UNIVERSITY OF SOUTHAMPTON

FACULTY OF NATURAL AND ENVIRONMENTAL SCIENCES

School of Chemistry

Computational predictions of structures, inclusion behaviour and properties of organic molecular crystals

by

Jonas Nyman

Thesis for the degree of Doctor of Philosophy

May 2017

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF NATURAL AND ENVIRONMENTAL SCIENCES
School of Chemistry

Doctor of Philosophy

COMPUTATIONAL PREDICTIONS OF STRUCTURES, INCLUSION BEHAVIOUR AND PROPERTIES OF ORGANIC MOLECULAR CRYSTALS

by Jonas Nyman

This thesis is about the computational prediction of crystal structures and their properties. Polymorphism, where the same molecule crystallizes in more than one structure is investigated. Several structures of polymorphs and clathrates, porous crystals that adsorb gases, are predicted without prior experimental data. Rigid-molecule lattice dynamics in an anisotropic force field is used to calculate temperature-dependent properties of large sets of crystals. Using the lattice-vibrational free energy as scoring function in crystal structure prediction is discussed. Brillouin zone sampling and convergence difficulties of lattice dynamics calculations are addressed, with a kernel density estimation method offered as a solution.

The advantage of computationally affordable force field methods over electronic structure methods will be described. I demonstrate that multipole-based force fields can be comparable in accuracy to dispersion-corrected generalized gradient approximation density functional theory and that such force fields are very suitable for crystal structure prediction as they allow the calculation of realistic free energies for hundreds of structures at a relatively small computational cost.

For the prediction of clathrate structures, I show that a combination of the free energy and the guest-to-cavity volume ratio can be a suitable scoring function. I suggest that porous structures in predicted crystal energy landscapes should be interpreted as possible solvates, clathrates and hydrates, and that these should be carefully considered in the analysis of prediction results.

Contents

List of Figures

List of Tables

Declaration of Authorship

I, Jonas Nyman, declare that the thesis entitled *Computational predictions of structures, inclusion behaviour and properties of organic molecular crystals* and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

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

Signed:.....

Date:.....

Acknowledgements

I thank Graeme M. Day for being an excellent supervisor and scientific mentor. Your deep knowledge of computational chemistry and attention to detail is widely known and you have taught me much. Perhaps unintentionally, you have also taught me the value of a serene and stoic mind in the pursuit of science.

I am grateful to Kim Bolton, Dragu Atanasiu and Peter Ahlström at the University of Borås, for introducing me to computational chemistry and encouraging me to pursue a career in science.

I thank Peter Bygrave for patiently answering the thousands of questions I bombarded him with during my first year in Southampton.

I got much pleasure out of supervising two talented undergraduate students. Orla Sheehan Punduke and Jacob Marshall performed preliminary calculations that proved important for the methods used in this thesis.

The summer school in Erice by the Ettore Majorana Foundation and Centre for Scientific Culture in 2015 was an amazing and unforgettable experience, for which I am deeply grateful.

I thank Ulrich J. Griesser for hosting me in his laboratory in Innsbruck for an exciting week in April 2016. I thank Ana Ivanović for instructing me in experimental techniques and for fruitful scientific collaboration.

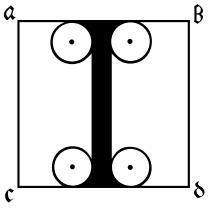
Perttu Lantto and Marcin Selent at the University of Oulu are acknowledged for their important work on fluorophenol clathrates. Without their long-term commitment to a challenging project, this thesis would be lacking significantly in scope, depth and quality.

I thank the Willborg and Bergman families for their support during the time I wrote this thesis.

During my time as a PhD student, I have been fully funded through a research grant by the European Research Council.

Chapter 1

Introduction and background



In this thesis I demonstrate how large-scale computer calculations can be used to predict new crystals and their properties. I explore computer algorithms that can be used for such purposes and apply them to real-world problems. Computer calculations will be used to predict possible crystal structures of organic molecules, and their properties. The results are then compared to experimental data. The accuracy of the calculations is assessed and several techniques are described that improve on currently used methods. Differences in properties between alternative crystal structures, polymorphs, of the same compound will be calculated and analysed in some detail.

More specifically, this thesis applies crystal structure prediction and lattice dynamical free energy calculations to predict the structure and properties of polymorphs and porous crystals. Computational results are compared with experimental observations, which in some cases have been obtained through collaborations with other research groups.

This chapter introduces the background and history of the research field, and familiarises the reader with common terms and concepts of computational crystallography used throughout the thesis.

1.1 Organic molecular crystal structures

1.1.1 Definitions and nomenclature

This thesis will only consider periodic crystals. A crystal structure is a solid phase of matter in which the constituent atoms are arranged in a pattern which is periodic in the three spatial dimensions. In this thesis all crystal structures will be modelled as ideal, defect-free and formally infinitely large, so that there are no surface effects.

The periodically repeating unit is a parallelepiped called the unit cell of the crystal. It is specified by the lengths of its sides a, b, c and the angles between them α, β, γ, or by the vectors spanning the parallelepiped **a**, **b**, **c**. The coordinates of the atoms in the unit cell are conventionally given as fractions of the lattice vectors, *i.e.* the lattice vectors are used as the basis for a coordinate system.

A molecular crystal consists of discrete molecules, and if these are organic molecules, the crystal is an organic molecular crystal. The atoms in each molecule are held together by relatively strong covalent bonds, while the forces acting between the molecules are much weaker.

