Crystal Structure Prediction and Thermodynamic Modelling of Chiral Molecules

Rebecca Kathryn Hylton

August 2017

Submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

to

University College London

Department of Chemistry

20 Gordon Street

London

WC1H 0AJ

United Kingdom

Declaration

I, Rebecca Kathryn Hylton, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

_____

Rebecca Kathryn Hylton

August 2017

Abstract

This thesis explores the potential of computed crystal energy landscapes as an aid in the rational design of chiral separation processes. Crystal structure prediction (CSP) methods have been used to explore the crystal energy landscapes of prototypical chiral systems and the corresponding lattice energies and properties are used to help explore the thermodynamics of these systems.

The crystal energy landscape of two very different chiral systems is explored. The small, but very flexible 3-chloromandelic acid molecule which can form strong hydrogen bonding motifs, and the rigid lactide molecule where the crystal structures are dominated by weak van der Waals forces. These crystal energy landscapes highlight the complexity of chiral molecules, particularly the enantiopure structures which tend to be high Z'. These systems demonstrate that the factors which influences the kinetics of crystallisation and growth are not yet adequately understood.

The accuracy of CSP methods was explored through the CCDC Blind Test on the supposedly rigid, pseudo-chiral structure XXII ([1,4]dithiino[2,3-c]isothiazole-3,5,6-tricarbonitrile). The crystal structure was successfully predicted within the submitted structures at a comparable rank to much more sophisticated prediction methods by other groups. This suggests that the CSP methods used in my research can give reliable results.

The sublimation cycle is an approach which can be used to support the rational design of chiral separation process by crystallisation. Lattice energy calculations and $k = 0$ phonon calculations were performed for the 3-chloromandelic acid, lactide and naproxen experimental structures. These results have been used in conjunction with experimental methods, performed by experimentalists at the MPI, Magdeburg, to explore the sublimation cycle. The methods proposed show promise for aiding chiral separation process design.

Acknowledgements

Firstly, I would like to thank my supervisors. Sally Price, for her support, advice and direction throughout the course of this project, and Matthias Stein for his continuous support and encouragement.

My greatest thanks to Louise Price for all she has done for me throughout my time at UCL, from providing invaluable knowledge and expertise on crystal structure prediction to her outstanding organisational and admin skills and constant support and encouragement. Thank you to Bob Lancaster for sharing with me your valuable industrial knowledge and all your wise words. My colleagues in G18 who have made my time at UCL both fun and interesting, and those at the MPI who have always made me feel most welcome. Thank you to my friends and family who have gone through the highs and lows with me and supported me through to the end.

Many thanks to my sponsors at the Max Plank Institute for Dynamics of Complex Technical Systems. Thank you to Hannes Buchholz who provided the experimental work on for the thermodynamics, I've enjoyed working with you and the opportunity we had to present our research together at European Symposium on Applied Thermodynamics (ESAT2014). Thank you to Andreas Seidel-Morgenstern and Heike Lorenz for all their time, energy and insight into the experimental work on chiral systems.

Thank you to my collaborators Simon Coles, Terry Threlfall and Amy Ellis at the University of Southampton and Colin Seaton (now at University of Bradford), who provided me with much of the experimental insights into my work on 3-chloromandelic acid and to Graham Tizzard for passing on his invaluable expertise in using XPac.

Thank you to Keith Refson who spent much time with me explaining CASTEP, your time was very much appreciated and to Rui Guo who has helped to find the optimal parameters for using CASTEP with organic crystal structures. Thank you to Gerit Brandenburg for your help and assistance with the electronic structure calculations and your contribution to the thermodynamics work.

Finally, this work would not have been possible without the financial support from the MPI and the UCL Impact Studentship. The computational infrastructure and support was funded by the Engineering and Physical Science Research Council (EPSRC) under grant number EP/K039229/1. The significant amount of travel between the groups at the MPI and UCL for collaboration throughout the project was funded by the ESPRC via a Global Engagements Grant to UCL. My travel to participate at ESAT2014 was funded by the UCL Graduate School and the UCL Department of Chemistry.

List of Publications

1. **Hylton, R. K.**; Tizzard, G. J.; Threlfall, T. L.; Ellis, A. L.; Coles, S. J.; Seaton, C. C.; Schulze, E.; Lorenz, H.; Seidel-Morgenstern, A.; Stein, M.; Price, S. L. Are the Crystal Structures of Enantiopure and Racemic Mandelic Acids Determined by Kinetics or Thermodynamics? *Journal of the American Chemical Society* **2015**, *137*, 11095-11104.

2. Reilly, A. M.; Cooper, R. I.; Adjiman, C. S.; Bhattacharya, S.; Boese, A. D.; Brandenburg, J. G.; Bygrave, P. J.; Bylsma, R.; Campbell, J. E.; Car, R.; Case, D. H.; Chadha, R.; Cole, J. C.; Cosburn, K.; Cuppen, H. M.; Curtis, F.; Day, G. M.; DiStasio Jr, R. A.; Dzyabchenko, A.; van Eijck, B. P.; Elking, D. M.; van den Ende, J. A.; Facelli, J. C.; Ferraro, M. B.; Fusti-Molnar, L.; Gatsiou, C.-A.; Gee, T. S.; de Gelder, R.; Ghiringhelli, L. M.; Goto, H.; Grimme, S.; Guo, R.; Hofmann, D. W. M.; Hoja, J.; **Hylton, R. K.**; Iuzzolino, L.; Jankiewicz, W.; de Jong, D. T.; Kendrick, J.; de Klerk, N. J. J.; Ko, H.-Y.; Kuleshova, L. N.; Li, X.; Lohani, S.; Leusen, F. J. J.; Lund, A. M.; Lv, J.; Ma, Y.; Marom, N.; Masunov, A. E.; McCabe, P.; McMahon, D. P.; Meekes, H.; Metz, M. P.; Misquitta, A. J.; Mohamed, S.; Monserrat, B.; Needs, R. J.; Neumann, M. A.; Nyman, J.; Obata, S.; Oberhofer, H.; Oganov, A. R.; Orendt, A. M.; Pagola, G. I.; Pantelides, C. C.; Pickard, C. J.; Podeszwa, R.; Price, L. S.; Price, S. L.; Pulido, A.; Read, M. G.; Reuter, K.; Schneider, E.; Schober, C.; Shields, G. P.; Singh, P.; Sugden, I. J.; Szalewicz, K.; Taylor, C. R.; Tkatchenko, A.; Tuckerman, M. E.; Vacarro, F.; Vasileiadis, M.; Vazquez-Mayagoitia, A.; Vogt, L.; Wang, Y.; Watson, R. E.; de Wijs, G. A.; Yang, J. Z.; Zhu, Q.; Groom, C. R. Report on the sixth blind test of organic crystal-structure prediction methods. *Acta Crystallographica Section B - Structural Science* **2016**, B72, 439-459.

3. Buchholz, H. K.; **Hylton, R. K.**; Brandenburg, J. G.; Seidel-Morgenstern, A.; Lorenz, H.; Stein, M.; Price, S. L. Thermochemistry of Racemic and Enantiopure Organic Crystals for Predicting Enantiomer Separation. *Crystal Growth & Design* **2017 ACS Editors' Choice,** 10.1021/acs.cgd.7b00582.

Table of Contents

List of Figures

List of Tables

List of Abbreviations and Symbols

CCDC	Cambridge Crystallographic Data Centre
CSD	Cambridge Structural Database
FDA	Food and Drugs Administration
API	Active Pharmaceutical Ingredient
CSP	Crystal Structure Prediction
CEL	Crystal Energy Landscape
MD	Molecular Dynamics
Z	The number of molecules in the asymmetric unit cell
Z'	The number of independent molecules in the asymmetric unit cell
RMSD	Root Mean Square Deviation
DFT	Density Functional Theory
DFT-D	Dispersion Corrected Density Functional Theory
TS	Tkatchenko-Sheffler Dispersion Correction
G06	Grimme 2006 Dispersion Correction
PBE	Perdew Burke and Ernzerhof exchange-correlation functional
PBE1PBE	Perdew Burke and Ernzerhof exchange-correlation functional with Hartree Fock exchange energy a.k.a PBE0
MP2	Møller-Plessett 2^{nd} order perturbation
B3LYP	Becke, three-parameter, Lee-Yang-Parr
PCM	Polarizable Continuum Model
DMA	Distributed Multipole Analysis
LAM	Local Approximate Model
Ψ_{mol}	General term for modelling methods that are derived from using electronic structure calculations on the molecule
Ψ_{crys}	General expression for periodic electronic structure calculations on the crystal
NMR	Nuclear Magnetic Resonance
IR	Infra Red
PXRD	Powder X-ray Diffraction
3ClMA	3-chloromandelic acid
Lactide	3,6-Dimethyl-1,4-dioxane-2,5-dione
Naproxen	2-(6-Methoxynaphthalen-2-yl)propanoic acid
XXII	[1,4]dithiino[2,3-c]isothiazole-3,5,6-tricarbonitrile
U_{latt}	Lattice Energy
U_{inter}	Intermolecular Energy
ΔE_{intra}	Intramolecular Energy Penalty
ZPE	Zero Point Energy
C_p	Isobaric heat capacity
C_V	Isochoric heat capacity
H_{corr}	Thermodynamic Correction
ΔH_{subl}	Enthalpy of Sublimation
ΔH_{solv}	Enthalpy of Solvation
x_{eu}	Eutectic composition
ΔG_{subl}	Free Energy of Sublimation

1 Introduction

1.1 Chirality

Chiral molecules are a prominent class of compound, commonly found in agrochemicals, flavours, fragrances and pharmaceuticals.[1] A molecule is considered chiral if there exists a second molecule that is identical in composition and functional structure, but has an alternate arrangement in 3D space such that they form non-superimposable mirror images of one another. Two chiral molecules that are mirror images of one another are said to be enantiomers. These are labelled R or S dependent on the arrangement of the functional groups surrounding the chiral centre. If a change in the configuration of a molecule brings a change in the chirality then the molecule, such as the bending in and out of plane of a ring system, is said to be pseudo-chiral (see Chapter 5.1.3.1).[2] Molecules that form superimposable mirror images of one another are said to be achiral.

When chiral molecules are placed in an achiral environment their chemical properties make it difficult to distinguish between them, for example they will have the same NMR and IR spectra. While the two enantiomers are almost identical this alternate arrangement in 3D space can lead to vastly different properties when placed in a chiral environment. For example, if chiral additives are placed in the solvent containing both enantiomers and a NMR spectrum taken, then the spectra of the two enantiomers will be different as the chiral additives will interact differently with each enantiomer. Our bodies are chiral environments and so when we are exposed to enantiomers their differing properties can be observed. For example, the chemical carvone is a small chiral molecule which is found naturally in essential oils, however the two enantiomers smell vastly different, with the R enantiomer smelling like spearmint and the S enantiomer smelling like caraway seeds.[3] While this is a relatively harmless, the differences in properties between different enantiomers are of particular importance for the pharmaceutical industry. The human body is a chiral environment and with many effective pharmaceuticals also being chiral it is important to understand the possible alternative effects each enantiomer might

have on the body. The potential negative effect of a drug molecule containing both enantiomers was highlighted by the use of the drug thalidomide to combat morning sickness in pregnant woman. Thalidomide was originally marketed as a relatively harmless sleeping pill as it was found during animal testing that it was virtually impossible to achieve a deadly overdose of the drug. However none of the animal tests looked into the effects of the drugs during pregnancy and it wasn't until a link was associated between birth defects and the drug that thalidomide was pulled from production in 1961.[4] Subsequent testing found that while one enantiomer has a sedative effect on the body, helping to ease the symptoms of morning sickness, the other causes teratogenic effects and that even if an enantiopure drug had been used, the drug is able to racemise in the body producing the harmful enantiomer.[4] Cases like this emphasised the potentially disastrous consequences of creating racemic drugs and the need to understand their behaviour in the body. In 1992 the Food and Drug Administration (FDA) declared that any new drug brought to the market must contain either a single enantiomer or it must be proved that neither enantiomer gives rise to harmful effects.[5] Thus economically viable methods in which enantiomers can be separated are required.

It is very rare that a substance will spontaneously resolve into a single enantiomer. More often than not, a mixture of the two enantiomers will form, a racemic mixture or racemate. The declaration by the FDA has led to extensive research in ways in which enantiopure molecules can be prepared, instead of a racemic mixture, in a manner which is both cost and time effective. There are a number of routes in which enantiopure molecules can be obtained,[6] as shown in Figure 1.1. These fall into three main categories: asymmetric synthesis, synthesis from the chiral pool and unselective synthesis followed by chiral separation. Asymmetric synthesis is the general term used to describe a reaction process whereby one of the enantiomers is produced more favourably than the other. The most effective of these types of reactions employ the use of a catalyst to enhance the outcome of the reaction.[7] One prominent example of asymmetric synthesis where a catalyst is employed is asymmetric hydrogenation.[8] An enantiomeric excess of greater than 90% can be achieved with this reaction. Unfortunately, the number of highly selective reactions that are both economically viable and lead to a pure enantiomer are rare. The

pharmaceutical industry requires a large array of different chiral molecules to formulate drugs. In order to synthesise the variety of chiral molecules required either a generic catalyst needs to be found which will allow any chiral molecule to be produced in the desired enantiomer, or a specific catalyst needs to be developed and optimised for each chiral molecule under development. Unfortunately, it is unlikely that a general catalyst will be found that can cater for the diverse nature of chemicals required. Often the ligands have to be specifically designed around the substrate which can also make finding suitable catalysts a lengthy and costly process.

An alternative synthetic method begins by taking fragments from the 'chiral pool'. These fragments are cheap molecules which already display the desired chirality. It is then hoped that the chirality is obtained throughout the entire synthesis, requiring extensive research into the development of the synthetic route. It simply may not be possible for the chirality to be maintained throughout the synthesis or that for 'new' chiral molecules, the starting block simply may not be available.

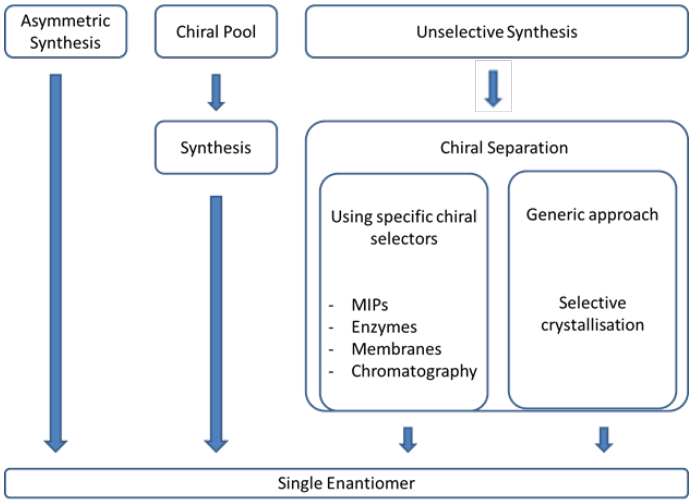


Figure 1-1: A summary of the methods in which single enantiomers can be obtained through experiment

The final method in which a single enantiomer may be obtained is via unselective synthesis followed by chiral separation. Chiral separation can occur via two ways, either by using a specific chiral selector or generic approaches such a selective crystallisation. Specific chiral selectors include the use of chiral chromatography, membranes, enzymes and molecular imprint polymers (MIPs) to separate out the two enantiomers. These methods tend to rely on the

selector being immobile which may mean this method is not always practically possible.[9,10]

All of the processes mentioned so far are specific to the molecule in question and cannot be readily adapted for a wide spread of chiral molecules. Separation via crystallisation methods, could potentially provide the most generic form of separation of enantiomers that is not compound specific. The physical properties of the two molecules (enantiomers) are the same and so experimental techniques for chiral separation by crystallisation often rely on the thermodynamic properties of the crystals involved. Therefore, effective crystallisation required the knowledge of kinetic and thermodynamic data of the system of interest as well as understanding the way in which a racemic mixture crystallises. Even so, separation via crystallisation can be more easily applied across a wide range of molecules as although each crystallisation will require specific conditions, the apparatus used is more generic.

There are three main ways in which a racemic mixture crystallises; either by forming a conglomerate, racemic compound or pseudoracemate. Depending on which of these crystal systems is obtained depends on whether a crystallisation technique can be used to separate the enantiomers.

A pseudoracemate occurs when both enantiomers coexist in an unordered manner, forming a solid solution. This is clearly the most difficult form to separate out the enantiomers, but it is very rare with less than 1% of racemic mixture able to form a pseudoracemate.

Some molecules spontaneously resolve due to the formation of a conglomerate; an equimolar mechanical mixture of crystals each one of which only contains one of the two enantiomers present in a racemate. In this case there are two main techniques that can be used to obtain the pure enantiomer: simultaneous crystallisation[6] or preferential crystallisation.[11] In simultaneous crystallisation a supersaturate solution of a racemic conglomerate is allowed to crystallise in two different crystallisers simultaneously. Each crystalliser is seeded with the respective enantiomer and filtered before the other enantiomer begins to crystallise. Preferential crystallisation is the process by which seeds of the R

enantiomer are added to a supersaturated enriched solution of R. This causes crystals of R to form in which creates a supersaturated solution of S. The cycle is then repeated with the S enantiomer used as a seed. This method of separation can rely heavily on kinetics which could lead to potential problems on industrial scale up.[12]

However, conglomerates only occur in approximately 10% of cases. The majority of racemates crystallise as racemic compounds in which both enantiomers are situated in crystals of a well ordered arrangement. It was traditionally thought that it was impossible to separate enantiomers which form racemic compounds, however recent advancements in separation techniques have now made this possible. Diastereomeric salt formation is the technique which is often used in industry. A diastereomeric salt is formed of each enantiomer with a chirally pure counter-ion. These salts will have different solubilities such that, provided the solubility difference is large enough, a solution can be formed such that one salt remains in solution while the second can be filtered off.[13] Thus extensive knowledge of the thermodynamic behaviour of the enantiomers in solution is required.

1.2 The connection between phase diagrams in the design of chiral separation processes

In order to design an efficient experiment in which enantiomers can be separated via crystallisation, detailed knowledge of the phase diagrams describing the melting (binary phase diagram) and/or solubility behaviour (ternary phase diagram) in the presence of a suitable solvent is required.

Binary phase diagrams describe the melting behaviour of a mixture as the ratios of the two components in the mixture are varied. In the case of chiral separation these components are the two hands of the molecule which can be obtained by measuring the melting profiles of enantiomeric mixtures. A key feature of the diagram is the eutectic point; this point denotes the ratio of mixtures at which a single composition is adopted. The position and number of eutectic points changes depending on the type of racemic mixture being studied. A conglomerate will have only one eutectic point, whereas a racemic mixture will have two eutectic points (see Figure 1.2).

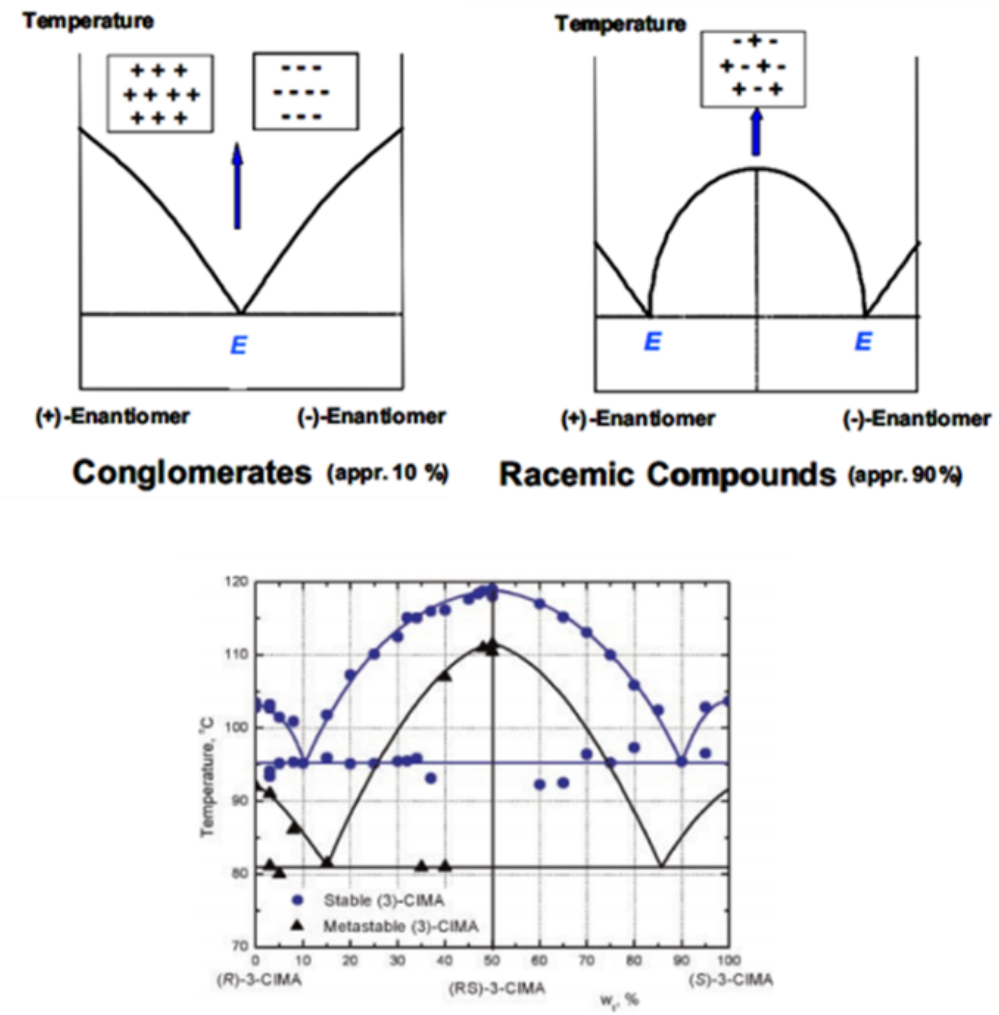


Figure 1-2: Ideal binary phase diagrams[14] for a conglomerate system and racemic compound (top) as well as the experimentally determined binary phase diagram for 3-chloromandelic acid (bottom). The circles and triangles represent the phase diagram for the stable and metastable compounds respectively[1].

A ternary phase diagram describes how the behaviour of a mixture in a solvent is varied with temperature, and is therefore solvent dependent. There are three distinct regions in a ternary phase diagram attributing to the conditions in which the mixture is in 1, 2 or 3 phases. The ternary phase diagram is constructed from solubility data which can be obtained from classical isothermal and/or polythermal methods. As ternary phase diagrams are an extension of binary phase diagrams, there is a difference in the diagrams representing conglomerates and racemic compounds (See Figure 1.3).[14] For a conglomerate system, the diagram consists of an under saturated 1-phase region, two 2-phase regions which under equilibrium conditions contain an enantiopure solid phase and a saturated liquid phase with one or both enantiomers and a 3-phase region where under equilibrium conditions the liquid phase is a saturated

solution of the two enantiomers. The ternary phase diagram for a racemic mixture is more complex with the presence of a second 2-phase region where the solid phase is the racemic compound and there are two separated 3-phase regions in which the solid phases are mechanical mixtures of a pure enantiomer and racemic compound.[14] It is the exploitation of these diagrams which allows efficient chiral separation techniques to be engineered.

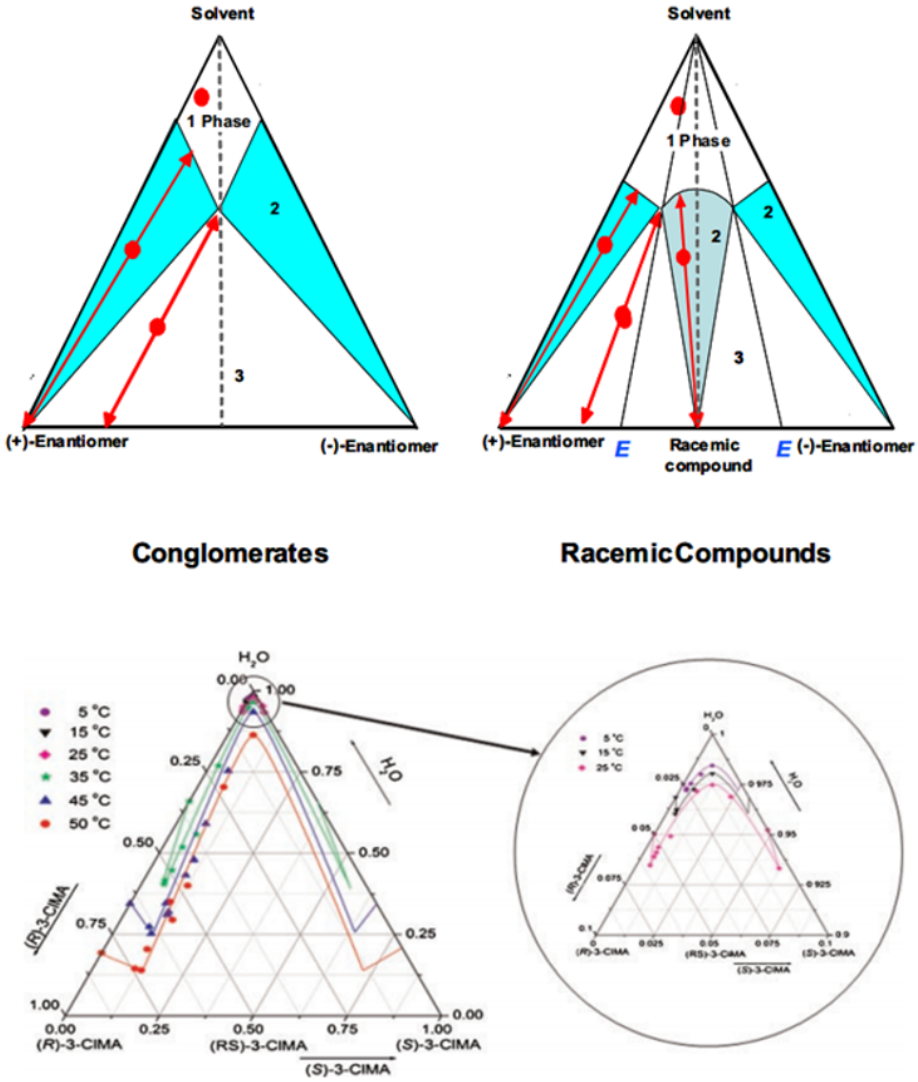


Figure 1-3: Ideal tertiary phase diagrams[14] for a conglomerate system and racemic compound (top) as well as the experimentally determined tertiary phase diagram for 3-chloromandelic acid in water[1] (bottom).

When studying binary and ternary phase diagrams it is paramount that you know exactly which crystal structure in which solvent is being studied. Different crystal structures will have different interactions with the solvent and so the binary and ternary phase diagrams will alter accordingly.

1.3 Polymorphism

Polymorphism is the ability of a crystal structure to crystallise or pack in a number of different structural forms.[15] It is a common phenomenon amongst crystal structures of organic molecules. The number of polymorphs within the Cambridge Structural Database (CSD) is ever increasing[16] as new experimental techniques, such as crystallisation under high pressure are used. One example of a molecule well known for the number of polymorphs it can form is 5-methyl-2-[(2-nitrophenyl)amino]-3-thiophene-carbonitrile, more commonly known as ROY due to the red, orange and yellow polymorphs it forms (see Figure 1.3). There are 10 known polymorphs of ROY,[17] which is one of the highest numbers of polymorphs recorded for a molecule currently in the CSD. The majority of these polymorphs are stable under atmospheric conditions allowing them to be widely studied.[17]

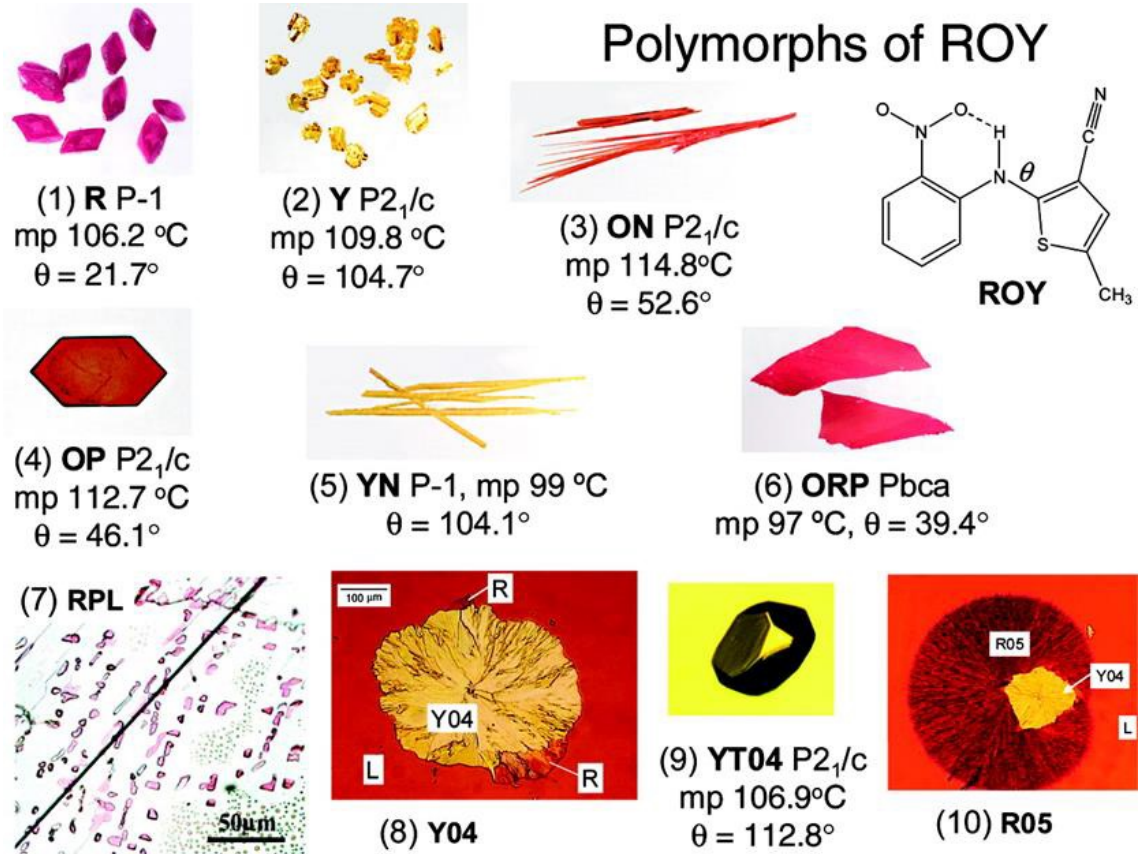


Figure 1-4: The 10 different polymorphs of ROY highlighting their differing colours, polymorphs and molecular conformations.[18]

While the molecules contained within the crystals of polymorphs are identical, the physical properties can vary widely. Properties such as stability, solubility and colour (particularly noticeable for ROY) can change dramatically between polymorphs, and, as a result, polymorphism is of great importance to many different branches of the chemical industry. One industry which is particularly

affected by polymorphism is the pharmaceutical industry. In addition to differences in melting point, physicochemical stability and mechanical properties, the solubilities and dissolution rates of different polymorphs can be significantly different, which in turn affects the bioavailablity of potential drug formulations.

Polymorphism is therefore an additional complication in the development of pharmaceuticals that needs to be carefully considered. Due to the differences in physiochemical properties, each polymorph could have significantly different binary and ternary phase diagrams. This can be observed with the experimental binary phase diagrams of 3-chloromandelic acid (shown in Figure 1.1) where the eutectic point shifts, depending on the polymorph in question and highlights the need to know exactly which polymorph is being targeted when designing chiral separation processes.

Polymorphism extends beyond the ability to reliably produce a drug molecule with the desired properties. Many drugs receive regulatory approval for only a single polymorph of the active pharmaceutical ingredient (API) and there is a need to protect the intellectual property of these APIs, with specific details of the polymorph being used required for patents. The discovery of late appearing polymorphs can have significant consequences for pharmaceutical companies, costing them millions of dollars and depriving patients with the medication in which they rely on. For example, a previously unknown, more stable, form of the drug Ritonavir was discovered late in development. This form was therapeutically ineffective, resulting in the drug needed to be reformulated in 1998.[19] Thus it is important from both a chiral resolution, medical and pharmaceutical perspective to be able to determine the stability of the targeted crystalline form and if there are potentially other polymorphs into which it might spontaneously resolve. One definition of polymorphism is that all polymorphs will have the same form in the melt and solution. Racemic and enantiopure crystals do not have the same melt and solution behaviour however it is possible for chiral compounds to racemise in solution further complicating the situation. Thus, more extensive experimental techniques and the use of computation is required to help explore the full extent of polymorphism in chiral systems.

Polymorphic screening is one of the tools that pharmaceutical companies use to help identify any other possible polymorphs. This involves recrystallizing the target compound under a large array of conditions to identify any additional polymorphs. This may include using different solvents, temperatures, pressures and additives. Due to the extensive nature of these screening experiments a sufficient amount of time and materials are required which may not always be possible and so computation can be used to help aid the screening process.

1.4 Identifying and predicting possible polymorphs

The structural differences between the racemic and enantiopure crystals, and each of their polymorphs, of a compound can be obtained via spectroscopic techniques and X-ray diffraction methods. Combined with experimental analysis, such as calorimetry to obtain the heat capacity, a complete picture of the atomic arrangement and thermodynamic stability can be built up across the entire phase diagram of the compound in question. However, there are limitations to the experimental work that can be done. A large amount of effort may be required to produce crystals of the desired polymorphs, let alone of a good enough quality to produce crystals suitable for single X-ray diffraction. It may not be possible to analyse a structure and its properties by experiment alone,[20] due to factors such as decomposition upon melting, for example. For a polymorphic screen to be carried out a large enough sample of the desired compound needs to be made, which may be difficult and costly to obtain and/or produce.

Computation can provide a means to aid in the identification and prediction of polymorphs. Crystal structure prediction (CSP) methods can be used to generate a set of thermodynamically feasible structures with an estimation of their relative stability at 0 K. These structures can be used to create a crystal energy landscape, in which each possible polymorph is plotted on a scatter diagram of energy vs. density. From the crystal energy landscape the most plausible stable polymorphs can be predicted. In computational studies, polymorphism is generally considered to occur within an energy range of 10 kJ mol^{-1} to account for any errors made during the calculations.[21] However this

energy range is also dependent on the compound in question, with larger, flexible molecules having a greater energy range than smaller, rigid molecules. If there are distinct structures found within the energy range of polymorphism it suggests that alternate stable crystal structures could be found.

There are three typical of crystal energy landscapes that can form, as depicted in Figure 1.5. The first case is that there is a single structure which has a far lower lattice energy than any of the other predicted polymorphs (Figure 1.5a). This is an ideal case where there is just one stable crystal structure with a low likelihood of an alternative stable crystal structure being able to form. Another possibility is that there is a second hypothetical crystal structure found lower in energy than the experimentally known structure (see in Figure 1.5b). This could indicate that there is an alternative, undiscovered polymorph of the target compound with which is more stable than the currently known experimental crystal structure. The final possibility is that there are a number of hypothetical crystal structures clustered around the global minimum of the crystal energy landscape (see Figure 1.5c). These crystal structures will be competitive and so the compound in question is likely to form different polymorphs.

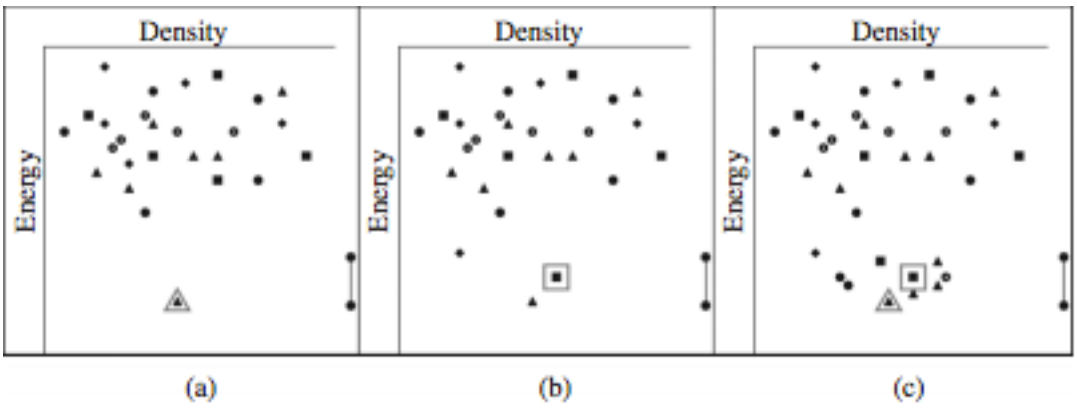


Figure 1-5: Stereotypical crystal energy landscapes.[22] Each symbol represents a unique, hypothetical crystal structure found in the CSP search. The large open symbols represent experimentally known crystal structures. The different symbols for the hypothetical crystal structures indicate unique structures with similar characteristics, such as space group or hydrogen bonding motif. The horizontal bar indicates the range of polymorphism.

The production of a crystal energy landscapes can be a useful tool to aid experimental screening processes. If there is only a small amount of the compound of interest available, then computed crystal energy landscapes can help indicate if there are any other competitive polymorphs which could possibly form as well as aid in designing experiments to find them. Any hypothetical crystal structures found lower in energy in the search than the experimentally

known polymorphs can provide potential targets for crystallisation experiments. Analysis of the packings and hydrogen bonding capabilities of the hypothetical crystal structures could additionally point experimentalists in the direction of additional metastable polymorphs. If the hypothetical structures show packings or motifs similar to those seen in analogous crystal structures to the target compound, templating experiments can be designed using seeds of the analogous structure to see if the targeted polymorph can be found. Experiments of this type have been used to successfully discover a new polymorph of cyheptamide.[23]

1.5 Determination of solubility differences of chiral compounds

There are a number of steps a potential drug molecule must go through in order to reach clinical trials. One of the main causes of for the failure of potential drug molecules is poor aqueous solubility. If it was possible to accurately predict the solubility of organic crystals then vast amounts of time and money could be saved in the drug development process.

In order to effectively separate out enantiomers, it is important to be able to accurately obtain the relative stability of the enantiopure and racemic crystal structures, as the separation methods mentioned in Section 1.1 often rely on these differences. Therefore, methods are required which allow accurate relative energies to be determined, either by experiment or computation methods.

Closely related to the relative energies are the relative solubilities of enantiopure and racemic compounds. New methods have been developed which exploit the differences in solubility of the enantiopure and racemic phases to achieve separation via crystallisation. This requires explicit knowledge of the solubility equilibria for both the enantiopure and racemic solid phases, and thus determination of the ternary phase diagram. Obtaining solubility data on a particular compound requires a performing a number of experiments which can be time consuming and requires sufficient amounts of material, therefore the ability to determine solubilities from calculation can aid in the design process, particularly in early stages of development.

Studies and blind tests[24] have been performed to assess the current situation in the ability to predict aqueous solubility from a crystal structure. There has been relative success in predicting solubility through the use of machine learning methods.[25,26] However, the ability to determine absolute solubilities, from first principles, of a compound in the range of solvents required for crystallisation studies is still relatively inaccurate. It is therefore hoped that by calculating solubility differences a cancellation of errors would occur such that a more accurate value can be determined.

1.6 Can we design chiral separation processes from the computational estimates of solubility differences?

It is now possible to determine the relative lattice energies of organic crystal structures to a reasonable degree of accuracy. But are these energies reliable enough to determine the solubility differences between racemic and enantiopure crystal structures?

Effective separation of chiral molecules relies on the exploitation of solubility differences. In principle, computational models could be used to predict the quantitative thermodynamic data needed for the design of chiral separation processes via crystallisation. As the molecules within racemic and enantiopure crystal structures are the same, the assumption is often made that many terms such as the zero point energy and heat capacity of the crystals will be the same, and therefore cancel. The enthalpy of solution can then simply be related to the difference in lattice energy between the two crystal structures.

$$\Delta\Delta H_{sol} = \Delta H_{sol}(S) - \Delta H_{sol}(RS) \approx \Delta U_{latt}$$

Equation 1-1

This relationship and the associated assumptions has been explored with the nonsteroidal anti-inflammatory drug naproxen.[20] In this study it was shown, by using a thermodynamic cycle, equating the solubility directly to the lattice energy to determine solubility differences is not accurate enough to predict the experimental values. This raised the question as to whether the small energy terms in the various thermodynamic cycles that can relate the thermodynamic properties of the two crystals and the gas phase, such as differences in the zero

point energies and heat capacities can make an important contribution (see Figure 1.6), in addition to the transfer energy and heats of mixing, which were already identified. It is typically more convenient to measure the Gibbs energy of solubility, ΔG_{sol}, which also includes entropic contributions that are also of significance when relating solubility differences to lattice energy (see Chapter 6).

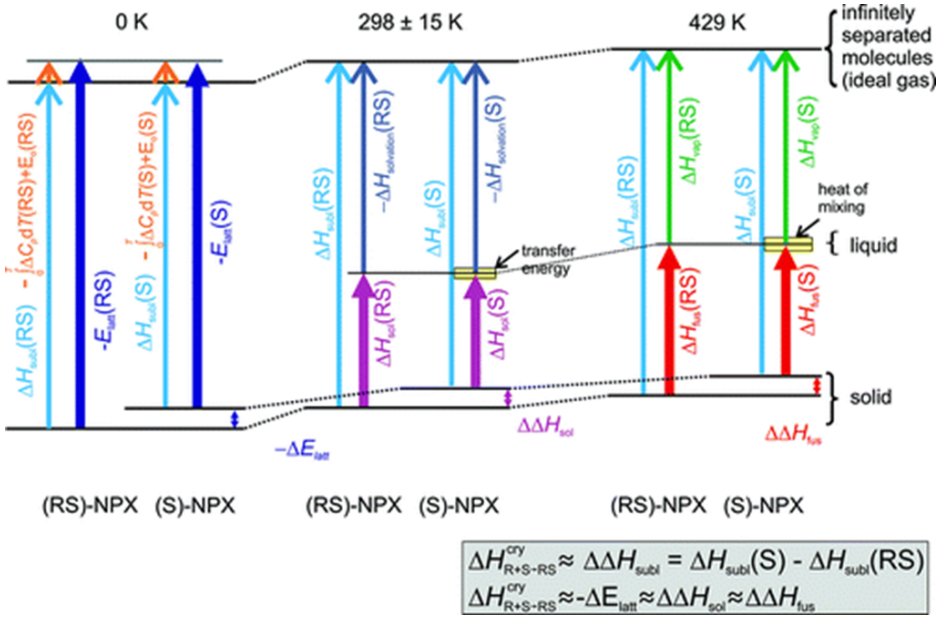


Figure 1-6: Summary of the measured and computationally determined thermodynamic quantities and other enthalpic contributions for the solid and gas phases of naproxen.[20]

The problems in predicting solubility differences can be exacerbated because there is an exponential relationship between solubility, x, and the free energy of solution ΔG_{sol}, (see Chapter 2.9), although it is conventional to predict $\log_{10}S$ values which are directly proportional to ΔG_{sol}. The accuracy of the sublimation enthalpy is not only dependent on the precision of the lattice energy but also on the small energy contributions within the thermodynamic cycle, such as heat capacity, zero point energy and solvation energy. These small contributions are often grouped into a –2RT thermal correction, H_{corr}, which is based on the assumption that $C_v^s = 6RT$ for all crystals between 0 K and the sublimation temperature and that the zero point vibrational energy is the same for the solid and gas, assuming that the gas is ideal. Recent determinations of this thermal correction term using electronic structure modelling has been explored on a set of 23 small, single component, organic crystal structures[27-29] referred to as the X23 benchmark set. Calculations using on the X23 set show that the –2RT thermal correction is a poor approximation, with the magnitude of the thermal

correction being dependent on the crystal being studied (see Figure 1.7). Not only do these results show the variance in the thermal correction away from the –2RT approximation across different crystal structures but they also show the lack of convergence in obtaining these values across different electronic structure methods. Even so, it is still hoped that many of the errors will cancel when determining differences in heat capacity between racemic and enantiopure crystal structures.

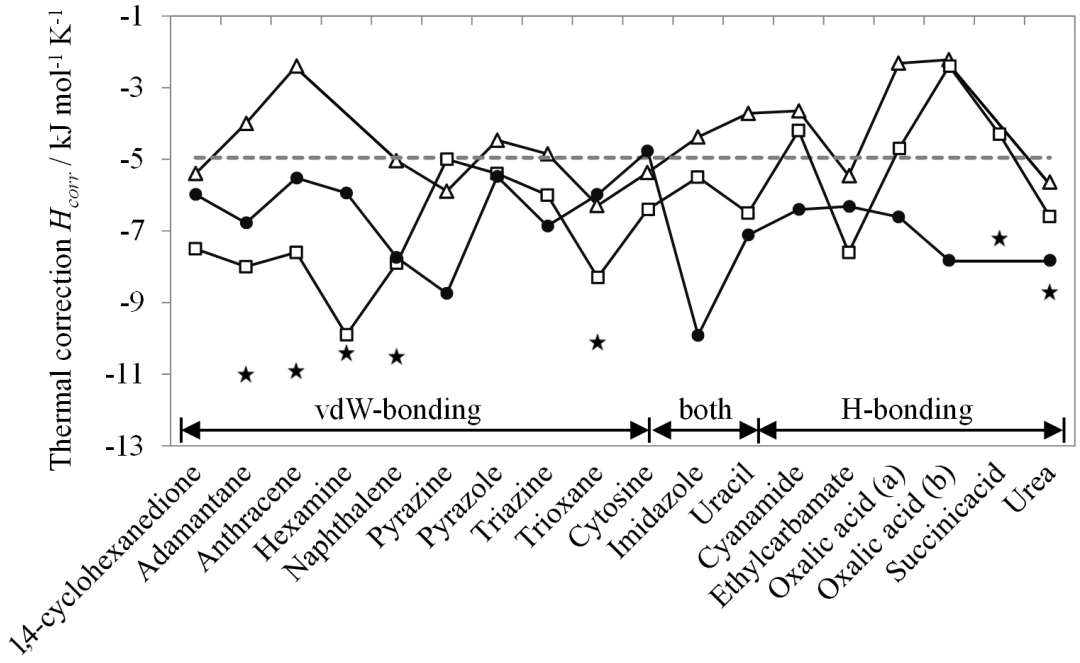


Figure 1-7: Calculated thermal contributions to the sublimation energy at 298 K for a set of 23 small organic crystals. Corrections in the harmonic approximation from different electronic structure methods are given: PBE-XDM (open triangles[28]), PBE-TS (filled squares[29]), and DFTB3-D3 (filled circles[27]) and compared to the –2RT-approximation (horizontal line): $\Delta \boldsymbol{H_{subl}(T)} = -\boldsymbol{U_{latt}} - \boldsymbol{2RT}$. For six systems estimations of anharmonic contributions are given (filled stars[29]).

1.7 Scope and Outline of thesis

This thesis aims to explore the crystal energy landscapes of three very different organic chiral systems, 3-choromandelic acid, lactide and Blind Test molecule XXII,[30] and evaluate how crystal structure prediction can be used to help design chiral separation processes.

Chapter 2 will explore the theoretical background behind the many crystal structure prediction techniques that have been used for my research. This will include discussions on the theory of intermolecular forces, crystal structure prediction methods, and density functional theory. It will also explore the many thermodynamic equations and their assumptions which are needed for

designing chiral separation processes. Chapters 3-6 contain the original research for this thesis. Chapter 3 will discuss the crystal energy landscape of the relatively small but very flexible 3-chloromandelic acid and rationalises the complexity. Chapter 4 will explore the crystal energy landscape of the rigid lactide system, and the accuracy of modelling structures which are dominated by weak intermolecular forces. Chapter 5 looks at the accuracy of CSP in predicting crystal structures without knowing the experimental result, as part of an international formal Blind Test study. Chapter 6 utilises the understanding developed in Chapters 3-5 to explore the use of crystal structure prediction to design chiral separation process as part of a collaborative research project with experimentalists at the Max Planck Institute, Magdeburg. Finally, Chapter 7 will summarise the findings of my PhD project and outline areas for future work.

1.8 References

(1) Le Minh, T.; Langermann, J. v.; Lorenz, H.; Seidel-Morgenstern, A. Enantiomeric 3-Chloromandelic Acid System: Binary Melting Point Phase Diagram, Ternary Solubility Phase Diagrams and Polymorphism. J*ournal of Pharmaceutical Sciences* **2010**, *99*, 4084-4095.

(2) Nourse, J. G. Pseudochirality. *Journal of the American Chemical Society* **1975**, *97*, 4594-4601.

(3) de Carvalho, C. C. C. R.; da Fonseca, M. M. R. Carvone: Why and how should one bother to produce this terpene. *Food Chemistry* **2006**, *95*, 413-422.

(4) Blaschke, G.; Kraft, H. P.; Fickentscher, K.; KÇôhler, F. Chromatographic separation of racemic thalidomide and teratogenic activity of its enantiomers (author's transl). *Arzneimittel-Forschung* **1979**, *29*, 1640-1642.

(5) Food and Drug Agency, R. *Development of New Stereoisomeric Drugs*1992.

(6) Crosby, J.: Chirality in industry - An overview. In *Chirality in industry*; Collins, A. N., Sheldrake, G. N., Crosby, J., Eds., 1992.

(7) Ault, A. The Nobel Prize in Chemistry for 2001. *Journal of Chemical Education* **2002**, *79*, 572.

(8) Noyori, R. Asymmetric Catalysis: Science and Opportunities (Nobel Lecture 2001). *Advanced Synthesis & Catalysis* **2003**, *345*, 15-32.

(9) Xie, R.; Chu, L. Y.; Deng, J. G. Membranes and membrane processes for chiral resolution. *Chemical Society Reviews* **2008**, *37*, 1243-1263.

(10) Stalcup, A. M. Chiral Separations. *Annual Review of Analytical Chemistry* **2010**, *3*, 341-363.

(11) Levilain, G.; Eicke, M. J.; Seidel-Morgenstern, A. Efficient Resolution of Enantiomers by Coupling Preferential Crystallization and Dissolution. Part 1: Experimental Proof of Principle. *Crystal Growth & Design* **2012**, *12*, 5396-5401.

(12) Le Minh, T.; Lorenz, H.; Seidel-Morgenstern, A. Enantioselective Crystallization Exploiting the Shift of Eutectic Compositions in Solid-Liquid Phase Diagrams. *Chemical Engineering & Technology* **2012**, *35*, 1003-1008.

(13) Lorenz, H.; Seidel-Morgenstern, A. Processes To Separate Enantiomers. *Angewandte Chemie International Edition* **2014**, *53*, 1218-1250.

(14) Lorenz, H.; Perlberg, A.; Sapoundjiev, D.; Elsner, M. P.; Seidel-Morgenstern, A. Crystallization of enantiomers. *Chemical Engineering and Processing* **2006**, *45*, 863-873.

(15) Price, S. L. Why don't we find more polymorphs? *Acta Crystallographica Section B: Structural Science, Crystal Engineering and Materials* **2013**, *69*, 313-328.

(16) Allen, F. H. The Cambridge Structural Database: A quarter of a million crystal structures and rising. *Acta Crystallographica Section B - Structural Science* **2002**, *58*, 380-388.

(17) Vasileiadis, M.; Kazantsev, A. V.; Karamertzanis, P. G.; Adjiman, C. S.; Pantelides, C. C. The polymorphs of ROY: application of a systematic crystal structure prediction technique. *Acta Crystallographica Section B-Structural Science* **2012**, *68*, 677-685.

(18) Yu, L. Polymorphism in Molecular Solids: An Extraordinary System of Red, Orange, and Yellow Crystals. *Accounts of Chemical Research* **2010**, *43*, 1257-1266.

(19) Bauer, J.; Spanton, S.; Henry, R.; Quick, J.; Dziki, W.; Porter, W.; Morris, J. Ritonavir: An extraordinary example of conformational polymorphism. *Pharmaceutical Research* **2001**, *18*, 859-866.

(20) Braun, D. E.; Ardid-Candel, M.; D'Oria, E.; Karamertzanis, P. G.; Arlin, J. B.; Florence, A. J.; Jones, A. G.; Price, S. L. Racemic Naproxen: A Multidisciplinary Structural and Thermodynamic Comparison with the Enantiopure Form. *Crystal Growth & Design* **2011**, *11*, 5659-5669.

(21) Cruz-Cabeza, A. J.; Reutzel-Edens, S. M.; Bernstein, J. Facts and fictions about polymorphism. *Chemical Society Reviews* **2015**, *44*, 8619-8635.

(22) Price, S. L.; Price, L. S.: Computational Polymorph Prediction. In *Solid State Characterisation of Pharmaceuticals*; Storey, R., Ymén, I., Eds.; John Wiley & Sons Ltd.: Chichester, West Sussex, UK, 2011; Vol. 3rd.

(23) Srirambhatla, V. K.; Guo, R.; Price, S. L.; Florence, A. J. Isomorphous template induced crystallisation: a robust method for the targeted crystallisation of computationally predicted metastable polymorphs. *Chemical Communications* **2016**, *52*, 7384-7386.

(24) Hopfinger, A. J.; Esposito, E. X.; Llinàs, A.; Glen, R. C.; Goodman, J. M. Findings of the Challenge To Predict Aqueous Solubility. *Journal of Chemical Information and Modeling* **2009**, *49*, 1-5.

(25) McDonagh, J. L.; Nath, N.; De Ferrari, L.; van Mourik, T.; Mitchell, J. B. Uniting Cheminformatics and Chemical Theory To Predict the Intrinsic Aqueous Solubility of Crystalline Druglike Molecules. *Journal of Chemical Information and Modeling* **2014**, *54*, 844-856.

(26) Palmer, D. S.; Llinas, A.; Morao, I.; Day, G. M.; Goodman, J. M.; Glen, R. C.; Mitchell, J. B. O. Predicting intrinsic aqueous solubility by a thermodynamic cycle. *Molecular Pharmaceutics* **2008**, *5*, 266-279.

(27) Brandenburg, J. G.; Grimme, S. Accurate Modeling of Organic Molecular Crystals by Dispersion-Corrected Density Functional Tight Binding (DFTB). *Journal of Physical Chemistry Letters* **2014**, *5*, 1785-1789.

(28) Otero-de-la-Roza, A.; Johnson, E. R. A benchmark for non-covalent interactions in solids. *Journal of Chemical Physics* **2012**, *137*, 054103.

(29) Reilly, A. M.; Tkatchenko, A. Understanding the role of vibrations, exact exchange, and many-body van der Waals interactions in the cohesive properties of molecular crystals. *The Journal of Chemical Physics* **2013**, *139*, 024705-024705.

(30) Reilly, A. M.; Cooper, R. I.; Adjiman, C. S.; Bhattacharya, S.; Boese, A. D.; Brandenburg, J. G.; Bygrave, P. J.; Bylsma, R.; Campbell, J. E.; Car, R.; Case, D. H.; Chadha, R.; Cole, J. C.; Cosburn, K.; Cuppen, H. M.; Curtis, F.; Day, G. M.; DiStasio Jr, R. A.; Dzyabchenko, A.; van Eijck, B. P.; Elking, D. M.; van den Ende, J. A.; Facelli, J. C.; Ferraro, M. B.; Fusti-Molnar, L.; Gatsiou, C.-A.; Gee, T. S.; de Gelder, R.; Ghiringhelli, L. M.; Goto, H.; Grimme, S.; Guo, R.; Hofmann, D. W. M.; Hoja, J.; Hylton, R. K.; Iuzzolino, L.; Jankiewicz, W.; de Jong, D. T.; Kendrick, J.; de Klerk, N. J. J.; Ko, H.-Y.; Kuleshova, L. N.; Li, X.; Lohani, S.; Leusen, F. J. J.; Lund, A. M.; Lv, J.; Ma, Y.; Marom, N.; Masunov, A. E.; McCabe, P.; McMahon, D. P.; Meekes, H.; Metz, M. P.; Misquitta, A. J.; Mohamed, S.; Monserrat, B.; Needs, R. J.; Neumann, M. A.; Nyman, J.; Obata, S.; Oberhofer, H.; Oganov, A. R.; Orendt, A. M.; Pagola, G. I.; Pantelides, C. C.; Pickard, C. J.; Podeszwa, R.; Price, L. S.; Price, S. L.; Pulido, A.; Read, M. G.; Reuter, K.; Schneider, E.; Schober, C.; Shields, G. P.; Singh, P.; Sugden, I. J.; Szalewicz, K.; Taylor, C. R.; Tkatchenko, A.; Tuckerman, M. E.; Vacarro, F.; Vasileiadis, M.; Vazquez-Mayagoitia, A.; Vogt, L.; Wang, Y.; Watson, R. E.; de Wijs, G. A.; Yang, J.; Zhu, Q.; Groom, C. R. Report on the sixth blind test of organic crystal structure prediction methods. *Acta Crystallographica Section B* **2016**, *72*, 439-459.

2 Theoretical background of Crystal Structure Prediction

2.1 Introduction

This chapter will focus on the theoretical models and procedures which have been used to perform the research included in this thesis. Fundamental to crystal structure prediction (CSP) is evaluating the thermodynamic stability of the crystal structures. This is usually approximated by the lattice energy, but can also include entropic and free energy contributions. There are different methods in which the relative energies in the energy landscape can be determined, with various methods of determining thermal contributions (see Section 1.9) but there are also a number of methods for determining the lattice energy, ranging from pure electronic structure methods to combinations of ab initio and semi-empirical force-fields.[1,2] Each method relies on the accurate representation of the inter and intra molecular forces within the molecular crystal to give the lattice energy.

2.2 Determining the Lattice Energy

The lattice energy of a crystal can be defined as the sum of the inter and intramolecular energies within the crystal structure:[3]

$$U_{latt} = U_{inter} + \Delta E_{intra}$$

Equation 2-1

U_{inter} is the intermolecular energy between the molecules in the lattice which can be thought of being comprised of electrostatic, dispersion and repulsion components that can be modelled by atom-atom intermolecular potentials (see Section 1.4). By modelling the intermolecular energy in this way, we are assuming pairwise interactions with a finite summation over all molecules in the lattice. U_{inter} is negative and the dominant component of the lattice energy.

ΔE_{intra} is the molecular energy conformation penalty. It is the difference in energy between the molecular conformation in the lattice and the optimum gas phase (isolated) molecular conformation. It is positive and generally very small, being zero for rigid molecules,[4] and typically in the order of only a few kJ mol^{-1} unless the molecule is capable of forming intramolecular hydrogen bonds that may become intermolecular in the crystal structure.[5]

U_{latt} is an internal energy, (and its commonly used equivalent, E_{latt}) which defines the energy of a static model of the crystal at 0 K relative to an energy baseline of infinitely separated molecules in their lowest energy conformation. This baseline implies that there is no difference in the molecular composition or covalent bonds in the thermodynamic cycles. It is worth noting that the lattice energy is usually defined in kJ mol^{-1} relative to a mole of molecules, not distinguishing chirality, so that the lattice energy is approximately equal for both racemic and chiral crystals – if we were to distinguish then the lattice energy of the racemic crystal would be approximately twice that of the enantiopure crystal.

2.3 Theory of Intermolecular Forces

Evaluation of intermolecular forces is central to any crystal structure prediction study. There are intermolecular forces between the molecules in any ensemble, with the number and strength of these interactions determining the physical properties of the molecules in the solid, liquid or gas.[3] As such, understanding the interactions between molecules is of great importance to the pharmaceutical industry where the physical properties must be under strict control.

2.4 Modelling Intermolecular Forces

The main challenge in calculating accurate lattice energies lies in the determination of intermolecular forces within the crystal structure, which are very weak compared with the interactions in traditional materials, such as metals and ceramics. Traditionally the intermolecular forces are considered to be very weak compared with the covalent bonding. However, the intermolecular forces can change the conformation of the molecule within the crystalline forms, mainly through changes in low energy torsion angles. Modelling intermolecular forces accurately enough is an ongoing challenge for crystal structure prediction methods, and so contrasting methods of evaluating the lattice energies has been a major focus of the Blind Tests of CSP (See Chapter 4).[2]

The intermolecular forces within a crystal structure can be modelled in two parts: the electrostatic component modelled by using a model of the molecular charge density obtained by an *ab initio* calculation on the molecule and all other terms represented by an empirical exp-6 potential. These models are denoted by Ψ_{mol} as the *ab initio* calculation provides ΔE_{intra} and the electrostatic model only needs to be done for each conformation of the molecule.

