.016
PENCENMOL	PENCEN11	0.016
PENCENMOL	PENCEN12	0.016
PENTQUMOL	PENTQU	0.027
PENTQUMOL	PENTQU01	0.05
PENTQUMOL	PENTQU02	0.029
PERLENMOL	PERLEN01	0.024
PERLENMOL	PERLEN03	0.021
PERLENMOL	PERLEN04	0.021
PERLENMOL	PERLEN05	0.021
PERLENMOL	PERLEN06	0.017
PERLENMOL	PERLEN07	0.006
PERLENMOL	PERLEN08	0.021
POVLUHMOL	POVLUH	0.063
RUNLAPMOL	RUNLAP	0.044
SANQIIMOL	SANQII	0.060

Table E.1: *continued*

Molecule	CSD Refcode	RMSD
SEVYEAMOL	SEVYEA	0.033
UPIRUIMOL	UPIRUI	0.051
UPIXAUMOL	UPIXAU	0.197
UPIXEYMOL	UPIXEY	0.081
VIBMAVMOL	VIBMAV	0.087
VIQTATMOL	VIQTAT	0.064
VUFKALMOL	VUFKAL	0.022
VUFNIWMOL	VUFNIW	0.039
WOTZUDMOL	WOTZUD	0.031
WUPYIQMOL	WUPYIQ	0.057
YOFROBMOL	YOFROB	0.087
YOFROBMOL	YOFROB01	0.092
ZETQUNMOL	ZETQUN	0.022
ZZZOYCMOL	ZZZOYC01	0.056
ZZZOYCMOL	ZZZOYC02	0.048
ZZZOYCMOL	ZZZOYC03	0.037
ZZZOYCMOL	ZZZOYC04	0.062

Table E.1: *continued*

E.2 CSP Sampling I

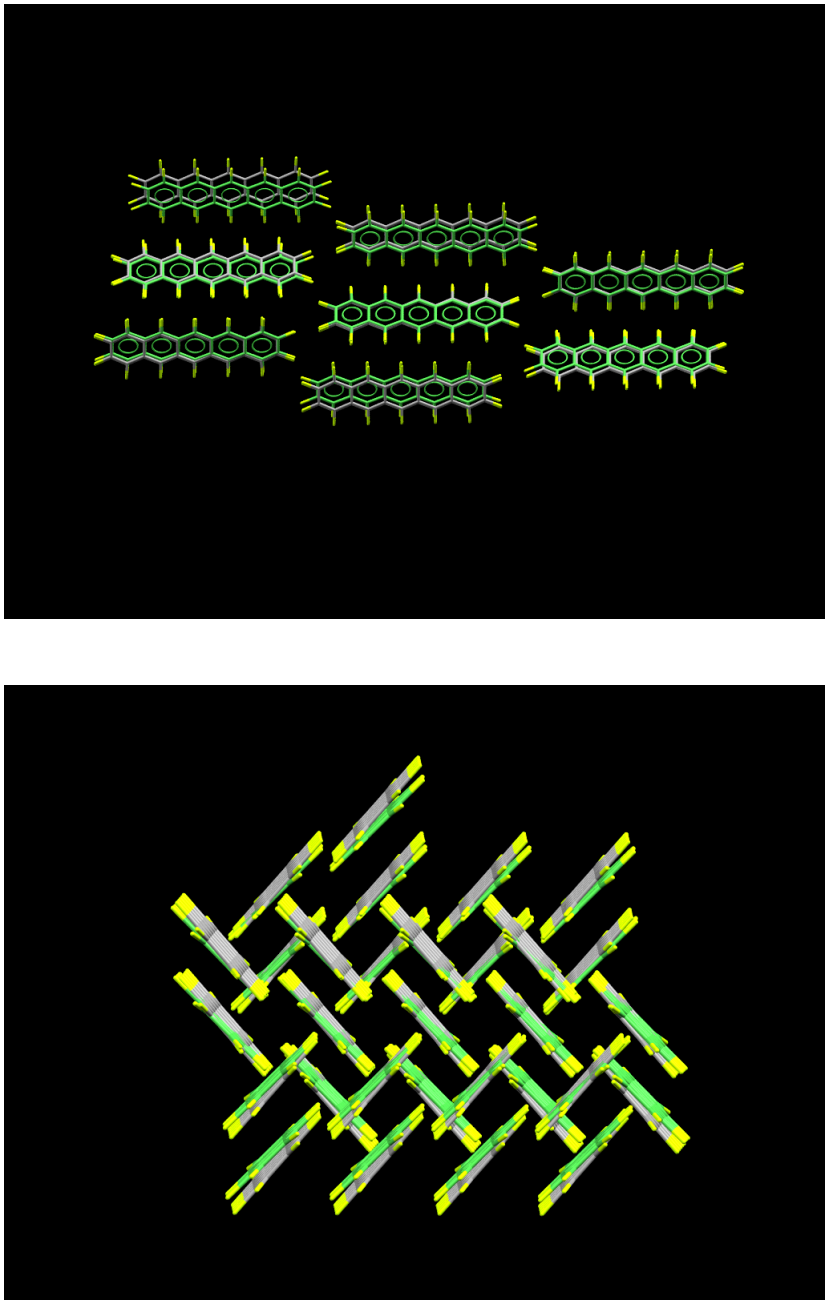


Figure E.1: Two views of the 30 out of 30 COMPACK overlay with an RMSD of 0.327 Å of BEZLUO for the predicted crystal structure with the best match to the experimental and the original experimental crystal structure (green).

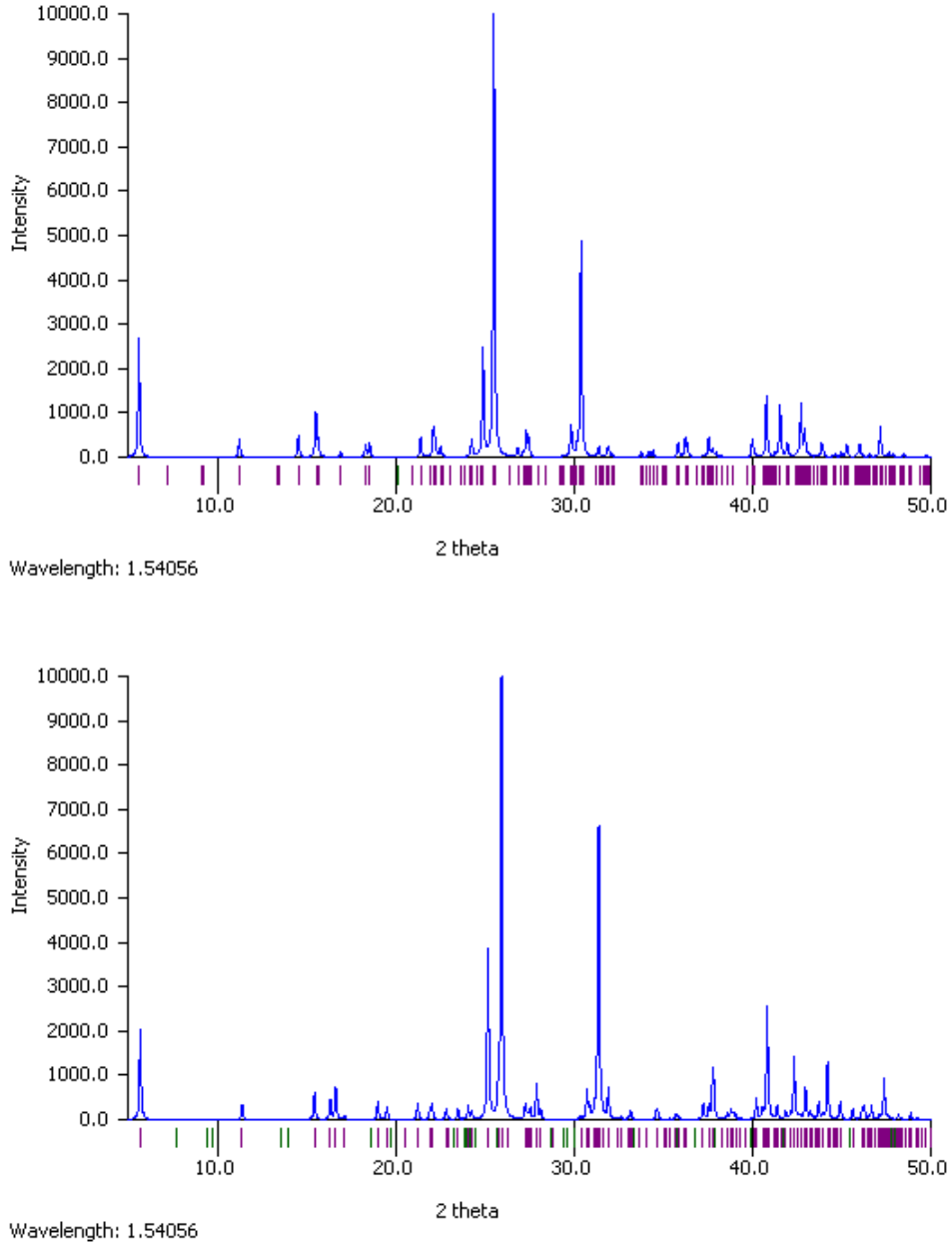


Figure E.2: Simulated PXRD patterns of BEZLUO generated in Mercury for the predicted crystal structure with the best match to the experimental (top) and the original experimental structure (bottom). X-ray powder patterns similarities between these two structures using the CSD Python API resulted in a similarity of 0.931.

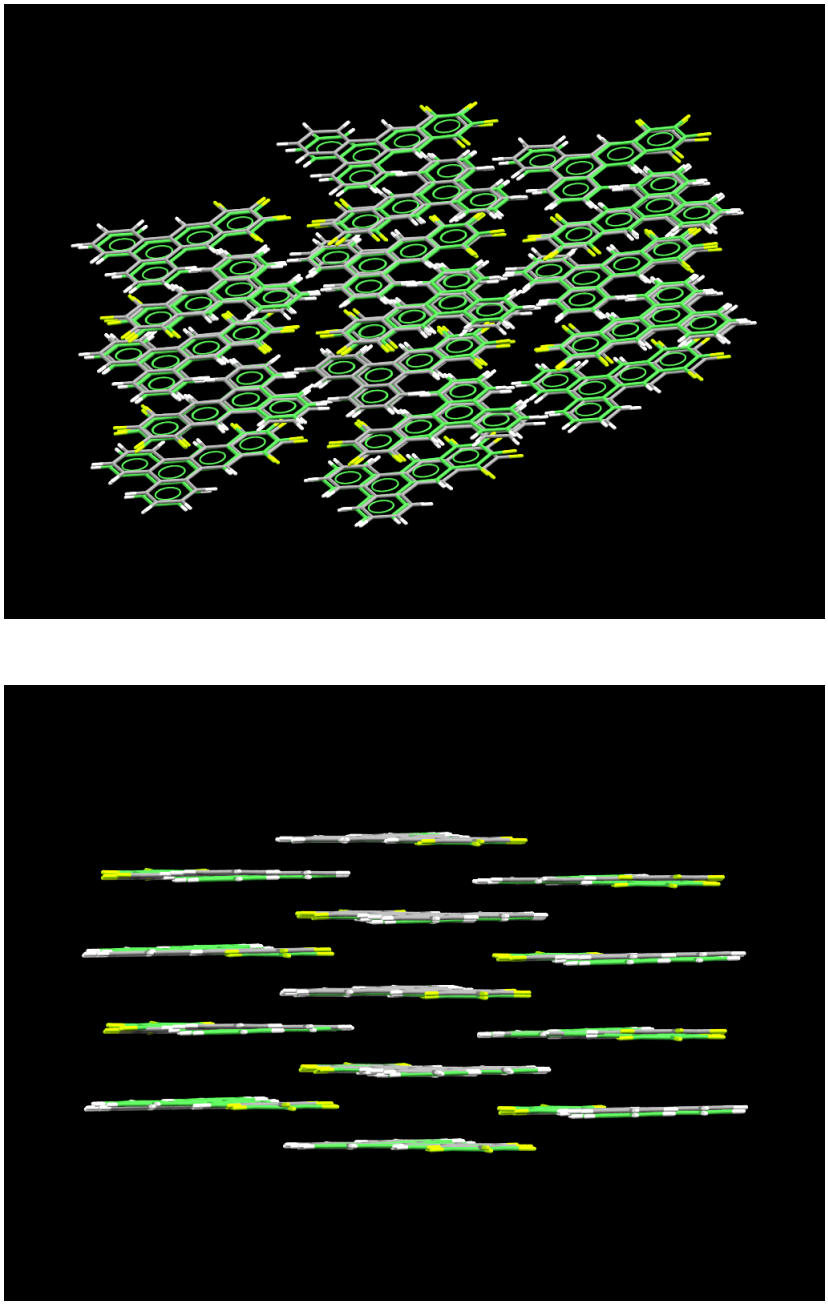
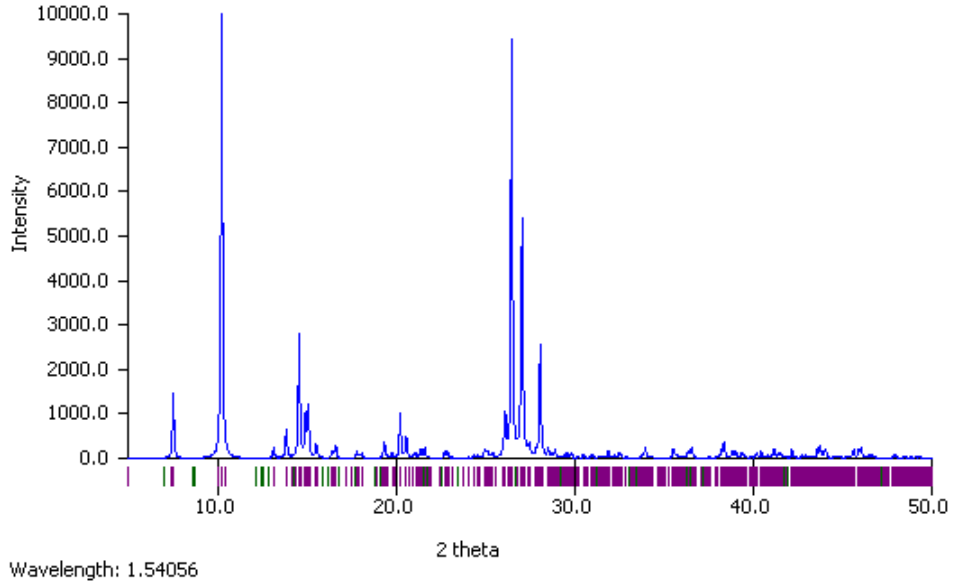


Figure E.3: Two views of the 29 out of 30 COMPACK overlay of DUXSAS for the predicted crystal structure with the best match to the experimental and the original experimental crystal structure (green).



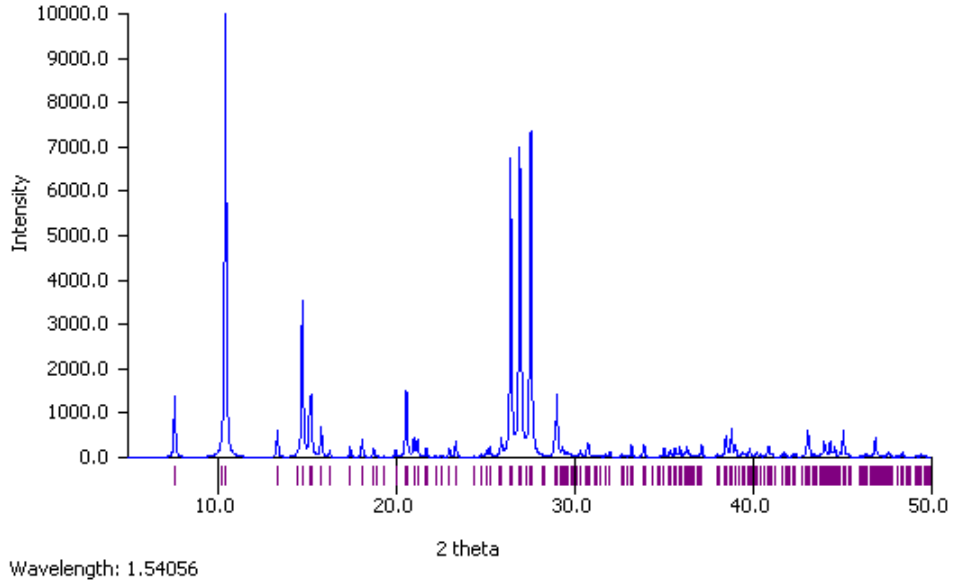


Figure E.4: Simulated PXRD patterns of DUXSAS generated in Mercury for the predicted crystal structure with the best match to the experimental (top) and the original experimental structure (bottom). X-ray powder patterns similarities between these two structures using the CSD Python API resulted in a similarity of 0.985.

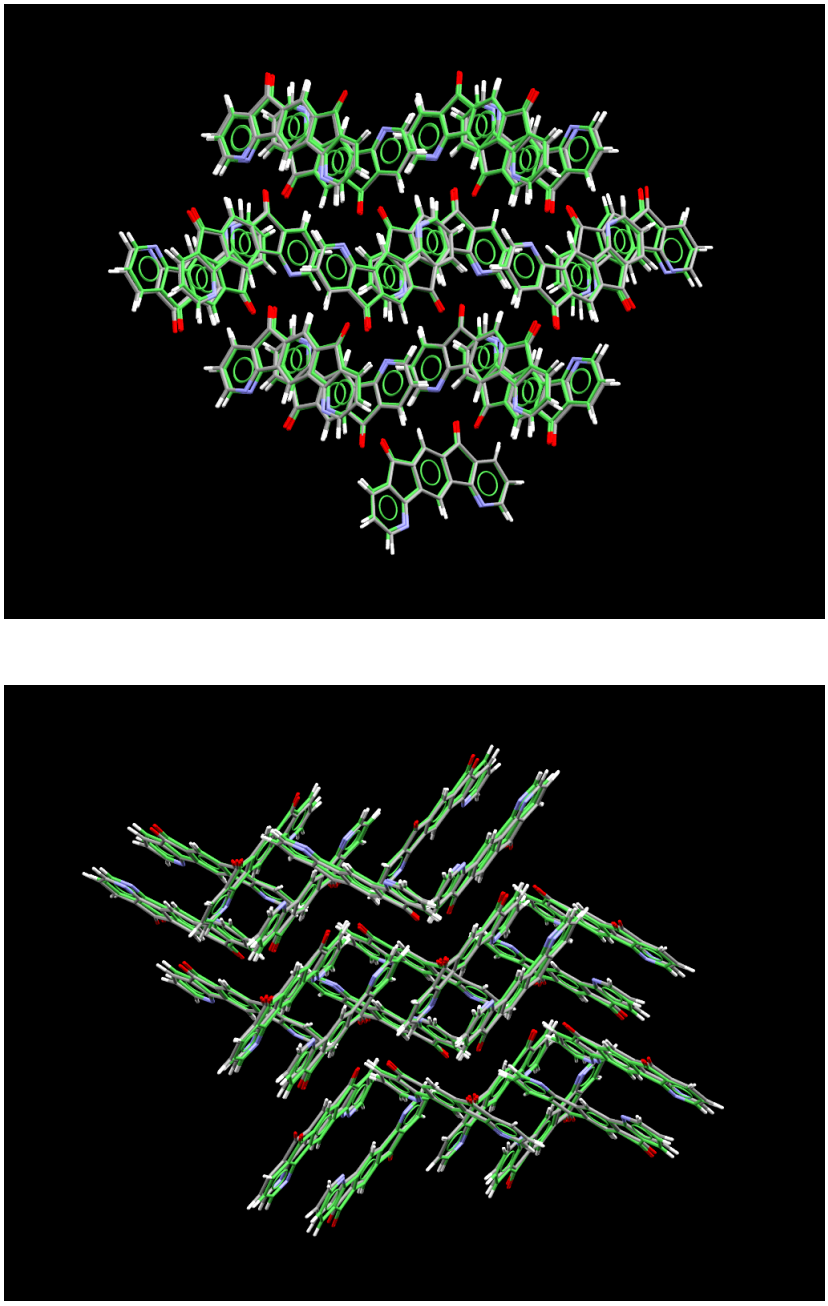
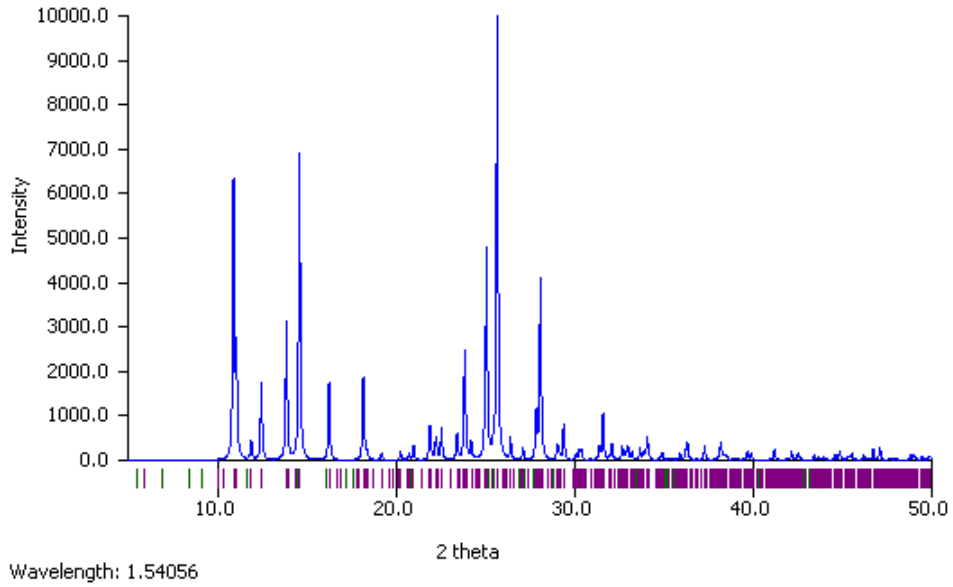


Figure E.5: Two views of the 30 out of 30 COMPACK overlay with an RMSD of 0.222 Å of MORROD for the predicted crystal structure with the best match to the experimental and the original experimental crystal structure (green).



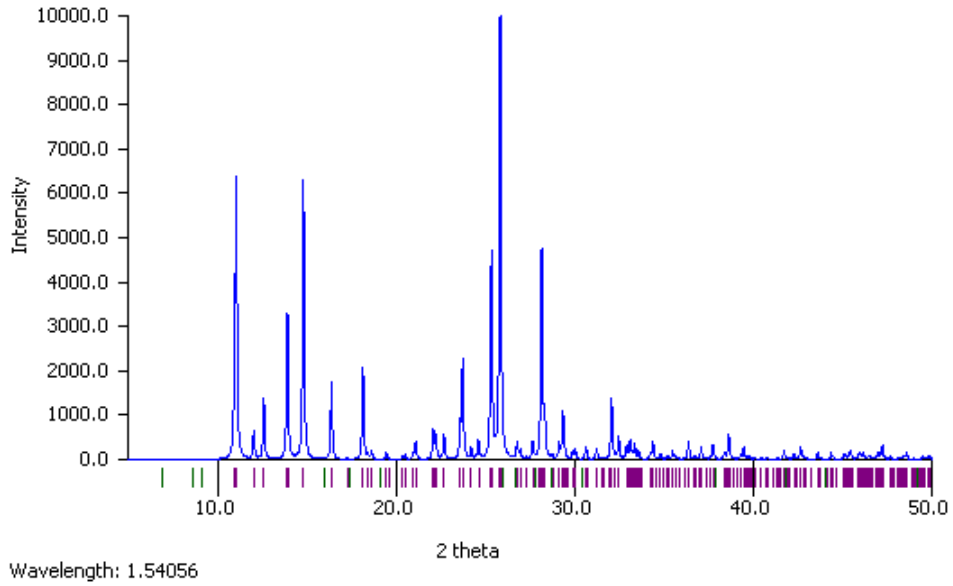


Figure E.6: Simulated PXRD patterns of MORROD generated in Mercury for the predicted crystal structure with the best match to the experimental (top) and the original experimental structure (bottom). X-ray powder patterns similarities between these two structures using the CSD Python API resulted in a similarity of 0.991.

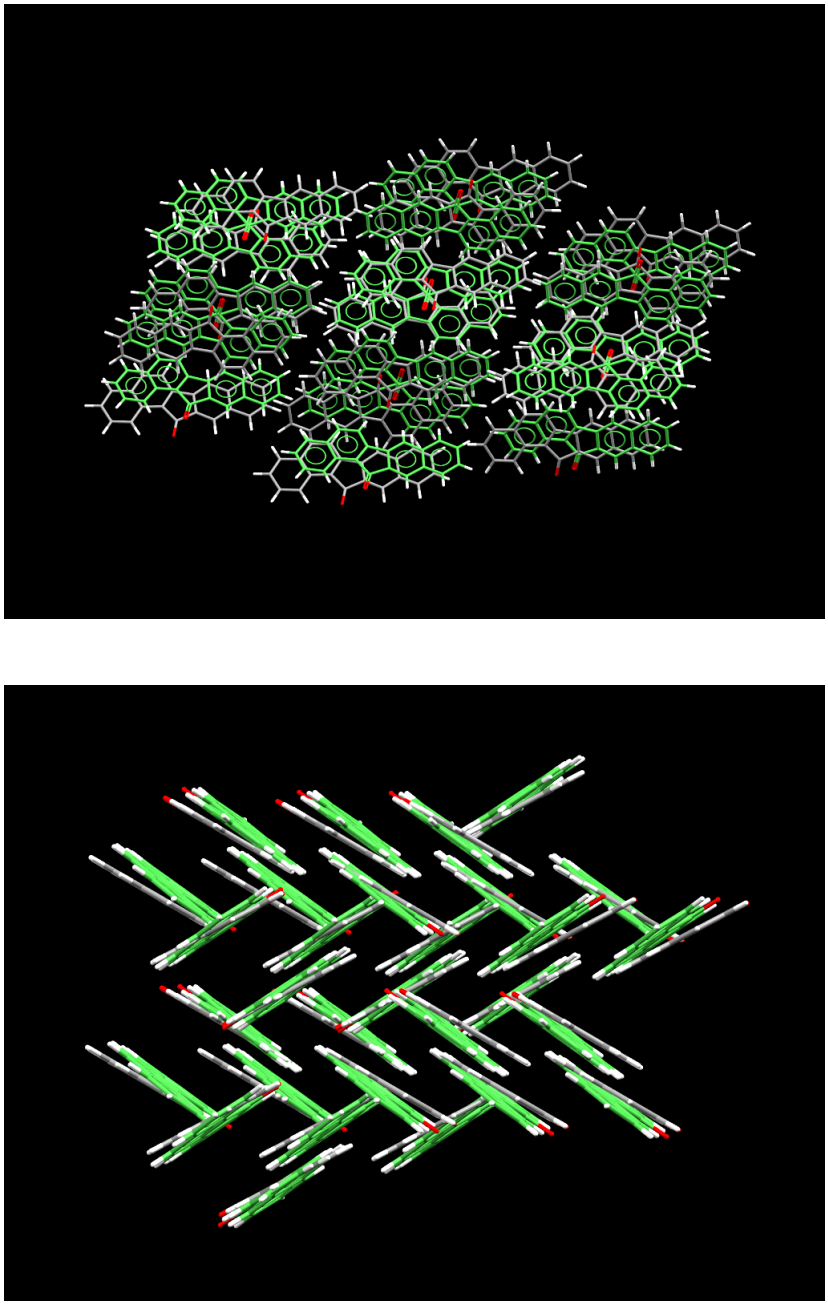


Figure E.7: Two views of the 30 out of 30 COMPACK overlay with an RMSD of 1.127 Å of NIJCEQ for the predicted crystal structure with the best match to the experimental and the original experimental crystal structure (green).

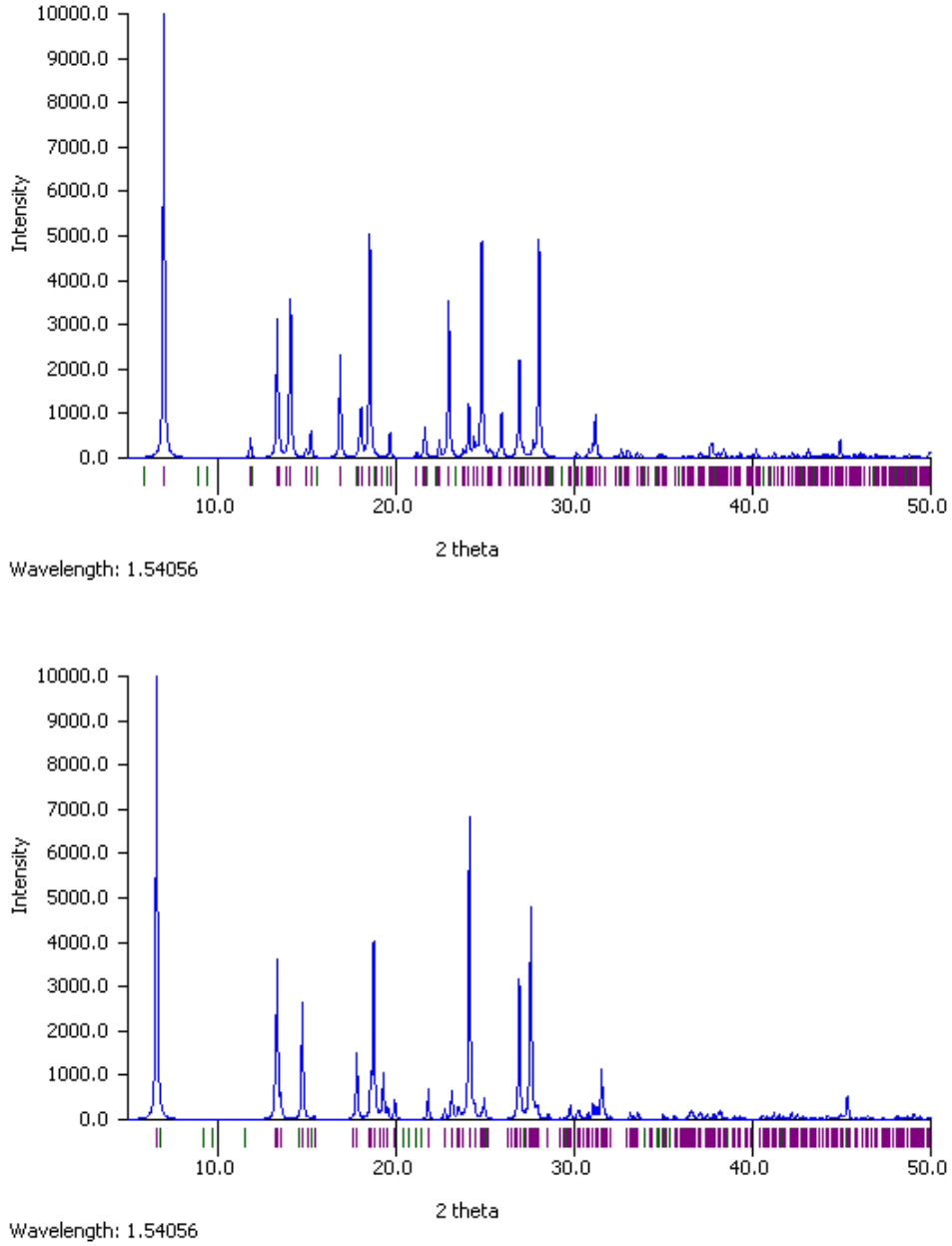


Figure E.8: Simulated PXRD patterns of NIJCEQ generated in Mercury for the predicted crystal structure with the best match to the experimental (top) and the original experimental structure (bottom). X-ray powder patterns similarities between these two structures using the CSD Python API resulted in a similarity of 0.951.

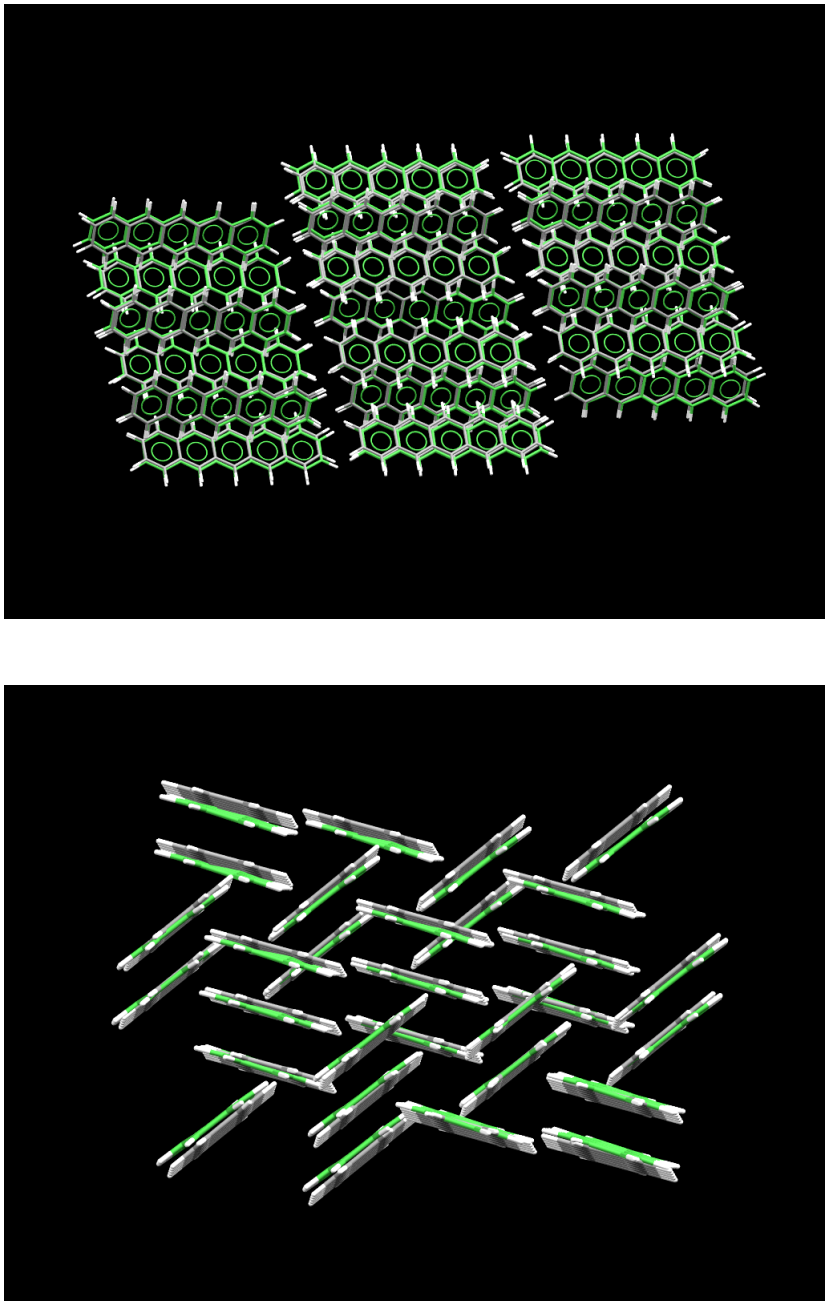


Figure E.9: Two views of the 30 out of 30 COMPACK overlay with an RMSD of 0.345 Å of PENCEN for the predicted crystal structure with the best match to the experimental and the original experimental crystal structure (green).