An organic molecular crystal that contains only a single chemical species is called a single-component crystal, otherwise a multicomponent crystal. If all species in a multicomponent crystal are each solid in isolation at room temperature and pressure conditions, it is said to be a co-crystal.

If one or more of the constituent compounds is liquid at ambient conditions, we say that the crystal is a solvate, unless if that solvent is water, then the crystal is referred to as a hydrate. Normally, solvates form in crystallisations from solutions of the same solvent, but hydrates can also form by absorbing moisture from the air or contaminations in organic solvents.

Clathrates are crystals that encapsulate, or 'enclathrate', an element or compound that is normally gaseous. The gas molecules are trapped in discrete cage-like structures. So-called inclusion compounds also absorb and hold gaseous or liquid guests, but in channel-type cavities [6].

A multicomponent crystal may contain electrically charged species (ions), in which case we call it a salt crystal or ionic crystal.

Almost any combination of the above special cases is also allowed. One can imagine solvates of salt crystals and clathrates of hydrates *etc.* This thesis only considers organic molecular crystals, the main reason is that the computational methods needed to model inorganic and ionic crystals are quite different.

1.1.2 Symmetry in crystals

The definition of a crystal used here depends entirely on the periodic arrangement of the constituent atoms in space. The atoms are arranged in a three-dimensional lattice. At each lattice point is one or more of the molecular building blocks. Usually, there are several molecules located around the lattice point, and these molecules are in a symmetric arrangement.

This makes it convenient to specify the crystal structure by explicitly listing the positions of only a finite set of atoms that are not related by symmetry (the asymmetric unit), and then specifying a set of symmetry operations to be performed on those atoms to generate the positions of all the other atoms in the crystal.

The molecules oriented around each lattice point are related to each other by a set of symmetry operations that leave the lattice point invariant. The set of symmetry operations form a group, in the mathematical sense of the word. There are 32 point groups that obey the restriction of being commensurate with a periodic lattice.

Combining the 32 crystallographic point groups with the 14 possible 3-dimensional Bravais lattices makes for 230 possible crystallographic space groups (some combinations are redundant). The space groups are tabulated in the International Tables of Crystallography [7], where they are numbered and their symmetry elements listed explicitly. However, the space group number reveals nothing about the symmetry and structure. It is often more convenient (and intuitive) to specify the symmetry of the crystal rather than the space group number.

The Hermann-Mauguin notation is universally used to specify crystal symmetry. In this notation, screw axes are denoted by integers, inversions by an overline bar and mirror planes by the letter *m*. The pointhe internal energyt group symbol is then combined with one of the seven Bravais lattice types, indicated by a capital letter. The Hermann-Mauguin notation depends on the (arbitrary) labelling of the lattice vectors, centring of the unit cell *etc.* Hence, each space group often have more than one possible Hermann-Mauguin symbol. For instance, choosing lattice vector **c** as the unique axis, space group 14 have the symbol $P2_1/c$.

Alternatively, one could simply list the symmetry operations in some other notation. In computer files for instance, symmetry can be represented by text strings; for example '$-x, y+1/2, -z+1/2$' is one of the symmetry operations in space group $P2_1/c$. Both space group numbers and Hermann-Mauguin-symbols will be used interchangeably to specify space groups.

Mathematically, or for computer arithmetic in particular, it is convenient to calculate atomic coordinates in crystals using a matrix representation of symmetry operations. Generally, a 3-dimensional space group can be represented as a set of affine mappings (U, u). Each affine mapping consists of a 3×3 rotation matrix U and a 3-vector u describing translation, in this case:

$$U = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \quad u = \begin{pmatrix} 0 \\ 1/2 \\ 1/2 \end{pmatrix} \tag{1.1}$$

Coordinates generated by symmetry can then be calculated by multiplying the coordinates of the asymmetric unit by the matrix and adding the translation vector. An alternative is to combine the rotation matrix and the translation vector to a single 4×4 matrix. Further introduction to crystallographic terminology, symmetry and the associated algebra can be found in text books [8, 9] and compendiums [10].

There are a few more concepts and ideas that the reader should be aware of. The first is that in experimental crystal structure determinations, the space group symmetry is 'maximized'. There is a general assumption that the 'true' space group of a crystal structure is the one with the most symmetry operations that can be used to describe it. Conversely, one strives to have as few atoms as possible in the asymmetric unit, although it is possible to represent the same physical structure in a less symmetric space group.

Another thing to be aware of is that the symmetry observed in experimental work is the symmetry of the electron density, averaged over time and vibrational motion. If the constituent molecules are themselves symmetric, this symmetry will also be included in the space group symmetry and the asymmetric unit will contain only a fraction of the real molecule. Hence, the perfect mathematical, maximal symmetry seen in experimental work is often not the same as that used in computations. It is tempting to argue that the symmetry seen in diffraction experiments is not necessarily the true symmetry of the physical crystal structure, a view held by Niggli [11].

This leads us to the concept of pseudo-symmetry, which is a spatial relationship between atoms that is nearly symmetric. The symmetry is slightly broken and some of the atoms in the structure appear to be displaced from the ideal symmetry-equivalent positions. Why some crystals are pseudo-symmetric is subject to discussion [12].

Some nomenclature for the contents of a crystal is needed. In each unit cell there will be one or more formula units of the constituent species. This integer is denoted Z in this thesis, and for single-component molecular crystals this is also the number of molecules per unit cell. By Z' is meant the number of formula units in the asymmetric unit. Note that since the principle of maximum symmetry applies, symmetric molecules frequently crystallise with $Z' < 1$. This nomenclature is commonly used in crystallography, but there is sometimes a need to define additional terms to denote the number of crystallographically or chemically distinct entities [13].