2.4.1 Electrostatics

2.4.1.1 Atomic Charges

The charge distribution surrounding a molecule could be represented by atomic charges. This is a relatively crude but simple model where Coulomb's law is used through pairwise interactions of isotropic point charges that are placed on each atom. It is assumed that the charge density of the molecule can be described by superimposing spherical atom electron densities onto the molecules. As the electron density is spherical, atomic charges cannot represent non-spherical features, such as lone pairs and π electrons, or short distance directional intermolecular interactions, such as hydrogen bonds, well. However, at long range the electrostatic potential becomes more accurate as the effect of the non-spherical atomic charge distribution becomes less important as the distances approach infinity. In order to achieve a better description of electrostatic interactions at van der Waals contact distances using just atomic charges, additional sites to represent π-electrons and lone pairs may be added.[6]

2.4.1.2 Distributed Multipoles

An alternative way in which the charge distribution can be represented is by an expansion of multipole moments about each nucleus. Distributed multipole moments have successfully predicted directional lone pair interactions, π-π stacking arrangements in aromatic rings and electron and hydrogen-bonded geometries in organic crystals. This method accounts for the anisotropy of the charge distribution and is essential for CSP.[7]

Distributed multipole analysis (DMA)[8] is a systematic way of determining distributed multipole moments from an *ab initio* wavefunction and describes the molecular charge density in terms of the distributed multipoles located at a number of sites in the molecule, usually all the atomic sites. DMA and point charges can be derived explicitly through ab initio calculations of the charge density of the molecule held in isolation. The program GDMA[9] obtains the DMA from a charge density described by a Gaussian basis set. This can be used as an input for organic crystal structure modelling using DMACRYS.[10] For modelling racemic crystals, the atomic multipoles on molecules generated by inversion are automatically adapted by DMACRYS so that the right handed axis

system is used throughout for evaluating the electrostatic interactions between the higher multipole moments.

2.4.2 Repulsion-dispersion

Within the Ψ_{mol} model, all other contributions to the intermolecular forces are usually modelled with an empirical Buckingham exp-6 atom-atom potential, which is assumed to be transferable. The repulsion is represented by an exponential based term and dispersion by an inverse sixth power term;

$$U_{rep-disp}^{MN} = \sum_{i\in M, k\in N} A_{ik}\exp(-B_{ik}R_{ik}) - \frac{C_{ik}}{R_{ik}^6}$$

Equation 2-2

where the interactions between atom *i* of type *ı* in molecule M and atom *k* of type *κ* in molecule N that are separated by a distance $R_{ik.}$ The potential is derived by fitting the parameters A, B and C to the heats of sublimation and structures of known organic crystals. Combining rules are often used to reduce the number of parameters that need to be determined by fitting e.g.

$$A_{ik} = (A_{ii}A_{kk})^{\frac{1}{2}}, B_{ik} = \frac{1}{2}(B_{ii} + B_{kk}), C_{ik} = (C_{ii}C_{kk})^{\frac{1}{2}}$$

Equation 2-3

The FIT intermolecular potential parameters[10] were fitted using a model that uses explicit electrostatic interactions. Here the parameters for carbon, non-polar hydrogen and nitrogen have been derived from fitting to azahydrocarbon crystal structures, while those for oxygen were derived from fitting to oxyhydrocarbon crystal structures.[11] The parameters for polar hydrogens were derived from simple molecules containing N-H hydrogen bonds.[12] There is a need to account for the difference between polar and non-polar hydrogens due to the smaller effective van der Waals radius when there is less charge density associated with the protons.

Other sets of potential parameters, such as those later derived by Williams using off-nuclear proton positions[13] or explicit O-H polar protons[14] have been derived. These were tested for the compounds studied in this thesis but gave less satisfactory results than the FIT potential.

The use of potentials which have been derived from fitting to crystal structures presents a number of problems for modelling crystal structures. The contributions to the intermolecular forces that are not explicitly represented (i.e. by Equation 2.2 and the accompanying electrostatic model),[3] such as the polarisation,[10,15] are absorbed into the fitting. The thermal effects are also absorbed by fitting to crystal structures which are usually measured at room temperature or more recently, at lower temperatures. Heats of sublimation are notoriously difficult to measure experimentally. These have often been equated to the lattice energy or assumed to differ by the –2RT correction (See Chapter 6). Thus, these potentials are not thermodynamically consistent with lattice energies corresponding to a static crystal at 0 K. The fitting process also means that only atoms in environments within the crystals used for the fitting are modelled well, for example the fluorine potential which is often used as part of the FIT parameter set[10] has been derived from fitting to just seven crystal structures and only one heat of sublimation.[16] While there are uncertainties in the level of accuracy of the potential it is hoped that any errors can be absorbed through careful fitting of the parameters used to determine U_{inter}. Indeed, this scheme of fitting potentials has recently been revisited[17] and appears to be as accurate for small organic molecule crystal structures as the periodic DFT-D Ψ_{crys} models in current use (see Section 1.7).[18]

2.5 Modelling Intramolecular Forces

The intramolecular energy penalty, ΔE_{intra}, can be considered to be the difference in energy between a molecule observed within the crystal structure and the molecule in its gas phase optimal conformation. For a rigid molecule this energy penalty will be negligible and so the lattice energy can be considered to be approximately the intermolecular energy of the crystal. However, for flexible molecules where the molecular conformation can drastically change the energy penalty can have a significant contribution to the lattice energy, potentially resulting in a number of different polymorphs.[5]

The intramolecular energy penalty can be determined through an electronic *ab initio* calculation of the molecule in the optimal gas phase configuration alongside that of the distorted molecule observed in the crystal. (Care must be taken to avoid the effect of experimental errors, particularly the underestimate of

bondlengths to hydrogen atoms in X-ray structure determination). The calculation of the energy of the isolated molecule is highly sensitive to the quality of the wavefunction used[19] and so sufficient forethought is required when deciding which wavefunction to use. The calculation of ΔE_{intra} and the corresponding DMA has to be repeated many times for flexible molecules (See Section 1.6.2) meaning that a compromise is required between accuracy and computational cost for flexible molecules.

2.6 CSP Methods

To create a global search, firstly a number of hypothetical crystal structures needs to be generated, and then their energies need to be accurately determined. These structures must be close packed, dense and energetically feasible. The algorithm used to produce the hypothetical crystal structures needs to be complete enough such that a thorough cross section of the sample space is achieved; if the experimental structure is missed in the initial set of structures produced then the accuracy of any subsequent steps is irrelevant. The final energy minimisation must be accurate enough to reliably rank the hypothetical crystal structures produced. The cost of this scales with the number of structures generated and so hierarchical methods are used.

2.6.1 Generating hypothetical crystal structures - *CrystalPredictor*

CrystalPredictor is one algorithm which can be used to generate hypothetical crystal structures. In order to produce a set of structures, the user needs to define a search space which includes information on the number of molecules in the asymmetric unit as well as the space groups which need to be considered. *CrystalPredictor* uses Sobol sequencing[20] to ensure that the search is complete by seeing how often each lattice energy minimum is found. By using Sobol sequencing a more uniform coverage of the sample space is obtained over pseudorandom number generation as well as providing a means to extend the search space without having to repeat the entire process.

CrystalPredictor uses atomic charges to account for the electrostatic forces within the crystal in order to obtain a rough estimate of the lattice energy. These lattice energies are later refined with distributed multipoles (see Section 1.6.2).

The way molecular flexibility is accounted for varies between the two different versions of *CrystalPredictor* which have been used in my research.

2.6.1.1 CrystalPredictor 1.6

In *CrystalPredictor* 1.6[21,22] molecular flexibility is accounted for by dividing the molecule into several independent fragments which are each assumed to be rigid. This allows a grid of ΔE_{intra} to be calculated describing how the flexible torsion angles are linked to the independent rigid fragments. The intramolecular energy is varied with a spline function based on the following parameters; the energy values and the energy derivatives calculated on the grid of flexible torsion angles. The atomic charges, as determined by the user defined potential, are either held constant or, in cases where there is a large variation on molecular geometry, allowed to vary and are also represented by a grid.

2.6.1.2 Crystal Predictor 2.0

In *CrystalPredictor* 2.0[23] the molecular geometry, atomic charges and intramolecular energies are given by a Local Approximate Model (LAM). This method allows the molecular flexibility to be handled in a much more accurate manner. The molecular geometry is updated to reflect the effect of changing the torsions on the geometry of rest of the molecule, the atomic charges are adapted to the changing molecular geometry without the need for potentially inaccurate splines and the intramolecular energy is handled using a model that is more accurate than the splines used in *CrystalPredictor* 1.6.

2.6.2 CrystalOptimizer

Once a set of crystal structures has been generated, the structures and their corresponding lattice energies need to be refined to account for any possible changes in molecular conformation due to the intermolecular forces. This can be done with the two step optimisation algorithm in *CrystalOptimizer*.[24] Firstly, the intramolecular energies and DMAs are evaluated, considering the flexible degrees of freedom as variables. Secondly the intermolecular energy and lattice variables are evaluated and refined using DMACRYS, keeping the conformation rigid. These two steps are repeated until a mechanically stable structure is determined via the lattice energy derived, along with the optimised cell

parameters of the crystal structure. The *ab initio* calculations used to determine the intramolecular energies, forces and DMAs are stored in a database and reused as appropriate in subsequent optimisations of crystal structures of the same molecules. This makes this method of optimisation feasible for CSP studies where thousands of crystal structures are generated.

2.6.3 Polarizable Continuum Model

The use of Equation 2.1 neglects the molecular polarisation of the molecules as, when the electrostatic model is derived from the isolated charge density, the rearrangement of electron density due to the crystalline environment is ignored. This causes the total energy of the crystal to be lowered. It is hoped that any effects due to polarisation have been absorbed with the empirically fitted potentials. One method to check the sensitivity of the molecule to any polarisation is to perform the molecular charge density calculation so that the multipole analysis is performed on an electron density that provides a better representation of the molecule in the crystal. This can be achieved by modelling the environment of the molecule in the crystal as a polarisable continuum.[25] Typically the dielectric constant $\varepsilon = 3$ is used for organic crystals and is not specific to the crystal structure under consideration.

2.6.4 Estimation of Helmholtz Free Energy

All of the methods for determining lattice energies mentioned previously, rely on the definition that the crystal structure being studied is a rigid body held at 0 K. In reality, crystal structures are not experimentally determined for substances held at 0 K and therefore the effects of temperature can potentially play a significant role in the relative stability of crystal structures. The vibrational properties of crystals are responsible for much of their thermal behaviour (expansion, polymorphic phase transitions and melting). These can be modelled through the calculation of the free energy. Within DMACRYS the vibrational contribution can be determined through calculation of the Helmholz free energy. The Helmholtz free energy is related to the Gibbs free energy by a PV term. If we assume there is no thermal expansion of the crystal structure then we can approximate the Gibbs and Helmholtz free energies to be the same. Although it has been assumed that the relative energy contribution from

the vibrational terms is small, it can become significant when considering the small energy differences observed between lattice energies of hypothetical crystal structures (as demonstrated in Chapters 3 and 5) and so it is important to assess the free energy contribution.

Within the Ψ_{mol} model, DMACRYS can be used to determine the intermolecular frequencies at the Gamma point (k = 0).[26] These frequencies are calculated as these are the modes which can be most easily compared with the low frequency Raman, infrared or terahertz frequencies of experimental crystal structures and so the errors can easily be estimated. By treating the crystals as rigid bodies the intramolecular modes can be ignored, an approximation which is assumed to be valid as the intramolecular modes are not expected to vary significantly between polymorphs of the same crystal structure. This expectation will be most realistic for crystals of small rigid molecules, such as carbon dioxide or benzene.

2.6.5 Advantages and disadvantages of the Ψ_{mol} model

The Ψ_{mol} model has been shown to provide a good description of the crystal energy landscapes of various crystal structures. This method provides a comparatively quick way of determining the lattice energy of organic crystal structures, which can be comparable to some popular DFT-D methods,[18] allowing a complete description of the CEL of a compound to be obtained in 1-12 months, depending on the computer infrastructure available and the size of the compound under investigation. Additionally, the lattice summations are independent of cell size which allows direct comparison of the energies of crystals structures of varying sizes, shapes and number of independent molecules.

However, there are some limitations with the Ψ_{mol} model. While the model is capable of calculating the lattice energy of flexible molecules, the flexibility is confined to only the most flexible torsion angles within a user defined energy range and so does not allow the full range of flexibility to be explored in the way that DFT-D models can. The Ψ_{mol} model on its own does not account for effects such as polarisation and repulsion anisotropy. These effects are partially accounted for by using empirically derived potentials by being absorbed into the parameter fitting process. As increasingly diverse systems are being studied

these empirically derived potentials are being refined,[17] highlighting the possible inaccuracies caused by double counting due to the measurements used for the fitting being at a different temperatures to the model in which they are being applied to. The Ψ_{mol} method for evaluating lattice energies has been shown to be very accurate but it is still far from the accuracy of state of the art calculations such as that used for benzene, which has been evaluated to within 1 kJ mol^{-1} of the experimentally determined value.[27] In the search for increasing accuracy, care still needs to be taken to ensure that the results are not obtained fortuitously rather than by thorough scientific method and testing. While there are some clear limitations to the Ψ_{mol} model it has still been shown to be a worthwhile model for obtaining the relative lattice energies of organic crystal structures.

2.7 Periodic Electronic Structure Methods

There are a number of alternative ways in which the lattice energy of a crystal can be calculated that have been emerging recently and have played a significant role in the recent Blind Tests.[2,28] One alternative method is periodic Density Functional Theory (DFT), which is an electronic structure method based on using the wavefunction of the entire crystal. These models will be denoted by Ψ_{crys}. The advantages of these methods are that they do not require the separation of the crystal into covalently bound molecules and so treat inter and intramolecular effects, such as polarisation in the same, non-empirical way. However, these methods are only just cheap enough to be used for large scale CSP for the final energy evaluation step.

2.7.1 Density Functional Theory

DFT is a widely-used application to describe atomic and molecular scale interactions using the wavefunction of the crystal. DFT is based on the theory that if the electronic density of the system in the ground state is known, then this is enough to determine the energy of the system through an appropriate functional.

DFT functionals that can be applied to the organic solid state do not typically encompass the long-range dispersive interactions which are of particular

importance for organic crystal structures. Since the intermolecular dispersion, whose leading term is C_6/R^6,[3] arises from electron correlation, it is only included in the more expensive types of electronic structure theories. Developments by Neumann and Perrin[29] have shown that it is possible to incorporate an empirical van der Waals correction to produce a method of evaluating the lattice energies of crystal structures at the electronic structure level. Further methods have been developed which allow the dispersion corrections to be derived without fitting to organic crystal structures, such as the DFT-D methods by Grimme (G06)[30] and Tkatchenko-Sheffler (TS).[31] This means that these types of calculation can be used more widely for organic crystal structures.

In this study a plane wave DFT-D approach is adopted using the CASTEP[32] code with the PBE functional, which is the most accurate that can be easily afforded for calculations on many organic crystal structures. The Ψ_{mol} approach can use better functionals (e.g. PBE0) or electronic structure methods, and so can provide a better description of the molecular charge distribution and hence the electrostatic interactions. The quality of either Ψ_{mol} or Ψ_{cry} calculations also depends on the basis set, which is comprises atomic orbitals for Ψ_{mol}, and plane waves for Ψ_{cry} within CASTEP.

2.7.2 Real space vs reciprocal space

We are trying to model an infinite lattice system and so DFT-D calculations need to take into account the cells adjacent to the unit cell that is being modelled and the convergence of the plane wave basis set. This can be done by sampling reciprocal space and interpolating between the k-points. In order to model the crystal structure well enough a sufficient sampling of the k-space is required to ensure that the relative energies of crystals of different shapes and sizes, containing different numbers of molecules, are independent of the k-point sampling (see Chapter 3 and 6). The k-point sampling needed for evaluating the lattice energy in Ψ_{cry} DFT-D calculations is distinct from the k-point sampling needed for determining the phonon density of states used to obtain thermodynamic properties.

2.8 Comparison of Structures

Identifying the similarities between two crystal structures is important when deciding if a new polymorph of a compound crystal structure has been discovered. This identification is particularly relevant for CSP studies where thousands of hypothetical crystal structures are generated and any crystal structures which converge to the same minimum need to be removed. However, the definition of polymorphism is complicated. In 1965 McCrone[33] defined polymorphism as 'a solid crystalline phase of a given compound resulting from the possibility of at least two crystalline arrangements of the molecules of that compound in the solid state'.[33] The key word in this definition is 'phase' as structural variation can vary without a change in phase, although a phase cannot change without some structural change.[34] Desiraju put forth the notion that for each individual molecule there is a structural landscape with different polymorphs, solvates, amorphous forms as minima on this highly multi-dimensional surface,[35] so how much difference does there need to be between two structures for them to be considered as two separate polymorphs? How easy is it to identify these potentially very subtle differences?

It is possible for crystal structures to be compared visually, and much of the analysis on packings throughout this project has been performed this way. There are benefits of this, including knowledge and experience of similar systems, however this method is subject to bias and is slow, indeed most of the hours put into analysing crystal energy landscapes are through visual inspection. Thus, researchers need practical computational tools which help them to identify identical or similar crystal structures.

2.8.1 Mercury

The Mercury[36] materials software is the primary tool that was used to view and compare crystal structures throughout this project. Mercury allows graphical representation of crystal structures and the similarities between them by visually overlaying the two structures. Additionally, the degree of similarity can be determined quantitatively by using the crystal packing similarity tool. This tool compares clusters of *n* molecules, calculating the level of packing similarities between two crystal structures, giving a root mean square deviation ($RMSD_n$) of

the distance between matched atoms in a cluster as an indication of the degree of similarity between the structures. The default size for a cluster is the first coordination "sphere" of *n* = 15 molecules, allowing for a distance tolerance of 20% and angle tolerance of 20%. Generally, an $RMSD_{15}$ of below 0.5˚ is considered to be an acceptable match.[2] If less than *n* molecules match, this indicates that the crystal structures have a motif such as a dimer (*n* = 2) or a layer (e.g. *n* = 6) in common.

2.8.2 XPac

XPac[37] is a program which allows the comparison of two structures to determine the degree of isostructurality (0D, 1D, 2D or 3D). XPac compares sets of vectors between selected atoms in a 'seed' molecule within each crystal structure and equivalent atoms of neighbouring molecules. How well these vectors match, as determined by the user, defines the degree of similarity between the molecules so that if all vectors match then the two structures are considered to be 3D isostructural. How well these vectors match is determined by the changes in angles (Δa), planes (Δp) and distances (Δd) with the acceptable values for tight tolerances on the structures being compared are Δa = 5 °, Δp = 7 ° and Δd = 0.5 Å. The set of vectors defining the isostructurality corresponds to the way the molecules are held in the corresponding 1D, 2D and 3D arrays. The arrangement of these molecules is called a supramolecular construct (SC). It is through analysis of these SCs from which a structural relationship diagram between crystal structures can be built.[38]

This type of analysis can help to identify plausible crystal structures of related molecules which have not be observed experimentally.[39] For example, if a 3D relationship is observed between crystals containing two similar molecules, then it is plausible that any other molecules which are structurally similar could also form the same crystal structure. Observations such as these can help design crystallisation experiments to produce these theoretically plausible crystal structures, for example via seeding or sublimation.[40,41]

2.8.3 Graph Sets

The packing and structure of crystals can be compared through the analysis of hydrogen bonding motifs. Graph set analysis, based on graph theory, allows hydrogen bonding motifs to be categorised and allowing them to be compared and contrasted for sets of molecules. A graph set is defined by the formula $G_d^a(n)$.[42] G indicates the type of hydrogen bonded chain, denoted by C for a chain, R for a ring, S for intramolecular hydrogen bonded patterns or D for other finite patterns. The a and d superscripts indicate the number of hydrogen bond acceptors (a) and hydrogen bond donors (d). The number of atoms in the pattern, called the degree of the pattern, is denoted by n.

2.9 Thermodynamic Properties

Accurate lattice energy calculations are a fundamental component to determine solubility differences of chiral organic crystals (see Chapter 1) as it is the main contribution to the free energy of solution. However, there are additional thermodynamic effects which, if included, may help to improve the accuracy of solubility difference calculations.

In order to determine the free energy of solution, ΔG_{sol}, via the sublimation cycle, it is necessary to understand the thermodynamic relationships contained within each stage of the cycle and any assumptions which have been made. The thermodynamic cycle, seen in Figure 2.1, has been used to predict absolute solubility.[43-45]

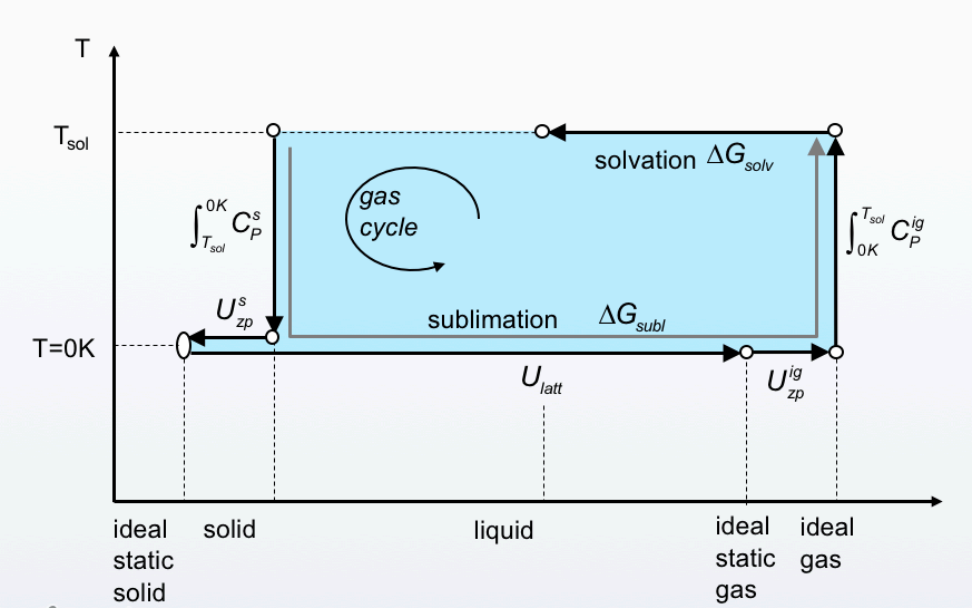


Figure 2-1: The sublimation cycle used to describe the free energy of solution, ΔG_{sol}, which uses the free energies of sublimation, ΔG_{subl}, and the solvation free energy, ΔG_{solv}.

The main contribution to the sublimation cycle is the lattice energy contribution to the free energy of sublimation, U_{latt}. However, there are other solid state thermodynamic quantities which need to be considered, such as the heat capacities and zero point energies.

Solubility is defined by the mole fraction, *x*, and is directly related to the free energy of solution. Thus, the solubility is calculated using the sublimation and solvation free energies from the following equations:

$$-RT\ln x = \Delta G_{sol} = \Delta G_{subl} + \Delta G_{solv}$$
$$= \Delta H_{subl} - T\Delta S_{subl} + \Delta G_{solv}$$

Equation 2-4

The solvation free energy, ΔG_{solv}, is identical for both the racemic and enantiopure crystals in infinite dilution. As we are considering the energy differences between enantiopure and racemic crystals this term can cancel out.

2.9.1 Estimation of thermal contributions to sublimation enthalpies

The Gibbs free energy of sublimation, ΔG_{subl}, can be calculated via the Gibbs-Helmoltz relation $\Delta G = \Delta H - T\Delta S$ from the enthalpy of sublimation, ΔH_{subl}. Assuming that there is no polymorphic change, ΔH_{subl} is related to the lattice energy, U_{latt}, via the following equation

$$\Delta H_{subl}(T) = -U_{latt} + \left(E^{g}_{ZPE} - E^{s}_{ZPE}\right) + \int_0^T \left(C_p^g(T) - C_p^s(T)\right) dT$$

Equation 2-5

where E_{ZPE} is the zero point vibrational energy and $C_p(T)$ is the heat capacity between zero Kelvin and a temperature T of the solid, s, and gas, g.

As mentioned in Chapter 1, there is a very approximate link between the lattice energy of a crystal and the enthalpy of sublimation, H_{sub}. This can be defined in two ways, either

$$\Delta H_{sub} \approx -U_{latt} = -E_{latt}$$

Equation 2-6

or

$$\Delta H_{sub} \approx -U_{latt} - 2RT$$

Equation 2-7

This equation is based on the following assumptions; $C_v^s = 6R$ for all crystals from 0 K to the sublimation temperature, the zero-point vibrational energies are the same in the gas and solid states, and that the gas is ideal.[46,47] The addition of the 2RT term has been widely used in calibrating the repulsion-dispersion potentials used in crystal structure prediction. It can also be ignored based on the fact 2RT ≈ 5 kJ mol^{-1} at room temperature is comparable to the experimental error in early heats of sublimation measurements.

It is important to critically assess the assumptions that have been made in these equations and how these might affect the accuracy of the final result. For the gas phase we continue to assume that C_p^g can be decoupled into translational, rotational and vibrational contributions. As we are considering non-linear molecules, the contribution from gas-phase translations and rotations can be approximated to be 4R. However, for larger very flexible molecules, molecular vibrations could be of sufficiently large amplitude to change the moments of inertia such that the separation into independent rotational and vibrational contributions would not be justified.[48] It is possible to obtain the gas phase molecular vibration contribution from *ab initio* calculation of the lowest energy conformation of the molecules in question. The solid phase is slightly more complex: if the thermal expansion of the unit cell is neglected we can approximate C_P and C_V. The isobaric and isochoric solid state heat capacities are related by the bulk modulus, K, and the thermal volume expansion coefficient, α_T:

$$C_p^s(T) = C_v^s(T) + TVK\alpha_T^2$$

Equation 2-8

In Chapter 6 experimental C_P values are being compared with computed C_V values and so only the lattice vibrations contribute to the heat capacity of the solid. Thus ΔH_{subl} can be considered to be combination of the lattice energy and a thermal correction, H_{corr}, comprising of the ZPE and heat capacity terms.

$$\Delta H_{subl}(T) = -U_{latt} + \Delta E_{vib}(T) + 4RT = -U_{latt} + H_{corr}$$

Equation 2-9

2.9.2 Modelling the lattice vibrations

It has been shown that the lattice energy differences may be accurately modelled using the Ψ_{mol} method described in Section 1.2, by separately modelling the isolated molecule and the molecules contained within the crystal lattice as being rigid. The crystal modes can then be evaluated in the rigid-body approximation using an atom-atom model intermolecular potential.[26] This separation is traditionally justified on the basis that the molecular vibrations are of a much higher frequency than the lattice vibrations. This is unlikely to be valid for low frequency intramolecular modes, such as low barrier C-C torsions, or out of plane modes, like R-NH_2 pyrimidalisation.

Using the rigid-body lattice modes for the heat capacities requires a hybrid Einstein-Debye model for the calculation of the intermolecular ZPEs and the thermal free energy of the crystal. For Z rigid molecules in the unit cell, the 6Z-3 optic frequencies are modelled with an Einstein model, with each mode being assigned its frequency at k=0 ($\omega_E^i = \omega^i(k=0)$). A Debye model is used to account for the acoustic modes by using a Debye frequency cut off (ω_D), estimated from the elastic stiffness matrix.[49] Taking the assumption that the thermal expansion of the crystal is negligible, the heat capacity can be expressed as:

$$C_V^s = k\sum_i \left(\frac{\omega_E^i}{kT}\right)^2 \exp\left(\frac{\omega_E^i}{kT}\right)\left[\exp\left(\frac{\omega_E^i}{kT}\right)-1\right]^{-2} + \frac{36k}{\left(\frac{\omega_D}{kT}\right)^3}\int_0^{\omega_D/kT}\frac{t^3}{\exp(t)-1}dt - \frac{9\omega_D}{T}\left[\exp\left(\frac{\omega_D}{kT}\right)-1\right]^{-1}$$

Equation 2-10

The division into optic and acoustic modes depends on how the unit cell is divided. The effects of free energies are only modelled around k = 0, instead of sampling across the full Brillouin zone which would be required for complete accuracy. The *k* = 0 cell frequencies correspond to measureable IR/Raman techniques and so can still be used to compare with experiment. Methods have been developed to help tackle this issue throughout the course of this project[50] however, sampling around *k* = 0 is still much easier. By using a large supercell the accuracy could be increased.[51] However, it can be difficult to select

equivalent unit cells when comparing unit cells of differing sizes, shapes, and containing different numbers of molecules.

Alternatively, the crystal modes can be evaluated using electronic structure calculations where the large unit cell sizes and number of molecules can potentially pose a problem, in addition to the accuracy of modelling the intermolecular forces. The equation for the heat capacities for periodic electronic structure phonon calculations uses the phonon density of states, F(ω), and performs a numerical integration over the Brillouin zone at every temperature:

$$C_V(T) = k \int \left(\frac{\omega}{kT}\right)^2 \exp\left(\frac{\omega}{kT}\right) \left[\exp\left(\frac{\omega}{kT}\right) - 1\right]^{-2} F(\omega)\ d\omega$$

Equation 2-11

whereas the ZPE is simply defined as:

$$E_{ZPE} = \frac{1}{2} \int \omega F(\omega)\, d\omega$$

Equation 2-12

The ZPE term is dominated by the high-frequency molecular modes. If these are estimated from the vibrational frequencies of the isolated molecule, ω_{mol}^{i}, as in the Ψ_{crys} method, the ZPE contribution to the solid is the same as in the gas, i.e.

$$E_{ZPE}^{g} = \frac{1}{2} \sum_{i}^{3N-6} \omega_{mol}^{i}$$

Equation 2-13

for a molecule of N atoms.

The two alternate methods of determining the lattice energies and phonon frequencies, Ψ_{mol} and Ψ_{crys}, will be compared alongside experimentally determined values to explore their accuracy in determining solubility differences of various organic crystal structures. There is still a large variability in the ranking of lattice energies when comparing various methods and so by extending this comparison to solubility difference calculations it will be possible to extend on the strengths and limitations of the Ψ_{mol} and Ψ_{crys} methods in the context of the other thermodynamic terms that are involved in free energy and solubility differences.

2.9.3 Modelling the Eutectic composition

At the eutectic point three phases are in equilibrium with each other: the pure solid enantiomer in excess, the solid racemic compound and the liquid phase with the eutectic composition (see Figure 1.3). Using the reference state of 1 mol of racemic compound contains ½ mol of each enantiomer, the phase equilibria of the pure enantiomer and pure racemic compound can be defined as:

$$S_{crystal} \leftarrow \Delta G_{solv,S} \rightarrow S_{solution}$$
$$RS_{crystal} \leftarrow \Delta G_{solv,RS} \rightarrow \tfrac{1}{2}S_{solution} + \tfrac{1}{2}R_{solution}$$

Equation 2-14

Klussmann *et. al.*[52] proposed a method to estimate the eutectic composition, x_{eu}, as a function of the solubility ratio, $\alpha = x_{RS} / x_S$, between the racemic compound and the pure enantiomer. From Equation 2.4 the solubility ratio and therefore the eutectic composition are related to the difference in the Gibbs free energy of solution between the racemic compound and the pure enantiomer, $\Delta_{RS\text{-}S}\Delta G_{solv} = \Delta G_{solv,RS} - \Delta G_{solv,S}$. If the assumption is made that the solutions are infinitely dilute, then the solubility of the enantiomers will not be affected by one another and $\Delta_{RS\text{-}S}\Delta G_{sol} = 0$. This leads to the following equation for the eutectic composition:

$$x_{eu} = \frac{1}{1 + \frac{\alpha^2}{4}} = \frac{1}{1 + \frac{1}{4}\exp\left(-2\frac{\Delta_{RS-S}\Delta G_{sol}}{RT}\right)}$$

Equation 2-15

The two thermodynamic cycles can be used to relate $\Delta_{RS\text{-}S}\Delta G_{sol}$ and hence the eutectic composition in Equation 2.15, to either the difference in the free energies of melting, $\Delta_{RS\text{-}S}\Delta G_{melt}$, or in the difference in the free energies of sublimation, $\Delta_{RS\text{-}S}\Delta G_{subl}$. The main focus of this thesis is on the sublimation cycle, and so the following relation between the solubility ratio, α, and the sublimation thermodynamics can be made:

$$\alpha = \frac{x_{RS}}{x_s} \approx \exp\left(-\frac{\Delta_{RS-S}\Delta G_{subl}}{RT}\right) \approx \exp\left(-\frac{\Delta_{RS-S}\Delta H_{subl}}{RT}\right) \approx \exp\left(\frac{\Delta_{RS-S}U_{latt}}{RT}\right)$$

Equation 2-16

There are several degrees of approximations which can be seen in Equation 2.16, with each step reducing the amount of computational effort involved. The first approximation is that the liquid phases are ideal and that the solvent is achiral. This is constituent with the assumptions used for deriving the thermodynamic model in Equation 2.15. The second approximation assumes that the entropy contributions are the same for both the enantiopure and racemic compounds in the solid and gas phase. The final approximation state that the thermal correction, H_{corr}, (see Chapter 1.6 and Chapter 6) is equal for the enantiopure and racemic compounds.

There is an alternative derivation of the eutectic composition,[53] which uses a reference state of a mole of molecules that are independent of chirality, defined by the equation:

$$x_{eu} = \frac{1}{1+\alpha^2} = \frac{1}{1+\exp\left(-2\frac{\Delta_{RS-S}\Delta G_{sol}}{RT}\right)}$$

Equation 2-17

However, it has been shown that the thermodynamic model derived in Equation 2.17 gives a poorer agreement with experiment than that derived in Equation 2.15, as shown in Figure 2.2.

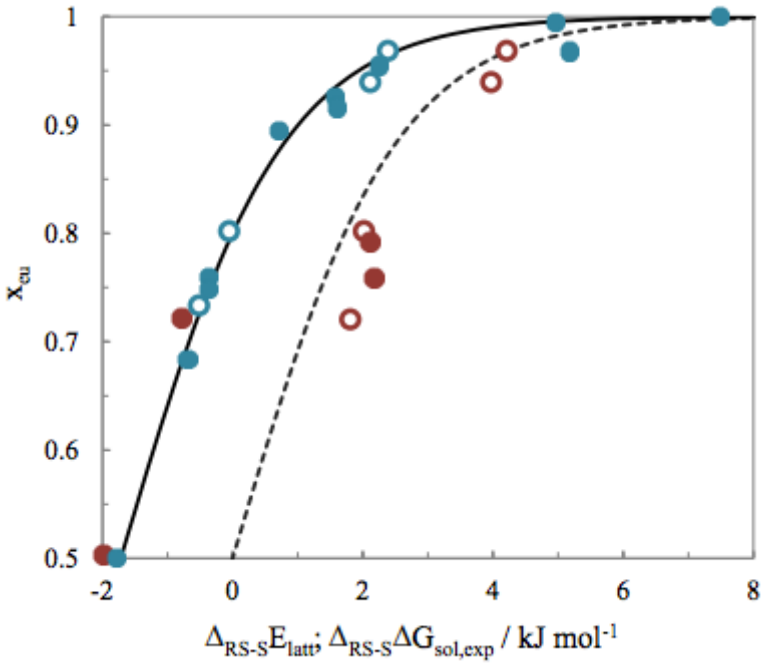


Figure 2-2: The eutectic composition as a function of the free energy differences, calculated using the model for x_{eu} from Equation 2.15 (thick line) or the alternative reference state used by Otero-de-la-Roza et. al. in Equation 2.17 (dashed line). Experimental values of x_{eu} are compared with free energy differences between the racemic and enantiopure compounds, $\Delta_{RS-S}\Delta G_{sol}$, which are derived from experimental

solubilities[52] (blue circles) with lattice energy differences[53] (red circles). Open symbols represent the systems evaluated in both studies.

2.10 References

(1) Kazantsev, A. V.; Karamertzanis, P. G.; Adjiman, C. S.; Pantelides, C. C.; Price, S. L.; Galek, P. T. A.; Day, G. M.; Cruz-Cabeza, A. J. Successful prediction of a model pharmaceutical in the fifth blind test of crystal structure prediction. *International Journal of Pharmaceutics* **2011**, *418*, 168-178.

(2) Reilly, A. M.; Cooper, R. I.; Adjiman, C. S.; Bhattacharya, S.; Boese, A. D.; Brandenburg, J. G.; Bygrave, P. J.; Bylsma, R.; Campbell, J. E.; Car, R.; Case, D. H.; Chadha, R.; Cole, J. C.; Cosburn, K.; Cuppen, H. M.; Curtis, F.; Day, G. M.; DiStasio Jr, R. A.; Dzyabchenko, A.; van Eijck, B. P.; Elking, D. M.; van den Ende, J. A.; Facelli, J. C.; Ferraro, M. B.; Fusti-Molnar, L.; Gatsiou, C.-A.; Gee, T. S.; de Gelder, R.; Ghiringhelli, L. M.; Goto, H.; Grimme, S.; Guo, R.; Hofmann, D. W. M.; Hoja, J.; Hylton, R. K.; Iuzzolino, L.; Jankiewicz, W.; de Jong, D. T.; Kendrick, J.; de Klerk, N. J. J.; Ko, H.-Y.; Kuleshova, L. N.; Li, X.; Lohani, S.; Leusen, F. J. J.; Lund, A. M.; Lv, J.; Ma, Y.; Marom, N.; Masunov, A. E.; McCabe, P.; McMahon, D. P.; Meekes, H.; Metz, M. P.; Misquitta, A. J.; Mohamed, S.; Monserrat, B.; Needs, R. J.; Neumann, M. A.; Nyman, J.; Obata, S.; Oberhofer, H.; Oganov, A. R.; Orendt, A. M.; Pagola, G. I.; Pantelides, C. C.; Pickard, C. J.; Podeszwa, R.; Price, L. S.; Price, S. L.; Pulido, A.; Read, M. G.; Reuter, K.; Schneider, E.; Schober, C.; Shields, G. P.; Singh, P.; Sugden, I. J.; Szalewicz, K.; Taylor, C. R.; Tkatchenko, A.; Tuckerman, M. E.; Vacarro, F.; Vasileiadis, M.; Vazquez-Mayagoitia, A.; Vogt, L.; Wang, Y.; Watson, R. E.; de Wijs, G. A.; Yang, J.; Zhu, Q.; Groom, C. R. Report on the sixth blind test of organic crystal structure prediction methods. *Acta Crystallographica Section B* **2016**, *72*, 439-459.

(3) Stone, A. J.: *The Theory of Intermolecular Forces*; Oxford University Press: Oxford, 2013; Vol. 2.

(4) Gavezzotti, A. Ten years of experience in polymorph prediction: what next? *CrystEngComm* **2002**, *4*, 343-347.

(5) Cruz-Cabeza, A. J.; Bernstein, J. Conformational Polymorphism. *Chemical Reviews* **2014**, *114*, 2170-2191.

(6) Williams, D. E.; Abraha, A. Site Charge Models for Molecular Electrostatic Potentials of Cycloalkanes and Tetrahedrane. *Journal of Computational Chemistry* **1999**, *20*, 579-585.

(7) Day, G. M.; Motherwell, W. D. S.; Jones, W. Beyond the isotropic atom model in crystal structure prediction of rigid molecules: Atomic multipoles versus point charges. *Crystal Growth & Design* **2005**, *5*, 1023-1033.

(8) Stone, A. J. Distributed multipole analysis: Stability for large basis sets. *Journal of Chemical Theory and Computation* **2005**, *1*, 1128-1132.

(9) Stone, A. J.: GDMA: A Program for Performing Distributed Multipole Analysis of Wave Functions Calculated Using the Gaussian Program System. 2.2 ed.; University of Cambridge: Cambridge, United Kingdom, 2010.

(10) Price, S. L.; Leslie, M.; Welch, G. W. A.; Habgood, M.; Price, L. S.; Karamertzanis, P. G.; Day, G. M. Modelling Organic Crystal Structures using Distributed Multipole and Polarizability-Based Model Intermolecular Potentials. *Physical Chemistry Chemical Physics* **2010**, *12*, 8478-8490.

(11) Cox, S. R.; Hsu, L. Y.; Williams, D. E. Nonbonded Potential Function Models for Crystalline Oxohydrocarbons. *Acta Crystallographica Section A - Crystal Physics, Diffraction, Theoretical and General Crystallography* **1981**, *37*, 293-301.

(12) Coombes, D. S.; Price, S. L.; Willock, D. J.; Leslie, M. Role of Electrostatic Interactions in Determining the Crystal Structures of Polar Organic Molecules. A Distributed Multipole Study. *Journal of Physical Chemistry* **1996**, *100*, 7352-7360.

(13) Williams, D. E. Improved intermolecular force field for molecules containing H, C, N, and O atoms, with application to nucleoside and peptide crystals. *Journal of Computational Chemistry* **2001**, *22*, 1154-1166.

(14) Beyer, T.; Price, S. L. Dimer or catemer? Low-energy crystal packings for small carboxylic acids. *Journal of Physical Chemistry B* **2000**, *104*, 2647-2655.

(15) Welch, G. W. A.; Karamertzanis, P. G.; Misquitta, A. J.; Stone, A. J.; Price, S. L. Is the induction energy important for modeling organic crystals? *Journal of Chemical Theory and Computation* **2008**, *4*, 522-532.

(16) Williams, D. E.; Houpt, D. J. Fluorine Nonbonded Potential Parameters Derived From Crystalline Perfluorocarbons. *Acta Crystallographica Section B - Structural Science* **1986**, *42*, 286-295.

(17) Pyzer-Knapp, E. O.; Thompson, H. P. G.; Day, G. M. An optimized intermolecular force field for hydrogen-bonded organic molecular crystals using

atomic multipole electrostatics. *Acta Crystallographica Section B-Structural Science Crystal Engineering and Materials* **2016**, *72*, 477-487.

(18) Nyman, J.; Pundyke, O. S.; Day, G. M. Accurate force fields and methods for modelling organic molecular crystals at finite temperatures. *Physical Chemistry Chemical Physics* **2016**, *18*, 15828-15837.

(19) Uzoh, O. G.; Galek, P. T. A.; Price, S. L. Analysis of the conformational profiles of fenamates shows route towards novel, higher accuracy, force-fields for pharmaceuticals. *Physical Chemistry and Chemical Physics* **2015**, *17*, 7936-7948.

(20) Sobol, I. M. *USSR Computational Mathematics and Mathematical Physics* **1967**, *7*, 86-112.

(21) Karamertzanis, P. G.; Pantelides, C. C. Ab initio crystal structure prediction - I. Rigid molecules. *Journal of Computational Chemistry* **2005**, *26*, 304-324.

(22) Karamertzanis, P. G.; Pantelides, C. C. Ab initio crystal structure prediction. II. Flexible molecules. *Molecular Physics* **2007**, *105*, 273-291.

(23) Kazantsev, A. V.; Karamertzanis, P. G.; Adjiman, C. S.; Pantelides, C. C. Efficient Handling of Molecular Flexibility in Lattice Energy Minimization of Organic Crystals. *Journal of Chemical Theory and Computation* **2011**, *7*, 1998-2016.

(24) Kazantsev, A. V.; Karamertzanis, P. G.; Adjiman, C. S.; Pantelides, C. C.: CrystalOptimizer. An efficient Algorithm for Lattice Energy Minimisation of Organic Crystal using Isolated-Molecule Quantum Mechanical Calculations. In *Molecular System Engineering*; Adjiman, C. S., Galindo, A., Eds.; Process Systems Engineering; WILEY-VCH Verlag GmbH & Co.: Weinheim, 2010; Vol. 6; pp 1-42.

(25) Cooper, T. G.; Hejczyk, K. E.; Jones, W.; Day, G. M. Molecular Polarization Effects on the Relative Energies of the Real and Putative Crystal Structures of Valine. *Journal of Chemical Theory and Computation* **2008**, *4*, 1795-1805.

(26) Day, G. M.; Price, S. L.; Leslie, M. Atomistic calculations of phonon frequencies and thermodynamic quantities for crystals of rigid organic molecules. *Journal of Physical Chemistry B* **2003**, *107*, 10919-10933.

(27) Yang, J.; Hu, W.; Usvyat, D.; Matthews, D.; Schutz, M.; Chan, H. Ab initio determination of the lattice energy in crystalline benzene to sub-kilojoule per mole accuracy. *Science* **2014**, *345*, 640-643.

(28) Bardwell, D. A.; Adjiman, C. S.; Arnautova, Y. A.; Bartashevich, E.; Boerrigter, S. X. M.; Braun, D. E.; Cruz-Cabeza, A. J.; Day, G. M.; Della Valle, R. G.; Desiraju, G. R.; van Eijck, B. P.; Facelli, J. C.; Ferraro, M. B.; Grillo, D.; Habgood, M.; Hofmann, D. W. M.; Hofmann, F.; Jose, K. V. J.; Karamertzanis, P. G.; Kazantsev, A. V.; Kendrick, J.; Kuleshova, L. N.; Leusen, F. J. J.; Maleev, A. V.; Misquitta, A. J.; Mohamed, S.; Needs, R. J.; Neumann, M. A.; Nikylov, D.; Orendt, A. M.; Pal, R.; Pantelides, C. C.; Pickard, C. J.; Price, L. S.; Price, S. L.; Scheraga, H. A.; van de Streek, J.; Thakur, T. S.; Tiwari, S.; Venuti, E.; Zhitkov, I. K. Towards crystal structure prediction of complex organic compounds - a report on the fifth blind test. *Acta Crystallographica Section B-Structural Science* **2011**, *67*, 535-551.

(29) Neumann, M. A.; Perrin, M. A. Energy ranking of molecular crystals using density functional theory calculations and an empirical van der Waals correction. *Journal of Physical Chemistry B* **2005**, *109*, 15531-15541.

(30) Grimme, S. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *Journal of Computational Chemistry* **2006**, *27*, 1787-1799.

(31) Tkatchenko, A.; Scheffler, M. Accurate Molecular Van Der Waals Interactions from Ground-State Electron Density and Free-Atom Reference Data. *Physical Review Letters* **2009**, *102*, 073005.

(32) Clark, S. J.; Segall, M. D.; Pickard, C. J.; Hasnip, P. J.; Probert, M. J.; Refson, K.; Payne, M. C. First principles methods using CASTEP. *Zeitschrift fur Kristallographie* **2005**, *220*, 567-570.

(33) McCrone, W. C.: Polymorphism. In *Physics and Chemistry of the Organic Solid State*; Fox, D., Labes, M. M., Weissberger, A., Eds.; Wiley Interscience: New York, 1965; Vol. II; pp 725-767.

(34) Threlfall, T. L.; Gelbrich, T. The crystal structure of methyl paraben at 118 K does not represent a new polymorph. *Crystal Growth & Design* **2007**, *7*, 2297-2297.

(35) Dubey, R.; Pavan, M. S.; Desiraju, G. R. Structural landscape of benzoic acid: using experimental crystal structures of fluorobenzoic acids as a probe. *Chemical Communications* **2012**, *48*, 9020-9022.

(36) Ccdc: Mercury. 2004.

(37) Gelbrich, T.; Hursthouse, M. B. A versatile procedure for the identification, description and quantification of structural similarity in molecular crystals. *CrystEngComm* **2005**, *7*, 324-336.

(38) Gelbrich, T.; Threlfall, T. L.; Hursthouse, M. B. XPac dissimilarity parameters as quantitative descriptors of isostructurality: the case of fourteen 4,5 '-substituted benzenesulfonamido-2-pyridines obtained by substituent interchange involving CF3/I/Br/Cl/F/Me/H. *CrystEngComm* **2012**, *14*, 5454-5464.

(39) Bhardwaj, R. M.; Price, L. S.; Price, S. L.; Reutzel-Edens, S. M.; Miller, G. J.; Oswald, I. D. H.; Johnston, B.; Florence, A. J. Exploring the Experimental and Computed Crystal Energy Landscape of Olanzapine. *Crystal Growth & Design* **2013**, *13*, 1602-1617.

(40) Srirambhatla, V. K.; Guo, R.; Price, S. L.; Florence, A. J. Isomorphous template induced crystallisation: a robust method for the targeted crystallisation of computationally predicted metastable polymorphs. *Chemical Communications* **2016**, *52*, 7384-7386.

(41) Arlin, J. B.; Price, L. S.; Price, S. L.; Florence, A. J. A strategy for producing predicted polymorphs: catemeric carbamazepine form V. *Chemical Communications* **2011**, *47*, 7074-7076.

(42) Bernstein, J.; Davies, R. E.; Shimoni, L.; Chang, N. Patterns in Hydrogen Bonding: Functionality and graph set analysis in crystals. *Angewandte Chemie-International Edition in English* **1995**, *34*, 1555-1573.

(43) McDonagh, J.; Palmer, D.; van Mourik, T.; Mitchell, J. Are the Sublimation Thermodynamics of Organic Molecules Predictable? *Journal of Chemical Information and Modeling* **2016**, *56*, 2162-2179.

(44) McDonagh, J. L.; Nath, N.; De Ferrari, L.; van Mourik, T.; Mitchell, J. B. Uniting Cheminformatics and Chemical Theory To Predict the Intrinsic Aqueous Solubility of Crystalline Druglike Molecules. *Journal of Chemical Information and Modeling* **2014**, *54*, 844-856.

(45) Palmer, D. S.; Llinas, A.; Morao, I.; Day, G. M.; Goodman, J. M.; Glen, R. C.; Mitchell, J. B. O. Predicting intrinsic aqueous solubility by a thermodynamic cycle. *Molecular Pharmaceutics* **2008**, *5*, 266-279.

(46) Pertsin, A. J.; Kitaigorodsky, A. I.: *The Atom-Atom Potential Method. Applications to Organic Molecular Solids*; Springer-Verlag: Berlin, 1987.

(47) Gavezzotti, A.: *Theoretical Aspects and Computer Modeling of the Molecular Solid State*; John Wiley: Chichester, 1997.

(48) Maiti, A.; Zepeda-Ruiz, L.; Gee, R.; Burnham, A. Vapor pressure and sublimation rate of molecular crystals: Role of internal degrees of freedom. *Journal of Physical Chemistry B* **2007**, *111*, 14290-14294.

(49) Lorenz, H.; Seidel-Morgenstern, A. Processes To Separate Enantiomers. *Angewandte Chemie International Edition* **2014**, *53*, 1218-1250.

(50) Nyman, J.; Day, G. Modelling temperature-dependent properties of polymorphic organic molecular crystals. *Physical Chemistry Chemical Physics* **2016**, *18*, 31132-31143.

(51) Nyman, J.; Day, G. Static and lattice vibrational energy differences between polymorphs. *Crystengcomm* **2015**, *17*, 5154-5165.

(52) Klussmann, M.; White, A. J. P.; Armstrong, A.; Blackmond, D. G. Rationalization and Prediction of Solution Enantiomeric Excess in Ternary Phase Systems. *Angewandte Chemie International Edition* **2006**, *45*, 7985-7989.

(53) Otero-de-la-Roza, A.; Cao, B. H.; Price, I. K.; Hein, J. E.; Johnson, E. R. Predicting the Relative Solubilities of Racemic and Enantiopure Crystals by Density-Functional Theory. *Angewandte Chemie-International Edition* **2014**, *53*, 7879-7882.

3 Chapter 3: Kinetics or Thermodynamics: A Crystal Structure Prediction Study of 3-Chloromandelic Acid

3.1 Introduction

Mandelic acid and its derivatives are prototypical chiral molecules. They are commonly used as precursors for a number of substances, including antibacterial agents and cosmetics as well as intermediates in the industrial synthesis of many target molecules.[1] As such, these molecules require carefully designed separation processes in order to obtain a single enantiomer.[2,3] 3-chloromandelic acid (3ClMA) is one such molecule.

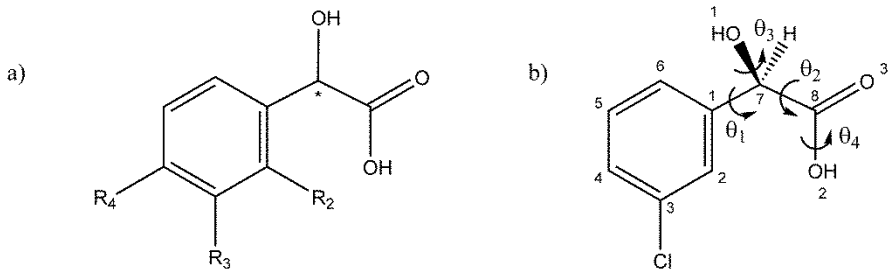


Figure 3-1: (a) The mandelic acid family, where the phenyl substituents R2, R3 or R4 include H, F, Cl, Br, I, CF_3, Me and OMe. The crystal structures are denoted by chirality (R/S/RS), the substituents, and polymorph, for example S-3ClMA_2. (b) The flexible torsion angles in 3ClMA considered in the computational generation of crystal structures.

3ClMA has been the focus of two projects; the recent European INTENANT (INTegrated synthesis and purification of single ENANTiomers) project[3,4] and a crystal engineering analysis of the structural systematics of racemic mandelic acid compounds.[5] At the start of this study only limited research had been performed on 3ClMA, and while previous studies had explored the tertiary phase diagram,[6] there was only a small amount of information on the polymorphs and solid solutions that were able to form.[1,2] Work at the MPI on the binary phase diagrams gave rise to four different polymorphs; two racemic (RS-stable and RS-metastable) and two enantiopure (R-stable and R-metastable). While PXRD patterns of each polymorph had been published,[6] the crystal structures had not been solved. Initially it was hoped that crystal structure prediction methods could be used to help characterise these structures. However, during the course of this study, five crystal structures of 3ClMA were solved by collaborators at the University of Southampton. These five crystal

structures consisted of three racemic structures (RS form1, RS form 2 and RS form 3) and two enantiopure structures (S form 1 and S form 2). Comparison of the PXRD patterns of the structures from SOUTHAMPTON with those generated at the MPI showed some overlap. For the racemic structures, the experimental PXRD pattern of RS-stable was found to overlay with the simulated powder pattern of SOUTHAMPTON's RS-form 3 and RS-metastable with RS form 1. RS form 2 was found to be structurally related to RS form 1.[7] For the enantiopure structures the experimental PXRD pattern of R-stable was found to overlay with the simulated powder pattern of S-3ClMA form 1 and that of R-metastable to overlay with the simulated powder pattern of S-3ClMA form 2. Thus, there are currently five known polymorphs of 3ClMA, the structures of which can be seen in Figure 3.2.

The polymorphs of 3ClMA are complex. Racemic forms 1 and 2 are structurally related, while enantiopure form 1 is disordered and enantiopure form 2 has four independent molecules in the asymmetric unit (Z'=4). The experimental heats of formation have been determined for four out of the five polymorphs.[6] Assuming the system is monotropic, these enthalpies of fusion allow the structures to be ranked in order of their stability and provides a cross reference for relative stability based on lattice energy calculation. RS form 3 has been determined to be the most stable crystal structure, with an enthalpy of fusion of 27.92 kJ mol^{-1}, relative to RS form 1, with an enthalpy of fusion of 17.77 kJ mol^{-1}. The relative stability of RS form 2 is unknown at present. For the enantiopure structures, S form 1 was found to have an enthalpy of fusion of 22.55 kJ mol^{-1}, relative to S form 2 with an enthalpy of fusion of 14.62 kJ mol^{-1}. All enthalpies of fusion were measured with an error of $\pm$2.4%.[6]

It is hoped that, by using CSP methods, it might be possible to explore the whole range of crystal packings available to 3ClMA through analysis of the hypothetical crystal structures produced. How likely these hypothetical crystal structures are to be observed experimentally can be determined through examination of the relative lattice energies and comparing the structures to experimentally known mandelic acid derivatives.

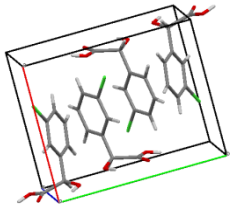
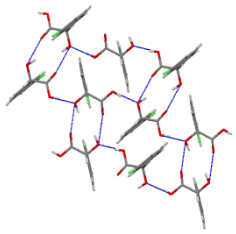
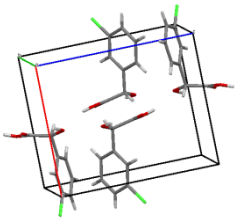
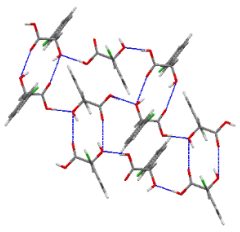
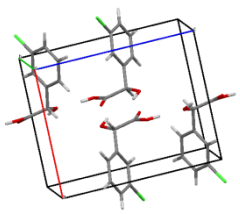
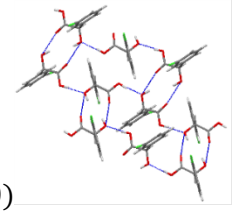
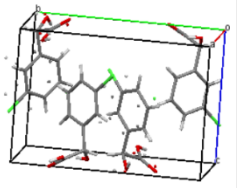
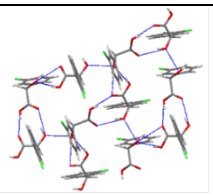
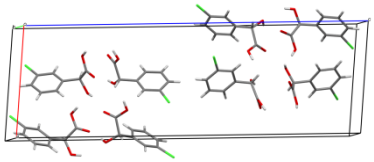
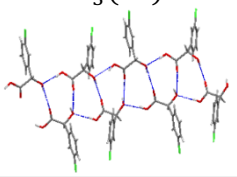
Form and XRD determination	Unit Cell and XPAC construct	Hydrogen Bonding Motif
RS-3-Chloromandelic Acid		
RS Form 1 $P2_1/c$, Z'=1 Single crystal, determined at RT Formed from dissolution in polar aprotic solvents ΔH_f = 17.77 kJ mol^{-1}	 A32	$R_2^2(10)$ 
RS Form 2 P-1, Z'=2 Single crystal, determined at 100 K Low temperature form, obtained by cooling a single crystal of form 1.	 A32	$R_2^2(10)$ 
RS Form 3 $P2_1/c$, Z'=1 Single crystal, determined at 100 K Obtained by dissolution in dichloromethane ΔH_f = 27.92 kJ mol^{-1}	 1DB	 $R_2^2(10)$
S-3-Chloromandelic Acid		
S Form 1 $P2_1$, Z'=2 Single crystal, determined at RT Exhibits phenyl disorder. Obtained by slow solvent evaporation from water ΔH_f = 22.55 kJ mol^{-1}	 3D4	 $R_2^2(9)$
S Form 2 $P2_1$, Z'=4 Single crystal, determined at 100 K Crystallized from toluene ΔH_f = 14.62 kJ mol^{-1}	 3D2	$R_3^3(11)$ 

Figure 3-2: An overview of the experimental crystal structures, crystallization conditions[6,7] and thermodynamic data[6] of 3ClMA. The cell packing diagram is labelled by the XPAC motif (defined in Figure 3.3) and the hydrogen bonding motif with its graph set (defined in Chapter 2.8.3).

Due to the complexity of the experimentally known 3ClMA crystal structures some considerations need to be taken into account when determining lattice energies. The disordered structure S-3ClMA form 1 particularly needs to be considered carefully as lattice energies of disordered cannot be evaluated exactly through current CSP methods.[8] S-3ClMA form 1 is a Z'=2 structure where one molecule is disordered in a 2:1 ratio. Hence, the structure can be represented by two individual crystal structure each with the same cell

dimensions and one molecule identical. The only difference will lie in the placement and the geometry of the second molecule. These structures will represent the two end points of the disordered structure and will be considered separately when performing lattice energy calculations.[9] For the purposes of this study the major component will be denoted S-3ClMA from 1_A and the minor component S-3ClMA form 1_B. It is thought that the overall lattice energy of the disordered structure will lie somewhere between the energies of these two end points and can be approximated to be a 2:1 average of the components.