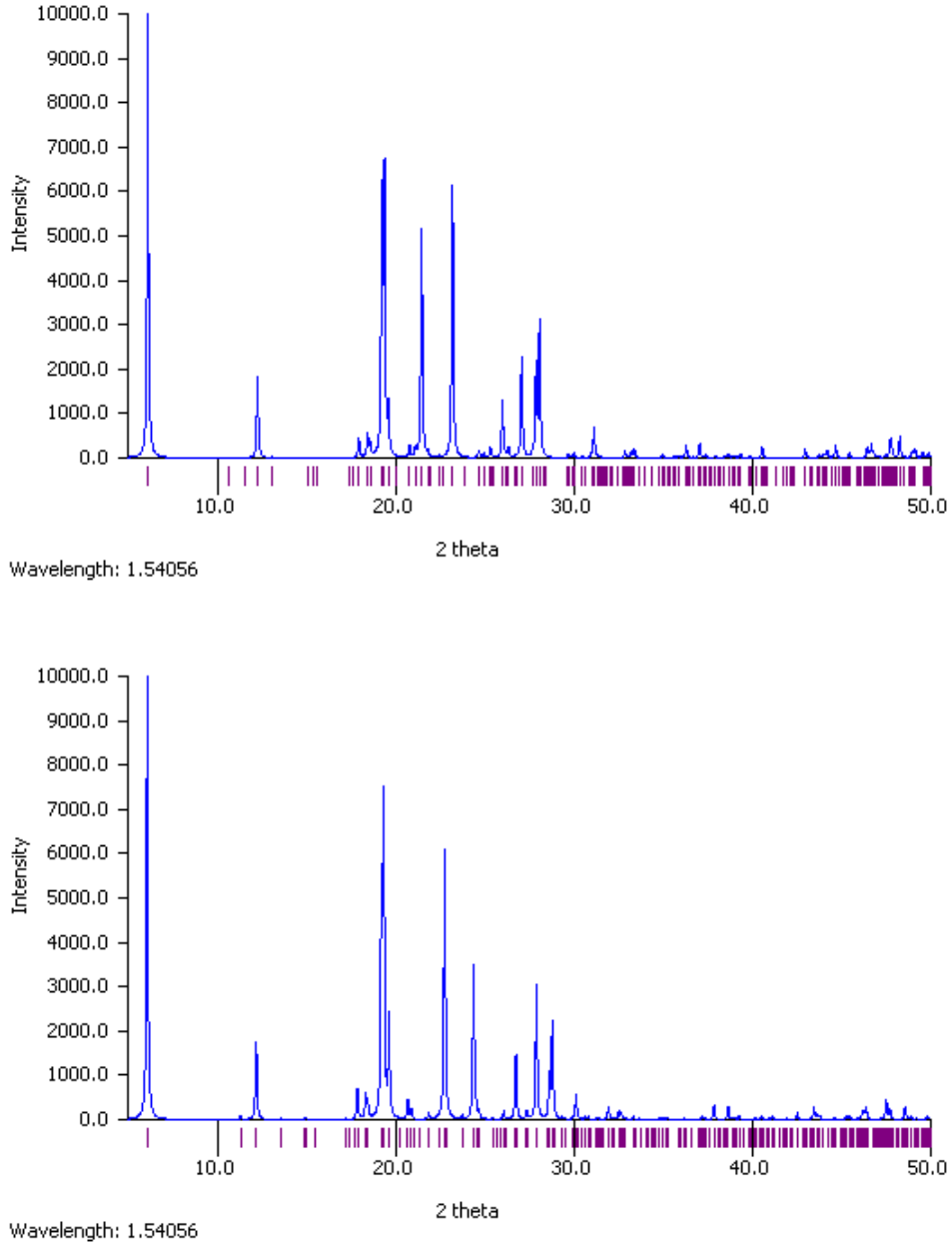


Figure E.10: Simulated PXRD patterns of PENCEN generated in Mercury for the predicted crystal structure with the best match to the experimental (top) and the original experimental structure (bottom). X-ray powder patterns similarities between these two structures using the CSD Python API resulted in a similarity of 0.967.

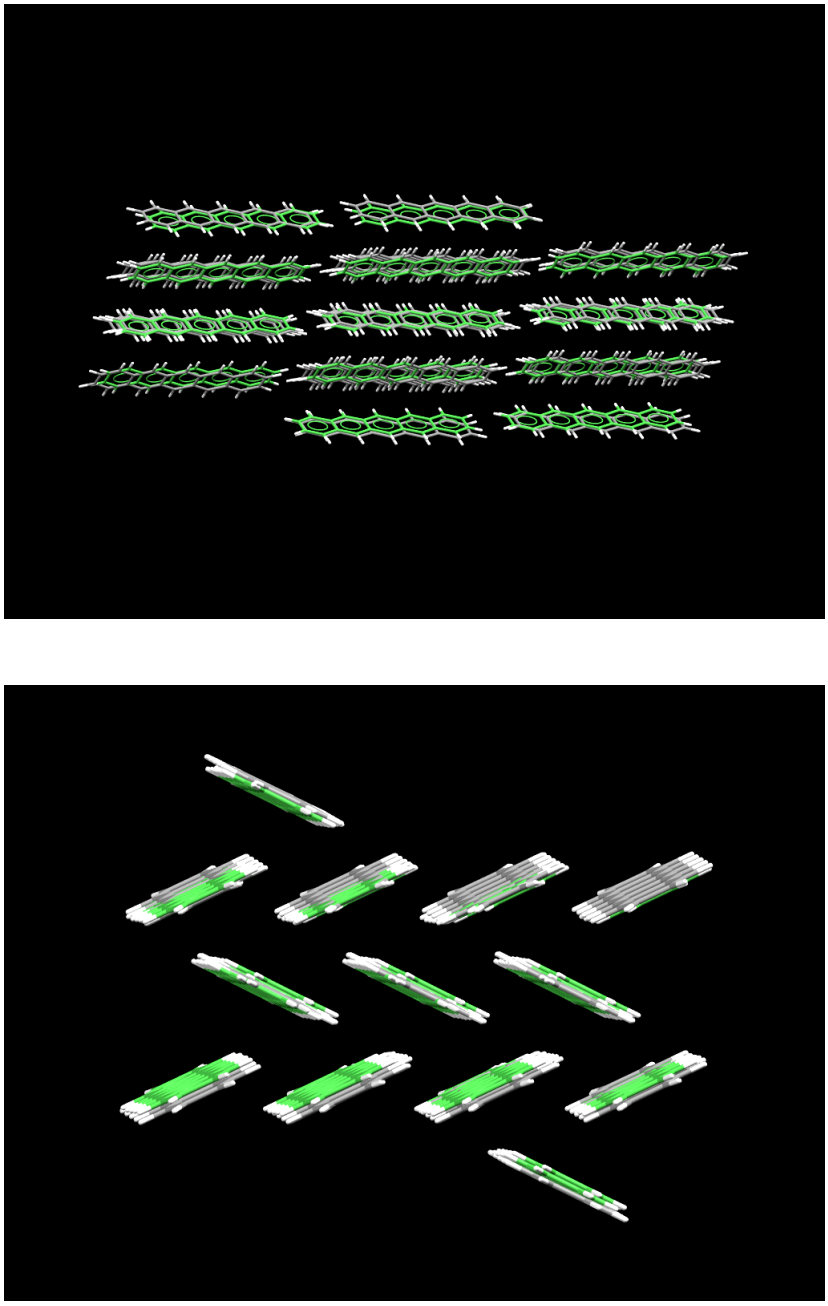
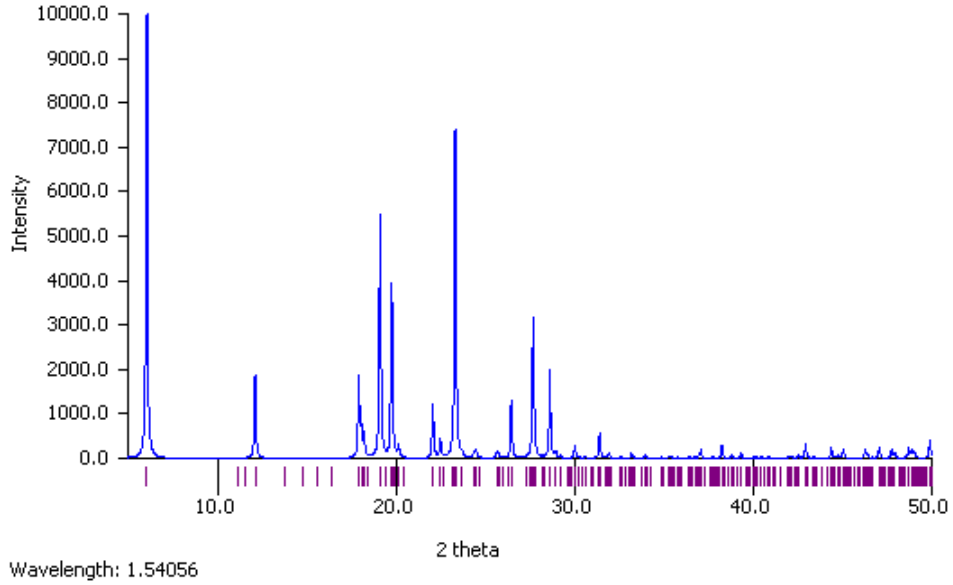


Figure E.11: Two views of the 30 out of 30 COMPACK overlay with an RMSD of 0.562 Å of PENCEN01 for the predicted crystal structure with the best match to the experimental and the original experimental crystal structure (green).



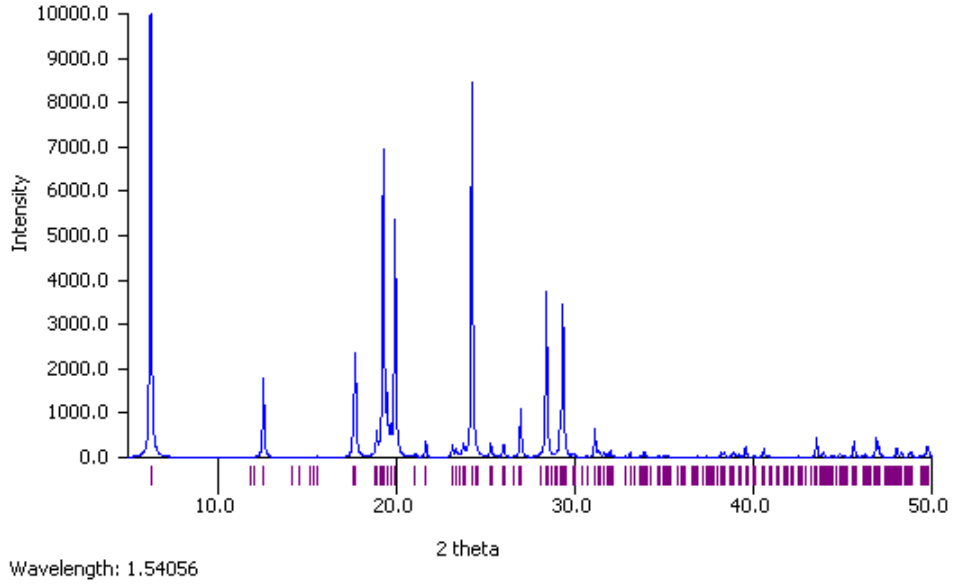


Figure E.12: Simulated PXRD patterns of PENCEN01 generated in Mercury for the predicted crystal structure with the best match to the experimental (top) and the original experimental structure (bottom). X-ray powder patterns similarities between these two structures using the CSD Python API resulted in a similarity of 0.929.

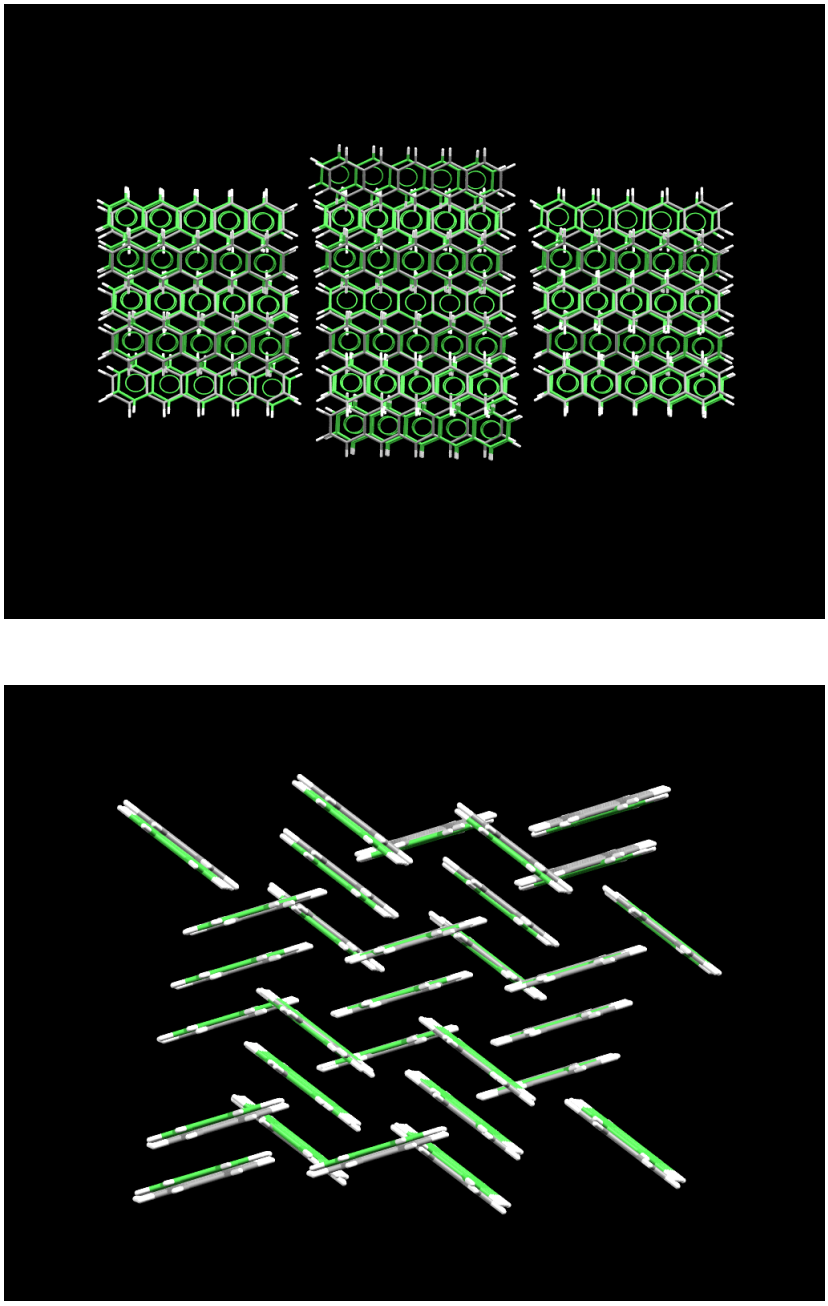


Figure E.13: Two views of the 30 out of 30 COMPACK overlay with an RMSD of 0.357 Å of PENCEN10 for the predicted crystal structure with the best match to the experimental and the original experimental crystal structure (green).

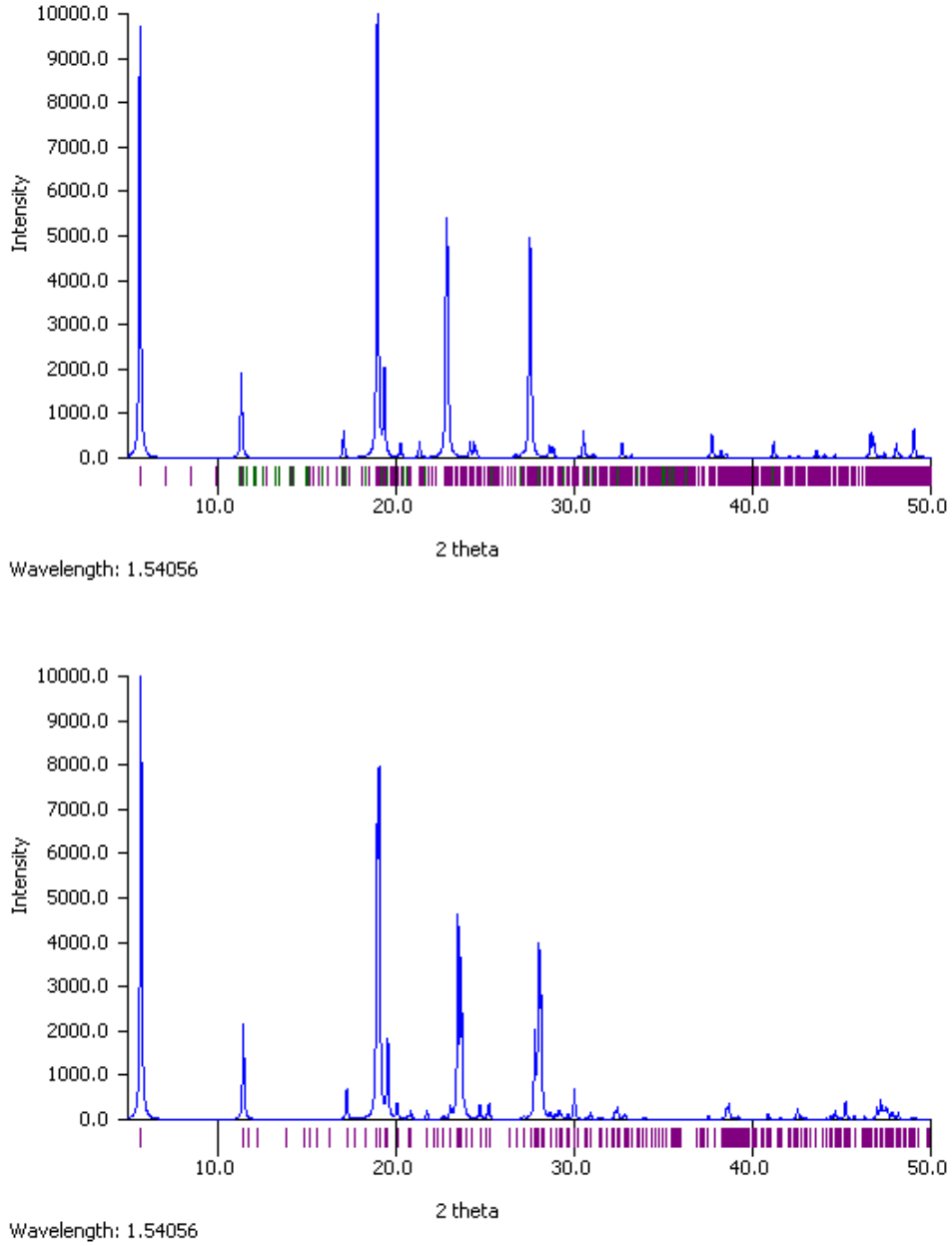


Figure E.14: Simulated PXRD patterns of PENCEN10 generated in Mercury for the predicted crystal structure with the best match to the experimental (top) and the original experimental structure (bottom). X-ray powder patterns similarities between these two structures using the CSD Python API resulted in a similarity of 0.952.

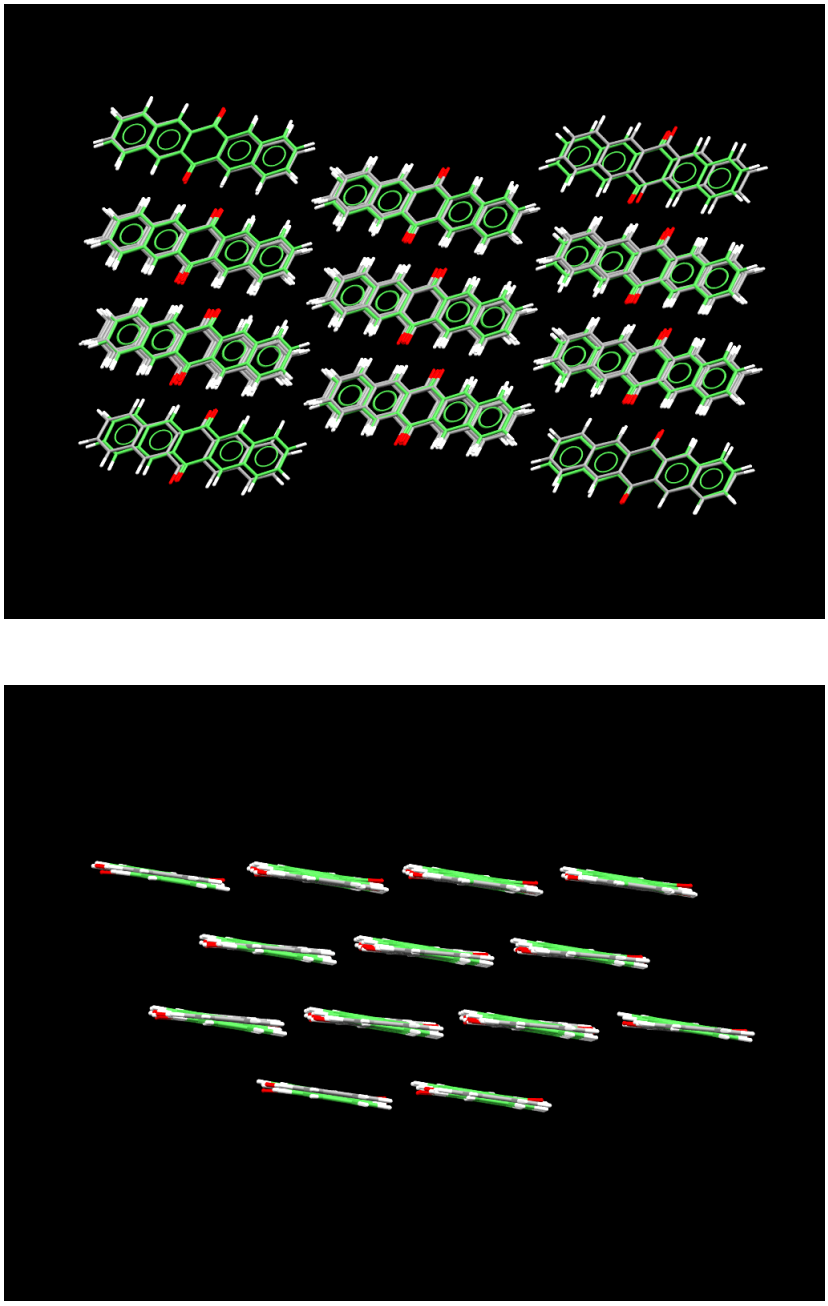


Figure E.15: Two views of the 28 out of 30 COMPACK overlay of PENTQU01 for the predicted crystal structure with the best match to the experimental and the original experimental crystal structure (green).

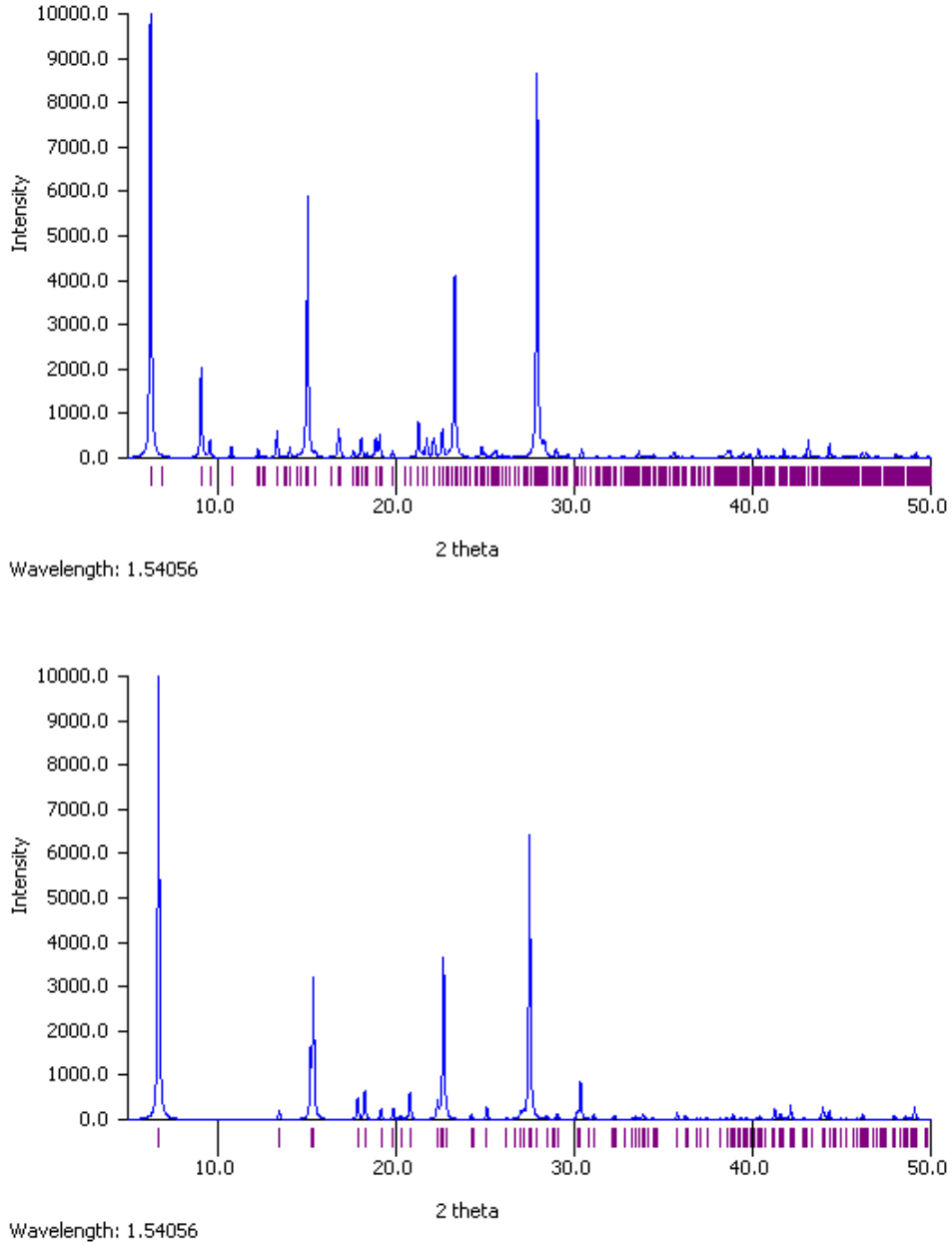


Figure E.16: Simulated PXRD patterns of PENTQU01 generated in Mercury for the predicted crystal structure with the best match to the experimental (top) and the original experimental structure (bottom). X-ray powder patterns similarities between these two structures using the CSD Python API resulted in a similarity of 0.913.

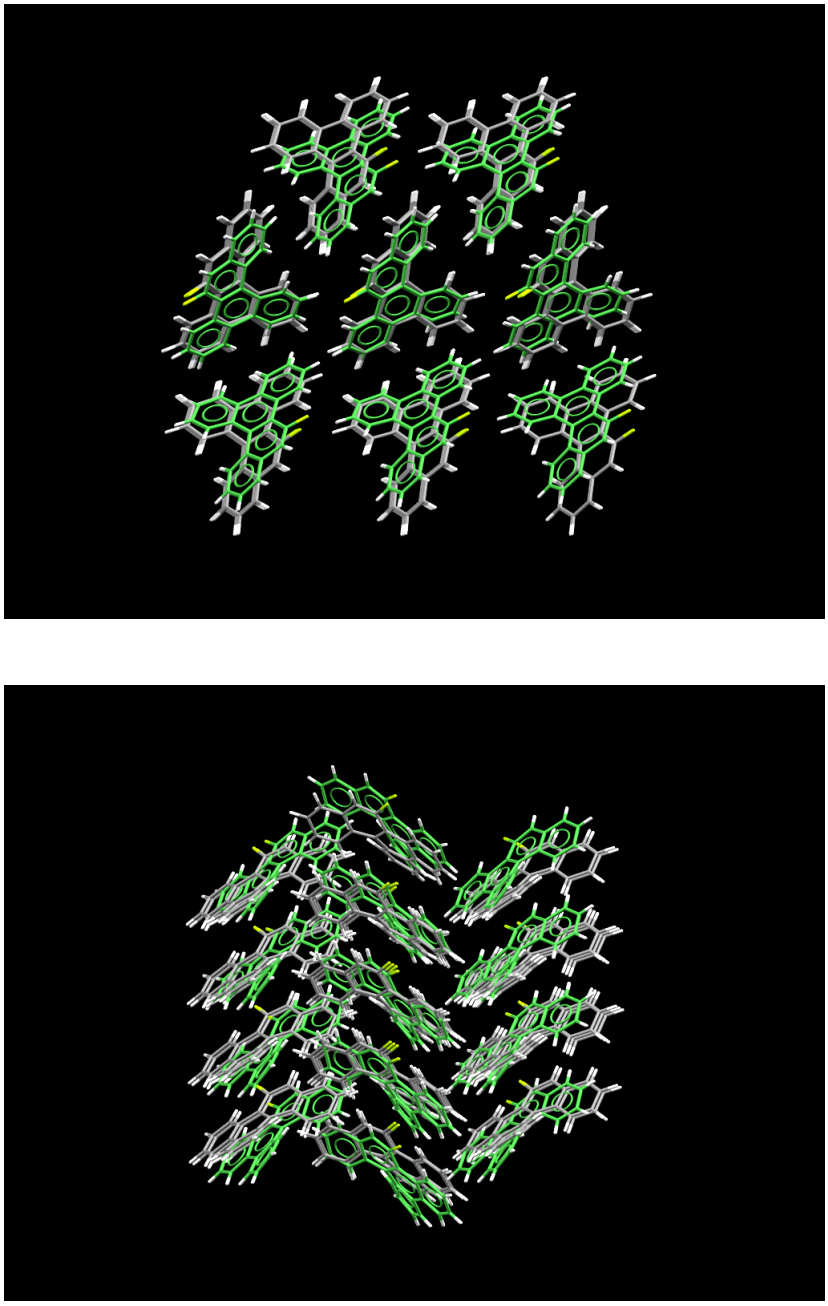


Figure E.17: Two views of the 30 out of 30 COMPACK overlay with an RMSD of 1.060 Å of UPIXAU for the predicted crystal structure with the best match to the experimental and the original experimental crystal structure (green).

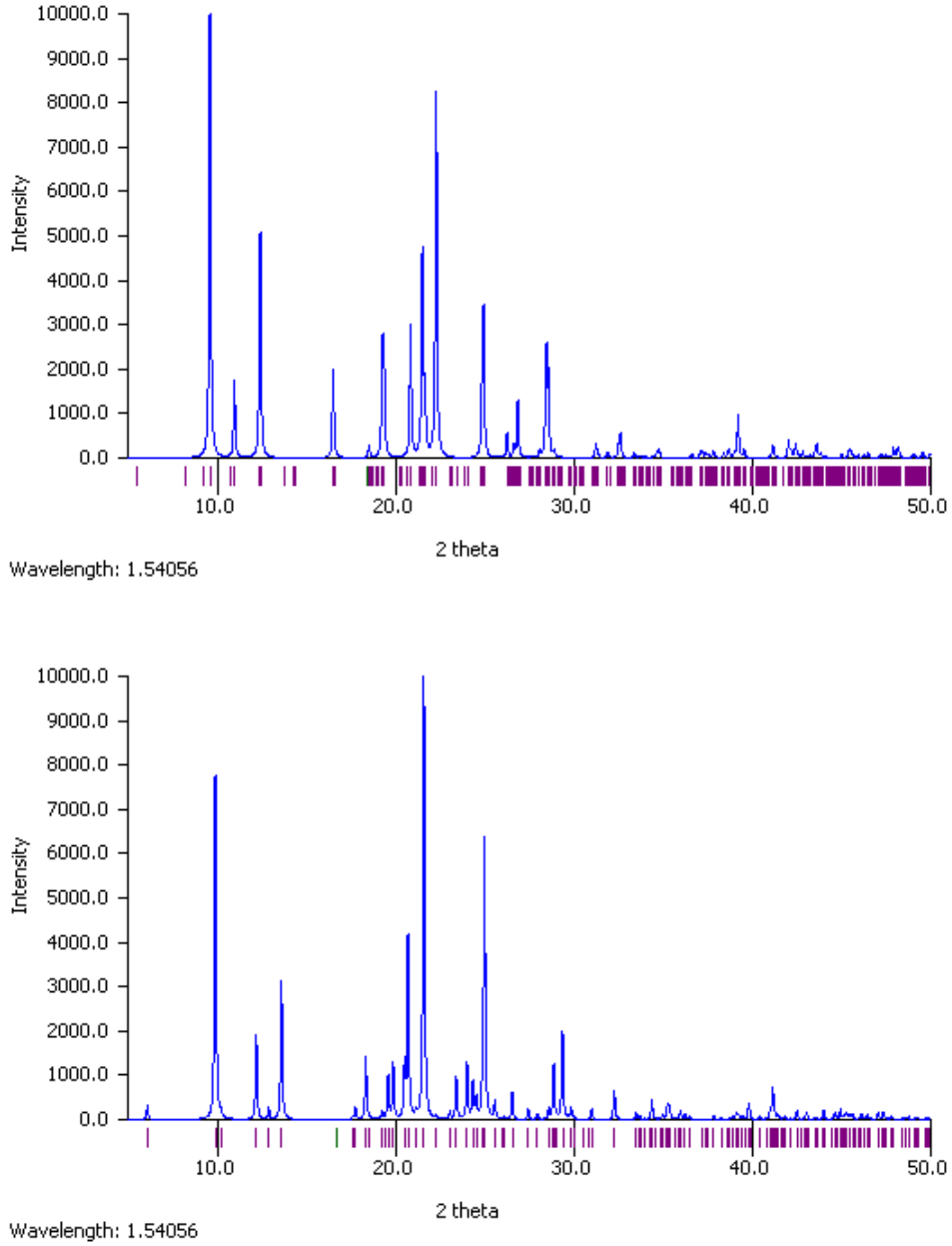


Figure E.18: Simulated PXRD patterns of UPIXAU generated in Mercury for the predicted crystal structure with the best match to the experimental (top) and the original experimental structure (bottom). X-ray powder patterns similarities between these two structures using the CSD Python API resulted in a similarity of 0.910.

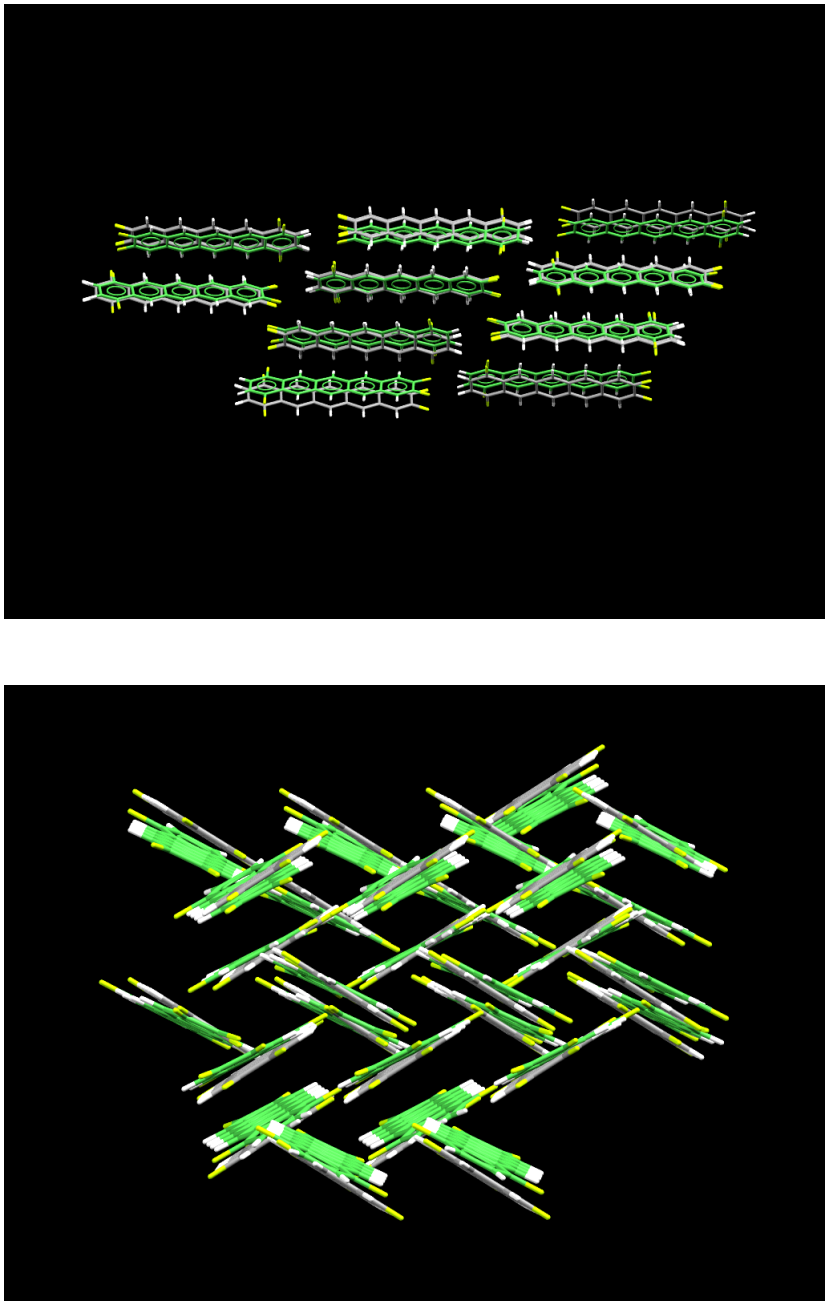


Figure E.19: Two views of the 28 out of 30 COMPACK overlay of VUFNIW for the predicted crystal structure with the best match to the experimental and the original experimental crystal structure (green).