It is very important to note that one should not assume trivial relationships between these quantities. In particular, it is deceptively easy to think that the number of formula units in the unit cell Z is equal to Z' times the number of symmetry operations in the space group, but that is not necessarily so. One symmetry operation can map a set of atoms to positions that have already been populated by another. One notable example is the crystal structure ACOGOO [14], a co-crystal between 1,8-naphthalenedicarboxylic acid and tetrakis(4-pyridyl)cyclobutane. The experimental structure was determined

in space group $P\bar{1}$ with $Z = 1$. The space group has only the identity and inversion symmetry operations, implying $Z' = 1/2$, and this is the structure deposited in the CSD. However, applying the inversion operation on the naphtalenedicarboxyl molecule leads to a mapping of the molecule onto itself! No second molecule is generated by the space group symmetry. In fact, the stoichiometry is 1:1 with $Z' = 1.5$.

1.1.3 Polymorphs

1.1.3.1 The basics of polymorphism

The term polymorphism was defined by McCrone as "*a solid crystalline phase of a given compound resulting from the possibility of at least two different arrangements of the molecules of that compound in the solid state*" [15].

This definition almost captures what is now meant by the term. The modification that needs to be made to the above definition is a clarification that if the molecules in different polymorphs have different conformations, we would refer to them as *conformational polymorphs* [16]. Polymorphs containing the same conformer would be called *packing polymorphs*. There are additional caveats regarding inter-converting isomers and tautomers and the nomenclature of polymorphs of multicomponent crystals and solvates [17, 18], but I will not go into such details.

McCrone also made the now famous remark that "*it is at least this author's opinion that every compound has different polymorphic forms and that, in general, the number of forms known is proportional to the time and money that has been spent in research on that compound*" [15]. Current estimates indeed show that polymorphism is common [19], but there are also counterexamples; sucrose and naphthalene for instance have never been observed to have alternative crystal structures [20].

Of the possible crystal forms, there must be one crystal packing that is the most thermodynamically favoured at any given temperature and pressure [21]. However, this does not exclude the possibility of additional crystal packings that may be meta-stable.

Exactly how commonly polymorphism occurs is subject to debate [22, 23, 19], but it is clear that many important compounds exhibit polymorphism. Since different polymorphs have different physical properties, and the properties of for instance pharmaceuticals and explosives need to be controlled with great precision, studies of polymorphism have become a large scientific endeavour.

The relative stability of polymorphs depends on the difference in Gibbs free energy ΔG between them, the more stable being the one with a lower (more negative) energy.

$$\Delta G = \Delta U + P\Delta V - T\Delta S, \tag{1.2}$$

where U is .

$$U = E_{\text{latt}} + ZPE + \int_0^T C_{\text{p}}\, dT \tag{1.3}$$

At 0 K the internal energy is equal to the lattice energy E_{latt} plus the zero point energy ZPE. At ambient pressure, the $P\Delta V$-term is negligible but can be quite important at elevated pressures [24].

A common simplification is to assume that the zero point energy does not differ significantly between polymorphs. If this holds, the lattice energy alone can be used to establish the relative stability of alternative packings at 0 K. At elevated temperatures the difference in entropy may become significant, but it is common to also neglect this and to estimate polymorph relative stabilities by temperature-independent lattice energies alone [25, 5]. It is a central objective of this thesis to provide hard data on whether these approximations are justified and what errors can be expected when relying on them.

The relationship between two or more polymorphs can be visualized by plotting their respective internal and free energies as functions of temperature [20], see Figures 1.1 and 1.2. Note the relationships $C_{\text{p}} = (\partial U/\partial T)_P$ and $S = -\partial G/\partial T$.

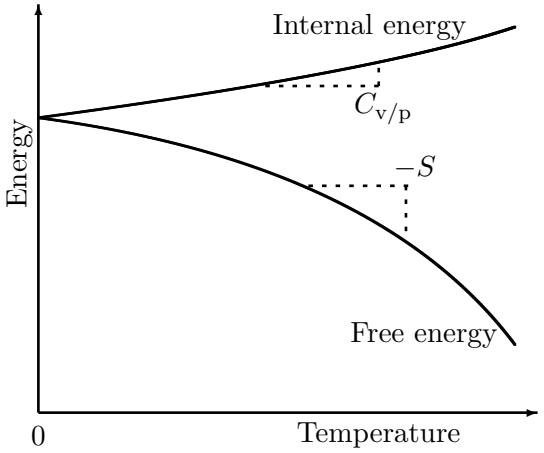


Figure 1.1: The internal energy increases slightly due to thermal effects, and the free energy decreases. The derivative of the internal energy is either C_{v} or C_{p}, depending on whether the volume of pressure is held constant. Likewise, the free energy is either the Helmholtz or the Gibbs energy.

Figure 1.2 shows the behaviour of two *enantiotropic* polymorphs. At a specific temperature, the free energy of polymorph II drops below that of I, possibly facilitating a spontaneous solid state transition between the two. If the second polymorph instead is always higher in energy than the first, the polymorphs are said to be *monotropic*. It is conceivable that a polymorph pair that is monotropic at one pressure can be enantiotropic at another.