3.1.1 Types of Packing in Mandelic Acids

In the determination of the crystal structures of RS forms 1 and 2 it has been demonstrated that the two polymorphs are isostructural with each other.[7] Traditionally the definition of isostructurality only encompasses the entire structure across three crystallographic translations; however it can be extended to 2D molecular arrangements (sheets or planes), 1D arrangements (chains or stacks) and 0D arrangements (dimers) (see Chapter 2.8.2). Collaborators at the University of Southampton have been exploring the notion of isostructurality and its relationship to polymorphism. As part of the study they explored the relationships of a number of mandelic acid derivatives. A systematic study was performed on 28 different racemic mandelic acid derivaties and the relationships between them. These structures include fluoro, chloro, bromo, iodo, trifluoromethyl, methyl and methoxy monosubstituted mandelic acids. XPac[10] was used to investigate the degree of isostructurality of racemic and enantiopure mandelic acid derivates. The tolerance in the discrepancy between the two vectors being compared for the structures shown in Figure 3.3, was determined by the default changes in angles (Δa), planes (Δp) and distances (Δd) for a tight tolerance as Δa = 5 °, Δp = 7 ° and Δd = 0.5 Å to determine the degree of isostructurality: 0D, 1D, 2D or 3D.

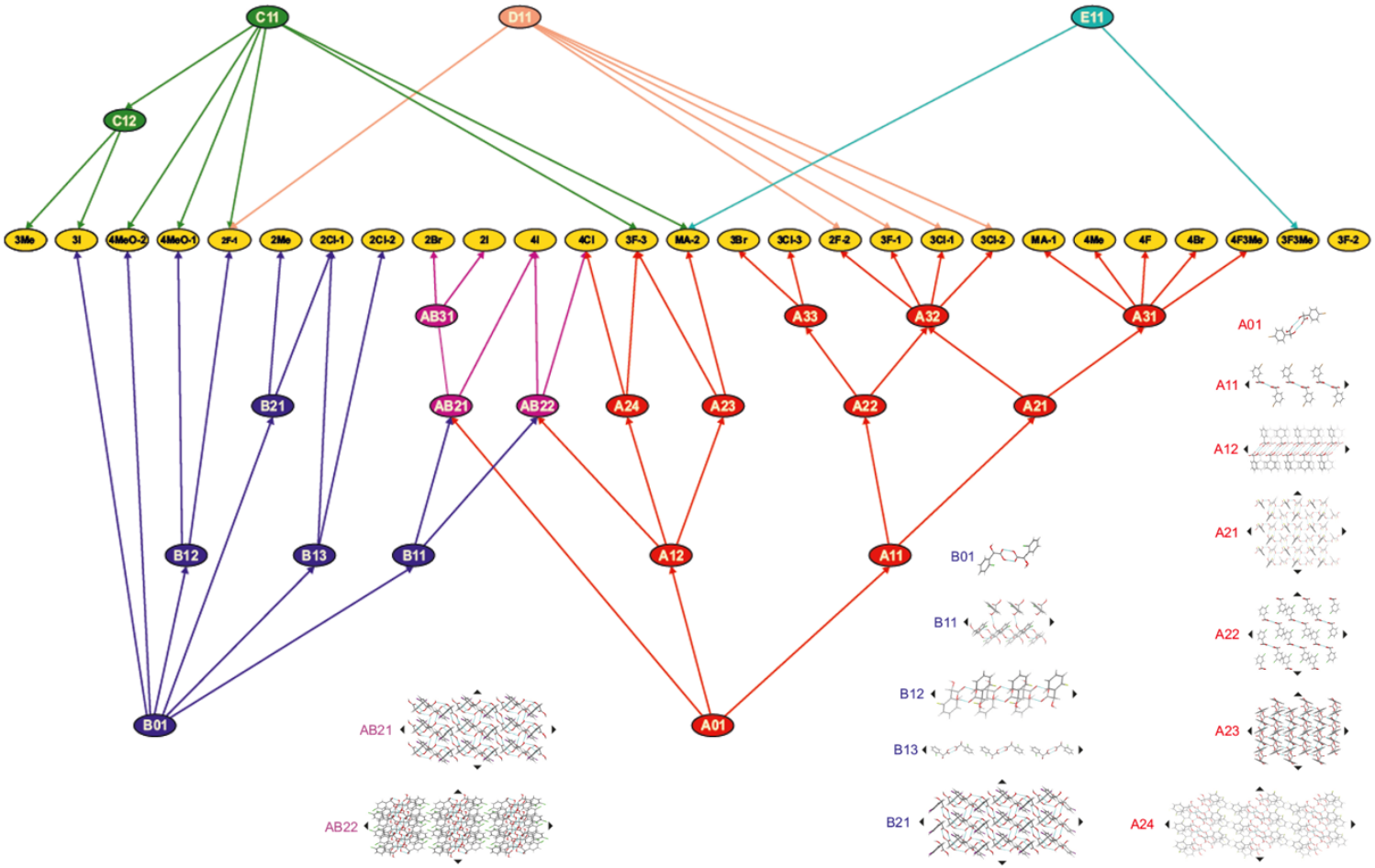


Figure 3-3: Structural relationship diagram of experimental racemic mandelic acid derivatives – yellow nodes represent crystal structures, all other nodes are supramolecular constructs (SC). Each SC is identified by a letter and colour denoting a related group and two numbers representing the dimensionality and identifier. SCs derived from others are linked by arrows. 0D, 1D and 2D SCs derived from the COOH dimer (B01) and/or carboxyl-hydroxyl dimer (A01). A31, A32, A33 and AB31 are 3D SCs represented by the crystal structures themselves. The C, D and E SCs are based on catemers. This plot is based on that published[11] for the A and B dimer-based structures.

Analysis of the crystal structures of the 28 racemic mandelic acid derivatives led to the Hasse plot shown in Figure 3.1. It was observed that, for the majority of the mandelic acid derivatives studied, the packings are based on COOH (A0) or carboxyl-hydroxyl dimers (B0) as previously described.[11] Additionally, the study shows that molecules with substituents that are similar in size, such as 3BrMA and 3ClMA form 3, tend to adopt crystal structures that are 3D isostructural. This type of analysis can help to identify plausible crystal structures which have not been observed experimentally. For example, if a 3D relationship is observed between crystals containing two similar molecules, then it is probable that any other molecules which are also structurally similar could also form the same crystal structure. Observations such as these can help design crystallisation experiments, such as seeding or templating experiments, to produce first crystallisations of these polymorphs which have been identified with CSP. Researchers at the University of Southampton used this structural relationship diagram to see if they could seed and find more structures of mandelic acid derivatives.

The investigation into the structural relationships between mandelic acid derivatives in Figure 3.3 was not a complete study and was limited to racemic compounds only. The hydrogen bonding motifs and structures observed enantiopure mandelic acid derviatives are very different from those seen in the racemic compounds. In Section 1.3.4 I extend this investigation of the mandelic acid derivatives to include enantiopure compounds. It is important to note that as the number of structures in an isostructural study is increased the potential of finding new relationships and avenues of investigation for finding new polymorphs experimentally increases.

3.2 Methods

3.2.1 Generation of hypothetical structures and subsequent refinement

CrystalPredictor[12,13] was used to generate Z'=1 structures of 3ClMA in the most common space groups: racemic $P\bar{1}$, $P2_1/c$, $Pna2_1$, $Pca2_1$, *Pbca*, *Pbcn*, $C2/c$, *Cc*, *Pc*, *Cm*, $P2_1/m$, $C2/m$, $P2/c$, $Pmn2_1$, $Cmc2_1$, $Pmn2_1$, $Cmc2_1$, *Aba*2, *Fdd*2, *Iba*2, *Pnna*, *Pccn*, *Pbcm*, *Pnnm*, *Pmmn*, *Pnma*, *Cmcm*, *Cmca*, *Fddd*, *Ibam*, $I\bar{4}$,

*P*4/*n*, *P*4$_2$/*n*, *I*4/*m*, *I*4$_1$/*a*, *P*$\bar{4}$2$_1$/*c*, *I*$\bar{4}$2*d*, *P*$\bar{3}$, *R*$\bar{3}$, *R*3/*c*, *R*$\bar{3}$/*c*, *P*6$_3$/*m* and *Pa*$\bar{3}$ and chiral *P*1, *P*2$_1$, *C*2, *P*2$_1$2$_1$2, *P*2$_1$2$_1$2$_1$, *C*222$_1$, *P*4$_1$, *P*4$_3$, *P*4$_1$2$_1$2, *P*4$_3$2$_1$2, *P*3$_1$, *P*3$_2$, *R*3, *P*3$_1$2$_1$, *P*322$_1$, *P*6$_1$, *P*6$_3$ and *P*2$_1$3.

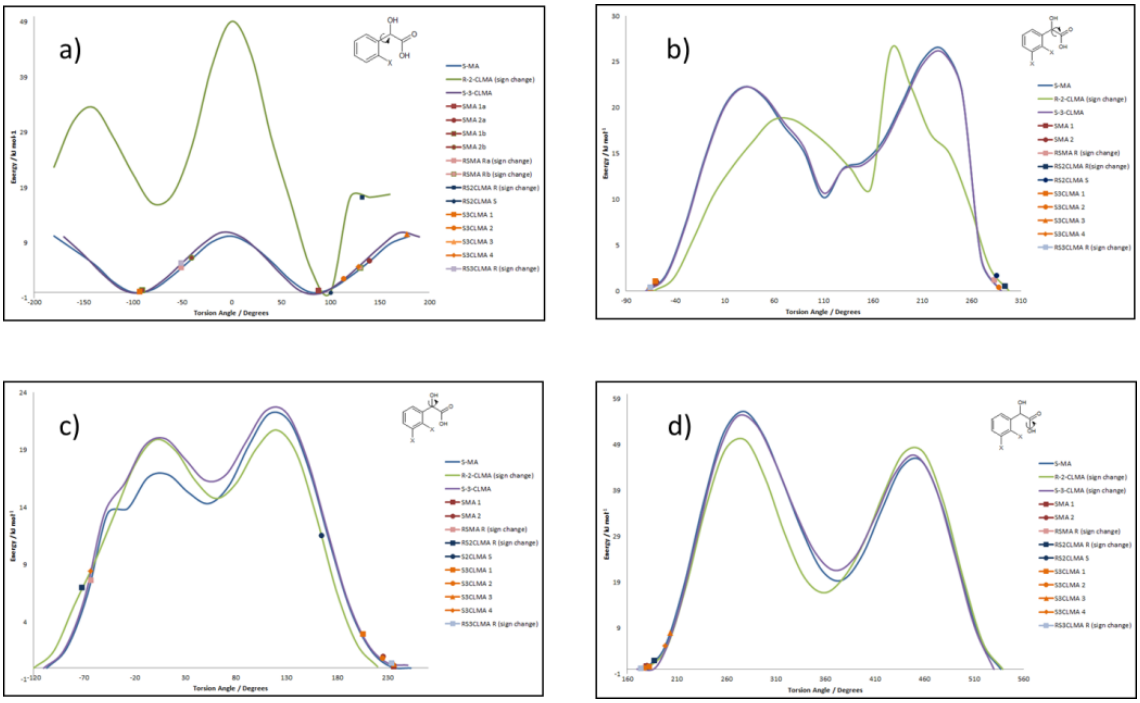


Figure 3-4: Torsion scans of S-mandelic acid (light blue), RS-2-chloromandelic acid (green) RS-3-chloromandelic acid (dark blue) and S-3-chloromandelic acid (purple) about a) θ1, b) θ2, c) θ3 and d) θ4. The experimentally determined dihedral angles of RS-3CIMA form 3 and S-3CIMA form 2 are indicated.

3CIMA is a highly flexible compound. Torsion angle scans were performed for potentially flexible bonds of 3CIMA (see Figure 3.4), with any angles having a conformational energy below 10 kJ mol^{-1} over a particular range being deemed to be flexible. Analysis of the torsion angle scans shows a very low energy barrier for rotation for the phenyl ring, allowing it to essentially freely rotate and pack in many different ways. Additionally, there are two possible positions in which the carboxylic acid could form an internal hydrogen bond, either between O1 and O2 or O1 and O3 (see Figure 3.1). Thus it was determined the following four flexible torsion angles should be considered for the search: θ1 = C2-C1-C7-C8, θ2 = C1-C7-C8-O2, θ3 = C1-C7-O1-H6 and θ4 = C7-C8-O2-H7. This gives rise to four different conformational minima which should be considered, denoted by A, B, C and D (see Figure 3.5). These four conformations define a conformational region used as a basis to generate the hypothetical crystal structures with *CrystalPredictor*, with each conformation considered as a separate search.

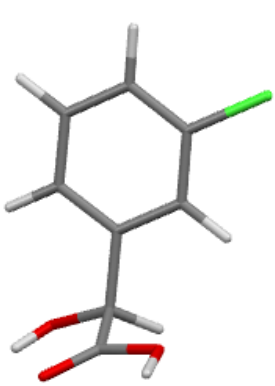
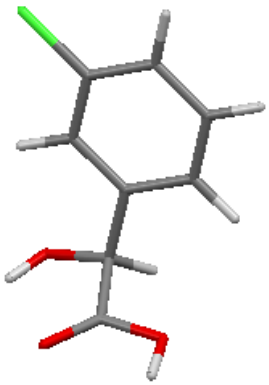
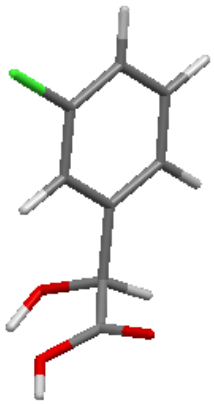
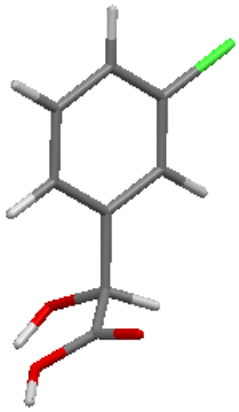
A	B	C	D
			
$\Delta E = 0.00$ kJ mol^{-1}	$\Delta E = 0.07$ kJ mol^{-1}	$\Delta E = 7.71$ kJ mol^{-1}	$\Delta E = 7.35$ kJ mol^{-1}
θ_1 = 0° to + 180° θ_2 = -90° to -30° θ_3 = +180° to +300° θ_4 = +150° to +210°	θ_1 = -180° to 0° θ_2 = -90° to -30° θ_3 = +180° to +300° θ_4 = +150° to +210°	θ_1 = 0° to + 180° θ_2 = +110° θ_3 = +180° to +300° θ_4 = +150° to +210°	θ_1 = -180° to 0° θ_2 = +110° θ_3 = +180° to +300° θ_4 = +150° to +210°
S-3ClMA form 1 S-3ClMA form 2	RS-3ClMA form 1 RS-3ClMA form 2 RS-3ClMA form 3	-	-

Figure 3-5: The four different fully optimized (MP2 6-31G(d.p)) conformations of 3ClMA, and their relative energies. The range of low energy torsions around these minima considered in the search in the four different regions are defined, and the experimental structures which have molecules within these conformational regions are listed.

A search was carried out for each conformational minimum, using a grid of ΔE_{intra} values and atomic point charges for variations in torsion angles, θ1, θ2, θ3 and θ4, in 20° steps using constrained optimization that had been constructed at the MP2/6-31G(d,p) level of theory using Gaussian03.[14] The lattice energies were evaluated from the atomic charges, interpolated from the grid, with all other terms represented by an isotropic atom-atom 6-exp potential, with the parameters of the FIT potential.[15]

The global *CrystalPredictor* search examined one million structures in each of the four conformational regions, A, B, C and D, with the calculated lattice energies refined further in a multi-stage process. All unique structures within 20 kJ mol^{-1} of the global minimum for regions A and B, and 10 kJ mol^{-1} for regions C and D were taken forward to be refined using *CrystalOptimizer*,[16] covering approximately 9600 structures. The *CrystalOptimizer*, algorithm refines the crystal structure by using a more accurate distributed multipole model for the intermolecular electrostatic forces while also allowing the torsion angles θ_1, θ_2, θ_3 and θ_4 to change in response to the crystal packing forces in order to minimize U_{latt}. The intermolecular energies, ΔE_{intra}, were evaluated with

Gaussian using the MP2 6-31G(d,p) wavefunction and analysed by GDMA[17] to give the distributed multipoles. The rigid molecule intermolecular lattice energy was optimised using DMACRYS.[18] The large number of iterations required to converge this two level optimisation problem, up to 200 interations, to determine a mechanically stable structure, is aided by using a database of the *ab initio* calculations. Due to the large number of structures optimised for 3ClMA the use of a database is exceptionally useful here at reducing the computational time taken to produce the final energy landscape.

The distributed multipole model for the electrostatic intermolecular forces assumed that the molecule has the same charge distribution in the crystal as in isolation, and that the crystalline environment does not change the relative energies of different conformations. This assumption can be poor for molecules with very different conformations and hydrogen bonding capabilities, such as 3ClMA. The effect of polarisation is estimated through the use of the polarisable continuum model (PCM) using the MP2 wavefunction with a dielectric constant of ε = 3 (see Chapter 2.6.3) This provided a polarized estimate of the intramolecular energy, ΔE_{intra}, and a set of distributed multipoles, for reminimising the lattice energy, starting from all structures within 10 kJ mol^{-1} of the global minimum (the lowest 1080 structures).

To estimate the effects of temperature, the rigid body elastic constants and k=0 phonons were calculated for all of the structures on the crystal energy structures (i.e. those which had been refined following PCM) using the same intermolecular potential and used to calculate the Helmholtz free energy and zero-point vibrational energy (see Chapter 2.6.4).

3.2.2 Method for periodic electronic structure calculations

A plane wave DFT-D approach, using the CASTEP code, with the PBE wavefunction was used to evaluate the lattice energy of selected structures. Using the Tkatchenko-Scheffler[19] (TS) dispersion correction, a full geometry optimisation was performed, starting from either the experimental structure (with hydrogen positions elongated to give standard neutron bond lengths), or the CSP crystal structures prior to the PCM correction. A k-point spacing of 0.05 Å^{-1}

and a k-cutoff of 750 eV were used for all structures. The TS optimised structures were then used as the starting point for reoptimisation with the G06[20] dispersion correction.

3.3 Results and Discussion

3.3.1 The Crystal Energy Landscape of 3ClMA

The crystal energy landscape of 3ClMA is exceptionally dense, with approximately 3050 unique hypothetical crystal structures, across the four different conformational regions, lying within 20 kJ mol^{-1}. There is no distinction between the relative energy and density of the enantiopure and racemic crystal structures of 3ClMA, as can be observed in the crystal energy landscape. This corresponds well with the recent survey[21] of the Cambridge Structural Database (CSD) that showed there are no systematic density differences between enantiopure and racemic compounds and that the energy differences between these are often very small, in the same order as polymorphic energy differences, and within the uncertainty of current computational methods.

Of the 3050 unique crystal structures produced the lowest 100 unique structures were examined further (see Section 1.3.3). These are the structures which fall below the dashed line in Figure 3.6. 100 structures was chosen as the cut off as it was a manageable number for post processing analysis by human comparison. These structures fell within an energy range of 6.59 kJ mol^{-1} which can be considered sufficient to cover the range of polymorphism in such a small molecule. The crystallographic details of these structures can be found in Appendix 3.6.1.

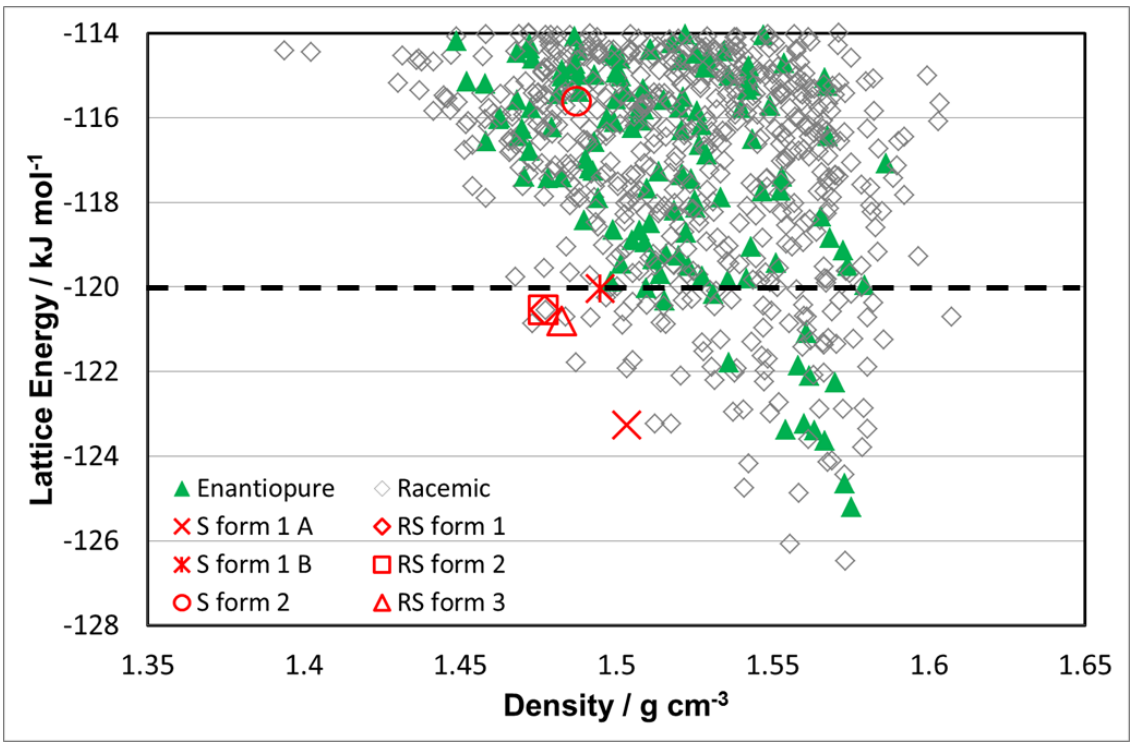


Figure 3-6: The crystal energy landscape of 3-chloromandelic acid. The enantiopure structures are denoted by green triangles, racemic by grey diamonds. The lattice energy minima corresponding to the experimental structures are given in red, and were found in the search if Z'=1. Structures shown below the dotted line (at -120 kJ mol^{-1}) were considered in greater detail.

3.3.2 Reproduction of the experimental crystal structures

The experimental crystal structures are well produced from the *CrystalOptimizer* refinement, all with acceptable RMS overlays (see Table 3.1). The only experimental structure not modelled particularly well is the minor component of S-3ClMA form 1. Here only 13 of the 15 molecules in the 15 molecule overlay were found to be in common. However, this is the minor component of the disordered structure, in which the hydrogen positions have not been fully refined. This may be the cause of the discrepancy between the experimental and refined structure.

All Z'=1 experimental crystal structures, namely RS forms 1 and 3, were found in the search. Although the energies of the Z' > 1 experimental crystal structures are shown on the crystal energy landscape, their structures were beyond the scope of the search and so a match was not expected to be found.

Structure	Z	Z'	U_{inter} / kJ mol^{-1}	ΔE_{intra} / kJ mol^{-1}	Total lattice energy (kJ mol^{-1})	RMS_{15} / Å
RS-3ClMA form 1	4	1	-124.79	+4.25	-120.54	0.146
RS-3ClMA form 2	4	2	-124.74	+4.20	-120.54	0.186
RS-3ClMA form 3	4	1	-125.52	+4.72	-120.81	0.308
S-3ClMA form 1_A	4	2	-131.11	+7.83	-123.28	0.262
S-3ClMA form 1_B	4	2	-129.70	+9.65	-120.05	0.353* (RMS_{13})
S-3ClMA form 2	8	4	-117.80	+2.18	-115.62	0.188
B2	4	1	-128.05	+1.56	-126.49	N/A
A315	2	1	-128.57	+3.36	-125.21	N/A

Table 3-1: The lattice energies of the experimentally known structures and computer generated structures B2 and A315 as determined with *CrystalOptimizer* with MP2 wavefunction, FIT exp-6 dispersion-repulsion model and PCM. B2 represents the lowest energy racemic structure from the search and A315 represents the lowest energy enantiopure structure from the search. The RMS_{15} error of comparing the optimised crystal structures with the experimental crystal structures is stated.

While the major component of disordered S form 1 structure was determined to be lower in energy than the most stable crystal structure RS form 3, the actual energy of the disordered complex is thought to lie somewhere between the two components. Experimentally determined RS form 3 was found to match structure B7 from the search. An RMS_{15} overlay of 0.371 Å was obtained when these two structures were overlaid (see Figure 3.7). While the RMS of the overlay is relatively high it is still acceptable. The simulated PXRD pattern of structure B7 closely matches the experimental PXRD pattern of RS form 3, further indicating that this structure was found in the search.

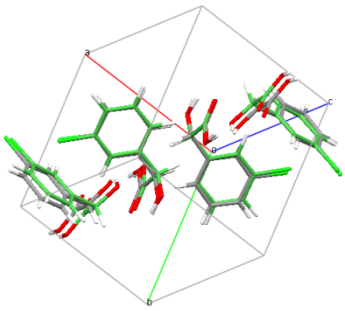


Figure 3-7: Overlay of experimental RS-3CIMA form 3 (grey) and computer generated structure B7 (green) from the search

RS form 1 is a Z'=1 structure that is related to Z'=2 RS form 2 by a phase transition. When the refined crystal structures of RS forms 1 and 2 were added to the crystal energy landscape, it was observed that both structures overlapped with the hypothetical structure B50. Structure B50 overlays with Z'=1 RS form 1 with an RMS_{15} of 0.147 Å and Z'=2 RS form 2 with an RMS_{15} of 0.19 Å. Thus, the hypothetical crystal structure B50 is a good match for both structures in addition to indicating that B50 is the same structure as RS form 1. As forms 1 and 2 have been found to be isostructural it is not surprising that they both overlay well with structure B50. A comparison of the experimental PXRD pattern of RS form 1, RS form 2 and the simulated pattern of B50 show that there are strong similarities between the structures (see Figure 3.8), again indicating that RS form 1 was produced and found in the search. The limitations of the search being Z'=1 mean that RS form 2 will not be found in the search as it is a Z'=2 structure, however the isostructurality which it exhibits with RS form 1 means that the powder patterns of the Z'=1 hypothetical crystal structure B50 are very similar.

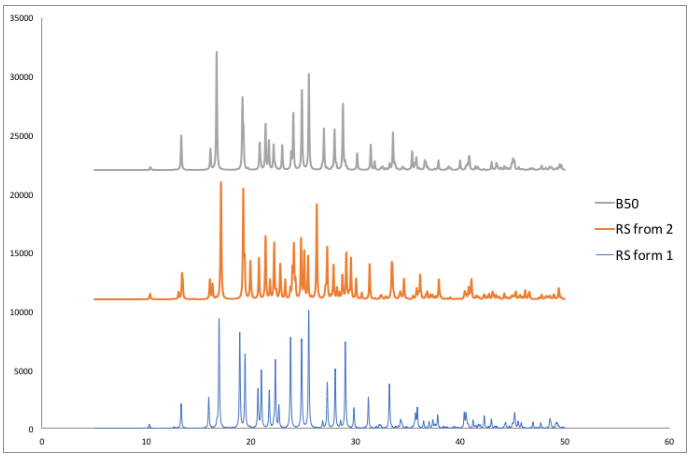


Figure 3-8: Comparison of the simulated powder pattern of hypothetical crystal structure B50 with the experimental PXRD patterns of 3CIMA RS forms 1 and 2

The racemic experimental structures all form crystals with the molecule in conformational region B, whereas the enantiopure structures all form crystals with the molecule in conformational region A. Conformational region A is relatively the most stable with conformational region B, with a difference in energy between the minimum conformations of just 0.07 kJ mol^{-1} see (Figure 3.5). Racemic molecules are generally easier to pack together than enantiopure molecules and so the energetic cost caused by having the molecules held in a less favourable conformation could be overcome by intermolecular interactions between the molecules in the crystal lattice. Due to the greater difficulty enantiopure crystals have in packing together the extra energetic stability gained by being in the most stable conformation could help to stabilise the crystal structure where intermolecular interactions cannot. The energy minima of conformational regions C and D are over 7 kJ mol^{-1} higher in energy than conformations A and B however there are hypothetical crystal structures within the lowest 100 structures with 3ClMA in these conformations (See Appendix 3.6.1). This further suggests that intermolecular forces within the crystal structures are far more dominant in determining the stability of the crystal structures, rather than the conformation of the molecule.

3.3.3 Analysis of the lowest structures in the search

The experimentally known crystal structures are found at much higher energy and lower density than a large number of the hypothetical crystal structures generated. Analysing the hydrogen bonding motifs of lowest 100 crystal structures, which lie within 7 kJ mol^{-1} of the global minimum, shows that there are 20 unique motifs that 3ClMA is able to form (see Table 3.2). Five of these hydrogen bonding motifs are also observed in the crystal structures of other known mandelic acid derivatives. The hydrogen bonding motifs observed amongst the hypothetical structures can stack together in a multitude of different ways allowing a large variety of structures to form. There are examples of the hydrogen bonding motives forming enantiopure ribbons or layers that, due to the various ways in which the phenyl rings can pack together, can produce both racemic and enantiopure structures. While there are groups of structures which share the same hydrogen bonding motif, there is no definitive area of the crystal energy landscape which is dominated by one single motif – the overall spread of that particular hydrogen bonding motif is relatively large.

This suggests that, in general, there is no overall preference for a molecule to adopt a particular hydrogen bonding motif across the crystal energy landscape.

The array of packing types observed amongst the hypothetical crystal structures of 3ClMA far outweighs the number seen experimentally (See Table 3.2). The most common hydrogen bonding pattern observed is motif 2, with 38 of the 100 structures adopting this motif. Motif 2 is a relatively open hydrogen bonding pattern, allowing for small variations in hydroxyl positions and phenyl packing to occur whilst retaining the overall hydrogen bonding network. However, the majority of packings observed are limited to just 1 or 2 structures as the hydrogen bonding network is much more rigid. The sheer number of hydrogen bonding motifs observed amongst the hypothetical crystal structures to highlight the variety of ways in which the phenyl rings can pack together.

Motifs adopted by large numbers of structures do not seem to be exclusive to either racemic or enantiopure crystal structures, this is of particular note for motifs 2, 3 and 5, but trends in their relative ranking can be observed. For motif 2 the first 9 structures, encompassing an energy range of 4 kJ mol^{-1}, are dominated by racemic structures, however after this point there is a mix of enantiopure and racemic structures some of which have less than 0.1 kJ mol^{-1} difference in energy between them. Motif 3 is adopted by the lowest energy enantiopure structure found in the search. All but one of the structures which adopts this motif lie within an energy range of less than 2 kJ mol^{-1} with an equal spread of racemic and enantiopure structures in this range indicating that motif 3 can be readily adopted by either racemic or enantiopure structures around the global minimum. Motif 5 is dominated by racemic structures. The enantiopure structures which adopt motif 5 were analysed to determine if any symmetry operations had been missed. It was found that the enantiopure structures which adopt motif 5 could be reduced to Z'=2 structures in racemic space groups. If the search had been expanded to Z'=2 then it is possible that these structures would correspond to Z'=2 minima.

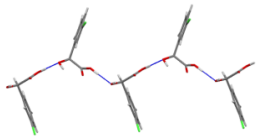
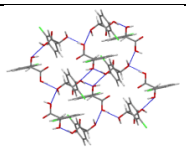
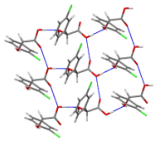
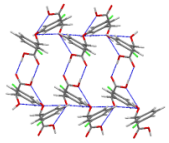
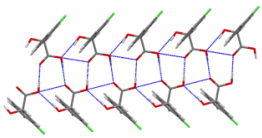
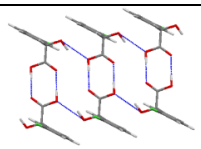
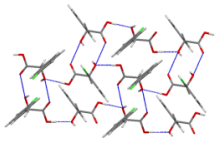
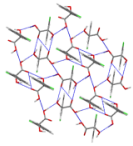
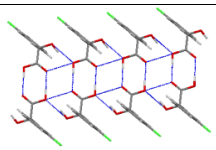
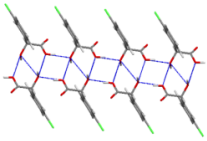
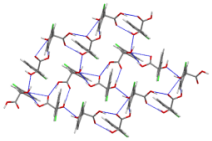
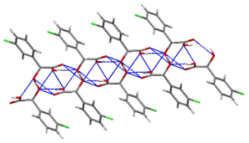
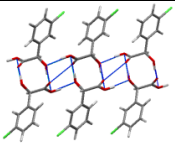
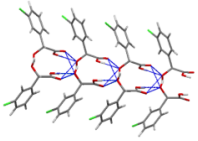
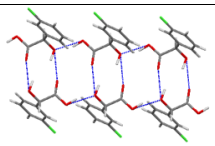
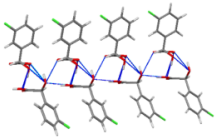
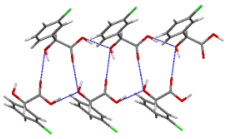
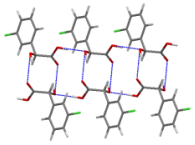
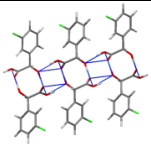
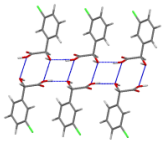
Motif No	Motif	Motif No	Motif
1 Adopted by 2 structures		11 Adopted by 2 structures	
2 Adopted by 38 structures		12 Adopted by 1 structure	
3 Adopted by 10 structures		13 Adopted by 1 structure	
4 Adopted by 8 structures including RS-3ClMA form 1 RS-3ClMA form 2 RS-3ClMA form 3		14 Adopted by 1 structure	
5 Adopted by 15 structures		15 Adopted by 2 structures	
6 Adopted by 1 structure including S-3ClMA form 1		16 Adopted by 1 structure	
7 Adopted by 4 structures		17 Adopted by 2 structures	
8 Adopted by 2 structures		18 Adopted by 2 structures	
9 Adopted by 5 structures including S-3ClMA form 2		19 Adopted by 1 structure	
10 Adopted by 1 structure		20 Adopted by 1 structure	

Table 3-2: The hydrogen bonding motifs observed within the lowest 100 structures from the search. The structures are ranked in order of stability, based on the relative energy of the structure which was first observed to form the motif. The number of structures which were observed to form each motif is listed as well as any experimental 3ClMA structures which have been observed show a particular motif.

S form 2, the high energy Z'=4 experimentally determined polymorph of 3ClMA, adopts hydrogen bonding motif 9. However, within the set of hypothetical crystal structures studied, motif 9 is only observed for racemic crystals, all containing molecules which adopt the A conformation only. Experimentally only enantiopure crystals have been observed to adopt this conformation, with racemic structures preferring the B conformation. Although there are racemic crystal structures that can adopt conformation A and appear to be able to form low energy thermodynamically plausible structures, the fact no racemic structures have been observed experimentally with this conformation suggests there is some barrier for their formation.

In the lowest 100 hypothetical crystal structures analysed, all four conformations, A, B, C and D, are observed. It is also noted that there are many pairs of structures are identical apart from the position of the Cl substituent (i.e. conformation A substituted for conformation B or D for C and vice versa) that are also close in energy. An example of this can be seen with hypothetical crystal structures A24 and B547 (see Figure 3.7). Both structures form identical hydrogen bonding networks, even though the chlorine atoms are situated on opposite sides of the phenyl ring, with an energy difference of just 0.856 kJ mol^{-1} between the two structures. The small energy difference between these two structures indicates that the position of the chlorine atom does not have a large effect on the overall lattice energy. The lack of preference energetically for the chlorine position suggests there could be further, almost identical, crystal structures within the search attributing to the density of the crystal energy landscape overestimating the number of truly distinct crystal structures by as much as 400%. This also leads to the suggestion that it could be relatively easy for Z'>1 disordered structures to form, comprising of molecules simply containing a difference in chlorine position only. This may help to explain why S form 1 is a disordered complex, where the differences in the positions of the atoms in the second molecule mainly lie with those in the phenyl ring (i.e. an approximate switch between conformation A and C or B and D).

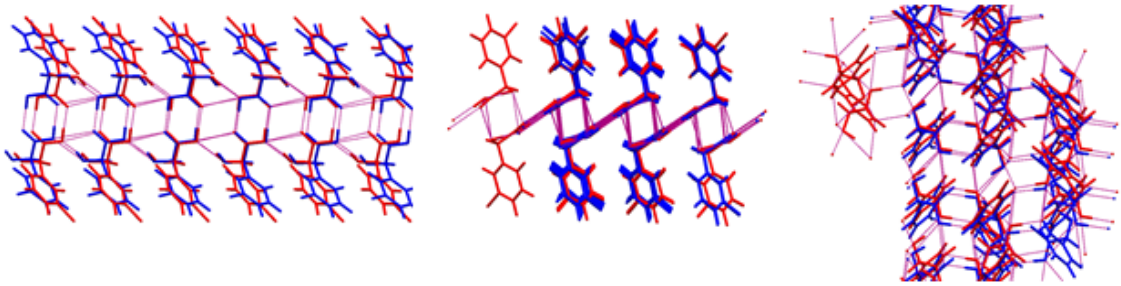


Figure 3-9: Overlays of theoretical crystal structures B49 (red) and A379 (blue) which exhibit the same hydrogen bonding motifs

3.3.4 Exploring the isostructurality of enantiopure mandelic acid derivatives

The isostructurality of mandelic acid derivatives has previously only been assessed for racemic compounds. This study has been extended to include 13 enantiopure mandelic acid derivatives through collaboration with both the University of Southampton and the Max Planck Institute. Comparison of these crystals shows that they adopt five distinct 3D structures, the majority of which are complex, with only S-4FMA having just one molecule in the asymmetric unit. While all enantiopure crystal structures are able to form catemeric crystal structures, the adoption of high Z' structures[22] could allow the formation of either of the two hydrogen-bonded dimers observed in 27 of the 30 racemic structures[23] in the previous study. Due to the inherent nature of enantiopure molecules, symmetry prevents the formation of the dimeric interactions forming for Z'=1 crystal structures due to symmetry. The Hasse plot of the racemic mandelic acid derivatives (see Figure 3.3), while mainly being focussed on dimeric interactions, shows some similarities between the catemeric chains observed with enantiopure structures. XPac[10] analysis looking for similarities between enantiopure and racemic mandelic acid crystal structures, has shown that there are two prevalent 1D packings of the phenyl groups (1D1 and 1D2 in Figure 3.10). However, if the number of enantiopure derivatives is increased it would be expected that more relationships would be observed. As such, it would appear that a key factor in the formation of crystal structures of mandelic acid derivatives, and thus the overall packing, lies in the way the phenyl groups are able to stack and pack together. These observations allow us to make speculation on the way in which mandelic acid derivatives come together in solution. Do the structures form catemeric chains which then hydrogen bond

together in solution, or do dimers form which then hydrogen bond together to form chains?

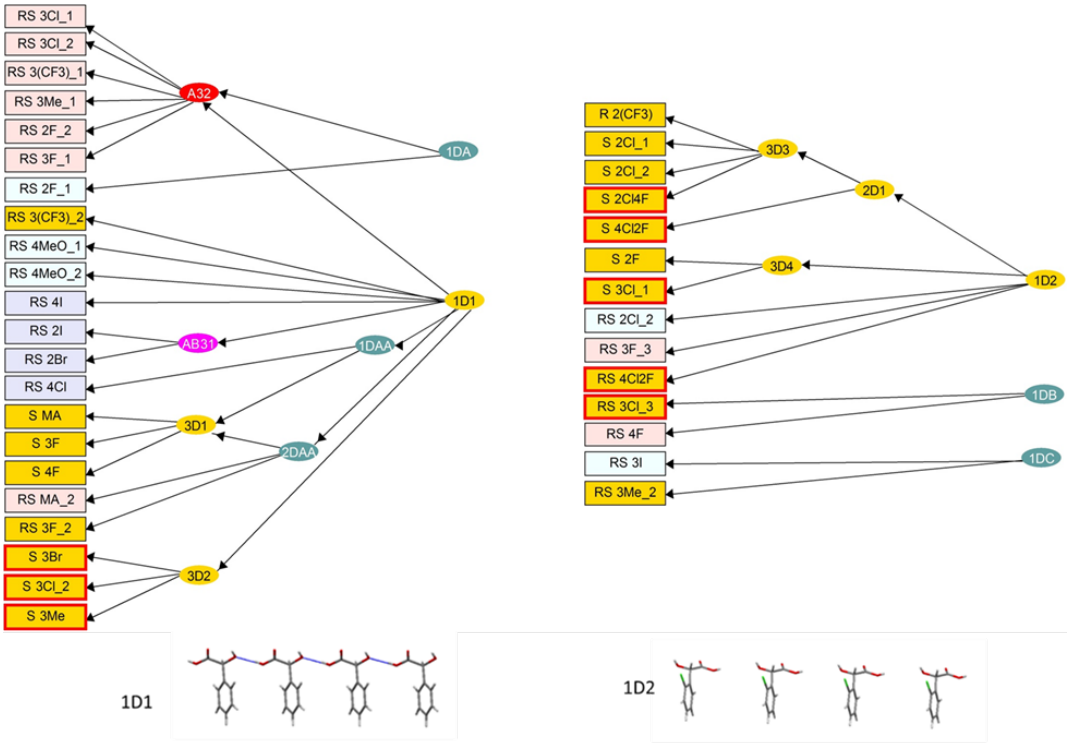


Figure 3-10: The common packing features of the mandelic acid crystal structures derived from pairwise XPac comparison of enantiopure and racemic crystals. The crystal structures are connected by any common 3D structures (i.e. isostructural ignoring the substituent), with 3Dx denoting the isotructural relationships for the enantiopure structures and A32 and AB31 being labels from the previous analysis of just the racemic structures.[23] These are related by any common 2-dimensional motifs, and finally to the depicted common 1D packing relationships. Packing motif 1D1 is hydrogen bonded, whilst 1D2 contains systematic absences. The new crystal structures are denoted by a red box. The background colour gives the hydrogen bonding motif, yellow for catemer, pink for $\mathbf{R}_2^2(\mathbf{8})$, light blue for $\mathbf{R}_2^2(\mathbf{10})$ and mauve for both. Only the substituents are given for each of the mandelic acid derivatives, with MA notation omitted for clarity.

3.3.5 Alternative lattice energy calculations

Through analysis of the low energy structures in the crystal energy landscape and the isostructurality of mandelic acid derivatives, the complexity of the structures that could be generated and leads to a number of questions. Is the model used to determine the lattice energies accurate enough? Are the experimental crystal structures kinetic artefacts and more stable thermodynamic structures not found yet?

These questions can be investigated by comparing the relative stability of the experimental crystal structures and a selection of the low energy hypothetical crystal structures (see Figure 3.9). The hypothetical crystal structures were

selected to try and represent the diverse range of structures that were predicted for 3ClMA. These include the lowest energy racemic and enantiopure structures and structures which have hydrogen bonding motifs that have been observed in experimental mandelic acid derivatives (see Figure 3.11).

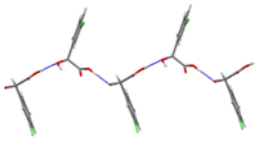
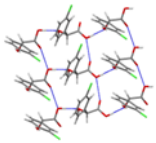
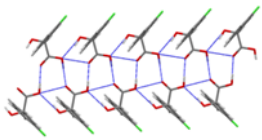
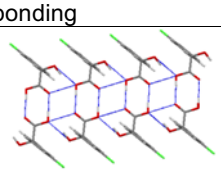
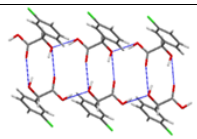
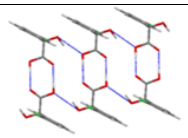
		
RS-B2 lowest energy structure, with only catemeric hydrogen bonding	RS-B1335 – enantiopure sheets can also pack in S structures.	S-A135 lowest enantiopure structure generated, but chain also in RS structures.
		
S-A379 $R_2^2(8)$ motif in enantiopure sheets which can also pack to give RS	RS-A50 $R_2^2(10)$ chain seen in RS-MA_2, RS-3F2MeMA, RS-3FMA_2	RS-B109 Chain containing $R_2^2(8)$ dimers seen in RS-4MeOMA_2.

Figure 3-11: The hydrogen boding motifs of the hypothetical crystal structures for 3ClMA, whose energies are compared in Figure 3.10. The letter gives the conformation (A or B) and the number from the energy ranking when the structures were generated by *CrystalPredictor*.

For each of these structures various lattice energy calculations were performed, either based on the wavefunction of the molecule or wavefunction of the crystal. By using different methods for evaluating the lattice energy it is possible to assess their sensitivity to the assumptions used within each approach. The crystal energy landscape in Figure 3.5 was generated using the wavefunction of the molecule. It assumes the charge distribution and conformational energy of the isolated molecule is a good approximation to that in the crystal. The polarisation of the charge density within the crystal was approximated by a polarisable continuum, but this is not reflective of the specific environment within each crystal. Additionally, while the intramolecular energy and electrostatic contribution are calculated for the specific molecular geometry observed within the crystal, all other terms for the intermolecular potential are represented by an exp-6 atom-atom model whose parameters have been fitted to room temperature crystal structures and so may not accurately reflect the crystal in question. DFT-D calculations are based on the wavefunction of the crystal and have the advantage of better structure optimisation and the automatic inclusion of induction and polarisation effects. However, DFT-D uses a poorer quality charge density and requires a dispersion correction to accurately describe the intermolecular forces within the crystal structure. In this

study two different dispersion corrections have been used and reveal how sensitive the crystal structures are to this small, but relevant energy contribution.

The representation of the crystal structures is fairly consistent between the the lattice energy calculations of *CrystalOptimizer* and using the G06 and TS dispersion corrections with DFT-D (see Figure 3.12). The TS DFT dispersion correction tended to cause the volume of the unit cell to increase relative to experimental determinations, indicating that the dispersion is underrepresented and is a common known artefact of this dispersion correction.

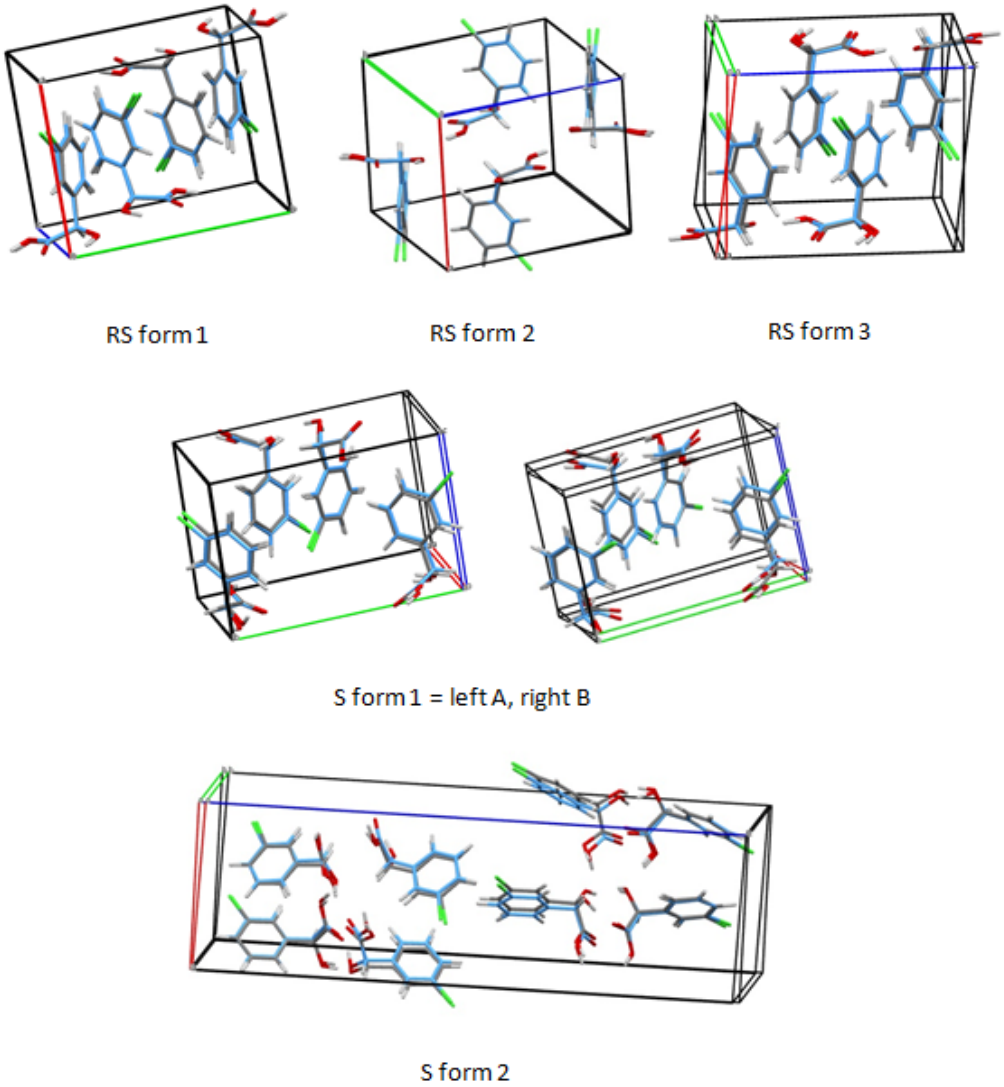


Figure 3-12: Overlays of the experimental crystal structures for 3CIMA (element colours) with those generated by *CrystalOptimizer* (blue). There is very little visual difference with the structures produced by the PBE+TS optimizations, though the G06 dispersion correction gives less dense cells.

Considering calculations based on the wavefunction of the molecule without including the effect of a PCM or free energies, the relative lattice energies of the known crystal structures of 3ClMA correlated with experimental heats of fusion, allowing for the fact that S-form 1 was treated as two extremes of the disordered complex. When the additional effects of a PCM and free energies are included, there is a small degree in re-ranking amongst the experimental and hypothetical crystal structures.

When the method of calculation is changed to the wavefunction of the crystal a more significant degree of re-ranking is observed (see Figure 3.13). The relative energies of the experimental crystal structures no longer reflect the rankings as given by the experimental heats of fusion. There is also significant re-ranking amongst all the structures studied when the two different dispersion corrections are used. It is important to note that different functionals have been used to evaluate the lattice energy for each of the methods considered – MP2 for *CrystalOptimizer* and PBE for DFT-D. This may result in some of the variation in ranking observed.

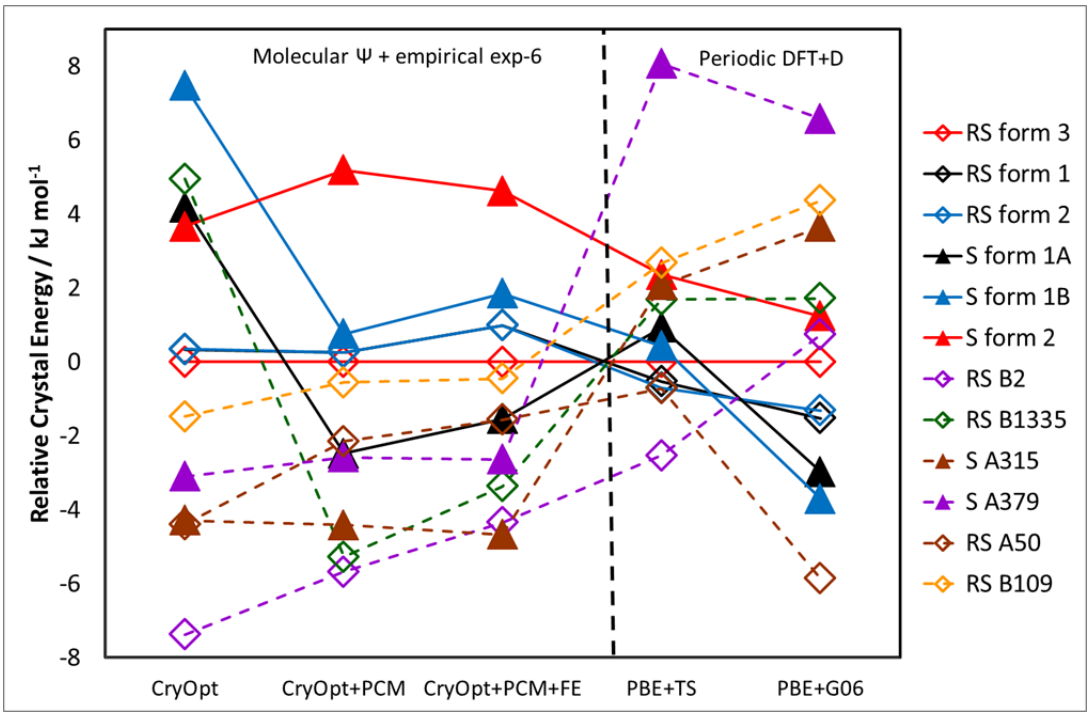


Figure 3-13: The energies of 3ClMA crystals relative to the most experimentally stable RS Form 3, calculated by different methods: CryOpt – isolated molecule wavefunctions using *CrystalOptimizer*; +PCM, molecular wave functions calculated in a polarizable continuum (as on Figure 3.3); +FE, with rigid body free energy estimates; DFT, periodic PBE density functional calculations with +TS (Tkatchenko-Scheffler) and +G06 (Grimme) dispersion corrections. Closed triangles denote enantiopure structures and open diamonds denote racemic structures, with RS forms 1 and 2 frequently superimposed. Observed structures are joined by solid lines, and hypothetical by dashed lines.

The variation in relative lattice energies across the different methods is consistent with the current challenges in obtaining accurate lattice energies.[24-26] The uncertainty in the dispersion component of the calculated lattice energies may account for structure RS-B2, the global minimum structure for many of the computational methods used, a highly dense structure with an unusual catemetic hydrogen bonding motif. However, there are other structures which are consistently calculated to be lower in energy than the experimental crystals that could be more plausible. Structure RS-A50 has the same hydrogen bonding motifs as the three experimentally determined racemic mandelic acid structures, based on the $R_2^2(10)$ dimer. Structures S-A379 and RS-B109 contain the hydrogen bonded $R_2^2(8)$ dimer which is commonly observed amongst other racemic mandelic acid derivatives and, while are not consistently lower in energy, are also competitive.[23] This suggests that there is a detailed and complicated balance between the various inter- and intramolecular forces, such as conformation, hydrogen bonding, phenyl π stacking, Cl---Cl halogen bonding etc) within the 3ClMA crystals. This delicate balance between the various forces at play explains why there is such a diverse range of thermodynamically competitive structures which are accessible for 3ClMA. It is plausible that there are alternative Z'=1 enantiopure and racemic structures that 3ClMA could form that are at least thermodynamically feasible, if not more stable than those structures which have already been observed experimentally.

The relationships between the mandelic acid derivatives as shown in the Hasse diagram could be used in conjunction with the crystal energy landscape of 3ClMA to design templating experiments to see if additional crystal structures of 3ClMA could be obtained. Amy Elis, from the University of Southampton performed several crystallisation experiments on 3ClMA as part of an investigation to try and identify any diffraction quality crystals or novel forms. A small selection of the low energy hypothetical crystal structures from crystal energy landscape of 3ClMA were given to Amy to be used in conjunction with Hasse diagrams in the hopes of finding a lead to design templating experiments. Unfortunately, none of the experiments performed at Southampton yielded any new polymorphs of 3ClMA.

3.3.6 Exploring the interactions of 3ClMA dimers

While the various lattice energy calculations have been unable to explain why the experimentally observed structures are found so high in energy in the crystal energy landscape, perhaps the association of dimers can help provide an explanation. The differences in hydrogen bonding capabilities and the association of molecules in solution has often been suggested, and occasionally demonstrated, to control the polymorph which is produced.[27,28]

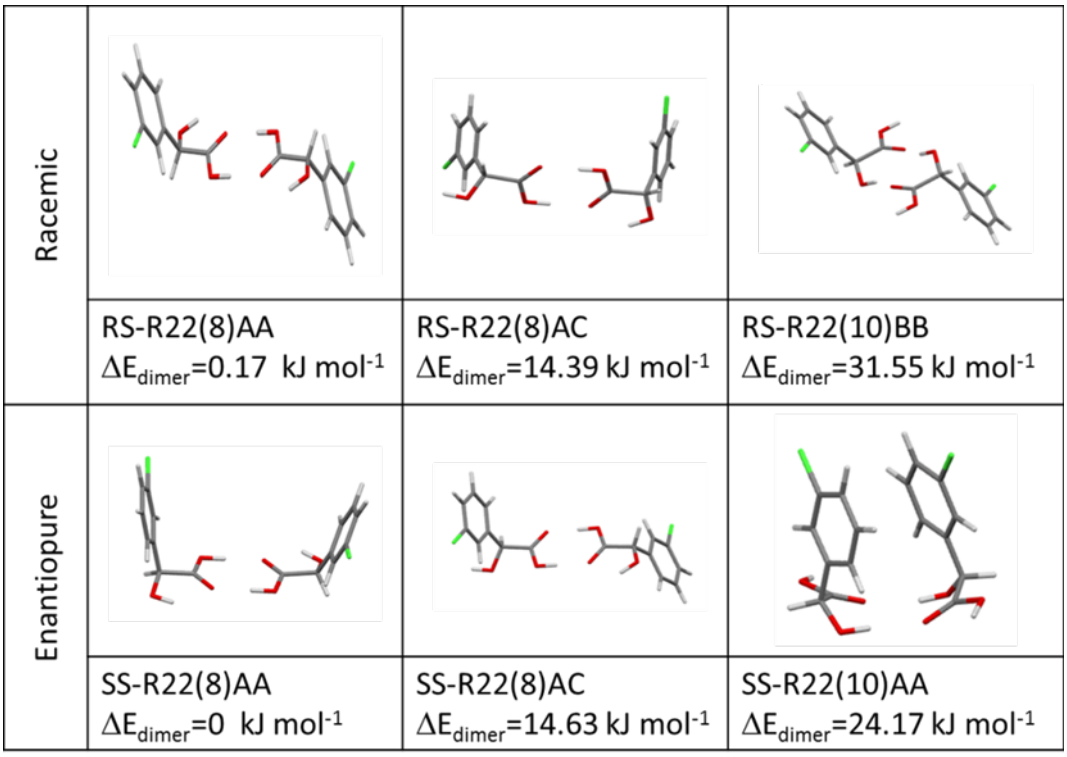


Figure 3-14: Low energy isolated dimer structures of 3ClMA, as optimized by counterpoise-corrected MP2 6-31G(d,p) calculations (See Table 3.3). The binding energies are given relative to the most stable dimer (SS-$\boldsymbol{R_2^2(8)}$), which has a binding energy of –56.81 kJ mol^{-1}. The chirality and graph set labels are followed by the two conformer labels, defined by the nearby isolated molecule conformation (Figure 3.4); A is the lowest energy conformation, B has the Cl on the opposite side which is only 0.07 kJ mol^{-1} higher for the isolated molecule, C and D have the two OH groups on the same side. These labels are also used for mandelic acid dimer types ignoring the Cl.

Quantum chemical calculations were performed on isolated dimers of 3ClMA. The dimers were selected such that they represented the full range of dimer interactions that could be possible amongst the four possible conformations of 3ClMA. These calculations showed that the $R_2^2(8)$ hydrogen bonding motif is more stable than the $R_2^2(10)$ for both the enantiopure and racemic dimers. However, it is the $R_2^2(10)$ motif which is more commonly observed amongst the observed crystal structures.

Structure	Optimized BSSE corrected energy / kJ mol^{-1}	BSSE correction / kJ mol^{-1}	Binding energy / kJ mol^{-1}	Binding energy from single point calculation using aug-cc-pVTz basis set / kJ mol^{-1}
RS-$R_2^2(8)$ AA	-5213466.38	19.82	-56.63	-66.8
RS-$R_2^2(10)$ BB	-5213434.99	15.42	-25.25	-39.7
RS-$R_2^2(8)$ AC	-5213452.16	20.35	-42.42	-55.2
S-$R_2^2(10)$ AA	-5213442.38	24.82	-32.64	-69.5
S-$R_2^2(8)$ AA	-5213466.55	19.82	-56.81	-67.0
S-$R_2^2(8)$ AC	-5213451.92	20.34	-42.18	-54.9
Isolated molecule A	-2606704.87			

Table 3-3: The MP2 6-31G(d.p) dimer energies of 3ClMA, relative to infinitely separated molecules in the most stable configuration A. A single point energy calculation on these dimers was performed using the MP2 wavefunction and aug-cc-pVTz basis set, with Turbomole.[29]

There is relatively little energy difference between the racemic and enantiopure dimeric interactions, suggesting that there is no preference over which dimer forms. This has also been observed for isolated mandelic acid dimers.[30] The distance between the phenyl rings is quite large in the dimers which explains why the energy difference between racemic and enantiopure dimeric interactions is so low. The only dimer which brings the phenyl rings close enough to exhibit a significant dispersion interaction is the SS-$R_2^2(10)$ dimer, which is sensitive to the treatment of the electron correlation.