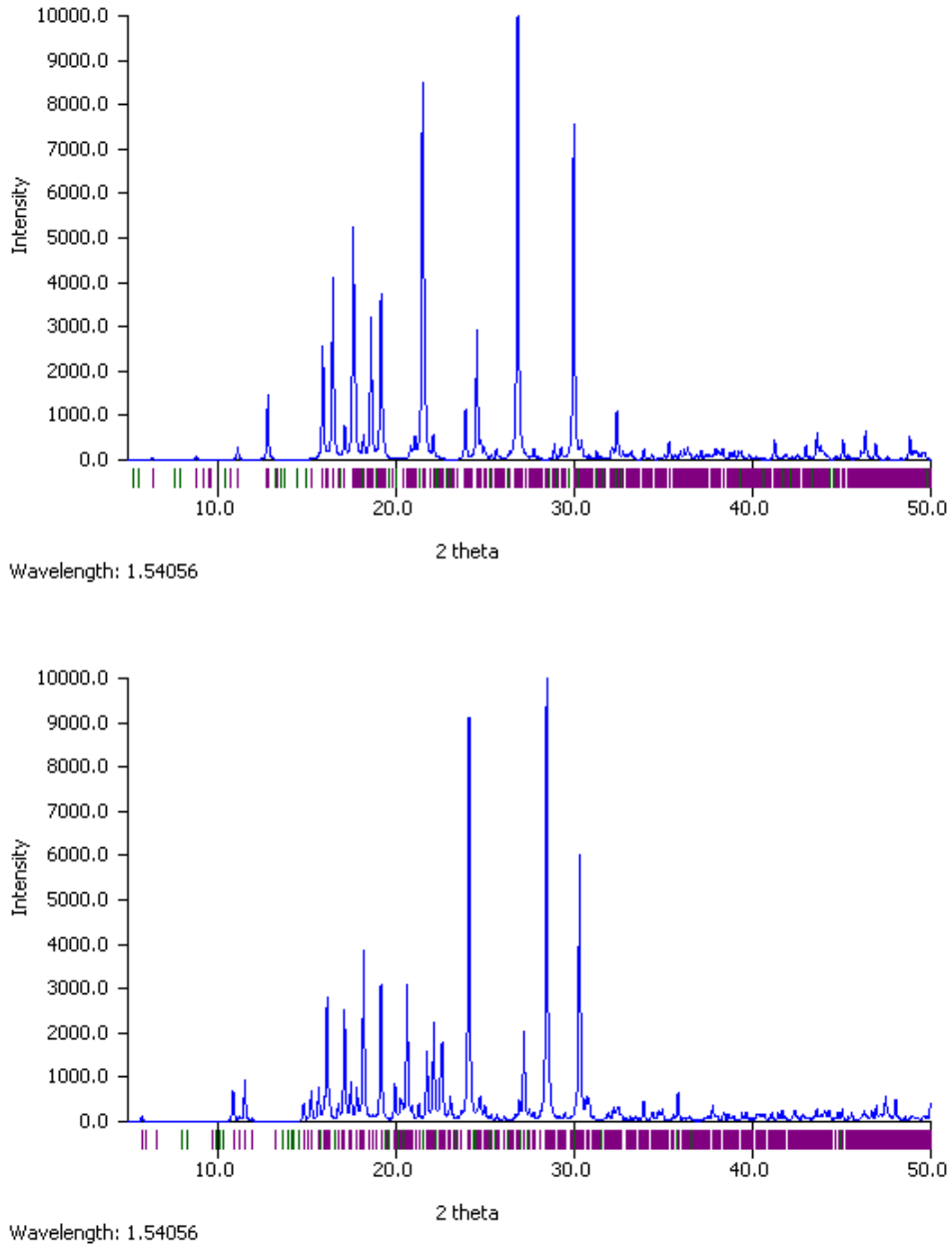


Figure E.20: Simulated PXRD patterns of VUFNIW generated in Mercury for the predicted crystal structure with the best match to the experimental (top) and the original experimental structure (bottom). X-ray powder patterns similarities between these two structures using the CSD Python API resulted in a similarity of 0.944.

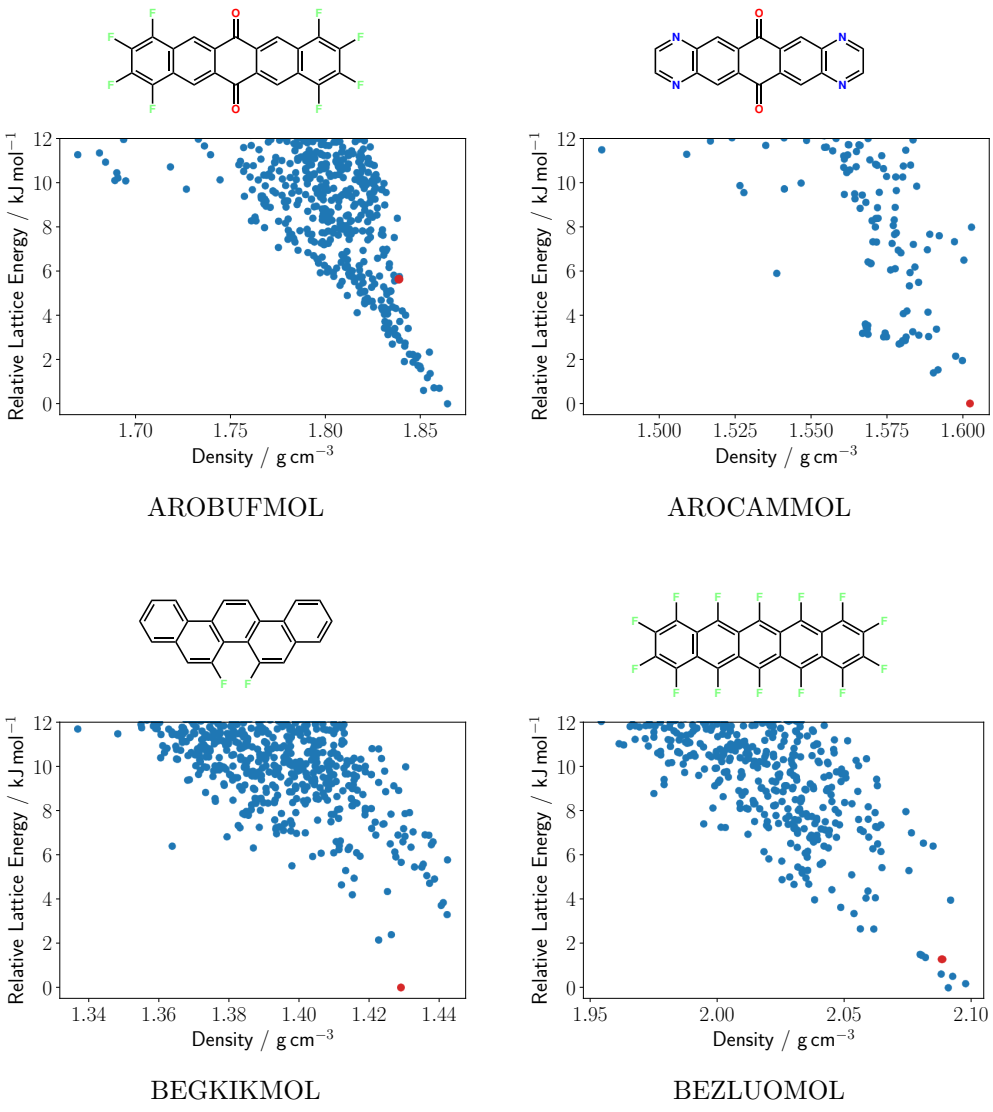
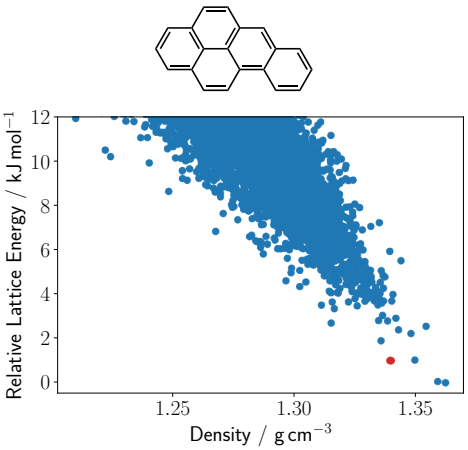
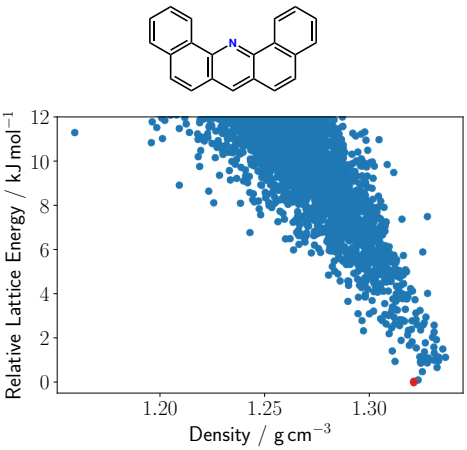


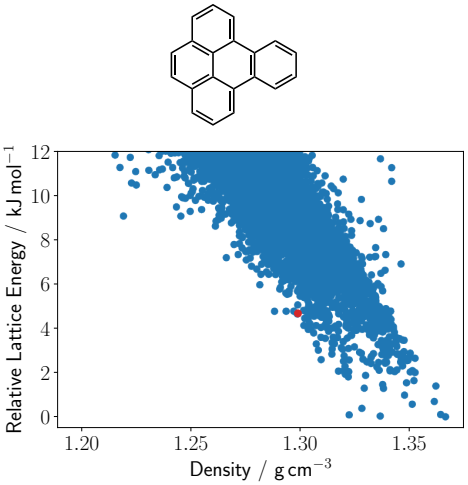
Figure E.21: Energy-density plots generated from the CSP searches described in Section 7.2 for all 40 molecules used to validate the potential model and generate data for the development of a reduced sampling scheme. Plot points coloured in red indicate matches with experimental crystal structures from the CSD.



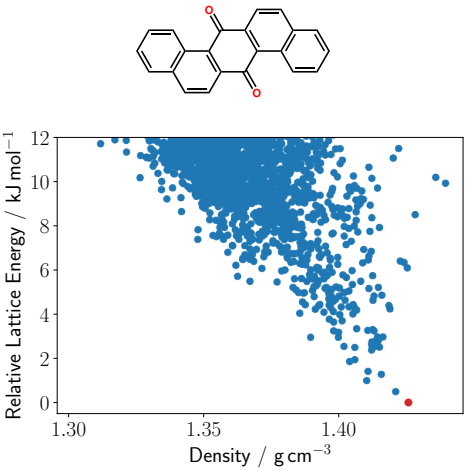
BNPYREMOL



BNZACRMOL

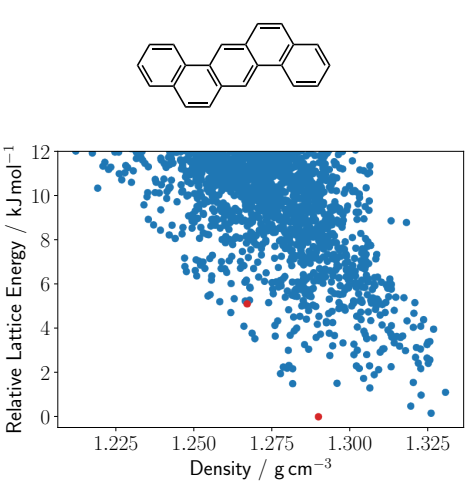


CEQGELMOL

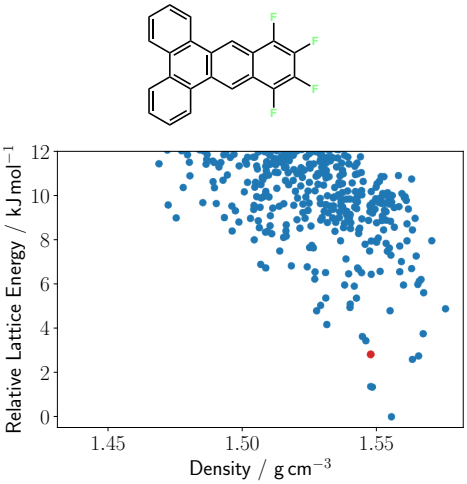


DBANQUMOL

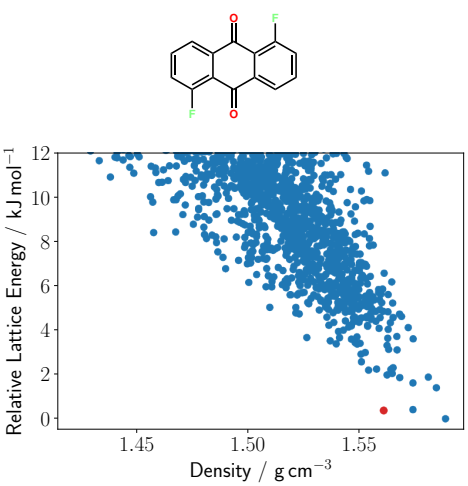
Figure E.21: *continued*



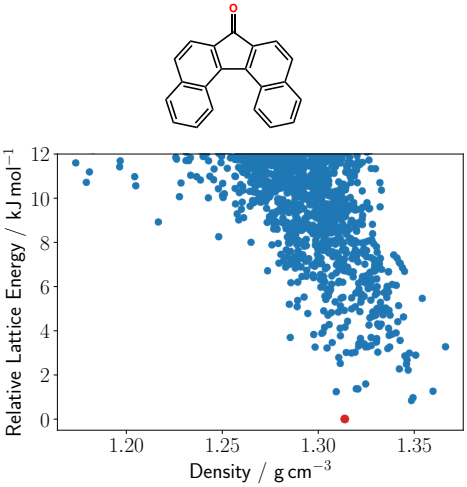
DBNTHRMOL



DUXSASMOL

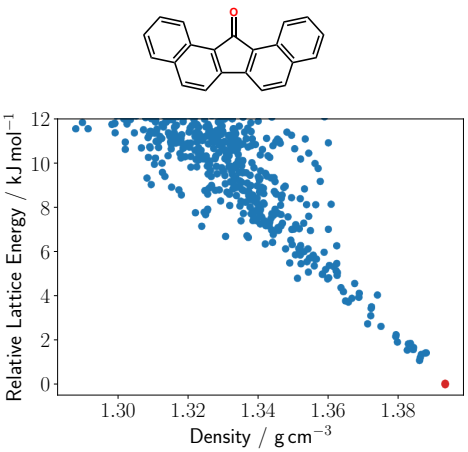


FANTRQMOL

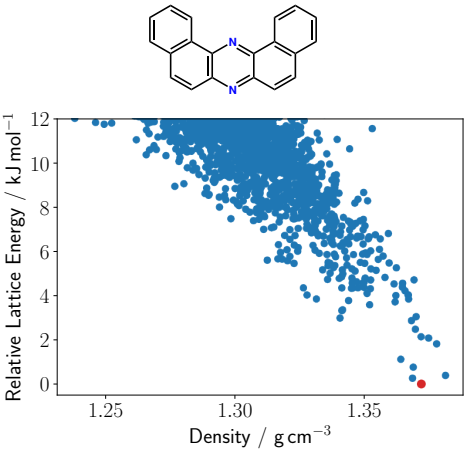


HAMDUWMOL

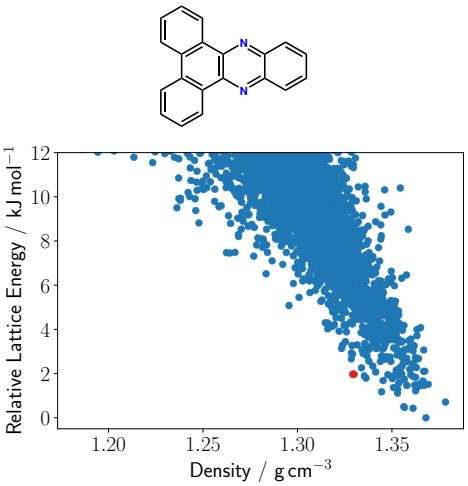
Figure E.21: *continued*



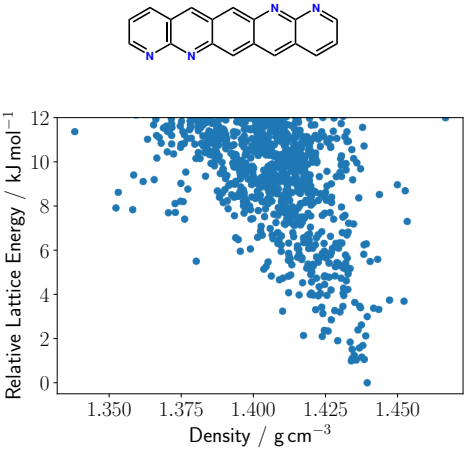
KONBAQMOL



LOHVAHMOL

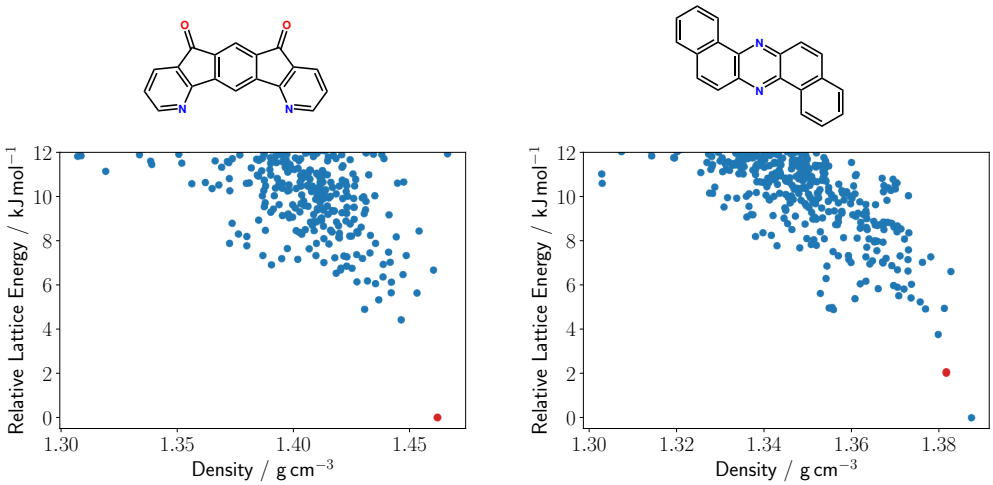


MIVRUEMOL

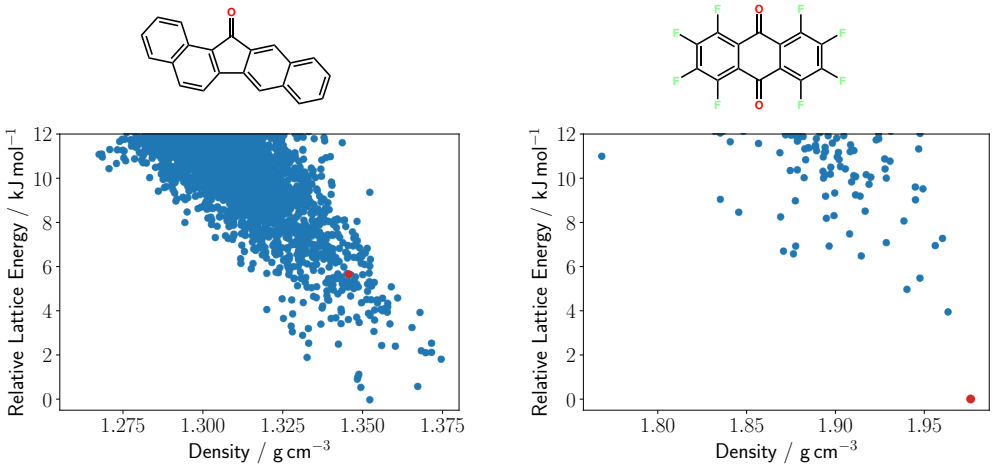


MIVSESMOL

Figure E.21: *continued*

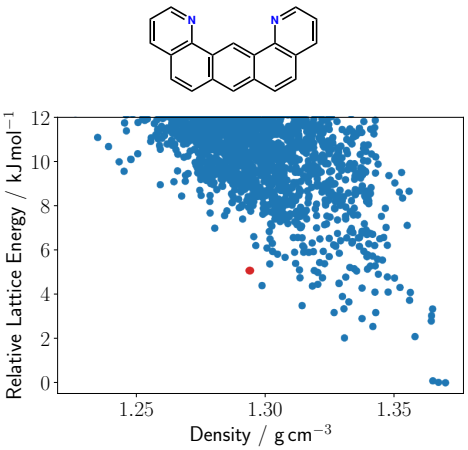


MORRODMOL NAAZASMOL

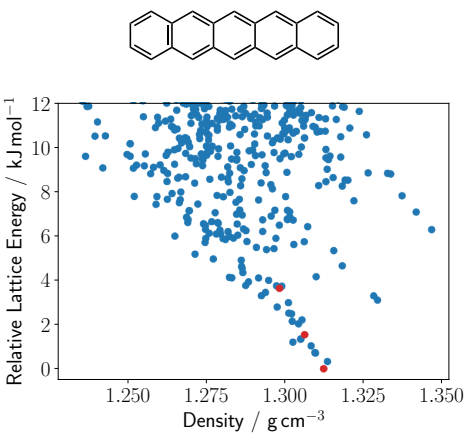


NIJCEQMOL OFANTQMOL

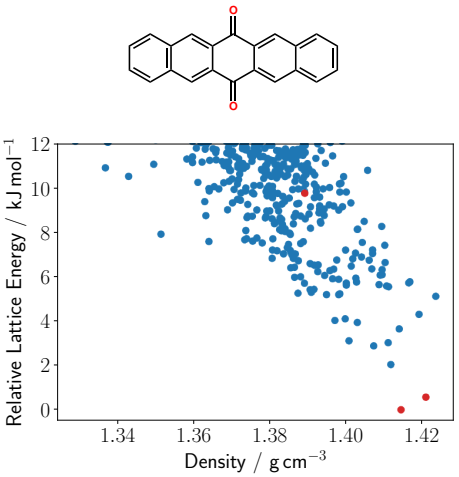
Figure E.21: *continued*



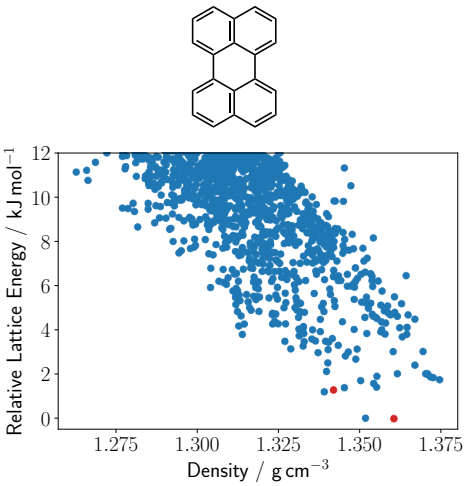
PAGMERMOL



PENCENMOL

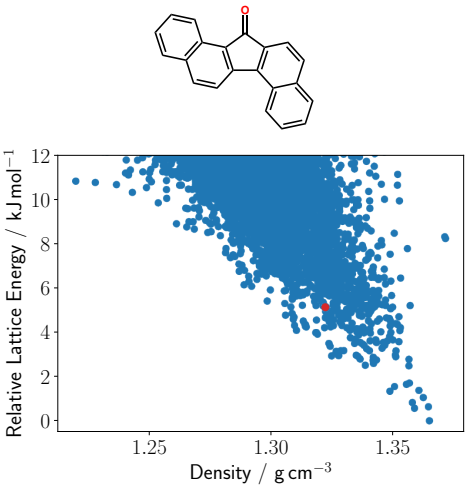


PENTQUMOL

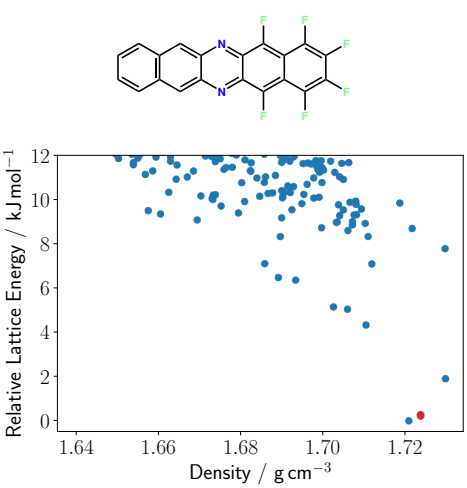


PERLENMOL

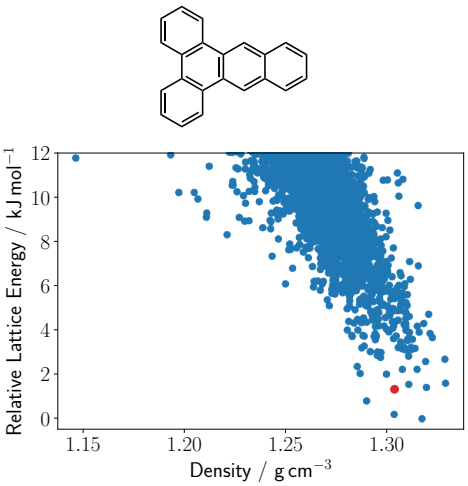
Figure E.21: *continued*



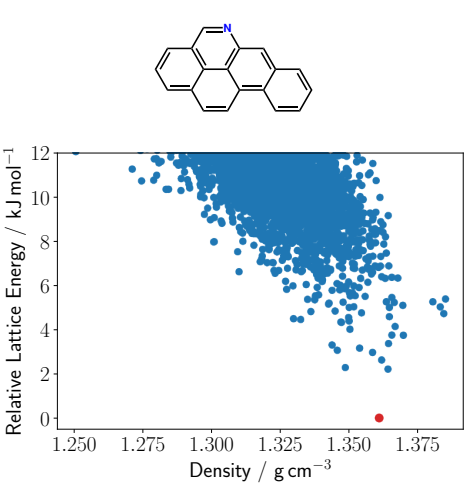
POVLUHMOL



RUNLAPMOL

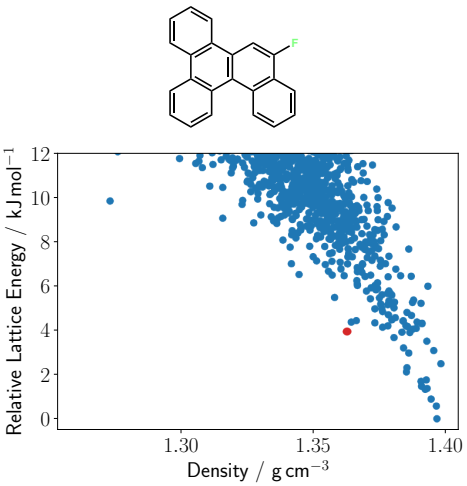


SANQIIMOL

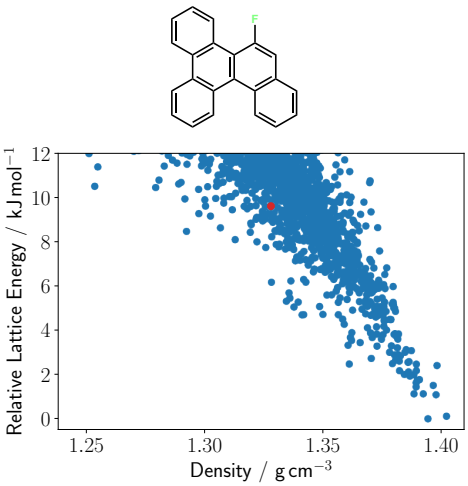


SEVYEAMOL

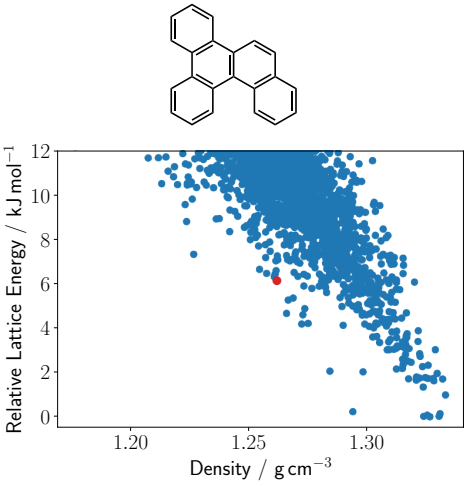
Figure E.21: *continued*



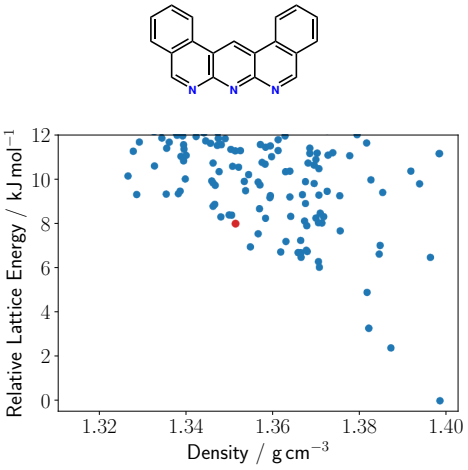
UPIRUIMOL



UPIXAUMOL

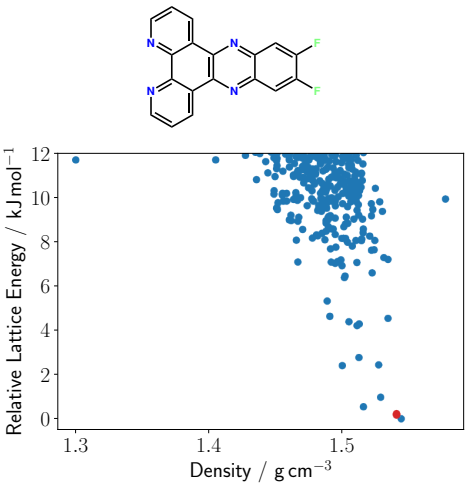


UPIXEYMOL

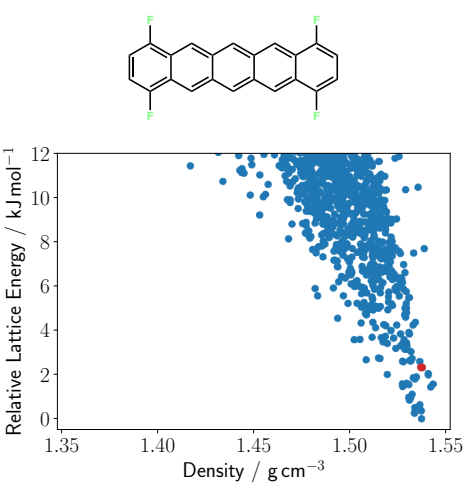


VIBMAVMOL

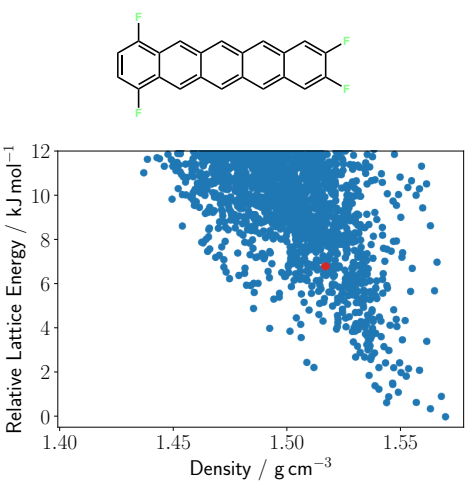
Figure E.21: *continued*



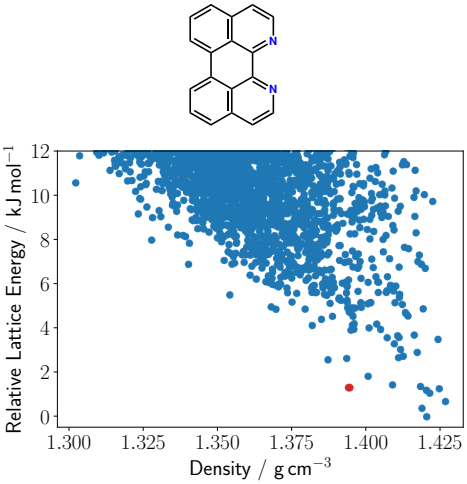
VIQTATMOL



VUFKALMOL

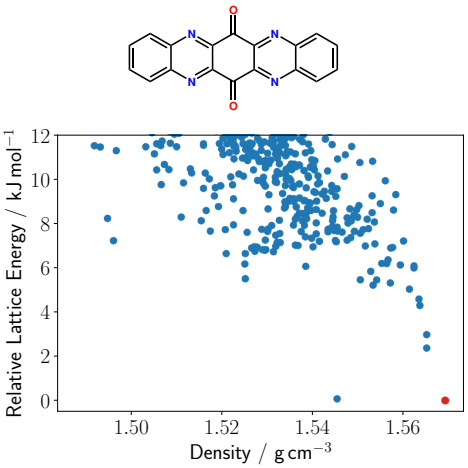


VUFNIWMOL

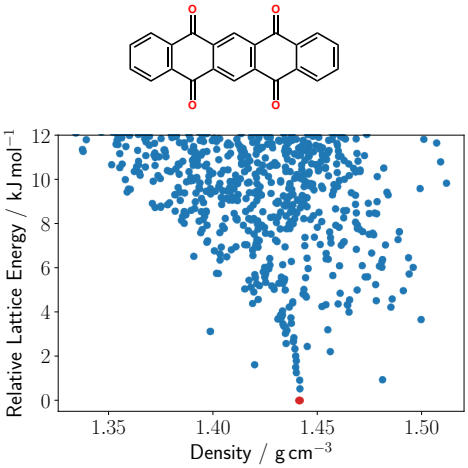


WOTZUDMOL

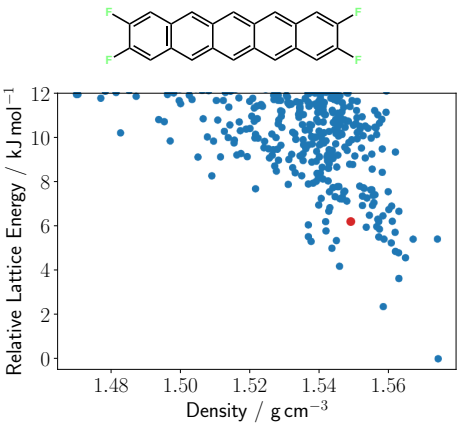
Figure E.21: *continued*



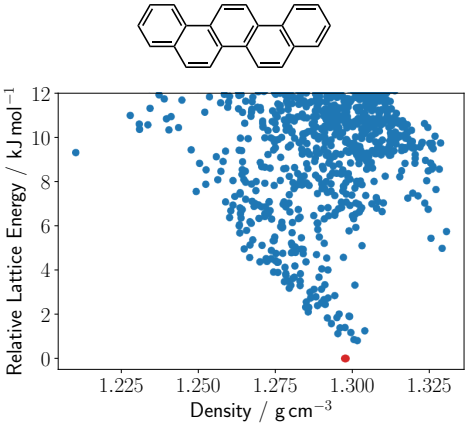
WUPYIQMOL



YOFROBMOL



ZETQUNMOL



ZZZOYCMOL

Figure E.21: *continued*

Predicted	CSD Refcode	RMSD30	RMSD60	PXRD	ΔE
AROBUF_optx2-QR-1-88962-3	AROBUF	0.458	0.641	0.956	5.627
AROCAM_opt-QR-2-12929-3	AROCAM	0.145	0.177	0.976	0.000
BEGKIK_opt-QR-61-7745-3	BEGKIK	0.179	0.202	0.993	0.000
BEZLUO_opt-QR-4-7633-3	BEZLUO	0.327	0.472	0.931	1.261
BNPYRE_optx2-QR-14-2060-3	BNPYRE10	0.285	0.352	0.989	0.968
BNPYRE_optx2-QR-14-2060-3	BNPYRE11	0.290	0.373	0.986	0.968
BNPYRE_optx2-QR-14-2060-3	BNPYRE12	0.301	0.356	0.987	0.968
BNZACR_opt-QR-33-1910-3	BNZACR	0.283	0.363	0.972	0.000
CEQGEL_optx2-QR-2-32200-3	CEQGEL	0.334	0.417	0.985	4.665
CEQGEL_optx2-QR-2-32200-3	CEQGEL01	0.234	0.282	0.978	4.665
DBANQU_opt-QR-29-901-3	DBANQU	0.436	0.490	0.968	0.000
DBNTHR_opt-QR-19-4321-3	DBNTHR01	0.551	0.689	0.957	0.000
DBNTHR_opt-QR-19-4321-3	DBNTHR02	0.555	0.691	0.957	0.000
DBNTHR_opt-QR-4-4352-3	DBNTHR10	0.292	0.304	0.987	5.091
DUXSAS_optx2-QR-14-13217-3	DUXSAS	-	-	0.985	2.821
FANTRQ_optx2-QR-4-11713-3	FANTRQ10	0.186	0.230	0.986	0.347
HAMDUW_opt-QR-14-1087-3	HAMDUW	0.399	0.502	0.968	0.000
HAMDUW_opt-QR-14-1087-3	HAMDUW01	0.385	0.484	0.970	0.000
HAMDUW_opt-QR-14-1087-3	HAMDUW02	0.465	0.578	0.949	0.000
KONBAQ_opt-QR-33-6642-3	KONBAQ	0.091	0.123	0.994	0.000
LOHVAH_optx2-QR-4-19292-3	LOHVAH	0.227	0.304	0.986	0.000
MIVRUE_opt-QR-19-169-3	MIVRUE	0.458	0.516	0.975	1.956

Table E.2: COMPACK RMSDs for a 30 out of 30 and a 60 out of 60 cluster with distance and angles tolerances of 30% and 30° or smaller and hydrogen positions ignored, X-ray similarities between the matched predicted crystal structures from the CSP searches described in Section 7.2 against experimental structures, and the relative lattice energies of the matched predicted structures. Matches to the experimental structure were not obtained for MIVSES and a full 30 out of 30 or 60 out of 60 match in COMPACK were not obtained for the experimental structures of DUXSAS, PENTQU01, and VUFNIW.