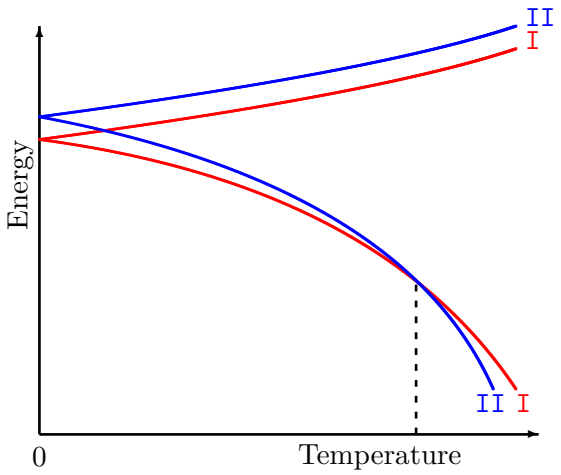


Figure 1.2: The relationship between two enantiotropic polymorphs I and II. At a specific transition temperature (dashed line) the free energy curves cross and polymorph II becomes thermodynamically more stable than form I.

The internal energy curves are unlikely to cross, as differences in heat capacity between polymorphs are expected to be small, see Chapter 5.

This introduction will only mention a few examples of polymorphism occurring in pharmaceuticals, explosives, pigments *etc.* For a more comprehensive review, the textbook by Bernstein is strongly recommended [20].

1.1.3.2 Polymorphism in pharmaceuticals

Pharmaceuticals are organic molecules with typically 50-300 atoms. Polymorphism in pharmaceutically active ingredients is common [19], and often seen as a problem in the formulation and production of drugs. Since different polymorphs of the same compound can have different solubility and consequently different bioavailability, it becomes crucial to control what polymorph forms in the manufacturing process. It is usually desirable to have the thermodynamically stable form in a drug formulation, so as to prevent transitions to more stable forms during storage and to increase the shelf life.

The solubility of a compound is directly related to the free energy of the crystal structure. To prevent ageing of a pharmaceutical formulation, it is desirable to have a stable crystal form, but this is also the least soluble. The limited solubility of pharmaceutically active compounds is frequently a problem in the formulation of drugs [26]. On the other hand, a metastable crystal form might have other properties making it favourable. Better elastic properties facilitating is an example [27].

Below, I briefly review a few well-known examples of polymorphism in pharmaceuticals, as an illustration of why control and prediction of polymorphs is important.

Ritonavir Ritonavir, an HIV protease inhibitor, was discovered by Abbot laboratories in 1992. After FDA-approval in 1996, the drug was released on the market under the trade name Norvir [28]. In 1998 the semi-solid drug formulation started to fail routine dissolution tests, the drug had become less soluble and precipitated out of the formulation [29]. The active pharmaceutical ingredient, ritonavir, had spontaneously crystallized into a previously unknown crystal structure, with respect to which the formulation was supersaturated.

The drug had to be withdrawn from the market at great cost. The new polymorph was named form II, and the older polymorph form I. The new stable polymorph acted as a nucleation site and catalysed the transition of all ritonavir into the new form [30]. The metastable form I could no longer be obtained once the production plant had been contaminated with trace amounts of form II, a phenomenon known as disappearing polymorphism [31, 32].

The accidental discovery of polymorphism in ritonavir, and the large and damaging economic effects it had, spurred further research into polymorphism and crystal structure prediction, as is evident from the many references to the story in CSP-related literature.

Thyroxine Hypothyroidism is typically treated with synthetic thyroxine. In 2004 several medical associations publicly issued their concerns that thyroxine produced by different manufacturers had different pharmacological efficacy [33]. The difference in solubility and bioavailability has been attributed to polymorphism [34].

Rotigotine The active pharmaceutical ingredient rotigotine suffered the same problem as ritonavir; the unexpected appearance of a more stable polymorph [35]. The drug was dispersed in a silicone adhesive in transdermal patches. A more thermodynamically stable polymorph then formed spontaneously during storage. Data published in patents [36, 37] allowed the determination of the relative thermodynamic stability of the two forms [38].

Cefdinir Another important aspect of polymorphism in pharmaceuticals is that patents cover a specific polymorph. A competing drug company may discover a new polymorph and patent the medicinal use of it. Cefdinir is a telling example. The cephalosporin antibiotic cefdinir was discovered by Fujisawa Pharmaceutical, which filed two patents for the drug in 1982 and 1987. Several new polymorphs and hydrates have since been discovered and patented by eight different companies [39]. Defending and challenging patent claims in a court of law with experimental evidence can be problematic [40].

Carbamazepine Carbamazepine is a mood-stabilizing drug used in the treatment of bipolar disorder and schizophrenia. Carbamazepine is known to form many solid forms, including at least five anhydrous crystal structures [41, 42]. Crystal structure prediction of carbamazepine has been performed by Cruz-Cabeza [43]. Some of the predicted

crystal structures have subsequently been confirmed to exist experimentally, for instance an acetic acid solvate [44].

Form II of carbamazepine is porous and is of particular interest for the thesis at hand. The $R\overline{3}$-symmetry of this polymorph results in an inefficient crystal packing [45]. A computational study revealed that this polymorph is found as a high-energy structure during computational crystal structure predictions. Computationally inserting solvent molecules into the voids allowed solvates to be predicted [46]. This rare computational prediction of inclusion behaviour of the porous crystal structure of polymorph II has been confirmed experimentally [47].