It is possible that the any hydrogen bonded dimers formed in solution could be used as a building block for the crystal structure. FTIR studies[30] have shown that there is no difference in the solutions of racemic and enantiopure crystals as well as no evidence for hydrogen bonding of dimers in select solvents. These solvents did not include water and toluene, the two solvents used by the MPI in their 3ClMA studies. Collaborators at the MPI performed molecular dynamics simulations[5] on racemic and enantiopure dimers of mandelic acid in aqueous solution and toluene. Simulations of mandelic acids in water show that the $R_2^2(8)$ and $R_2^2(10)$ dimers readily break up into perfectly solvated separate

monomers while simulations in toluene show fluctuations between the $R_2^2(8)$ and $R_2^2(10)$ dimers with no real preference for either. This implies that mandelic acid can readily change hydrogen bonding during crystallisation and that the hydrogen bonding in solution is not present long enough to have a significant effect on the hydrogen bonding observed in the solid state.

Overall, the interaction energies of 3ClMA dimers and the solution behaviour does not appear to account for the observed hydrogen bonding motifs that have been observed amongst racemic 3ClMA crystal structures, nor why enantiopure structure appear to favour catemeric hydrogen bonded motifs.

3.4 Conclusions

3ClMA is a small yet highly flexible molecule. This flexibility leads to four distinct, energetically plausible conformations, each with a wide range of torsion angles, which can be adopted. Crystal structure prediction has allowed the thermodynamically plausible structures to be investigated. A wide range of crystal packings have been observed for 3ClMA, with many hypothetical crystal structures shown to be thermodynamically competitive to the experimentally observed structures.

All experimentally known structures of 3ClMA within the limitations of the search (i.e. Z'=1) were successfully found within the CSP search, matching the ranking determined from experimental heats of formation, although at relatively high energies and low densities with many structures being observed at lower energies. The crystal energy landscape is particularly dense. Analysis of low energy structures shows that 3ClMA in different conformational regions can form practically identical structures with relatively little energy penalty. This suggests that 3ClMA could form disordered structures relatively easily, as with S form 1, as well as indicating potential difficulty in forming crystals due to the flexible nature of the molecule.

The different lattice energy calculation performed on the test set of experimental and hypothetical crystal structures highlights how sensitive these types of

calculation are to the method used. In order to generate accurate crystal energy landscapes for molecules lattice energy calculations need to be accurate, particularly when energy differences between polymorphs can be as small as 1-2 kJ mol^{-1}.[31] Although it has been shown that lattice energies can be calculated to accuracies of less than 1kJ mol^{-1} for benzene,[24] these type of calculation are far from routine.[26] Even so there are a number of structures which are consistently lower in energy across all methods which suggests that there are alternate structures which have been generated by CSP which are competitive with the currently experimentally known structures.

Analysis of the packings and hydrogen bonding motifs of enantiopure and racemic crystal structures shows that 3ClMA is capable of forming a wide range of hydrogen bonding motifs and packings. Contrasting the racemic and enantiopure crystal structures and examining the dimeric interactions and solution behaviour shows that phenyl packing is likely to play the largest role in the formation of crystal structures of mandelic acids. While some screening has been performed to find further polymorphs of 3ClMA, it has not been extensive. Analysis of the structural relationships of known mandelic acid derivatives combined with the crystal energy landscape of 3ClMA could help experimentalists to design experiments, specifically templating experiments, to identify potentially lower energy structures of 3ClMA.

The crystallisation of 3ClMA, and organic crystals in general, is a fine balance between the kinetic effects of nucleation and growth and the most thermodynamically stable structure that is available. To design an effective separation method, it is imperative that the exact crystal structure is known and so experiments can be designed with the specific properties of that structure in mind. Understanding these kinetic effects can aid in the design of chiral separation processes by enabling experiments to be performed which are tailored to the specific conditions in which a crystal structure is favoured.

3.5 References

(1) Zhang, Y.; Ray, A.; Rohani, S. Measurement and prediction of phase diagrams of the enantiomeric 3-chloromandelic acid system. *Chemical Engineering Science* **2009**, *64*, 192-197.

(2) Lorenz, H.; Seidel-Morgenstern, A. Processes To Separate Enantiomers. *Angewandte Chemie International Edition* **2014**, *53*, 1218-1250.

(3) von Langermann, J.; Temmel, E.; Seidel-Morgenstern, A.; Lorenz, H. Solid Phase Behavior in the Chiral Systems of Various 2-Hydroxy-2-phenylacetic Acid (Mandelic Acid) Derivatives. *Journal of Chemical & Engineering Data* **2015**, *60*, 721-728.

(4) Seidel-Morgenstern, A. INTENANT Report Summary. http://cordis.europa.eu/result/report/rcn/45395_en.html **2010**.

(5) Hylton, R. K.; Tizzard, G. J.; Threlfall, T. L.; Ellis, A. L.; Coles, S. J.; Seaton, C. C.; Schulze, E.; Lorenz, H.; Seidel-Morgenstern, A.; Stein, M.; Price, S. L. Are the Crystal Structures of Enantiopure and Racemic Mandelic Acids Determined by Kinetics or Thermodynamics? *Journal of the American Chemical Society* **2015**, *137*, 11095-11104.

(6) Le Minh, T.; Langermann, J. v.; Lorenz, H.; Seidel-Morgenstern, A. Enantiomeric 3-Chloromandelic Acid System: Binary Melting Point Phase Diagram, Ternary Solubility Phase Diagrams and Polymorphism. *Journal of Pharmaceutical Sciences* **2010**, *99*, 4084-4095.

(7) Coles, S. J.; Threlfall, T. L.; Tizzard, G. J. The Same but Different: Isostructural Polymorphs and the Case of 3-Chloromandelic Acid. *Crystal Growth & Design* **2014**, *14*, 1623-1628.

(8) Habgood, M.; Grau-Crespo, R.; Price, S. L. Substitutional and orientational disorder in organic crystals: a symmetry-adapted ensemble model. *Physical Chemistry Chemical Physics* **2011**, *13*, 9590-9600.

(9) Habgood, M. Form II Caffeine: A Case Study for Confirming and Predicting Disorder in Organic Crystals. *Crystal Growth & Design* **2011**, *11*, 3600-3608.

(10) Gelbrich, T.; Hursthouse, M. B. A versatile procedure for the identification, description and quantification of structural similarity in molecular crystals. *CrystEngComm* **2005**, *7*, 324-336.

(11) Tizzard, G. J.; Coles, S. J.; Leung, K.; Sarson, J.; Threlfall, T. In *Tilte*, 2013.

(12) Karamertzanis, P. G.; Pantelides, C. C. Ab initio crystal structure prediction - I. Rigid molecules. *Journal of Computational Chemistry* **2005**, *26*, 304-324.

(13) Karamertzanis, P. G.; Pantelides, C. C. Ab initio crystal structure prediction. II. Flexible molecules. *Molecular Physics* **2007**, *105*, 273-291.

(14) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A.: Gaussian 03. Gaussian Inc.: Wallingford CT, 2004.

(15) Coombes, D. S.; Price, S. L.; Willock, D. J.; Leslie, M. Role of Electrostatic Interactions in Determining the Crystal Structures of Polar Organic Molecules. A Distributed Multipole Study. *Journal of Physical Chemistry* **1996**, *100*, 7352-7360.

(16) Kazantsev, A. V.; Karamertzanis, P. G.; Adjiman, C. S.; Pantelides, C. C.: CrystalOptimizer. An efficient Algorithm for Lattice Energy Minimisation of Organic Crystal using Isolated-Molecule Quantum Mechanical Calculations. In *Molecular System Engineering*; Adjiman, C. S., Galindo, A., Eds.; Process Systems Engineering; WILEY-VCH Verlag GmbH & Co.: Weinheim, 2010; Vol. 6; pp 1-42.

(17) Stone, A. J.: GDMA: A Program for Performing Distributed Multipole Analysis of Wave Functions Calculated Using the Gaussian Program System. 2.2 ed.; University of Cambridge: Cambridge, United Kingdom, 2010.

(18) Price, S. L.; Leslie, M.; Welch, G. W. A.; Habgood, M.; Price, L. S.; Karamertzanis, P. G.; Day, G. M. Modelling Organic Crystal Structures using Distributed Multipole and Polarizability-Based Model Intermolecular Potentials. *Physical Chemistry Chemical Physics* **2010**, *12*, 8478-8490.

(19) Tkatchenko, A.; Scheffler, M. Accurate Molecular Van Der Waals Interactions from Ground-State Electron Density and Free-Atom Reference Data. *Physical Review Letters* **2009**, *102*, 073005.

(20) Grimme, S. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *Journal of Computational Chemistry* **2006**, *27*, 1787-1799.

(21) Gavezzotti, A.; Rizzato, S. Are Racemic Crystals Favored over Homochiral Crystals by Higher Stability or by Kinetics? Insights from Comparative Studies of Crystalline Stereoisomers. *Journal of Organic Chemistry* **2014**, *79*, 4809-4816.

(22) Steed, K. M.; Steed, J. W. Packing Problems: High Z' Crystal Structures and Their Relationship to Cocrystals, Inclusion Compounds, and Polymorphism. *Chemical Reviews* **2015**, *115*, 2895-2933.

(23) Coles, S. J.; Ellis, A. L.; Leung, K.; Sarson, J.; Threlfall, T. L.; Tizzard, G. J. Relationships between the racemic structures of substituted mandelic acids containing 8-and 10-membered hydrogen bonded dimer rings. *Crystengcomm* **2014**, *16*, 10816-10823.

(24) Yang, J.; Hu, W.; Usvyat, D.; Matthews, D.; Schutz, M.; Chan, H. Ab initio determination of the lattice energy in crystalline benzene to sub-kilojoule per mole accuracy. *Science* **2014**, *345*, 640-643.

(25) Reilly, A. M.; Tkatchenko, A. Role of Dispersion Interactions in the Polymorphism and Entropic Stabilization of the Aspirin Crystal. *Physical Review Letters* **2014**, *113*.

(26) Reilly, A. M.; Cooper, R. I.; Adjiman, C. S.; Bhattacharya, S.; Boese, A. D.; Brandenburg, J. G.; Bygrave, P. J.; Bylsma, R.; Campbell, J. E.; Car, R.; Case, D. H.; Chadha, R.; Cole, J. C.; Cosburn, K.; Cuppen, H. M.; Curtis, F.; Day, G. M.; DiStasio Jr, R. A.; Dzyabchenko, A.; van Eijck, B. P.; Elking, D. M.; van den Ende, J. A.; Facelli, J. C.; Ferraro, M. B.; Fusti-Molnar,

L.; Gatsiou, C.-A.; Gee, T. S.; de Gelder, R.; Ghiringhelli, L. M.; Goto, H.; Grimme, S.; Guo, R.; Hofmann, D. W. M.; Hoja, J.; Hylton, R. K.; Iuzzolino, L.; Jankiewicz, W.; de Jong, D. T.; Kendrick, J.; de Klerk, N. J. J.; Ko, H.-Y.; Kuleshova, L. N.; Li, X.; Lohani, S.; Leusen, F. J. J.; Lund, A. M.; Lv, J.; Ma, Y.; Marom, N.; Masunov, A. E.; McCabe, P.; McMahon, D. P.; Meekes, H.; Metz, M. P.; Misquitta, A. J.; Mohamed, S.; Monserrat, B.; Needs, R. J.; Neumann, M. A.; Nyman, J.; Obata, S.; Oberhofer, H.; Oganov, A. R.; Orendt, A. M.; Pagola, G. I.; Pantelides, C. C.; Pickard, C. J.; Podeszwa, R.; Price, L. S.; Price, S. L.; Pulido, A.; Read, M. G.; Reuter, K.; Schneider, E.; Schober, C.; Shields, G. P.; Singh, P.; Sugden, I. J.; Szalewicz, K.; Taylor, C. R.; Tkatchenko, A.; Tuckerman, M. E.; Vacarro, F.; Vasileiadis, M.; Vazquez-Mayagoitia, A.; Vogt, L.; Wang, Y.; Watson, R. E.; de Wijs, G. A.; Yang, J.; Zhu, Q.; Groom, C. R. Report on the sixth blind test of organic crystal structure prediction methods. *Acta Crystallographica Section B* **2016**, *72*, 439-459.

(27) Kulkarni, S. A.; McGarrity, E. S.; Meekes, H.; ter Horst, J. H. Isonicotinamide self-association: the link between solvent and polymorph nucleation. *Chemical Communications* **2012**, *48*, 4983-4985.

(28) Khamar, D.; Zeglinski, J.; Mealey, D.; Rasmuson, A. C. Investigating the Role of Solvent-Solute Interaction in Crystal Nucleation of Salicylic Acid from Organic Solvents. *Journal of the American Chemical Society* **2014**, *136*, 11664-11673.

(29) Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. Electronic structure calculations on workstation computers: The program system turbomole. *Chemical Physics Letters* **1989**, *162*, 165-169.

(30) Davey, R. J.; Dent, G.; Mughal, R. K.; Parveen, S. Concerning the relationship between structural and growth synthons in crystal nucleation: Solution and crystal chemistry of carboxylic acids as revealed through IR spectroscopy. *Crystal Growth & Design* **2006**, *6*, 1788-1796.

(31) Price, S. L. Why don't we find more polymorphs? *Acta Crystallographica Section B: Structural Science, Crystal Engineering and Materials* **2013**, *69*, 313-328.

3.6 Appendix

3.6.1 Low energy structures of 3ClMA produced by CSP

Appendix Figure 3.6.1: The lowest 100 structures of the crystal energy landscape of 3ClMA. The lattice energy minima corresponding to the experimental structures using the same computational model are highlighted in green. The hydrogen bond motifs are defined in Table 3.2

Label	Space Group	Z	a /Å	b /Å	c /Å	α /°	β /°	γ /°	Density /g cm^{-3}	Intermolecular Lattice Energy /kJmol^{-1}	Repulsion Dispersion Energy /kJmol^{-1}	Lattice Energy /kJ mol^{-1}	Zero-point energy /kJ mol^{-1}	Hydrogen bonding motif
B2	P21/c	4	5.68	16.00	10.37	90	56.8	90	1.573	-128.05	-58.94	-126.49	3.12	1
B1335	P21/c	4	15.07	4.90	10.97	90	79.2	90	1.556	-141.96	-53.97	-126.08	3.29	2
A315	P21	2	4.97	4.87	16.32	90	85.5	90	1.575	-128.57	-61.81	-125.21	2.91	3
B106	Pbca	8	10.94	29.58	4.92	90	90.0	90	1.559	-140.92	-54.11	-124.87	3.25	2
B38	P21/c	4	8.48	8.88	10.68	90	88.7	90	1.541	-131.66	-53.41	-124.75	3.25	4
A2	P212121	4	5.00	4.84	32.57	90	90.0	90	1.573	-128.15	-61.97	-124.66	2.72	3
A3	P21/c	4	5.01	4.83	32.78	90	97.2	90	1.573	-127.87	-61.8	-124.43	2.74	3
B101	C2/c	4	33.74	4.90	10.96	90	62.4	90	1.542	-139.58	-53.27	-124.17	3.24	2
A6	Pca21	4	32.84	4.95	4.87	90	90.0	90	1.568	-127.61	-61.32	-124.14	2.72	3
A5	P21/c	4	4.94	4.87	32.89	90	86.0	90	1.569	-127.48	-62.01	-124.10	2.71	3
D19	P21/c	4	15.33	5.40	9.64	90	80.0	90	1.579	-135.63	-55.87	-123.80	3.03	2
A21	P41	8	4.90	4.90	65.80	90	90.0	90	1.567	-127.17	-123.71	-123.64	2.63	3

A7	Pna21	4	4.98	32.96	4.83	90	90.0	90	1.562	-127.14	-60.81	-123.60	2.71	3
A379	P212121	8	10.21	31.77	4.89	90	90.0	90	1.563	-126.85	-115.15	-123.39	3.03	5
A8	C2	2	33.37	4.83	4.99	90	96.9	90	1.554	-126.95	-60.94	-123.37	2.83	3
B49	P21/c	4	15.23	5.15	10.01	90	86.5	90	1.580	-126.44	-60.39	-123.36	3.13	5
ES1A	P21	4	8.06	12.18	8.41	90	93.0	90	1.503	-131.11	-106.98	-123.28	3.09	6
A1373	P21/c	4	16.62	5.12	10.83	90	62.8	90	1.512	-138.94	-47.44	-123.24	3.34	7
A1313	Pn21a	8	10.90	29.65	5.06	90	90.0	90	1.518	-138.9	-95.59	-123.23	3.29	7
D604	C2/c	4	32.10	5.41	9.63	90	73.0	90	1.549	-134.62	-54.04	-123.00	3.04	2
A50	P21/n	4	12.04	5.87	11.41	90	91.9	90	1.538	-124.45	-52.72	-122.95	3.02	8
B162	P21/c	4	15.78	4.93	10.94	90	109.2	90	1.541	-138.43	-52.24	-122.90	3.28	2
C2	P21/c	4	5.35	8.77	17.29	90	77.4	90	1.565	-129.53	-55.92	-122.89	2.96	2
D109	P21/c	4	15.49	5.40	9.62	90	101.4	90	1.573	-134.33	-54.21	-122.88	3.07	2
C45	P21/c	4	5.86	16.22	9.87	90	56.9	90	1.579	-132.59	-57.11	-122.87	3.19	2
A56	P21/c	4	8.16	5.83	16.83	90	85.7	90	1.552	-124.1	-54.37	-122.72	3.03	9
D872	P21	8	5.40	30.37	9.63	90	89.8	90	1.570	-133.7	-215.7	-122.26	2.99	2
D53	Pbca	8	10.64	9.47	15.91	90	90.0	90	1.547	-132.83	-52.94	-122.24	3.24	2
B159	Pna21	8	4.91	30.10	10.94	90	90.0	90	1.532	-137.69	-104.35	-122.22	3.18	10
D1346	P21	8	9.66	30.46	5.39	90	90.0	90	1.562	-133.66	-214.35	-122.11	2.97	2
B1100	Pn21a	8	10.95	30.31	4.91	90	90.0	90	1.521	-137.67	-102.38	-122.11	3.11	2
A24	P21/c	4	4.97	4.83	33.12	90	94.0	90	1.563	-125.37	-60.29	-122.08	2.67	5
A55	P21/c	4	16.13	4.86	10.27	90	79.5	90	1.567	-124.93	-57.87	-122.03	3.03	5
B12	P21/c	4	9.03	9.93	9.82	90	113.7	90	1.538	-128.32	-51.52	-122.03	3.34	11
C11	P21/c	4	16.00	5.31	9.61	90	78.7	90	1.548	-127.91	-51.95	-121.98	2.94	2

A1740	P21/b	8	16.86	5.80	16.86	90.0	90.0	90	1.503	-123.64	-96.95	-121.94	2.97	9
A93	P21/c	4	16.13	4.89	10.31	90	82.3	90	1.539	-125.66	-56.1	-121.93	2.98	5
B6	Pbca	8	15.93	10.71	9.49	90	90.0	90	1.531	-126.24	-52.33	-121.90	3.25	11
D6	P21/n	4	8.43	16.68	5.58	90	87.2	90	1.581	-132.05	-58.11	-121.89	3.12	2
A46	P21	2	8.76	9.07	5.05	90	82.3	90	1.558	-125.52	-56.14	-121.86	3.17	2
A526	Pbca	8	10.89	30.33	5.05	90	90.0	90	1.487	-137.04	-45.44	-121.79	3.26	7
C118	P212121	8	9.40	32.40	5.30	90	90.0	90	1.536	-127.94	-99.57	-121.78	2.92	2
C1034	P21/c	4	9.02	5.57	16.67	90	79.1	90	1.505	-136.23	-50.32	-121.74	3.24	12
C40	P21/c	4	16.07	5.31	9.50	90	99.4	90	1.549	-127.82	-51.76	-121.71	3	2
C569	Pc	8	5.30	9.57	31.73	90	84.1	90	1.547	-127.91	-200.05	-121.62	2.95	2
A85	P21/c	4	8.60	5.87	16.07	90	89.3	90	1.528	-122.76	-53.58	-121.52	2.9	9
B573	Pbca	8	8.56	10.91	17.03	90	90.0	90	1.560	-130.81	-57.6	-121.42	3.15	4
B109	P21/c	4	16.42	5.19	9.85	90	70.0	90	1.571	-124.78	-58.3	-121.35	3.14	13
A521	P-1	2	5.00	18.39	4.79	93.74	88.8	64.34	1.566	-124.76	-61.47	-121.32	2.85	5
C35	Pca21	4	10.50	8.20	9.42	90	90.0	90	1.528	-127.84	-48.67	-121.31	3.24	2
C14	P21/c	4	16.38	5.30	9.28	90	85.0	90	1.543	-127.85	-50.8	-121.30	2.95	2
A16	P21/c	4	4.80	5.00	33.14	90	86.3	90	1.562	-124.92	-60.81	-121.26	2.66	5
A91	P21/c	4	16.06	4.88	10.04	90	96.4	90	1.586	-124.85	-65.43	-121.24	3	2
A1	P21/c	4	9.46	9.35	9.76	90	65.5	90	1.576	-123.58	-63.89	-121.23	3.15	14
B547	Pna21	8	30.71	5.14	10.02	90	90.0	90	1.568	-124.14	-117.41	-121.22	3.05	5
A104	P21	4	4.93	33.33	4.83	90	88.9	90	1.561	-124.44	-121.24	-121.09	2.65	5
D63	P21/c	4	15.45	5.40	9.62	90	98.4	90	1.561	-132.96	-54.17	-121.03	3.03	2
C15	C2/c	4	31.85	5.34	9.64	90	77.6	90	1.547	-127.7	-52.84	-120.96	3.03	2

C111	P-1	2	14.90	5.34	7.26	47.40	103.5	101.99	1.502	-136.32	-47.47	-120.90	2.96	15
B3	P21/c	4	9.53	9.31	11.21	90	124.7	90	1.516	-122.31	-53.85	-120.89	3.06	4
A26	P21/c	4	4.79	33.40	4.98	90	90.8	90	1.555	-124.42	-60.12	-120.87	2.65	5
A605	Pbcn	8	5.10	10.84	30.45	90	90.0	90	1.473	-136.61	-44.94	-120.87	3.23	7
ERS3	P21/c	4	8.71	8.89	10.86	90	95.9	90	1.483	-125.52	-48.41	-120.81	3.04	4
C1978	P-1	2	5.41	5.34	17.95	115.99	110.2	96.35	1.492	-134.85	-46.26	-120.73	2.99	15
C256	P21/c	4	14.10	5.88	9.73	90	76.3	90	1.580	-142.68	-58.98	-120.73	3.31	16
B7	P21/c	4	8.69	8.89	10.89	90	82.8	90	1.484	-125.26	-48.2	-120.71	3.04	4
C109	Pbca	8	27.57	5.95	9.67	90	90.0	90	1.561	-142.46	-57.04	-120.70	3.16	17
B157	P21/c	4	5.73	16.40	9.90	90	56.0	90	1.607	-129.49	-61.46	-120.70	3.09	1
A522	P-1	2	18.41	4.93	4.85	89.08	73.7	70.57	1.560	-123.79	-60.81	-120.70	2.84	5
C347	C2/c	4	7.26	7.85	28.95	90	89.6	90	1.503	-134.93	-47.57	-120.59	2.84	18
A23	Cc	4	6.99	6.83	33.13	90	86.7	90	1.569	-123.72	-123.25	-120.58	2.7	5
A54	P21/c	4	16.49	4.81	10.28	90	100.9	90	1.550	-123.93	-55.33	-120.57	3	3
B50	P21/c	4	8.54	10.61	9.27	90	87.2	90	1.477	-124.82	-47.6	-120.55	3.02	4
ERS1	P21/c	4	8.54	10.61	9.27	90	92.9	90	1.477	-124.79	-47.57	-120.54	3.02	4
ERS2	P-1	4	8.54	9.27	10.61	89.98	90.0	92.87	1.477	-124.74	-95.02	-120.54	3.02	4
A34	P2/c	4	4.99	4.79	33.76	90	82.7	90	1.547	-124.03	-59.65	-120.45	2.62	5
A120	P21/c	4	8.64	16.56	5.79	90	80.8	90	1.515	-123.2	-51.92	-120.44	2.89	9
A288	Pbcn	8	5.85	16.85	16.64	90	90.0	90	1.510	-122.21	-50.66	-120.43	2.93	8
A40	Cc	4	6.83	6.97	33.36	90	90.0	90	1.560	-123.75	-122.02	-120.42	2.69	5
A97	P21/c	4	16.38	4.86	10.03	90	79.0	90	1.582	-124.05	-64.83	-120.38	3.02	2
C1143	A2	4	29.13	7.27	7.84	90	90.0	82.77	1.505	-134.83	-95.62	-120.35	2.81	18

A153	P21/c	4	14.76	4.98	11.41	90	76.3	90	1.521	-135.6	-48.96	-120.33	3.13	2
C155	P21	4	5.36	28.82	5.31	90	86.7	90	1.516	-136.38	-97.87	-120.33	2.71	19
A31	P21/c	4	15.59	5.02	10.41	90	75.6	90	1.572	-123.51	-64.47	-120.32	2.96	2
C164	Pca21	8	5.30	9.61	31.65	90	90.0	90	1.537	-126.56	-99.2	-120.31	2.9	2
C408	P21/c	4	5.24	5.36	29.99	90	105.0	90	1.524	-137.57	-49.11	-120.28	2.73	2
A281	P21/c	8	8.84	5.80	32.38	90	94.0	90	1.497	-122.04	-94.27	-120.25	2.86	9
A35	Pbca	8	10.39	30.19	5.02	90	90.0	90	1.575	-123.3	-64.82	-120.17	2.9	2
C713	P21	8	9.39	32.62	5.29	90	89.8	90	1.531	-126.69	-191.1	-120.17	2.88	2
D1607	C2/c	4	36.92	5.42	9.59	90	57.4	90	1.532	-132.15	-52.51	-120.10	3.03	2
C113	P21/c	8	9.56	27.70	6.01	90	90.0	90	1.558	-141.42	-111.11	-120.07	3.15	17
ES1B	P21	4	8.28	12.08	8.42	90	99.8	90	1.495	-129.7	-107.16	-120.05	3.11	6
C126	P21212	8	5.31	9.54	32.42	90	90.0	90	1.510	-126.13	-93.36	-120.02	2.93	20
A663	Pn21a	8	10.07	32.27	4.83	90	90.0	90	1.578	-123.74	-129.48	-120.02	2.93	2
A99	P21/c	4	8.56	16.80	5.84	90	80.18	90	1.499	-121.39	-48.77	-120.02	2.92	21
A81	P-1	2	5.84	8.012	8.58	92.78	84.38	90.41	1.552	-121.60	-57.93	-120.01	3.04	8
B245	P21/c	4	10.77	9.05	10.89	90	51.09	90	1.50	-124.80	-50.33	-112.00	3.14	11
C433	P21	2	16.47	4.91	4.86	90	86.00	90	1.58	-133.59	-62.60	-119.96	2.96	22
A199	Pna21	8	4.87	32.68	10.25	90	90	90	1.52	-123.32	-109.41	-119.90	3.17	5
A174	C2/c	4	35.63	4.99	11.38	92	125.87	90	1.51	-134.83	-47.55	-119.90	3.06	23
ES2	P21	8	9.78	5.48	31.16	90	93.71	90	1.49	-117.8	-203.84	-115.62	2.92	9

4 Crystal Structure Prediction of Lactides

4.1 Introduction

The crystal structure prediction study of 3-chloromandelic acid has shown just how complex the crystallisation of a supposedly simple chiral molecule can be. Is the complexity of the crystal energy landscape of 3-ClMA purely down to the flexibility of the molecule or an inherent property of chiral molecules?

The lactides, 3,6-dimethyl-1,4-dioxane-2,5-dione, are internal cyclic esters of lactic acid. Lactides are chiral molecules with two chiral centres, therefore there are two different diastereoisomers which are possible; a R,R (S,S) form, where the methyl groups are both on the same side of the ring and a R,S form (also referred to as *meso*-lactide), where the methyl groups are on the opposite side of the ring. For this study I will be concentrating on the R,R (S,S) form and not the *meso* form, and so any reference to lactide will not concern the *meso* form. Due to the biological nature of the lactides, the diastereoisomers are more commonly referred to by the D,L nomenclature, with R,R- lactide corresponding to D,D-lactide. Throughout this chapter I will be referring to the lactides by the R and S nomenclature so more direct comparisons can be made with my other systems of interest.

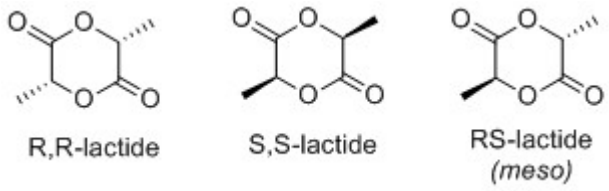


Figure 4-1: The diastereoisomers of the lactide molecule

The lactide[1,2] is used in the preparation of polylactide, a biologically decomposable polymer with many complex properties that make it ideal for medical uses.[3] The mechanical properties of polylactide are dependent on the proportion of R- and S- lactide within the structure. Polylactide comprising of the S-enantiomer only will give rise to a semi-crystalline material, while the copolymers of polylactide with a high proportion of R-monomer exhibit lower melting points and a far slower crystallisation behaviour. Thus, efficient separation of the enantiomers is required to obtain a polylactide with the desired properties.[4,5]

Due to the importance of the polymerisation of the lactide molecule, the main focus of research has been on the structure of the final polymer rather than on the monomer units. The monomer unit is widely produced in the fermentation of sugar sources, such as corn, sugarcane and tapioca. As such, crude lactide usually contains many impurities, including acids, alcohols, esters and traces of sugars and nutrients.[6] Thus, purification of the initial monomer units is also of great importance. Although the possible polymorphism of the monomer units have not been the prime focus of research on the lactide system, crystal structures of both the racemic and enantiopure compounds have been structurally characterised.[1,2]

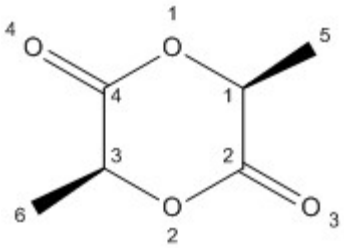


Figure 4-2: The S,S-lactide molecule, including atom labelling.

The experimental racemic crystal structure, BICVIS[1], is a $P2_1/c$, Z'=1 crystal structure with 4 molecules in the unit cell. The structure was obtained from a single crystal diffraction using crystals that were obtained by controlled recrystallization from dry ether and had dimensions that were greater than 0.4 mm.[1] The structure has also been solved independently via sublimation crystallisation under vacuum.[2]

Figure 4-3: Left: Racemic RR,SS-lactide, BICVIS. Right: Enantiopure R,R (S,S)-lactide, NAHNOZ

The experimental enantiopure structure, NAHNOZ[2], is a more complicated, $P2_12_12_1$, Z'=3 structure with 12 molecules in the asymmetric unit, determined via sublimation of the purified product under vacuum. The crystal structure of NAHNOZ was determined as part of an investigation into CH···O hydrogen bonds and their observance in organic crystal structures. CH···O hydrogen bonds are in the energy range of 0.5 – 1 kJ mol^{-1}, which is much weaker than 'traditional' hydrogen bonds which have an energy of ~4.7 kJ mol^{-1}. Even

though CH···O bonds are relatively weak they could be important in determining the final crystal structure, particularly when there are no other strong directional forces within the crystals.

The lactide system were studied intensively by my collaborators at the MPI, Magdeburg, looking at its crystallisation properties both dry (melting) and in solution (solubilities and eutectic composition) for several solvent systems.[7,8] It has also been chosen as an example system to study the thermodynamic properties of chiral systems (See Chapter 6) During these studies on solid liquid equilibria (SLE) and determination of heat capacities at low temperatures, no polymorphs of solvents were discovered. However, this was by no means an exhaustive polymorph search, such as those done in industry.

Monomeric R-lactide is a small molecule and, due to its cyclic nature, it can be considered to be rigid. The properties of lactide means this molecule provides an excellent basis for crystal structure prediction and in the design of chiral separation processes as there are no known polymorphs or solvates as well as a large differences in the enthalpies of formation (4.83 kJ mol^{-1}) between the racemic and enantiopure forms.[9]

A previous study investigating the reproducibility of the geometry, IR, NMR and electronic spectra of the molecule using DFT- methods[10] has been performed. This study has shown that DTD-D can be used to investigate the geometries and frequencies of the lactide molecules to a good degree of accuracy. This study focused only on the experimentally known forms of lactide and there has been little investigation into any potential polymorphism of lactide. The lattice energies and thermodynamic properties, both absolute and energy differences, of experimental lactide crystal structures have been determined via Ψ_{mol} methods, using CrystalOptimizer, and Ψ_{crys} methods using CASTEP, as part of the study on thermodynamic cycles (see Chapter 6). The absolute lattice energies obtained with *CrystalOptimizer* are of a similar magnitude to those evaluated with the DFT-D methods. The DFT-D calculations gave a lattice energy approximately 4 kJ mol^{-1} lower, for both the RS and S structures, than obtained with *CrystalOptimizer*. The discrepancies in absolute lattice energies

between the two methods is relatively large but within the variation seen across crystal structure prediction methods.[11]

4.2 Methods

A comparison of different basis sets and functionals for the molecular calculations to determine the molecular structure and distributed multipoles was performed to determine which would reproduce the experimental crystal structures the most accurately and economically. The MP2, PBE1PBE, PBE, B3LYP wavefunctions with the 6-31G(d,p) and aug-cc-pVTz basis sets were tested. The PBE1PBE and PBE wavefunctions with the aug-cc-pVTz basis set took a long time to complete one stage of the geometry optimisation and so were deemed to be unsuitable for use within *CrystalOptimizer*. Of the remaining wavefunctions, the experimental crystal structures were reproduced with similar degrees of accuracy. The PBE1PBE wavefunction with the 6-31G(d,p) basis set was selected as being the best compromise between accuracy and computational cost.

CrystalPredictor[12] was used to generate Z'=1 structures of lactide in the most common space groups: racemic $P\bar{1}$, $P2_1/c$, $Pna2_1$, $Pca2_1$, *Pbca, Pbcn, C2/c, Cc, Pc, Cm,* $P2_1/m$, *C2/m, P2/c,* $Pmn2_1$, $Cmc2_1$, $Pmn2_1$, $Cmc2_1$, *Aba2, Fdd2, Iba2, Pnna, Pccn, Pbcm, Pnnm, Pmmn, Pnma, Cmcm, Cmca, Fddd, Ibam,* $I\bar{4}$, *P4/n,* $P4_2/n$, *I4/m,* $I4_1/a$, $P\bar{4}2_1/c$, $I\bar{4}2d$, $P\bar{3}$, $R\bar{3}$, *R3/c,* $R\bar{3}/c$, $P6_3/m$ and $Pa\bar{3}$ and chiral *P1,* $P2_1$, *C2,* $P2_12_12$, $P2_12_12_1$, $C222_1$, $P4_1$, $P4_3$, $P4_12_12$, $P4_32_12$, $P3_1$, $P3_2$, *R3,* $P3_121$, $P3_221$, $P6_1$, $P6_3$ and $P2_13$. Due to the lack of flexibility of the lactide molecule a rigid search was performed using the gas phase optimum structure, evaluated at the PBE1PBE/6-31G(d,p) level of theory using Gaussain03[13]. The lattice energies were evaluated from the atomic charges with all other terms represented by an isotropic atom-atom 6-exp potential, with parameters of the FIT[14] potential. Any refinement of the molecular structure due to flexibility was taken into account during the *CrystalOptimizer*[15] stage.

The global *CrystalPredictor* search examined one million structures with the calculated lattice energies refined further in a multi-stage process. All unique

structures within 20 kJ mol^{-1} of the global minimum were taken forward to be refined using *CrystalOptimizer*, covering approximately 1200 structures. *CrystalOptimizer* was used to assess if any of the bonds, angles or torsions within the lactide molecule changed significantly when the whole molecule was allowed to relax across a selection of 30 of the structures generated from the search. As a result the following dihedral angles were treated as flexible within *CrystalOptimizer*: C1-C2-C5-H3, O1-C1-C2-C5, C4-O1-C1-C2, C4-O1-C1-C2, C3-C4-O1-C1, O4-C3-C4-O1 (see Figure 4.1). The lattice energy of each of the structures was evaluated using the PBE1PBE/6-31G(d,p) molecular wavefunction for the conformational energy penalty and distributed multipoles and the FIT exp-6 atom-atom repulsion dispersion.

The effect of polarisation was estimated using a polarisable continuum, ε =3. The cutoff for the repulsion-dispersion interactions was increased from the typical 15 Å to 30 Å for this stage due to the relatively shallow potential energy well. This provided a polarized estimate of the intermolecular energy, ΔE_{intra}, and a set of distributed multipoles, for reminimising the lattice energy for all structures in the search.

To estimate the effects of temperature, the rigid body elastic constants and k=0 phonons were calculated for all of the structures in the crystal energy landscape using the same intermolecular potential and used to calculate the Helmholtz free energy and zero point vibrational energy at 298K.

4.3 Results and Discussion

4.3.1 The Crystal Energy Landscape of Lactide

The crystal energy landscape of lactide (Figure 4.4) contains approximately 1200 structures within an energy range of 24.072 kJ mol^{-1}. Of these 1200 structures, 255 fall within the range of polymorphism (10 kJ mol^{-1}). There appears to be no distinction between the energy and density of the racemic and enantiopure structures, although there is a negative correlation between the energy and density of the structures which is expected. As the predominant interactions between the lactide molecules in the crystal structures are weak van der Waals forces it is expected that, as the distance between the molecules

is increased the stabilising interactions weaken causing a reduction in lattice energy.

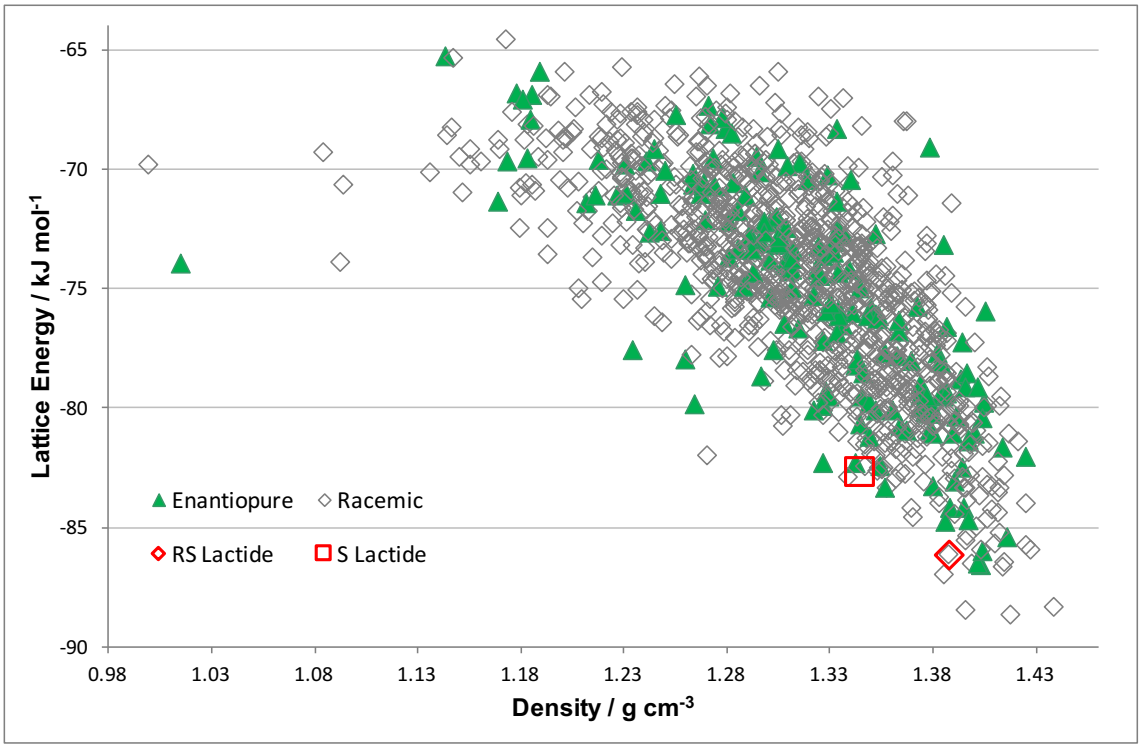


Figure 4-4: The crystal energy landscape of Lactide with lattice energies refined by *Crystaloptimizer*. Racemic structures represented by grey diamonds and enantiopure structures represented by green triangles. The lattice energy minima corresponding to the experimental structures are given in red. The Z'=1 structure of RS lactide was found in the search.

Similarly to the crystal energy landscape of 3-CIMA (see Figure 3.6) the landscape of lactide (Figure 4.4) is also very dense, with the experimental crystal structures lying towards the less dense region. The experimental racemic structure, BICVIS, was found in the search corresponding to structure A9 in the search. As the search was limited to Z'=1, the experimental enantiopure structure was not expected to be found. Both of the experimental structures were found relatively close to the global minimum, with BICVIS ranked as the 8th racemic structure (ranked 10th in the search) and the lattice energy minimum of NAHNOZ ranked just above the 11th enantiopure structure found in the search (ranked 59th in the search).

4.3.2 The inclusion of PCM and the Helmholtz Free Energies

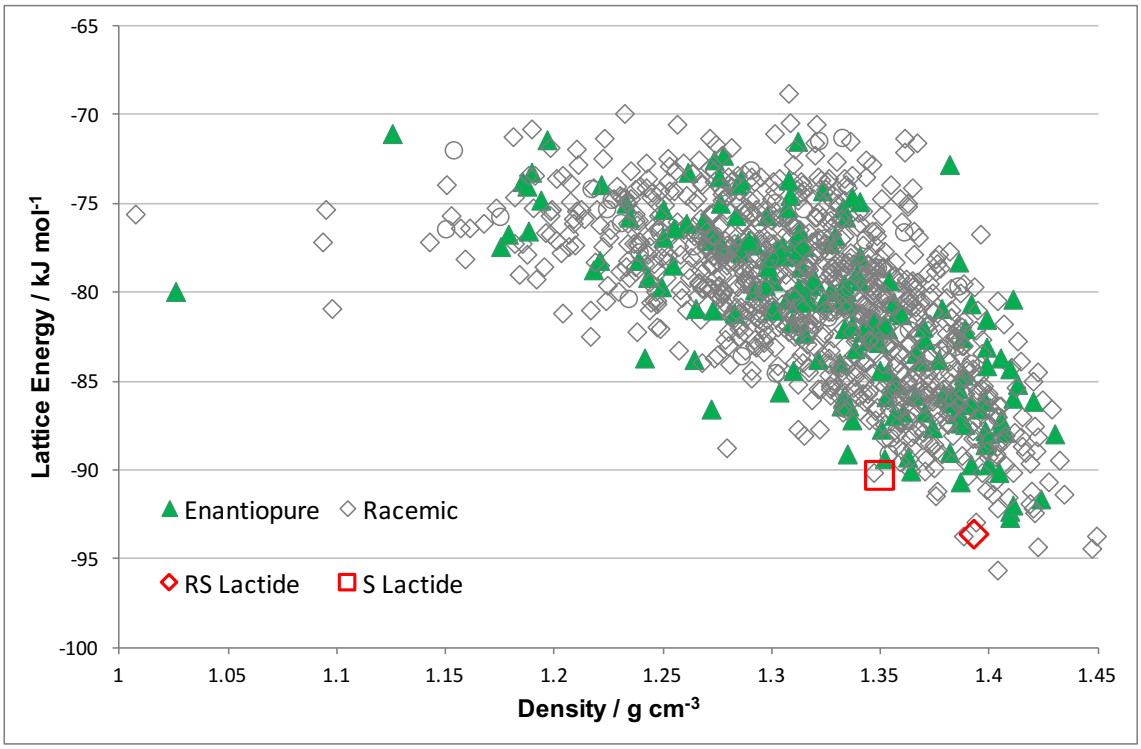


Figure 4-5: The crystal energy landscape of Lactide with lattice energies refined by *Crystaloptimizer* and using PCM (ε=3). Racemic structures represented by grey diamonds and enantiopure structures represented by green triangles. The lattice energy minima corresponding to the experimental structures are given in red, and were found in the search if Z'=1.

The crystal energy landscape was reproduced including the addition of a polarisable continuum model (ε=3) and with the inclusion of Helmholtz free energies. Due to the lack of hydrogen bonding ability of the lactide molecules it was not expected that the crystal energy landscape would show any real change. With the addition of the PCM (Figure 4.5), BICVIS became the 6th racemic structure (ranked 6th in the search) and NAHNOZ became the 6th enantiopure structure (ranked 27th in the search), showing that the PCM model had significantly improved the ranking of the observed structures.

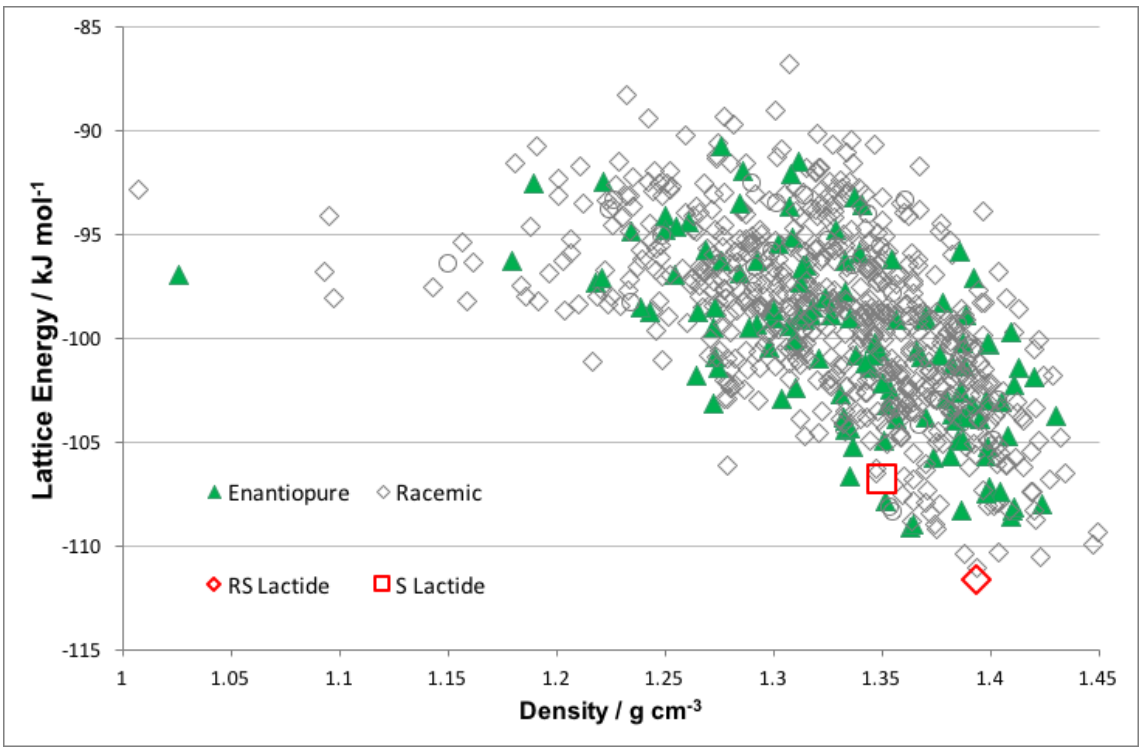


Figure 4-6: The crystal energy landscape of lactide including the effects of the PCM (ε=3) and the Helmholtz free energies. Racemic structures represented by grey diamonds and enantiopure structures represented by green triangles. The lattice energy minima corresponding to the experimental structures are given in red, and that of RS lactide Z'=1 was found in the search.

Upon inclusion of the Helmholtz free energy the ranking of the structures around the global minimum changes even more (Figure 4.6). This brings experimental racemic crystal structure to the global minimum, and re-ranks the enantiopure crystal structure to the 6th most stable enantiopure structure. This lowering in rank of the experimental structures indicates that the inclusion of free energies plays an important role in determining which crystal structure is most favoured at room temperature. However, the crystal energy landscape shows that there are still several crystal structures that are competitive in energy and no large energy gap between the most stable and next ranked crystal structures. Hence the potential for polymorphism of the lactide system needs consideration.

4.3.3 Analysis of the crystal energy landscape

4.3.3.1 Reproduction of the experimental crystal structures

The experimental crystal structures of lactide were reasonably well produced following refinement with *CrystalOptimizer* using the FIT potential with the PBE1PBE wavefunction, 6-31G(d,p) basis set and PCM (see Table 4.1 and Figure 4.7). The RMS overlay of the refined experimental structures were

reasonable, but slightly high for the racemic structure BICVIS. When the unit cell of BICVIS was refined a deformation of the unit cell occurred caused by a decrease in the β angle. This is thought to be due to a very shallow energy potential energy surface. Investigations into the phonon modes of the lactide crystal structures, as detailed in Chapter 6, show that there is a low frequency vibrational mode corresponding to the direction in which the unit cell was deformed. This vibrational mode could account for the slipping of the molecules observed due to the relatively flat potential energy surface.

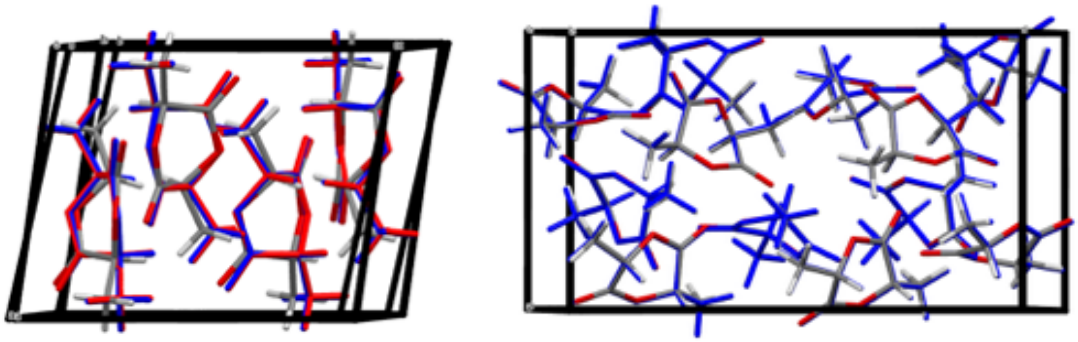


Figure 4-7: Left - Overlay of the experimental racemic crystal structure BICVIS (atom colours) with the crystal structure refined by *CrystalOptimizer* (red), $RMSD_{15}$ = 0.217 and with a PCM (blue), $RMSD_{15}$ = 0.355. Right - Overlay of the enantiopure crystal structure NAHNOZ (atom colours), with the crystal structure refined by *CrystalOptimizer* (red), $RMSD_{15}$ = 0.384, and with a PCM (blue), $RMSD_{15}$ = 0.156.

While the enantiopure structure could not be found in the search as it is not a Z'=1 structure, the experimental racemic structure was found to correspond to hypothetical crystal structure A9. There is good agreement between the experimental crystal structure of RS-lactide and structure A9, with an RMS_{15} of 0.369 Å determined when the crystal packing of these structures were compared (see Figure 4.8). The small differences in energy and packings between the lattice energy minimum found in the search and that found starting from the experimental structure are also thought to be due to a very shallow potential energy surface. The main energy difference is in the intermolecular energy contribution. With no relatively strong hydrogen bonding interactions between the molecules to hold the packing in a more rigid fashion, planes of molecules are more easily able to slide over each other. The deformation of the experimental crystal structure due to a decrease in the β angle could be as a result of slight inaccuracies in the modelling of the van der Waals interactions within the crystal structure. These differing interactions could account for the small energy and density differences observed. Additionally, the experimental crystal structure of S-lactide was measured at room temperature (298-303 K) whereas the hypothetical crystal structures are generated using a 0 K model. It

is possible that the experimentally determined structure is a thermal average and so is not in fact a true minimum, which will account for the discrepancies observed when the experimental crystal structure and hypothetical crystal structure A9 are overlaid.

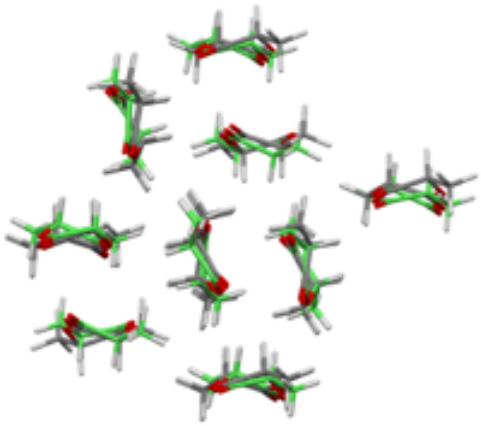


Figure 4-8: Crystal structure overlay of BICVIS (green) with structure A9 from the search (grey). $RMSD_{15}$ = 0.369 Å

Structure	Z	a / Å	b / Å	c / Å	α / °	β / °	γ / °	Z'	U_{inter} / kJ mol^{-1}	ΔE_{intra} / kJ mol^{-1}	U_{latt} / kJ mol^{-1}	RMS_{15} / Å	Packing Coef.	Density / g cm^{-3}
S-lactide (NAHNOZ)	12	9.329	13.615	16.822	90.000	90.000	90.000	3	-	-	-	-	-	-
S-lactide_CO	12	9.485	13.192	17.076	90.000	90.000	90.000	3	-86.42	+0.249	-86.171	0.384	0.7110	1.3881
S-lactide_PCM	12	9.424	13.377	16.907	90.000	90.000	90.000	3	-90.83	+0.571	-90.259	0.156	0.6890	1.3499
RS-lactide (BICVIS)	4	8.050	9.086	9.713	90.000	102.860	90.000	1	-	-	-	-	-	-
RS-lactide_CO	4	7.963	8.714	10.042	90.000	98.181	90.000	1	-83.25	+0.610	-82.64	0.217	0.6879	1.3441
RS-lactide_PCM	4	7.966	8.697	10.014	90.000	98.481	90.000	1	-93.81	+0.215	-93.595	0.355	0.7143	1.3932
A9_PCM	4	7.965	8.701	10.016	90.000	81.547	90.000	1	-93.13	+0.180	-92.950	0.369	0.7130	1.3942
A4_PCM	4	8.068	10.149	9.033	90.000	111.190	90.000	1	-94.47	+0.696	-93.774	0.595 (RMS_{11})	0.7107	1.3883

Table 4-1: The lattice parameters and energies of the experimentally known structures and computer generated structure A9 and A4 as determined with *CrystalOptimizer* with the PBE1PBE wavefunction, FIT exp-6 dispersion-repulsion model, PCM and Helmholtz free energy contribution. A9 represents the structure corresponding to BICVIS found in the search and A4 a closely related structure. The RMS_{15} error of comparing the optimised crystal structures with the experimental crystal structures is stated.

4.3.4 Analysis of the packings observed in the crystal energy landscape

As the energy range of polymorphism is considered to be 10 kJ mol^{-1}, ideally all hypothetical crystals structures within this range of the global minimum would be analysed further. There are approximately 230 structures which fall within this energy range, and so it was decided that a more manageable set of 100 crystal structures would be examined. This set of structures fell within an energy range of 7.6 kJ mol^{-1}.

Of the structures examined it was noticed that there were pairs of structures which had identical packings and similar densities that were ranked very close in energy, although determined to be distinct crystal structures (e.g Figure 4.9). Upon further examination of these crystal structures they are virtually identical, although have slight deviations in crystal structure resulting in energy differences large enough for the structures considered to be independent upon clustering.

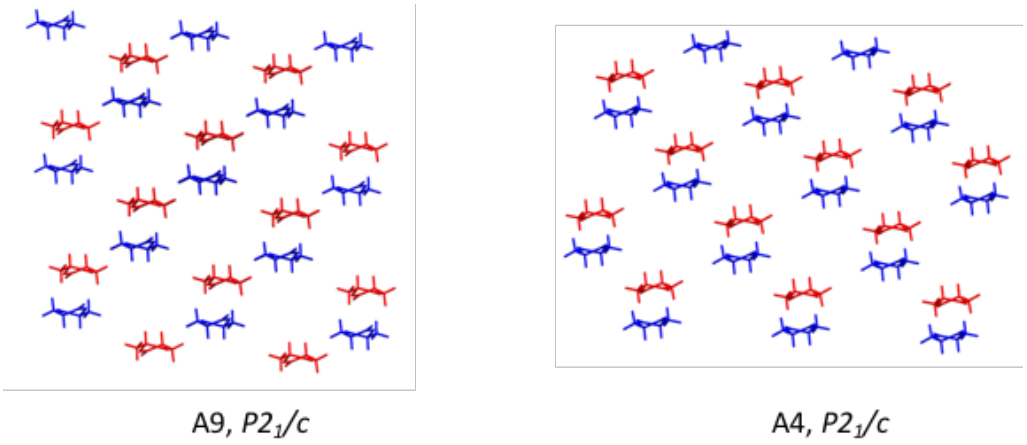


Figure 4-9: The hypothetical crystal structures, A9 and A4 of the lactide molecule. The two structures are related by a sliding of the blue layer of molecules and overlay with 6 molecules out of 15 in common to give an $RMSD_6$ of 0.883 Å. A9 is the closest structure to RS-lactide (BICVIS) but A4 is very similar in energy (Table 4.1)

The packings of lactide emphasise the spherical like nature of the molecule with many packings differing in only the rotation of a stack or plane of molecules within the crystal structure. For example structures A464 and A610 which differ only in the stacking of a layer of molecules (see Figure 4.10). These energy difference between these two structures is just 0.226 kJ mol^{-1}, when free energies are considered.

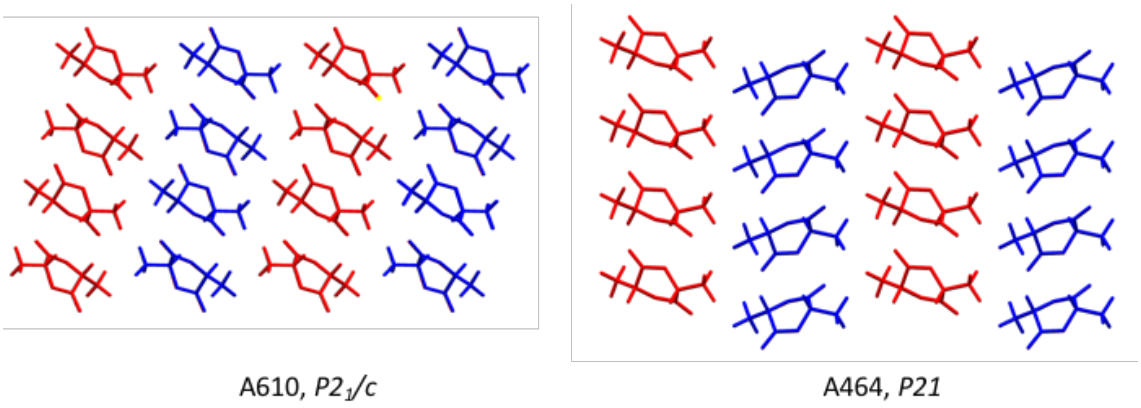


Figure 4-10: The packing in hypothetical crystal structures A610 and A464 which are related by a 180 degree turn in the blue layers

The presence of structures with related packings close in energy and structures which could be considered identical due to a slip in a plane of molecules emphasises the shallow potential energy surface of this molecule. These weak van der Waals forces which dominate in the crystals of lactide allow for relative free rotation of the lactide molecule with little energy penalty. This has the potential to lead to disordered structures.

4.3.5 Why are the experimentally determined crystal structures of lactide not the global minimum racemic and enantiopure structures?

To date, experimentally there has been no indication of any other polymorphs of either the racemic or enantiopure crystal structures of lactide. Is there an explanation for why these structures are not found at the global minimum?