Predicted	CSD Refcode	RMSD30	RMSD60	PXRD	ΔE
-	MIVSES	-	-	-	-
MORROD_optx2-QR-14-30352-3	MORROD	0.222	0.248	0.991	0.000
NAAZAS_optx2-QR-14-23290-3	NAAZAS	0.150	0.182	0.991	2.034
NIJCEQ_opt-QR-14-7929-3	NIJCEQ	1.127	1.328	0.951	5.646
OFANTQ_opt-QR-14-5875-3	OFANTQ	0.171	0.201	0.990	0.000
OFANTQ_opt-QR-14-5875-3	OFANTQ01	0.229	0.271	0.951	0.000
PAGMER_optx2-QR-14-10607-3	PAGMER	0.184	0.209	0.993	5.073
PENCEN_optx2-QR-1-107678-3	PENCEN	0.345	0.481	0.967	0.000
PENCEN_optx2-QR-1-244620-3	PENCEN01	0.562	0.710	0.929	1.529
PENCEN_optx2-QR-1-244620-3	PENCEN02	0.508	0.656	0.948	1.529
PENCEN_optx2-QR-1-244620-3	PENCEN03	0.514	0.638	0.947	1.529
PENCEN_optx2-QR-1-244620-3	PENCEN04	0.601	0.762	0.901	1.529
PENCEN_optx2-QR-1-107678-3	PENCEN05	0.302	0.413	0.965	0.000
PENCEN_optx2-QR-1-244620-3	PENCEN06	0.590	0.744	0.919	1.529
PENCEN_optx2-QR-1-244620-3	PENCEN07	0.501	0.631	0.957	1.529
PENCEN_optx2-QR-1-244620-3	PENCEN08	0.441	0.589	0.977	1.529
PENCEN_optx2-QR-1-107678-3	PENCEN09	0.304	0.409	0.964	0.000
PENCEN_optx2-QR-14-11695-3	PENCEN10	0.357	0.460	0.952	3.613
PENCEN_optx2-QR-14-11695-3	PENCEN11	0.371	0.479	0.950	3.613
PENCEN_optx2-QR-14-11695-3	PENCEN12	0.359	0.461	0.951	3.613
PENTQU_optx2-QR-2-48396-3	PENTQU	0.165	0.209	0.980	0.000
PENTQU_optx2-QR-2-70612-3	PENTQU01	-	-	0.913	9.77
PENTQU_optx2-QR-1-52560-3	PENTQU02	0.575	0.659	0.924	0.569
PERLEN_optx2-QR-2-51571-3	PERLEN01	0.273	0.347	0.991	1.266
PERLEN_optx2-QR-2-51571-3	PERLEN03	0.272	0.345	0.989	1.266

Table E.2: *continued*

Predicted	CSD Refcode	RMSD30	RMSD60	PXRD	ΔE
PERLEN_optx2-QR-2-51571-3	PERLEN04	0.271	0.340	0.991	1.266
PERLEN_optx2-QR-2-51571-3	PERLEN05	0.303	0.385	0.988	1.266
PERLEN_opt-QR-4-2473-3	PERLEN06	0.631	0.842	0.916	0.000
PERLEN_opt-QR-4-2473-3	PERLEN07	0.631	0.833	0.908	0.000
PERLEN_optx2-QR-2-51571-3	PERLEN08	0.294	0.373	0.990	1.266
POVLUH_opt-QR-15-11115-3	POVLUH	0.649	0.813	0.965	5.123
RUNLAP_opt-QR-14-3143-3	RUNLAP	0.397	0.457	0.966	0.230
SANQII_optx2-QR-4-32846-3	SANQII	0.458	0.596	0.982	1.311
SEVYEA_opt-QR-19-1708-3	SEVYEA	0.223	0.290	0.985	0.000
UPIRUI_opt-QR-61-6185-3	UPIRUI	0.206	0.234	0.988	3.928
UPIXAU_optx2-QR-4-11875-3	UPIXAU	1.060	1.349	0.910	9.611
UPIXEY_opt-QR-29-4519-3	UPIXEY	0.336	0.409	0.980	6.132
VIBMAV_opt-QR-14-559-3	VIBMAV	0.307	0.385	0.977	7.993
VIQTAT_opt-QR-19-1876-3	VIQTAT	0.364	0.426	0.977	0.195
VUFKAL_opt-QR-14-7845-3	VUFKAL	0.686	0.816	0.935	2.299
VUFNIW_optx2-QR-14-55723-3	VUFNIW	-	-	0.944	6.790
WOTZUD_opt-QR-14-5639-3	WOTZUD	0.249	0.309	0.972	1.280
WUPYIQ_optx2-QR-2-41353-3	WUPYIQ	0.274	0.338	0.983	0.000
YOFROB_optx2-QR-1-16182-3	YOFROB	0.348	0.513	0.940	0.000
YOFROB_optx2-QR-1-16182-3	YOFROB01	0.289	0.418	0.966	0.000
ZETQUN_optx2-QR-1-243541-3	ZETQUN	0.262	0.340	0.947	6.206
ZZZOYC_opt-QR-4-2556-3	ZZZOYC01	0.386	0.510	0.946	0.000
ZZZOYC_opt-QR-4-2556-3	ZZZOYC02	0.427	0.536	0.920	0.000
ZZZOYC_opt-QR-4-2556-3	ZZZOYC03	0.386	0.487	0.929	0.000
ZZZOYC_opt-QR-4-2556-3	ZZZOYC04	0.446	0.592	0.940	0.000

Table E.2: *continued*

CSP	Z'=1	Z'=2
AROBUFMOL	9h 14m 56s	19h 7m 44s
AROCAMMOL	8h 3m 28s	17h 6m 35s
BEGKIKMOL	7h 34m 19s	17h 4m 0s
BEZLUOMOL	8h 56m 31s	18h 34m 46s
BNPYREMOL	7h 18m 14s	15h 58m 52s
BNZACRMOL	7h 57m 56s	17h 16m 15s
CEQGELMOL	7h 21m 1s	15h 44m 56s
DBANQUMOL	8h 55m 44s	19h 20m 9s
DBNTHRMOL	8h 12m 48s	17h 56m 5s
DUXSASMOL	7h 43m 24s	17h 11m 40s
FANTRQMOL	5h 18m 37s	11h 41m 6s
HAMDUWMOL	6h 59m 37s	16h 6m 54s
KONBAQMOL	7h 59m 28s	17h 20m 27s
LOHVAHMOL	7h 48m 48s	17h 2m 26s
MIVRUEMOL	7h 54m 57s	16h 49m 59s
MIVSESMOL	7h 45m 54s	16h 29m 35s
MORRODMOL	6h 36m 21s	14h 24m 57s
NAAZASMOL	8h 5m 44s	17h 7m 58s
NIJCEQMOL	7h 59m 5s	17h 4m 6s
OFANTQMOL	4h 39m 24s	10h 5m 49s
PAGMERMOL	7h 32m 19s	16h 9m 1s
PENCENMOL	8h 20m 39s	17h 19m 42s
PENTQUMOL	8h 20m 33s	17h 38m 39s
PERLENMOL	7h 11m 7s	15h 10m 42s
POVLUHMOL	7h 52m 19s	17h 11m 36s
RUNLAPMOL	7h 57m 52s	17h 4m 59s
SANQIIMOL	7h 55m 29s	17h 19m 38s
SEVYEAMOL	7h 4m 30s	15h 26m 28s
UPIRUIMOL	7h 25m 28s	17h 17m 42s
UPIXAUMOL	7h 24m 3s	17h 20m 56s

Table E.3: Total wall times for the CSP searches carried out in Section 7.2 using 4 nodes on Iridis 5 with dual Intel® Xeon® Gold 6138 CPUs @ 2.00GHz on each node for the Z'=1 searches for 5,000 successfully minimised structures in space groups ($P2_1/c$, $P2_12_12_1$, $P\bar{1}$, $P2_1$, $Pbca$, $C2/c$, $Pna2_1$, Cc, $Pca2_1$, $C2$) and the Z'=2 searches for 10,000 successfully minimised structures in the space groups ($P\bar{1}$, $P2_1/c$, $P2_1$, $P2_12_12_1$, $P1$).

CSP	Z'=1	Z'=2
UPIXEYMOL	7h 31m 41s	17h 44m 36s
VIBMAVMOL	7h 54m 2s	16h 48m 1s
VIQTATMOL	7h 50m 25s	16h 36m 4s
VUFKALMOL	8h 13m 54s	16h 54m 11s
VUFNIWMOL	7h 58m 0s	17h 14m 16s
WOTZUDMOL	6h 49m 25s	14h 20m 49s
WUPYIQMOL	7h 34m 46s	16h 3m 11s
YOFROBMOL	8h 29m 42s	18h 46m 2s
ZETQUNMOL	7h 44m 7s	16h 12m 16s
ZZZOYCMOL	8h 44m 10s	18h 41m 17s

Table E.3: *continued*

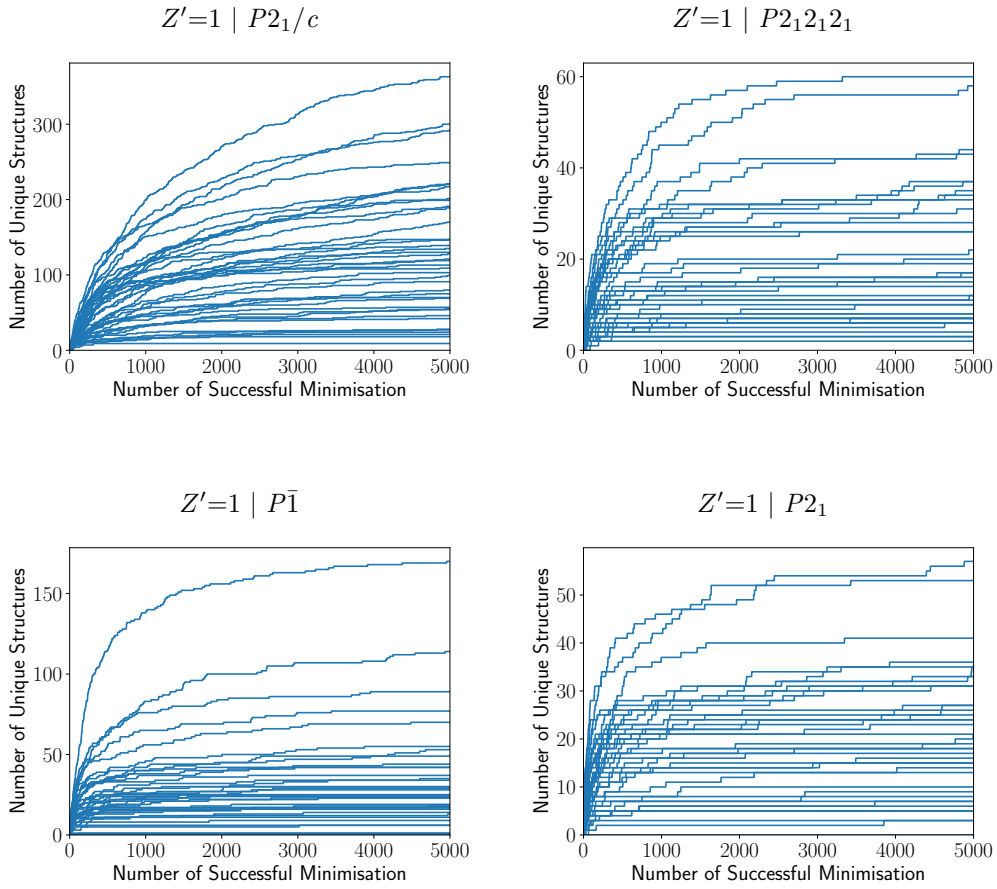


Figure E.22: The number of unique structures found in the CSP searches described in Section 7.2 within 12 kJ mol^{-1} from the global lattice energy minimum of the full search as a function of the number of successful minimisations for a given Z' | SG. Each line plot shows the progress of the CSP for a given molecule in the test set.

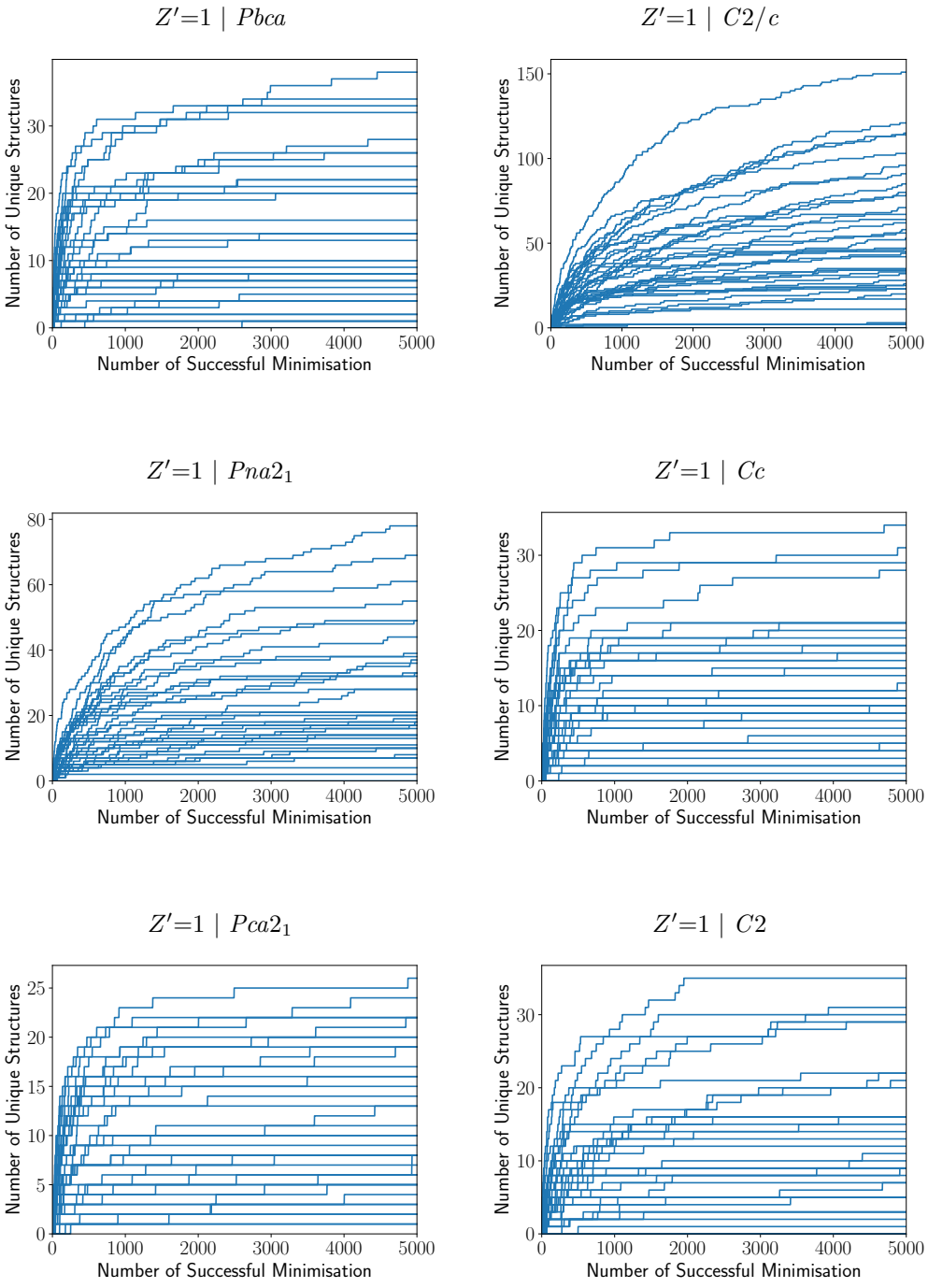


Figure E.22: *continued*

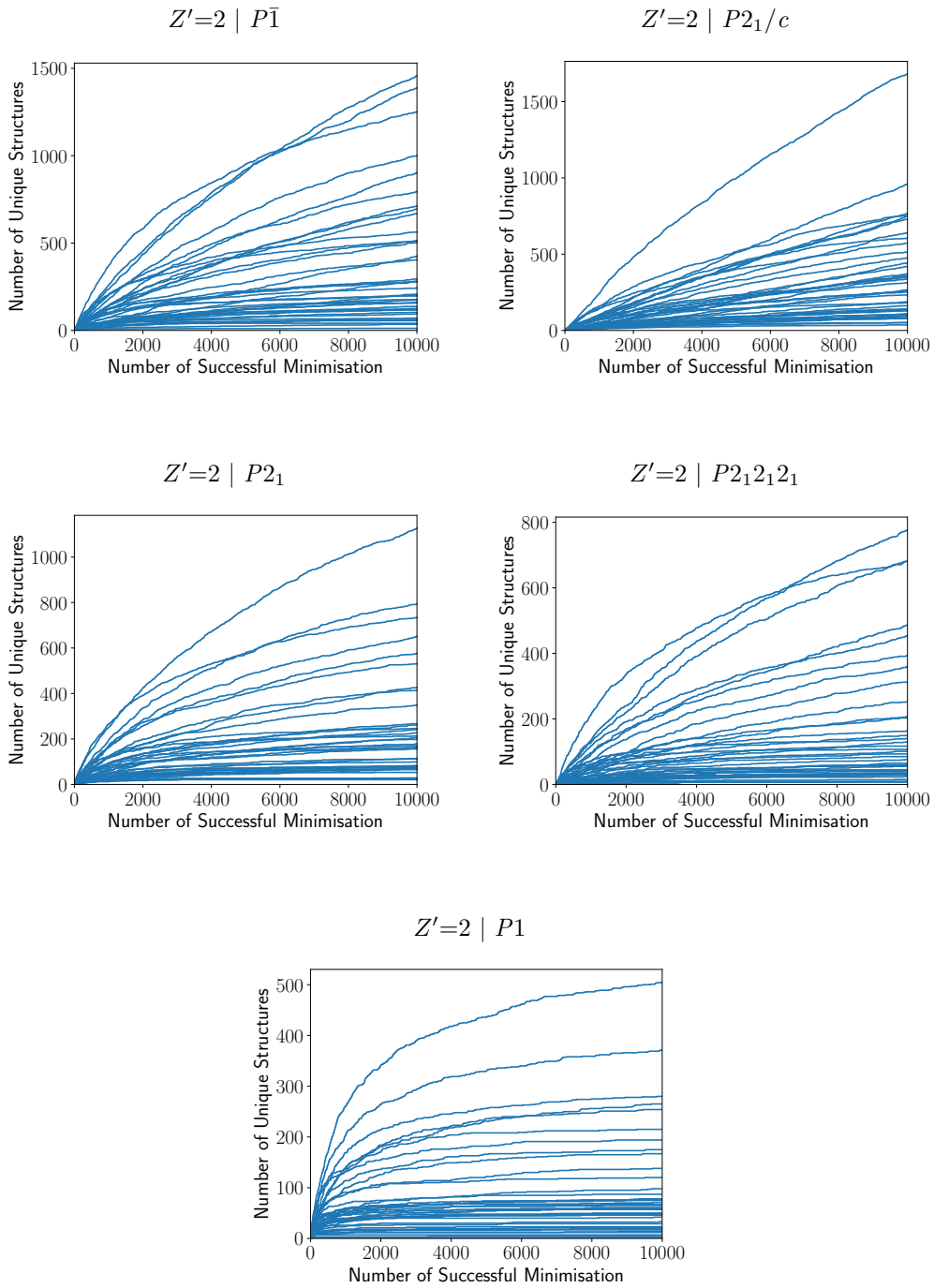


Figure E.22: *continued*

CSP	Reference	14	19	2	4	61	15	33	9	29	5
AROBUFMOL	GM	5	0	0	7	0	0	0	0	0	0
AROBUFMOL	AROBUF	0	0	249	0	0	0	0	0	0	0
AROCAMMOL	AROCAM	185	0	1092	423	0	146	0	134	0	214
BEGKIKMOL	BEGKIK	0	0	0	0	81	0	0	0	0	0
BEZLUOMOL	GM	15	0	0	74	0	0	0	0	0	0
BEZLUOMOL	BEZLUO	11	0	0	36	0	0	0	0	0	0
BNPYREMOL	GM	3	0	0	0	0	0	0	0	0	0
BNPYREMOL	BNPYRE10	25	0	0	0	0	0	0	0	0	0
BNZACRMOL	BNZACR	0	0	0	0	0	0	6	0	0	0
CEQGELMOL	GM	8	0	0	0	0	0	0	0	0	0
CEQGELMOL	CEQGEL	0	0	0	0	0	0	0	0	0	0
DBANQUMOL	DBANQU	0	0	0	0	0	0	0	0	21	0
DBNTHRMOL	DBNTHR01	0	274	0	0	0	0	0	0	445	0
DBNTHRMOL	DBNTHR10	0	0	0	112	0	0	0	0	0	0
DUXSASMOL	GM	0	0	1731	0	0	0	0	0	0	0
DUXSASMOL	DUXSAS	28	0	0	0	0	0	0	0	0	0
FANTRQMOL	GM	10	0	0	47	0	0	0	0	0	0
FANTRQMOL	FANTRQ10	0	0	0	0	0	0	0	0	0	0
HAMDUWMOL	HAMDUW	28	0	0	0	0	0	0	0	0	0
KONBAQMOL	KONBAQ	12	54	0	0	0	0	15	0	0	0
LOHVAHMOL	LOHVAH	0	21	0	0	0	0	0	0	0	0
MIVRUEMOL	GM	13	0	0	0	0	0	0	0	0	0
MIVRUEMOL	MIVRUE	0	87	0	0	0	0	0	0	0	0
MIVSESMOL	GM	0	0	0	0	0	0	0	0	0	0
MORRODMOL	MORROD	56	0	0	0	0	0	0	0	0	0

Table E.4: The number of times the reference crystal structures were sampled in the CSP searches described in Section 7.2 for a given space group with Z'=1.

CSP	Reference	14	19	2	4	61	15	33	9	29	5
NAAZASMOL	GM	2	0	0	36	0	0	0	0	0	0
NAAZASMOL	NAAZAS	1	0	0	21	0	0	0	0	0	0
NIJCEQMOL	GM	10	0	0	0	0	0	0	0	0	0
NIJCEQMOL	NIJCEQ	37	0	0	0	0	0	0	0	0	0
OFANTQMOL	OFANTQ	57	0	0	0	0	0	0	0	0	0
PAGMERMOL	GM	0	41	0	0	0	0	0	0	0	0
PAGMERMOL	PAGMER	34	0	0	0	0	0	0	0	0	0
PENCENMOL	PENCEN	0	0	0	0	0	0	0	0	0	0
PENCENMOL	PENCEN01	0	0	0	0	0	0	0	0	0	0
PENCENMOL	PENCEN10	10	0	0	29	0	0	0	0	0	0
PENTQUMOL	PENTQU	30	0	0	31	0	0	0	0	0	0
PENTQUMOL	PENTQU01	0	0	0	0	0	0	0	0	0	0
PENTQUMOL	PENTQU02	28	0	0	20	0	0	0	0	0	0
PERLENMOL	PERLEN01	250	0	0	0	0	0	0	0	0	0
PERLENMOL	PERLEN06	87	0	0	738	0	0	0	0	0	0
POVLUHMOL	GM	7	0	0	0	0	0	0	0	0	0
POVLUHMOL	POVLUH	0	0	0	0	0	32	0	0	0	0
RUNLAPMOL	GM	80	0	0	0	0	0	0	0	0	0
RUNLAPMOL	RUNLAP	40	0	0	0	0	0	0	0	0	0
SANQIIMOL	GM	0	0	0	0	0	0	0	0	0	0
SANQIIMOL	SANQII	0	0	0	0	0	0	0	0	0	0
SEVYEAMOL	SEVYEA	0	124	0	0	0	0	0	0	0	0
UPIRUIMOL	GM	6	0	0	0	0	0	0	0	0	0
UPIRUIMOL	UPIRUI	0	0	0	0	33	0	0	0	0	0
UPIXAUMOL	GM	2	0	0	0	0	0	0	0	0	0

Table E.4: *continued*

CSP	Reference	14	19	2	4	61	15	33	9	29	5
UPIXAUMOL	UPIXAU	0	0	0	20	0	0	0	0	0	0
UPIXEYMOL	GM	1	0	0	0	0	0	0	0	0	0
UPIXEYMOL	UPIXEY	0	0	0	0	0	0	0	0	18	0
VIBMAVMOL	GM	228	0	0	0	0	0	0	0	0	0
VIBMAVMOL	VIBMAV	94	0	0	0	0	0	0	0	0	0
VIQTATMOL	GM	54	0	0	0	0	0	0	0	0	0
VIQTATMOL	VIQTAT	0	244	0	0	0	0	0	0	0	0
VUFKALMOL	GM	57	0	0	55	0	0	0	0	0	0
VUFKALMOL	VUFKAL	22	0	0	0	0	0	0	0	0	0
VUFNIWMOL	GM	1	0	0	0	0	0	0	0	0	0
VUFNIWMOL	VUFNIW	17	0	0	0	0	0	0	0	0	0
WOTZUDMOL	GM	126	0	0	0	0	0	0	0	0	0
WOTZUDMOL	WOTZUD	193	0	0	0	0	0	0	0	0	0
WUPYIQMOL	WUPYIQ	74	0	0	104	0	0	0	0	0	0
YOFROBMOL	YOFROB	0	0	1207	0	0	0	0	0	0	0
ZETQUNMOL	GM	70	0	0	97	0	0	0	0	0	0
ZETQUNMOL	ZETQUN	0	0	0	0	0	0	0	0	0	0
ZZZOYCMOL	ZZZOYC01	0	0	0	227	0	0	0	0	0	0

Table E.4: *continued*

CSP	Reference	2	14	4	19	1
AROBUFMOL	GM	5	0	5	0	9
AROBUFMOL	AROBUF	56	0	0	0	131
AROCAMMOL	AROCAM	436	69	153	0	1234
BEGKIKMOL	BEGKIK	0	3	0	0	0
BEZLUOMOL	GM	64	2	20	0	425
BEZLUOMOL	BEZLUO	25	0	14	0	93
BNPYREMOL	GM	0	0	1	0	0
BNPYREMOL	BNPYRE10	3	1	19	0	0
BNZACRMOL	BNZACR	0	0	0	0	0
CEQGELMOL	GM	1	0	4	0	0
CEQGELMOL	CEQGEL	6	0	0	0	0
DBANQUMOL	DBANQU	0	0	0	0	0
DBNTHRMOL	DBNTHR01	0	0	24	0	0
DBNTHRMOL	DBNTHR10	0	0	11	0	11
DUXSASMOL	GM	343	0	0	0	378
DUXSASMOL	DUXSAS	8	6	9	0	0
FANTRQMOL	GM	24	1	11	0	87
FANTRQMOL	FANTRQ10	0	0	10	0	0
HAMDUWMOL	HAMDUW	0	0	0	0	0
KONBAQMOL	KONBAQ	0	3	14	0	0
LOHVAHMOL	LOHVAH	0	0	7	0	0
MIVRUEMOL	GM	0	0	0	0	0
MIVRUEMOL	MIVRUE	0	0	22	0	0
MIVSESMOL	GM	0	0	0	0	3
MORRODMOL	MORROD	20	10	25	0	0
NAAZASMOL	GM	8	0	0	0	22
NAAZASMOL	NAAZAS	5	1	1	0	22
NIJCEQMOL	GM	1	0	0	0	0
NIJCEQMOL	NIJCEQ	0	2	9	0	0
OFANTQMOL	OFANTQ	34	3	29	0	0
PAGMERMOL	GM	0	0	7	0	0
PAGMERMOL	PAGMER	7	3	14	0	0
PENCENMOL	PENCEN	69	0	0	0	585
PENCENMOL	PENCEN01	16	0	0	0	539
PENCENMOL	PENCEN10	0	1	12	0	0
PENTQUMOL	PENTQU	45	7	54	0	103
PENTQUMOL	PENTQU01	1	0	0	0	0
PENTQUMOL	PENTQU02	53	1	48	0	106
PERLENMOL	PERLEN01	45	22	157	0	0

Table E.5: The number of times the reference crystal structures were sampled in the CSP searches described in Section 7.2 for a given space group with Z'=2.

CSP	Reference	2	14	4	19	1
PERLENMOL	PERLEN06	49	16	125	0	356
POVLUHMOL	GM	1	1	3	0	0
POVLUHMOL	POVLUH	2	0	0	0	0
RUNLAPMOL	GM	25	2	24	0	0
RUNLAPMOL	RUNLAP	9	5	19	0	0
SANQIIMOL	GM	8	0	0	0	0
SANQIIMOL	SANQII	0	0	7	0	0
SEVYEAMOL	SEVYEA	0	0	19	0	0
UPIRUIMOL	GM	2	0	0	0	0
UPIRUIMOL	UPIRUI	0	0	0	0	0
UPIXAUMOL	GM	0	0	0	0	0
UPIXAUMOL	UPIXAU	0	0	5	0	30
UPIXEYMOL	GM	0	0	0	0	0
UPIXEYMOL	UPIXEY	0	0	0	0	0
VIBMAVMOL	GM	190	18	48	0	0
VIBMAVMOL	VIBMAV	57	15	87	0	0
VIQTATMOL	GM	6	4	23	0	0
VIQTATMOL	VIQTAT	0	0	38	0	0
VUFKALMOL	GM	39	6	43	0	51
VUFKALMOL	VUFKAL	0	10	0	0	0
VUFNIWMOL	GM	0	0	1	0	0
VUFNIWMOL	VUFNIW	0	2	4	0	0
WOTZUDMOL	GM	33	12	106	0	0
WOTZUDMOL	WOTZUD	38	23	80	0	0
WUPYIQMOL	WUPYIQ	233	8	79	0	440
YOFROBMOL	YOFROB	548	0	0	0	1357
ZETQUNMOL	GM	27	33	186	0	31
ZETQUNMOL	ZETQUN	11	0	0	0	600
ZZZOYCMOL	ZZZOYC01	0	0	52	0	233

Table E.5: *continued*

CSP	Reference	14	19	2	4	61	15	33	9	29	5
AROBUFMOL	GM	931	-	-	331	-	-	-	-	-	-
AROBUFMOL	AROBUF	-	-	2	-	-	-	-	-	-	-
AROCAMMOL	AROCAM	43	-	4	66	-	14	-	56	-	26
BEGKIKMOL	BEGKIK	-	-	-	-	22	-	-	-	-	-
BEZLUOMOL	GM	180	-	-	76	-	-	-	-	-	-
BEZLUOMOL	BEZLUO	502	-	-	59	-	-	-	-	-	-
BNPYREMOL	GM	2287	-	-	-	-	-	-	-	-	-
BNPYREMOL	BNPYRE10	87	-	-	-	-	-	-	-	-	-
BNZACRMOL	BNZACR	-	-	-	-	-	-	529	-	-	-
CEQGELMOL	GM	375	-	-	-	-	-	-	-	-	-
CEQGELMOL	CEQGEL	-	-	-	-	-	-	-	-	-	-
DBANQUMOL	DBANQU	-	-	-	-	-	-	-	-	29	-
DBNTHRMOL	DBNTHR01	-	64	-	-	-	-	-	-	9	-
DBNTHRMOL	DBNTHR10	-	-	-	38	-	-	-	-	-	-
DUXSASMOL	GM	-	-	1	-	-	-	-	-	-	-
DUXSASMOL	DUXSAS	58	-	-	-	-	-	-	-	-	-
FANTRQMOL	GM	203	-	-	22	-	-	-	-	-	-
FANTRQMOL	FANTRQ10	-	-	-	-	-	-	-	-	-	-
HAMDUWMOL	HAMDUW	398	-	-	-	-	-	-	-	-	-
KONBAQMOL	KONBAQ	242	104	-	-	-	-	186	-	-	-
LOHVAHMOL	LOHVAH	-	51	-	-	-	-	-	-	-	-
MIVRUEMOL	GM	319	-	-	-	-	-	-	-	-	-
MIVRUEMOL	MIVRUE	-	129	-	-	-	-	-	-	-	-
MIVSESMOL	GM	-	-	-	-	-	-	-	-	-	-
MORRODMOL	MORROD	50	-	-	-	-	-	-	-	-	-

Table E.6: The number of successful minimisations required for the CSP searches described in Section 7.2 to obtain the reference crystal structure in a given space group with Z'=1.