Theophylline The asthma-drug theophylline is a derivative of caffeine and has several known polymorphs [48, 49, 50, 51]. The different polymorphs have different hydrogen bonding motifs [52], facilitated by several hydrogen bond donor and acceptors in the molecule. The stable form at room temperature is designated form IV and has $Z' = 2$. Form II was incorrectly believed to be stable until form IV was discovered. Form II spontaneously transitions into the high-temperature form I at temperatures above 541 K [53], implying an enantiotropic relationship between them.

Theophylline has been the subject of crystal structure prediction calculations. 73 hypothetical crystal structures with one molecule in the asymmetric unit ($Z' = 1$) were found within 15 kJ/mol from the global minimum. Form II was identified as the second structure from the global minimum [50].

Two new polymorphs were discovered by the combined use of CSP and transmission electron microscopy diffraction. The new polymorph VI was ranked as number 8, some 4 kJ/mol from the global minimum in the CSP results [50]. A second novel polymorph VII was ranked as number 14 at 6 kJ/mol above the minimum [51]. The combination of CSP and electron diffraction allowed the determination of the crystal structures from only nanogram quantities of the compound.

1.1.3.3 Polymorphism in other molecular materials

Polymorphism also affects many other classes of materials, such as explosives, food-related compounds, pigments and organic semiconductors. Here, a few examples of such industrially important products are presented.

The properties of high explosives must be precisely controlled. Sensitivity to humidity, shock, temperature and ageing are all important and affect the safety of handling and using explosives. Polymorphism in high explosives is therefore of interest [54, 55, 20].

Cyclotetramethylene tetranitramine (HMX) has four known polymorphs that have been extensively studied by many different methods. The crystal structures of the different forms have been the subject of electronic structure calculations in the local density

approximation [56]. Also molecular dynamics calculations with force fields have been performed to study the properties of several high explosives [57, 58].

Polymorphism also occurs in some edible substances. Chocolate is a widely known example. Cocoa butter, despite not being a pure substance, has six known polymorphs with different melting points [59, 60]. The desired form is polymorph V, which has a melting point of 33.8°C, making it melt in the mouth. Unfortunately, the thermodynamically stable form VI has a melting point of 36.3°C, high enough to make it far less pleasant to eat. The spontaneous transition of form V to form VI greatly limits the shelf life of chocolate and causes large economic losses. Melting and recrystallizing chocolate while gently agitating it makes the fats crystallise in the metastable form again. This 'tempering' of the chocolate is routinely done in professional confectioneries.

The commonly used sweetener aspartame has several known polymorphs. The original manufacturing process resulted in fine needle-like crystals, which caused problems in filtration and down-stream processing [61, 20]. Several patents were issued for various improvements to the crystallization method. One such patent, by Ajinomoto, was challenged by the Opposition Division of the European Patent Office, which claimed that Ajinomoto's crystals were not a new polymorph but just a different shape/habit of the same crystal structure. Ajinomoto responded by listing several important differences between the original fine needles and his new habit. Ajinomoto won the case and was issued a new patent specification.

Polymorphism is common also in pigments and dyes [20, 62]. That different polymorphs can have different colours is beautifully demonstrated by the compound ROY, so named because of its many red, orange and yellow polymorphs [63, 64].

The pigment and organic semiconductor quinacridone has at least four polymorphs, denoted α, β, γ and α' [20, 65, 66]. The α' form was discovered in 2007 by an elegant use of crystal structure prediction in combination with experimental powder X-ray diffraction [66].

There are many pigments based on perylene, several of which are polymorphic. Pigment Red 149 has three known polymorphs, while Pigment red 179 has only one, despite being intensely studied and used industrially [20]. Another apparently monomorphic compound that has been studied and used extensively is chloridazon, which we will take a closer look at in Chapter 6.

1.1.4 Clathrates

1.1.4.1 The close-packing principle

As a thought-experiment, assume that there are no directional interactions between the molecules in a crystal, but that the only interactions are a short-distance hard repulsion and a long-range softer attraction. Then it is easily concluded that the energetically most favourable structure is one in which each atom interacts with as many other atoms as possible. The consequence is that the energetic minimum is a close-packed structure, *i.e.* the molecules are as compactly arranged as possible, leaving no empty voids between them [67, 21]. The packing coefficient of closely packed equally sized spheres is $\pi/\sqrt{18} \approx 74\%$ [68], and practically all molecular solids have a similar packing efficiency.

It is important to note that the close-packing principle does not exclude the possibility of porous crystal structures. It is conceivable that some of the higher-energy crystal packings are porous. These will almost always be thermodynamically unstable, but perhaps they can be kinetically stable and stabilized by entropic contributions. Remember also that the close-packing principle is based on the assumption of only non-directional interactions.

The known porous molecular crystals can be split into two groups depending on whether the cavities are located between the molecules (extrinsically porous), or if the molecules themselves contain empty space (intrinsically porous) [69, 70, 71]. This thesis deals only with extrinsically porous clathrates of small molecules. Two prominent examples are given below.

1.1.4.2 Dianin's compound

Porous crystal structures exist, but they are quite rare. One of the best known is Dianin's compound [69, 72]. This is an almost unique molecule in that its thermodynamically stable structure has cavities large enough to host small molecules. The reason for this seems to be that the awkward shape of the molecule prevents close-packing. Dianin's compound is able to absorb a large number of different guest molecules to form clathrates and solvates.

1.1.4.3 Hydroquinone

Hydroquinone (quinol, 1,4-dihydroxybenzene) is probably the most studied and best known clathrate-forming compound. Its metastable polymorph β easily forms stable clathrates with an unusually large number of gases. The clathrates are isostructural with a characteristic $R\bar{3}$ symmetry, which results in porosity, as is also seen in other phenolic clathrates [73, 74, 4].