4.3.5.1 Is the current computational model inadequate?

The racemic structure of lactide is only determined to be the global minimum structure when the effects of free energies are included. These provide an estimate of the energetic landscape at room temperature, as opposed to the 0 K model our crystal structure prediction techniques are based on. The experimental crystal structures were obtained at room temperature and so it would appear that the thermal effects included in the free energy calculations play an important role in the stability of lactide crystals. These free energy calculations are modelled using the harmonic approximation and it is likely that for the lactide system the harmonic approximation is not completely valid as it does not allow for thermal expansion. Although experiments on both RS and S lactide have been performed across a broad temperature range (Chapter 6.2), it is not possible to rule out that the system could be enantiotropic. Crystal

structure prediction studies have successfully been able to guide experimentalists in designing experiments to investigate the changing stabilities of polymorphs across a temperature range, such as with 5-flurouracil.[16] Even if the lactide system is not enantiotropic, it is clear that the effects of free energies are important in the stabilising the crystal structures of lactide.

Even when including the effects of free energies on the crystal energy landscape the experimental structure of S-lactide is found to be at a higher energy and density then the Z'=1 crystal structures produced in the search. The refined structure of NAHNOZ, including the effect of free energies, is ~2.3 kJ mol^{-1} above the global minimum enantiopure structure. The 11 Z'=1 hypothetical crystal structures which lie beneath NAHNOZ in the crystal energy landscape are generally much denser than the experimental structure. Analysis of the CCDC shows that there is a tendency for chiral molecules to adopt high Z' structures[17], with the enantiopure crystal structures of 3ClMA being a good example of this (see Chapter 3).

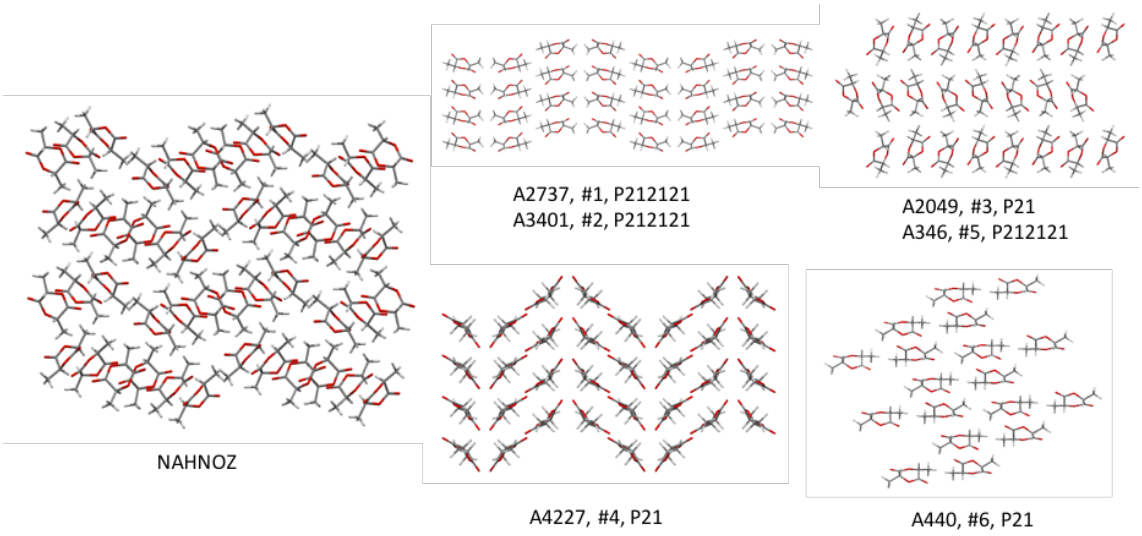


Figure 4-11: The structural packings of the low energy Z'=1 enantiopure structures of lactide produced in the search compared with the experimental Z'=3 NAHNOZ.

Structural analysis of the hypothetical Z'=1 enantiopure crystal structures of lactide show no similarity to that of the Z'=3 experimental structure of S-lactide (see Figure 4.11). The structure of NAHNOZ appears to be dominated by clusters of three lactide molecules which cannot be easily replicated in a Z'=1 structure. As the structure of NAHNOZ was determined from sublimation, it would suggest that the trimeric clusters of lactide molecules would have to come together in the gas phase before crystallising and that most of the gas

phase would have to be formed from these clusters. The statistical likelihood of 3 gas phase molecules clustering together is small, let alone the entirety of the gas phase to be formed by trimers. However, if there was a specific stabilising interaction between these trimers, which made them more preferable over a single molecule in the gas phase then this cannot be ruled out. Molecular dynamics calculations could be used to investigate the behaviour and interactions between molecules of lactide in the gas phase and potentially give an insight into the nucleation and overall crystal structure of lactide.

4.3.5.2 Was it due to the way the crystals were obtained?

The crystals of RS lactide were obtained by selecting crystals of an appropriate size (>0.4mm), with no apparent analysis of the bulk sample. The sample could contain a mixture of crystals, with the ones which were not selected due to their small size being of a different/disordered crystal structure. The phenomenon of Ostwald's ripening, where during the crystallisation process unstable forms are obtained first which then transform into a more stable form, could be occurring in the sample. However, without analysis of the bulk it is impossible to tell this. Additionally, crystals of the racemic and enantiopure lactide were obtained via sublimation under vacuum of the purified compound. Crystallisation via sublimation usually takes place at relatively high temperatures, which means that within the system there is a lot of energy when the crystal grows. At high temperatures differences in molecular conformation and/or orientation can be far more insignificant than at lower temperatures, which can lead to an increased likelihood of disordered and complex crystal structures. Additionally, at higher temperatures, crystals tend to grow faster, increasing the chances of obtaining kinetic artefacts. Both experimental techniques lend to crystals which grow fast and therefore the most kinetically accessible.

It is possible that, due to the experimental methods in which the crystal structures were obtained, that these are long lived kinetic artefacts, which could slowly transition into a more stable crystal structure (c.f. graphite and diamond). Questions such as these cannot be answered quickly as long term studies over years of decades would be required to investigate such theories.

4.3.5.3 Has the correct experiment been performed yet?

Although a number of experiments have been performed on the enantiopure and racemic systems of lactide producing no other crystal structures, this has by no means been a fully exhaustive search. There are numerous experiments in which polymorphs can be obtained such as long term slurry experiments, to spray drying. The type of experiment used can favour stable over metastable polymorphs and vice versa and the type of experiment which can be performed depends entirely on the properties of the compound under investigation. In addition to this there are copious factors which affect crystallisation from the seed used, to temperature and humidity. The Z'=1 hypothetical crystal structures of enantiopure lactide could exist, however the right experiment and conditions might not yet have been found. Or indeed the right crystallisation experiment might not be able to be performed at all due to the thermal and/or solubility properties of the molecule itself. [18] The crystal structure obtained can be down to a number of factors, such as specific impurities in the sample[19], specific solvent, temperature and pressure conditions to name a few. There are an increasing number of reports of disappearing polymorphs,[20-23] where a new more stable crystal structure has formed and since that polymorph's appearance it is no longer possible to obtain the original crystal structure. As well as examples of a computational searches suggesting more stable crystal structures which have later been found experimentally, such as with 5-fluoruracil.[24] To date, these experimental crystal structures may be the most stable crystal structures of lactide. However, there is still only limited research on the monomers of lactide and it is not possible to rule out a more exhaustive polymorph screen giving rise to further structures.

4.3.5.4 Are there errors any other calculation errors?

Although it is possible that more stable crystal structures of the lactides could exist according to the crystal energy landscape, it is also feasible that there are significant errors in the crystal energy landscape such that the global minimum structures generated are not, in fact, the true global minimum. The crystal structures of lactide are dominated by weak CH···O interactions, with no stronger directional intermolecular forces, such as hydrogen bonds. These weak CH···O interactions are in the energy range of 0.5 – 1 kJ mol^{-1} and therefore any small errors made in calculating them, including those within the

typical error range, could have a significant impact in the overall calculated lattice energy, and therefore ranking of the crystal structures in the landscape. There is only an energy difference of 2.3 kJ mol^{-1} between the experimentally determined S-lactide and the most stable enantiopure hypothetical crystal structure. This energy range is well within the uncertainty of our calculations.

There is also a tendency of CSP studies to over predict the number of polymorphs that could be accessible experimentally. This could be due to the fact that CSP does not currently accurately take into account the thermal motions of crystal structures which are measured at 298 K or the kinetic routes to low energy crystal structures are not taken into account. Molecular dynamics simulations would need to be performed in addition to producing a crystal energy landscape to fully explore the thermal motions of the hypothetical crystal structures produced. To ensure the entire search space is sampled, crystal structures could be generated that are simply not plausible to obtain experimentally. It is the role of the researcher examining the crystal energy landscape to be able to draw on experimental knowledge to assess the plausibility of the crystal structures produced.

4.4 Conclusions

The crystal energy landscape of lactide was produced, successfully placing the racemic experimental structure at the global minimum of the crystal energy landscape, when free energies are included. Although the Z'=3 enantiopure structure is beyond the scope of the search it was found to be located close to the global minimum, however the presence of lower energy, less dense, Z'=1 enantiopure structures leads to questions surrounding the potential polymorphism and accuracy in the computational model of the lactide system. This is a similar situation to what was observed with 3ClMA despite the differing dominant intermolecular forces within the crystal structure.

One of the issues with this structure has been with the clustering of crystal structures. The relatively flat potential energy surface caused many structures to slide and therefore be determined to be distinct crystal structures. The majority of the time spent taken to do a CSP study comes from the human

analysis of the structures produced. Therefore, an improved computational method of automatically clustering these structures would be hugely beneficial. A key concern for this is clustering the structures so much so that the too many structures are thrown away and the experimental crystal structures are missed.

The similarity between the two lowest energy racemic structures could be explored through molecular dynamics simulations which would provide an insight into which of these structures are uniquely present in solution or if the structures slide between one another. The low potential energy surface and number of related crystal structures produced suggests that lactide could form disordered systems. Molecular dynamics simulations, again could provide insight into the potential for disorder. Alternatively, investigations into predicted growth rates and morphologies could provide added insight into some of the hypothetical crystal structures which have been predicted to be more stable than the experimental structures. If growth rates were found to be unrealistically slow or give rise to mechanically unstable structures then these could be ruled out as potential competitive polymorphs of lactide.

The crystallisation processes of organic molecules are highly complex and still relatively poorly understood. Crystal structure prediction methods can help to highlight potential polymorphs, and even help experimentalists find new structures. However, it is not without its faults. The crystal structures of lactides are dominated by weak intermolecular forces, and so any small error in the calculation of these forces within the crystal structures can lead to a misleading crystal energy landscape.

Crystal structure prediction methods can accurately predict the most stable crystal structure of a compound. But care and a good understanding of experiment needs to be taken when interpreting the results.

4.5 References

(1) van Hummel, G. J.; Harkema, S.; Kohn, F. E.; Feijen, J. Structure of 3,6-dimethyl-1,4-dioxane-2,5-dione [d-,d-(l-,l-)lactide]. *Acta Crystallographica Section B* **1982**, *38*, 1679-1681.

(2) Belenkaya, B. G.; Belsky, V. K.; Dementev, A. I.; Sakharova, V. I.; Chernikova, N. T. Crystal and Molecular Structures of Glycolide and Lactide: Association through CH...O Hydrogen Bonds. *Crystallography Reports* **1997**, *42*, 449-452.

(3) Ouchi, T.; Ohya, Y. Design of lactide copolymers as biomaterials. *Journal of Polymer Science Part A: Polymer Chemistry* **2004**, *42*, 453-462.

(4) Schindler, A.; Harper, D. Poly (lactic acid). I. Stereosequence distribution in the polymerization of racemic dilactide. *Journal of Polymer Science: Polymer Letters Edition* **1976**, *14*, 729-734.

(5) Jalabert, M.; Fraschini, C.; Prud'homme, R. E. Synthesis and characterization of poly(L-lactide)s and poly(D-lactide)s of controlled molecular weight. *Journal of Polymer Science Part A: Polymer Chemistry* **2007**, *45*, 1944-1955.

(6) Garlotta, D. A Literature Review of Poly(Lactic Acid). *Journal of Polymers and the Environment* **2001**, *9*, 63-84.

(7) Buchholz, H. Personal Communication, 2017.

(8) Le Minh, T. Designing crystallization based-enantiomeric separation for chiral compound-forming systems in consideration of polymorphism and solvate formation. Otto-von-Guericke-Universität, Magdeburg, 2014.

(9) Emel'yanenko, V.; Verevkin, S.; Pimerzin, A. The thermodynamic properties of DL- and L-lactides. *Russian Journal of Physical Chemistry A* **2009**, *83*, 2013-2021.

(10) Wu, W.; Li, W.; Wang, L.; Zhang, P.; Zhang, J. Density functional theory study on lactides: Geometries, IR, NMR and electronic spectra. *Journal of Molecular Structure: THEOCHEM* **2007**, *816*, 13-19.

(11) Reilly, A. M.; Cooper, R. I.; Adjiman, C. S.; Bhattacharya, S.; Boese, A. D.; Brandenburg, J. G.; Bygrave, P. J.; Bylsma, R.; Campbell, J. E.; Car, R.; Case, D. H.; Chadha, R.; Cole, J. C.; Cosburn, K.; Cuppen, H. M.; Curtis, F.; Day, G. M.; DiStasio Jr, R. A.; Dzyabchenko, A.; van Eijck, B. P.; Elking, D. M.; van den Ende, J. A.; Facelli, J. C.; Ferraro, M. B.; Fusti-Molnar,

L.; Gatsiou, C.-A.; Gee, T. S.; de Gelder, R.; Ghiringhelli, L. M.; Goto, H.; Grimme, S.; Guo, R.; Hofmann, D. W. M.; Hoja, J.; Hylton, R. K.; Iuzzolino, L.; Jankiewicz, W.; de Jong, D. T.; Kendrick, J.; de Klerk, N. J. J.; Ko, H.-Y.; Kuleshova, L. N.; Li, X.; Lohani, S.; Leusen, F. J. J.; Lund, A. M.; Lv, J.; Ma, Y.; Marom, N.; Masunov, A. E.; McCabe, P.; McMahon, D. P.; Meekes, H.; Metz, M. P.; Misquitta, A. J.; Mohamed, S.; Monserrat, B.; Needs, R. J.; Neumann, M. A.; Nyman, J.; Obata, S.; Oberhofer, H.; Oganov, A. R.; Orendt, A. M.; Pagola, G. I.; Pantelides, C. C.; Pickard, C. J.; Podeszwa, R.; Price, L. S.; Price, S. L.; Pulido, A.; Read, M. G.; Reuter, K.; Schneider, E.; Schober, C.; Shields, G. P.; Singh, P.; Sugden, I. J.; Szalewicz, K.; Taylor, C. R.; Tkatchenko, A.; Tuckerman, M. E.; Vacarro, F.; Vasileiadis, M.; Vazquez-Mayagoitia, A.; Vogt, L.; Wang, Y.; Watson, R. E.; de Wijs, G. A.; Yang, J. Z.; Zhu, Q.; Groom, C. R. Report on the sixth blind test of organic crystal-structure prediction methods. *Acta Crystallographica Section B - Structural Science* **2016**, *submitted*.

(12) Habgood, M.; Sugdan, I. J.; Kazantsev, A. V.; Adjiman, C. S.; Pantelides, C. Efficient Handling of Molecular Flexibility in Ab Initio Generation of Crystal Structures. *Journal of Chemical Theory and Computation* **2015**, *11*, 1957-1969.

(13) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A.: Gaussian 03. Gaussian Inc.: Wallingford CT, 2004.

(14) Coombes, D. S.; Price, S. L.; Willock, D. J.; Leslie, M. Role of Electrostatic Interactions in Determining the Crystal Structures of Polar Organic Molecules. A Distributed Multipole Study. *Journal of Physical Chemistry* **1996**, *100*, 7352-7360.

(15) Kazantsev, A. V.; Karamertzanis, P. G.; Adjiman, C. S.; Pantelides, C. C.: CrystalOptimizer. An efficient Algorithm for Lattice Energy Minimisation of Organic Crystal using Isolated-Molecule Quantum Mechanical Calculations. In *Molecular System Engineering*; Adjiman, C. S., Galindo, A., Eds.; Process Systems Engineering; WILEY-VCH Verlag GmbH & Co.: Weinheim, 2010; Vol. 6; pp 1-42.

(16) Karamertzanis, P. G.; Raiteri, P.; Parrinello, M.; Leslie, M.; Price, S. L. The Thermal Stability of Lattice Energy Minima of 5-Fluorouracil: Metadynamics as an Aid to Polymorph Prediction. *Journal of Physical Chemistry B* **2008**, *112*, 4298-4308.

(17) Anderson, K. M.; Goeta, A. E.; Steed, J. W. Supramolecular synthon frustration leads to crystal structures with Z '> 1. *Crystal Growth & Design* **2008**, *8*, 2517-2524.

(18) Price, S. L. Why don't we find more polymorphs? *Acta Crystallographica Section B - Structural Crystallography and Crystal Chemistry* **2013**, *69*, 313-328.

(19) Lancaster, R. W.; Harris, L. D.; Pearson, D. Fifty-year old samples of progesterone demonstrate the complex role of synthetic impurities in stabilizing a metastable polymorph. *CrystEngComm* **2011**, *13*, 1775-1777.

(20) Lancaster, R. W.; Karamertzanis, P. G.; Hulme, A. T.; Tocher, D. A.; Lewis, T. C.; Price, S. L. The Polymorphism of Progesterone: Stabilization of a 'Disappearing' Polymorph by Co-Crystallization. *Journal of Pharmaceutical Sciences* **2007**, *96*, 3419-3431.

(21) Henck, J. O.; Bernstein, J.; Ellern, A.; Boese, R. Disappearing and reappearing polymorphs. The benzocaine : picric acid system. *Journal of the American Chemical Society* **2001**, *123*, 1834-1841.

(22) Bombicz, P.; Czugler, M.; Tellgren, R.; Kalman, A. A classical example of a disappearing polymorph and the shortest intermolecular H center dot center dot center dot H separation ever found in an organic crystal structure. *Angewandte Chemie-International Edition* **2003**, *42*, 1957-1960.

(23) Rubin-Preminger, J. M.; Bernstein, J. 3-aminobenzenesulfonic acid: A disappearing polymorph. *Crystal Growth & Design* **2005**, *5*, 1343-1349.

(24) Hulme, A. T.; Price, S. L.; Tocher, D. A. A New Polymorph of 5-Fluorouracil Found Following Computational Crystal Structure Predictions. *Journal of the American Chemical Society* **2005**, *127*, 1116-1117.

5 Participation in the 6th Blind Test of Crystal Structure Prediction

5.1 Introduction

While it has been shown that CSP is a valuable tool to work alongside studies in which experimental structures have already been determined, what happens if we do not have *a priori* knowledge of what the crystal structure might be? Can we actually use these computational methods to predict crystal structures without any knowledge of experimentally known forms?

5.1.1 CCDC Blind Tests

Since 1999 the CCDC have held six 'Blind Tests' in which research groups are invited to predict the crystal structures of a selection of predetermined organic molecules. These organic crystal structures are unpublished, fully determined, high quality structures with no disorder, held in confidence until the end of the Blind Test. Molecular diagrams and the crystallization solvents used to obtain the crystals are provided to each of the participants and it is from this information only that hypothetical crystal structures are put forward. The first three tests were held via invitation only, being opened up to any group that wished to participate for all subsequent tests. The Blind Tests allow an unbiased evaluation of the current methods used for crystal structure prediction and provide a unique opportunity to show researchers just how much left there is to learn on the topic of crystal structure prediction.

Most of the methods used in the Blind Tests use the assumption that the crystal structure determined is the most thermodynamically favourable form and that the compounds selected are monomorphic. However, it is not always guaranteed that these structures are, indeed, monomorphic. This was highlighted after the 2001[1] Blind Test when it was observed that for two of the test molecules alternative competitive hydrogen bonding motifs to the released experimental crystal structure were consistently predicted amongst the applicants. Subsequently new polymorphs have been discovered based on the alternative hydrogen bonding motifs predicted.[1,2] This occurred again in the 4th Blind Test with gallic acid monohydrate where a more stable polymorph was

predicted, and subsequently crystallised.[3] Therefore, one of the main problems in crystal structure prediction is with polymorphism. It is not possible to know whether the crystal structure being targeted is indeed the most thermodynamically stable polymorph and it could be a metastable structure resulting from kinetics. If the crystal structure obtained is, indeed, a metastable polymorph then the chances of it being missed, or ranked far above the global minimum within a Blind Test situation are high.

It has only been with the 5th and 6th Blind Tests that the notion of polymorphism has been explored. For the 5th Blind Test the structure of the third and fourth polymorphs of gallic acid monohydrate were to be predicted with two previously characterised polymorphs already in the CCDC database. For the 6th Blind Test competitors were asked to predict the five polymorphs of molecule XXIII (see Figure 1) however this time there were no previously characterised polymorphs available. A polymorph screen of XXIII had been carried out prior to the start of the Blind Test while a polymorph screen of flexible XXVI was carried out during the course of the Blind Test. Hence the rules for submission for the 6th Blind Test were changed from submitting 3 structures plus an extended list of 100 structures, to two lists of 100 structures for each target compound which could be generated or ranked using different methods. Altering the submissions in this way removes the arbitrary cut off that the structures must lie around the global minimum of the crystal energy landscapes, as well as allowing a greater depth of information to be gathered about each of the structures investigated.

5.1.2 Considerations of the crystal energy landscapes used for Blind Tests

To add to the complications arisen by polymorphism, crystal energy landscapes generate a large number of thermodynamically plausible crystal structures, a far larger number than possible polymorphs. Some of these additional crystal structures arise from the approximations made during the generation of the crystal energy landscape, particularly as the lattice energies of the structures are determined at 0 K, neglecting molecular motion of molecules. Therefore, it is imperative that the crystal energy landscape is determined with the highest possible accuracy, preferably with the consideration of thermal effects and free energies (see Chapter 2.6.4). For many groups who took part in the 6th Blind

Test the addition of a free energy estimation was what distinguished between the two submitted lists.

The Blind Tests have highlighted areas of improvement in crystal prediction studies and consequently the methods used for CSP studies have evolved since the first Blind Test in 1999. The key difficulty in crystal structure prediction lies with the accuracy in determining intermolecular interactions. Initially, the only option was to use empirical potentials to give the intermolecular interactions with purely force-field based methods, but by the end of the third blind test it was realised that significant improvements were unlikely to be made using this approach. Molecule specific potentials have been used in CSP studies since the first Blind Test,[4] however with advances in computational power and methods, methods of deriving molecule-specific potentials from ab initio data have been developed and are more commonly used.[5-7]The Blind Tests have shown that the more accurate energy evaluations have been derived from the electronic structure of the either the crystal or molecule (see Chapter 2.7), however, while this type of calculation could be performed for small simple crystal structures, the cost of the calculation scales badly with complexity and so for large, flexible molecules and crystals containing more than one dependent molecule this option is not always feasible.

5.1.3 The 6th Blind Test (2014-2015)

The 2014-2015 Blind Test[8] contained 5 different categories of molecule: a rigid molecule, a salt hydrate, a co-crystal, a large flexible molecule and a partially flexible molecule with five known but unpublished polymorphs, two of which were Z'=2 (Figure 5.1). Participants were invited to submit two lists of 50-100 hypothetical crystal structures for each target system.

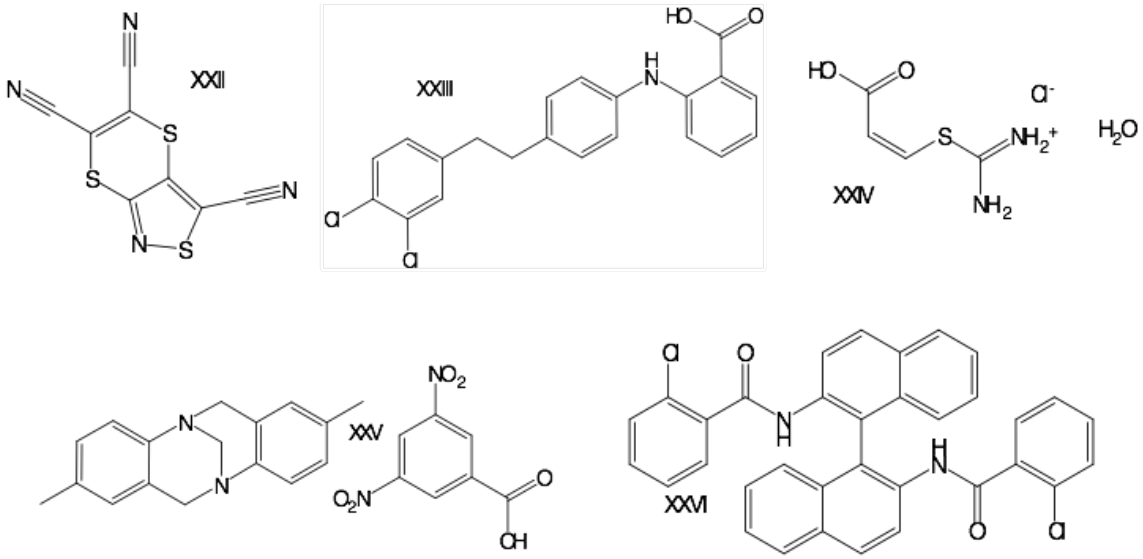


Figure 5-1: The 5 target systems considered for the 6th CCDC Blind Test

The Price group submitted predictions for all the target molecules, with each researcher doing one system, using the same overall approach adapted to the specific molecule. A decision was made not to use periodic DFT+D, but to pass sets of 1000 structures to other participants who were developing such methods but did not have the capability of generating structures. This chapter will concentrate on the molecule XXII. Full details of the group submission can be found in the ESI of the paper reporting the overall results of the Blind Test.[8]

5.1.3.1 Molecule under investigation

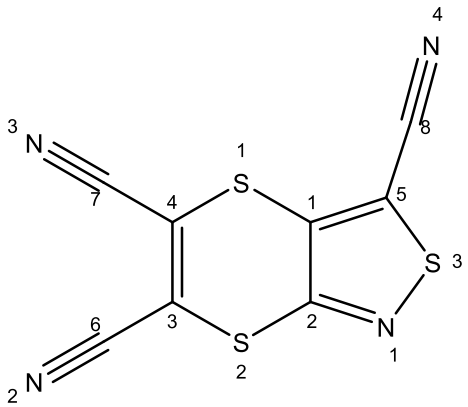


Figure 5-2: The XXII molecule showing the enumeration of the atoms used.

The molecule under investigation for the rigid molecule category was [1,4]dithiino[2,3-c]isothiazole-3,5,6-tricarbonitrile, molecule XXII (Figure 5.2). While XXII was initially stated to be rigid, participants were later informed that the molecule could be considered as having chiral like character because of potential flexibility of the six membered ring although no chiral precursors were

used in the synthesis. Conformations could be adopted which position the two nitrile groups and connected ring atoms in the six membered ring above or below the plane of the 5 membered ring. XXII can therefore be considered to be pseudo-chiral and adopt either a bent or planar configuration (see Figure 5.3)

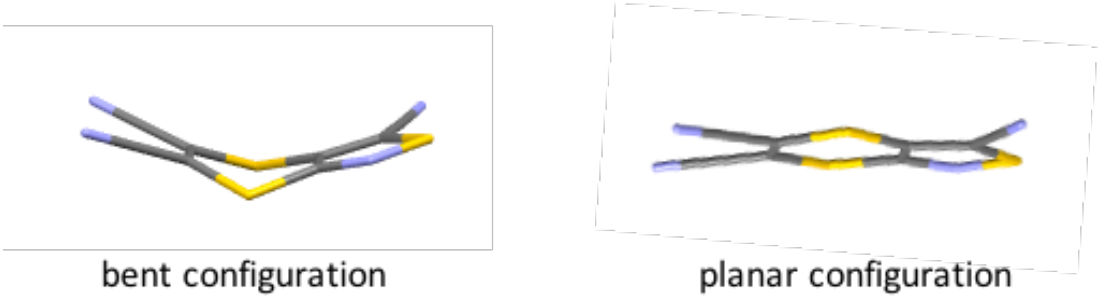


Figure 5-3: The bent and planar configurations of XXII. The bent conformation is pseudo-chiral

There are a number of challenges associated with this relatively simple looking molecule, aside from the potential flexibility. The main challenge lies with the lack of suitable sulphur potential within the FIT potential.[9] Two potentials have previously been developed to work alongside the FIT potential; the Halgren[10] potential which was developed on fitting to thiols and thioethers and the Scheraga[1] potential which was developed on fitting to thioazides. Both are isotropic repulsion-dispersion potentials, fitted to S in a different covalent bonding environment. The cyanide group similarly has a different type of N atom to most organic molecules, and both groups are exposed at the perimeter of the molecule. Hence, using a model that is only anisotropic and specific to the molecule in the electrostatic term could be a gross approximation: a molecule specific anisotropic repulsion-dispersion potential would be desirable, such as that used for 1,3-dibromo-2-chloro-5-fluorobenzene[6,11] As an additional complication, an introductory search of the CCDC revealed no molecules similar to XXII that could be used to test how well either of these potentials reproduced the crystal structure. Therefore, an investigation into the sensitivity of the results to the sulphur potential had to be done after the generation of the crystal energy landscapes.

5.2 Methodology

5.2.1 Generation of hypothetical crystal structures

A molecule of XXII was generated using Molden[12] with the molecular structure optimised in the gas phase with Gaussian03[13] at the PBE1PBE/6-31G(d,p) level

of theory. The MP2, PBE, B3LYP wavefunctions with the 6-31G(d,p) and aug-cc-pVTz basis sets were also tested however it was determined that the PBE1PBE/6-31G9(d,p) wavefunction and basis set provided the best compromise for the cost of calculation as all combinations of wavefunction and basis set optimised XXII to the same geometry.

The gas phase optimized molecule was used as the precursor for the *CrystalPredictor*[14] search. A rigid, Z'=1 search was carried out, with energies evaluated using the PBE1PBE/6-31G(d,p) molecular wavefunction, amongst the most common space groups: *P*1, *P*-1, $P2_1$, $P2_1/c$, $P2_12_12$, $P2_12_12_1$, $Pna2_1$, $Pca2_1$, *Pbca*, *Pbcn*, *C*2/*c*, *Cc*, *C*2, *Pc*, *Cm*, $P2_1/m$, *C*2/*m*, *P*2/*c*, $C222_1$, $Pmn2_1$, $Cmc2_1$, *Aba*2, *Fdd*2, *Iba*2, *Pnna*, *Pccn*, *Pbcm*, *Pnnm*, *Pmmn*, *Pnma*, *Fddd*, *Ibam*, $P4_1$, *P*43, *I*-4, *P*4/*n*, *P*42/*n*, *I*4/*m*, $I4_1/a$, $P4_12_12$, $P432_12$, $P\text{-}42_1c$, *I*-42*d*, $P3_1$, *P*32, *R*3, *P*-3, *R*-3, $P3_12_1$, $P322_1$, *R*3c, *R*-3c, $P6_1$, *P*63, *P*63/*m*, $P2_13$, *Pa*-3, *Cmcm* and *Cmca*. The lattice energies were evaluated from the atomic charges with all other terms represented by an isotropic atom-atom 6-exp potential, with parameters of the FIT potential. The FIT potential was used with the Halgren[10] (-S-) parameter to represent sulphur at this point. One million structures were generated and subsequently clustered to remove any duplicates based on closest intermolecular atom-atom distances.

The energies of the structures generated by *CrystalPredictor* were then improved upon by replacing the atomic charges with the distributed multipoles of the corresponding charge density and the lattice energy minimised using DMACRYS.[15] Subsequently to this the structures were re-clustered to meet the following criteria: ΔEnergy < 0.5 kJ mol^{-1}, ΔDensity < 0.01 g cm^{-3}, PXRD Sim. (5-40 2θ) > 0.97, and $RMSD_{20}$ (20% distance / 20° angle tolerance) < 0.25 Å. This reduced the total number of structures to 8306 lying within 45 kJ mol^{-1} of the global minimum. 1000 structures, which were all within 20 kJ mol^{-1} of the global minimum were taken forward for further refinement by *CrystalOptimizer*[16] to take into account any flexibility in the molecule.

5.2.2 Flexibility of XXII

Subsequently to performing the rigid search for hypothetical structures, the organisers informed all participants that the molecule was not completely rigid.

A torsion angle scan between the two rings indicated that the molecule had a very low barrier to flip between two conformations, one bent, one more planar.

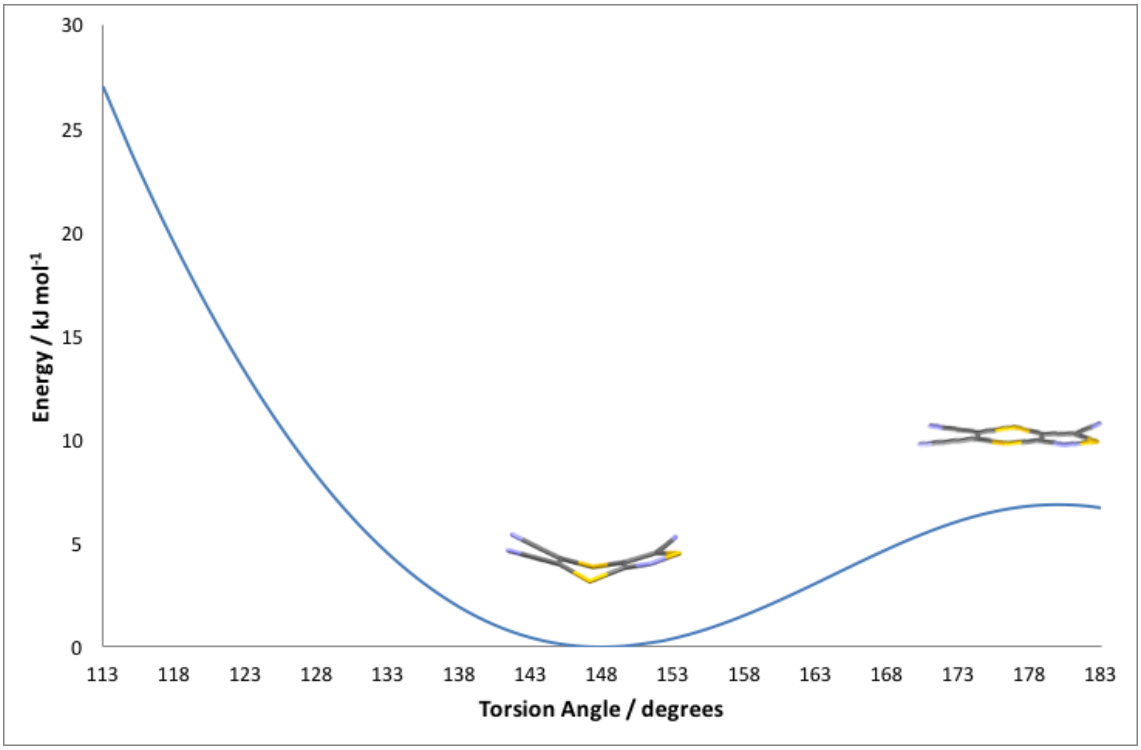


Figure 5-4: The relaxed torsion angle scan between 113 and 183 degrees in 1 degree steps about C3-S2-C2-N1, using the PBE1PBE/6-31G(d,p) level of theory, showing the low energy barrier to flip between the planar and bent configurations of XXII

A torsion angle scan around C3-S2-C2-N1 (see Figure 5.2) showed a very low energy barrier between placing the six membered ring above or in line with the plane of the five membered ring. This implies that the molecule could easily deviate from its gas phase minimum configuration with relatively little energy penalty. It was therefore decided that flexibility should be taken into account during further refinement of the structures.

Full atomistic *CrystalOptimizer* runs were performed on the 30 most stable crystal structures following the addition of multipoles with DMACRYS. Here all angles, bonds and torsions were allowed to vary and those which consistently changed amongst all 30 structures were chosen to be modelled as flexible during all other refinements as these were affected by the packing forces. The angles C8_C5_C1, N2_C6_C4, N3_C7_S1 and N4_C8_C1, and the torsions C4_S1_C1_C2, C7_C4_S1_C1, N2_C6_C4_C3, N3_C7_S1_C4 and N4_C8_C1_C5 (see Figure 5.5) were not held rigid during the optimisation of all 1000 structures carried forward to this stage. The intermolecular energies, ΔE_{intra}, were evaluated with Gaussian03[13] using the PBE1PBE/6-31G(d,p)

wavefunction and analysed by GDMA[17] to give the distributed multipoles. The rigid molecule intermolecular lattice energy was optimised using DMACRYS.[15] The large number of iterations required to converge this two level optimisation problem, up to 200 iterations, to determine a mechanically stable structure, is aided by using a database of the *ab initio* calculations. Once all structures had been refined with *CrystalOptimizer*, the structures were clustered again (subject to the following constraints: ΔEnergy < 0.3 kJ mol^{-1}, ΔDensity < 0.01 g cm^{-3}, PXRD Sim. (5-40 2θ) >0.97, $RMSD_{30}$ (30% distance / 30° angle tolerance) < 0.1Å) and all structures were checked to be a true minima within the space group constraints and those that were not had to be symmetry reduced resulting in the generation of some Z'=2 structures. This comprises the method used to generate and rank the structures submitted as list 1.

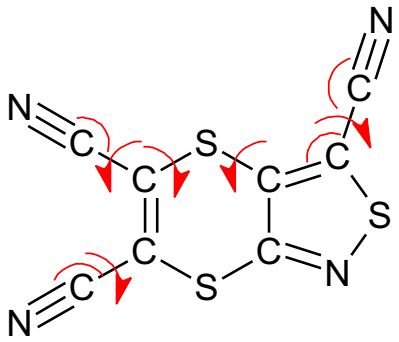


Figure 5-5: XXII highlighting the torsion and bond angles considered to be flexible during *CrystalOptimizer* refinement

The structures refined with *CrystalOptimizer* using the Halgren potential were then re-optimised using the Scheraga[1] potential for sulphur (S=O). A polarisable continuum model (PCM) was also applied (with ε = 3), followed by a rigid body optimisation as well as calculating the k=0 rigid body phonon modes to give an estimate of the free energy at 298 k (See Chapter 2.6.4). This comprises the method used to generate and rank the structures submitted as list 2.
All packing diagrams and short contacts were visualised and calculated using Mercury CSD 3.6.

5.3 Results and Discussion

5.3.1 *CrystalPredictor* Search

Due to the first assumption that XXII was a rigid molecule, a rigid search was performed to generate the hypothetical crystal structures. This step only crudely estimates the lattice energy of the structure, using an interpolation of the atomic

charges. Structures were labelled from their ranking at this stage in the search e.g. A467 is the 467th most stable structure in the search. DMACRYS was used to replace the atomic charges with distributed multipoles to give a more refined estimation of the lattice energy of the hypothetical crystals. The crystal energy landscape, at this stage of the search (see Figure 5.6), shows that there are a small number of structures around the global minimum which are somewhat distinct from the main bulk of the rest of the hypothetical structures.

At this stage the lowest 100 structures fall within an energy range of 9.28 kJ mol^{-1} and only 18 structures fall within 5kJ mol^{-1} of the global minimum, the energy range considered for polymorphism for this small 'rigid' molecule. The global minimum structure at this stage is structure A467 with an estimated energy of -107.5 kJ mol^{-1}.

The largest contribution to the lattice energy comes from the repulsion-dispersion and so the accuracy of the ranking is dependent on how well the repulsion-dispersion is modelled. Indeed, it is the contribution from the repulsion-dispersion to the lattice energy which differs between the two structures at the global minimum. The choice of sulphur potential used in determining the lattice energies is therefore very important and has not been tested fully at this stage of the search. The conformation of the molecule will also have an effect on the intermolecular forces, and so the effect of assuming XXII is rigid needs to be considered. Both the uncertainty in potential and conformation mean a range of structures need to be considered that is wider than the range of polymorphism.

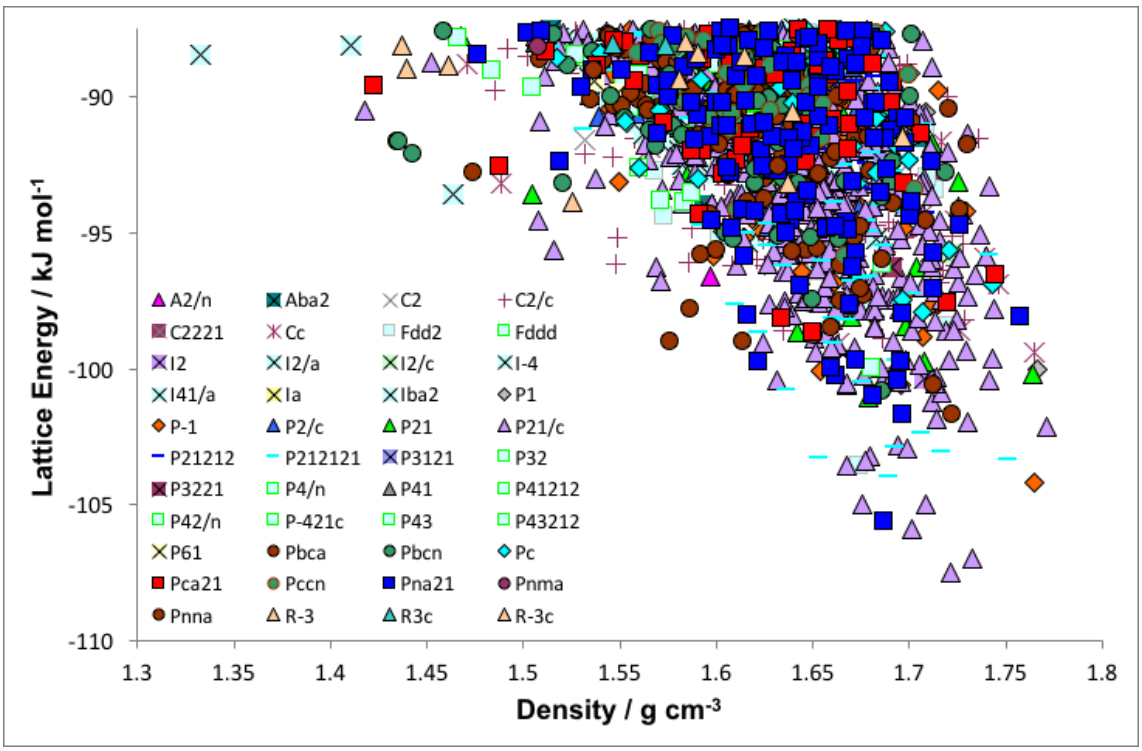


Figure 5-6: Summary of the CSP results for XXII at the CrystalPredictor stage, after the addition of multipoles, evaluated using the PBE1PBE/6-31G(d,p) wavefunction and the FIT potential with Halgren sulphur potential. Each point represents a unique crystal structure with its symbol corresponding to the space group.

5.3.2 *CrystalOptimizer* refinement

The addition of molecular refinement from the *CrystalOptimizer* algorithm caused a degree of re-ranking within the low energy structures from the search as can be seen in Figure 5.7. While the global minimum structure remains A467, there is a significant degree of reranking when the molecule is allowed to deviate from the gas phase configuration. The lowest 100 structures now cover an energy range of 7.749 kJ mol^{-1} with only 65 structures in common with the original 100 structures after the *CrystalPredictor* stage. There are now 24 structures within the energy range considered for polymorphism and grouped around the global minimum, including all the low energy structures from the *CrystalPredictor* stage plus three structures which were previously outside the lowest 100 structures but have improved stability with changing the conformation. These structures remain somewhat separate from the main bulk of structures generated, though are less clearly delineated.

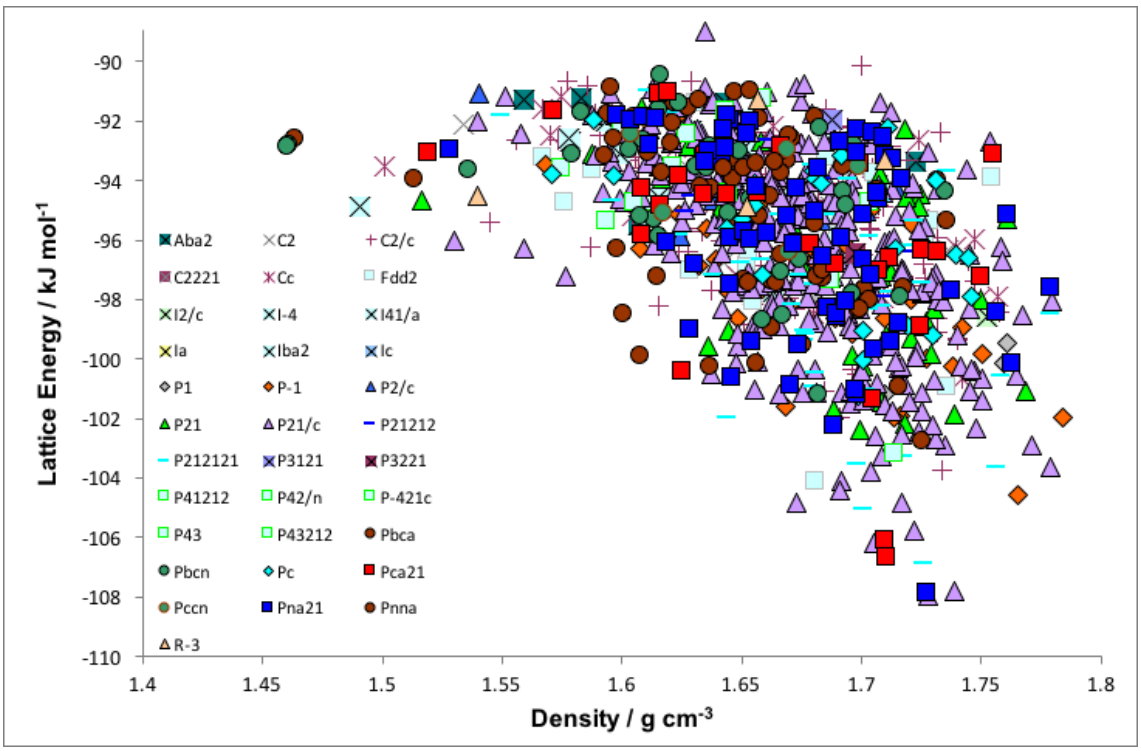


Figure 5-7: Summary of the CSP results for XXII at the *CrystalOptimizer* stage evaluated using the PBE1PBE/6-31G(d,p) wavefunction and the FIT potential with Halgren sulphur potential. Each point represents a unique crystal structure with its symbol corresponding to the space group. This CEL corresponds to list 1.

A more significant degree of reranking can be observed when the structures are reoptimised using the Scheraga potential and the effects of the PCM and free energies are considered (see Figure 5.8). The change in sulphur potential had little effect on the molecular position and geometries within the crystal structure, with the majority of crystal structures having a good overlay with those optimised with the Halgren sulphur potential. However it was noted that there is a general decrease in density in these reoptimised structures. This is evident when comparing Figure 5.7 and Figure 5.8 and upon inspection of the crystal structures it can be noted that crystals which have been refined with the Halgren potential contain many more short close contacts, as defined by Mercury (i.e. shorter than the sum of van der Waals radii -0.2 Å). As a significant proportion of the lattice energy is determined by the intermolecular forces, the degree of reranking could be explained by the degree of change in these close contacts.

The inclusion of the PCM and free energy contributions had little effect on the ranking of the structures around the global minimum, however across the lowest 100 structures there is a significant degree of reranking, with only half of the structures being in common with the lowest 100 structures refined using the

Halgren sulphur potential. This suggests that a vibrational contribution is important for obtaining accurate lattice energies of XXII.

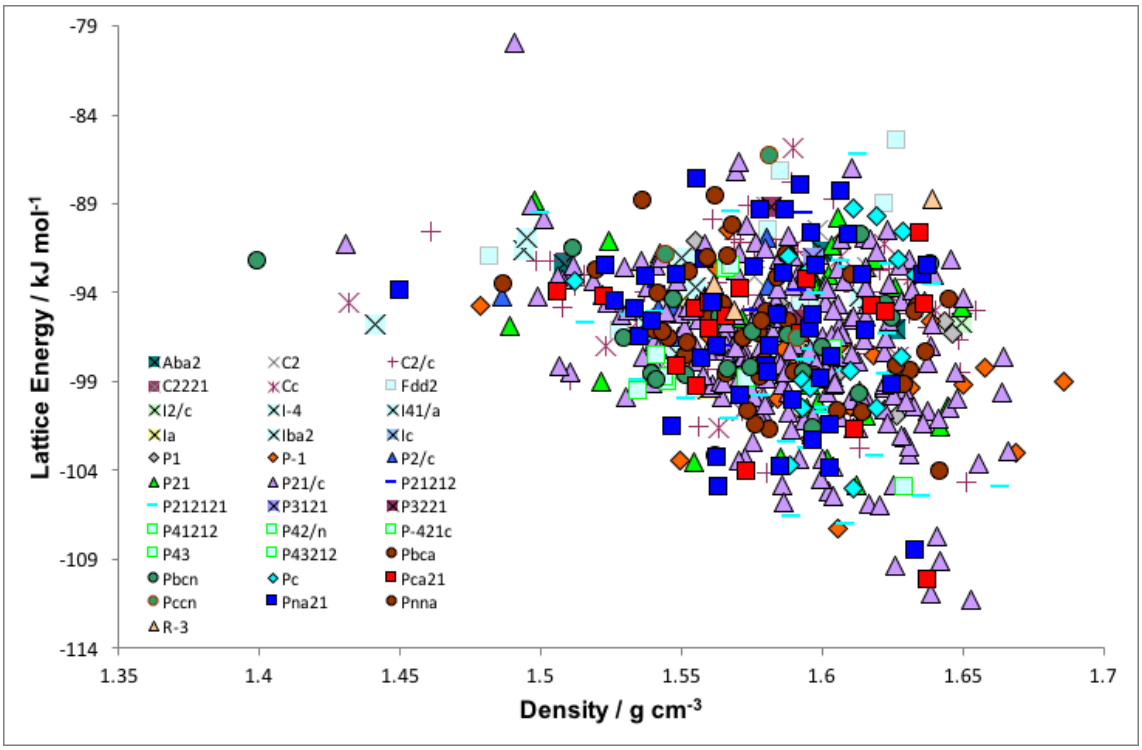


Figure 5-8: Summary of the CSP results for XXII at the CryststalOptimizer stage evaluated using the PBE1PBE/6-31G(d,p) wavefunction and the FIT potential with Scheraga sulphur potential with structures modelled in a PCM (ε = 3) and the effect of free energies. Each point represents a unique crystal structure with its symbol corresponding to the space group it belongs to. This CEL forms the basis of list 2.

5.3.3 Analysis of packings

Inspection of the packings of XXII in the the lowest energy crystal structures from the two refined energy landscapes (Figure 5.7 and Figure 5.8) show the variety of ways in which XXII molecules can pack together. Many of the packings have similar layers but different stackings and orientations of the layers. The degree of bending and flexibility between the five and six membered rings in XXII also adds to the number of possible ways in which the structures can pack together.

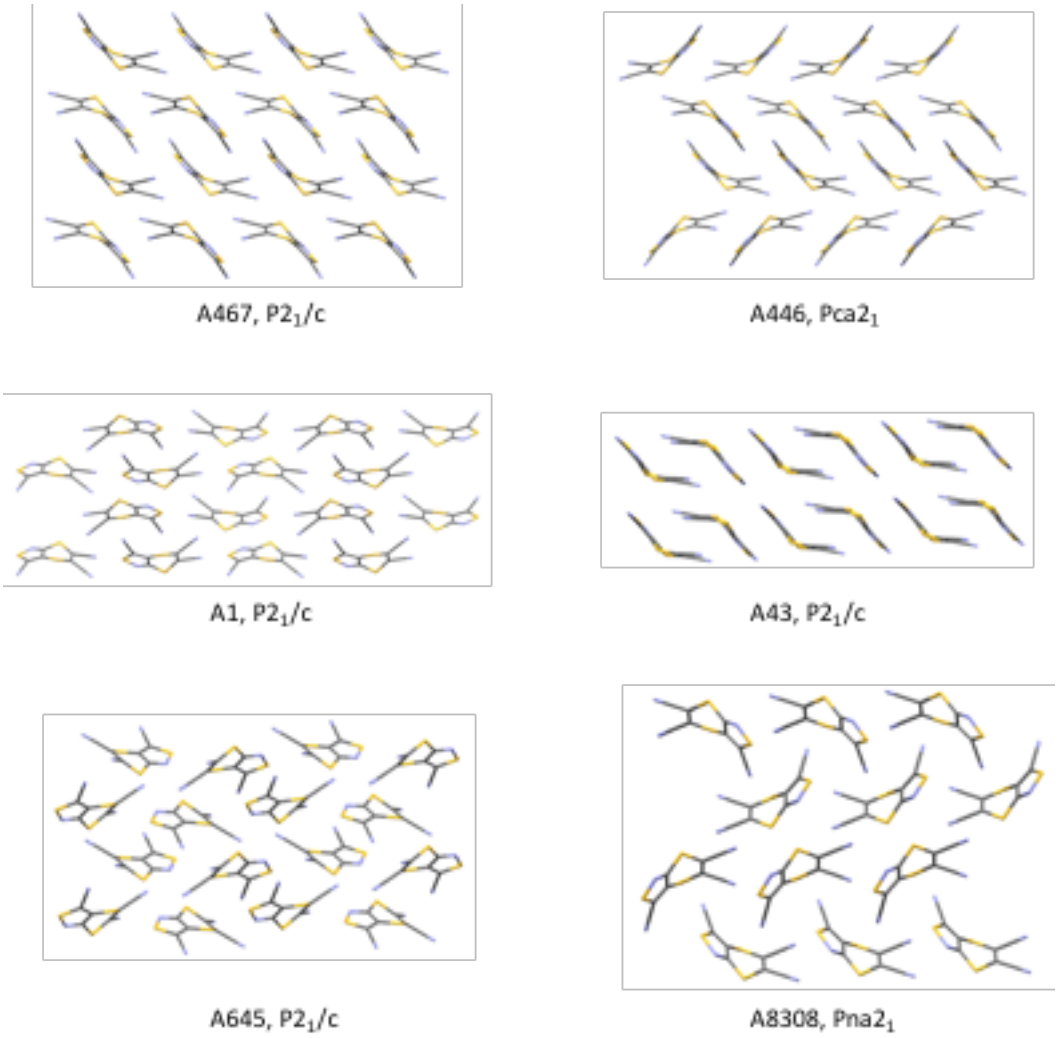


Figure 5-9: A selection of crystal structures around the global minimum, highlighting the diversity in the packings contained within the crystal structures.

The number of different ways in which XXII can pack together within a small energy range leads to the conclusion that there is no strong preference to the orientation in which the molecules crystallise. There is some variation in how bent or planar the XXII molecules are within the structures surrounding the global minimum (see Figure 5.9), however the bent configuration does dominate which would be expected as this is the energetically most favourable configuration. The lack of energetically competitive structures with strongly defined rival motifs suggests that XXII is likely to be able to crystallise in the most thermodynamically stable form. It is possible that the XXII could form a high temperature dynamically disordered polymorph with the molecule flexing or sliding.

5.3.4 Selection of structures and search limitations

As mentioned in 6.1.1, the submissions 6th Blind Test comprised two lists comprised 100 structures. The first list of structures submitted for XXII was based on the lowest 100 structures from the *CrystalOptimizer* search with the Halgren sulphur potential, after the manual removal of any duplicate structures missed during the clustering stage (see Appendix Table 5.9.1). The structures covered an energy range of 7.9 kJ mol^{-1}.

The second list comprised structures optimised using the Scheraga potential with the addition of PCM and free energies. Many of the structures were identical after the use of the alternate sulphur potential although a few structures showed a marked difference in structure. The structures in list 2 followed the energy ranking of list one, up to 5 kJ mol^{-1} (equating to 18 structures). To optimise the chances of success, the rest of list 2 comprised those structures which were not virtually identical to those in list 1. It was hoped that by selecting the structures this way the structural diversity contained within the search could be better represented (see Appendix Table 5.9.2). The 100 structures in list 2 covered an energy range of 11.156 kJ mol^{-1}.

At this point of the submission my analysis suggested I could have moderate confidence that the structure of XXII would be found within the two lists of submitted structures. There are a number of limitations in the search which reduces the confidence of the experimental structure being amongst the crystal structures contained within the two submission lists.

The first is with the configuration of molecule XXII and whether it adopts the gas phase optimum bent configuration or the energetically less favourable planar configuration. The search was initially performed on the rigid bent configuration and then allowed to relax and change during the *CrystalOptimizer* stage. There is a chance that packings may have been missed during the initial generation of structures as the planar configuration was not included, if the structures produced with the bent configuration were unable to relax sufficiently in the *CrystalOptimizer* refinement to reach a structure that contained the planar conformation.

The uncertainty in the adequacy of the sulphur potential also adds a significant degree of uncertainty to the problem. As lattice energy is dominated by the intermolecular energy any inaccuracies in the sulphur potential used could have a significant effect on the precision of the ranking of the structures. With there being no similar structures in the CCDC database in which to test the accuracy of the potential on there is a degree of uncertainty in which potential will provide better results. It was expected that the environment surrounding the sulphur atoms in XXII would be better represented by the Halgren (sulphur single bonded to two carbon side chains). However, the presence of three nitrile functional groups in XXII could affect the environment of S to a significant degree that the Scheraga potential would be more appropriate. Neither potential reflects the anisotropy in the repulsion, so neither potential may be accurate enough.

Since a diverse range of packings and molecular configurations are observed amongst the lowest energy structures generated within relatively small energy windows, this reduces the confidence that the correct structure was submitted. It was hoped that by selecting structures for list 2 that were not observed in list 1 will help to cover the diversity of the packings and configurations observed.

5.4 Alternative methods investigated prior to submission

5.4.1 *CrystalPredictor* + DMACRYS using Custom Delaware Potential

A short time before the submission deadline, Szalewicz and Metz from University of Delaware, were having difficulty finding a program to use the customised potential that they had developed for XXII. The intermolecular potentials were fitted to the energies of dimers calculated using symmetry-adapted perturbation theory based on Kohn-Sham density functional theory, SAPT(DFT).[7] The monomer geometry was obtained from optimisation at the PBE0+D3 level of theory with the aug-cc-pVTz basis set and was assumed rigid. They had fitted an isotropic atom-atom intermolecular potential as very few Molecular Dynamics simulation programs can use anisotropic atom-atom potentials with distributed multipoles.[18] However the form of isotropic atom-atom potential generated for this study was also constrained by limitations of *CrystalPredictor* and DMACRYS.

5.4.1.1 Methodology

Two potentials were constructed from the following functional form

$$V = \sum_{a \in A} \sum_{b \in B} u_{ab}(r_{ab})$$

Equation 5-1

where a and b are the atoms in the monomers, A and B, respectively, and r_{ab} is the distance between them. For this work an isotropic atom-atom function was used, which has the form

$$u_{ab}(r_{ab}) = e^{\alpha_{ab} - \beta_{ab} r_{ab}} + \frac{q_a q_b}{r_{ab}} + \frac{\sqrt{C_a C_b}}{(r_{ab})^6}$$

Equation 5-2

were α_{ab}, β_{ab}, q_i and C_i are optimised parameters. It may be considered as a generalised form of a Buckingham-type potential.