CSP	Reference	14	19	2	4	61	15	33	9	29	5
NAAZASMOL	GM	1405	-	-	271	-	-	-	-	-	-
NAAZASMOL	NAAZAS	3883	-	-	293	-	-	-	-	-	-
NIJCEQMOL	GM	630	-	-	-	-	-	-	-	-	-
NIJCEQMOL	NIJCEQ	326	-	-	-	-	-	-	-	-	-
OFANTQMOL	OFANTQ	312	-	-	-	-	-	-	-	-	-
PAGMERMOL	GM	-	143	-	-	-	-	-	-	-	-
PAGMERMOL	PAGMER	165	-	-	-	-	-	-	-	-	-
PENCENMOL	PENCEN	-	-	-	-	-	-	-	-	-	-
PENCENMOL	PENCEN01	-	-	-	-	-	-	-	-	-	-
PENCENMOL	PENCEN10	561	-	-	170	-	-	-	-	-	-
PENTQUMOL	PENTQU	73	-	-	141	-	-	-	-	-	-
PENCENMOL	PENCEN01	-	-	-	-	-	-	-	-	-	-
PENTQUMOL	PENTQU02	236	-	-	248	-	-	-	-	-	-
PERLENMOL	PERLEN01	24	-	-	-	-	-	-	-	-	-
PERLENMOL	PERLEN06	63	-	-	1	-	-	-	-	-	-
POVLUHMOL	GM	98	-	-	-	-	-	-	-	-	-
POVLUHMOL	POVLUH	-	-	-	-	-	17	-	-	-	-
RUNLAPMOL	GM	36	-	-	-	-	-	-	-	-	-
RUNLAPMOL	RUNLAP	139	-	-	-	-	-	-	-	-	-
SANQIIMOL	GM	-	-	-	-	-	-	-	-	-	-
SANQIIMOL	SANQII	-	-	-	-	-	-	-	-	-	-
SEVYEAMOL	SEVYEA	-	108	-	-	-	-	-	-	-	-
UPIRUIMOL	GM	926	-	-	-	-	-	-	-	-	-
UPIRUIMOL	UPIRUI	-	-	-	-	243	-	-	-	-	-
UPIXAUMOL	GM	821	-	-	-	-	-	-	-	-	-

Table E.6: *continued*

CSP	Reference	14	19	2	4	61	15	33	9	29	5
UPIXAUMOL	UPIXAU	-	-	-	66	-	-	-	-	-	-
UPIXEYMOL	GM	2148	-	-	-	-	-	-	-	-	-
UPIXEYMOL	UPIXEY	-	-	-	-	-	-	-	-	87	-
VIBMAVMOL	GM	132	-	-	-	-	-	-	-	-	-
VIBMAVMOL	VIBMAV	72	-	-	-	-	-	-	-	-	-
VIQTATMOL	GM	62	-	-	-	-	-	-	-	-	-
VIQTATMOL	VIQTAT	-	54	-	-	-	-	-	-	-	-
VUFKALMOL	GM	180	-	-	54	-	-	-	-	-	-
VUFKALMOL	VUFKAL	99	-	-	-	-	-	-	-	-	-
VUFNIWMOL	GM	2429	-	-	-	-	-	-	-	-	-
VUFNIWMOL	VUFNIW	519	-	-	-	-	-	-	-	-	-
WOTZUDMOL	GM	2	-	-	-	-	-	-	-	-	-
WOTZUDMOL	WOTZUD	1	-	-	-	-	-	-	-	-	-
WUPYIQMOL	WUPYIQ	44	-	-	25	-	-	-	-	-	-
YOFROBMOL	YOFROB	-	-	3	-	-	-	-	-	-	-
ZETQUNMOL	GM	29	-	-	165	-	-	-	-	-	-
ZETQUNMOL	ZETQUN	-	-	-	-	-	-	-	-	-	-
ZZZOYCMOL	ZZZOYC01	-	-	-	26	-	-	-	-	-	-

Table E.6: *continued*

CSP	Reference	2	14	4	19	1
AROBUFMOL	GM	2164	-	718	-	1622
AROBUFMOL	AROBUF	93	-	-	-	1
AROCAMMOL	AROCAM	21	1	-	-	3
BEGKIKMOL	BEGKIK	-	2064	-	-	-
BEZLUOMOL	GM	58	6	1402	-	7
BEZLUOMOL	BEZLUO	384	-	2116	-	129
BNPYREMOL	GM	-	-	5430	-	-
BNPYREMOL	BNPYRE10	3203	545	132	-	-
BNZACRMOL	BNZACR	-	-	-	-	-
CEQGELMOL	GM	3087	-	385	-	-
CEQGELMOL	CEQGEL	3452	-	-	-	-
DBANQUMOL	DBANQU	-	-	-	-	-
DBNTHRMOL	DBNTHR01	-	-	586	-	-
DBNTHRMOL	DBNTHR10	-	-	176	-	26
DUXSASMOL	GM	-	-	-	-	18
DUXSASMOL	DUXSAS	547	676	1087	-	-
FANTRQMOL	GM	152	687	1108	-	26
FANTRQMOL	FANTRQ10	-	-	1765	-	-
HAMDUWMOL	HAMDUW	-	-	-	-	-
KONBAQMOL	KONBAQ	-	14	1310	-	-
LOHVAHMOL	LOHVAH	-	-	1602	-	-
MIVRUEMOL	GM	-	-	-	-	-
MIVRUEMOL	MIVRUE	-	-	652	-	-
MIVSESMOL	GM	-	-	-	-	5409
MORRODMOL	MORROD	384	263	252	-	-
NAAZASMOL	GM	1590	-	-	-	30
NAAZASMOL	NAAZAS	1398	5018	7910	-	2125
NIJCEQMOL	GM	5356	-	-	-	-
NIJCEQMOL	NIJCEQ	-	5173	900	-	-
OFANTQMOL	OFANTQ	92	1910	391	-	-
PAGMERMOL	GM	-	-	3360	-	-
PAGMERMOL	PAGMER	483	2083	1258	-	-
PENCENMOL	PENCEN	128	-	-	-	-
PENCENMOL	PENCEN01	450	-	-	-	11
PENCENMOL	PENCEN10	-	1722	507	-	-
PENTQUMOL	PENTQU	450	932	206	-	68
PENTQUMOL	PENTQU01	9351	-	-	-	-
PENTQUMOL	PENTQU02	287	2546	115	-	55
PERLENMOL	PERLEN01	278	642	47	-	-

Table E.7: The minimum number of successful minimisations required for the CSP searches described in Section 7.2 to obtain the reference crystal structure in a given space group with Z'=2.

CSP	Reference	2	14	4	19	1
PERLENMOL	PERLEN06	29	556	36	-	12
POVLUHMOL	GM	8726	6789	769	-	-
POVLUHMOL	POVLUH	1301	-	-	-	-
RUNLAPMOL	GM	173	6372	192	-	-
RUNLAPMOL	RUNLAP	2320	132	1101	-	-
SANQIIMOL	GM	351	-	-	-	-
SANQIIMOL	SANQII	-	-	278	-	-
SEVYEAMOL	SEVYEA	-	-	317	-	-
UPIRUIMOL	GM	516	-	-	-	-
UPIRUIMOL	UPIRUI	-	-	-	-	-
UPIXAUMOL	GM	-	-	-	-	-
UPIXAUMOL	UPIXAU	-	-	2473	-	19
UPIXEYMOL	GM	-	-	-	-	-
UPIXEYMOL	UPIXEY	-	-	-	-	-
VIBMAVMOL	GM	75	874	41	-	-
VIBMAVMOL	VIBMAV	3	139	26	-	-
VIQTATMOL	GM	1373	437	487	-	-
VIQTATMOL	VIQTAT	-	-	206	-	-
VUFKALMOL	GM	145	3014	102	-	93
VUFKALMOL	VUFKAL	-	414	-	-	-
VUFNIWMOL	GM	-	-	1261	-	-
VUFNIWMOL	VUFNIW	-	4240	14	-	-
WOTZUDMOL	GM	814	237	61	-	-
WOTZUDMOL	WOTZUD	175	25	31	-	-
WUPYIQMOL	WUPYIQ	21	1178	26	-	31
YOFROBMOL	YOFROB	9	-	-	-	3
ZETQUNMOL	GM	2252	78	1	-	148
ZETQUNMOL	ZETQUN	2293	-	-	-	47
ZZZOYCMOL	ZZZOYC01	-	-	34	-	8

Table E.7: *continued*

E.3 CSP Sampling II

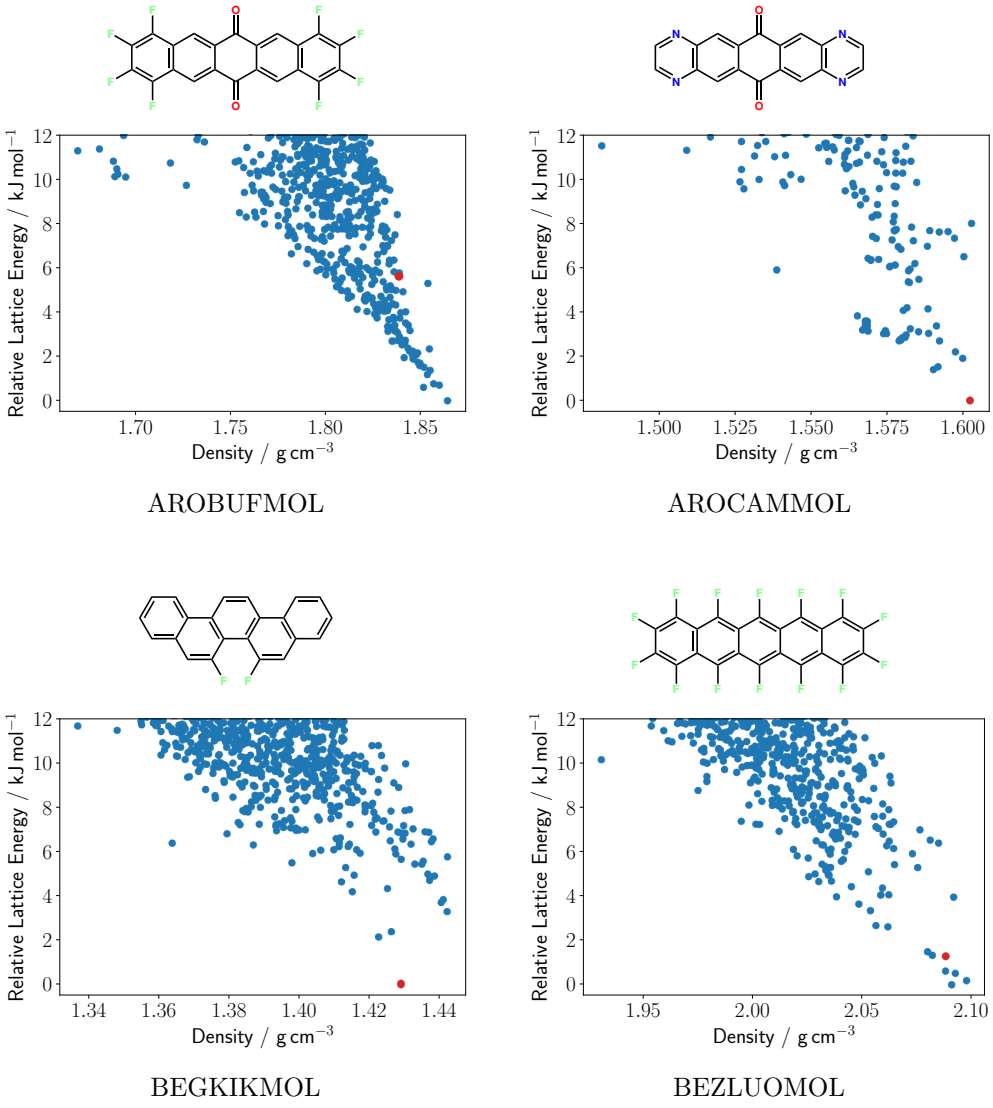
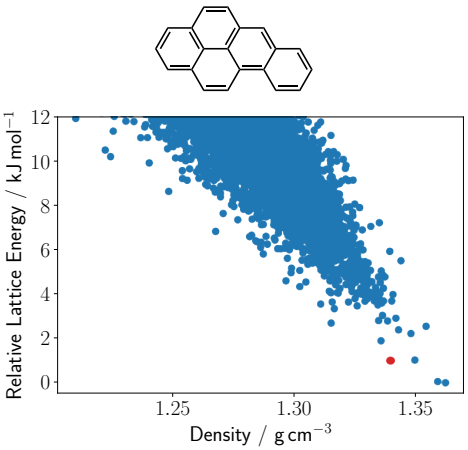
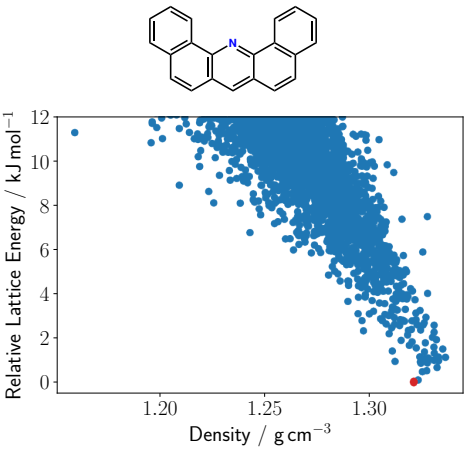


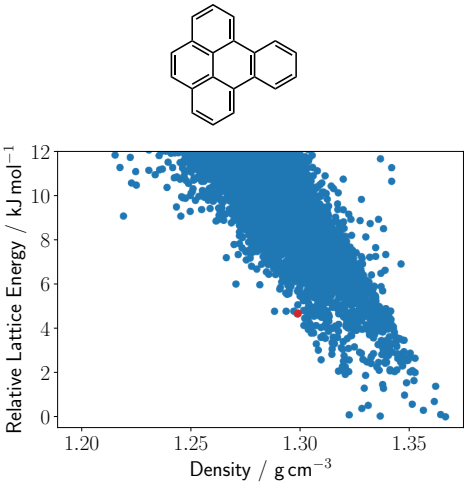
Figure E.23: Energy-density plots generated from the CSP searches described in Section 7.3 for all 40 molecules used to validate the potential model and generate data for the development of a reduced sampling scheme. Plot points coloured in red indicate matches with experimental crystal structures from the CSD.