Remarkably, also the thermodynamically stable form, named α, is porous and has $R\bar{3}$-symmetry. In this case, it is perhaps the strong hydrogen bonds that allows it to violate the close-packing principle.

Hydroquinone has been the subject of intense experimental and computational studies. Computational crystal structure prediction of hydroquinone was performed by Cruz-Cabeza [75], who also considered the possibility of guest inclusion. Lattice dynamics calculations of hydroquinone clathrates were performed by Belosludov [76]. Carbon capture by means of hydroquinone enclathration of CO_2 from flue gases was recently suggested [77] and the crystal structure of that clathrate was recently determined [78].

The hydroquinone hydrogen clathrate has an interesting history. A computational study predicted that such clathrates can form and that they may have multiple H_2 occupancy [79], naturally of great interest for gas storage applications. Experimental work has recently confirmed that the predictions were correct, the clathrate has double and triple occupancy of H_2 at elevated pressures [80, 81].

Solid-state ^{129}Xe NMR spectroscopy has been used to study the hydroquinone xenon clathrate [82]. Xenon chemical shift has been suggested as a very sensitive probe for the study of clathrates [83, 84], and this will be explored in Chapter 7 of this thesis.

1.2 Computational techniques

1.2.1 Energy models

1.2.1.1 The need of energy models

A fundamental prerequisite for modelling chemical systems is that the energy of the systems can be calculated. The total energy of a system with several quantum-mechanically interacting particles cannot be calculated exactly, but different approximations must be done. In short, we will look at classical methods using analytical atom-atom pairwise potentials, and we will look at density functional theory (DFT) which models interactions between electrons in the system. For a comprehensive treatment of the theory of intermolecular interactions, see the excellent book by Stone [85].

A very brief introduction to these methods is given below. A slightly more detailed description is given in Chapter 2. In this thesis, a hybrid method combining classical potentials and DFT will be used. The individual molecules in a crystal structure will be modelled with a density functional method, while interactions between molecules are calculated with analytical potentials. Electrostatic intermolecular interactions are calculated from the molecule's electron density, obtained from the DFT calculation.

1.2.1.2 Force fields

The potential energy between molecules $k, l, m \ldots$ can be written

$$E = \sum_{k<l} E_{kl} + \sum_{k<l<m} E_{klm} + \sum_{k\ldots n} E_{k\ldots n} + \ldots \tag{1.4}$$

This expression rapidly becomes prohibitively complicated as the number of interacting molecules grow, and it is usually necessary to truncate the right hand side and only consider pairwise interactions, and occasionally also three-body terms. Hence, we will assume that the interaction energy can be expressed with an *effective* potential that is pairwise additive [86, 85], that is a force field. This is arguably a crude simplification but it works reasonably well for the repulsion, electrostatic and dispersion interactions. Other effects, like the mutual polarization of molecules in a crystal structure or charge transfer, are poorly represented however.

The interactions between molecules are modelled with analytical model potentials between the individual atoms. The pairwise atom-atom potential method has been reviewed in detail by Pertsin & Kitaigorodskii [87].

The most common pairwise additive atom-atom potential is the Lennard-Jones potential

$$V_{\iota\kappa}(r) = 4\epsilon_{\iota\kappa}\left(\frac{\sigma_{\iota\kappa}^{12}}{r^{12}} - \frac{\sigma_{\iota\kappa}^{6}}{r^{6}}\right). \tag{1.5}$$

Here $\epsilon_{\iota\kappa}$ is the potential well depth, and $\sigma_{\iota\kappa}$ is the distance were the potential is zero for two atoms of types ι and κ. These parameters are determined for each type of atom. Parametrization is performed by fitting to experimental or *ab initio* calculations. Usually, electrostatic interactions are modelled separately with a Coulomb-potential and atomic partial charges, which are either parametrised or calculated.

The Lennard-Jones potential has been generalised to the Mie potential, where other values of the exponents are used. Common exponents are 14/7 and 9/6, the latter being used in the COMPASS force field.

By the term force field we mean a self-consistent set of potential functions and parameters. Force fields with the Mie functional form are commonly used for molecular dynamics and Monte Carlo simulations, mainly because they are computationally very fast to evaluate. In this thesis the COMPASS [88] and OPLS2005 [89] force fields are used in Monte Carlo simulations.

The attractive r^{-6} term in Mie and Lennard-Jones potentials is physically justified by the distance-dependence of the leading term in the London dispersion [90, 85]. The exponent in the repulsive term (9, 12, 14) has little physical justification and is chosen to be computationally convenient, *i.e.* a simple multiple of the exponent for the attractive term.

The repulsion between atoms at short distances is nearly proportional to the electron density overlap, leading to an exponential repulsion term. For two atoms of types ι and κ at a distance r we have

$$V_{\iota\kappa}(r) = A_{\iota\kappa} \exp\left(-B_{\iota\kappa} r\right) - C_{\iota\kappa} r^{-6} \tag{1.6}$$

where A, B, C are empirical parameters that are different for each atom type pair. This so-called Buckingham potential is used in the Williams force fields [91, 92, 93], which are used in this thesis. More accurate model potentials for intermolecular interactions exist and are described in the Theory chapter.

Intramolecular bonds and geometries can also be modelled with force fields. Covalent bonds are often modelled with a Hookean potential. Similarly, bond angles and dihedrals are modelled with simple analytical functions with empirically fitted parameters.