Two potentials were constructed, the first, denoted exp-6_com1, was constructed specifically to be used with *CrystalPredictor* as this requires the following arithmetic combination rules for the exponential parameter:

$$\alpha_{ab} = \alpha_a + \alpha_b$$
$$\beta_{ab} = \beta_a + \beta_b$$

Equation 5-3

It was constructed from data fitted to a set of 981 dimer SAPT(DFT) configurations of the same molecule. The second potential, denoted LL+LR+crys was constructed for use with DMACRYS and uses a more general form with separate α_{ab} and β_{ab} parameters for each atom pair. This potential was fitted to a larger data set of 1423 dimer configurations, the same 981 configurations which were used for the *CrystalPredictor* potential as well as an additional 400 points computed from LR dimers and a further 42 dimer configurations taken from our trial crystal structure predictions. Thus, we contributed to their methodology development by showing the need to consider the dimer interaction of both hands of this molecule, which is only slightly chiral. The final optimisation of the potential was tested using 881 training dimer configurations with attractive interactions. The RMSE of the potentials exp-6_com1 and LL+LR+crys were 4.18 and 0.255 kJ mol^{-1}, respectively. These RMSE's are larger than ideal however, relating to errors in the lattice energy of approximately 10 kJ mol^{-1}, but it was not possible to improve upon these within

the assumed functional form. Ideally the creation of a damped potential was desired, however a form of damped potential which was compatible with DMACRYS could not be created in the time frame for the investigation.
Unlike the FIT potential which tends to be used alongside *CrystalPredictor* and DMACRYS, the custom potential contained a unique potential for each individual atom within XXII, rather than a potential for each atom type (C, N, S). The charges and C6 coefficients were fitted to the potential at long range, and both the *α* and *β* parameters were simultaneously fitted to shorter range SAPT(DFT) points. Hence, exp-6_com1 and LL+LR+crys had a common 15 charges and C6 parameters, and *15* and *105* repulsion parameters, respectively. The difference in the number of repulsion parameters arises because *CrystalPredictor* applies the combining rules and so only the single atom terms need to be produced with the cross-potentials calculated internally by the *CrystalPredictor*. DMACRYS, on the other hand, is not able to internally calculate the cross-terms and so these need to be calculated independently by the user, requiring a greater degree of fitting to generate a potential which can be used with DMACRYS.

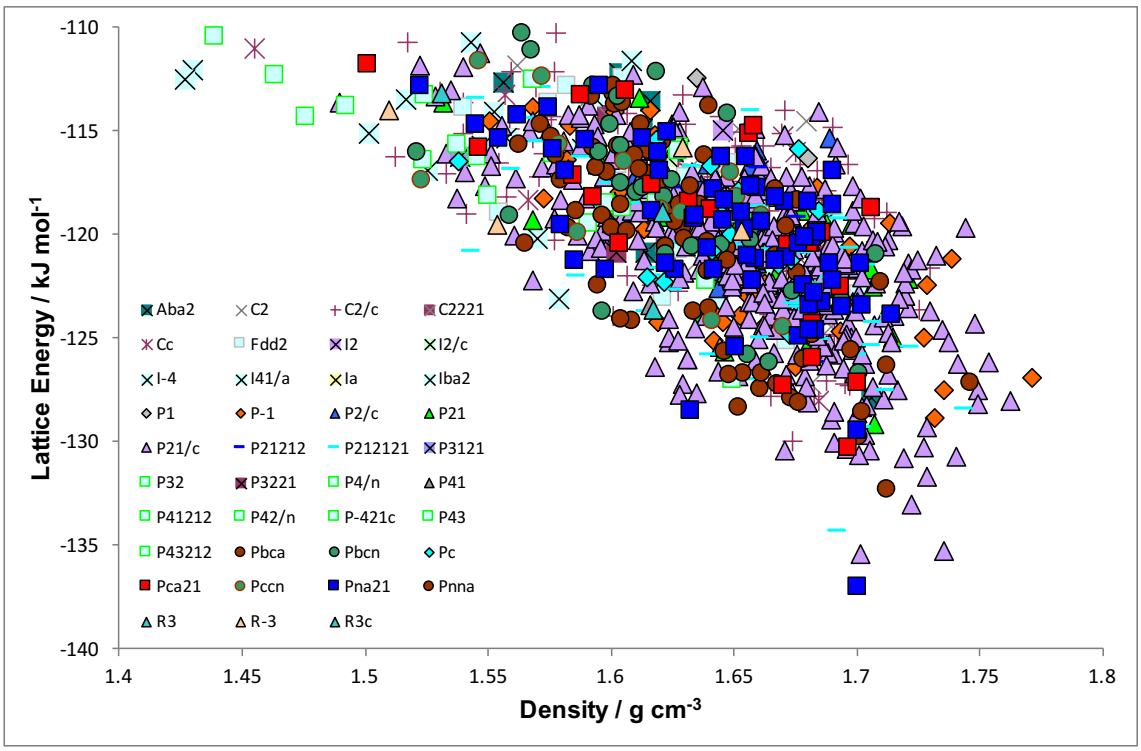


Figure 5-10: Summary of the CSP results for XXII at the CryststalOptimizer stage evaluated using the PBE1PBE/6-31G(d,p) wavefunction and followed by refinement with DMACRYS using the custom isotonic Delaware LL+LR+crys potential. Each point represents a unique crystal structure with its symbol corresponding to the space group it belongs to.

The lowest 100 structures generated from this search were submitted with minimal post-processing due to the time constraints on the task. Much of the

analysis of the crystal energy landscape was done after the submission deadline.

5.4.2 *CrystalPredictor* + DFT-D collaborations

The Tkatchenko group and Gerit Brandenburg had no means by which to generate putative crystal structures which lattice minimisation calculations could be performed on. As a result, the first 1000 structures from my *CrystalPredictor* search with the addition of multipoles were sent to the both of these groups to be used as a starting point for various types of periodic DFT-D calculations that they were developing. The final lists and rankings submitted by these groups were a result of their own post processing analysis as described in the ESI of the Blind Test paper.[8]

5.5 Discussion in Light of Experimental Results

5.5.1 CrystalPredictor + CrystalOptimizer using FIT Potential

Shortly after the submission deadline the experimental structures for each of the systems in the Blind Test were released. The structure of XXII was determined to be in space group *P21/n* with the molecule adopting a bent configuration such that the individual molecules pack together like 'clasped hands'. The full details of the experimental structure can be seen in Table 1, along with the successful predictions of this structure.

Using the crystal packing similarity function within the Mercury program, the experimental crystal structure of XXII was compared with each of the structures submitted within the two lists. The experimental structure was found amongst the submitted structures as hypothetical crystal A4 (i.e. the 4th energetically stable structure after the *CrystalPredictor* stage, see Figure 5.6) from the search. This was ranked 6th in list 1 (1.80 kJ mol^{-1} above the global minimum), and 2nd in list 2 (0.27 kJ mol^{-1} above the global minimum), overlaying with the experimental crystal structure with an RMD_{15} of 0.26 Å and 0.21 Å, respectively. Overlays of the experimental molecular configuration with structure A4 show that the preferred molecular configuration is indeed that determined during the gas phase optimisation of the molecule at the PBE1PBE/6-31G(d,p) level of

theory. There was little deviation from this geometry upon crystallisation as reflected in the refinement with *CrystalOptimizer* (Figure 5.11).

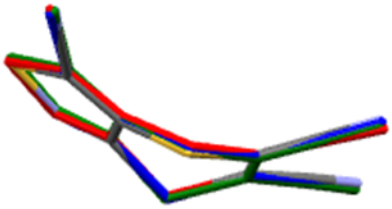


Figure 5-11: Overlays of the experimental structure of XXII with those in the matching structure A4 at different stages of calculation. Experimental: atom colours, *CrystalPredictor* (rigid conformation used for search): red, *CrystalOptimizer* (Halgren sulphur potential, refined conformation): green, *CrystalOptimizer* (Scheraga, suphur potential): blue.

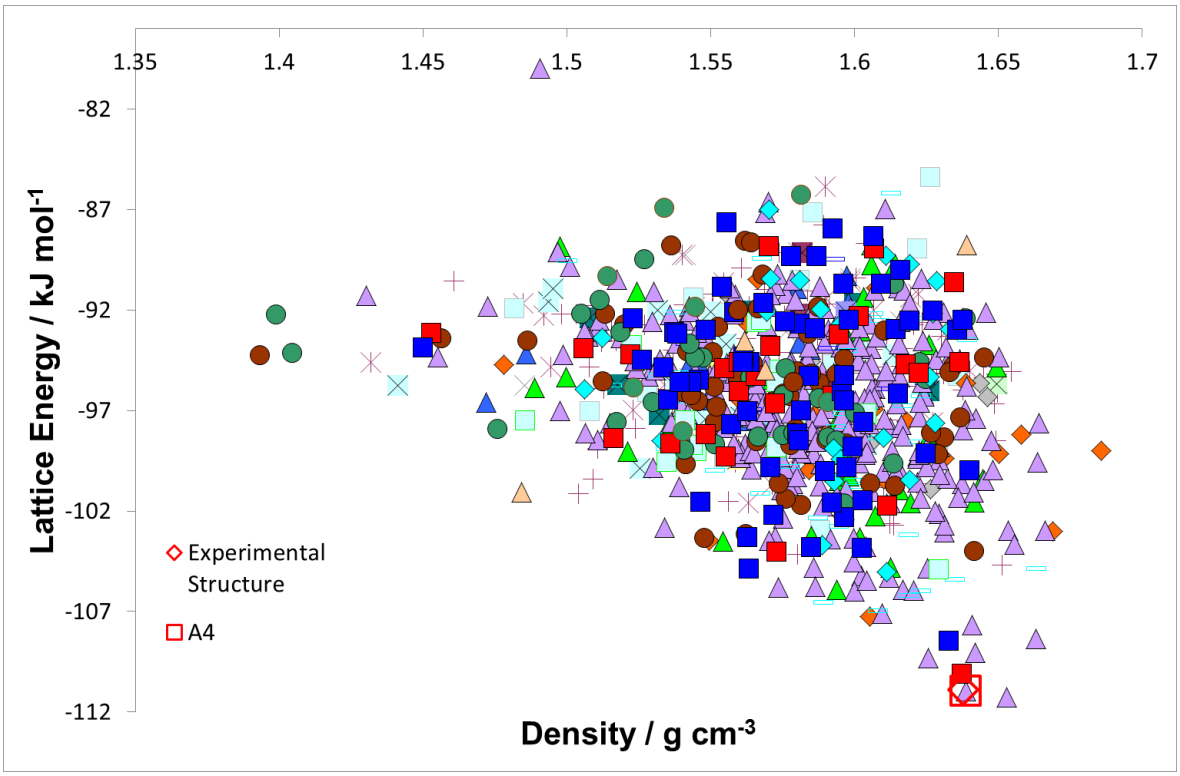


Figure 5-12: Summary of the CSP results for XXII at the *CrystalOptimizer* stage evaluated using the PBE1PBE/6-31G(d,p) wavefunction and the FIT potential with Scheraga sulphur potential with structures modelled in a PCM (ε = 3) and the effect of free energies. This CEL forms the basis of list 2. The experimentally determined structure is highlighted with a red diamond

Structure	Space Group	Cell Parameters						ΔE* / kJ mol^{-1}	Density / g cm^{-3}	RMSD$_{15}$ / Å
		a / Å	b / Å	c / Å	α / °	β / °	γ / °			
Exptl.	P2$_1$/*n*	11.947	6.696	12.598	90	108.6	90	-	1.73	-
Exptl. CO (-S-)	P2$_1$/*n*	12.178	6.636	12.426	90	105.8	90	0	1.71	0.28
A4 CO (-S-)	P2$_1$/*n*	12.432	6.651	12.178	90	106.1	90	−0.264	1.71	0.26
Exptl. PCM + FE (S=O)	P2$_1$/*n*	11.852	6.839	13.669	90	110.0	90	0	1.59	0.42
A4 PCM + FE (S=O)	P2$_1$/*n*	12.200	6.800	12.800	90	107.3	90	−0.049	1.62	0.225

Table 5-1: The cell parameters of the experimental crystal structure and the corresponding minima found either by minimising from the experimental structure or as submitted for the blindtest. The energies of the hypothetical crystal structure are given relative to the experimental structure which has been minimised using the same method. (-S) indicated the Halgren potential and (S=0) the Scheraga potential) N.B The cell dimensions of A4 PCM + FE (S=O) were converted to the same setting using isocif.

The global minimum from list 2, structure A467, was just 0.27 kJ mol^{-1} lower in energy than structure A4 which corresponds to the experimental crystal structure. As can be seen in Figure 5.13, the packing of A467 and the experimental structure show strong similarities. Structure A467 is in space group *P*2$_1$/*c*, and overlays with the experimental structure with an RMSD$_{11}$ of 0.208 Å. While the packing in the two structures is similar on the 2-dimensional plane, the differences in space group reflects a difference in 3D packing. The similarities and small energy difference between the global minimum structure and the experimental structures suggests that a high temperature disordered phase could exist.

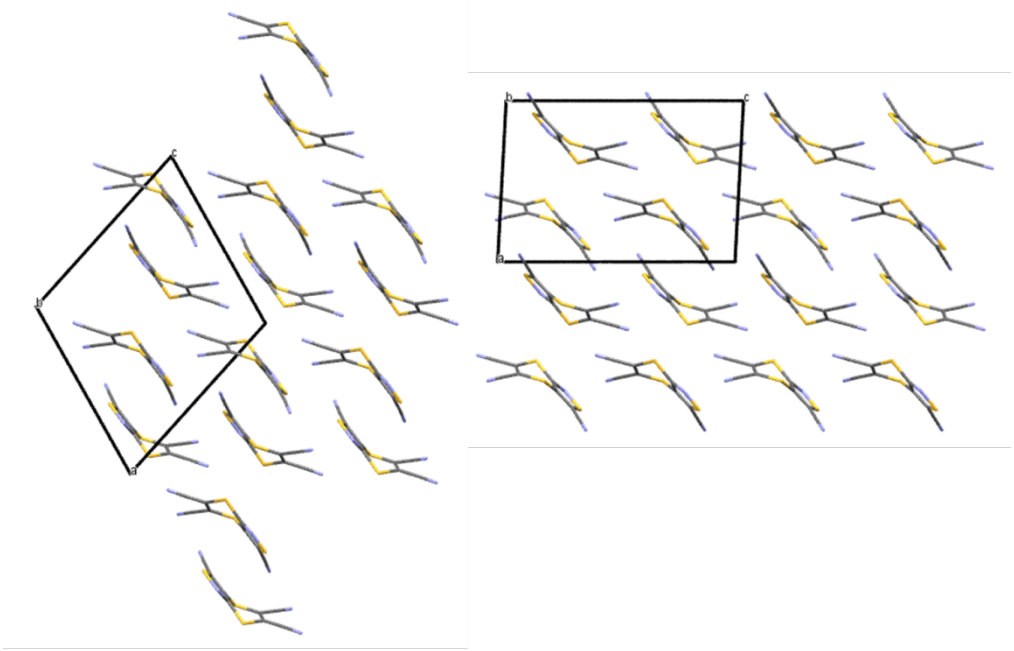


Figure 5-13: Comparisons of the packings in the global minimum structure, A467, from Scheraga+PCM+FE crystal energy landscape (list 2) with the experimentally determined structure

It is interesting to note that it is the Scheraga potential which ranked structure A4 closer to the global minimum than the Halgren potential when molecular flexibility is accounted for. It was expected that the Halgren sulphur potential would model XXII more closely due to the similarities of the environments for the sulphur – sulphur single bonded to two carbon side chains. However, thioazides, which were used as a basis for the developments of the Scheraga potential, have a nitrogen group in close proximity to the sulphur. While XXII contains no SO_2 groups, each sulphur atom does have a nitrogen atom within either one or two bond distances. This could contribute to the Scheraga potential placing the experimental structure closer to the global minimum than the Halgren potential. While the energetic ranking of the structures in the search is better modelled using the Scheraga potential, the density of the XXII is more closely modelled by the Halgren potential (see Table 5.1). This is supported by the $RMSD_{15}$ values which indicated that the experimental structures are reproduced better with the Halgren potential. The structure A4, corresponding to the experimentally determined structure of XXII, was ranked 6th however this was only 1.80 kJ mol^{-1} above the global minimum and so is within the uncertainty of empirically fitted isotropic repulsion-dispersion potentials. Overall, the crude isotropic empirical potentials have performed well in this CSP study, possibly because of how the low energy structures sample the S repulsion.

5.5.2 *CrystalPredictor* + DMACRYS using Custom Delaware Potential

The conformation that had been generated for the rigid search was a good match for the experimental conformation of XXII. Unfortunately the experimental structure was not found within the lowest 100 structures in the joint submission with Szalewicz et al. A search through all structures generated found the experimental structure to correspond to structure 139 (i.e. the 139th energetically stable structure after the *CrystalPredictor* stage, see Figure 5.10) ranked 139th with a lattice energy of -124.04 kJ mol^{-1}, after the *CrystalPredictor* search. The ranking then dropped to 162 following refinement in DMACRYS with the revised custom potential, 12.59 kJ mol^{-1} above the global minimum with a lattice energy of -124.42 kJ mol^{-1}, significantly outside of the 10 kJ mol^{-1} considered to be the energy range of polymorphism.

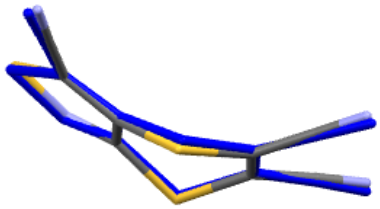


Figure 5-14: Overlay of experimental structure (atom colours) with global minimum from Delaware search (structure 3, blue)

The drop in ranking following the DMACRYS run leads to questions in the accuracy of the second potential formed from LL and LR pairs of dimers. In principle, the removal of the combining rules should have produced a more accurate potential as it fitted the dimer data better. The low ranking of the experimental structure as a whole leads to question on whether the isotropic potential which was provided was able to accurately determine the lattice energies of the crystal structures from the search. This suggests that the directionality in the potential, as given in our model by the distributed multipoles, is important. It is known that the potential used was not the optimal model that the Delaware group had wanted to use and this is shown to be reflected in the results. It was originally hoped that a damped dispersion could be used for the potential. Damped dispersion has never been fully tested in DMACRYS and doubt arose about its implementation following a comparison with the lattice energies given when compared with MD calculations performed in DL_POLY of

the same crystal structures. The MD calculations were performed using the same potential at 1K, and so should be equivalent to the optimisations performed with DMACRYS, however there was a difference in the minimised energies of approximately 15 kJ mol^{-1}. Due to the time constraints in getting results for submission it was not possible to investigate this discrepancy further and so an undamped dispersion had to be used for the calculations. (Since submission, DMACRYS has been tested with damped dispersion on pyridine crystals with success, so this was not the problem.) In collaboration with another group, Szalewicz et al submitted lists of structures in which damped dispersion was used. In this case, the experimental structures came out ranked 3 in the list. Thus, it seems likely that the poor results with the potential that I used was due to the importance of using a damped dispersion in fitting this isotropic potential.

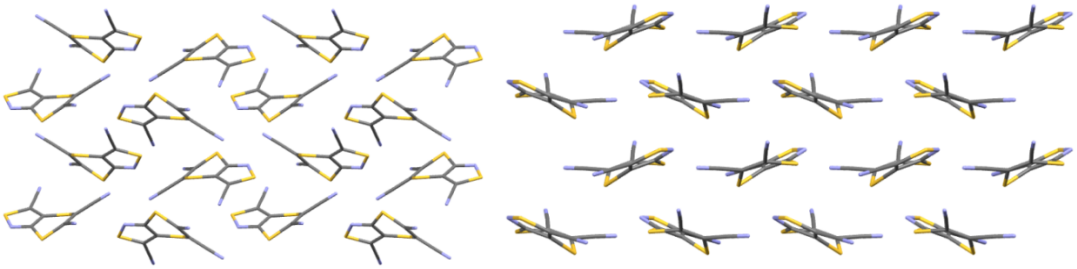


Figure 5-15: The packing motifs of the global minimum structure from the Delaware search: left after CrystalPredictor, right after DMACRYS

The global minimum structures from both the *CrystalPredictor* and DMACRYS stages, with the custom Delaware isotropic potential, of the search can be seen in Figure 5.15. While both packings are observed within the low energy structures from my *CrystalOptimizer* search, no structures within the lowest 100 of the Delaware search had packings similar to that observed in the experimental structure (see Figure 5.15 and Appendix Figure 5.9.1). Many of the motifs observed in the search, particularly within the low energy structures, form sheets of molecule XXII stacked together rather than the clasped hands observed within the experimental structure. There are a significant number of close contacts between the molecules in these sheets which may be being modelled inaccurately. Further analysis revealed two duplicates in the lowest 100 structures which were missed upon clustering (1224 and 267, and 12 and 821 – in 2nd 100: 436 and 1288 duplicates). This would have brought the experimental structure down to rank 167. While this does not make a significant

difference to the overall results, it highlights the lack of time which was available to process the results prior to submission.

After the submission, it was revealed that an error had been made by Szalewicz et al when calculating the additional SAPT(DFT) points used to generate the LL+LR+crys potential used for the DMACRYS run.[8] This could contribute to the drop in ranking of the experimental structure between *CrystalPredictor* and DMACRYS runs. Unfortunately, a corrected potential was not provided in order to determine how large the effect of this error was on the overall ranking of the structures. In conclusion, the time constraints meant that we are unable to reach any firm conclusions about the non-empirical potential from my part of the collaboration. The alternative submission based on the Delaware approach suggests that fitting many isotropic coefficients in a flexible functional form was able to predict the structure successfully.

5.5.3 *CrystalPredictor* + DFT-D methods

For the first list submitted by the Tkatchenko group, they performed DFT-D calculations with PBE wavefunction and a new many body dispersion.[19] This placed structure A4, corresponding to the experimental structure, as rank #3. The second list consists of the lowest 20 structures from list1 with the addition of the harmonic vibrational free energies calculated using the PBE wavefunction and the many body dispersion correction at 300 K. This brought structure A4 down to be the global minimum and only 0.11 kJ mol^{-1} below the second structure, A467, in the landscape. A more significant energy gap can be seen between the global minimum and the structure ranked 3rd of 1.08 kJ mol^{-1}.

The more expensive DFT-D calculations were able to accurately predict the experimental structure to be the most thermodynamically stable structure in the crystal energy landscape over the similarly packed A467. Although this more expensive method was able to accurately predict the correct experimental crystal structure as the global minimum, the small energy difference between structures A4 and A467 is consistent with the observations from my refinements, again suggesting the possibility of a high temperature disordered structure.

Gerit Brandenburg also had the 1000 structures. He reduced the number of structures being considered by the initial energy, thus eliminating many structures that were high up the repulsive wall with his functionals. However, in post analysis, his alternative function TPSS with the D3[20] dispersion correction, with or without a zero-point energy and thermal correction gave the experimental structure as the lowest in energy.

Hence, DFT methods using different dispersion corrections and wavefunctions were able to rank the observed structure closer to the global minimum than ours semi-empirical method. This confirms that energy criteria can predict XXII. However, our much less sophisticated, computationally cheaper method gave comparable results. All DFT-D methods showed that the inclusion of free energies was important in reranking the predicted structure of XXII to the global minimum structure, which supports the observations shown in our submitted results.

5.6 Overall Results for XXII in the Blind Test

Due to the relative simplicity of XXII, 24 of the 25 submissions included an attempt at predicting this crystal structure. Of those, 13 groups were successful, all placing the structure within the top 10 in their list of 100 structures, with 3 groups placing it as their first choice. Even though this molecule was the simplest it shows the difficulty in predicting the structures of even relatively simple, small molecules.

5.6.1 Comparison with other groups

Across the submissions there were a number of different ways in which the generation of the lists of molecules was tackled. Some groups generated an optimised version of XXII and then used this structure to perform a rigid search. Some groups took into account the pseudo-chiral nature of the molecule and used two different conformations, one planar and one bent, using both conformations of the molecule to perform a rigid search. A handful of groups performed a rigid search and then factored in flexibility at a later stage when the molecule was optimised. Approximately half of the submissions treated XXII as fully flexible in their search. There is no link between whether XXII was treated as rigid or flexible and the success rate of the predictions put forward for the Blind Test, with approximately equal numbers of successes and failures across

each method. Two of the groups who obtained the experimental structure as the global minimum structure performed a fully rigid search while the third performed a rigid search followed by a flexible search minimisation and the addition of Helmholtz free energies. It is also noticed that any groups which entered two lists, one containing an estimate of the free energies and one without, the list which contained the free energies placed the experimental structure at a lower ranking than the list that did not include this estimate. This suggests that the vibrational contribution to the lattice energy is of importance for XXII – vibrational motions affecting the many different ways in which the molecules can pack together. The importance of including the free energies when considering polymorphs with negligible energy difference has shown to be important, particularly in the case of aspirin.[21]

5.6.2 Comparison with closely related methods by other groups

The Pantelides and Adjiman group used the same *CrystalPredictor* program to generate the crystal structures as was used in my study of XXII, however there are slight differences between the methods which has led to a difference in the structures produced and their relative ranking. The Pantelides and Adjiman group used the M06 functional as opposed to the PBE1PBE used in my study. Additionally, a custom potential with a new set of parameters for all atoms involved was used to refine the structures as opposed to the FIT potential with Halgren sulphur parameters. The best comparison between the two lists would be to compare the *CrystalPredictor* + DMACRYS structures, prior to being reoptimised with *CrystalOptimizer*, with the structure submitted by the Pantelides and Adjiman groups. The differences in potential and functional are shown in the differences in ranking of the structures.

Structure 3 from the Pantelides and Adjiman set of structures was not found in amongst the structures generated in my search. A *CrystalOptimizer* refinement using the parameters used in my search placed the energy of this structure at −96.4 kJ mol^{-1}, approximately 11 kJ mol^{-1} higher in energy than the global minimum structure. Even if it had been found in my search it would not have been competitive in energy. However, it shows very similar packing to A467, A4 (both have an $RMSD_{11}$ of 0.139 Å with structure 3) and therefore the experimental crystal structure albeit being in a different space group.

Upon application of the free energies structure A1, which was determined to be the global minimum structure for the Pantelides and Adjiman group, was determined to be unstable at room temperature. All other structures within the lowest 100 structures which adopted the same structure as A1 were also found to be unstable once the free energies were applied suggesting that XXII would not readily form a crystal structure in this manner.

Even though similar techniques were used to determine the lists submitted for the Blind Test, the differences in the two submissions show how experience, knowledge and analysis of crystal energy landscape by the researcher can have an impact on the final outcome.

5.7 Conclusion

Molecule XXII was successfully predicted from the molecular diagram, with the experimental form appearing in both lists of structures submitted to the CCDC. While there were no molecules within the CCDC database to test the sulphur potential prior to generation of the crystal energy landscape, the models chosen for this study were able to accurately reproduce the experimental crystal structure.

The results for XXII across all submissions for the Blind Test indicate that the addition of free energies lowers the ranking of the experimental structure within the crystal energy landscape, suggesting that the vibrational energy is an important consideration with regards to XXII. The collaboration with Szalewicz group highlights the importance of using an anisotropic potential for this molecule, or alternatively the need for the extra flexibility in using a damping of the dispersion term when an isotropic potential is used.

Molecule XXII was sufficiently small that a wide range of different methods could be applied. Our search method and relatively simple potential were adequate to find the experimental structure close to the global minimum. While more complex DFT-D based methods were able to determine the experimental structure of XXII as the most stable, our comparatively far computationally cheaper method was able to provide comparable results. There was no reason

to believe that this molecule was not able to crystallise in the most thermodynamically stable form.

The 6th Blind Test provided a useful means for assessing the current strengths and weaknesses of crystal structure prediction methods; whilst XXII was within the capabilities of all participants, there were considerably fewer entrants, let alone successes for the other molecules. Nonetheless, the 6th Blind Test showed that there have been considerable advances in the complexity of the molecular systems that can be tackled.

5.8 References

(1) Motherwell, W. D. S.; Ammon, H. L.; Dunitz, J. D.; Dzyabchenko, A.; Erk, P.; Gavezzotti, A.; Hofmann, D. W. M.; Leusen, F. J. J.; Lommerse, J. P. M.; Mooij, W. T. M.; Price, S. L.; Scheraga, H.; Schweizer, B.; Schmidt, M. U.; van Eijck, B. P.; Verwer, P.; Williams, D. E. Crystal structure prediction of small organic molecules: a second blind test. *Acta Crystallographica Section B - Structural Science* **2002**, *58*, 647-661.

(2) Hulme, A. T.; Johnston, A.; Florence, A. J.; Fernandes, P.; Shankland, K.; Bedford, C. T.; Welch, G. W. A.; Sadiq, G.; Haynes, D. A.; Motherwell, W. D. S.; Tocher, D. A.; Price, S. L. Search for a predicted hydrogen bonding motif - A multidisciplinary investigation into the polymorphism of 3-azabicyclo[3.3.1]nonane-2,4-dione. *Journal of the American Chemical Society* **2007**, *129*, 3649-3657.

(3) Braun, D. E.; Bhardwaj, R. M.; Florence, A. J.; Tocher, D. A.; Price, S. L. Complex Polymorphic System of Gallic Acid-Five Monohydrates, Three Anhydrates, and over 20 Solvates. *Crystal Growth & Design* **2013**, *13*, 19-23.

(4) Tsui, H. H. Y.; Price, S. L. A non-empirical method of determining atom-atom repulsion parameters: application to the crystal structure prediction of an oxyboryl derivative. *CrystEngComm* **1999**, *1*, 7.

(5) Neumann, M. A. Tailor-made force fields for crystal-structure prediction. *Journal of Physical Chemistry B* **2008**, *112*, 9810-9829.

(6) Misquitta, A. J.; Welch, G. W. A.; Stone, A. J.; Price, S. L. A first principles solution of the crystal structure of C6Br2ClFH2. *Chemical Physics Letters* **2008**, *456*, 105-109.

(7) Szalewicz, K. Determination of Structure and Properties of Molecular Crystals from First Principles. *Accounts of Chemical Research* **2014**, *47*, 3266-3274.

(8) Reilly, A. M.; Cooper, R. I.; Adjiman, C. S.; Bhattacharya, S.; Boese, A. D.; Brandenburg, J. G.; Bygrave, P. J.; Bylsma, R.; Campbell, J. E.; Car, R.; Case, D. H.; Chadha, R.; Cole, J. C.; Cosburn, K.; Cuppen, H. M.; Curtis, F.; Day, G. M.; DiStasio Jr, R. A.; Dzyabchenko, A.; van Eijck, B. P.; Elking, D. M.; van den Ende, J. A.; Facelli, J. C.; Ferraro, M. B.; Fusti-Molnar, L.; Gatsiou, C.-A.; Gee, T. S.; de Gelder, R.; Ghiringhelli, L. M.; Goto, H.; Grimme, S.; Guo, R.; Hofmann, D. W. M.; Hoja, J.; Hylton, R. K.; Iuzzolino, L.;

Jankiewicz, W.; de Jong, D. T.; Kendrick, J.; de Klerk, N. J. J.; Ko, H.-Y.; Kuleshova, L. N.; Li, X.; Lohani, S.; Leusen, F. J. J.; Lund, A. M.; Lv, J.; Ma, Y.; Marom, N.; Masunov, A. E.; McCabe, P.; McMahon, D. P.; Meekes, H.; Metz, M. P.; Misquitta, A. J.; Mohamed, S.; Monserrat, B.; Needs, R. J.; Neumann, M. A.; Nyman, J.; Obata, S.; Oberhofer, H.; Oganov, A. R.; Orendt, A. M.; Pagola, G. I.; Pantelides, C. C.; Pickard, C. J.; Podeszwa, R.; Price, L. S.; Price, S. L.; Pulido, A.; Read, M. G.; Reuter, K.; Schneider, E.; Schober, C.; Shields, G. P.; Singh, P.; Sugden, I. J.; Szalewicz, K.; Taylor, C. R.; Tkatchenko, A.; Tuckerman, M. E.; Vacarro, F.; Vasileiadis, M.; Vazquez-Mayagoitia, A.; Vogt, L.; Wang, Y.; Watson, R. E.; de Wijs, G. A.; Yang, J.; Zhu, Q.; Groom, C. R. Report on the sixth blind test of organic crystal structure prediction methods. *Acta Crystallographica Section B* **2016**, *72*, 439-459.

(9) Coombes, D. S.; Price, S. L.; Willock, D. J.; Leslie, M. Role of Electrostatic Interactions in Determining the Crystal Structures of Polar Organic Molecules. A Distributed Multipole Study. *Journal of Physical Chemistry* **1996**, *100*, 7352-7360.

(10) Day, G. M.; Cooper, T. G.; Cruz-Cabeza, A. J.; Hejczyk, K. E.; Ammon, H. L.; Boerrigter, S. X. M.; Tan, J.; Della Valle, R. G.; Venuti, E.; Jose, J.; Gadre, S. R.; Desiraju, G. R.; Thakur, T. S.; van Eijck, B. P.; Facelli, J. C.; Bazterra, V. E.; Ferraro, M. B.; Hofmann, D. W. M.; Neumann, M.; Leusen, F. J. J.; Kendrick, J.; Price, S. L.; Misquitta, A. J.; Karamertzanis, P. G.; Welch, G. W. A.; Scheraga, H. A.; Arnautova, Y. A.; Schmidt, M. U.; van de Streek, J.; Wolf, A.; Schweizer, B. Significant progress in predicting the crystal structures of small organic molecules - a report on the fourth blind test. *Acta Crystallographica Section B - Structural Science* **2009**, *65*, 107-125.

(11) Tremayne, M.; Grice, L.; Pyatt, J. C.; Seaton, C. C.; Kariuki, B. M.; Tsui, H. H. Y.; Price, S. L.; Cherryman, J. C. Characterization of complicated new polymorphs of chlorothalonil by X-ray diffraction and computer crystal structure prediction. *Journal of the American Chemical Society* **2004**, *126*, 7071-7081.

(12) Schaftenaar, G.; Noordik, J. H. Molden: a pre- and post-processing program for molecular and electronic structures. *Journal of Computer-Aided Molecular Design* **2000**, *14*, 123-134.

(13) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr.; Vreven, T.; Kudin, K. N.;

Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A.: Gaussian 03. Gaussian Inc.: Wallingford CT, 2004.

(14) Habgood, M.; Sugdan, I. J.; Kazantsev, A. V.; Adjiman, C. S.; Pantelides, C. Efficient Handling of Molecular Flexibility in Ab Initio Generation of Crystal Structures. *Journal of Chemical Theory and Computation* **2015**, *11*, 1957-1969.

(15) Price, S. L.; Leslie, M.; Welch, G. W. A.; Habgood, M.; Price, L. S.; Karamertzanis, P. G.; Day, G. M. Modelling Organic Crystal Structures using Distributed Multipole and Polarizability-Based Model Intermolecular Potentials. *Physical Chemistry Chemical Physics* **2010**, *12*, 8478-8490.

(16) Kazantsev, A. V.; Karamertzanis, P. G.; Adjiman, C. S.; Pantelides, C. C.: CrystalOptimizer. An efficient Algorithm for Lattice Energy Minimisation of Organic Crystal using Isolated-Molecule Quantum Mechanical Calculations. In *Molecular System Engineering*; Adjiman, C. S., Galindo, A., Eds.; Process Systems Engineering; WILEY-VCH Verlag GmbH & Co.: Weinheim, 2010; Vol. 6; pp 1-42.

(17) Stone, A. J.: GDMA: A Program for Performing Distributed Multipole Analysis of Wave Functions Calculated Using the Gaussian Program System. 2.2 ed.; University of Cambridge: Cambridge, United Kingdom, 2010.

(18) Leslie, M. DL_MULTI - A molecular dynamics program to use distributed multipole electrostatic models to simulate the dynamics of organic crystals. *Molecular Physics* **2008**, *106*, 1567-1578.

(19) Tkatchenko, A.; DiStasio, R. A. J.; Car, R.; Scheffler, M. Accurate and efficient method for many-body van der Waals interactions. *Physical Review Letters* **2012**, *108*, 236402-236402.

(20) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *The Journal of Chemical Physics* **2010**, *132*, 154104.

(21) Reilly, A. M.; Tkatchenko, A. Role of Dispersion Interactions in the Polymorphism and Entropic Stabilization of the Aspirin Crystal. *Physical Review Letters* **2014**, *113*.

5.9 Appendix

5.9.1 Lists of structures submitted to the CCDC

Structure	Space Group	Z	a /Å	b /Å	c /Å	α /°	β /°	γ /°	Lattice Energy /kJmol^{-1}
A467	$P2_1/c$	4	9.8296	6.5977	14.7253	90	91.368	90	-107.987
A8	$Pna2_1$	4	13.1204	10.1993	7.1357	90	90	90	-107.877
A1	$P2_1/c$	4	6.0892	7.9691	20.1613	90	75.869	90	-107.821
A5293	$P2_12_12_1$	4	13.0395	7.1738	10.2161	90	90	90	-106.851
A446	$Pca2_1$	8	6.7439	14.5924	19.5986	90	90	90	-106.656
A4	$P2_1/n$	4	12.4318	6.6508	12.1782	90	73.882	90	-106.188
A645	$P2_1/c$	4	8.3514	8.0089	14.5098	90	99.246	90	-105.773
A22	$P2_12_12_1$	4	9.9234	11.8411	8.2534	90	90	90	-105.056
A519	$P2_1/c$	4	9.7425	10.7683	9.4341	90	85.089	90	-104.836
A43	$P2_1/c$	4	10.9627	6.9471	12.6878	90	83.833	90	-104.812
A67	P-1	2	9.7272	7.6452	6.4511	94.307	95.606	78.548	-104.599
A98	$P2_1/c$	4	7.0108	11.548	13.6093	90	62.288	90	-104.414
A5160	$Fdd2$	4	14.3222	39.799	6.8881	90	90	90	-104.123
A2846	$P2_1/n$	4	9.9532	9.3425	10.6024	90	81.495	90	-104.122
A134	$P2_1/c$	4	7.57	6.6547	19.2278	90	92.015	90	-103.769
A818	$C2/c$	4	21.7518	9.9987	10.5652	90	55.901	90	-103.736
A1393	$P2_12_12_1$	4	13.5458	6.8709	10.0917	90	90	90	-103.618
A3230	$P2_1/n$	4	9.6586	9.9005	9.9328	90	77.472	90	-103.617
A5146	$P2_12_12_1$	4	19.4359	7.4583	6.6997	90	90	90	-103.54
A2848	$P2_1/c$	4	9.0198	9.9044	12.0895	90	116.617	90	-103.275
A58	$P2_12_12_1$	4	6.743	12.0025	11.8658	90	90	90	-103.262
A394	$P4_12_12$	8	6.7359	6.7359	42.4348	90	90	90	-103.17

A279	$P2_12_12_1$	4	4.0387	18.8179	12.6632	90	90	90	-103.121
A124	$P2_1/c$	4	6.0606	19.4371	8.0792	90	87.497	90	-102.923
A641	$P2_1/c$	4	9.8739	9.925	12.0855	90	51.829	90	-102.895
A7	*Pbca*	8	13.7123	7.1081	19.6202	90	90	90	-102.764
A40	$P2_1/c$	4	10.1009	7.7856	14.9199	90	54.29	90	-102.712
A83	$P2_1/c$	4	9.9717	7.1466	13.7604	90	78.388	90	-102.551
A9315	$P2_1/n$	4	14.7069	9.9521	8.4122	90	50.999	90	-102.432
A442	$P2_1$	4	12.0575	11.9646	6.7417	90	90	93.588	-102.398
A160	$P2_1/c$	4	3.7609	27.0741	9.3291	90	83.418	90	-102.355
A3719	$Pna2_1$	4	12.2232	19.4295	4.1139	90	90	90	-102.253
A464	$P2_1/c$	4	11.2548	7.1761	13.2371	90	63.112	90	-102.241
A78	$P2_1$	2	7.0872	6.6443	10.2262	90	85.448	90	-102.161
A165	*P*-1	2	6.9215	29.8696	19.3998	60.033	114.063	54.676	-101.98
A381	*P*-1	2	6.9936	12.1166	10.4355	116.181	86.771	45.347	-101.98
A84	$P2_12_12_1$	4	6.0507	19.7799	8.3855	90	90	90	-101.966
A131	*C*2/*c*	4	18.3492	6.5725	16.191	90	92.392	90	-101.966
A164	*P*-1	2	7.8912	9.7339	6.922	107.222	71.188	96.515	-101.933
A133	$P2_1$	4	6.9135	13.3109	11.9228	90	59.85	90	-101.881
A244	$P2_1/c$	4	6.8954	11.6079	11.984	90	94.71	90	-101.774
A53	$P2_1/c$	4	7.9802	6.044	20.0398	90	94.938	90	-101.763
A32	$P2_1$	2	7.5647	6.5857	9.8054	90	90.979	90	-101.703
A389	*P*-1	2	7.1483	10.5546	6.8262	103.218	80.477	93.608	-101.607
A177	$P2_1/c$	4	10.3133	7.4569	12.771	90	79.219	90	-101.539
A196	$P2_1/c$	4	12.1605	6.6673	14.3352	90	56.814	90	-101.509
A752	$P2_1$	4	6.9591	11.8289	13.4874	90	60.417	90	-101.469
A282	$P2_1/c$	4	3.9876	18.7092	13.1132	90	78.585	90	-101.441
A96	$P2_1/n$	4	7.5164	18.775	6.9027	90	95.421	90	-101.405

A817	$P2_1/n$	4	6.8906	20.3495	6.9245	90	76.043	90	-101.395
A23	$Pca2_1$	4	13.644	9.8661	7.1875	90	90	90	-101.349
A738	Pa	4	19.2645	7.477	6.7407	90	90.134	90	-101.223
A36	$P2_1/c$	4	6.1388	8.1557	20.3674	90	76.167	90	-101.205
A253	$P1$	2	17.5641	6.7147	8.8538	67.735	87.042	89.989	-101.197
A488	$Pbcn$	8	6.9358	11.4861	24.6156	90	90	90	-101.172
A1772	$P2_1/c$	4	6.9466	7.1897	22.7023	90	120.097	90	-101.156
A128	$P2_1/c$	4	11.2582	6.959	13.8581	90	61.721	90	-101.137
A664	$P2_1/c$	4	9.6291	6.6635	15.4628	90	97.064	90	-101.124
A19	$P2_1/n$	4	12.4622	6.6497	11.882	90	81.963	90	-101.114
A424	$P2_1$	2	3.815	12.5911	9.8757	90	79.388	90	-101.1
A1898	$C2/c$	4	16.0303	6.7477	18.1244	90	87.299	90	-101.065
A1809	$P2_1/c$	4	6.8698	10.0399	16.0507	90	58.625	90	-101.064
A76	$P2_1/c$	4	12.0696	6.8836	13.1543	90	62.912	90	-101.064
A103	$P2_12_12_1$	4	7.0955	6.8125	20.0522	90	90	90	-101.045
A5931	$Pna2_1$	4	20.0777	7.1007	6.8166	90	90	90	-101.025
A178	$P2_1/c$	4	10.3387	13.6831	7.3654	90	73.012	90	-101.01
A2464	$P2_1$	2	9.7161	4.0671	12.2493	90	86.946	90	-100.962
A1338	$Fdd2$	4	18.7857	54.9391	3.683	90	90	90	-100.951
A587	$Pbca$	8	13.589	7.5244	18.8044	90	90	90	-100.938
A492	$P2_12_12_1$	4	6.5167	9.5978	15.7003	90	90	90	-100.923
A10	$P2_1/n$	4	7.4873	15.6186	8.2156	90	87.867	90	-100.875
A190	$P2_1/n$	4	7.5361	18.6424	6.9307	90	95.436	90	-100.874
A1874	$Pna2_1$	4	17.217	9.3447	6.1376	90	90	90	-100.869
A2051	$P2_1/c$	4	12.0525	4.1419	19.3926	90	93.123	90	-100.834
A322	$P2_1/c$	4	7.4563	6.8712	18.6852	90	95.38	90	-100.713
A3213	$P2_1/n$	4	11.8201	7.2298	11.6805	90	70.734	90	-100.683

A300	*Cc*	2	3.6992	27.3982	9.591	90	76.938	90	-100.666
A95	$Pna2_1$	4	12.3626	9.9793	8.1245	90	90	90	-100.587
A347	$P2_12_12_1$	4	9.2782	26.6963	3.7871	90	90	90	-100.577
A573	$P2_1/c$	4	7.2841	6.3466	20.2686	90	86.1	90	-100.572
A1663	$P2_1$	2	7.0439	7.1435	10.7621	90	63.68	90	-100.55
A298	$P2_1/c$	4	6.1379	20.2492	10.0145	90	54.04	90	-100.531
A405	*C*2/*c*	4	20.842	9.9041	10.8615	90	119.94	90	-100.493
A426	$P2_12_12_1$	4	7.3515	6.8144	19.5887	90	90	90	-100.461
A574	$P2_1/c$	4	7.2044	6.6153	20.4412	90	84.789	90	-100.423
A238	$P2_1/c$	4	6.5029	9.7637	14.8764	90	92.114	90	-100.419
A82	$Pca2_1$	4	19.4128	5.2831	9.8963	90	90	90	-100.379
A580	$P2_1/c$	4	7.3426	6.8627	19.5648	90	88.574	90	-100.377
A38	$P2_1/c$	4	8.7958	10.3776	10.3508	90	90.756	90	-100.363
A540	*C*2/*c*	4	18.6304	6.8434	15.2573	90	89.35	90	-100.346
A161	$P2_1/c$	4	12.4625	6.7129	14.7304	90	52.777	90	-100.334
A104	$P2_1/c$	4	6.9234	7.1235	19.6749	90	97.807	90	-100.277
A29	$P2_1/c$	4	12.0048	6.06	15.5067	90	58.396	90	-100.27
A8608	$P2_1$	2	13.1995	6.2003	7.3286	90	53.948	90	-100.255
A4355	*Pbca*	8	13.4285	7.348	20.4299	90	90	90	-100.249
A1515	*P*-1	2	10.3358	6.9259	8.5544	54.507	107.358	96.578	-100.228
A224	$Pna2_1$	4	9.268	26.2938	3.8396	90	90	90	-100.139
A325	*Pbca*	8	6.9914	11.6036	24.557	90	90	90	-100.133
A981	*P1*	1	3.8266	6.6368	9.7648	97.762	79.32	77.916	-100.126
A5857	$P2_1/a$	4	10.4371	20.2468	6.8907	90	41.352	90	-100.084

Appendix Table 5.9.1: The 100 low energy structures from the CrystalOptimizer serach using the Halgren sulphur potential which comprised the structures submitted as List 1

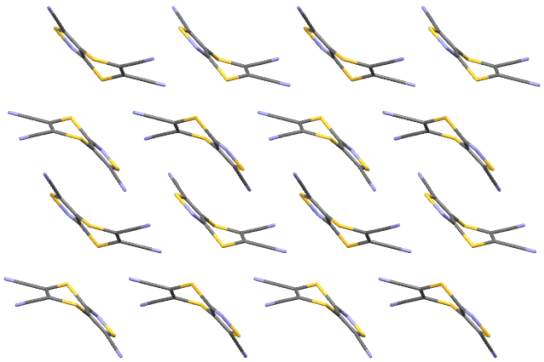
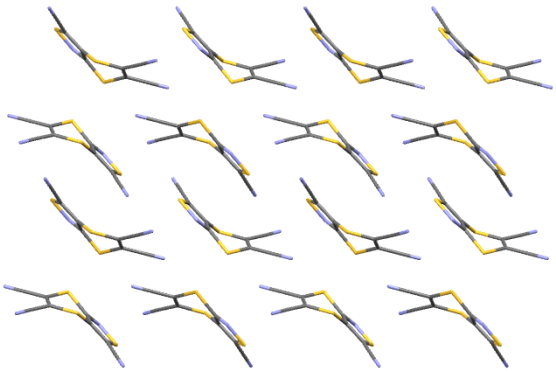
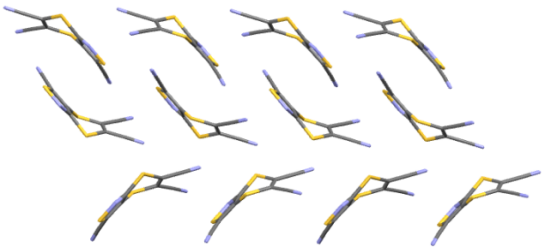
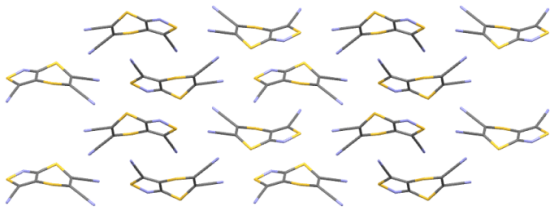
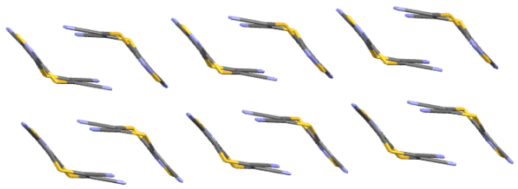
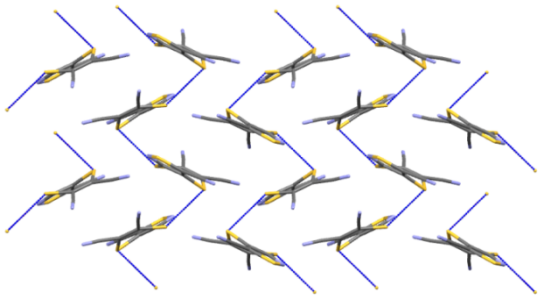
Structure	Space Group	Z	a /Å	b /Å	c /Å	α /°	β /°	γ /°	Lattice Energy /kJmol^{-1}
A467	$P2_1/c$	4	10.0336	6.7252	14.8097	90	93.049	90	-135.012
A4	$P2_1/n$	4	12.7601	6.7887	12.1682	90	72.729	90	-134.742
A172	$Pca2_1$	8	6.839	14.5446	20.2526	90	90	90	-134.569
A43	$P2_1/c$	4	11.1212	7.0832	12.8062	90	84.74	90	-133.936
A98	$P2_1/c$	4	7.0845	11.7227	13.7915	90	62.347	90	-133.121
A8	$Pna2_1$	4	13.3432	10.2427	7.3911	90	90	90	-131.982
A389	P-1	2	7.2211	10.4226	6.988	100.149	83.735	94.422	-131.672
A76	$P2_1/c$	4	12.342	7.0112	13.2669	90	62.448	90	-131.472
A8308	$Pca2_1$	4	20.742	7.9297	6.4145	90	90	90	-131.286
A84	$P2_12_12_1$	4	6.2734	19.7872	8.3604	90	90	90	-131.191
A279	$P2_12_12_1$	4	4.2232	18.8818	12.7754	90	90	90	-131.148
A394	$P4_12_12$	8	6.8055	6.8055	43.7108	90	90	90	-130.973
A22	$P2_12_12_1$	4	9.8988	12.4201	8.342	90	90	90	-130.887
A298	$P2_1/c$	4	6.2968	20.4181	10.1785	90	52.614	90	-130.649
A36	$P2_1/c$	4	6.3613	8.1929	20.3077	90	76.534	90	-130.324
A471	$P2_1/c$	4	7.101	20.6258	8.4915	90	55.765	90	-130.199
A4800	$P2_1/c$	4	7.3548	19.7143	9.8522	90	45.265	90	-129.945
A34	$Pna2_1$	4	8.8083	19.7916	5.9692	90	90	90	-129.941
A450	Pn	4	12.3237	6.8267	12.7902	90	72.056	90	-129.266
A41	$P2_1/c$	4	6.8387	6.9313	21.7641	90	88.123	90	-129.071
A824	$P2_1$	4	7.3736	22.3497	6.4396	90	89.977	90	-128.878
A2511	Pc	4	7.28	6.8636	20.799	90	87.421	90	-128.851
A82	$Pca2_1$	4	19.2465	6.6257	8.2225	90	90	90	-128.795
A21	$P2_12_12_1$	4	12.1202	12.522	6.8458	90	90	90	-128.535
A250	$P2_12_12_1$	4	8.6639	6.0305	20.1412	90	90	90	-128.442
A14	$Pna2_1$	4	12.1154	12.2819	6.9159	90	90	90	-128.125

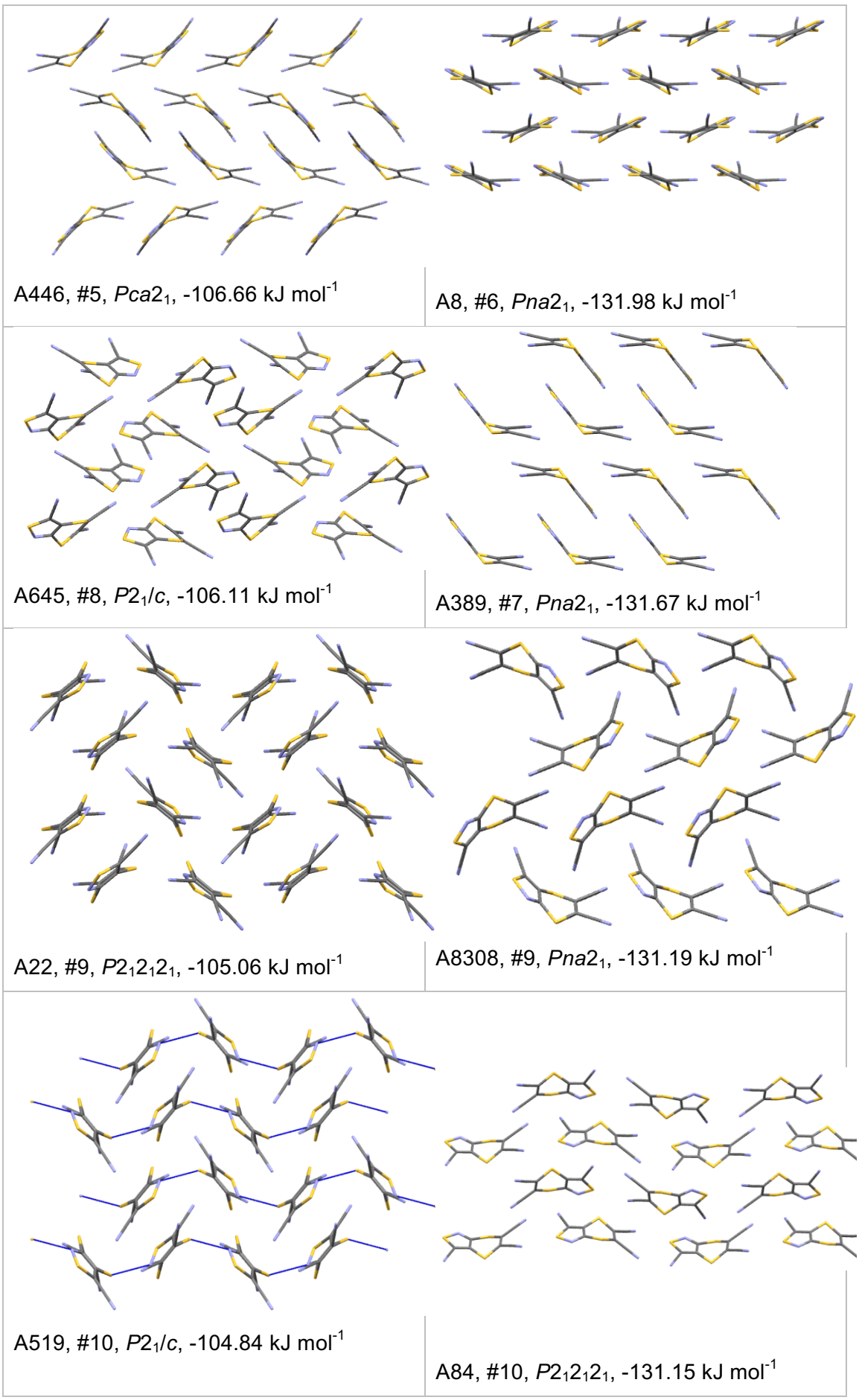
A1598	$P2_1$	4	4.4924	25.7193	9.7729	90	66.257	90	-128.015
A2502	$Pna2_1$	4	25.8344	9.8092	4.1185	90	90	90	-127.9
A143	*Pbca*	8	20.2675	13.319	7.8217	90	90	90	-127.647
A6650	*Pbca*	8	13.9666	7.4735	19.2507	90	90	90	-127.559
A5945	$P2_1/c$	4	7.2506	13.7733	10.5173	90	89.88	90	-127.488
A218	$P2_1$	2	9.6693	8.9016	6.0894	90	83.057	90	-127.435
A1991	$P2_1/c$	4	4.1344	18.8567	14.0734	90	112.335	90	-127.249
A7940	$P2_12_12_1$	4	4.8434	24.8916	8.5767	90	90	90	-127.183
A191	$P2_1/c$	4	4.4529	10.0289	22.6883	90	87.451	90	-126.997
A935	*Pc*	4	21.2211	8.4276	11.5781	89.516	90	90	-126.935
A198	$P2_12_12_1$	4	24.6243	4.2175	9.9519	90	90	90	-126.881
A423	$Pna2_1$	4	7.5929	24.6775	5.6916	90	90	90	-126.805
A858	$Pna2_1$	4	8.5866	25.0934	4.786	90	90	90	-126.694
A333	$P2_1/c$	4	9.2354	10.1266	10.8471	90	85.393	90	-126.396
A56	$P2_1/c$	4	10.9727	6.4623	14.65	90	89.355	90	-126.392
A3645	$P2_1/c$	4	10.3318	13.8283	7.6885	90	107.262	90	-126.285
A682	*C2/c*	4	14.1426	13.4044	13.7226	90	53.365	90	-126.171
A2476	*Cc*	4	19.9292	7.0669	14.9065	90	90.583	90	-126.042
A713	$P2_1/n$	4	16.691	4.0774	16.2274	90	69.374	90	-125.837
A5541	$Pna2_1$	4	12.7791	20.3949	4.1215	90	90	90	-125.829
A414	$Pna2_1$	4	22.2454	7.1326	6.4849	90	90	90	-125.759
A3115	$P2_1/c$	4	10.9537	7.0757	13.7394	90	79.5	90	-125.756
A1032	*I2/a*	8	19.4309	4.2206	25.3308	90	89.481	90	-125.687
A2801	*P*-1	2	7.2889	9.7331	8.8888	95.306	74.048	64.756	-125.599
A1138	*Pc*	4	9.73	4.2031	25.3234	90	90.843	90	-125.593
A1855	*Pbca*	8	18.8531	9.4149	11.7515	90	90	90	-125.517
A188	$P2_1$	4	7.0738	21.1425	8.3517	90	124.103	90	-125.462

A200	$P2_1$	2	7.2003	7.4516	9.7434	90	77.563	90	-125.402
A489	$P2_1/n$	4	12.73	6.8183	11.8646	90	83.303	90	-125.334
A8685	*A2/n*	8	13.7074	15.3567	12.7405	90	52.227	90	-125.28
A461	*Cc*	4	15.4002	10.9832	13.1318	90	108.167	90	-125.265
A3726	$P2_1/n$	4	13.6573	10.5843	7.0058	90	82.83	90	-125.246
A621	*C2/c*	4	19.1748	4.0709	26.8909	90	90.814	90	-125.237
A347	$P2_12_12_1$	4	9.7156	25.2122	4.1497	90	90	90	-125.142
A7686	$P2_1/c$	4	11.9306	6.9926	14.3851	90	120.125	90	-125.137
A825	$P2_1/c$	4	6.0581	8.3144	22.7648	90	63.593	90	-125.108
A239	$P2_1/c$	4	9.8843	11.2791	11.1688	90	55.294	90	-125.08
A337	$Pna2_1$	4	22.6115	4.5352	10.033	90	90	90	-124.991
A4108	$P2_12_12_1$	4	7.2955	9.7304	14.5158	90	90	90	-124.962
A127	$P2_12_12_1$	4	9.147	19.1887	5.8524	90	90	90	-124.929
A3	*C2/c*	4	18.5266	8.0392	13.9974	90	101.296	90	-124.916
A815	$P2_1$	4	7.1088	13.9072	10.3379	90	83.072	90	-124.874
A80	*Cc*	4	15.156	6.9895	19.5475	90	88.788	90	-124.858
A793	*Fdd2*	4	53.2324	19.3543	4.1191	90	90	90	-124.809
A952	*P42/n*	8	9.2396	9.2396	25.1776	90	90	90	-124.754
A1384	*Ia*	8	14.8371	7.1047	20.0668	90	90.048	90	-124.733
A3102	$P2_1/c$	4	6.9771	6.8919	22.1826	90	96.048	90	-124.655
A537	$P2_12_12_1$	4	6.8184	22.5369	6.9179	90	90	90	-124.632
A959	*C2/c*	4	25.7386	4.1826	19.4989	90	90.103	90	-124.62
A958	$P2_1/c$	4	9.6889	9.6587	11.2148	90	86.556	90	-124.57
A1134	$Pca2_1$	4	24.9532	5.929	7.1679	90	90	90	-124.55
A201	$P2_1/c$	4	7.1157	13.0188	10.9898	90	81.979	90	-124.495
A493	$P2_1/c$	4	9.6714	4.2441	25.1263	90	88.016	90	-124.484
A1965	$P2_1/c$	4	15.8943	4.2513	19.3158	90	52.906	90	-124.48

A1161	*P2/c*	4	7.7484	5.42	26.0536	90	103.27	90	-124.474
A5240	$P2_1/n$	4	6.9365	15.7124	9.7136	90	103.744	90	-124.474
A812	$P2_1$	2	6.6768	7.2792	12.608	90	117.783	90	-124.46
A362	*Pbcn*	8	15.5315	11.2725	11.8017	90	90	90	-124.451
A729	*C2/c*	4	15.3179	6.9861	19.1974	90	85.688	90	-124.383
A103	$P2_12_12_1$	4	7.4291	6.7881	20.4826	90	90	90	-124.366
A353	*Pbca*	8	7.5003	23.7351	11.7744	90	90	90	-124.317
A97	*Cc*	2	18.0596	7.1515	8.9948	90	64.239	90	-124.277
A1963	*Pc*	4	21.3339	11.5224	8.4085	90.111	90	90	-124.27
A316	$P2_1/c$	4	5.803	7.6464	24.3677	90	97.786	90	-124.267
A716	$P2_1/c$	4	7.3665	10.709	13.9035	90	114.162	90	-124.241
A878	$P2_12_12_1$	4	13.2981	7.205	10.9441	90	90	90	-124.206
A194	*I-4*	4	14.6732	14.6732	9.6008	90	90	90	-124.144
A6991	$Pna2_1$	4	24.5671	9.9728	4.2492	90	90	90	-124.112
A1697	$Pna2_1$	4	13.9777	17.8411	4.1823	90	90	90	-124.109
A1911	$P2_1/c$	4	6.9888	7.2455	21.501	90	81.923	90	-124.019
A175	*C2/c*	4	10.6041	9.514	20.6982	90	85.484	90	-123.963
A8432	$P2_1/n$	4	11.4988	10.0549	9.8653	90	117.005	90	-123.894
A415	*Ic*	8	21.9675	6.9547	13.781	90	86.822	90	-123.877
A1970	$P2_1$	2	12.0761	10.0794	4.759	90	62.965	90	-123.856

Appendix Table 5.9.1: The 100 low energy structures from the CrystalOptimizer+PCM+FE serach using the Scheraga sulphur potential which comprised the structures submitted as List 2

Motif in order of stability with Halgren S	Motif in order of stability with Scheraga S
	
A467, #1, $P2_1/c$, -107.99 kJ mol^{-1} A4, #6, $Pna2_1$, -106.19 kJ mol^{-1}	A467, #1, $P2_1/c$, -135.01 kJ mol^{-1} A4, #2, $Pna2_1$, -134.74 kJ mol^{-1}
	
A8, #2, $Pna2_1$, -107.88 kJ mol^{-1}	A172, #3, $Pna2_1$, -134.57 kJ mol^{-1} This has same dimer chain as #1, with alternative stacking
	
A1, #3, $P2_1/c$, -107.82 kJ mol^{-1}	A43, #4, $P2_1/c$, -133.94 kJ mol^{-1}
	
A5293, #4, $P2_12_12_1$, -106.85 kJ mol^{-1}	A98, #5, $P2_1/c$, -133.12 kJ mol^{-1}



Appendix Figure 5.9.1: The packings of the 10 lowest energy structure submitted as list 1 and list 2 of XXII

6 Testing the ability to calculate heat capacity and solubility differences between enantiopure and racemic crystal structures.