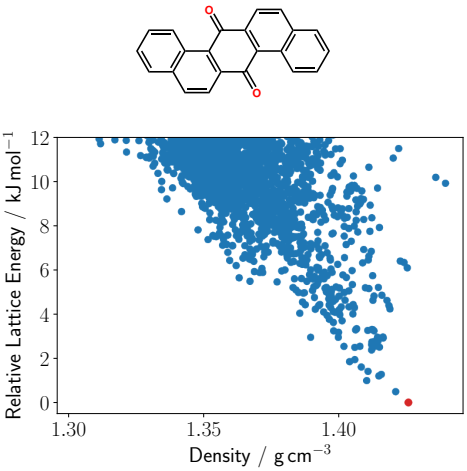
BNPYREMOL



BNZACRMOL

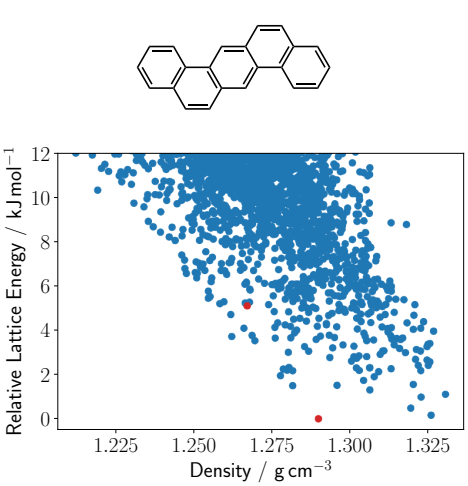


CEQGELMOL

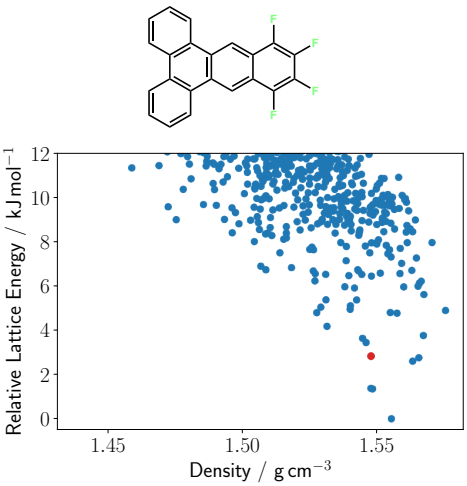


DBANQUMOL

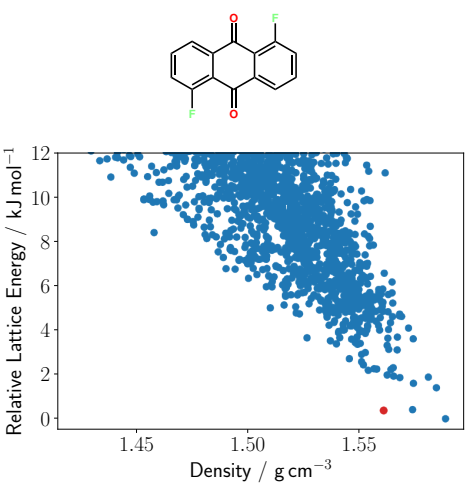
Figure E.23: *continued*



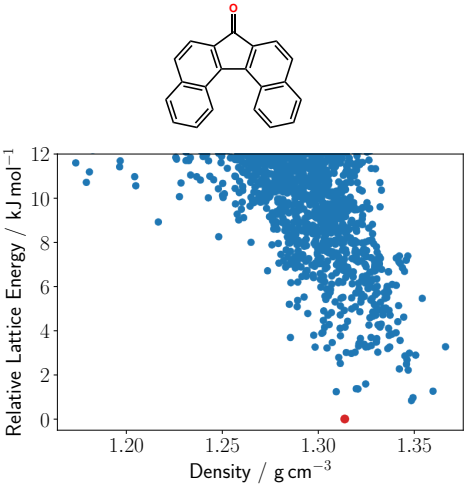
DBNTHRMOL



DUXSASMOL

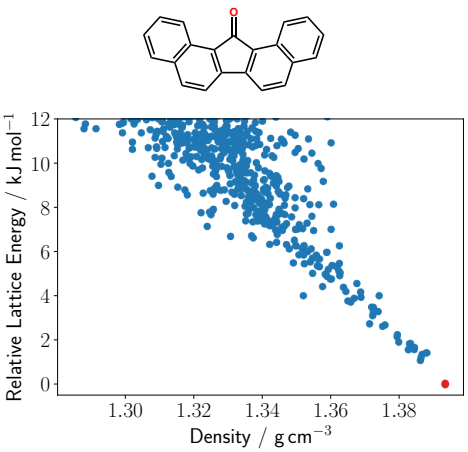


FANTRQMOL

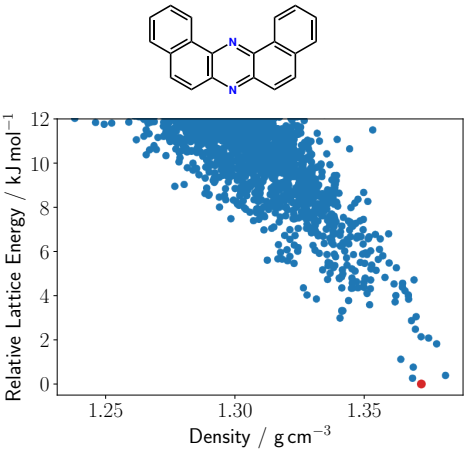


HAMDUWMOL

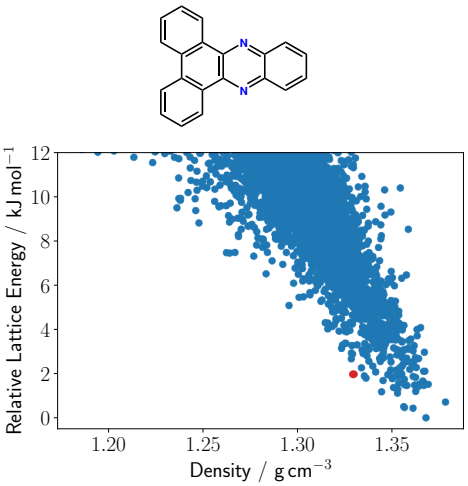
Figure E.23: *continued*



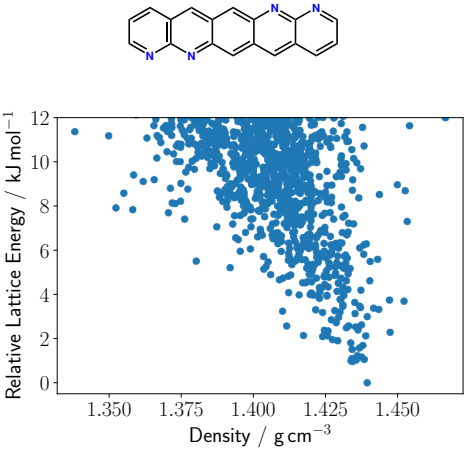
KONBAQMOL



LOHVAHMOL

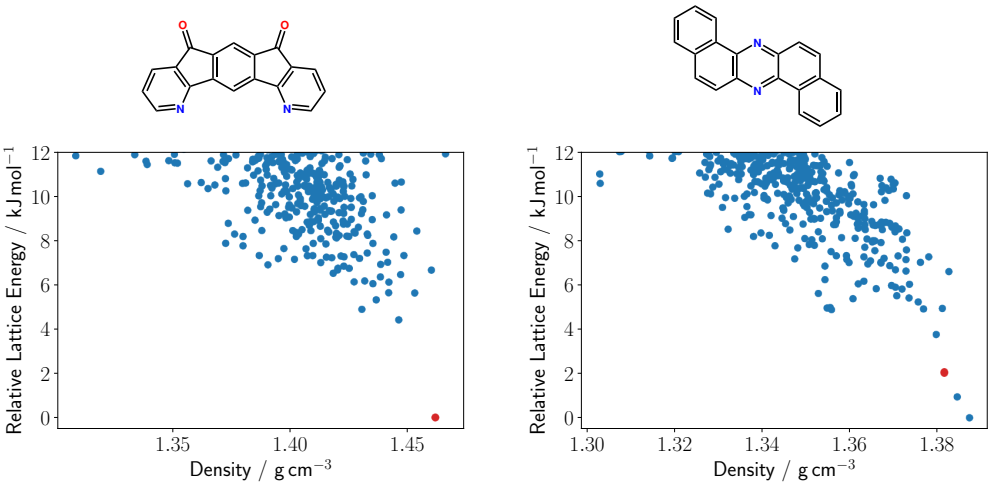


MIVRUEMOL

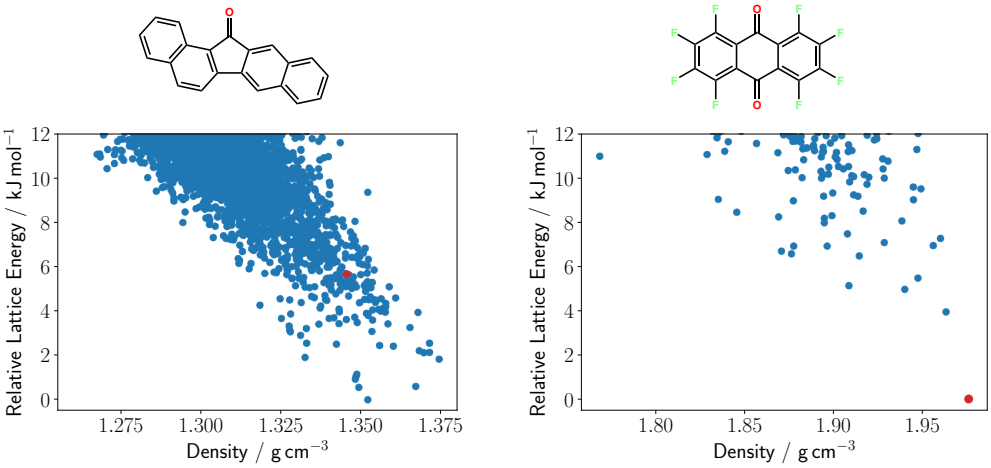


MIVSESMOL

Figure E.23: *continued*

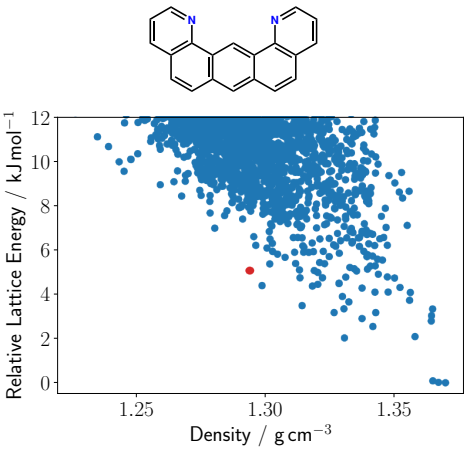


MORRODMOL NAAZASMOL

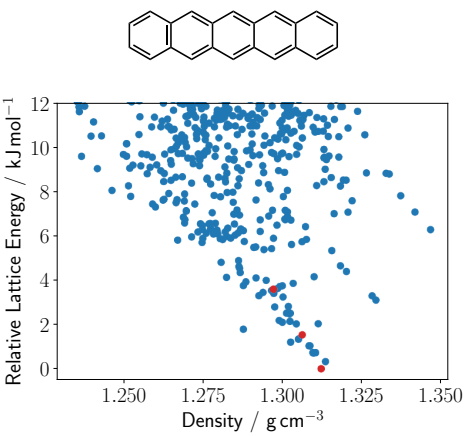


NIJCEQMOL OFANTQMOL

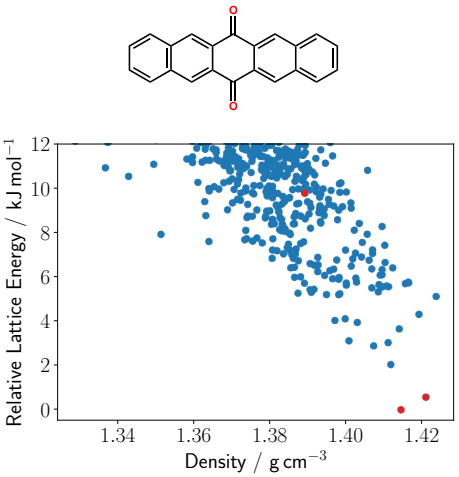
Figure E.23: *continued*



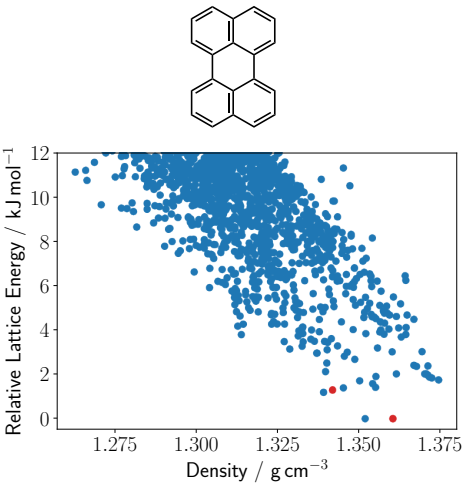
PAGMERMOL



PENCENMOL

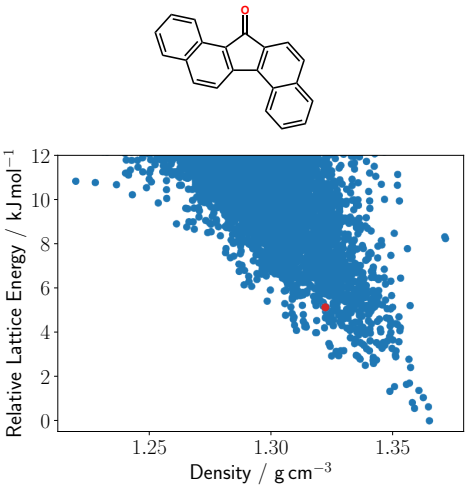


PENTQUMOL

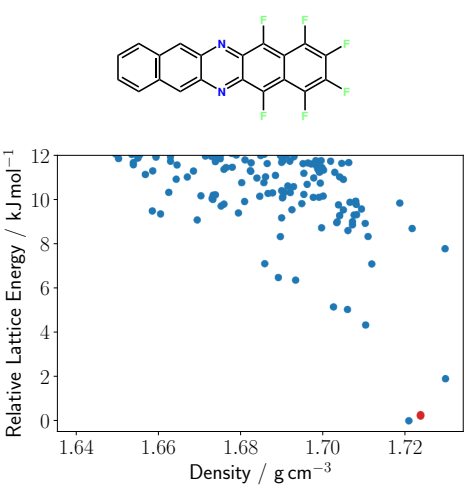


PERLENMOL

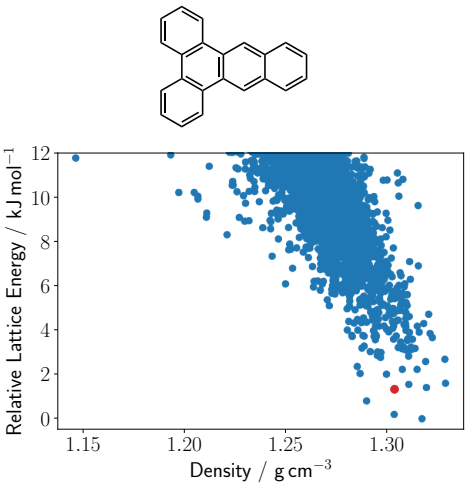
Figure E.23: *continued*



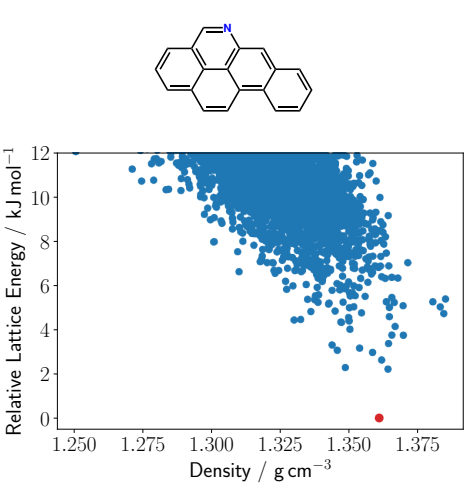
POVLUHMOL



RUNLAPMOL

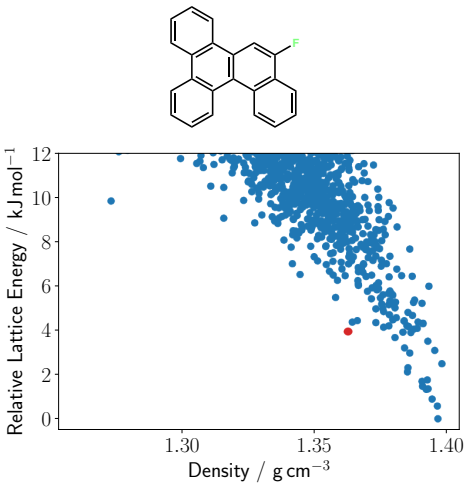


SANQIIMOL

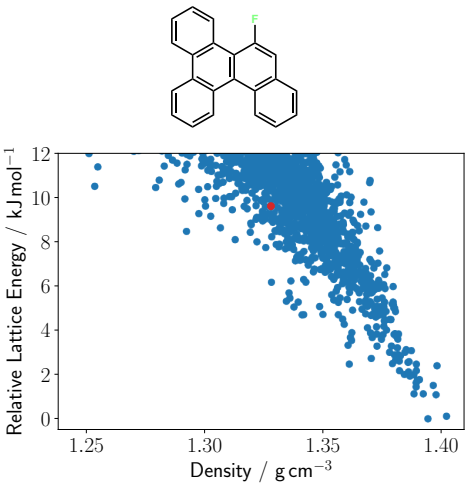


SEVYEAMOL

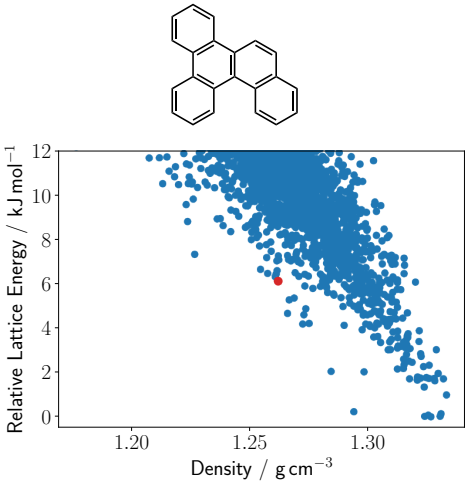
Figure E.23: *continued*



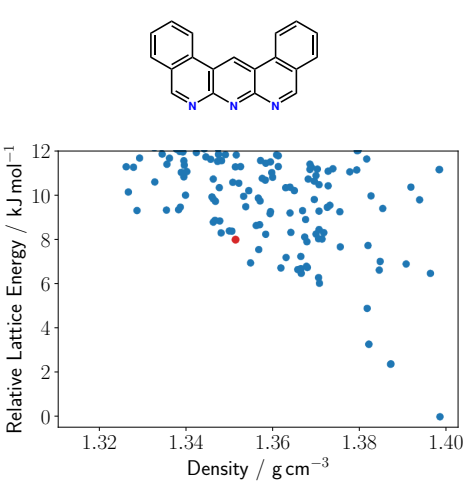
UPIRUIMOL



UPIXAUMOL

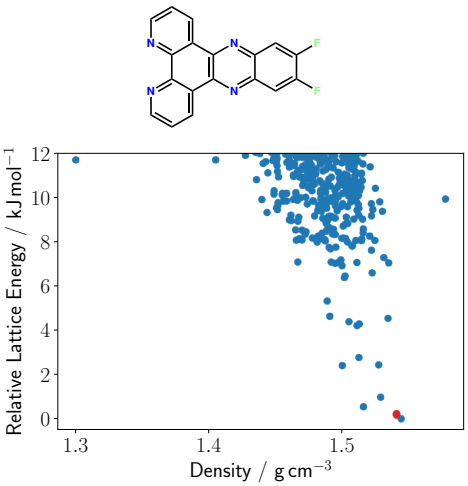


UPIXEYMOL

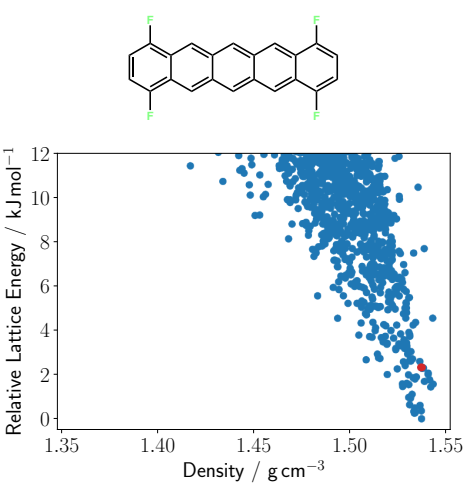


VIBMAVMOL

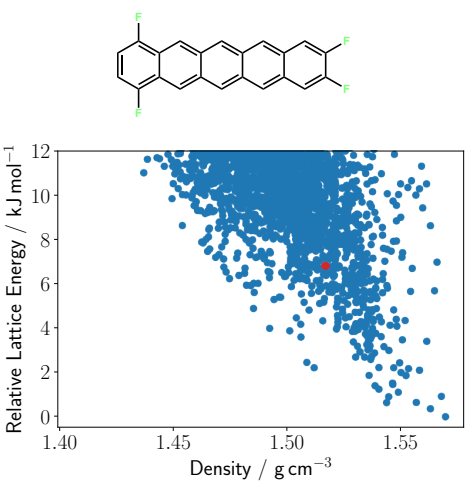
Figure E.23: *continued*



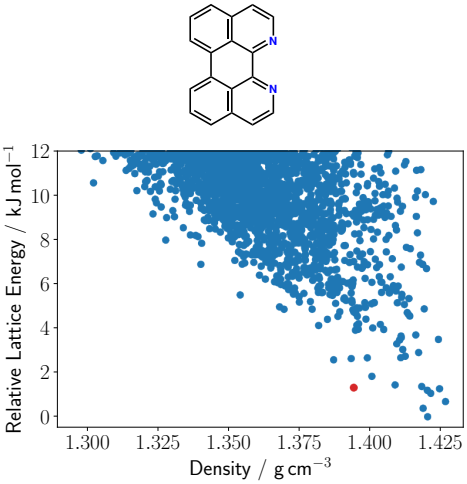
VIQTATMOL



VUFKALMOL

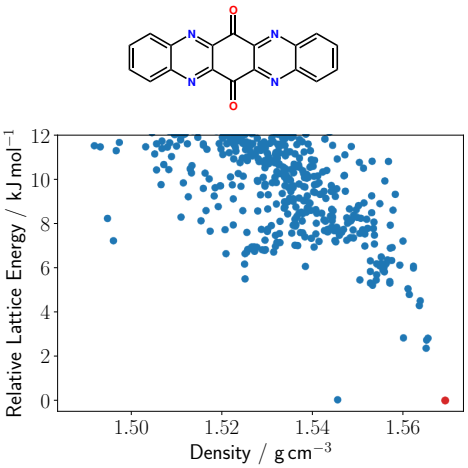


VUFNIWMOL

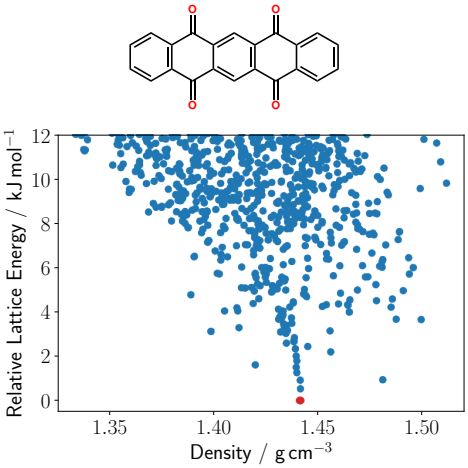


WOTZUDMOL

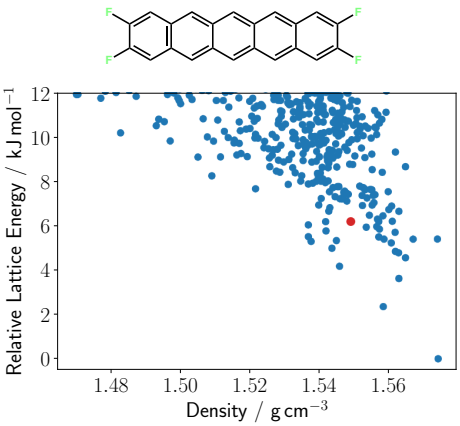
Figure E.23: *continued*



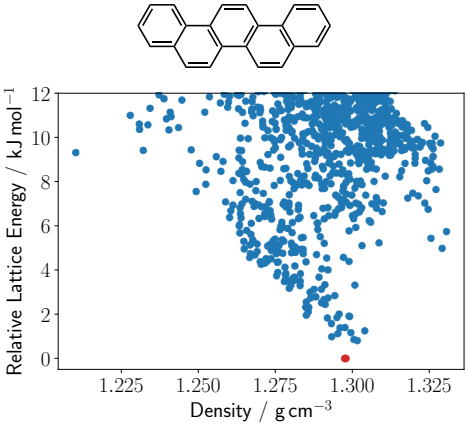
WUPYIQMOL



YOFROBMOL



ZETQUNMOL



ZZZOYCMOL

Figure E.23: *continued*

Predicted	CSD Refcode	RMSD30	RMSD60	PXRD	ΔE
AROBUF_optx2-QR-1-88962-3	AROBUF	0.458	0.641	0.956	5.627
AROCAM_opt-QR-2-12929-3	AROCAM	0.145	0.177	0.976	0.000
BEGKIK_opt-QR-61-7745-3	BEGKIK	0.179	0.202	0.993	0.000
BEZLUO_opt-QR-4-7633-3	BEZLUO	0.327	0.472	0.931	1.261
BNPYRE_optx2-QR-14-2060-3	BNPYRE10	0.281	0.352	0.989	0.968
BNPYRE_optx2-QR-14-2060-3	BNPYRE11	0.289	0.373	0.986	0.968
BNPYRE_optx2-QR-14-2060-3	BNPYRE12	0.297	0.356	0.987	0.968
BNZACR_opt-QR-33-2364-3	BNZACR	0.283	0.363	0.972	0.000
CEQGEL_optx2-QR-2-32200-3	CEQGEL	0.334	0.417	0.985	4.665
CEQGEL_optx2-QR-2-32200-3	CEQGEL01	0.234	0.282	0.978	4.665
DBANQU_opt-QR-29-2442-3	DBANQU	0.436	0.490	0.968	0.000
DBNTHR_opt-QR-19-4321-3	DBNTHR01	0.551	0.689	0.957	0.000
DBNTHR_opt-QR-19-4321-3	DBNTHR02	0.555	0.691	0.957	0.000
DBNTHR_opt-QR-4-4352-3	DBNTHR10	0.292	0.304	0.987	5.091
DUXSAS_optx2-QR-14-13217-3	DUXSAS	-	-	0.985	2.830
FANTRQ_optx2-QR-4-11713-3	FANTRQ10	0.186	0.230	0.986	0.347
HAMDUW_opt-QR-14-1087-3	HAMDUW	0.399	0.502	0.968	0.000
HAMDUW_opt-QR-14-1087-3	HAMDUW01	0.385	0.484	0.970	0.000
HAMDUW_opt-QR-14-1087-3	HAMDUW02	0.465	0.578	0.949	0.000
KONBAQ_opt-QR-33-6642-3	KONBAQ	0.091	0.123	0.994	0.000
LOHVAH_optx2-QR-4-19292-3	LOHVAH	0.227	0.304	0.986	0.000
MIVRUE_opt-QR-19-169-3	MIVRUE	0.458	0.516	0.975	1.956

Table E.8: COMPACK RMSDs for a 30 out of 30 and a 60 out of 60 cluster with distance and angles tolerances of 30% and 30° or smaller and hydrogen positions ignored, X-ray similarities between the matched predicted crystal structures from the CSP searches described in Section 7.3 against experimental structures, and the relative lattice energies of the matched predicted structures. Matches to the experimental structure were not obtained for MIVSES and a full 30 out of 30 or 60 out of 60 match in COMPACK were not obtained for the experimental structures of DUXSAS, PENTQU01, and VUFNIW.

Predicted	CSD Refcode	RMSD30	RMSD60	PXRD	ΔE
-	MIVSES	-	-	-	-
MORROD_optx2-QR-14-30352-3	MORROD	0.222	0.248	0.991	0.000
NAAZAS_optx2-QR-14-23290-3	NAAZAS	0.150	0.182	0.991	2.034
NIJCEQ_opt-QR-14-7929-3	NIJCEQ	1.127	1.328	0.951	5.646
OFANTQ_opt-QR-14-5875-3	OFANTQ	0.170	0.201	0.990	0.000
OFANTQ_opt-QR-14-5875-3	OFANTQ01	0.229	0.271	0.951	0.000
PAGMER_optx2-QR-14-10607-3	PAGMER	0.184	0.209	0.993	5.073
PENCEN_optx2-QR-1-107678-3	PENCEN	0.345	0.457	0.967	0.000
PENCEN_optx2-QR-1-244620-3	PENCEN01	0.562	0.710	0.929	1.529
PENCEN_optx2-QR-1-244620-3	PENCEN02	0.508	0.656	0.948	1.529
PENCEN_optx2-QR-1-244620-3	PENCEN03	0.514	0.638	0.947	1.529
PENCEN_optx2-QR-1-244620-3	PENCEN04	0.601	0.762	0.901	1.529
PENCEN_optx2-QR-1-107678-3	PENCEN05	0.302	0.408	0.965	0.000
PENCEN_optx2-QR-1-244620-3	PENCEN06	0.590	0.744	0.919	1.529
PENCEN_optx2-QR-1-244620-3	PENCEN07	0.501	0.631	0.957	1.529
PENCEN_optx2-QR-1-244620-3	PENCEN08	0.441	0.589	0.977	1.529
PENCEN_optx2-QR-1-107678-3	PENCEN09	0.304	0.417	0.964	0.000
PENCEN_optx2-QR-4-18660-3	PENCEN10	0.241	0.316	0.977	3.583
PENCEN_optx2-QR-4-18660-3	PENCEN11	0.252	0.332	0.975	3.583
PENCEN_optx2-QR-4-18660-3	PENCEN12	0.247	0.324	0.975	3.583
PENTQU_optx2-QR-2-48396-3	PENTQU	0.165	0.209	0.980	0.000
PENTQU_optx2-QR-2-70612-3	PENTQU01	-	-	0.913	9.77
PENTQU_optx2-QR-1-52560-3	PENTQU02	0.575	0.659	0.924	0.569
PERLEN_optx2-QR-2-51571-3	PERLEN01	0.273	0.347	0.991	1.266
PERLEN_optx2-QR-2-51571-3	PERLEN03	0.271	0.345	0.989	1.266

Table E.8: *continued*

Predicted	CSD Refcode	RMSD30	RMSD60	PXRD	ΔE
PERLEN_optx2-QR-2-51571-3	PERLEN04	0.267	0.340	0.991	1.266
PERLEN_optx2-QR-2-51571-3	PERLEN05	0.299	0.385	0.988	1.266
PERLEN_opt-QR-4-2473-3	PERLEN06	0.631	0.842	0.916	0.000
PERLEN_opt-QR-4-2473-3	PERLEN07	0.631	0.833	0.908	0.000
PERLEN_optx2-QR-2-51571-3	PERLEN08	0.291	0.373	0.990	1.266
POVLUH_opt-QR-15-11115-3	POVLUH	0.649	0.812	0.965	5.119
RUNLAP_opt-QR-14-3143-3	RUNLAP	0.397	0.457	0.966	0.230
SANQII_optx2-QR-4-32846-3	SANQII	0.458	0.596	0.982	1.311
SEVYEA_opt-QR-19-1708-3	SEVYEA	0.223	0.289	0.985	0.000
UPIRUI_opt-QR-61-6185-3	UPIRUI	0.206	0.234	0.988	3.928
UPIXAU_optx2-QR-4-11875-3	UPIXAU	1.060	1.349	0.910	9.611
UPIXEY_opt-QR-33-2353-3	UPIXEY	0.337	0.409	0.980	6.114
VIBMAV_opt-QR-14-559-3	VIBMAV	0.307	0.385	0.977	7.993
VIQTAT_opt-QR-19-4689-3	VIQTAT	0.364	0.426	0.977	0.195
VUFKAL_opt-QR-14-7845-3	VUFKAL	0.684	0.816	0.935	2.299
VUFNIW_opt-QR-14-7090-3	VUFNIW	-	-	0.944	6.804
WOTZUD_opt-QR-14-5639-3	WOTZUD	0.249	0.309	0.972	1.280
WUPYIQ_optx2-QR-2-41353-3	WUPYIQ	0.274	0.338	0.983	0.000
YOFROB_optx2-QR-1-16182-3	YOFROB	0.348	0.513	0.940	0.000
YOFROB_optx2-QR-1-16182-3	YOFROB01	0.289	0.418	0.966	0.000
ZETQUN_optx2-QR-1-243541-3	ZETQUN	0.262	0.340	0.947	6.206
ZZZOYC_opt-QR-4-2556-3	ZZZOYC01	0.386	0.497	0.946	0.000
ZZZOYC_opt-QR-4-2556-3	ZZZOYC02	0.427	0.536	0.920	0.000
ZZZOYC_opt-QR-4-2556-3	ZZZOYC03	0.386	0.487	0.929	0.000
ZZZOYC_opt-QR-4-2556-3	ZZZOYC04	0.446	0.580	0.940	0.000

Table E.8: *continued*

CSP	Z'=1	Z'=2
AROBUFMOL	12h 29m 23s	20h 43m 59s
AROCAMMOL	11h 27m 59s	18h 42m 56s
BEGKIKMOL	9h 47m 16s	17h 36m 26s
BEZLUOMOL	13h 0m 43s	20h 18m 33s
BNPYREMOL	9h 41m 47s	16h 28m 26s
BNZACRMOL	10h 18m 38s	17h 42m 5s
CEQGELMOL	9h 33m 45s	16h 24m 59s
DBANQUMOL	12h 4m 23s	19h 55m 34s
DBNTHRMOL	10h 4m 53s	18h 17m 2s
DUXSASMOL	10h 24m 16s	18h 0m 2s
FANTRQMOL	6h 57m 26s	12h 20m 17s
HAMDUWMOL	9h 6m 48s	16h 43m 18s
KONBAQMOL	10h 23m 5s	18h 4m 26s
LOHVAHMOL	10h 23m 51s	17h 38m 25s
MIVRUEMOL	10h 34m 48s	17h 31m 24s
MIVSESMOL	10h 32m 21s	17h 18m 24s
MORRODMOL	8h 36m 19s	15h 0m 3s
NAAZASMOL	10h 19m 43s	17h 29m 51s
NIJCEQMOL	10h 32m 28s	17h 52m 20s
OFANTQMOL	7h 2m 47s	10h 57m 46s
PAGMERMOL	9h 44m 36s	16h 37m 51s
PENCENMOL	10h 25m 20s	17h 59m 49s
PENTQUMOL	10h 43m 28s	18h 52m 21s
PERLENMOL	9h 17m 22s	15h 48m 42s
POVLUHMOL	10h 26m 47s	17h 49m 13s
RUNLAPMOL	10h 23m 45s	18h 11m 11s
SANQIIMOL	10h 7m 45s	17h 57m 20s
SEVYEAMOL	9h 24m 50s	16h 9m 37s
UPIRUIMOL	9h 18m 50s	17h 49m 12s
UPIXAUMOL	9h 13m 48s	17h 53m 8s

Table E.9: Total wall times for the CSP searches carried out in Section 7.3 using 4 nodes on Iridis 5 with dual Intel® Xeon® Gold 6138 CPUs @ 2.00GHz on each node for the Z'=1 searches for 5,000 successfully minimised structures in the space groups ($P2_1/c$, $P2_12_12_1$, $P\bar{1}$, $P2_1$, $Pbca$, $C2/c$, $Pna2_1$, Cc, $Pca2_1$, $C2$) and the Z'=2 searches for 10,000 successfully minimised structures in the space groups ($P\bar{1}$, $P2_1/c$, $P2_1$, $P2_12_12_1$, $P1$).

CSP	Z'=1	Z'=2
UPIXEYMOL	9h 15m 19s	17h 54m 5s
VIBMAVMOL	10h 43m 22s	17h 22m 39s
VIQTATMOL	11h 17m 55s	17h 22m 52s
VUFKALMOL	11h 13m 26s	17h 51m 56s
VUFNIWMOL	10h 52m 2s	18h 12m 38s
WOTZUDMOL	9h 26m 39s	14h 49m 11s
WUPYIQMOL	10h 9m 0s	17h 14m 53s
YOFROBMOL	11h 23m 53s	19h 46m 30s
ZETQUNMOL	10h 17m 25s	16h 55m 57s
ZZZOYCMOL	11h 26m 16s	19h 23m 34s

Table E.9: *continued*

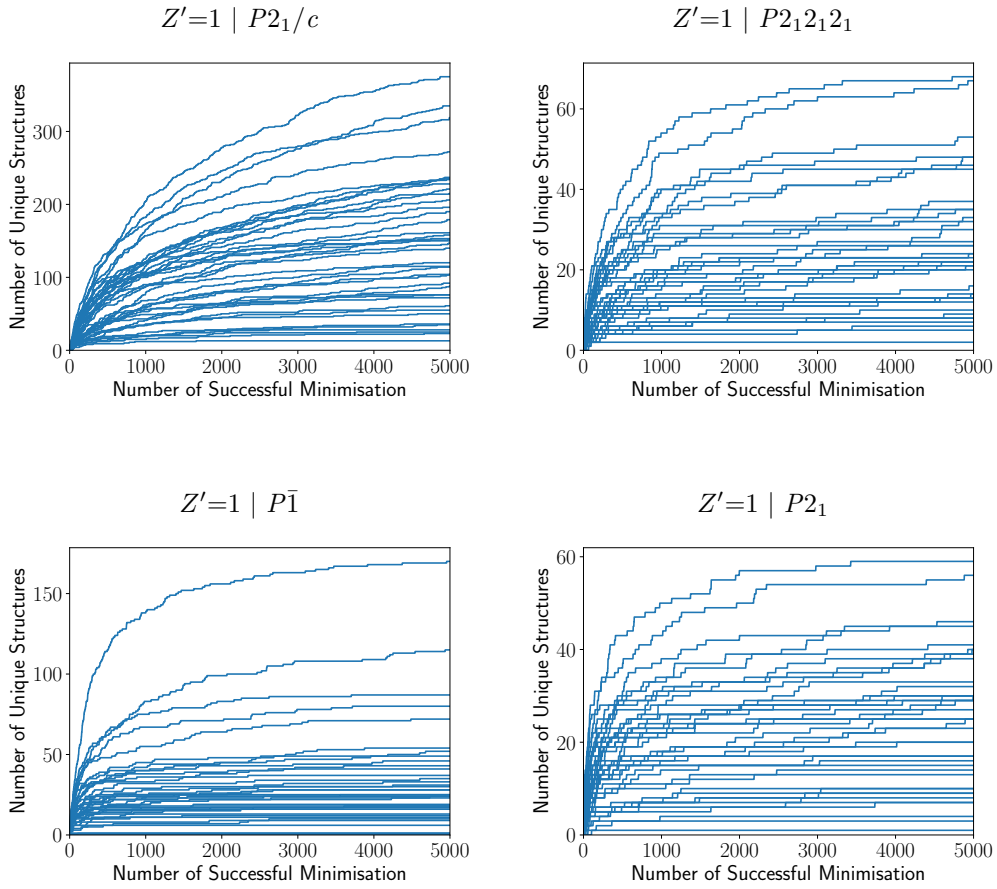


Figure E.24: The number of unique structures found in the CSP searches described in Section 7.3 within 12 kJ mol^{-1} from the global lattice energy minimum of the full search as a function of the number of successful minimisations for a given Z' | SG. Each line plot shows the progress of the CSP for a given molecule in the test set.

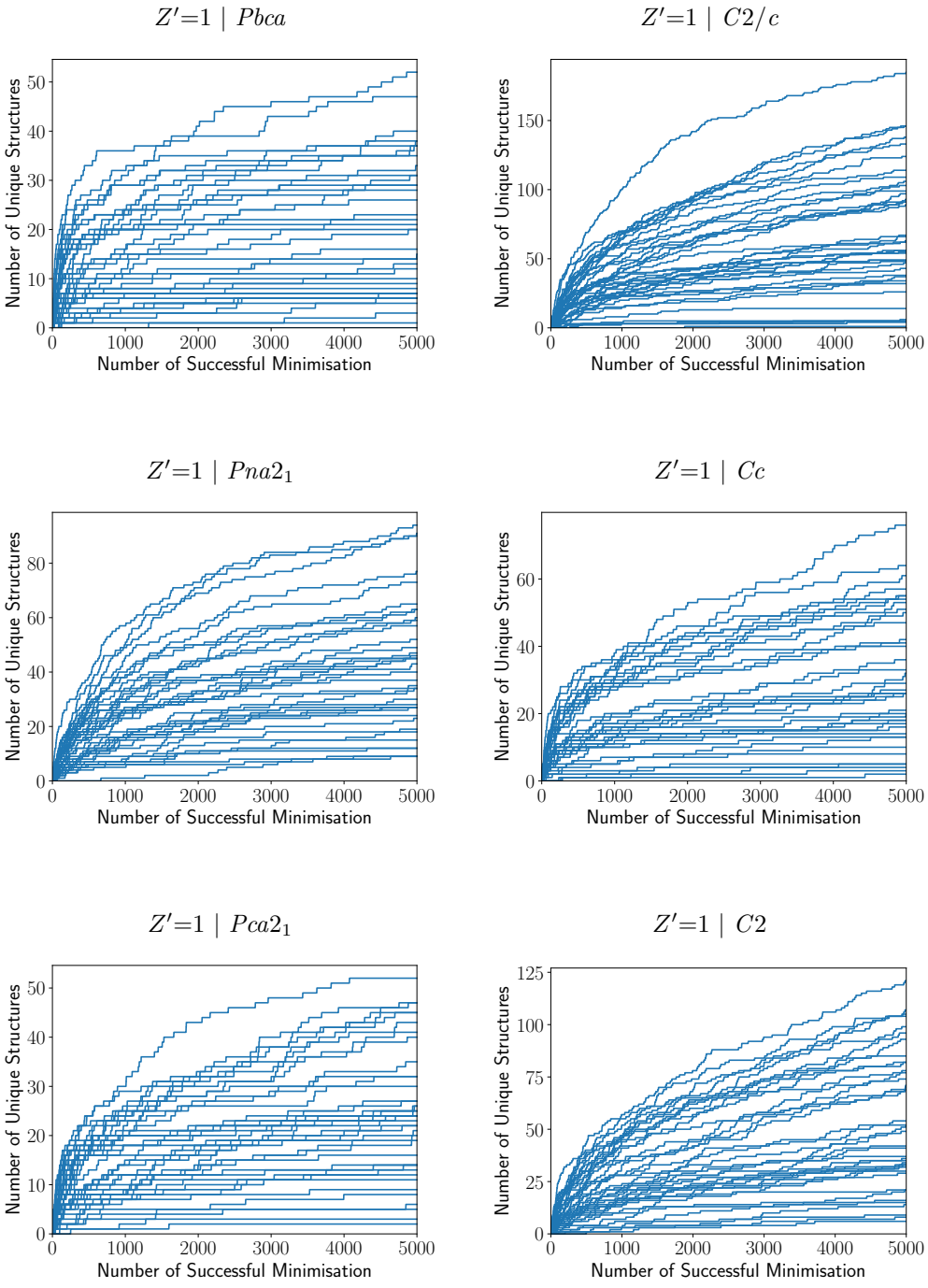


Figure E.24: *continued*

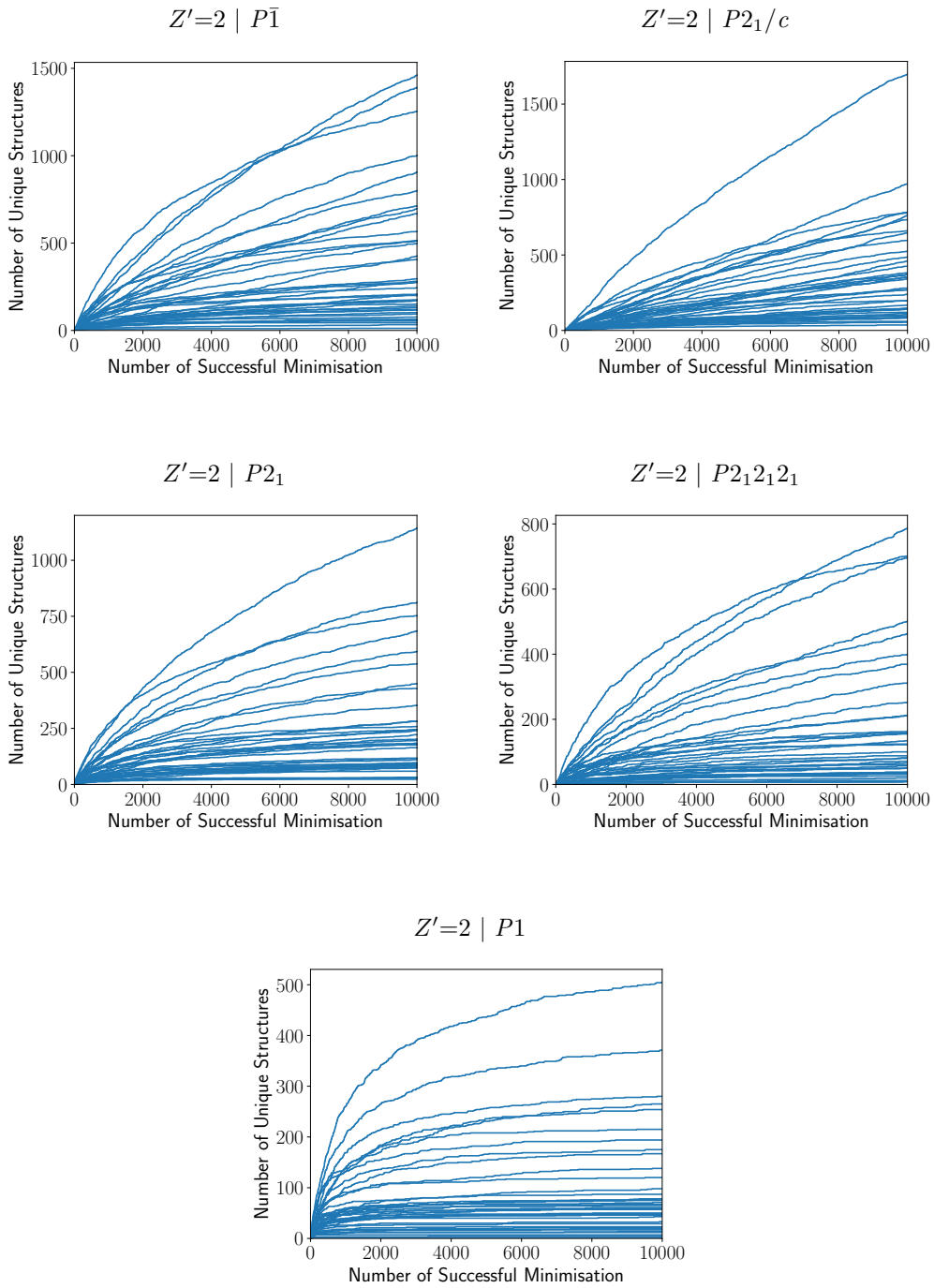


Figure E.24: *continued*

CSP	Reference	14	19	2	4	61	15	33	9	29	5
AROBUFMOL	GM	5	0	0	7	0	1	0	0	0	18
AROBUFMOL	AROBUF	1	0	240	0	0	0	0	1	0	2
AROCAMMOL	AROCAM	191	1	1092	437	4	157	18	142	34	222
BEGKIKMOL	BEGKIK	0	0	0	0	81	0	0	0	0	0
BEZLUOMOL	GM	15	0	0	68	0	0	0	0	1	0
BEZLUOMOL	BEZLUO	11	0	0	36	0	0	0	0	0	3
BNPYREMOL	GM	3	0	0	0	0	0	0	0	0	0
BNPYREMOL	BNPYRE10	25	0	0	0	0	0	0	0	0	0
BNZACRMOL	BNZACR	0	0	0	0	0	0	5	0	0	0
CEQGELMOL	GM	8	0	0	0	0	0	0	0	0	0
CEQGELMOL	CEQGEL	0	0	0	0	0	0	0	0	0	0
DBANQUMOL	DBANQU	0	0	0	0	0	0	0	0	26	0
DBNTHRMOL	DBNTHR01	0	274	0	0	0	0	0	0	445	0
DBNTHRMOL	DBNTHR10	0	0	0	112	0	0	0	0	0	0
DUXSASMOL	GM	0	0	1731	0	0	0	0	0	2	9
DUXSASMOL	DUXSAS	28	0	0	0	0	1	0	0	0	1
FANTRQMOL	GM	10	0	0	47	0	0	0	0	0	1
FANTRQMOL	FANTRQ10	1	9	0	0	0	0	9	0	6	11
HAMDUWMOL	HAMDUW	28	0	0	0	0	0	0	0	0	0
KONBAQMOL	KONBAQ	12	55	0	0	0	0	15	0	0	0
LOHVAHMOL	LOHVAH	0	21	0	0	0	0	0	0	0	0
MIVRUEMOL	GM	12	0	0	0	0	0	0	0	0	0
MIVRUEMOL	MIVRUE	0	86	0	0	0	0	0	0	0	0
MIVSESMOL	GM	1	0	0	5	0	0	0	0	0	0
MORRODMOL	MORROD	56	0	0	0	0	0	0	0	0	0

Table E.10: The number of times the reference crystal structures were sampled in the CSP searches described in Section 7.3 for a given space group with Z'=1.

CSP	Reference	14	19	2	4	61	15	33	9	29	5
NAAZASMOL	GM	2	0	0	36	0	0	0	0	0	0
NAAZASMOL	NAAZAS	1	0	0	21	0	0	0	0	0	0
NIJCEQMOL	GM	10	0	0	0	0	0	0	0	0	0
NIJCEQMOL	NIJCEQ	37	0	0	0	0	0	0	0	0	0
OFANTQMOL	OFANTQ	56	0	0	0	0	2	1	1	0	1
PAGMERMOL	GM	0	42	0	0	0	0	0	0	0	0
PAGMERMOL	PAGMER	34	0	0	0	0	0	0	0	0	0
PENCENMOL	PENCEN	0	0	0	1	0	0	0	0	1	8
PENCENMOL	PENCEN01	0	1	0	0	0	0	0	0	0	3
PENCENMOL	PENCEN10	1	0	0	0	0	0	0	0	0	0
PENTQUMOL	PENTQU	31	0	0	31	0	0	0	0	0	0
PENTQUMOL	PENTQU01	0	0	0	0	0	0	0	0	0	0
PENTQUMOL	PENTQU02	27	4	0	20	0	0	3	0	0	0
PERLENMOL	PERLEN01	251	0	0	0	0	0	0	0	0	0
PERLENMOL	PERLEN06	87	0	0	739	0	0	0	0	0	1
POVLUHMOL	GM	6	0	0	0	0	0	0	0	0	0
POVLUHMOL	POVLUH	0	0	0	0	0	30	0	0	0	0
RUNLAPMOL	GM	79	0	0	0	0	0	0	0	0	0
RUNLAPMOL	RUNLAP	41	0	0	0	0	0	0	0	0	0
SANQIIMOL	GM	0	0	0	0	0	0	0	0	0	0
SANQIIMOL	SANQII	0	0	0	0	0	0	0	0	0	0
SEVYEAMOL	SEVYEA	0	125	0	0	0	0	0	0	0	0
UPIRUIMOL	GM	6	0	0	0	0	0	0	0	0	0
UPIRUIMOL	UPIRUI	0	0	0	0	33	0	0	0	0	0
UPIXAUMOL	GM	2	0	0	0	0	0	0	0	0	0

Table E.10: *continued*

CSP	Reference	14	19	2	4	61	15	33	9	29	5
UPIXAUMOL	UPIXAU	0	0	0	20	0	0	0	0	0	0
UPIXEYMOL	GM	1	0	0	0	0	0	0	0	0	0
UPIXEYMOL	UPIXEY	0	0	0	0	0	0	1	0	18	0
VIBMAVMOL	GM	229	0	0	0	0	0	2	0	0	0
VIBMAVMOL	VIBMAV	94	0	0	0	0	0	0	0	0	0
VIQTATMOL	GM	52	0	0	0	0	0	0	0	0	0
VIQTATMOL	VIQTAT	0	244	0	0	0	0	0	0	0	0
VUFKALMOL	GM	61	8	0	55	0	17	6	1	0	1
VUFKALMOL	VUFKAL	22	0	0	0	0	0	0	0	0	0
VUFNIWMOL	GM	1	0	0	0	0	0	0	0	0	0
VUFNIWMOL	VUFNIW	17	0	0	0	0	0	0	0	0	0
WOTZUDMOL	GM	126	0	0	0	0	0	2	0	0	0
WOTZUDMOL	WOTZUD	193	0	0	0	0	0	0	0	0	0
WUPYIQMOL	WUPYIQ	72	4	0	104	0	0	0	0	0	1
YOFROBMOL	YOFROB	6	0	1208	4	0	1	8	3	0	14
ZETQUNMOL	GM	71	9	0	97	0	0	0	1	1	0
ZETQUNMOL	ZETQUN	0	0	0	15	0	0	0	0	0	6
ZZZOYCMOL	ZZZOYC01	0	0	0	225	0	0	32	4	0	6

Table E.10: *continued*

CSP	Reference	2	14	4	19	1
AROBUFMOL	GM	5	0	5	0	9
AROBUFMOL	AROBUF	54	0	3	0	131
AROCAMMOL	AROCAM	437	71	157	1	1234
BEGKIKMOL	BEGKIK	0	3	0	0	0
BEZLUOMOL	GM	64	2	19	0	425
BEZLUOMOL	BEZLUO	27	1	15	0	94
BNPYREMOL	GM	0	0	1	0	0
BNPYREMOL	BNPYRE10	3	1	19	0	0
BNZACRMOL	BNZACR	0	0	0	0	0
CEQGELMOL	GM	1	0	4	0	0
CEQGELMOL	CEQGEL	6	0	0	0	0
DBANQUMOL	DBANQU	0	0	0	0	0
DBNTHRMOL	DBNTHR01	0	0	24	0	0
DBNTHRMOL	DBNTHR10	0	0	11	0	11
DUXSASMOL	GM	345	0	0	0	378
DUXSASMOL	DUXSAS	8	5	9	0	0
FANTRQMOL	GM	24	1	11	0	87
FANTRQMOL	FANTRQ10	0	0	14	0	0
HAMDUWMOL	HAMDUW	0	0	0	0	0
KONBAQMOL	KONBAQ	0	3	14	0	0
LOHVAHMOL	LOHVAH	0	0	7	0	0
MIVRUEMOL	GM	0	0	0	0	0
MIVRUEMOL	MIVRUE	0	0	23	0	0
MIVSESMOL	GM	0	0	1	0	3
MORRODMOL	MORROD	19	10	25	0	0
NAAZASMOL	GM	8	0	0	0	22
NAAZASMOL	NAAZAS	5	1	1	0	22
NIJCEQMOL	GM	2	0	0	0	0
NIJCEQMOL	NIJCEQ	0	2	9	0	0
OFANTQMOL	OFANTQ	34	3	28	0	0
PAGMERMOL	GM	0	0	7	0	0
PAGMERMOL	PAGMER	7	3	14	0	0
PENCENMOL	PENCEN	69	0	1	0	585
PENCENMOL	PENCEN01	16	0	1	0	539
PENCENMOL	PENCEN10	0	0	1	0	0
PENTQUMOL	PENTQU	44	6	54	0	103
PENTQUMOL	PENTQU01	1	0	0	0	0
PENTQUMOL	PENTQU02	55	1	48	0	106
PERLENMOL	PERLEN01	45	22	156	0	0

Table E.11: The number of times the reference crystal structures were sampled in the CSP searches described in Section 7.3 for a given space group with Z'=2.