It must be noted that the energy of covalent bonds are about two or three orders of magnitude larger than the weak interactions between molecules in a crystal. Hence, if the energy of a crystal structure is modelled as a sum of intra- and intermolecular energies, the intramolecular terms (and their errors) will completely dominate, causing large errors in the calculated lattice energies of crystals. This is the reason why classical force fields are unsuitable for molecular crystals.

In this thesis, a hybrid energy model will be used to circumvent this problem. The intramolecular energy and geometry will not be modelled with force fields but by density functional theory.

1.2.1.3 Density functional theory

An alternative to force fields is to explicitly model the electronic structure by quantum-mechanical methods. Kohn-Sham density functional theory (DFT) is currently the most popular method for such calculations on individual molecules and the organic solid state.

Density functional theory is based on the Hohenberg-Kohn theorems [94]. These in turn rest on the Born-Oppenheimer approximation. In addition we will constrain ourselves to consider only ground state electronic structures. The theorems provide a theoretical way to circumvent explicitly solving the Schrödinger equation. Everything that can be known about the system, its energy and properties, can be expressed as functionals of the electron density.

Only a very brief introduction is given here and in Chapter 2. For a full treatment, the reader should consult the review article by Becke [95] and the textbook by Koch [96].

The basic idea of Kohn-Sham density functional approximations is to replace the system of interacting electrons with a system of non-interacting electrons having the same density and energy.

Exchange-correlation functionals based on the local density and gradient fail to account for long-range electron correlation which causes attractive dispersion interactions between atoms. Hence, density functional theory in the GGA approximation, including hybrid functionals, underestimate the dispersion. This error is quite significant in condensed phases and cannot be ignored. Several methods have been proposed to correct for this. The earliest were empirically parametrized force fields to be added to the density functional energy [97, 98]. Recently, sophisticated algorithms for calculating the missing dispersion have been published [99, 100].

1.2.2 Molecular dynamics

The methods mentioned so far has been for calculating static, formally 0 K properties of chemical systems. Thermal motion and other effect of temperature must also be modelled to obtain the *free* energy.

It is possible to simulate the time-evolution of a molecular system at finite temperatures by integrating over Newton's equations of motion in short time-steps. Such molecular dynamics (MD) simulations are commonly used to study biological macromolecules, such as the motion and activity of enzymes, and the binding of ligands.

Because of the need for long simulation times to adequately sample the phase space that is available to the system, it is essential that the energy of the system can be computed rapidly. Because of this, classical force fields are almost always used.

Molecular dynamics can also be applied to molecular crystals, and allows thermal vibrations to be computed. This is an attractive method since the vibrational frequency spectrum can be used to calculate the vibrational free energy of the crystal structure [101].

An important advantage of molecular dynamics is that the unit cell can be made flexible, allowing compression under pressure and thermal expansion to be observed directly. The thermal motion is also not restrained to the lattice, so that strongly anharmonic vibrations or even structural changes can be modelled. In principle, solid-to-solid phase transitions can be observed during simulations.

However, there are also important drawbacks of molecular dynamics. The need to calculate the energy at every time step makes it necessary to use force fields with simple model potentials. Classical force fields are known not to be accurate enough for predicting the stability and properties of molecular crystals [25, 5].

A further complication is that long simulation times are needed to observe rare events, such as a complete transition to a different polymorph. For vibrational calculations and to be able to see large-scale cooperative motion, it is also necessary to use simulation boxes that include many unit cells, as vibrations with a wavelength longer than the size of the box cannot be observed [102].

Molecular dynamics simulations with the aim to observe solid-to-solid phase transitions in the amino acid norleucine have recently been performed by van den Ende [103, 104].

1.2.3 Monte Carlo methods

The temperature-dependent properties of a chemical system can also be calculated by Monte Carlo simulations. In Monte Carlo simulations, the positions and orientations of atoms or molecules are randomly perturbed. Hence, unlike MD-simulations, the trajectory of the simulation is not continuous in time, but aims to be a more representative sample of phase space, the thermodynamic ensemble of states available to the system at the given temperature. This is often difficult to achieve in MD-simulations. To achieve this in Monte Carlo simulations, each new simulation frame is accepted only with a certain probability that depends on the change in energy of the system, and the temperature. This so-called Metropolis sampling method yields a Boltzmann-distributed ensemble of states [105].

The Boltzmann-distributed ensemble of states provides the free energy of the system and other properties through standard statistical mechanicals ensemble averages. The system can also change in chemical composition during a simulation, the random moves may include the addition or deletion of atoms or molecules. Such 'grand-canonical' simulations provide a way to simulate for instance the adsorption of gases or water in porous materials.

Monte Carlo simulations can also be used for sampling possible crystal structures. Monte Carlo simulated annealing and parallel tempering methods have been used to generate crystal structures for CSP purposes. In such cases, a classical force field is used to calculate the energy of the system, while the temperature is artificially allowed to first increase to a very high temperature, and then the system is cooled, to allow the molecules to condense to a crystalline structure. Parallel tempering improves the sampling of phase space by running several simulations in parallel and occasionally swapping states between simulations.

Monte Carlo simulated annealing was one of the first methods employed for crystal structure prediction and such methods has been successful in the sampling of possible crystal structures [106, 107, 108]. In Chapter 7, simulated annealing is used to generate crystal structures of clathrates.