6.1 Introduction

In the previous chapters, different methods for obtaining the lattice and free energies of crystal structures have been used. Are these calculated energies accurate enough to be used to predict heat capacities and solubilities? In order to design chiral resolution processes, we need the ternary phase diagrams (See Chapter 1.2) which require the solubility energy differences between enantiopure and racemic crystals. Such as differences may require less accurate absolute calculations of lattice and free energies as some errors could cancel between crystals of the same molecule. Different methods of calculating these energies, as described in Chapters 1 and 2, will be used and compared to experimentally determined heat capacities and solubility energy differences to assess how well these values can be predicted computationally.

6.1.1 Thermodynamic cycles

In collaboration with my sponsors at the MPI, Magdeburg I have been assisting in a project on the thermodynamic cycles of crystal structures. Hannes Buchholz has been experimentally investigating alternative thermodynamic cycles to determine solubility differences of crystal structures (see Figure 6.1) and link these to the ability to model the chiral separation processes that have been developed in Magdeburg via the ternary phase diagram. Both cycles consider properties of the crystal (melting and sublimation free energies) which are dependent on the structure of the crystal under investigation and thus there will be differences between enantiomeric and racemic crystals, as well as between different polymorphic forms.

The most common method for estimating solubilities via other, easier to measure, properties are based on the melt cycle, and thus the thermodynamic relationship between the free energy of solution, ΔG_{sol}, and a thermodynamic description of the melting process of the solute, including knowledge of the

melting temperature and any thermal corrections of the solid and the sub-cooled melt. The energy of mixing the melt with the solvent is also required, but this is often assumed to be the same for the enantiopure and racemic melt, and thus does not affect the solubility. However, it is not always possible to access this pathway experimentally as many samples decompose or a phase transition occurs before the sample melts which places severe restrictions on the wide spread use of this cycle.

Instead of following the thermodynamic pathway of melting, an alternative has been proposed by Palmer[1] in which an estimation of ΔG_{sol} is obtained from the free energy of sublimation, ΔG_{subl}, and the free energy of solvation, ΔG_{solv}. This cycle, referred to as the sublimation cycle, can be contrasted with the first, traditional, melt cycle for determining ΔG_{sol} as it relies on the thermodynamics of the solid-gas phase transition and solvation of a molecule.

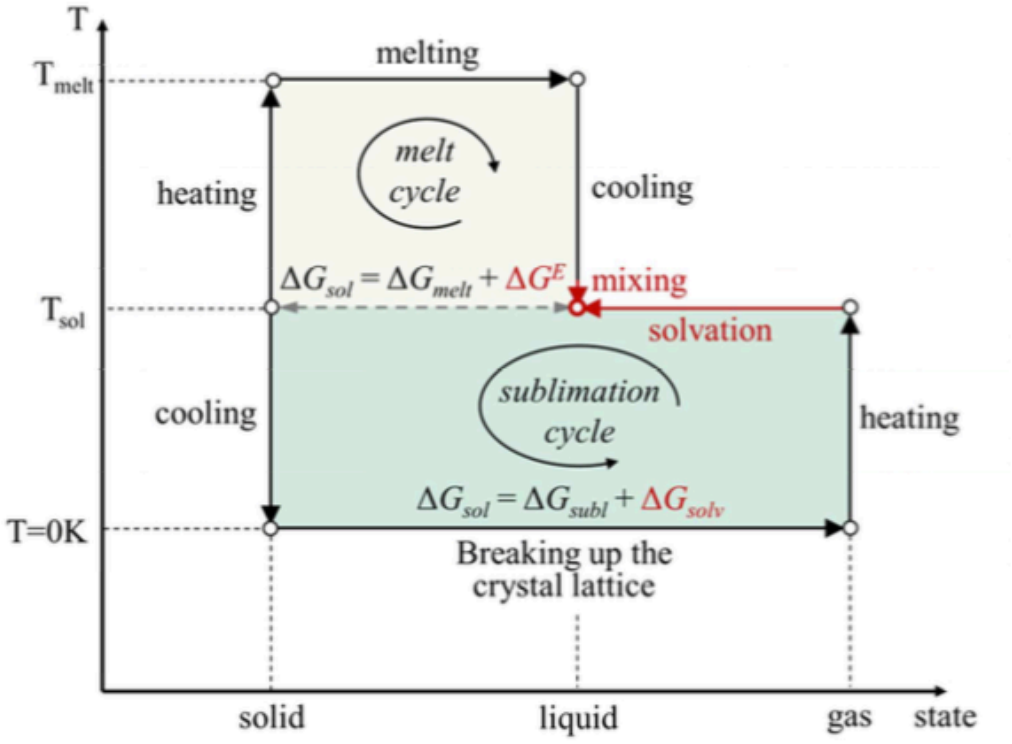


Figure 6-1: Two thermodynamic cycles which can be used to obtain the free energy of solution, ΔG_{sol}. The melt cycle, top, relies on the free energy of melting, ΔG_{melt}, and the free energy of mixing, $\Delta G^E = RT \ln \gamma$. The sublimation cycle relies on the free energy of sublimation, ΔG_{subl}, and the solvation of free energy ΔG_{solv}. Properties which are solvent dependent are highlighted in red.

During my collaborations with Hannes Bucholtz, he performed a comprehensive set of measurements on a set of three pairs of enantiopure and racemic crystals, 3-choloromandelic acid (3ClMA) and lactide which have featured in my thesis, and naproxen (Chapter 1), which was the study which influenced this

collaboration. In measuring the heat capacities of the pair of crystals at both high and low temperature regimes, Hannes provides the data for testing the calculation of this phonon-dependent thermodynamic property. At the MPI they have also analysed literature and other measurements, to extend the "expt/theory" comparisons to heats of sublimation (which also tests the lattice energies) and eutectic excess. This enables a comparison of how calculations and more readily accessible experimental data can be combined to understand and eventually help support the design of chiral separation processes. To determine the solution energy from the thermodynamic cycles, many different thermodynamic energies need to be either measured or calculated. These include the sublimation or melting enthalpies, solvation or mixing of melt and solute and solvent heating and cooling. In addition to understanding the thermodynamic equations for each of these processes, all approximations made in each step need to be considered. The parameters required to determine the elements of the melt cycle can be found experimentally via DSC experiments and measuring the Gibbs energy of mixing directly, where it is possible to melt the crystals avoiding decomposition.

The sublimation cycle, however, is more complicated and requires the use of computational calculations, as discussed in Chapter 2, either in entirety or mixed with experimental results, unlike the melt cycle which can be determined via experiment alone. The solubility depends on the free energy of sublimation (See Equation 2.4 and 2.5), which involved the lattice energy, difference in zero-point energy between the gas and solid, and integration over the heat capacity differences between the gas and solid. Assuming that there is no polymorphic phase transition between 0 K and the temperature of interest, T, the sublimation free energy is defined as

$$\Delta G_{subl} = \Delta H_{subl} - \Delta T S_{subl} = -U_{latt} + H_{corr} - T\Delta S_{subl}$$

with $$H_{corr} = \left(E^{g}_{ZPE} - E^{s}_{ZPE}\right) + \int_{T'=0}^{T} \Delta C_p^{g-s}(T')\, dT'$$

and $$\Delta T S_{subl} = T \int_{T'=0}^{T} \frac{\Delta C_p^{g-s}(T')}{T'} dT'$$

Equation 6-1

where U_{latt} is the lattice energy, E_{ZPE}, is the zero point vibrational energy and $\Delta C_p^{g-s}(T) = C_p^{g}(T) - C_p^{s}(T)$ is the temperature dependent isobaric heat capacity

differences between the gas (g) and solid (s). The thermal correction term, H_{corr}, is the sum of the zero point energies (ZPEs) and the integrals of the C_p.
The heat capacity of the solid and enthalpy of solvation for the solid can be measured by low temperature DSC experiments and direct measurement, respectively. The lattice energy at 0 K and zero point energies and all need to be determined by computational calculation as they are not experimentally accessible. The heat capacities can be determined from the calculated phonon modes of the crystal structures which can be further analysed for then origin of the experimentally observed differences. Calculating the solubility requires accuracy in the determination of these terms.

6.1.2 Estimation of the thermodynamic quantities

The main contribution to both cycles requires the breaking up of a crystal lattice, to the melt in one cycle, and to the ideal gas in the sublimation cycle This quantity, the energy required to separate the static perfect lattice into infinitely separated static molecules in their lowest energy configuration can be estimated computationally using a number of different techniques which have been described previously (See Chapters 1 and 2). However, the small contributions from zero-point energies and heat capacity differences are important in obtaining accurate solubility energy differences (See Chapter 1.9 and Chapter 2.6). While it is possible to determine the heat capacity terms via experiment, it may not always be feasible to do so, especially when other phase transitions or decomposition takes place, e.g. if another polymorph is stable at low temperatures. Therefore, if a direct comparison between computationally and experimentally determined heat capacities can be made an idea of the accuracy of solely computational thermodynamic quantity estimates can be made.

6.2 Methodology

Experimentally determined crystal structures of racemic and enantiopure 3-ClMA (FIZPEL03,[2] Z'=1, Z=4, P$2_1/c$ and TUYBIA,[2] Z'=2, Z=4, P21 with a 2:1 disorder in the phenyl ring orientation), lactide (BICVIS,[3] Z'=1, Z=4, P$2_1/c$ and NAHNOZ,[4] Z'=3, Z=12, P$2_12_12_1$) and naproxen (PAPTUX,[5] Z=1, Z=8, Pbca and COYRUD,[6] Z'=1, Z=2, P21) were obtained from the Cambridge Structural

Database[7]. These crystal structures were used as a basis for all calculations as they were the forms used in the experiments and are believed to be the most thermodynamically stable experimentally determined crystal structures of each of the molecules under investigation.

6.2.1 Modelling the heat capacities separating the molecular and lattice modes.

CrystalOptimizer[8] was used to determine the geometry optimised lattice energy for each of the structures of lactide, 3ClMA and naproxen. The torsions considered to be flexible (see Figure 6.2) were allowed to adapt to the packing forces within the crystal in order to minimise the lattice energy. A database of ab initio calculations on the isolated molecule using Gaussian03[9]/09 was used to determine the change in intramolecular energy, ΔE_{intra}, at the PBE1PBE/6-31G(d,p) level of theory for naproxen and lactide (see Chapter 4.2) and MP2/6-31G(d,p) for 3ClMA. For the disordered S-3ClMA form 1 structure, the two components were treated separately as S-3ClMA1_A and S-3ClMA1_B (see Chapter 3.2). The molecular charge densities were analysed by GDMA[10] to give the distributed multipole representation of the charge density that was subsequently used with DMACRYS[11] to obtain the electrostatic component of the intermolecular lattice energy, U_{inter}. The remaining intermolecular energies were represented by an isotropic atom-atom exp-6 Buckingham-type potential, using the parameters from the FIT potential.[11]

For S-naproxen (PAPTUX) the presence of non-zero eigenvalues in the final Hessian matrix of the refined crystal structure meant that the structure had relaxed to a saddle point and thus there is a lower energy structure with a lower symmetry S-naproxen was therefore symmetry reduced to a Z'=2 structure in order to reach an energy minimum.

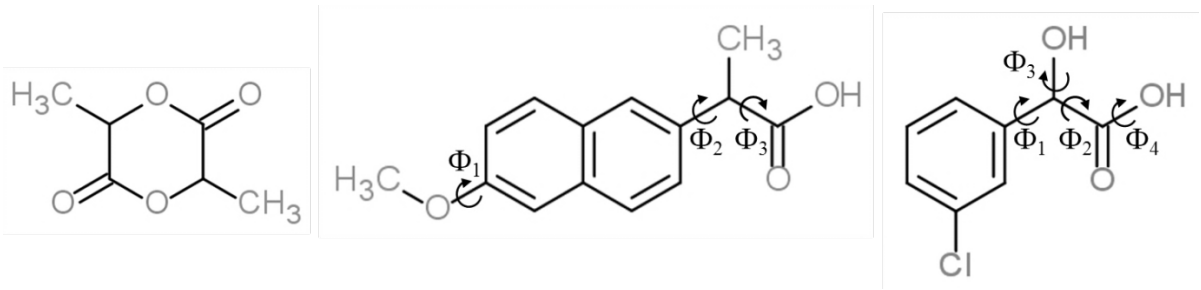


Figure 6-2: The lactide, naproxen and 3-chloromandelic acid molecules, highlighting the flexible torsion angles considered for CrystalOptimizer lattice energy minimisation.

To estimate the effects of temperature within the separated model, the *k*=0 phonon[12] and elastic properties[13] were determined. Under the harmonic approximation, the rigid-body lattice modes were calculated using an Einstein-Debye model which assumes that the thermal expansion of the crystal is negligible (see Chapter 2.9). This gave the zero point energy and phonon frequencies for a specific temperature (typically 298 K).
The molecular vibrations of the geometry optimized crystal structures, determined by Hannes Buchholz, were calculated at the PBE/def2-TZVP level of theory using the D3 dispersion correction[14] with TURBOMOLE V6.4.[15] These molecular vibrations were used in conjunction with the phonon frequencies to determine the heat capacity at constant volume of the solid crystal and thermal correction can be estimated (See Chapter 2.9).

6.2.2 Electronic modelling

Unlike with the separated model, periodic electronic structure calculations do not assume that the molecular and rigid-molecule phonons need to be separated. These calculations should therefore represent the effect of the crystal packing on the molecular modes, which is expected to be more important for flexible molecules. The following methods were used to provide the Ψ_{crys} estimates of the heat capacities, using the general approach given in Chapter Theory.

6.2.2.1 DFT-D modelling using CASTEP

I originally intended to calculate the phonon modes using plane-wave DFT-D calculation using CASTEP[16]. This proved to be much more challenging than expected, as these methods have not previously been used for larger organic crystal structures where the need to perform calculations to an accuracy such that the errors from having different sized unit cells did not affect the differences in the calculated energies. DFT-D plane wave pseudopotential calculations were performed on each of the crystal structures. Starting from the refined *CrystalOptimizer* structure, the geometry of each structure was re-optimised using the PBE function with the Tkatchenko-Scheffler[17] (TS) dispersion correction using CASTEP version 16.1[16] to ensure the structure was fully converged prior to calculating the phonon frequencies. A cut off of 1000 eV was

used for all crystal structures. A *k*-point grid was set up such that a *k*-point spacing of approximately 0.2 $Å^{-1}$ was used in all 3 dimensions of the unit cell. These values were determined from convergence calculations on the high symmetry urea crystal. Unfortunately, due to the size of the crystal structures being optimized, particularly the Z'=3 structure of lactide containing 216 atoms, convergence of the geometry to a stable minimum could not be obtained and this is required to calculate the phonons.

Where convergence of the geometry was obtained sufficiently, electronic phonon calculations were attempted. In order to make a direct comparison to the $\Psi_{mol,}$ *k=0* phonon frequencies, initially the phonon modes at *k=0* were calculated. It was found that the *k*=0 phonon modes did not lead to sufficiently accurate heat capacities. This is in agreement with work on vibrational energy differences by Day et al.[18] Thus phonon calculations across the full Brillouin zone were attempted. Unfortunately, electronic structure calculations across the full Brillouin zone led to difficulties in obtaining a sufficient level of convergence for an optimised geometry at other k-points leading to imaginary phonon frequencies. These imaginary phonon frequencies made it impossible to calculate heat capacities.[19,20] I worked closely with Keith Refson, one of the CASTEP developers, to see if this problem could be overcome however it proved to be much more complex than first anticipated. This work is being continued by Dr Rui Guo in conjunction with Prof. Refson. Performing phonon calculations on all but the smallest organic crystal structures is still far from routine. Our attempts to generate sufficiently accurate results shows the inadequacies that remain, even for these relatively small organic molecules and their diverse crystals.

6.2.2.2 Alternative periodic electronic structure calculations

Due to the unforeseen complications associated with the CASTEP calculations, alternative electronic structure calculations based on a minimal basis set Hartree-Fock (HF-3c) and on a tight binding Hamiltonian (DFTB3-D3) calculations were performed by Gerit Brandenburg using a developer version of CRYSTAL14[21] and dftb+.[22] We worked together to analyse the output of these calculations, developing the use of Jmol[23] to visualise the phonon modes, which could be done only using the output of the CRYSTAL electronic structure calculations. These electronic structure calculations provide a means of

comparing different theoretical calculations of the lattice energies and heat capacities to experimental methods to provide an insight into areas of improvement for theoretical calculations. It was found that the differences between the heat capacities from the much cheaper DFTB3-D3 calculations and the HF-3c results (quoted as Ψ_{crys} or electronic throughout) were significant (see Appendix 6.5.1) which indicates that a good description of the interactions and a fully coupled treatment of all modes is important. As both DFTB3-D3 an HF-3c methods give the full dynamical matrix where all degrees of freedom are treated explicitly in the period boundary conditions a separation of the inter- and intermolecular modes cannot be given. To achieve this, and be able to compare the effects of changing the quality of the model for the inter- and intramolecular forces from the effect of separating the molecular and lattice modes, better theoretical models, such as using DFT-D calculations in CASTEP, would be required. This has been shown to currently be too demanding for routine applications.

6.2.3 Experimental results

All experimental results were obtained by Hannes Buchholz from the MPI Magdeburg, unless otherwise stated.

6.2.3.1 Heat Capacity Measurements

Low temperature heat capacity measurements for racemic and enantiopure 3ClMA and naproxen were measured using direct heat pulse calorimetry (DHPC) between 2 and 200 K. The measurements were obtained using a commercially available physical property measurement system (PPMS) relaxation calorimeter from Quantum Design. This method of measurement provides a high-vacuum environment in which the measurements are obtained. Low temperature heat capacity measurements for lactide could therefore not be obtained in the PPMS due to the high vapour pressures of lacide[24] and so low temperature heat capacity measurements obtained in an adiabatic vacuum calorimeter were taken from the literature.[25,26]

Heat capacity measurements above room temperature of 3ClMA, lactide and naproxen were measured using a DSC 111 calorimeter from Setaram.[27] 3-ClMA

decomposes at 92°C [28] and so any thermodynamic measurements were only obtained up to this temperature. To support experimental high temperature measurements of 3ClMA, measurements of the closely related enantiopure and racemic mandelic acid were also obtained.

6.2.3.2 Eutectic composition and solubilities

The eutectic composition of 3ClMA[28,29] and lactide[29,30] in various solvents were taken from the literature, including previous work at Magdeburg. The eutectic composition of RS- and S-naproxen in ethanol, toluene and ethanol/acetenitrile were determined experimentally.[31]

The solubilities of lactide and naproxen were measured in additional solvents to supplement pre-existing literature data. Melting properties, vapour pressures and corresponding enthalpies of free energies of sublimation were taken from the literature,[24,25,28,32] where available. Due to the decomposition of 3ClMA the sublimation enthalpies of the crystal structures could not be determined experimentally.

6.2.4 Comparison of experimental and theoretical values

To maximise the comparison of the experimental and theoretical values and test the extent to which some components in the cycle could be substituted by experimental measurements, so-called 'exp/theory' methods were developed. The experimental heat capacities were interpolated and numerically integrated using the in-built spline functions within the MATLAB software package to give experimental solid state heat capacity contribution to the heat of sublimation or entropic contribution to the free energy of sublimation (see Equation 2.1). These can be used to compare with calculated values. Additionally these values can then be combined with the computationally determined ideal-gas heat capacities from the isolated molecular frequencies and the calculated zero-point energies to give the thermal correction, H_{corr}, and sublimation entropies (See Equation 2.4) in a mixed "exp/theory" model. The free energy of solvation, ΔG_{solv}, was calculated using the thermodynamic model within COSMO-RS taking into account the distinct flexibility of 3ClMA, naproxen and lactide by using a Boltzmann weighting of all possible conformers[31] to assess how

accurately this quantity could be calculated, and whether it accounted for the variation in solubility with solvent.

6.3 Results

6.3.1 Determination of absolute heat capacities and from computed phonon frequencies

The crude rigid body k=0 phonon calculations were used to determine the heat capacities, which were compared with other calculated and experimentally determined heat capacities. The calculated absolute heat capacities of each of the individual crystal structures, and subsequent enthalpies of sublimation (estimated using the lattice energy) are in good agreement with those that were obtained from experimentally determined heat capacities, as can be seen in Figure 6.3. Naproxen has a significantly greater heat capacity than both 3ClMA and lactide which is a result of the greater number of atoms per molecule.
At low temperatures, the heat capacities follow Debye's cubic law ($C_p \sim AT^3$) and then increases linearly with temperature. If the –2RT thermal correction was valid then mode saturation would occur at high temperatures, causing the heat capacity to level out. As this is not observed both with the theoretical and experimental results the –2RT assumption is not valid for the organic systems under investigation.

At high temperatures, the calculated heat capacities appear to be systematically underestimated. This is consistent with a recent study which compared experimental solid state heat capacities with calculated values for crystals comprised of small molecules, where an underestimation was noted in 85% of the systems studied.[19] This systematic difference could also be attributed to the gap between the low temperature DHCP and high temperature DSC measurements. The gap between the DHCP and DSC measurements is around 4% for all systems which is within the range of precision for heat capacity measurements[33]. While the error is not large enough to account for the complete systematic underestimation of the calculated heat capacities it could still contribute to the differences in the experimental and computational heat

capacity values. Hence, we need to consider whether the theoretical models used may also be making an assumption that would underestimate the high temperature heat capacities.

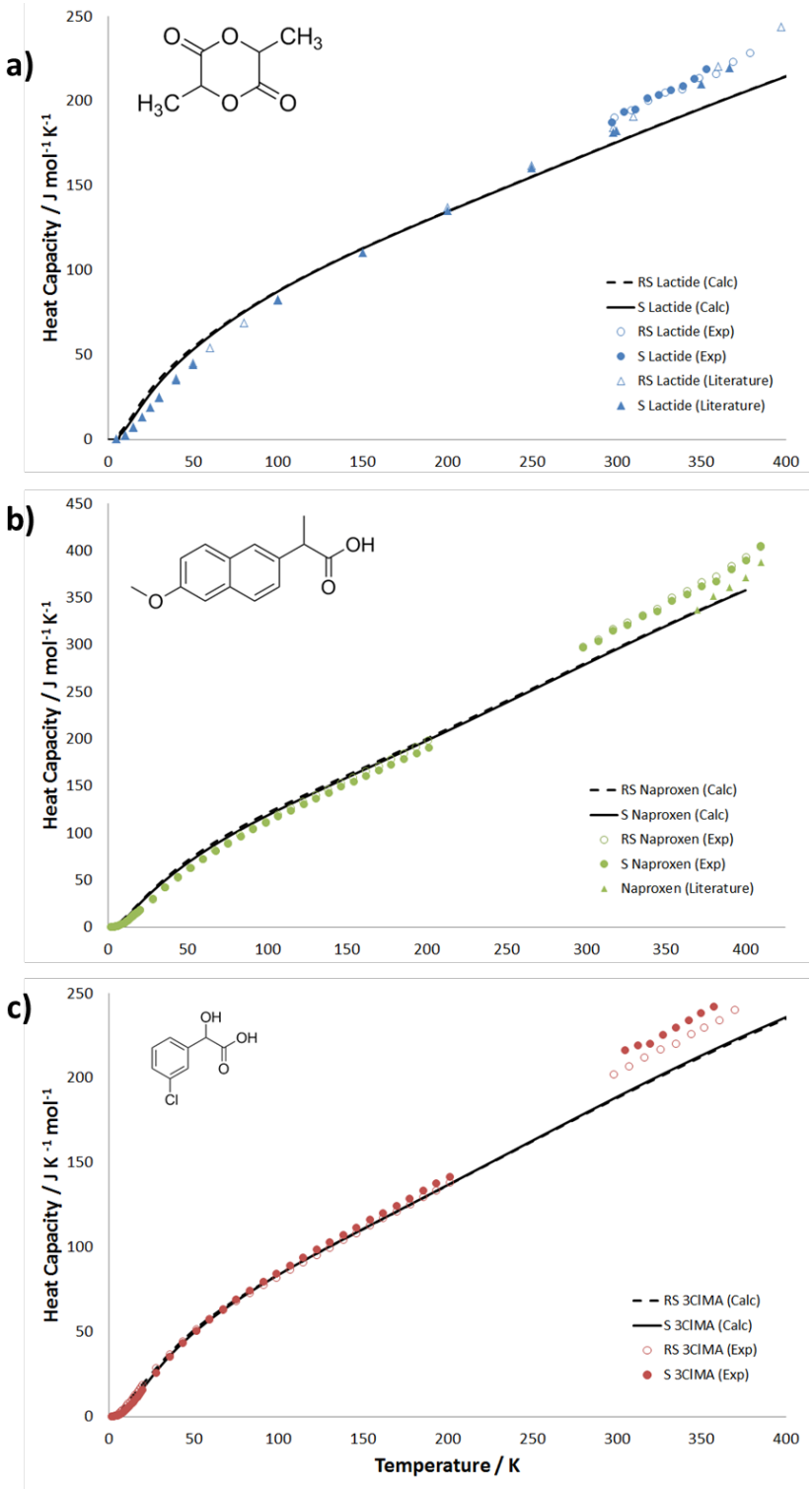


Figure 6-3: Experimental heat capacities of (a) (●) (S)-Lactide and (○) (RS)-Lactide measured with differential scanning calorimetry in comparison to literature data[25,26] (b) (●) (S)-Naproxen and (○) (RS)-Naproxen measured with PPMS (T = 2 K – 200 K) and differential scanning calorimetry (T > 298 K). Grey ● represent literature data of an unspecified naproxen sample.[34] (c) (●) (R)-3ClMA and (○) (RS)-3ClMA measured with PPMS (T = 2 K – 200 K) and differential scanning calorimetry (T > 298 K). The calculated heat capacities using the rigid-body lattice and molecular modes are given by a full line for the enantiomer, and a dashed line for the racemic compound, and on (b) in green the lattice modes. On this scale, the heat capacities estimated from the periodic HF-3c calculations are almost insdistinguishable from the DFTB3 and Ψ_{mol} calculations and are compared with the experimental values in the Appendix 6.5.1

6.3.2 Heat capacity differences

There are apparent discrepancies between the theoretical and experimental absolute heat capacities. However, for the design of chiral separation processes we are more concerned with looking at heat capacity differences between enantiopure and racemic crystal structures. It is hoped that any errors incurred through calculating the absolute heat capacities will cancel when looking at heat capacity differences.

The calculated heat capacity differences, with both the separated and electronic approaches, are not able to quantitatively represent the experimental values at low temperatures below 100 K and indicate that the heat capacity differences are system dependent (see Figure 6.4). The separated model, using a harmonic, rigid body approximation, which from 100 K tends to zero for all three systems. This is unrepresentative of the experimental heat capacity measurements.

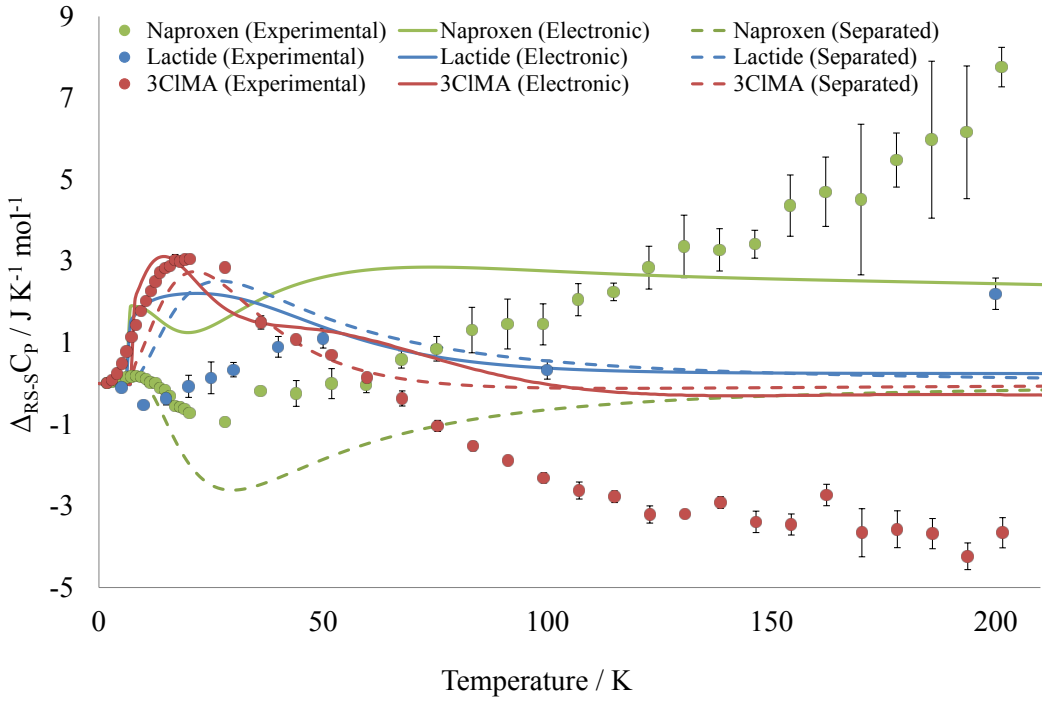


Figure 6-4: The low temperature (0 – 200 K) heat capacity differences, $\Delta_{RS\text{-}S}C_P$, between the enantiopure and racemic crystal structures of 3ClMA, lactide and naproxen. Experimental results are shown with experimental uncertainty (dots) and compared with calculated $\Delta_{RS\text{-}S}C_v$ values from electronic Ψ_{crys} HF-3c (solid lines) and Ψ_{mol} rigid body phonon modes (dashed lines)

As the temperature is increased to process relevant temperatures (see Figure 6.5), the same discrepancies can be observed. The periodic electronic structure calculations provide a slightly better representation of the heat capacity differences above 100K, particularly for naproxen, although these calculations

still do not show the variation in heat capacity differences with temperature that are observed experimentally.

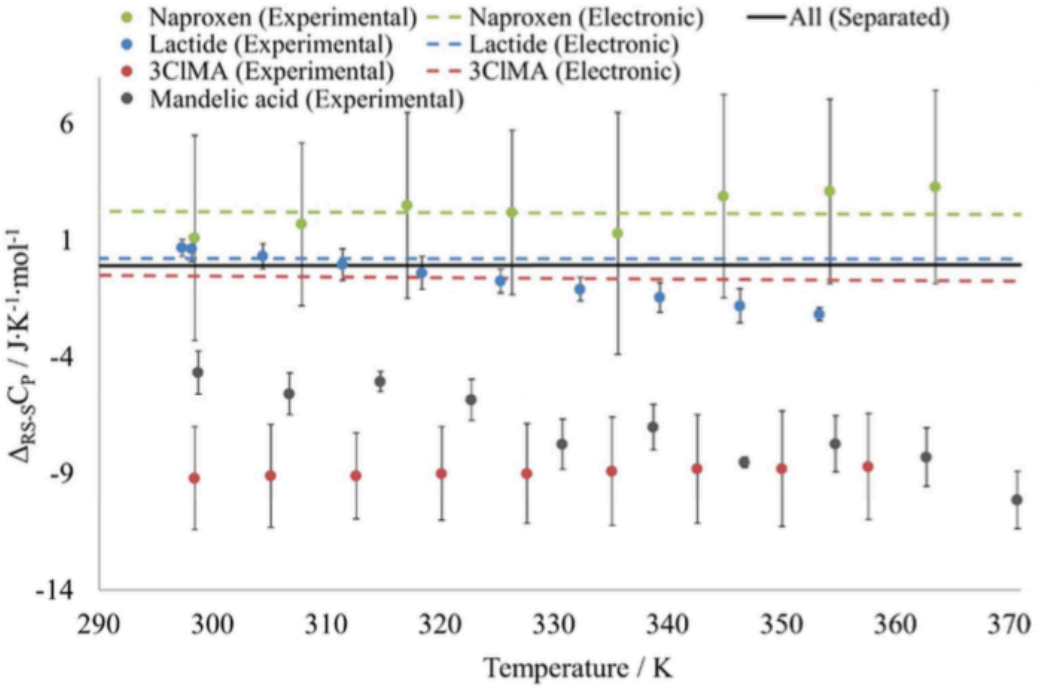


Figure 6-5: The high temperature (300 – 370 K) calculated and experimental heat capacity differences between the enantiopure and racemic crystal structures of 3ClMA, lactide and naproxen. Computed electronic HF-3c $\Delta_{RS\text{-}S}$ C_v are given as dashed lines in the same colour; atomistic $\Delta_{RS\text{-}S}$ C_v are zero (black straight line).

6.3.2.1 Origins of the differences between computed and experimentally determined heat capacities

The deviation between the calculated and experimental heat capacities of 3ClMA, naproxen and lactide suggests that there are sources of errors in the methods of calculation. As the main contribution to the calculated heat capacities is from the phonon modes then the assumptions made during their calculation should be considered carefully. The separated model assumes that the lattice modes are sufficiently separated in energy from the molecular modes such that no coupling between these two modes should occur. However, in reality, this cannot be said about any of the three systems being investigated (see Figure 6.6). For the 'rigid' lactide system there are ring bending modes which are of a similar frequency to the lattice modes. With naproxen there are low frequency molecular modes which correspond to rotations of the propionic acid side chain and the bending of the naphthyl group, which is particularly prominent in the enantiopure crystal.[6] For 3ClMA there are low frequency

molecular modes which correspond to rotations of the flexible alpha-hydroxy acid side chain and out-of-plane vibrations of the chlorine atoms.

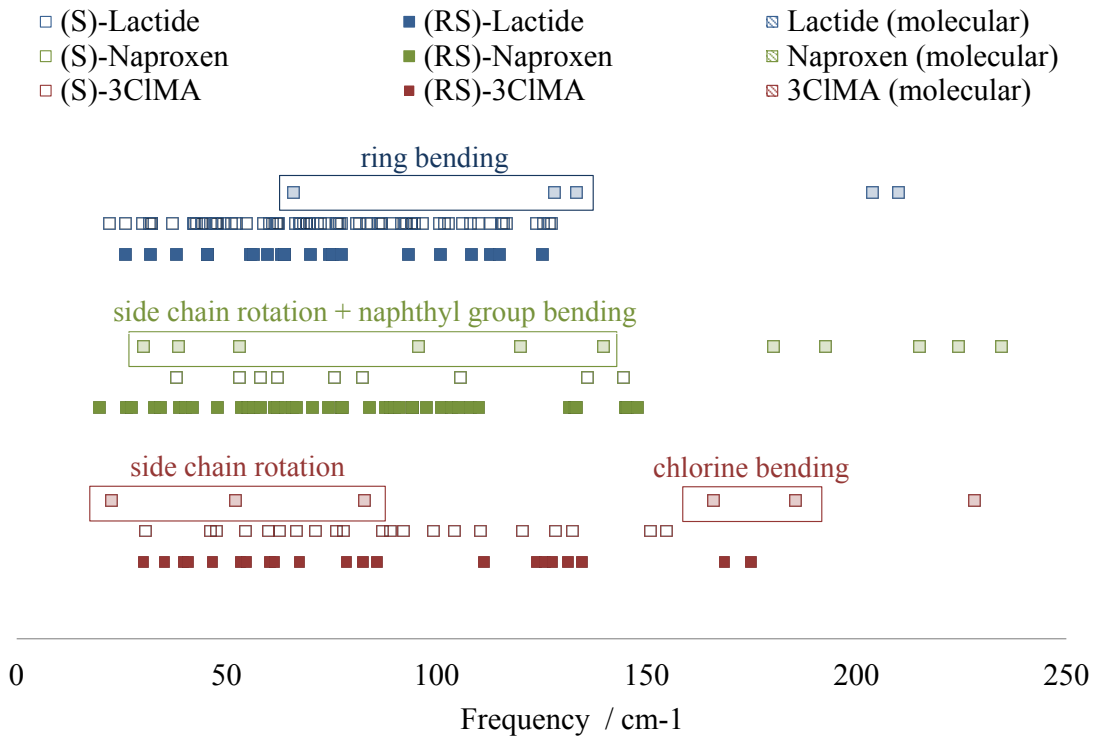


Figure 6-6: A comparison of the rigid-molecule k=0 lattice frequencies, evaluated using the separated model, of the enantiopure and racemic crystal structures of 3ClMA, lactide and naproxen compared with their isolated molecule low frequency modes.

At low temperatures, the heat capacities are determined by the low frequency lattice modes. The packing of the molecules of 3ClMA, lactide and naproxen are very different in the enantiopure and racemic structures and so the low frequency lattice modes are very different. The low frequency modes of RS-naproxen correspond to a relative sliding of the layers of hydrogen bonded dimers whereas in S-naproxen there are hydrogen-bonded chains which rotate relative to one another. For RS-3ClMA the low frequency modes, like naproxen, are due to pairs of hydrogen-bonded layers sliding relative to each other and the low frequency modes in S-3ClMA are due to phenyl-sidechain bending. These lattice frequencies can be observed in Figure 6.7. Both the low frequency modes of RS naproxen and 3ClMA involve the van der Waals forces between layers of molecules and so will be very sensitive to the underlying potential energy surface used to evaluate the lattice frequencies. The crystal structures of lactide are dominated by weak intermolecular van der Waals interactions with no hydrogen bonded interactions that can be seen in naproxen and 3ClMA and

so is particularly affected by the choice of potential used to model the potential energy surface.

In addition to this, the low frequency modes for all systems were evaluated, in the atomistic model, by assuming the molecules are rigid. Lactide is the only system which can be considered rigid and so there are potential for errors to incur for the more flexible naproxen and 3ClMA. Therefore the error induced by neglecting the coupling of molecular and lattice vibrational modes is very dependent on the molecule and crystal structure being generated and careful consideration of the potential and degree of flexibility of the molecule are needed.

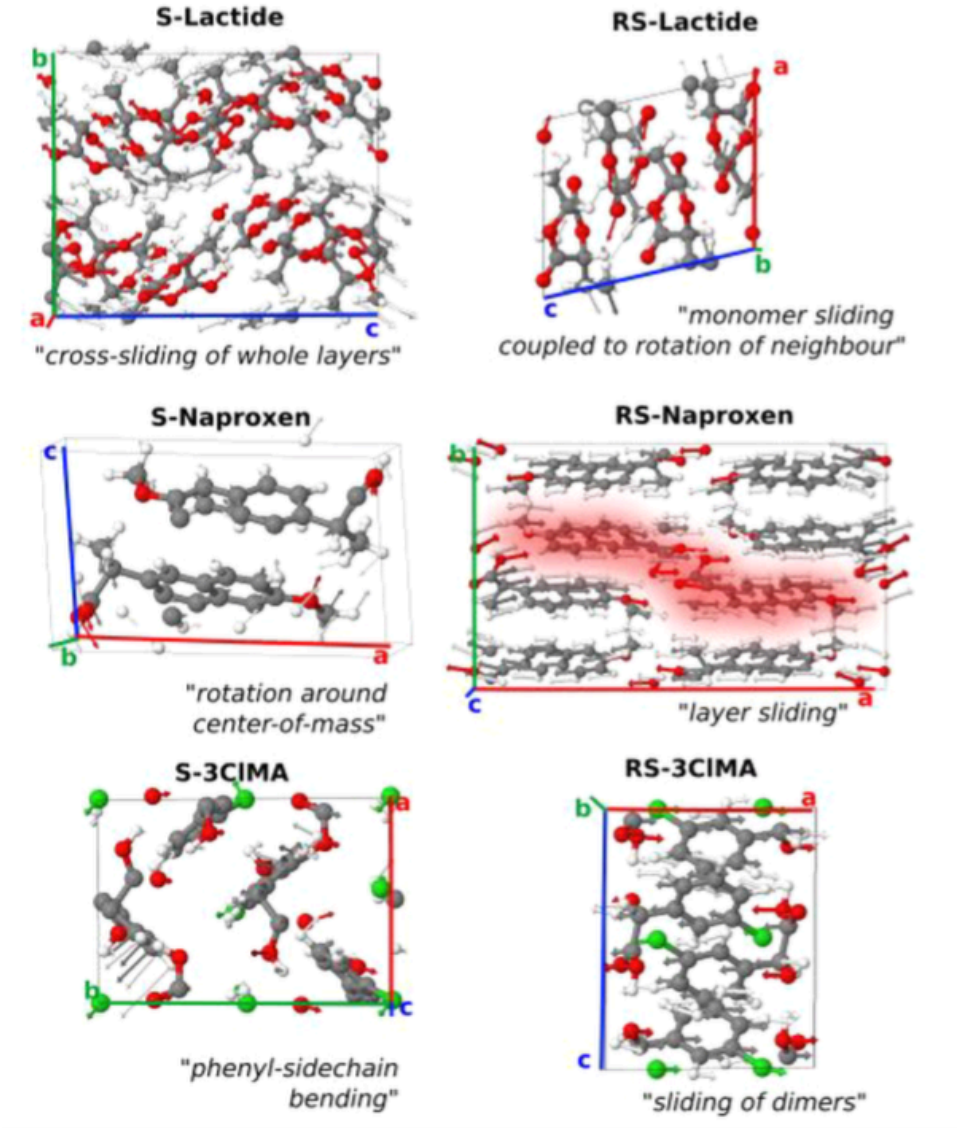


Figure 6-7: The low frequency lattice modes of 3ClMA, lactide and naproxen visualised using a vector representation of the atomic molecule in Jmol. The lattice modes were derived using HF-3c k=0 electronic structure calculations. The red shading represents the hydrogen bonded layers in the RS-naproxen crystal.

At high temperatures the low energy lattice modes become saturated and so only the high frequency molecular modes are relevant when considering the heat capacities at high temperatures. The separated atomistic model therefore predicts that the heat capacity differences tend to zero with increasing temperature (see Figure 6.5) as the molecular modes are assumed to be unaffected by the crystal structure. The electronic calculations however, predict that there are differences in the high frequency modes of the enantiopure and racemic crystals for each system, which lead to differences in the heat capacities for both naproxen and 3ClMA but not lactide. This is more in line with what is observed with the experimental heat capacity differences.

These deviations in heat capacity could be accounted for by differences in bonding between the racemic and enantiopure crystal structures. Differences in bonding environment affect the molecular frequencies from which the heat capacities are derived. The lattice frequency modes measured by Raman spectroscopy are an accessible way to assess the validity of this statement as the experimentally determined vibrations can be compared with calculated dynamic properties of crystal structures.[35] Indeed, small shifts in peaks on the Raman spectra can be observed for all three systems (see Figure 6.8). For naproxen and 3ClMA there are noticeable differences between the crystal forms, particularly at higher frequencies and in regions which would be correspond to changes in hydrogen bonding environments i.e. the C-O around 1600-1800 cm^{-1} and O-H around 3000 cm^{-1}. The Raman spectra of racemic and enantiopure and racemic lactide are almost identical which is expected as the crystal structures are dominated by van der Waals forces, and lactide does not have the ability to form hydrogen bonds. A change in the higher frequency molecular modes of racemic and enantiopure crystal structures is a requirement for the difference in heat capacities observed at process relevant temperatures and so the assumption that the molecular modes are unaffected by the crystal structure is not strictly valid and differences in hydrogen bonding environment between racemic and enantiopure crystals needs to be considered.

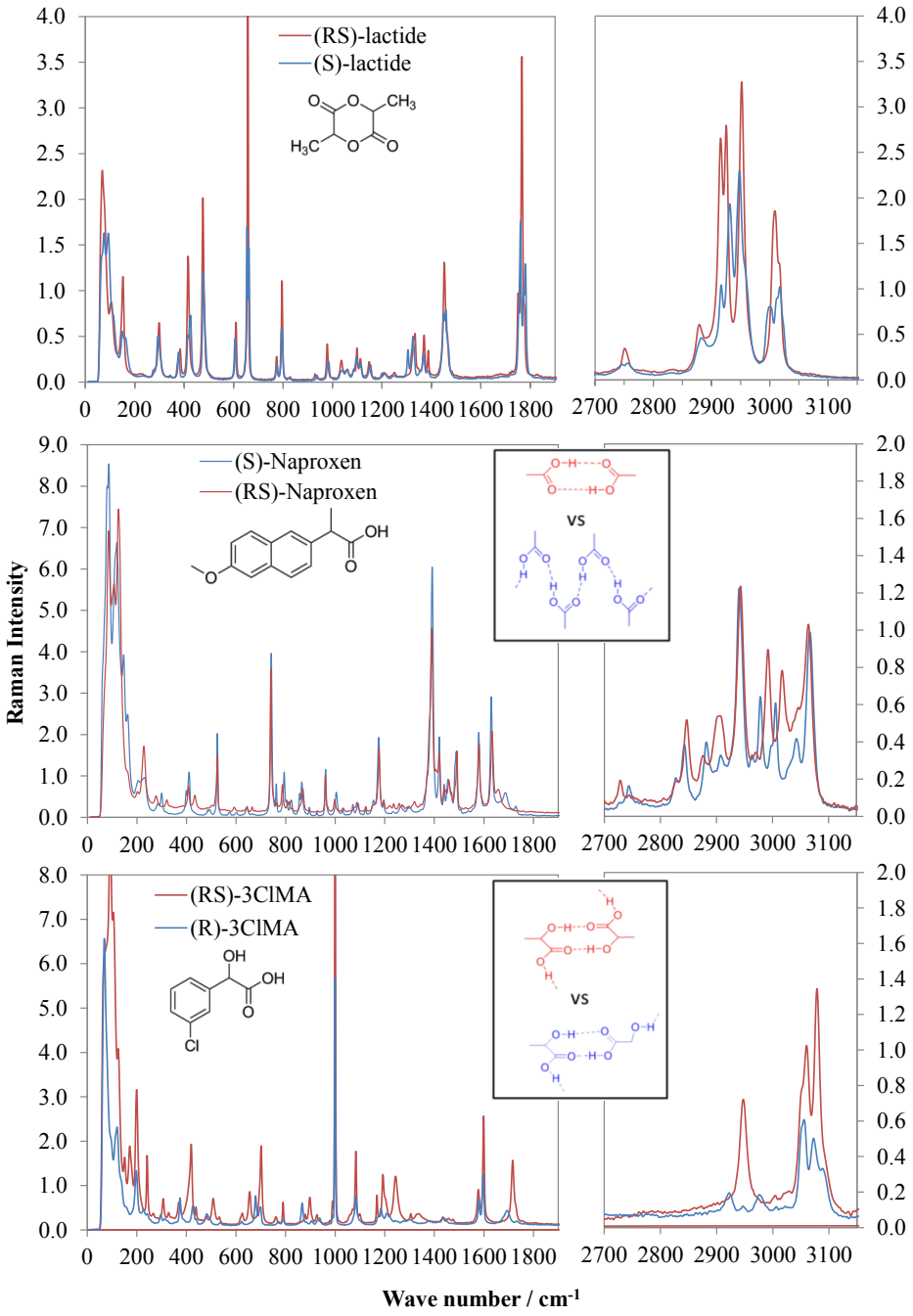


Figure 6-8: Overlay of the solid-state Raman spectra of the racemic and enantiopure crystals of lactide naproxen and 3ClMA showing the regions between 0-1800 cm^{-1} and 2700-3150 cm^{-1}. The inset shows the relevant differences in hydrogen bonding motifs.

While the electronic structure calculations provide a better representation of the molecular modes of the systems, there is still a systematic under-estimation of the high temperature heat capacities. For the work done in this chapter the harmonic approximation has been assumed. This approximation is unlikely to be valid at process relevant temperatures, close to the melting points (see Appendix 6.5.2). However, it is the anharmonicity of the vibrations which leads to a thermal expansion of a crystal structure. The harmonic approximation completely neglects the thermal expansion of the crystal which can be very

anisotropic, depending on the differences in molecular packing. The effects of thermal expansion will be negligible at low temperatures, but the approximation that solid-state heat capacities at constant volume and pressure are the same, i.e. the thermal volume expansion coefficient is the same (see Equation 2.10) will become increasingly poor as the temperature increases. Gerit Brandenberg used the quasi-harmonic approximation and DFT-D3 calculations to estimate the effect of thermal expansion on the heat capacities of the naproxen crystals. This showed that harmonic approximation assumption was responsible for underestimating the room temperature C_p by 7 and 8 J K^{-1} mol^{-1} for RS and S naproxen, respectively. Thus the breakdown of the harmonic approximation at higher, process-relevant temperatures could lead to deviations observed between the experimental and calculated heat capacities by as much as 5%.[31] This is consistent with recent accurate calculations on acetic acid and imidazole crystals which showed that the zero-point energy and thermal expansion affect the value of the free energies by a few kJ mol^{-1}.[36] Thermal expansion will decrease the value of the lattice frequencies and so it follows that the harmonic approximation will cause a systematic underestimate of the heat capacities at higher temperatures. The extent of this underestimation will depend on the degree and anisotropy of thermal expansion which is individual for each crystal structure being investigated.

The differences in phonon modes, could in principle, cancel out in the differences in thermodynamic quantities involved within the sublimation cycle. The absolute thermochemical properties, as can been seen in Table 1, show that there is no cancellation between any of the contributions or for any of the molecules investigated. Unless there is a sufficiently large difference in the lattice energies, which would render chiral separation by crystallisation impossible, then all of the small thermodynamic terms will have an effect on the eutectic composition. Therefore it is necessary to evaluate all of the solid-state contributions to the thermodynamics of the ternary phase diagrams, either computationally or experimentally, to the highest degree of accuracy.

	Lactide			Naproxen			3CIMA		
	RS	S	RS-S	RS	S	RS-S	RS	S	RS-S
Lattice Energy: U_{latt} / kJ mol^{-1}									
Separated (Ψ_{mol})	-86.17	-82.64	-3.53	-133.48	-125.41	-8.07	-110.40	-105.11	-5.28
Electronic (Ψ_{crys})	-90.1	-87.25	-2.85	-135.71	-134.26	-1.45	-127.43	-126.20	-1.23
Zero Point Energy $\left(E^{g}_{ZPE} - E^{s}_{ZPE}\right)$/ kJ mol^{-1}									
Separated (Ψ_{mol})	-2.40	-2.59	0.19	-2.83	-2.76	-0.07	-2.90	-2.95	0.05
Electronic (Ψ_{crys})	-2.75	-2.99	0.24	-4.45	-5.39	0.94	-6.04	-7.38	1.34
Heat capacity correction to enthalpy: $\int_{T'=0}^{T} \left(C_P^g(T') - C_P^s(T')\right)dT'$ / kJ mol^{-1}									
Separated (Ψ_{mol})	-2.72	-2.55	-0.17	-2.37	-2.55	0.18	-2.31	-2.25	-0.06
Electronic (Ψ_{crys})	-2.18	-2.02	-0.16	-0.92	-0.21	-0.71	0.60	0.65	-0.05
Experimental	-3.11	-2.73	-0.37	-1.84	-1.55	-0.29	-0.61	-1.52	0.91
Entropic correction: $T\Delta S_{subl} = T\int_{T'=0}^{T} \left(C_P^g(T')/T' - C_P^s(T')/T'\right)dT'$ / kJ mol-1									
Separated (Ψ_{mol})	51.55	52.72	-1.17	59.78	58.64	1.14	57.00	57.93	-0.94
Electronic (Ψ_{crys})	47.88	51.85	-3.97	62.30	65.44	-3.14	63.19	64.62	-1.43
Experimental	56.53	57.07	-0.54	69.76	70.50	-0.74	65.80	66.14	-0.33

Table 6-1: The absolute thermochemical properties and differences between the enantiopure and racemic crystals of lactide, naproxen and 3CIMA. Contributions to the sublimation enthalpies, ΔH_{subl}, and free energies, ΔG_{subl}, determined at 298 K are shown for all compounds.

6.3.3 Sublimation enthalpies and thermal corrections

The lattice energies determined from DMACRYS and *CrystalOptimizer* calculations were used as a first estimate for determining the absolute

solubilities and solubility differences of the crystal structures being investigated. As shown previously with naproxen (see Chapter 1.6), the lattice energies alone do not accurately reproduce the enthalpies of sublimation which were determined from experiment and so it is hoped that by including the small contributions from the thermal corrections a more accurate representation of the sublimation enthalpy can be obtained.

6.3.3.1 Absolute and relative sublimation enthalpies

The absolute and relative sublimation enthalpies are calculated from a combination of the calculated lattice energies, zero point energies and the calculated or experimental heat capacities (see Equation 2.1). Two different approaches were used to determine solid state heat capacities. The first was a combination of determined experimental and theoretical values, the 'exp/theory model. This model combines the numerical integrations of the experimentally determined solid state heat capacities of each compound with their corresponding calculated gas phase heat capacities to estimate the thermal correction. The solid state heat capacities obtained from pure calculation were combined with the zero-point energies, and the gas phase heat capacities derived from combining the unscaled molecular frequencies with the calculated zero-point and lattice energies. The molecular frequencies were used within the RRHO approximation, with both the atomistic and electronic calculated zero-point energies and lattice energies used to estimate the thermal corrections so to provide a comparison of the different theoretical models used, see Table 6.1.

Figure 6.9 shows the differences between using just the lattice energies and combining the lattice energies and thermal energies to estimate the absolute sublimation energy and the sublimation energy differences in comparison to experiment. The main contribution to the sublimation energy is the lattice energy and so any improvement in the accuracy of obtaining the lattice energy would improve the accuracy of the calculated sublimation enthalpies. However the zero-point energies and heat capacity terms, while small, still provide a significant contribution and so should not be neglected.

Comparing the 'exp/theory' model with the purely theoretical model generally shows only small differences between the two models for both. The only

differences between these two methods is whether the experimental or calculated heat capacities are used, indicating that the calculated heat capacities provide a reasonable estimate to the experimental heat capacities in determining absolute sublimation and sublimation differences. The only exception to this is for the sublimation difference of 3ClMA where the heat capacities are poorly reproduced at higher temperatures. Therefore, in cases where there is a significant difference in the heat capacities over a large temperature range it would be more advantageous to use the 'exp/theory' model to determine the sublimation differences, if experimental heat capacities are able to be obtained.

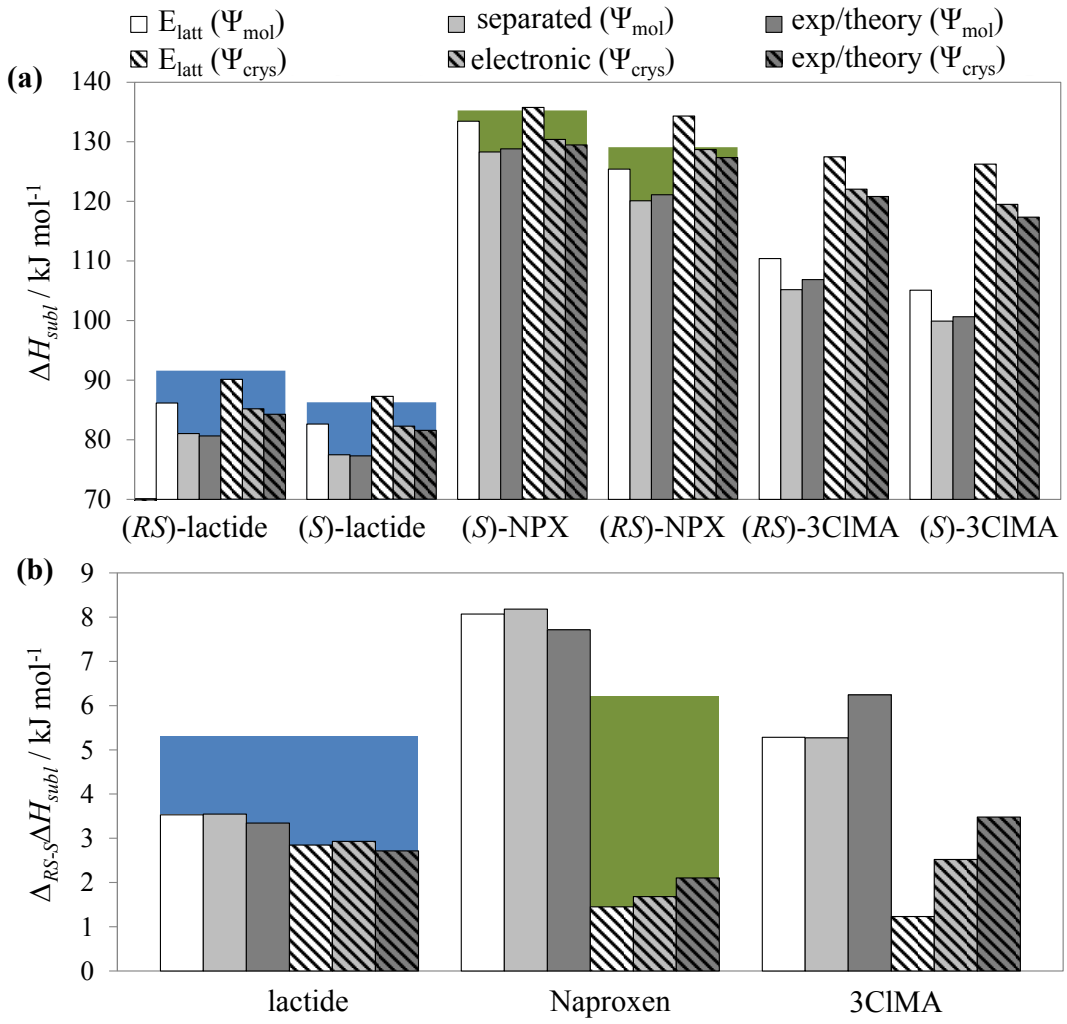


Figure 6-9: (a) Absolute sublimation enthalpies, ΔH_{subl}, and (b) sublimation enthalpy differences between the racemate and enantiopure structures compared to experiment as coloured background (blue = lactide,[24] green = naproxen,[32] 3ClMA decomposes), with the plain bars corresponding to the separated model and the striped bars to the electronic model.