CSP	Reference	2	14	4	19	1
PERLENMOL	PERLEN06	49	16	125	0	356
POVLUHMOL	GM	1	1	3	0	0
POVLUHMOL	POVLUH	2	0	0	0	0
RUNLAPMOL	GM	24	2	24	0	0
RUNLAPMOL	RUNLAP	10	5	19	0	0
SANQIIMOL	GM	8	0	0	0	0
SANQIIMOL	SANQII	0	0	7	0	0
SEVYEAMOL	SEVYEA	0	0	19	0	0
UPIRUIMOL	GM	2	0	0	0	0
UPIRUIMOL	UPIRUI	0	0	0	0	0
UPIXAUMOL	GM	0	0	0	0	0
UPIXAUMOL	UPIXAU	0	0	5	0	30
UPIXEYMOL	GM	0	0	0	0	0
UPIXEYMOL	UPIXEY	0	0	0	0	0
VIBMAVMOL	GM	191	17	48	0	0
VIBMAVMOL	VIBMAV	58	15	87	0	0
VIQTATMOL	GM	5	4	23	0	0
VIQTATMOL	VIQTAT	0	0	40	1	0
VUFKALMOL	GM	39	8	43	0	51
VUFKALMOL	VUFKAL	0	7	0	0	0
VUFNIWMOL	GM	0	0	1	0	0
VUFNIWMOL	VUFNIW	0	1	4	0	0
WOTZUDMOL	GM	33	12	108	0	0
WOTZUDMOL	WOTZUD	38	23	80	0	0
WUPYIQMOL	WUPYIQ	236	10	81	0	440
YOFROBMOL	YOFROB	554	0	1	0	1357
ZETQUNMOL	GM	28	34	185	0	31
ZETQUNMOL	ZETQUN	11	0	8	0	600
ZZZOYCMOL	ZZZOYC01	0	0	52	0	232

Table E.11: *continued*

CSP	Reference	14	19	2	4	61	15	33	9	29	5
AROBUFMOL	GM	930	-	-	331	-	2090	-	-	-	294
AROBUFMOL	AROBUF	1257	-	2	-	-	-	-	1543	-	3152
AROCAMMOL	AROCAM	43	147	4	66	518	82	136	56	87	25
BEGKIKMOL	BEGKIK	-	-	-	-	21	-	-	-	-	-
BEZLUOMOL	GM	180	-	-	76	-	-	-	-	2104	-
BEZLUOMOL	BEZLUO	502	-	-	59	-	-	-	-	-	371
BNPYREMOL	GM	2286	-	-	-	-	-	-	-	-	-
BNPYREMOL	BNPYRE10	86	-	-	-	-	-	-	-	-	-
BNZACRMOL	BNZACR	-	-	-	-	-	-	522	-	-	-
CEQGELMOL	GM	375	-	-	-	-	-	-	-	-	-
CEQGELMOL	CEQGEL	-	-	-	-	-	-	-	-	-	-
DBANQUMOL	DBANQU	-	-	-	-	-	-	-	-	29	-
DBNTHRMOL	DBNTHR01	-	64	-	-	-	-	-	-	9	-
DBNTHRMOL	DBNTHR10	-	-	-	38	-	-	-	-	-	-
DUXSASMOL	GM	-	-	1	-	-	-	-	-	1519	300
DUXSASMOL	DUXSAS	58	-	-	-	-	4674	-	-	-	1079
FANTRQMOL	GM	203	-	-	22	-	-	-	-	-	1306
FANTRQMOL	FANTRQ10	307	843	-	-	-	-	112	-	578	399
HAMDUWMOL	HAMDUW	398	-	-	-	-	-	-	-	-	-
KONBAQMOL	KONBAQ	242	104	-	-	-	-	186	-	-	-
LOHVAHMOL	LOHVAH	-	51	-	-	-	-	-	-	-	-
MIVRUEMOL	GM	319	-	-	-	-	-	-	-	-	-
MIVRUEMOL	MIVRUE	-	129	-	-	-	-	-	-	-	-
MIVSESMOL	GM	803	-	-	859	-	-	-	-	-	-
MORRODMOL	MORROD	50	-	-	-	-	-	-	-	-	-

Table E.12: The number of successful minimisations required for the CSP searches described in Section 7.3 to obtain the reference crystal structure in a given space group with Z'=1.

CSP	Reference	14	19	2	4	61	15	33	9	29	5
NAAZASMOL	GM	1403	-	-	271	-	-	-	-	-	-
NAAZASMOL	NAAZAS	3880	-	-	293	-	-	-	-	-	-
NIJCEQMOL	GM	630	-	-	-	-	-	-	-	-	-
NIJCEQMOL	NIJCEQ	326	-	-	-	-	-	-	-	-	-
OFANTQMOL	OFANTQ	58	-	-	-	-	389	2425	4071	-	897
PAGMERMOL	GM	-	143	-	-	-	-	-	-	-	-
PAGMERMOL	PAGMER	165	-	-	-	-	-	-	-	-	-
PENCENMOL	PENCEN	-	-	-	1812	-	-	-	-	4401	431
PENCENMOL	PENCEN01	-	4964	-	-	-	-	-	-	-	3081
PENCENMOL	PENCEN10	4574	-	-	-	-	-	-	-	-	-
PENTQUMOL	PENTQU	73	-	-	141	-	-	-	-	-	-
PENTQUMOL	PENTQU01	-	-	-	-	-	-	-	-	-	-
PENTQUMOL	PENTQU02	237	83	-	248	-	-	173	-	-	-
PERLENMOL	PERLEN01	24	-	-	-	-	-	-	-	-	-
PERLENMOL	PERLEN06	63	-	-	1	-	-	-	-	-	4320
POVLUHMOL	GM	98	-	-	-	-	-	-	-	-	-
POVLUHMOL	POVLUH	-	-	-	-	-	13	-	-	-	-
RUNLAPMOL	GM	36	-	-	-	-	-	-	-	-	-
RUNLAPMOL	RUNLAP	139	-	-	-	-	-	-	-	-	-
SANQIIMOL	GM	-	-	-	-	-	-	-	-	-	-
SANQIIMOL	SANQII	-	-	-	-	-	-	-	-	-	-
SEVYEAMOL	SEVYEA	-	108	-	-	-	-	-	-	-	-
UPIRUIMOL	GM	924	-	-	-	-	-	-	-	-	-
UPIRUIMOL	UPIRUI	-	-	-	-	237	-	-	-	-	-
UPIXAUMOL	GM	821	-	-	-	-	-	-	-	-	-

Table E.12: *continued*

CSP	Reference	14	19	2	4	61	15	33	9	29	5
UPIXAUMOL	UPIXAU	-	-	-	66	-	-	-	-	-	-
UPIXEYMOL	GM	2149	-	-	-	-	-	-	-	-	-
UPIXEYMOL	UPIXEY	-	-	-	-	-	-	1843	-	87	-
VIBMAVMOL	GM	132	-	-	-	-	-	1265	-	-	-
VIBMAVMOL	VIBMAV	72	-	-	-	-	-	-	-	-	-
VIQTATMOL	GM	62	-	-	-	-	-	-	-	-	-
VIQTATMOL	VIQTAT	-	54	-	-	-	-	-	-	-	-
VUFKALMOL	GM	180	250	-	54	-	293	74	4189	-	568
VUFKALMOL	VUFKAL	99	-	-	-	-	-	-	-	-	-
VUFNIWMOL	GM	2428	-	-	-	-	-	-	-	-	-
VUFNIWMOL	VUFNIW	517	-	-	-	-	-	-	-	-	-
WOTZUDMOL	GM	2	-	-	-	-	-	1579	-	-	-
WOTZUDMOL	WOTZUD	1	-	-	-	-	-	-	-	-	-
WUPYIQMOL	WUPYIQ	43	1574	-	25	-	-	-	-	-	243
YOFROBMOL	YOFROB	936	-	3	940	-	4338	93	2352	-	642
ZETQUNMOL	GM	29	709	-	165	-	-	-	2969	671	-
ZETQUNMOL	ZETQUN	-	-	-	782	-	-	-	-	-	1534
ZZZOYCMOL	ZZZOYC01	-	-	-	26	-	-	628	770	-	46

Table E.12: *continued*

CSP	Reference	2	14	4	19	1
AROBUFMOL	GM	2165	-	717	-	1622
AROBUFMOL	AROBUF	92	-	6123	-	1
AROCAMMOL	AROCAM	21	1	-	5739	3
BEGKIKMOL	BEGKIK	-	2047	-	-	-
BEZLUOMOL	GM	58	5	1400	-	7
BEZLUOMOL	BEZLUO	384	9736	2114	-	129
BNPYREMOL	GM	-	-	5430	-	-
BNPYREMOL	BNPYRE10	3203	546	132	-	-
BNZACRMOL	BNZACR	-	-	-	-	-
CEQGELMOL	GM	3087	-	385	-	-
CEQGELMOL	CEQGEL	3452	-	-	-	-
DBANQUMOL	DBANQU	-	-	-	-	-
DBNTHRMOL	DBNTHR01	-	-	586	-	-
DBNTHRMOL	DBNTHR10	-	-	176	-	26
DUXSASMOL	GM	-	-	-	-	18
DUXSASMOL	DUXSAS	547	2499	1086	-	-
FANTRQMOL	GM	152	678	1109	-	26
FANTRQMOL	FANTRQ10	-	-	1766	-	-
HAMDUWMOL	HAMDUW	-	-	-	-	-
KONBAQMOL	KONBAQ	-	14	1307	-	-
LOHVAHMOL	LOHVAH	-	-	1604	-	-
MIVRUEMOL	GM	-	-	-	-	-
MIVRUEMOL	MIVRUE	-	-	652	-	-
MIVSESMOL	GM	-	-	458	-	5409
MORRODMOL	MORROD	385	259	252	-	-
NAAZASMOL	GM	1591	-	-	-	30
NAAZASMOL	NAAZAS	1399	4944	7910	-	2125
NIJCEQMOL	GM	5357	-	-	-	-
NIJCEQMOL	NIJCEQ	-	5063	900	-	-
OFANTQMOL	OFANTQ	92	1910	391	-	-
PAGMERMOL	GM	-	-	3359	-	-
PAGMERMOL	PAGMER	483	2053	1257	-	-
PENCENMOL	PENCEN	129	-	2502	-	-
PENCENMOL	PENCEN01	451	-	5200	-	11
PENCENMOL	PENCEN10	-	-	3036	-	-
PENTQUMOL	PENTQU	450	894	205	-	68
PENTQUMOL	PENTQU01	9349	-	-	-	-
PENTQUMOL	PENTQU02	287	2468	114	-	55
PERLENMOL	PERLEN01	279	632	47	-	-

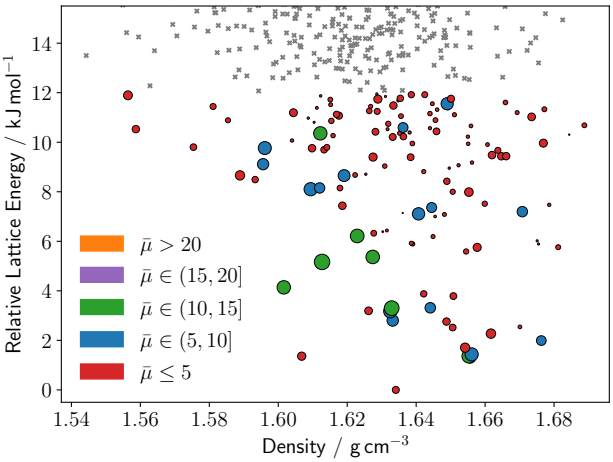
Table E.13: The minimum number of successful minimisations required for the CSP searches described in Section 7.3 to obtain the reference crystal structure in a given space group with Z'=2.

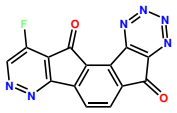
CSP	Reference	2	14	4	19	1
PERLENMOL	PERLEN06	29	546	36	-	12
POVLUHMOL	GM	8725	6686	768	-	-
POVLUHMOL	POVLUH	1302	-	-	-	-
RUNLAPMOL	GM	173	6185	192	-	-
RUNLAPMOL	RUNLAP	2319	129	1101	-	-
SANQIIMOL	GM	351	-	-	-	-
SANQIIMOL	SANQII	-	-	277	-	-
SEVYEAMOL	SEVYEA	-	-	317	-	-
UPIRUIMOL	GM	516	-	-	-	-
UPIRUIMOL	UPIRUI	-	-	-	-	-
UPIXAUMOL	GM	-	-	-	-	-
UPIXAUMOL	UPIXAU	-	-	2472	-	19
UPIXEYMOL	GM	-	-	-	-	-
UPIXEYMOL	UPIXEY	-	-	-	-	-
VIBMAVMOL	GM	75	851	41	-	-
VIBMAVMOL	VIBMAV	3	134	26	-	-
VIQTATMOL	GM	4766	421	487	-	-
VIQTATMOL	VIQTAT	-	-	206	1464	-
VUFKALMOL	GM	145	2950	102	-	93
VUFKALMOL	VUFKAL	-	2643	-	-	-
VUFNIWMOL	GM	-	-	1260	-	-
VUFNIWMOL	VUFNIW	-	4155	14	-	-
WOTZUDMOL	GM	814	237	61	-	-
WOTZUDMOL	WOTZUD	175	25	31	-	-
WUPYIQMOL	WUPYIQ	21	1152	26	-	31
YOFROBMOL	YOFROB	9	-	4780	-	3
ZETQUNMOL	GM	2252	76	1	-	148
ZETQUNMOL	ZETQUN	2293	-	274	-	47
ZZZOYCMOL	ZZZOYC01	-	-	34	-	8

Table E.13: *continued*

Appendix F

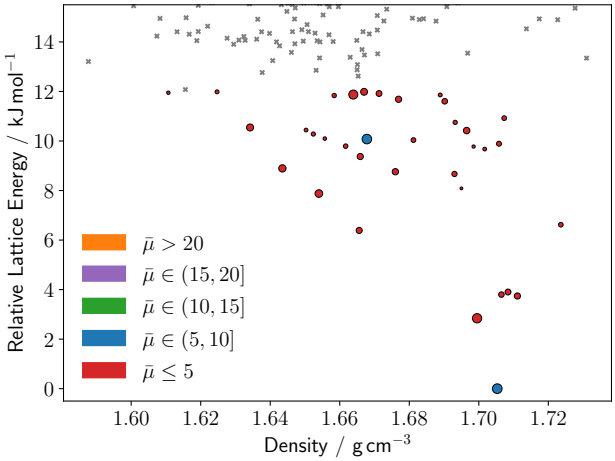
IFO Derivatives ESF Maps

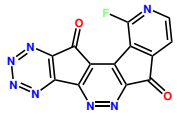




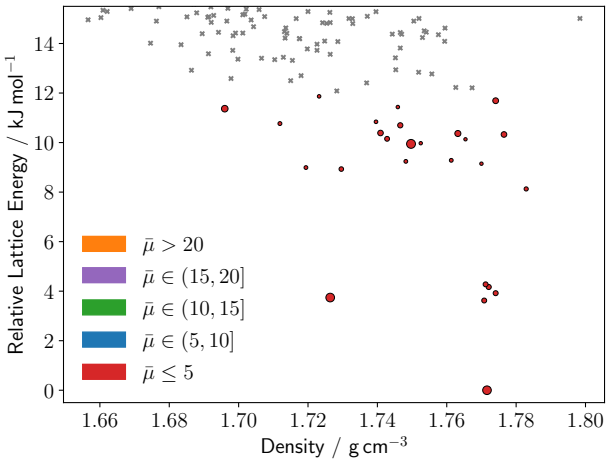
1

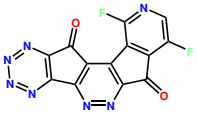
Figure F.1: ESF maps of the CSP searches described in Chapter 8 for the 18 best molecules out of all 320 molecules taken through the mobility evaluations stage of the computational screening workflow and 6 experimental molecules which had previously been used in the fabrication of n-type OFETs. Electron mobilities were calculated for all structures within 12 kJ mol^{-1} from the global minimum, plot points are scaled and coloured by their calculated mobilities. Ranges of the mobilities defined in the ESF map legends are given in units of cm^2(Vs)$^{-1}$.



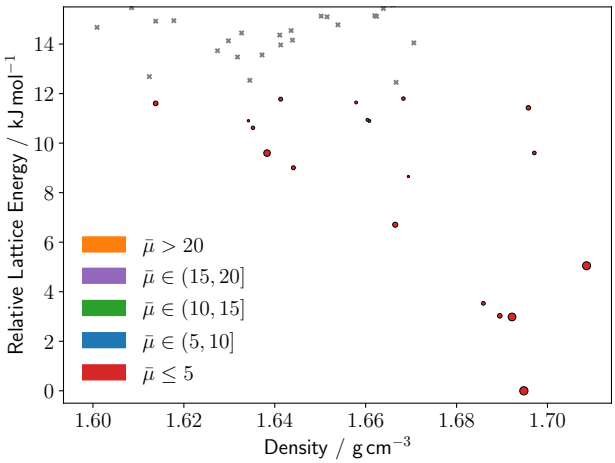


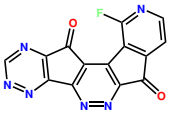
2





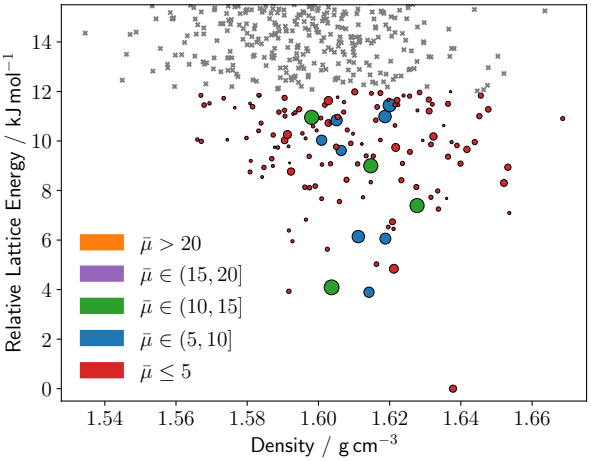
3

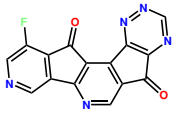




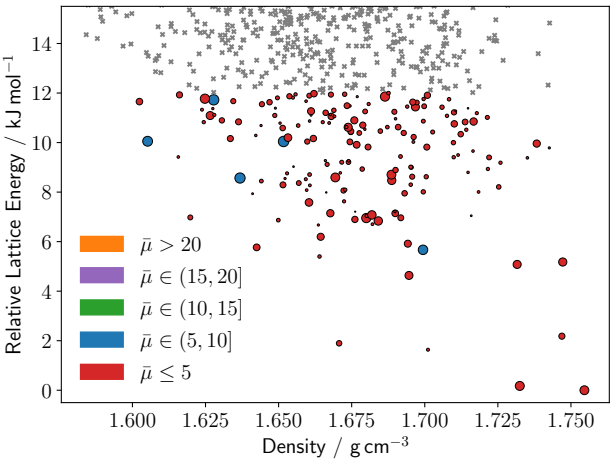
4

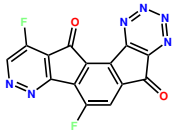
Figure F.1: *continued*



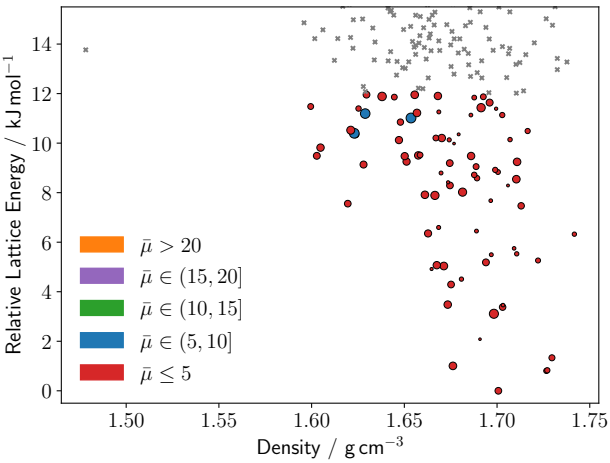


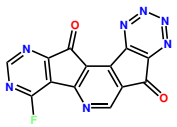
5





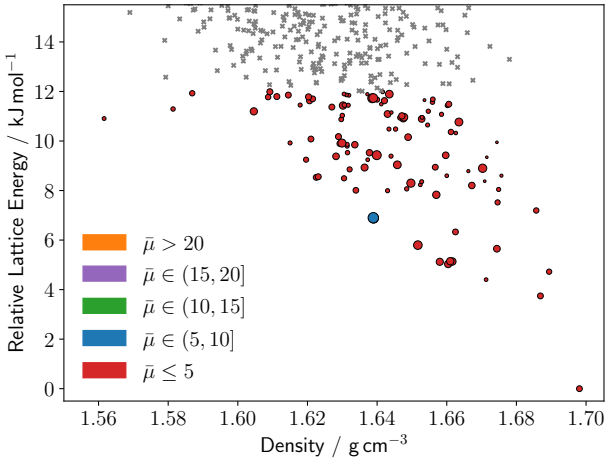
6

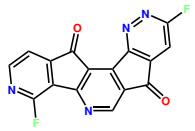




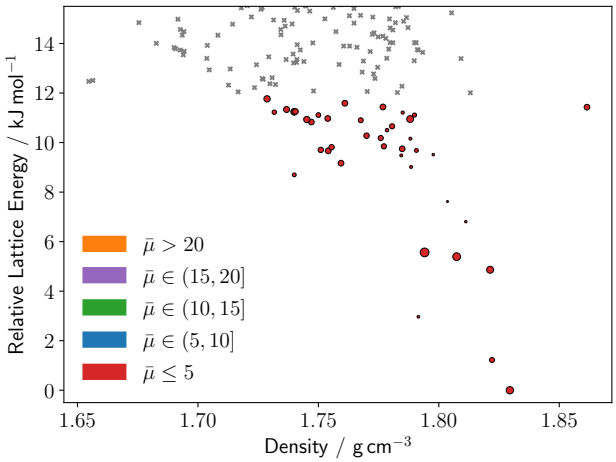
7

Figure F.1: *continued*

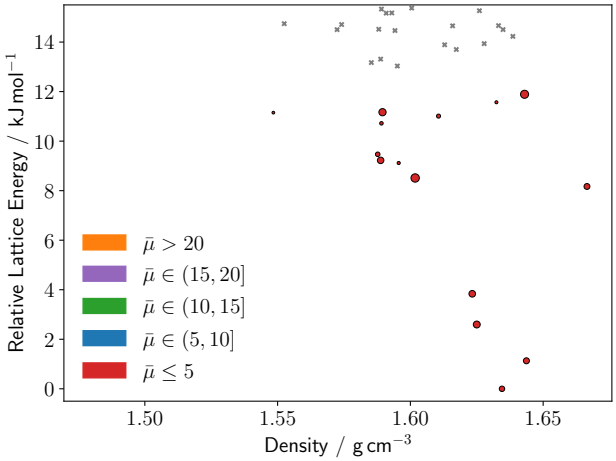


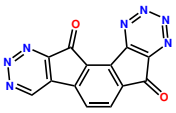


8



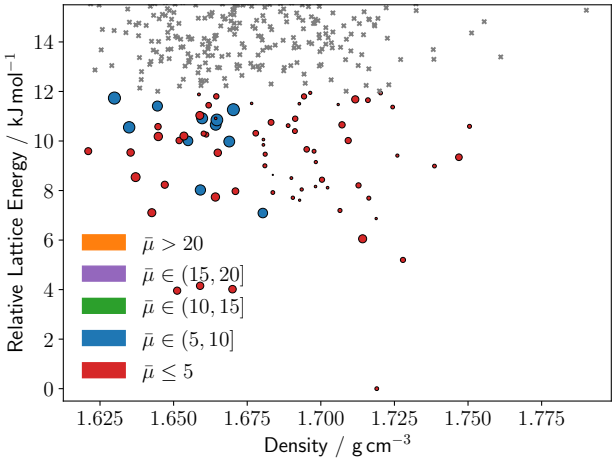
9

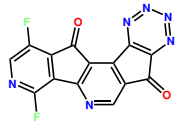




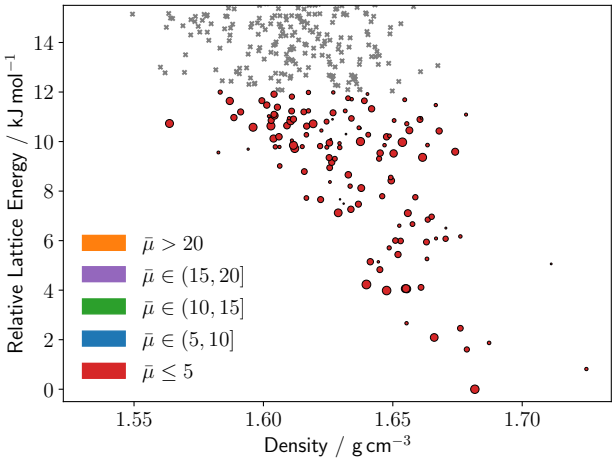
10

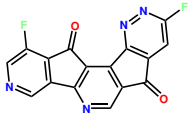
Figure F.1: *continued*



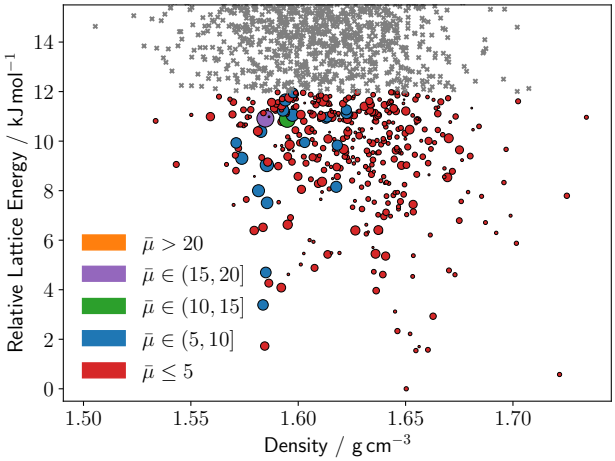


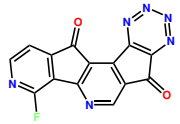
11





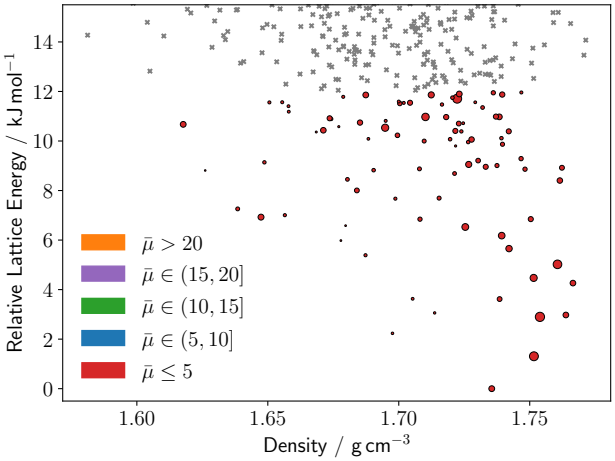
12

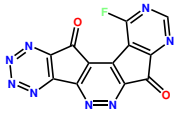




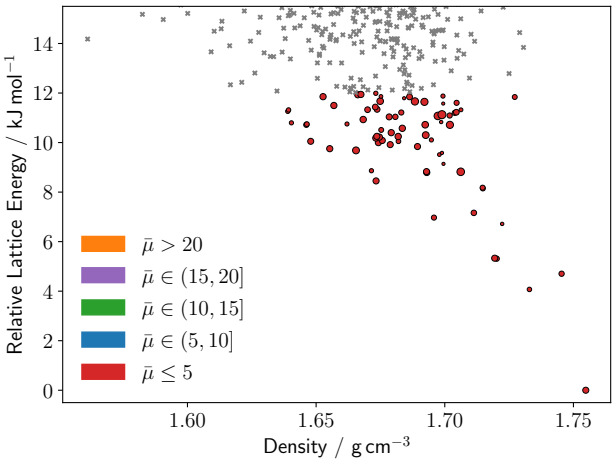
13

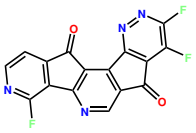
Figure F.1: *continued*



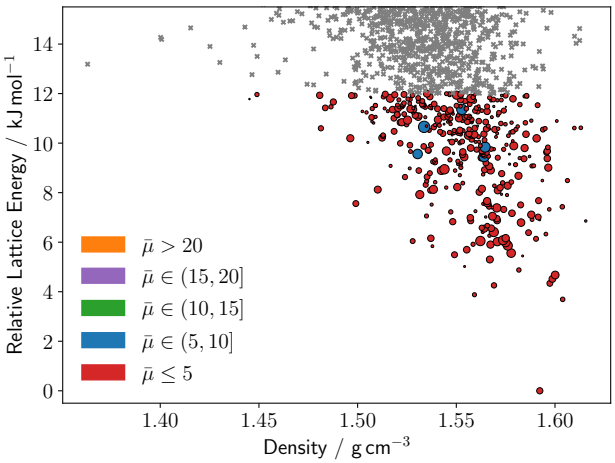


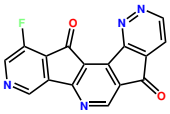
14





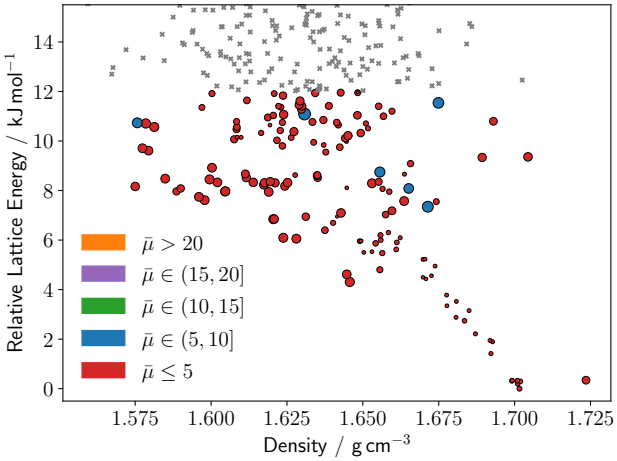
15



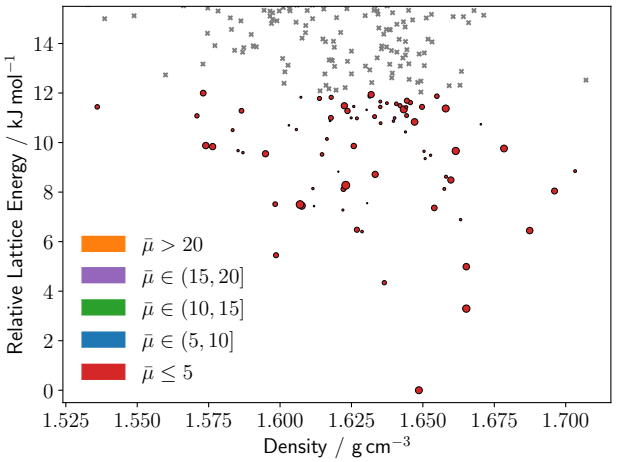


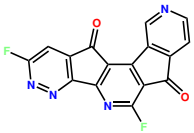
16

Figure F.1: *continued*

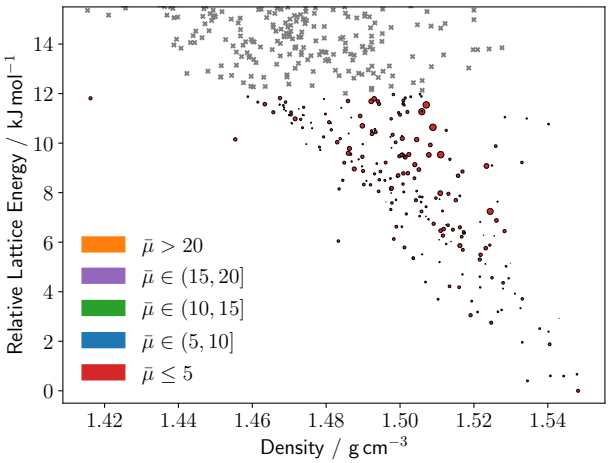


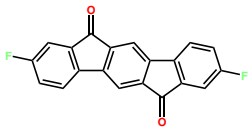
17





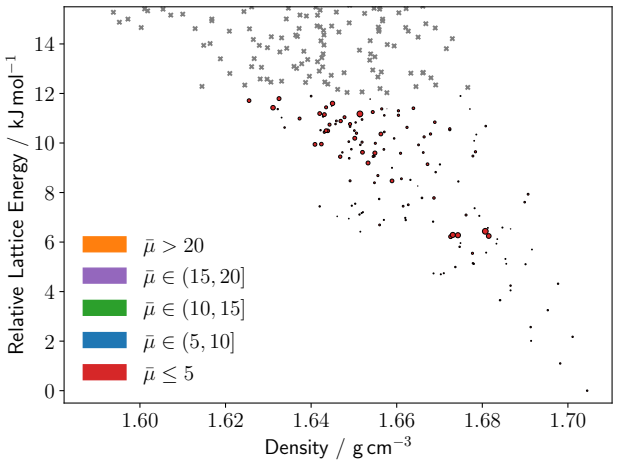
18

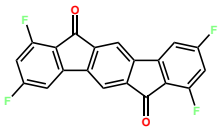




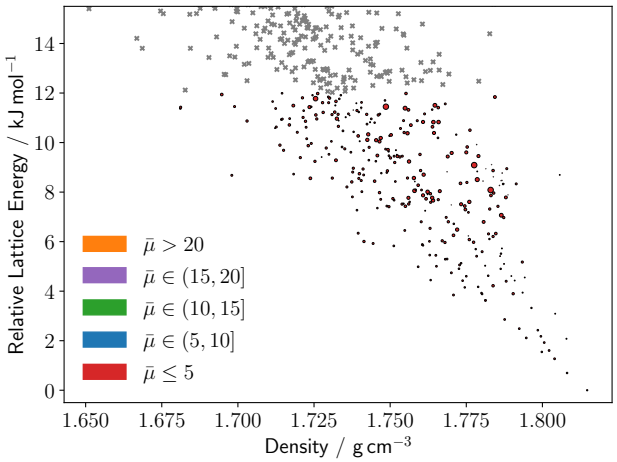
MonoF-IFO

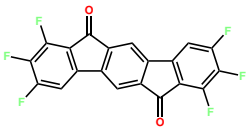
Figure F.1: *continued*



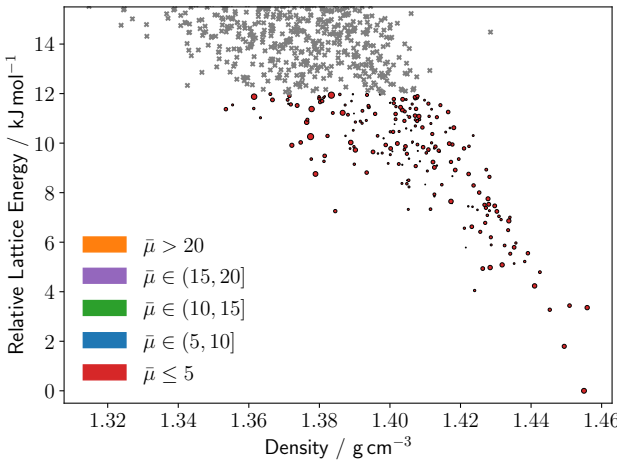


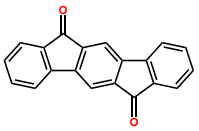
DiF-IFO





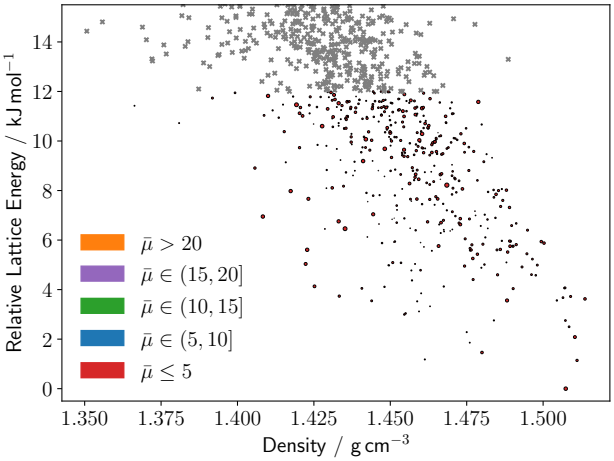
TriF-IFO

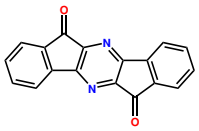




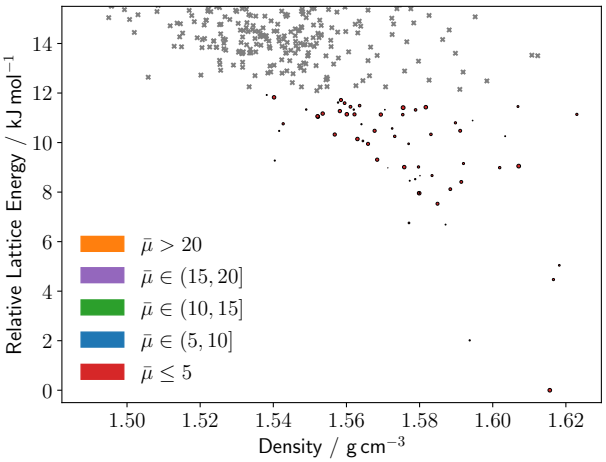
1a

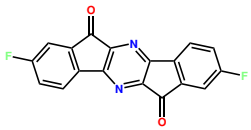
Figure F.1: *continued*





2a





2b

Figure F.1: *continued*

References

[1] M. Vasilopoulou, *Semiconducting Organic Molecules* (John Wiley & Sons, Inc., 2017), .