6.3.3.2 Thermodynamic corrections

The –2RT thermal correction has often been used to account for the thermodynamic contributions arising from molecule and crystal-specific phonon

frequencies. It has previously been shown (see Chapter 1.6) that the –2RT approximation is inadequate for small organic crystal structures and my results show that this conclusion extends to the larger molecular crystals of 3ClMA, lactide and naproxen (see Figure 6.10). It appears that the thermal correction is influenced by both the intermolecular forces in the crystal structure and degree of flexibility of the molecule in question. Lactide is the small rigid molecule, with no hydrogen bonding capability and so the thermal corrections evaluated using the 'exp/theory' model and both the theoretical atomistic or electronic methods are quite close to the –2RT approximation, as well as having limited variability across all three methods used to evaluate the thermal correction. However, as the flexibility of the molecule increases and differences in hydrogen bonding capability of the enantiopure and racemic crystals are more apparent so are the differences in the thermal corrections. These differences are extended to the method in which the thermal corrections are calculated, with a significant difference between the 'exp/theory' and purely theoretical methods.

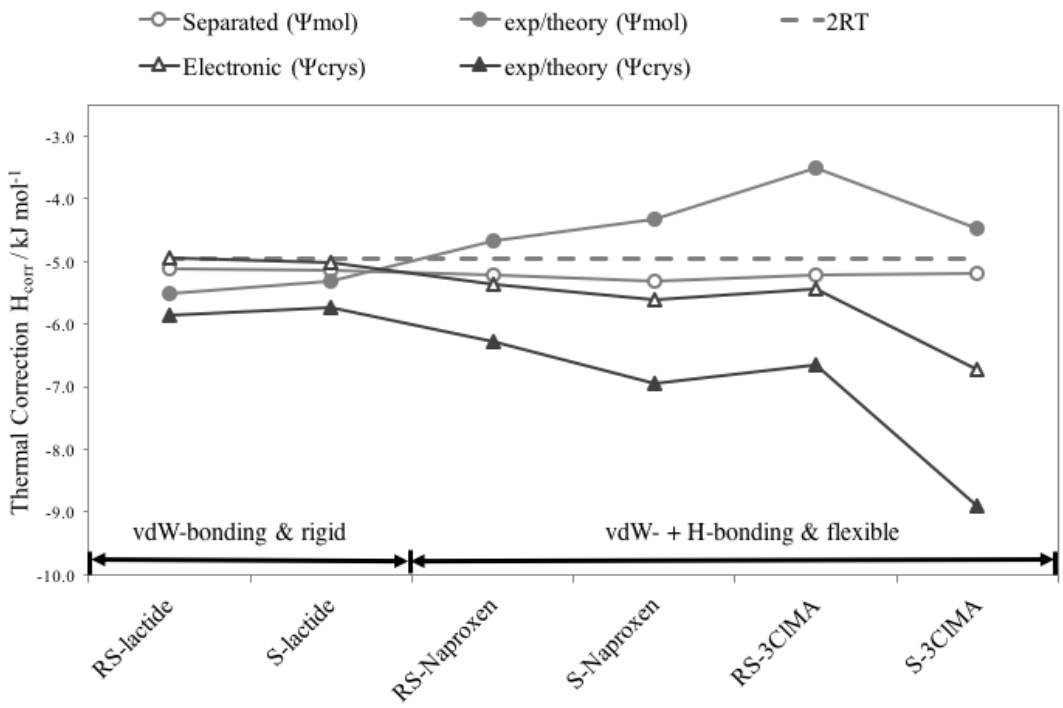


Figure 6-10: The thermal corrections of lactide, naproxen and 3ClMA, classified by intrinsic molecular interactions in the crystal using the zero-point energies and heat capacities determined by the 'exp/theory' model and purely theoretical models using the separated and electronic (HF-3c) methods, compared to the –2RT approximation.

The most significant source of error in the determination of the sublimation energies is in the heat capacity, with an error of 1.5 kJ mol^{-1} and 3.3 kJ mol^{-1} for

naproxen and 3ClMA, respectively. The general model-dependent variations in the sublimation differences are up to 2.3 kJ mol^{-1}. Even with modern, state of the art techniques, sublimation enthalpies are still difficult to obtain experimentally, if at all, with a typical margin of error of ±5 kJ mol^{-1} is associated with absolute sublimation calculations.[37] For sublimation differences this error range extents to 2-8 kJ mol^{-1}. Therefore as a large proportion of the error originates from the calculated heat capacities, it would suggest that using the 'exp/theory' model, in which the experimental heat capacities are used may provide a more accurate result until more accurate theoretical models can be determined to measure the heat capacity. This is reliant on heat capacity measurements being available experimentally. If this is not viable then the purely theoretical models still provide a reasonable approximation to the absolute and relative sublimation energies.

6.3.4 Estimation of the eutectic composition

Knowledge of the eutectic composition is key for designing effective chiral separation methods (see Chapter 2.9). In both the sublimation and melt cycle, it is usual to assume that the solvation energy or free energy of mixing is the same for both the enantiopure and racemic crystal, and so independent of solvent. This is clearly not the case, as can be seen in Table 2, where the eutectic composition is very system dependent and changing with either the temperature, the solvent or both. To help make the comparison, experimentally measured solubilities of the racemic and enantiopure crystals have been corrected by the calculated COSMO-RS estimate of ΔG_{solv} (at infinite dilution and therefore the same for both crystals) to provide a solvent-dependent non-ideality "error bar" for the experimental eutectic composition at 298 K (see Figure 6.11). The other methods of estimating the eutectic composition using predominantly experimental measurements, i.e. the melt cycle from the experimental differences in the free energies of melting or using the experimental heats of sublimation, have also been compared by Hannes (Figure 6.11).

The eutectic composition can be computationally estimated using the sublimation thermodynamic cycle (Figure 6.1) and so the calculated sublimation

energies, from the sublimations cycle are used. In estimating the eutectic composition, it is assumed that the free energy of mixing with the solvent is identical for both the racemic and enantiopure compounds. These calculated values are compared with experimentally determined solubilities which have been used to estimate the eutectic composition (see Figure 6.11). As, expected, experimentally determined solubilities show variation with solvent for all three systems. This indicates that the solutions of the molecules are not ideal and that it is not accurate to assume the free energy of mixing is identical for the racemic and enantiopure compounds of lactide, naproxen and 3ClMA, as would be expected if the solutions were ideal.

Once again, the main contribution to the accuracy of the eutectic composition comes from the estimation of the lattice energy differences, and so obtaining accurate lattice energies is critical. The computed lattice energy differences systematically overestimate the eutectic composition, which is in agreement with a previous study on amino acids in water.[38] The thermodynamic correction, H_{corr}, only appears to make a significant contribution to the estimation of the eutectic composition for 3ClMA with the heat capacity and zero point energy terms are evaluated using the electronic theory. However entropic differences can account for up for 4 kJ mol^{-1} of the Gibbs free energy difference of sublimation, and so will have an effect on the eutectic composition of all molecules considered.

Where direct sublimation energies could be obtained by experiment, these values correlate well with the experimentally determined eutectic compositions for all molecules. The best estimate of the eutectic composition is from combining the experimental heat capacities with the electronic structure calculations in the ‘exp/theory’ model, which is expected as these results correlated more closely with the experimental results for the sublimation energies (see Chapter 3.3)

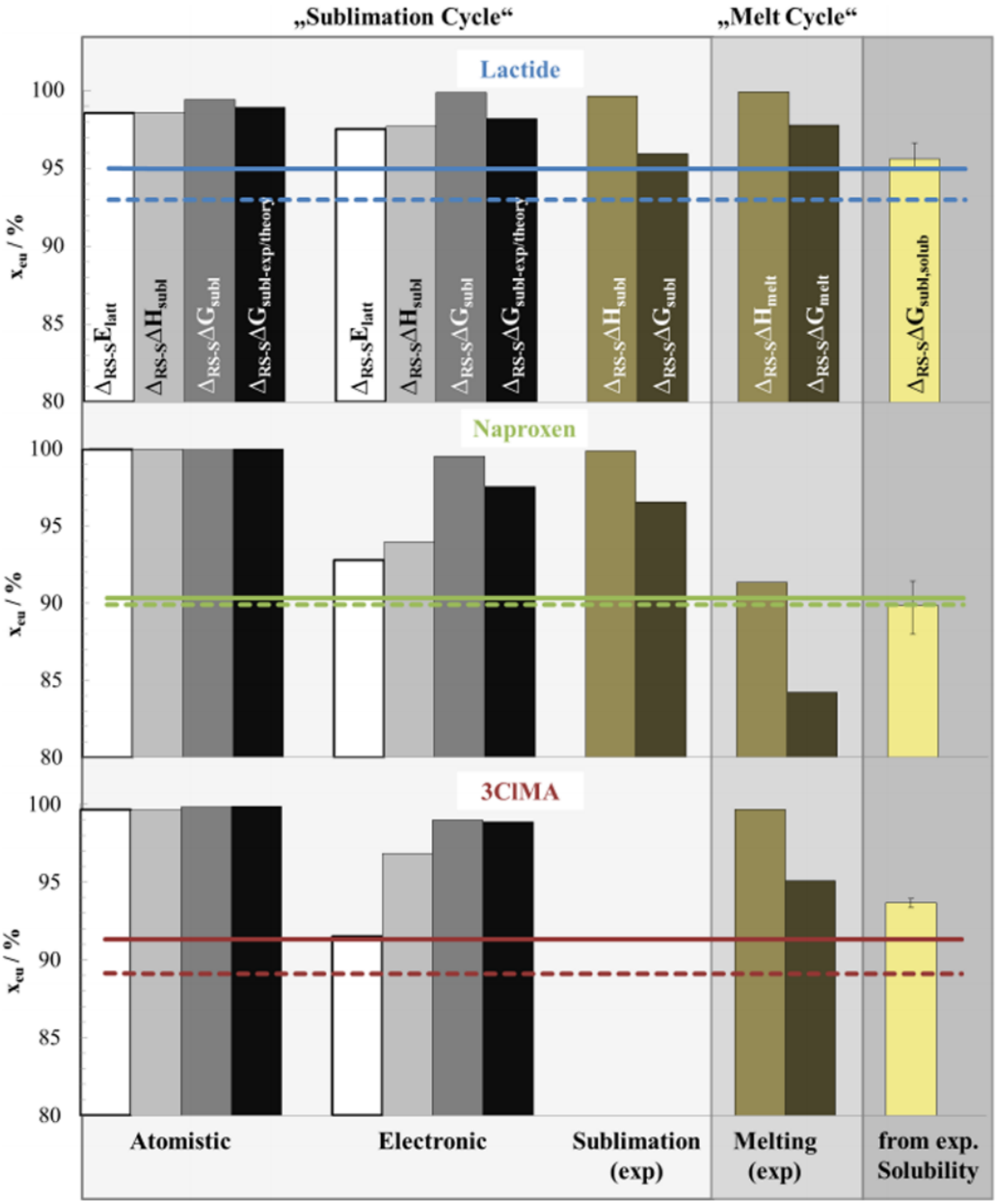


Figure 6-11: The eutectic compositions of lactide, 3ClMA and naproxen estimated at 298 K. The calculated lattice energy, enthalpy corrections and entropy differences are given in white, light grey and dark grey respectively, for both atomistic (separated) and electronic calculations from the sublimation cycle. These are compared with values experimental sublimations (where available) and the experimental melt cycle in green, and measured solubilities (yellow). The maximum and minimum values of the eutectic composition in various solvents are denoted by the solid and dashed line, respectively.

6.3.5 Temperature dependence of the eutectic composition

Experiments investigating the effect of the eutectic composition with temperature have shown that there is a temperature dependence for all three systems considered in this study (see Table 6.2). This temperature dependence cannot be accounted for in the calculations of the solid-state properties, since neither the separated nor electronic models give any variation in the heat

capacity differences at process-relevant temperatures (Figure 6.5). However, it is possible to estimate this by integrating the measured heat capacity differences over a defined temperature range.

Substance	Solvent	Temperature / °C	$x_{eu,exp}$ / %
Lactide	Toluene[39]	5 to 45	94
	Isopropanol[39]	25 to 45	93.4
	Acetone[39]	5 to 35	**95**
	Ethanol[39]	5 to 45 (25)	97.1 to 93.6 (**95.0**)
	Ethyl acetate[39]	15 to 45 (25)	95.8 to 90.2 (**93.2**)
Naproxen	Ethanol[31]	15 to 45 (25)	90.5 to 88.9 (**90.1**)
	Toluene[31]	15 to 45 (25)	90.5 to 88.7 (**89.9**)
	EtOH/ACN 50/50 wt/wt[31]	10 to 40 (25)	90.9 to 89.1 (90.0)
3ClMA	Toluene[30]	40 to 80	91
	Toluene/EA 80/20 v/v[30]	25 to 45	**91.3** to 92.6
	water[30]	5 to 25 (25)	89.4 to 89.1 (**89.1**)
	water[28]	5 to 50	90 to 84
	water / IPA 90/10 wt/wt[40]	20 to 40	0.88 to 0.85

Table 6-2: Experimental eutectic compositions of lactide, naproxen and 3ClMA in various solvent systems and temperatures. The eutectic compositions at 298 K (25 °C) are given in bold.

Integrating the experimental heat capacity differences between 298 and 318 K shows that there is a small temperature effect on the Gibbs free energy of sublimation differences. This causes a shift in the eutectic composition of lactide by -0.4 mol%, naproxen by -0.9 mol% and 3ClMA by +0.6 mol%. The degree of this shift is also solvent dependent for lactide and 3ClMA, where the shift ranges between 0 and -3.0 mol% for lactide and 0 and 1.3 mol% for 3ClMA. The eutectic composition of naproxen is solvent independent.

It is therefore important to note that the choice of solvent used is important when predicting the eutectic composition. The nature of the solvent can not only affect the eutectic composition itself, but can also promote or prevent shifts when the temperature is varied as can be observed see Table 6.2 and with the error bars in Figure 6.11. The eutectic composition and its temperature dependence is complex and relies on more than just the solid-state properties

of the system being investigated as the temperature dependence requires a high temperature variation in the heat capacities, which cannot be reproduced by calculation. To accurately predict the eutectic composition, understanding of the differences in solvation free energies of racemic and enantiopure solutions and extending calculations to account for the non-ideality of the solution in terms of the specific molecule, the solvent used and the solute-solvent interactions is required.

6.3.6 Conclusions

Computed and experimental thermodynamic properties of racemic and enantiopure structures of lactide, naproxen and 3ClMA have been generated and compared. The 'sublimation' thermodynamic cycle provides a plausible alternative means to determine sublimation energies compared to the traditionally used 'melt' cycle, and thus provides a promising method of designing chiral resolution methods by crystallisation.

The sublimation cycle can be evaluated using a combination of calculation and experimental data. While the contribution of the lattice energy is the most dominant term in predicting the sublimation energies, lattice energies alone are not enough to accurately determine the solubility differences between racemic and enantiopure crystals. Current state-of-the-art electronic structure calculations can generate lattice energies with an error range of 3-7 kJ mol^{-1} for small organic crystals, but this may not be accurate enough. The addition of zero-point energy and heat capacities difference is a small, yet important contribution, which allows a more accurate estimation of the absolute and relative sublimation energies to be determined. However, discrepancies between the calculated and experimental heat capacities suggests that the combined 'exp/theory' would provide the best estimation.

The degree of cancellation of errors between racemic and enantiopure compounds is very system specific. Small rigid molecules, such as lactide, with weaker intermolecular interactions that are not affected by molecular structure or vibrations are modelled quite well by both the atomistic and electronic calculations. However, as the molecule becomes more flexible and the degree of hydrogen bonding increases, as with naproxen and 3ClMA, effects such as

thermal expansion need to be considered in order to obtain a more accurate result. Here electronic calculations perform much better than the separated model. The effects caused by thermal expansion indicate that considerations which move away from a simple harmonic approximation should be accounted for.

While it was hoped that the more complete DFT-D CASTEP calculations would be able to provide heat capacity results which more closely matched the experimental data. Unfortunately performing these phonon calculations proved to be far more complicated than expected with many of the structures giving imaginary frequencies which prevent the calculation of heat capacities.
This investigation into the thermodynamic properties of three relatively simple small organic systems has shown that relatable computation of solubilities and eutectic shifts are dependent on more than just the solid-state properties of the system. Molecular dynamics calculations investigating the effects of the system in different solvents and temperatures could help to improve the prediction of the non-ideality of solvation energies and the effect of solvent on eutectic compositions. Lattice energy calculations can be used as a cheap approximation to determine whether the energy differences between the racemic and enantiopure compounds are too large for separation to be possible. However, if the energy differences are small enough then the small thermodynamic quantities become increasingly relevant in making accurate predictions to the eutectic composition. This result shows the current challenges to theory for obtaining accurate lattice energy calculations and thermodynamic properties which can be used to accurately predict experiment.

6.4 References

(1) Palmer, D. S.; Llinas, A.; Morao, I.; Day, G. M.; Goodman, J. M.; Glen, R. C.; Mitchell, J. B. O. Predicting intrinsic aqueous solubility by a thermodynamic cycle. *Molecular Pharmaceutics* **2008**, *5*, 266-279.

(2) Hylton, R. K.; Tizzard, G. J.; Threlfall, T. L.; Ellis, A. L.; Coles, S. J.; Seaton, C. C.; Schulze, E.; Lorenz, H.; Seidel-Morgenstern, A.; Stein, M.; Price, S. L. Are the Crystal Structures of Enantiopure and Racemic Mandelic Acids Determined by Kinetics or Thermodynamics? *Journal of the American Chemical Society* **2015**, *137*, 11095-11104.

(3) van Hummel, G. J.; Harkema, S.; Kohn, F. E.; Feijen, J. Structure of 3,6-dimethyl-1,4-dioxane-2,5-dione [d-,d-(l-,l-)lactide]. *Acta Crystallographica Section B* **1982**, *38*, 1679-1681.

(4) Belenkaya, B. G.; Belsky, V. K.; Dementev, A. I.; Sakharova, V. I.; Chernikova, N. T. Crystal and Molecular Structures of Glycolide and Lactide: Association through CH...O Hydrogen Bonds. *Crystallography Reports* **1997**, *42*, 449-452.

(5) Braun, D. E.; Ardid-Candel, M.; D'Oria, E.; Karamertzanis, P. G.; Arlin, J. B.; Florence, A. J.; Jones, A. G.; Price, S. L. Racemic Naproxen: A Multidisciplinary Structural and Thermodynamic Comparison with the Enantiopure Form. *Crystal Growth & Design* **2011**, *11*, 5659-5669.

(6) Ravikumar, K.; Rajan, S. S.; Pattabhi, V. Structure of Naproxen, C14H14O3. *Acta Crystallographica Section C - Crystal Structure Communications* **1985**, *41*, 280-282.

(7) Allen, F. H. The Cambridge Structural Database: A quarter of a million crystal structures and rising. *Acta Crystallographica Section B - Structural Science* **2002**, *58*, 380-388.

(8) Kazantsev, A. V.; Karamertzanis, P. G.; Adjiman, C. S.; Pantelides, C. C. Efficient Handling of Molecular Flexibility in Lattice Energy Minimization of Organic Crystals. *Journal of Chemical Theory and Computation* **2011**, *7*, 1998-2016.

(9) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A.: Gaussian 03. Gaussian Inc.: Wallingford CT, 2004.

(10) Stone, A. J.: GDMA: A Program for Performing Distributed Multipole Analysis of Wave Functions Calculated Using the Gaussian Program System. 2.2 ed.; University of Cambridge: Cambridge, United Kingdom, 2010.

(11) Price, S. L.; Leslie, M.; Welch, G. W. A.; Habgood, M.; Price, L. S.; Karamertzanis, P. G.; Day, G. M. Modelling Organic Crystal Structures using

Distributed Multipole and Polarizability-Based Model Intermolecular Potentials. *Physical Chemistry Chemical Physics* **2010**, *12*, 8478-8490.

(12) Day, G. M.; Price, S. L.; Leslie, M. Atomistic calculations of phonon frequencies and thermodynamic quantities for crystals of rigid organic molecules. *Journal of Physical Chemistry B* **2003**, *107*, 10919-10933.

(13) Day, G. M.; Price, S. L.; Leslie, M. Elastic constant calculations for molecular organic crystals. *Crystal Growth & Design* **2001**, *1*, 13-27.

(14) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *The Journal of Chemical Physics* **2010**, *132*, 154104.

(15) Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. Electronic structure calculations on workstation computers: The program system turbomole. *Chemical Physics Letters* **1989**, *162*, 165-169.

(16) Clark, S. J.; Segall, M. D.; Pickard, C. J.; Hasnip, P. J.; Probert, M. J.; Refson, K.; Payne, M. C. First principles methods using CASTEP. *Zeitschrift fur Kristallographie* **2005**, *220*, 567-570.

(17) Tkatchenko, A.; Scheffler, M. Accurate Molecular Van Der Waals Interactions from Ground-State Electron Density and Free-Atom Reference Data. *Physical Review Letters* **2009**, *102*, 073005.

(18) Nyman, J.; Day, G. Static and lattice vibrational energy differences between polymorphs. *Crystengcomm* **2015**, *17*, 5154-5165.

(19) Cervinka, C.; Fulem, M.; Stoffel, R. P.; Dronskowski, R. Thermodynamic Properties of Molecular Crystals Calculated within the Quasi-Harmonic Approximation. *Journal of Physical Chemistry A* **2016**, *120*, 2022-2034.

(20) Abdulla, M.; Refson, K.; Friend, R. H.; Haynes, P. D. A first-principles study of the vibrational properties of crystalline tetracene under pressure. *Journal of Physics-Condensed Matter* **2015**, *27*.

(21) Dovesi, R.; Orlando, R.; Erba, A.; Zicovich-Wilson, C. M.; Civalleri, B.; Casassa, S.; Maschio, L.; Ferrabone, M.; De La Pierre, M.; D'Arco, P.; Noël, Y.; Causà, M.; Rérat, M.; Kirtman, B. CRYSTAL14: A program for the ab initio investigation of crystalline solids. *International Journal of Quantum Chemistry* **2014**, *114*, 1287-1317.

(22) Aradi, B.; Hourahine, B.; Frauenheim, T. DFTB+, a Sparse Matrix-Based Implementation of the DFTB Method. *The Journal of Physical Chemistry A* **2007**, *111*, 5678-5684.

(23) Jmol: an open-source Java viewer for chemical structures in 3D.

(24) Emel'yanenko, V.; Verevkin, S.; Pimerzin, A. The thermodynamic properties of DL- and L-lactides. *Russian Journal of Physical Chemistry A* **2009**, *83*, 2013-2021.

(25) Lebedev, Y. G.; Kalugina, T. G.; Kiparisova, E. G. The Thermodynamic Properties of L-Lactide in the Temperature Range 0-430 K. *Russian Journal of Physical Chemistry* **1999**, *73*, 521-527.

(26) Kalugina, T. G.; Lebedev, Y. G.; Lyudvig, Y. B.; Barskaya, I. G. Thermodynamics of dl-Lactide, Polylactide and Polymerization of dl-Lactide in the range of 0-430 K. *Polymer Science U.S.S.R.* **1982**, *24*, 1702-1708.

(27) Le Parlouër, P. Simultaneous TG-DSC: a new technique for thermal analysis. *Thermochimica Acta* **1987**, *121*, 307-322.

(28) Le Minh, T.; Langermann, J. v.; Lorenz, H.; Seidel-Morgenstern, A. Enantiomeric 3-Chloromandelic Acid System: Binary Melting Point Phase

Diagram, Ternary Solubility Phase Diagrams and Polymorphism. *Journal of Pharmaceutical Sciences* **2010**, *99*, 4084-4095.

(29) Le Minh, T. Designing crystallization based-enantiomeric separation for chiral compound-forming systems in consideration of polymorphism and solvate formation. PhD Thesis, Dr. Hut, 2014.

(30) Le Minh, T.; Lorenz, H.; Seidel-Morgenstern, A. Enantioselective Crystallization Exploiting the Shift of Eutectic Compositions in Solid-Liquid Phase Diagrams. *Chemical Engineering & Technology* **2012**, *35*, 1003-1008.

(31) Buchholz, H. K.; Hylton, R. K.; Brandenburg, J. G.; Seidel-Morgenstern, A.; Lorenz, H.; Stein, M.; Price, S. L. Thermochemistry of Racemic and Enantiopure Organic Crystals for Predicting Enantiomer Separation. *Crystal Growth & Design* **2017**.

(32) Buchholz, H.; Emel'yanenko, V. N.; Lorenz, H.; Verevkin, S. P. An Examination of the Phase Transition Thermodynamics of (S)- and (RS)-Naproxen as a Basis for the Design of Enantioselective Crystallization Processes. *Journal of Pharmaceutical Sciences* **2016**, *105*, 1676-1683.

(33) de Barros, T. M. V. R.; Santos, R. C.; Fernandes, A. C.; da Piedade, M. E. M. Accuracy and precision of heat capacity measurements using a heat flux differential scanning calorimeter. *Thermochimica Acta* **1995**, *269–270*, 51-60.

(34) Neau, S. H.; Bhandarkar, S. V.; Hellmuth, E. W. Differential Molar Heat Capacities to Test Ideal Solubility Estimations. *Pharmaceutical Research*, *14*, 601-605.

(35) Donahue, M.; Botonjic-Sehic, E.; Wells, D.; Brown, C. W. Understanding Infrared and Raman Spectra of Pharmaceutical Polymorphs. *American Pharmaceutical Review* **2011**, *14*.

(36) Heit, Y.; Beran, G. How important is thermal expansion for predicting molecular crystal structures and thermochemistry at finite temperatures? *Acta Crystallographica Section B-Structural Science Crystal Engineering and Materials* **2016**, *72*, 514-529.

(37) Chickos, J. S.; Acree, W. E. Enthalpies of Sublimation of Organic and Organometallic Compounds. 1910–2001. *Journal of Physical and Chemical Reference Data* **2002**, *31*, 537-698.

(38) Otero-de-la-Roza, A.; Cao, B. H.; Price, I. K.; Hein, J. E.; Johnson, E. R. Predicting the Relative Solubilities of Racemic and Enantiopure Crystals by Density-Functional Theory. *Angewandte Chemie-International Edition* **2014**, *53*, 7879-7882.

(39) Lorenz, H.; Le Minh, T.; Kaemmerer, H.; Buchholz, H.; Seidel-Morgenstern, A. Exploitation of shifts of eutectic compositions in crystallization-based enantioseparation. *Chemical Engineering Research & Design* **2013**, *91*, 1890-1902.

(40) Zhang, Y.; Ray, A.; Rohani, S. Measurement and prediction of phase diagrams of the enantiomeric 3-chloromandelic acid system. *Chemical Engineering Science* **2009**, *64*, 192-197.

6.5 Appendix

6.5.1 Comparison of Experimental and computed heat capacities

Summary of the experimental heat capacities including the experimental error estimated as average absolute deviation of each measurement point. Calculations using the separated (Ψ_{mol}) and electronic ($\Psi_{crys,}$ HF3c-D3) results are either given exactly at measurement temperature or at the nearest integer value. The cheaper DFTB3-D3 periodic electronic structure estimates are also given for comparison.

6.5.1.1 Lactide

Appendix Table 6.5.1.1.1: Experimental (DSC) and calculated high temperature heat capacities of (*S*)-lactide.

T / K	$C_{P,exp}$ / J mol^{-1} K^{-1}	Abs dev. / J mol^{-1} K^{-1}	$C_{V,calc}$ Separated (Ψ_{mol}) / J mol^{-1} K^{-1}	$C_{V,calc}$ Electronic (Ψ_{crys}) / J mol^{-1} K^{-1}	$C_{V,calc}$ DFTB3-D3 / J mol^{-1} K^{-1}
297.3	187.4	0.3	173.8	174.6	172.2
304.4	193.4	0.5	176.8	177.5	175.3
311.4	194.8	0.6	179.7	180.3	178.2
318.4	201.3	0.7	182.6	183.1	181.2
325.3	203.3	0.5	185.5	185.9	184.1
332.3	206.3	0.5	188.3	188.7	187.0
339.3	208.6	0.6	191.2	191.5	189.9
346.3	212.8	0.7	194.1	194.2	192.7
353.3	218.7	0.2	196.9	196.9	195.6

Appendix Table 6.5.1.1.2: Experimental (from literature[25]) and calculated low temperature heat capacities of (*S*)-lactide.

T / K	$C_{P,exp}$ / J mol^{-1} K^{-1}	Abs dev. / J mol^{-1} K^{-1}	$C_{V,calc}$ Separated (Ψ_{mol}) / J mol^{-1} K^{-1}	$C_{V,calc}$ Electronic (Ψ_{crys}) / J mol^{-1} K^{-1}	$C_{V,calc}$ DFTB3-D3 / J mol^{-1} K^{-1}
5	0.404	0.01	0	0	0
10	2.95	0.06	2.4	6.2	5.0
15	7.58	0.15	7.3	13.3	10.7
20	13.2	0.26	13.6	20.6	16.6
25	18.95	0.38	20.4	27.4	22.4
30	24.67	0.12	27.0	33.5	27.9
40	35.2	0.18	38.8	44.1	38.1
50	44.36	0.22	48.8	53.1	47.4
100	82.3	0.16	84.3	87.5	82.9
150	110.5	0.22	110.0	112.9	108.4
200	135.1	0.27	132.2	134.6	130.6
250	160.3	0.32	153.6	155.3	152.0
298.15	181.5	0.36	174.1	175.0	172.6
300	182.3	0.36	174.9	175.3	173.0
350	210	3.15	195.6	195.7	194.3
366.6	219.5	3.29	202.3	202.1	201.0

Appendix Table 6.5.1.1.3: Experimental (DSC) and calculated high temperature heat capacities of (*RS*)-lactide.

T / K	$C_{P,exp}$ / J mol^{-1} K^{-1}	Abs dev. / J mol^{-1} K^{-1}	$C_{V,calc}$ Separated (Ψ_{mol}) / J mol^{-1} K^{-1}	$C_{V,calc}$ Electronic (Ψ_{crys}) / J mol^{-1} K^{-1}	$C_{V,calc}$ DFTB3-D3 / J mol^{-1} K^{-1}
298.8	190.2	0.2	174.5	175.5	173.6
308.8	194.2	0.2	178.7	179.5	177.8
318.8	200.2	0.3	182.9	183.5	182.0
328.8	204.6	0.1	187.0	187.5	186.2
338.8	206.5	0.1	191.1	191.5	190.3
348.8	213.5	0	195.2	195.4	194.4
358.8	215.9	0.2	199.2	199.3	198.4
368.8	223.1	0.2	203.2	203.1	202.4
378.8	228.1	0.2	207.1	206.9	206.4

Appendix Table 6.5.1.1.4: Experimental (from literature[26]) and calculated low temperature heat capacities of (*RS*)-lactide.

T / K	$C_{P,exp}$ / J mol^{-1} K^{-1}	Abs dev. / J mol^{-1} K^{-1}	$C_{V,calc}$ Separated (Ψ_{mol}) / J mol^{-1} K^{-1}	$C_{V,calc}$ Electronic (Ψ_{crys}) / J mol^{-1} K^{-1}	$C_{V,calc}$ DFTB3-D3 / J mol^{-1} K^{-1}
5	0.307	0	0	6.8	5.8
10	2.42	0.01	3.4	8.1	7.1
15	7.211	0.04	9.3	15.4	13.4
20	13.13	0.07	16.3	22.8	19.6
25	19.09	0.1	23.2	29.6	25.5
30	25.01	0.13	29.7	35.7	31.2
40	36.1	0.18	41.0	45.9	41.5
50	45.47	0.09	50.5	54.5	50.8
60	54.06	0.11	58.7	62.2	59.0
80	69.18	0.14	72.8	75.9	73.1
100	82.64	0.17	84.9	87.9	85.0
200	137.3	0.27	132.4	134.9	131.6
250	162	0.32	153.8	155.6	152.9
298.15	184.3	0.37	174.2	175.2	173.3
310	190.8	0.38	179.2	180.0	178.3
360	220.5	0.44	199.7	199.7	198.9
397	243.7	0.49	214.1	213.7	213.4

6.5.1.2 Naproxen

Appendix Table 6.5.1.2.1: Experimental (DSC) and calculated high temperature heat capacities of (*S*)-naproxen.

T / K	$C_{P,exp}$ / J mol^{-1} K^{-1}	Abs dev. / J mol^{-1} K^{-1}	$C_{V,calc}$ Separated (Ψ_{mol}) / J mol^{-1} K^{-1}	$C_{V,calc}$ Electronic (Ψ_{crys}) / J mol^{-1} K^{-1}	$C_{V,calc}$ DFTB3-D3 / J mol^{-1} K^{-1}
298.4	296.3	4.1	277.4	270.5	269.7
307.8	303.2	3.1	284.8	277.8	277.0
317	314.3	3.7	292.0	285.0	284.1
326.3	320.7	3.2	299.3	292.2	291.2
335.6	329.7	4.9	306.4	299.3	298.3
344.9	335.2	4.1	313.5	306.4	305.3
354.2	346.4	3.7	320.6	313.4	312.2
363.5	353.5	3.9	327.5	320.3	319.0
372.7	362.1	5.7	334.2	327.0	325.7
382	366.7	5.3	341.0	333.8	332.4
391.3	379.5	5	347.6	340.4	339.0
400.6	389.6	6.7	354.2	347.0	345.5
409.9	404.9	6	360.6	353.6	352.0

Appendix Table 6.5.1.2.2: Experimental (DHPC) and calculated low temperature heat capacities of (*S*)-naproxen.

T / K	$C_{P,exp}$ / J mol^{-1} K^{-1}	Abs dev. / J mol^{-1} K^{-1}	$C_{V,calc}$ Separated (Ψ_{mol}) / J mol^{-1} K^{-1}	$C_{V,calc}$ Electronic (Ψ_{crys}) / J mol^{-1} K^{-1}	$C_{V,calc}$ DFTB3-D3 / J mol^{-1} K^{-1}
1.83	0.026	0.0006	0	0	0
2.94	0.108	0.0003	0	0	0
4.04	0.277	0.0009	0	0	0
5.13	0.584	0.0005	0	0	0
6.21	1.052	0.0086	0	0	0
7.29	1.702	0.0042	0	3.6	5.5
8.36	2.509	0.006	3.5	5.2	7.3
9.43	3.481	0.0266	5.2	6.8	9.0
10.48	4.572	0.0076	7.2	8.5	10.8
11.55	5.841	0.0224	9.3	10.3	12.6
12.64	7.15	0.0192	1.2	12.2	14.6
13.7	8.587	0.0295	14.0	14.1	16.5
14.8	10.156	0.0346	16.4	16.1	18.5
15.89	11.706	0.0378	18.8	18.1	20.5
16.99	13.382	0.0322	21.2	20.0	22.5
18.08	15.077	0.1025	23.5	22.0	24.4
19.19	16.631	0.0167	25.8	24.0	26.4
20.31	18.337	0.0507	28.0	25.9	28.3
20.24	18.309	0.0131	27.9	25.8	28.2
28.25	30.364	0.0748	42.3	38.8	40.8
36.2	41.823	0.017	54.2	50.4	51.9
44.12	52.722	0.0082	64.7	61.0	61.9
52.05	62.677	0.245	74.3	70.8	71.3
59.81	71.979	0.0537	82.9	79.6	79.8
67.68	80.499	0.0075	91.1	88.0	87.9
75.52	88.708	0.0386	98.7	95.8	95.4
83.38	96.217	0.4746	106.0	103.1	102.6
91.31	103.502	0.4718	113.0	110.2	109.4
99.22	110.426	0.4506	119.7	116.9	116.0
107.14	117.252	0.3308	126.2	123.4	122.4
115.05	123.764	0.0202	132.6	129.8	128.6
122.91	130.266	0.0634	138.9	135.9	134.6
130.74	136.46	0.5387	145.0	141.9	140.55
138.62	142.562	0.5062	151.1	147.9	146.5
146.49	148.781	0.3258	157.3	153.8	152.4
154.39	154.372	0.5915	163.4	159.7	158.3
162.25	160.456	0.1456	169.5	165.6	164.2
170.22	166.402	0.7301	175.6	171.5	170.2
178.05	172.75	0.1308	181.7	177.4	176.1
185.93	178.462	0.1536	187.9	183.3	182.1
193.76	184.657	0.4778	194.0	189.2	188.1
201.61	190.72	0.4803	200.2	195.1	194.1

Appendix Table 6.5.1.2.3: Experimental (DSC) and calculated high temperature heat capacities of (*RS*)-naproxen.

T / K	$C_{P,exp}$ / J mol^{-1} K^{-1}	Abs dev. / J mol^{-1} K^{-1}	$C_{V,calc}$ Separated (Ψ_{mol}) / J mol^{-1} K^{-1}	$C_{V,calc}$ Electronic (Ψ_{crys}) / J mol^{-1} K^{-1}	$C_{V,calc}$ DFTB3-D3 / J mol^{-1} K^{-1}
298.4	297.4	1.6	277.3	272.7	268.8
307.8	304.9	1.6	284.7	280.0	276.1
317.1	316.8	1.5	292.0	287.3	283.2
326.3	322.9	1.5	299.2	294.3	290.2
335.6	331	1.7	306.4	301.5	297.3
344.9	338.1	1.5	313.5	308.5	304.3
354.2	349.5	1.4	320.5	315.5	311.2
363.5	356.8	1.4	327.4	322.4	318.0
372.7	366.2	1.5	334.2	329.1	324.7
382	372.2	1	340.9	335.9	331.4
391.3	383.5	1.5	347.6	342.5	338.0
400.6	393	1	354.1	349.1	344.4
409.9	404.2	1	360.6	355.6	351.0

Appendix Table 6.5.1.2.4: Experimental (DHPC) and calculated low temperature heat capacities of (*RS*)-naproxen.

T / K	$C_{P,exp}$ / J mol^{-1} K^{-1}	Abs dev. / J mol^{-1} K^{-1}	$C_{V,calc}$ Separated (Ψ_{mol}) / J mol^{-1} K^{-1}	$C_{V,calc}$ Electronic (Ψ_{crys}) / J mol^{-1} K^{-1}	$C_{V,calc}$ DFTB3-D3 / J mol^{-1} K^{-1}
1.82	0.041	0.0009	0	0	0
2.93	0.144	0.0001	0	0	0
4.02	0.344	0.0007	0	0	0
5.11	0.691	0.0015	0	0	0
6.18	1.202	0.0019	0	0	0
7.26	1.876	0.0011	0	5.5	5.9
8.33	2.695	0.0045	3.6	7.0	7.6
9.39	3.647	0.0064	5.3	8.6	9.3
10.44	4.688	0.0106	7.1	10.2	10.9
11.52	5.863	0.0173	9.1	11.9	12.7
12.61	7.16	0.0168	1.1	13.8	14.5
13.67	8.483	0.0156	13.3	15.6	16.3
14.76	9.997	0.0501	15.4	17.5	18.2
15.87	11.4	0.0194	17.6	19.4	20.1
16.94	12.839	0.0633	19.7	21.3	21.9
18.04	14.505	0.0295	21.8	23.2	23.8
19.12	16.003	0.041	23.8	25.1	25.6
20.23	17.621	0.0074	25.9	27.0	27.4
28.15	29.419	0.017	39.5	40.3	38.3
36.01	41.639	0.0783	51.5	52.2	49.6
43.99	52.479	0.3166	62.4	63.2	60.1
51.89	62.674	0.2753	72.3	73.2	69.7
59.65	71.938	0.1768	81.3	82.2	78.5
67.46	81.09	0.2129	89.6	90.6	86.6
75.32	89.56	0.2995	97.5	98.5	94.3
83.18	97.528	0.2897	104.9	105.8	101.4
91.13	104.962	0.39	112.0	112.8	108.3
99.06	111.877	0.2272	118.9	119.6	115.7
106.9	119.308	0.212	125.5	126.0	121.9
114.8	126.012	0.2131	131.9	132.3	128.1
122.66	133.108	0.5202	138.2	138.4	134.1
130.51	139.823	0.5438	144.4	144.4	150.4
138.43	145.839	0.1142	150.6	150.3	145.9
146.29	152.196	0.1019	156.8	156.2	151.8
154.19	158.735	0.4662	162.9	162.1	157.7
162.1	165.16	0.8372	169.1	168.0	163.6
169.98	170.913	1.6973	175.2	173.9	169.5
177.89	178.229	0.6492	181.4	179.7	175.4
185.74	184.442	1.9177	187.5	185.6	181.4
193.57	190.818	1.5547	193.7	191.5	187.3
201.34	198.479	0.049	199.8	197.4	193.3

6.5.1.3 3ClMA

Appendix Table 6.5.1.3.1: Experimental high temperatures (DSC) heat capacities of (*R*)-3ClMA and calculated high temperatures heat capacities of (*S*)-3ClMA.

T / K	$C_{P,exp}$ / J mol^{-1} K^{-1}	Abs dev. / J mol^{-1} K^{-1}	$C_{V,calc}$ Separated (Ψ_{mol}) / J mol^{-1} K^{-1}	$C_{V,calc}$ Electronic (Ψ_{crys}) / J mol^{-1} K^{-1}	$C_{V,calc}$ DFTB3-D3 / J mol^{-1} K^{-1}
305.1	216.2	1.4	200.1	191.4	191.7
312.6	218.9	1.6	203.8	195.1	195.5
320.1	220.2	1.7	207.5	198.9	199.2
327.6	225.1	1.6	211.1	211.1	202.8
335.1	229.8	1.7	214.7	206.2	206.4
342.6	233.8	1.7	218.2	209.8	210.0
350	238.2	1.8	221.7	213.3	213.5
357.6	241.8	1.5	225.1	216.9	217.0

Appendix Table 6.5.1.3.2: Experimental low temperatures (DHPC) heat capacities of (*R*)-3ClMA and calculated low temperatures heat capacities of (*S*)-3ClMA.

T / K	$C_{P,exp}$ / J mol^{-1} K^{-1}	Abs dev. / J mol^{-1} K^{-1}	$C_{V,calc}$ Separated (Ψ_{mol}) / J mol^{-1} K^{-1}	$C_{V,calc}$ Electronic (Ψ_{crys}) / J mol^{-1} K^{-1}	$C_{V,calc}$ DFTB3-D3 / J mol^{-1} K^{-1}
1.83	0.019	0.0004	0	0	0
2.94	0.089	0.0004	0	0	0
4.04	0.252	0.0005	0	0	0
5.14	0.56	0.0004	0	0	0
6.23	1.013	0.0219	0	0	0
7.3	1.609	0.0182	0	0	0
8.37	2.339	0.0021	3.9	2.81	2.6
9.44	3.171	0.0188	5.2	4.07	3.9
10.5	4.096	0.0088	6.6	5.34	5.2
11.57	5.131	0.0394	8.0	6.46	6.4
12.65	6.278	0.0018	9.5	7.64	7.6
13.72	7.436	0.0195	11.1	8.88	9.0
14.86	8.779	0.0132	12.8	10.24	10.4
15.92	10.081	0.0092	14.4	11.54	11.8
17.03	11.385	0.1359	16.1	12.93	13.2
18.12	12.791	0.0057	17.7	14.32	14.6
19.19	14.159	0.0082	19.4	15.70	16.0
20.28	15.547	0.006	21.0	17.11	17.5
28.2	25.616	0.0843	32.3	27.19	27.3
36.07	35.269	0.1234	42.0	36.25	36.0
43.95	43.32	0.059	50.2	44.21	43.7
51.84	50.66	0.0379	57.4	51.27	50.5
59.73	57.258	0.0224	63.8	57.64	56.7
67.61	63.415	0.1527	69.7	63.47	62.5
75.52	69.076	0.1089	75.1	68.92	67.9
83.42	74.214	0.1055	80.2	74.05	72.9
91.31	79.363	0.0166	85.1	78.91	77.8
99.19	84.29	0.1131	89.8	83.58	82.4
107.07	89.198	0.1942	94.4	88.08	86.9
114.96	93.763	0.0797	98.9	92.47	91.3
122.83	98.397	0.1877	103.3	96.75	95.6
130.7	102.855	0.0481	107.7	100.96	99.9
138.6	107.178	0.1289	112.1	105.12	104.1
146.48	111.444	0.1976	116.4	109.24	108.3
154.36	116.123	0.1625	120.7	113.33	112.5
162.24	119.855	0.0477	125.0	117.41	116.7
170.12	124.477	0.5773	129.3	121.48	120.9
178	128.704	0.1196	133.6	125.55	125.1
185.86	133.404	0.144	137.8	129.61	129.3
193.67	137.772	0.2761	142.0	133.65	133.4
201.48	141.691	0.2558	146.2	137.71	137.6

Appendix Table 6.5.3.3: Experimental (DSC) and calculated high temperature heat capacities of (*RS*)-3ClMA.

T / K	$C_{P,exp}$ / J mol^{-1} K^{-1}	Abs dev. / J mol^{-1} K^{-1}	$C_{V,calc}$ Separated (Ψ_{mol}) / J mol^{-1} K^{-1}	$C_{V,calc}$ Electronic (Ψ_{crys}) / J mol^{-1} K^{-1}	$C_{V,calc}$ DFTB3-D3 / J mol^{-1} K^{-1}
298.4	202.2	1.2	196.8	187.5	190.0
307.6	206.6	1.2	201.4	192.1	194.6
316.7	211.8	1.2	205.8	196.6	199.0
325.9	216.6	1.3	210.3	201.1	203.5
335.1	220.2	1.7	214.7	205.5	207.9
344.2	225.9	1.6	218.9	209.9	212.2
351.7	229.5	1.8	222.4	213.4	215.7
360.9	233.8	1.4	226.6	217.7	219.9
370.1	239.9	1.7	230.7	221.9	224.1
379.3	250.5	0.7	234.8	226.0	228.1

Appendix Table 6.5.3.1: Experimental (DHPC) and calculated low temperature heat capacities of (*RS*)-3CIMA.

T / K	$C_{P,exp}$ / J mol^{-1} K^{-1}	Abs dev. / J mol^{-1} K^{-1}	$C_{V,calc}$ Separated (Ψ_{mol}) / J mol^{-1} K^{-1}	$C_{V,calc}$ Electronic (Ψ_{crys}) / J mol^{-1} K^{-1}	$C_{V,calc}$ DFTB3-D3 / J mol^{-1} K^{-1}
1.83	0.038	0.0004	0	0	0
2.93	0.178	0.002	0	0	0
4.03	0.495	0.0057	0	0	0
5.12	1.05	0.0054	0	0	0
6.2	1.805	0.003	0	0	0
7.28	2.744	0.0196	0	0	0
8.35	3.779	0.0046	4.4	4.3	1.9
9.43	4.941	0.0241	6.0	6.3	2.9
10.5	6.115	0.0316	7.6	8.0	3.9
11.58	7.403	0.0464	9.4	9.3	4.9
12.66	8.78	0.0086	11.2	10.7	6.0
13.71	10.165	0.0113	13.0	11.9	7.1
14.84	11.615	0.0051	14.9	13.3	8.4
15.91	12.968	0.0103	16.7	14.6	9.7
16.99	14.412	0.02	18.5	15.9	10.9
18.09	15.78	0.0255	20.3	17.2	12.3
19.18	17.201	0.0132	22.0	18.5	13.6
20.27	18.601	0.0203	23.7	19.8	15.0
28.18	28.46	0.0183	34.7	29.0	24.8
36.13	36.768	0.1083	43.7	37.8	34.0
44.01	44.399	0.058	51.3	45.6	42.2
51.92	51.356	0.0773	58.1	52.6	49.5
59.75	57.405	0.0115	64.1	58.7	56.1
67.63	63.049	0.105	69.8	64.3	62.1
75.54	68.038	0.0812	75.1	69.5	67.7
83.46	72.679	0.0397	80.2	47.4	72.9
91.34	77.474	0.0393	85.0	79.1	77.9
99.23	81.977	0.0479	89.7	83.6	82.7
107.12	86.573	0.0769	94.3	88.0	87.4
115.02	90.988	0.1168	98.8	92.3	91.9
122.87	95.187	0.0945	103.2	96.5	96.4
130.75	99.649	0.0613	107.6	100.7	100.8
138.66	104.261	0.0591	112.0	104.9	105.2
146.56	108.051	0.1736	116.4	109.0	109.5
154.43	112.668	0.2015	120.7	113.1	113.8
162.3	117.122	0.2598	125.0	117.2	118.1
170.18	120.819	0.1211	129.3	121.2	122.4
178.08	125.131	0.4356	133.5	125.3	126.7
185.93	129.724	0.3405	137.8	129.4	130.9
193.78	133.535	0.1761	142.0	133.4	135.1
201.56	138.033	0.2691	146.2	137.5	139.3

6.5.2 Experimental melting, sublimation and heat capacity data

6.5.2.1 Melting properties

Appendix Table 6.5.2.1.1: Experimental melting temperatures and enthalpies. The Gibbs free energies of melting are calculated from $\Delta G_{melt} = \Delta H_{melt} - T \cdot \Delta H_{melt} / T_{melt}$, with T = 298.15 K.

	ΔH_{melt} / kJ/mol	T_{melt} / K	ΔG_{melt} / kJ/mol
(*S*)-lactide[25]	16.9	366.6	3.16
(*RS*)-lactide[25]	24.7	397.1	6.15
(*RS*-*S*)	7.8		2.99
(*S*)-Naproxen[32]	31.6	429.2	9.65
(*RS*)-Naproxen[32]	32.8	429.0	10.00
(*RS*-*S*)	1.2		0.36
(*R*)-3ClMA[28]	22.6	376.4	4.69
(*RS*)-3ClMA[28]	27.9	391.1	6.64
(*RS*-*S*)	5.4		1.95

7 Conclusions and Future Work

The work in this thesis has involved exploring computed crystal energy landscape of chiral organic crystal structures as an aid to the rational design of chiral separation processes. The ability of CSP to predict and reliably reproduce experimental crystal structures has been assessed through the study of prototypical chiral molecules. The assumption of cancellation of errors between racemic and enantiopure crystals has been investigated to build a more complete understanding of how computation can be used to help aid designing chiral resolution processes.

7.1 Can CSP be used to predict the crystal structures of chiral molecules?

Chapters 3 and 4 used CSP methods to generate crystal energy landscapes of two chiral molecules, 3ClMA and lactide, where experimental crystal structures were already known. The 3ClMA and lactide molecules are very different. 3ClMA is small, flexible, able to hydrogen bond and with a relatively large number of known polymorphs, some of which were discovered and structurally characterised during the course of this thesis. Lactide, on the other hand, is small, rigid, has no hydrogen bonding capabilities and has no known polymorphs which have been discovered experimentally.

In chapter 3 the crystal energy landscape of 3ClMA was investigated. The flexible nature of the 3ClMA resulted in a very dense crystal energy landscape with many hypothetical crystal structures observed lower in energy than the experimentally known structures. The diversity of the hypothetical structures generated are a balance of different phenyl group packings, hydrogen bonded motifs and conformational flexibility which goes beyond simple explanations of nucleation and crystal growth and has not been explained by MD simulation. The phenyl packing disorder of S-3ClMA form 1 and its relative stability to the most stable experimentally determined racemic form is a temperature dependent, kinetic effect rather than a result of thermodynamics. The CEL of 3ClMA highlights the fine balance between the kinetic effects of crystal

nucleation and growth and obtaining a thermodynamically stable crystal structure.

In Chapter 4 the crystal energy landscape of lactide was explored. The experimentally known racemic crystal structure, BICVIS, was successfully found to be the global minimum in the search if free energies are included. While the Z'=3 enantiopure structure, NAHNOZ, was beyond the scope of the search lower energy, less dense, Z'=1 hypothetical structures were observed. Analysis into the packings of the structures generated in the CEL show that there are a number of structures, close in energy, which appear to be related, for example by the sliding of layers within the crystal structure. This leads to questions surrounding the accuracy of the potential used to model the weak van der Waals forces which dominate the packing forces of lactides as well as indicating the possibility of disordered systems.

For both 3ClMA and lactide it was observed that the experimental enantiopure structures were high Z' structures. The tendency of chiral molecules to be high Z' means that the exact crystal structure would only be found in a CSP study if the search space was expanded to structures that were Z'>1. While only Z'=1 structures are considered in the search these still provide us with the idea of the range of packings and their relative energies which could be accessible to enantiopure molecules. Despite this there are many enantiopure Z'=1 structures found in the searches of 3ClMA and lactide which have not been observed experimentally and are estimated to be lower in energy. This leads to questions as to whether the experimentally determined enantiopure crystal structures are actually metastable kinetic artefacts and their high Z' is a result of difficulties in growing enantiopure crystals.

In Chapter 5 I successfully used CSP methods to predict the crystal packing of molecule XXII despite the uncertainty in the flexibility and accuracy of the sulphur potential used. Due to the small size and relative rigidity of molecule XXII many different methods could be used. Our comparatively computationally cheap method was shown to provide comparative results with more complex DFT-D methods. The results for XXII across all submissions for the Blind Test indicate that the addition of free energies was an important factor in placing the

experimental structure at the minimum of the crystal energy landscape yet were not routinely being performed by many of the groups who participated in the Blind Test. The 6th Blind Test provides a useful means for assessing the current strengths and weaknesses in crystal structure prediction, with this Blind Test showing that there have been considerable advances in the complexity of the molecular systems that can be tackled by CSP methods. However, the variation in the relative lattice energies, let alone reranking with harmonic free energy estimates, of the five known polymorphs of a pharmaceutical show that even the most sophisticated periodic electronic structure methods are far from converging on the stability order of the polymorphs.

7.2 Is chiral separation of crystal structures predictable?

In Chapter 6 computed and experimental thermodynamic properties of racemic and enantiopure crystal structures were compared through the 'sublimation' thermodynamic cycle. The most dominant contribution to the sublimation enthalpy is the lattice energy, however this alone is not enough to determine solubility differences between racemic and enantiopure crystal structures. Considering thermodynamic contributions, beyond the –2RT approximation, by including heat capacity differences and zero point energies is important to an extent that is very dependent on the crystal structures. Discrepancies between calculated and experimental heat capacities, particularly at high temperatures, indicate that a combined 'exp/theory' model using measured heat capacities would provide a better estimation of crystal thermodynamics than pure computation.

When comparing enantiopure and racemic crystal structures, it was hoped that there would be a large cancellation of errors as both structures contain the same molecule. However, it has been found that the degree of cancellation is very system specific as differences in intermolecular bonding between racemic and enantiopure crystal structures and resulting effects such as anisotropic thermal expansion mean that assuming the simple harmonic approximation used to estimate the thermodynamic effects is not always valid.

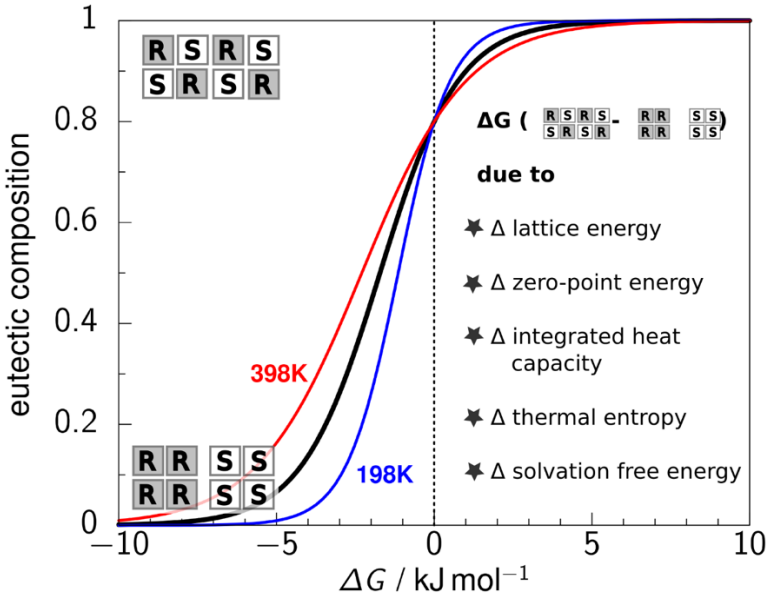


Figure 7-1: The dependence of the eutectic composition on the energy difference between the racemic and enantiopure crystal, for different temperatures. This figure was drawn from the equations in Chapter 2 by Gerit Brandenburg as a potential Table of Contents graphic for the paper from Chapter 6.

The solubility differences between enantiopure and racemic crystals are directly related to the eutectic composition. As seen in Figure 7.1, the idealised enantiomeric excess at 298 K changes rapidly over a very small difference in the free energy of solution. This emphasises the sensitivity of these systems and how accurate and reliable the theoretical models need to be to readily predict the eutectic composition. Lattice energy calculations can be used as a cheap approximation to determine whether the energy differences between the racemic and enantiopure compounds are too large for separation to be possible. However, if the energy differences are small enough then the small thermodynamic quantities become increasingly relevant in making accurate predictions of the eutectic composition

7.3 Future work

This was a first investigation in a collaboration with theoreticians and experimentalists at the MPI into seeing if current theoretical models could be used to accurately model lattice energies and thermodynamic effects in order to predict solubility differences as an aid to chiral separation techniques. It had been proposed on the basis of previous work on naproxen which showed that just considering lattice energy differences was not enough. While some

advancements and insights have been made on this topic there is still progress to be made.

Currently CSP prediction methods are good for small, relatively rigid molecules. However, many pharmaceutical-type molecules, which would benefit from CSP studies, are still too large and/or flexible to be readily studied. By improving current CSP techniques to handle more degrees of freedom to account for flexibility we will be able to generate more accurate crystal energy landscapes and energetic ranking of polymorphs. However, the observation that enantiopure crystal structures may be Z'>1 and metastable relative to the lowest CSP generated structure raises many questions as to whether their crystallisation is determined by thermodynamics (as assumed by CSP) or by kinetics.

This study has shown that the simple –2RT approximation to predicting the thermodynamic models of organic crystal structures is not satisfactory and there is not a simple cancellation of errors between racemic and enantiopure crystal structures. In order to more accurately determine the free energies and phonon modes required to calculate the thermodynamic quantities we need to go beyond calculating k=0 harmonic phonon modes, both with the separated atomistic model and with electronic DFT-D methods. This is now possible for the separated atomistic model, but this approach is limited by assuming that the molecules are rigid. Prof. Keith Refson and Dr Rui Guo are still working on methods to reliably accurately model phonons across the full Brillouin zone for this size of organic crystals with Castep.

The cheaper electronic structure HF-3c methods used in Chapter 6 were only just affordable enough to go beyond k=0 phonon modes and yet still cannot readily model the thermal expansion required for studying free energies of solution and eutectic composition. Dr Gerit Brandenberg has recently used the cheaper DFTB3-D3 method to model the thermal expansion of form III of carbamazepine. It is clear that considerable more work is required to be able to routinely determine free energy differences between polymorphs or enantiopure and racemic crystals, which have very different size and shapes of unit cell.

The ability to use CSP methods to aid chiral separation processes has been much harder than anticipated. There is still a need to validate all of the assumptions that are made in thermodynamic calculations against what is seen experimentally. However, this study has provided useful insights into what is required by theory in order to predict absolute and relative solubitilites, for example the reproduction of experimental high temperature heat capacity and the IR spectra differences, and the direction in which theory and computation need to advance in order to achieve this.