[2] X. Ren *et al.*, Advanced Electronic Materials **3**, 1700018 (2017).

[3] M. Matta *et al.*, Mater. Horiz. **5**, 41 (2018).

[4] S. M. Gali *et al.*, J. Mater. Chem. C **7**, 4382 (2019).

[5] M. T. Ruggiero, S. Ciuchi, S. Fratini and G. D'Avino, The Journal of Physical Chemistry C **123**, 15897 (2019).

[6] A. Landi, A. Peluso and A. Troisi, Advanced Materials **33**, 2008049 (2021).

[7] J. Elsner, S. Giannini and J. Blumberger, The Journal of Physical Chemistry Letters **12**, 5857 (2021).

[8] H. H. Choi *et al.*, Advanced Science **7**, 1901824 (2020).

[9] R. D. McCullough, Advanced Materials **10**, 93 (1998).

[10] H. Sirringhaus *et al.*, Nature **401**, 685 (1999).

[11] P. J. Fagan, J. C. Calabrese and B. Malone, Accounts of Chemical Research **25**, 134 (1992).

[12] J. Zaumseil, C. L. Donley, J.-S. Kim, R. Friend and H. Sirringhaus, Advanced Materials **18**, 2708 (2006).

[13] M. C. Gwinner *et al.*, Advanced Materials **24**, 2728 (2012).

[14] M. Weis, *Organic Field-Effect Transistors* (John Wiley & Sons, Inc., 2017), pp. 1–39.

[15] Z. A. Lamport, H. F. Haneef, S. Anand, M. Waldrip and O. D. Jurchescu, Journal of Applied Physics **124**, 071101 (2018).

[16] H. H. Choi, K. Cho, C. D. Frisbie, H. Sirringhaus and V. Podzorov, Nature Materials **17**, 2 (2018).

[17] V. Podzorov, MRS Bulletin **38**, 1524 (2013).

[18] M. Waldrip, O. D. Jurchescu, D. J. Gundlach and E. G. Bittle, Advanced Functional Materials **30**, 1904576 (2020).

[19] H. H. Choi *et al.*, Advanced Functional Materials **28**, 1707105 (2018).

[20] J. P. M. Lommerse *et al.*, Acta Crystallographica Section B **56**, 697 (2000).

[21] W. D. S. Motherwell *et al.*, Acta Crystallographica Section B **58**, 647 (2002).

[22] G. M. Day *et al.*, Acta Crystallographica Section B **61**, 511 (2005).

[23] G. M. Day *et al.*, Acta Crystallographica Section B **65**, 107 (2009).

[24] D. A. Bardwell *et al.*, Acta Crystallographica Section B **67**, 535 (2011).

[25] A. M. Reilly *et al.*, Acta Crystallographica Section B **72**, 439 (2016).

[26] Schrödinger, LLC, New York, NY, Macromodel, v9.9.013, 2014.

[27] W. L. Jorgensen, D. S. Maxwell and J. Tirado-Rives, Journal of the American Chemical Society **118**, 11225 (1996).

[28] G. A. Kaminski, R. A. Friesner, J. Tirado-Rives and W. L. Jorgensen, The Journal of Physical Chemistry B **105**, 6474 (2001).

[29] S. Grimme, J. Antony, S. Ehrlich and H. Krieg, The Journal of Chemical Physics **132**, 154104 (2010).

[30] S. Grimme, S. Ehrlich and L. Goerigk, Journal of Computational Chemistry **32**, 1456 (2011).

[31] J. C. Cole, C. R. Groom, O. Korb, P. McCabe and G. P. Shields, Journal of Chemical Information and Modeling **56**, 652 (2016).

[32] V. E. Bazterra, M. B. Ferraro and J. C. Facelli, The Journal of Chemical Physics **116**, 5984 (2002).

[33] V. E. Bazterra, M. B. Ferraro and J. C. Facelli, The Journal of Chemical Physics **116**, 5992 (2002).

[34] V. E. Bazterra, M. B. Ferraro and J. C. Facelli, International Journal of Quantum Chemistry **96**, 312 (2004).

[35] S. Kim, A. M. Orendt, M. B. Ferraro and J. C. Facelli, Journal of Computational Chemistry **30**, 1973 (2009).

[36] D. H. Case, J. E. Campbell, P. J. Bygrave and G. M. Day, Journal of Chemical Theory and Computation **12**, 910 (2016).

[37] S. L. Price *et al.*, Phys. Chem. Chem. Phys. **12**, 8478 (2010).

[38] A. V. Kazantsev, P. G. Karamertzanis, C. C. Pantelides and C. S. Adjiman, *CrystalOptimizer: An Efficient Algorithm for Lattice Energy Minimization of Organic Crystals Using Isolated-Molecule Quantum Mechanical Calculations* (John Wiley & Sons, Ltd, 2011), chap. 1, pp. 1–42.

[39] A. M. Lund, G. I. Pagola, A. M. Orendt, M. B. Ferraro and J. C. Facelli, Chemical Physics Letters **626**, 20 (2015).

[40] P. Giannozzi *et al.*, Journal of Physics: Condensed Matter **21**, 395502 (2009).

[41] A. Tkatchenko and M. Scheffler, Phys. Rev. Lett. **102**, 073005 (2009).

[42] T. A. Halgren, Journal of Computational Chemistry **17**, 490 (1996).

[43] A. Tkatchenko, R. A. DiStasio, R. Car and M. Scheffler, Phys. Rev. Lett. **108**, 236402 (2012).

[44] A. Ambrosetti, A. M. Reilly, R. A. DiStasio and A. Tkatchenko, The Journal of Chemical Physics **140**, 18A508 (2014).

[45] G. M. Day, S. L. Price and M. Leslie, The Journal of Physical Chemistry B **107**, 10919 (2003).

[46] J. Nyman and G. M. Day, CrystEngComm **17**, 5154 (2015).

[47] B. Monserrat, N. D. Drummond and R. J. Needs, Phys. Rev. B **87**, 144302 (2013).

[48] M. E. Tuckerman, D. Yarne, S. O. Samuelson, A. L. Hughes and G. J. Martyna, Computer Physics Communications **128**, 333 (2000).

[49] S. X. M. Boerrigter *et al.*, The Journal of Physical Chemistry A **108**, 5894 (2004).

[50] M. A. Deij, J. H. ter Horst, H. Meekes, P. Jansens and E. Vlieg, The Journal of Physical Chemistry B **111**, 1523 (2007).

[51] A. N. Sokolov *et al.*, Nature Communications **2**, 437 (2011).

[52] T. Yamamoto and K. Takimiya, Journal of Photopolymer Science and Technology **20**, 57 (2007).

[53] T. Yamamoto and K. Takimiya, Journal of the American Chemical Society **129**, 2224 (2007).

[54] S. L. Mayo, B. D. Olafson and W. A. Goddard, The Journal of Physical Chemistry **94**, 8897 (1990).

[55] W.-Q. Deng and W. A. Goddard, The Journal of Physical Chemistry B **108**, 8614 (2004).

[56] C. Wang, F. Wang, X. Yang, Q. Li and Z. Shuai, Organic Electronics **9**, 635 (2008).

[57] S.-H. Wen *et al.*, The Journal of Physical Chemistry B **113**, 8813 (2009).

[58] J. Yang *et al.*, Chemistry of Materials **30**, 4361 (2018).

[59] P. Gómez *et al.*, Crystal Growth & Design **17**, 3371 (2017).

[60] C. Schober, K. Reuter and H. Oberhofer, The Journal of Physical Chemistry Letters **7**, 3973 (2016).

[61] C. R. Groom, I. J. Bruno, M. P. Lightfoot and S. C. Ward, Acta Crystallographica Section B **72**, 171 (2016).

[62] T. Nematiaram, D. Padula, A. Landi and A. Troisi, Advanced Functional Materials **30**, 2001906 (2020).

[63] D. Padula, ö. H. Omar, T. Nematiaram and A. Troisi, Energy Environ. Sci. **12**, 2412 (2019).

[64] S. Fratini, D. Mayou and S. Ciuchi, Advanced Functional Materials **26**, 2292 (2016).

[65] T. Nematiaram, S. Ciuchi, X. Xie, S. Fratini and A. Troisi, The Journal of Physical Chemistry C **123**, 6989 (2019).

[66] O. G. Ziogos, S. Giannini, M. Ellis and J. Blumberger, J. Mater. Chem. C **8**, 1054 (2020).

[67] C. Reese, W.-J. Chung, M.-m. Ling, M. Roberts and Z. Bao, Applied Physics Letters **89**, 202108 (2006).

[68] C. Kitamura *et al.*, Chemistry A European Journal **16**, 890 (2010).

[69] G. Nan, X. Yang, L. Wang, Z. Shuai and Y. Zhao, Phys. Rev. B **79**, 115203 (2009).

[70] T. Nematiaram and A. Troisi, The Journal of Chemical Physics **152**, 190902 (2020).

[71] A. Szabo and N. Ostlund, *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory* (Dover Publications, 1996).

[72] D. Cook, *Handbook of Computational Quantum Chemistry* (Dover Publications, 2005).

[73] T. Helgaker, P. Jorgensen and J. Olsen, *Molecular Electronic-Structure Theory* (Wiley, 2014).

[74] J. F. Stanton, J. Gauss, J. D. Watts and R. J. Bartlett, The Journal of Chemical Physics **94**, 4334 (1991).

[75] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).

[76] R. Parr and Y. Weitao, *Density-Functional Theory of Atoms and Molecules* (Oxford University Press, 1994).

[77] K. Burke, The ABC of DFT, https://dft.uci.edu/doc/g1.pdf, Accessed: 2021-04-08.

[78] E. Engel and R. Dreizler, *Density Functional Theory: An Advanced Course* Theoretical and Mathematical Physics (Springer Berlin Heidelberg, 2011).

[79] R. Martin, *Electronic Structure: Basic Theory and Practical Methods* (Cambridge University Press, 2020).

[80] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).

[81] S. H. Vosko, L. Wilk and M. Nusair, Canadian Journal of Physics **58**, 1200 (1980).

[82] J. P. Perdew, K. Burke and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).

[83] A. D. Becke, The Journal of Chemical Physics **98**, 1372 (1993).

[84] A. D. Becke, The Journal of Chemical Physics **98**, 5648 (1993).

[85] P. J. Stephens, F. J. Devlin, C. F. Chabalowski and M. J. Frisch, The Journal of Physical Chemistry **98**, 11623 (1994).

[86] A. D. Becke, Phys. Rev. A **38**, 3098 (1988).

[87] C. Lee, W. Yang and R. G. Parr, Phys. Rev. B **37**, 785 (1988).

[88] R. A. Buckingham, Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences **168**, 264 (1938).

[89] B. Jeziorski, R. Moszynski and K. Szalewicz, Chemical Reviews **94**, 1887 (1994).

[90] A. Stone, *The Theory of Intermolecular Forces* (OUP Oxford, 2013).

[91] T. C. Scott, A. Dalgarno and J. D. Morgan, Phys. Rev. Lett. **67**, 1419 (1991).

[92] A. Nitzan, *Chemical Dynamics in Condensed Phases: Relaxation, Transfer, and Reactions in Condensed Molecular Systems* (OUP Oxford, 2013).

[93] M. Born and R. Oppenheimer, Annalen der Physik **389**, 457 (1927).

[94] J. W. Ochterski, Vibrational Analysis in Gaussian, https://gaussian.com/wp-content/uploads/dl/vib.pdf, Accessed: 2021-06-20.

[95] M. Baer, *Mathematical Introduction* (John Wiley & Sons, Ltd, 2006), chap. 1, pp. 1–25.

[96] C. A. Mead and D. G. Truhlar, The Journal of Chemical Physics **77**, 6090 (1982).

[97] J. E. Subotnik, S. Yeganeh, R. J. Cave and M. A. Ratner, The Journal of Chemical Physics **129**, 244101 (2008).

[98] S. Fatehi, E. Alguire and J. E. Subotnik, The Journal of Chemical Physics **139**, 124112 (2013).

[99] R. J. Cave and M. D. Newton, Chemical Physics Letters **249**, 15 (1996).

[100] R. J. Cave and M. D. Newton, The Journal of Chemical Physics **106**, 9213 (1997).

[101] C. Schober, K. Reuter and H. Oberhofer, The Journal of Chemical Physics **144**, 054103 (2016).

[102] H. Oberhofer and J. Blumberger, Phys. Chem. Chem. Phys. **14**, 13846 (2012).

[103] H. Oberhofer, K. Reuter and J. Blumberger, Chemical Reviews **117**, 10319 (2017).

[104] G. Grosso and G. Parravicini, *Solid State Physics* (Elsevier Science, 2013).

[105] G. Mahan, *Many-Particle Physics* (Springer US, 2000).

[106] S. Sauer, *Molecular Electromagnetism: A Computational Chemistry Approach* Oxford Graduate Texts (OUP Oxford, 2011).

[107] K. Sickafus, E. Kotomin and B. Uberuaga, *Radiation Effects in Solids* (Springer Netherlands, 2007).

[108] A. L. Spek, Acta Crystallographica Section D **65**, 148 (2009).

[109] P. Pracht, F. Bohle and S. Grimme, Phys. Chem. Chem. Phys. **22**, 7169 (2020).

[110] G. Landrum, RDKit: Open-source cheminformatics.

[111] G. Kresse and J. Hafner, Phys. Rev. B **47**, 558 (1993).

[112] G. Kresse and J. Hafner, Phys. Rev. B **49**, 14251 (1994).

[113] G. Kresse and J. Furthmüller, Computational Materials Science **6**, 15 (1996).

[114] G. Kresse and J. Furthmüller, Phys. Rev. B **54**, 11169 (1996).

[115] A. J. Stone, Journal of Chemical Theory and Computation **1**, 1128 (2005).

[116] A. D. Becke, The Journal of Chemical Physics **88**, 2547 (1988).

[117] A. J. Stone, Distributed Multipole Analysis of Gaussian wavefunctions: GDMA version 2.3.0, http://www-stone.ch.cam.ac.uk/documentation/gdma/manual.pdf, Accessed: 2020-08-17.

[118] G. G. Ferenczy, Journal of Computational Chemistry **12**, 913 (1991).

[119] C. Chipot, J. G. Angyan, G. G. Ferenczy and H. A. Scheraga, The Journal of Physical Chemistry **97**, 6628 (1993).

[120] P. J. Winn, G. G. Ferenczy and C. A. Reynolds, The Journal of Physical Chemistry A **101**, 5437 (1997).

[121] G. G. Ferenczy, P. J. Winn and C. A. Reynolds, The Journal of Physical Chemistry A **101**, 5446 (1997).

[122] S. Joe and F. Y. Kuo, ACM Trans. Math. Softw. **29**, 4957 (2003).

[123] S. Joe and F. Y. Kuo, SIAM Journal on Scientific Computing **30**, 2635 (2008), Journal Article.

[124] D. E. Williams, Journal of Computational Chemistry **22**, 1154 (2001).

[125] MOLPAK (MOLecular PAcKing), https://sourceforge.net/projects/molpak, Accessed: 2020-08-21.

[126] NEIGHCRYS, http://www.chem.ucl.ac.uk/cposs/dmacrys/manuals/neighcrys.pdf, Accessed: 2021-02-24.

[127] D. Williams, Journal of Molecular Structure **485-486**, 321 (1999).

[128] D. E. Williams, Journal of Computational Chemistry **22**, 1 (2001).

[129] D. S. Coombes, S. L. Price, D. J. Willock and M. Leslie, The Journal of Physical Chemistry **100**, 7352 (1996).

[130] S. R. Cox, L.-Y. Hsu and D. E. Williams, Acta Crystallographica Section A **37**, 293 (1981).

[131] D. E. Williams and S. R. Cox, Acta Crystallographica Section B **40**, 404 (1984).

[132] D. E. Williams and D. J. Houpt, Acta Crystallographica Section B **42**, 286 (1986).

[133] J. Nyman, O. S. Pundyke and G. M. Day, Phys. Chem. Chem. Phys. **18**, 15828 (2016).

[134] W. Smith and C. Daresbury, (2007).

[135] B. Stamm, L. Lagardère, É. Polack, Y. Maday and J.-P. Piquemal, The Journal of Chemical Physics **149**, 124103 (2018).

[136] C. G. Broyden, IMA Journal of Applied Mathematics **6**, 76 (1970).

[137] R. Fletcher, The Computer Journal **13**, 317 (1970).

[138] D. Goldfarb, Mathematics of Computation **24**, 23 (1970).

[139] D. Shanno, Mathematics of Computation **24**, 647 (1970).

[140] DMAREL 3.02, https://www.ucl.ac.uk/~ucca17p/dmarelmanual/full.html, Accessed: 2021-02-05.

[141] D. J. Willock, S. L. Price, M. Leslie and C. R. A. Catlow, Journal of Computational Chemistry **16**, 628 (1995).

[142] J. Nocedal and S. Wright, *Numerical Optimization* Springer Series in Operations Research and Financial Engineering (Springer New York, 2006).

[143] C. Y. Cheng, J. E. Campbell and G. M. Day, Chem. Sci. **11**, 4922 (2020).

[144] J. E. Campbell, *Crystal structure prediction of organic semiconductors*, PhD thesis, University of Southampton, 2017.

[145] D. Weininger, Journal of Chemical Information and Computer Sciences **28**, 31 (1988).

[146] D. Weininger, A. Weininger and J. L. Weininger, Journal of Chemical Information and Computer Sciences **29**, 97 (1989).

[147] Daylight Chemical Information Systems, Inc., http://www.daylight.com/dayhtml/doc/theory/theory.smarts.html, Accessed: 2020-08-14.

[148] R. Schmidt *et al.*, Journal of Chemical Information and Modeling **59**, 2560 (2019).

[149] S. R. Heller, A. McNaught, I. Pletnev, S. Stein and D. Tchekhovskoi, Journal of Cheminformatics **7**, 23 (2015).

[150] J. E. Campbell, J. Yang and G. M. Day, J. Mater. Chem. C **5**, 7574 (2017).

[151] H. Yoshida, Journal of Electron Spectroscopy and Related Phenomena **204**, 116 (2015), Organic Electronics.

[152] H. Yoshida, K. Yamada, J. Tsutsumi and N. Sato, Phys. Rev. B **92**, 075145 (2015).

[153] M. J. Frisch *et al.*, Gaussian 09 Revision E.01, Gaussian Inc. Wallingford CT 2009.

[154] S. Wang *et al.*, Chem. Commun. **51**, 11972 (2015).

[155] S. Wang *et al.*, CCDC 1052646: Experimental crystal structure determination, 2015.

[156] M. Brinkmann *et al.*, Journal of the American Chemical Society **122**, 5147 (2000).

[157] M. Brinkmann *et al.*, CCDC 147303: Experimental crystal structure determination, 2001.

[158] J. Wang, J.-W. Ye and Y. Wang, Acta Crystallographica Section E **63**, o2007 (2007).

[159] J. Wang, J.-W. Ye and Y. Wang, CCDC 643064: Experimental crystal structure determination, 2007.

[160] P. J. Low *et al.*, J. Mater. Chem. **15**, 2304 (2005).

[161] P. Low *et al.*, CCDC 257078: Experimental crystal structure determination, 2005.

[162] W. J. Begley and M. Rajeswaran, Acta Crystallographica Section E **62**, m1200 (2006).

[163] W. Begley and M. Rajeswaran, CCDC 610806: Experimental crystal structure determination, 2006.

[164] G. Paternò *et al.*, J. Mater. Chem. C **1**, 5619 (2013).

[165] G. Paternò *et al.*, CCDC 942996: Experimental crystal structure determination, 2013.

[166] H. Yoshida and K. Yoshizaki, Org. Electron. **20**, 24 (2015).

[167] H. Yoshida, The Journal of Physical Chemistry C **118**, 24377 (2014).

[168] Y. Zhong *et al.*, The Journal of Physical Chemistry C **119**, 23 (2015).

[169] W. Han, H. Yoshida, N. Ueno and S. Kera, Applied Physics Letters **103**, 123303 (2013).

[170] K. Yamada *et al.*, Phys. Rev. B **97**, 245206 (2018).

[171] H.-Y. Chen and I. Chao, ChemPhysChem **7**, 2003 (2006).

[172] M. Winkler and K. N. Houk, Journal of the American Chemical Society **129**, 1805 (2007).

[173] G. te Velde *et al.*, J. Comput. Chem. **22**, 931 (2001).

[174] A. Kubas *et al.*, The Journal of Chemical Physics **140**, 104105 (2014).

[175] A. Kubas *et al.*, Phys. Chem. Chem. Phys. **17**, 14342 (2015).

[176] W. Kabsch, Acta Cryst. A **32**, 922 (1976).

[177] Z. Liang *et al.*, Chemistry of Materials **22**, 6438 (2010).

[178] Z. Liang *et al.*, CCDC 829330: Experimental crystal structure determination, 2011.

[179] Z. Liang *et al.*, CCDC 829331: Experimental crystal structure determination, 2011.

[180] K. Fuchibe *et al.*, Journal of Fluorine Chemistry **203**, 173 (2017), Special issue in honor of Antonio Togni.

[181] K. Fuchibe *et al.*, CCDC 1545329: Experimental crystal structure determination, 2017.

[182] Y. Sakamoto *et al.*, Journal of the American Chemical Society **126**, 8138 (2004).

[183] Y. Sakamoto *et al.*, Molecular Crystals and Liquid Crystals **444**, 225 (2006).

[184] Y. Sakamoto *et al.*, CCDC 234729: Experimental crystal structure determination, 2004.

[185] J. Iball, S. N. Scrimgeour and D. W. Young, Acta Crystallographica Section B **32**, 328 (1976).

[186] G. Reck and B. Schulz, CCDC 613217: Experimental crystal structure determination, 2006.

[187] C. J. Carrell, T. G. Carrell, H. L. Carrell, K. Prout and J. P. Glusker, Carcinogenesis **18**, 415 (1997).

[188] C. J. Carrell, T. G. Carrell, H. L. Carrell, K. Prout and J. P. Glusker, CCDC 1830498: Experimental crystal structure determination, 2018.

[189] R. Mason and K. Y. Lonsdale, Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences **258**, 302 (1960).

[190] C. J. Carrell, T. G. Carrell, H. L. Carrell, K. Prout and J. P. Glusker, CCDC 1830499: Experimental crystal structure determination, 2018.

[191] R. F. Entwistle, J. Iball, W. D. S. Motherwell and B. P. Thompson, Acta Crystallographica Section B **25**, 770 (1969).

[192] J. Iball, C. H. Morgan and D. E. Zacharias, J. Chem. Soc., Perkin Trans. 2 , 1271 (1975).

[193] J. M. Robertson and J. G. White, J. Chem. Soc. , 925 (1956).

[194] K. C. Rippy *et al.*, Chemistry A European Journal **22**, 874 (2016).

[195] K. C. Rippy *et al.*, CCDC 1412574: Experimental crystal structure determination, 2015.

[196] L. A. Chetkina, Kristallografiya (1967).

[197] G. L. Eakins, J. S. Alford, B. J. Tiegs, B. E. Breyfogle and C. J. Stearman, Journal of Physical Organic Chemistry **24**, 1119 (2011).

[198] G. Eakins, J. Alford, B. Tiegs, B. Breyfogle and C. Stearman, CCDC 801984: Experimental crystal structure determination, 2012.

[199] T. Mala'bi, S. Pogodin, S. Cohen and I. Agranat, RSC Adv. **3**, 21797 (2013).

[200] T. Mala'bi, S. Pogodin, S. Cohen and I. Agranat, CCDC 920382: Experimental crystal structure determination, 2013.

[201] J. T. Mague and R. A. Pascal, CCDC 1404068: Experimental crystal structure determination, 2015.

[202] D. G. Morris, S. Higgins, K. S. Ryder, R. A. Howie and K. W. Muir, Acta Crystallographica Section C **56**, 570 (2000).

[203] D. Morris, S. Higgins, K. Ryder, R. Howie and K. Muir, CCDC 145535: Experimental crystal structure determination, 2000.

[204] Y. Takeda, M. Okazaki and S. Minakata, Chem. Commun. **50**, 10291 (2014).

[205] Y. Takeda, M. Okazaki and S. Minakata, CCDC 1004407: Experimental crystal structure determination, 2014.

[206] M. Day, X. Amashukeli and H. Gray, CCDC 142042: Experimental crystal structure determination, 2002.

[207] N. Ukwitegetse *et al.*, Chemistry A European Journal **25**, 1472 (2019).

[208] N. Ukwitegetse *et al.*, CCDC 1866946: Experimental crystal structure determination, 2019.

[209] S. Tan, X. Wu, Y. Zheng and Y. Wang, Chinese Chemical Letters **30**, 1951 (2019), Organic & Polymeric Luminescent Materials.

[210] S. Tan, X. Wu, Y. Zheng and Y. Wang, CCDC 1877247: Experimental crystal structure determination, 2019.

[211] B. Bovio and S. Locchi, Zeitschrift für Kristallographie **121**, 306 (1965).

[212] T. Mala'bi, S. Pogodin, S. Cohen and I. Agranat, CCDC 920378: Experimental crystal structure determination, 2013.

[213] L. A. Chetkina and E. G. Popova, Kristallografiya (1973).

[214] O. Kataeva *et al.*, Crystal Growth & Design **19**, 5123 (2019).

[215] K. J. Young, X. Bu and W. C. Kaska, Journal of Organometallic Chemistry **696**, 3992 (2011), SI:Small Molecule Activation and Catalysis Invoking Metal-Carbon Multiple Bonds.

[216] K. Young, X. Bu and W. Kaska, CCDC 838937: Experimental crystal structure determination, 2012.

[217] R. B. Campbell, J. M. Robertson and J. Trotter, Acta Crystallographica **15**, 289 (1962).

[218] D. Holmes, S. Kumaraswamy, A. J. Matzger and K. P. C. Vollhardt, Chemistry A European Journal **5**, 3399 (1999).

[219] D. Holmes, S. Kumaraswamy, A. Matzger and K. Vollhardt, CCDC 114447: Experimental crystal structure determination, 2000.

[220] T. Siegrist *et al.*, CCDC 145333: Experimental crystal structure determination, 2001.

[221] C. C. Mattheus *et al.*, Acta Crystallographica Section C **57**, 939 (2001).

[222] C. C. Mattheus *et al.*, CCDC 170186: Experimental crystal structure determination, 2001.

[223] C. C. Mattheus *et al.*, CCDC 170187: Experimental crystal structure determination, 2001.

[224] T. Siegrist *et al.*, Advanced Materials **19**, 2079 (2007).

[225] T. Siegrist *et al.*, CCDC 619981: Experimental crystal structure determination, 2007.

[226] S. Haas *et al.*, Phys. Rev. B **76**, 205203 (2007).

[227] T. Siegrist *et al.*, CCDC 619978: Experimental crystal structure determination, 2007.

[228] T. Siegrist *et al.*, CCDC 619979: Experimental crystal structure determination, 2007.

[229] T. Siegrist *et al.*, CCDC 619980: Experimental crystal structure determination, 2007.

[230] T. Siegrist *et al.*, CCDC 619982: Experimental crystal structure determination, 2007.

[231] S. Schiefer, M. Huth, A. Dobrinevski and B. Nickel, Journal of the American Chemical Society **129**, 10316 (2007).

[232] S. Schiefer, M. Huth, A. Dobrinevski and B. Nickel, CCDC 665900: Experimental crystal structure determination, 2008.

[233] S. Schiefer, M. Huth, A. Dobrinevski and B. Nickel, CCDC 674030: Experimental crystal structure determination, 2008.

[234] S. Schiefer, M. Huth, A. Dobrinevski and B. Nickel, CCDC 674031: Experimental crystal structure determination, 2008.

[235] A. V. Dzyabchenko, V. E. Zavodnik and V. K. Belsky, Acta Crystallographica Section B **35**, 2250 (1979).

[236] I. Salzmann *et al.*, Crystal Growth & Design **11**, 600 (2011).

[237] I. Salzmann *et al.*, CCDC 835494: Experimental crystal structure determination, 2011.

[238] H. J. Nam, Y. J. Kim and D.-Y. Jung, Bulletin of the Korean Chemical Society **31**, 2413 (2010).

[239] A. Camerman, J. Trotter and J. M. Robertson, Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences **279**, 129 (1964).

[240] T. M. Krygowski, A. Ciesielski, B. Swirska and P. Leszczynski, Polish Journal of Chemistry (1994).

[241] C. Näther, H. Bock, Z. Havlas and T. Hauck, Organometallics **17**, 4707 (1998).

[242] M. Botoshansky, F. Herbstein and M. Kapon, Helvetica Chimica Acta **86**, 1113 (2003).

[243] M. Botoshansky, F. Herbstein and M. Kapon, CCDC 198723: Experimental crystal structure determination, 2003.

[244] M. Botoshansky, F. Herbstein and M. Kapon, CCDC 198724: Experimental crystal structure determination, 2003.

[245] A. Ranganathan and G. U. Kulkarni, Journal of Chemical Sciences **115**, 637 (2003).

[246] A. Ranganathan and G. U. Kulkarni, CCDC 215338: Experimental crystal structure determination, 2004.

[247] M. Bolte, CCDC 1546182: Experimental crystal structure determination, 2017.

[248] D. G. Morris, K. S. Ryder and R. A. Howie, Acta Crystallographica Section C **54**, 1542 (1998).

[249] D. Morris, K. Ryder and R. Howie, CCDC 130472: Experimental crystal structure determination, 1998.

[250] J. Schwaben *et al.*, Chemistry A European Journal **21**, 13758 (2015).

[251] J. Schwaben *et al.*, CCDC 1051997: Experimental crystal structure determination, 2015.

[252] P. Fernandes, A. Florence, K. Shankland and W. I. F. David, Acta Crystallographica Section E **61**, o1483 (2005).

[253] P. Fernandes, A. Florence, K. Shankland and W. David, CCDC 272043: Experimental crystal structure determination, 2005.

[254] C. J. Carrell, T. G. Carrell, H. L. Carrell, K. Prout and J. P. Glusker, CCDC 1830500: Experimental crystal structure determination, 2018.

[255] S. Banerjee *et al.*, The Journal of Organic Chemistry **81**, 3983 (2016).

[256] S. Banerjee *et al.*, CCDC 1483904: Experimental crystal structure determination, 2016.

[257] S. Banerjee *et al.*, CCDC 1483905: Experimental crystal structure determination, 2016.

[258] S. Banerjee *et al.*, CCDC 1483906: Experimental crystal structure determination, 2016.

[259] S. Djurdjevic *et al.*, Journal of the American Chemical Society **129**, 476 (2007).

[260] S. Djurdjevic *et al.*, CCDC 642357: Experimental crystal structure determination, 2007.

[261] V. García-López *et al.*, Tetrahedron **74**, 5615 (2018).

[262] V. García-López *et al.*, CCDC 1535910: Experimental crystal structure determination, 2018.

[263] T. Geiger *et al.*, Chemistry A European Journal **26**, 3420 (2020).

[264] T. Geiger *et al.*, CCDC 1915510: Experimental crystal structure determination, 2020.

[265] T. Geiger *et al.*, CCDC 1967949: Experimental crystal structure determination, 2020.

[266] M. Kirste, T. Brietzke, H.-J. Holdt and U. Schilde, Zeitschrift für Kristallographie - New Crystal Structures **234**, 1255 (2019).

[267] M. Kirste, T. Brietzke, H.-J. Holdt and U. Schilde, CCDC 1944261: Experimental crystal structure determination, 2019.

[268] Q. Tang, Z. Liang, J. Liu, J. Xu and Q. Miao, Chem. Commun. **46**, 2977 (2010).

[269] Q. Tang, Z. Liang, J. Liu, J. Xu and Q. Miao, CCDC 761663: Experimental crystal structure determination, 2010.

[270] D. Käfer, M. El Helou, C. Gemel and G. Witte, Crystal Growth & Design **8**, 3053 (2008).

[271] D. Käfer, M. El Helou, C. Gemel and G. Witte, CCDC 703584: Experimental crystal structure determination, 2009.

[272] D. Käfer, M. El Helou, C. Gemel and G. Witte, CCDC 703585: Experimental crystal structure determination, 2009.

[273] B. Shen *et al.*, The Journal of Organic Chemistry **83**, 3149 (2018).

[274] B. Shen *et al.*, CCDC 1812131: Experimental crystal structure determination, 2018.

[275] A. De, R. Ghosh, S. Roychowdhury and P. Roychowdhury, Acta Crystallographica Section C **41**, 907 (1985).

[276] M. Bolte, CCDC 991788: Experimental crystal structure determination, 2014.

[277] B. Mahns *et al.*, Crystal Growth & Design **14**, 1338 (2014).

[278] B. Mahns *et al.*, CCDC 972624: Experimental crystal structure determination, 2014.

[279] Z. Zhao, L. H. Britt and G. K. Murphy, Chemistry A European Journal **24**, 17002 (2018).

[280] Z. Zhao, L. H. Britt and G. K. Murphy, CCDC 1829172: Experimental crystal structure determination, 2018.

[281] C. Kitamura *et al.*, CCDC 723050: Experimental crystal structure determination, 2010.

[282] C. Kitamura *et al.*, CCDC 723051: Experimental crystal structure determination, 2010.

[283] C. Kitamura *et al.*, CCDC 723052: Experimental crystal structure determination, 2010.

[284] C. Kitamura *et al.*, CrystEngComm **9**, 644 (2007).

[285] C. Kitamura *et al.*, CCDC 642337: Experimental crystal structure determination, 2007.

[286] C. Kitamura *et al.*, CCDC 723053: Experimental crystal structure determination, 2010.

[287] C. Kitamura *et al.*, CCDC 723054: Experimental crystal structure determination, 2010.

[288] C. Kitamura *et al.*, European Journal of Organic Chemistry **2010**, 3033 (2010).

[289] Kitamura, C. *et al.*, CCDC 746563: Experimental crystal structure determination, 2010.

[290] Kitamura, C. *et al.*, CCDC 746565: Experimental crystal structure determination, 2010.

[291] C. Kitamura *et al.*, European Journal of Organic Chemistry **2010**, 2571 (2010).

[292] Kitamura, C. *et al.*, CCDC 761527: Experimental crystal structure determination, 2010.

[293] J. Tanaka, Bulletin of the Chemical Society of Japan **36**, 1237 (1963).

[294] P. E. Blöchl, Phys. Rev. B **50**, 17953 (1994).

[295] G. Kresse and D. Joubert, Phys. Rev. B **59**, 1758 (1999).

[296] J. P. Perdew, K. Burke and M. Ernzerhof, Phys. Rev. Lett. **78**, 1396 (1997).

[297] C. F. Macrae *et al.*, Journal of Applied Crystallography **53**, 226 (2020).

[298] I. J. Bruno *et al.*, Journal of Computer-Aided Molecular Design **11**, 525 (1997).

[299] A. N. Chernyshev, M. V. Chernysheva, P. Hirva, V. Y. Kukushkin and M. Haukka, Dalton Trans. **44**, 14523 (2015).

[300] A. N. Chernyshev, M. V. Chernysheva, P. Hirva, V. Y. Kukushkin and M. Haukka, CCDC 1029125: Experimental crystal structure determination, 2015.

[301] A. Gavezzotti, CrystEngComm **10**, 389 (2008).

[302] F. Gajdos *et al.*, J. Chem. Theory Comput. **10**, 4653 (2014).

[303] O. G. Ziogos and J. Blumberger, The Journal of Chemical Physics **155**, 244110 (2021).

[304] O. G. Ziogos and J. Blumberger, pyAOMlite, https://github.com/oziogos/pyAOMlite, Accessed: 2021-07-05.

[305] S. Fratini and S. Ciuchi, Phys. Rev. Research **2**, 013001 (2020).

[306] Y.-I. Park *et al.*, Chemistry of Materials **23**, 4038 (2011).

[307] T. Nakagawa, D. Kumaki, J.-i. Nishida, S. Tokito and Y. Yamashita, Chemistry of Materials **20**, 2615 (2008).

[308] S. L. Price, Acta Crystallographica Section B **69**, 313 (2013).