1.2.4 Lattice dynamics

An alternative method for the study of thermal vibrations in solids is lattice dynamics (LD). In LD, harmonic force constants for infinitesimal translations (and rotations) of atoms (and molecules) are constructed from the second derivative of the potential energy of the system. The pairwise force constants and the masses of the constituent atoms or molecules are used to construct a mass-weighted Hessian matrix. This so called *dynamical matrix* is then diagonalized. The resulting eigenvectors are called the normal modes of the system and the eigenvalues are the square of the phonon frequencies. For a system corresponding to a local energy-minimum, the eigenvalues will be positive and the frequencies real [109, 110, 111].

For a macroscopic crystal the phonon modes will be populated according to the Boltzmann distribution, allowing a partition function to be calculated, and from this the thermal vibrational free energy can be calculated [112]. Explicit expressions for the thermodynamic functions are given below.

Lattice dynamics has a couple of caveats that can make it challenging to use. One is that the method rests on an assumption that the system is an energy minimum. It is imperative to use crystal structures that have first been energy-minimised with the same energy model used for the phonon calculation. Another is that there is actually one dynamical matrix for every point in the reciprocal lattice. In order to calculate frequencies for phonons propagating in different directions, the dynamical matrix must be diagonalised for different wave vectors $\mathbf{k}$ [113, 114].

A third issue is that in calculating the force constants, it is assumed that the potential energy is harmonic near the equilibrium. This is not strictly true, anharmonic vibrations cause a substantial thermal expansion of the crystal structure. A common approximate solution is to still calculate harmonic phonons, but account for the thermal expansion in some other way, leading to the so-called quasi-harmonic approximation (QHA) [115, 110].

Lattice dynamics is used extensively in this thesis and the difficulties mentioned here will be addressed in detail in Chapters 2 and 3.

1.3 Crystal Structure Prediction

1.3.1 A continuing scandal

In 1988, the editor-in-chief of Nature, John Maddox, wrote a deliberately provoking editorial [116]. He began by stating that: "*One of the continuing scandals in the physical sciences is that it remains in general impossible to predict the structure of even the simplest crystalline solids from a knowledge of their chemical composition.*"

He further writes that "*one would have thought that, by now, it should be possible to equip a sufficiently large computer by a sufficiently large program, type in the formula of the chemical and obtain, as output, the coordinates of the atoms in a unit cell.*"

The latter quote can in fact serve as the definition of (computational[1]) crystal structure prediction.

1.3.2 History of crystal structure prediction

Here a brief review of the development of crystal structure prediction up to approximately year 2000 is presented. Predictions were at that time primarily based on classical methods and applied to small rigid molecules. Work after 2000 increasingly often include intramolecular flexibility, vibrational energy contributions and quantum-chemical methods.

The first serious reflection on the problem of close-packing and its predictability was probably by Hilbert. He included the problem of proving what the densest packing is of equal spheres and other geometric shapes in his list of 23 unsolved fundamental problems in mathematics [118].

Williams pioneered the use of computer calculations of crystal structures in the 1960s. He realized that geometry-optimisation by lattice-energy minimisation can yield crystal structures that can be used as starting points for indexing and refinement of experimental X-ray diffraction results [119, 120, 121].

Kitaigorodskii attempted to predict crystal structures in the 1960s by non-computational methods [67, 21]. He noticed that molecular crystals pack *"in such a way that the 'projections' formed by the atoms of one molecule fill the 'hollows' between the atoms of adjacent molecules, the molecules dovetailing in a pattern of close packing"*. Using such empirical principles, he predicted crystal structures by tangibly packing space-filling models of molecules as densely as possible. Kitaigorodskii also performed force field lattice energy calculations and foresaw the usefulness of computational crystal structure prediction [122, 21].

The first attempt at what we would now call crystal structure prediction was by Dzyabchenko [123]. In 1984 he searched for lattice energy minima of randomly generated benzene crystals in four space groups. A force field was used to calculate the lattice energy. The global energy minimum corresponded to the known stable crystal structure and also the high-pressure form of benzene was reproduced.

Gavezzotti and Desiraju studied the packing of aromatic hydrocarbons in 1989. The crystal structures were classified and rationalised in terms of geometric descriptors and

[1]Non-computational methods for CSP have been attempted. See reference [21] and [117].

energy calculations. Two years later Gavezzotti attempted the prediction of halobenzene and polythienol crystals using his own force field [124, 125]. This was followed by a discussion about the difficulty of crystal structure prediction in 1994 [126] and an important study of polymorph pairwise differences in which also thermal contributions were considered [23].

Monte Carlo simulated annealing (see below) was used early for structure predictions. In 1992 crystal structures of hexamethylbenzene and ethylene were successfully found by this method [127]. A year later, the first true predictions of previously unknown crystal structures were demonstrated using the same method [106].

Other early works include a prediction of monosaccharides by van Eijck [128] and a computer program by Hofmann that implemented a grid search over a discretized search space [129]. Hofmann advocated scoring functions derived from empirical data, such as the commonality of packing motifs in databases [130, 131, 132, 133].

One can argue that the term crystal structure prediction should preferably be applied only to blind predictions. When experimental data is available it is all too easy, and tempting, to choose or adjust the computational methods until agreement with the experiments is reached. For this reason, the CSP community has jointly conducted a series of blind tests of predictions. The first such blind test was initialised in 1999 at a workshop organised by the CCDC [107].

The six blind tests conducted so far provide a wide and deep description of the different computational methods used and their performance [107, 130, 134, 135, 25, 5]. Each blind test contains 3–5 target systems for which crystal structures are to be predicted. The targets may be single molecules, a pair of molecules that form a co-crystal, ions that form a salt *etc.* In total, 26 target systems have been used for such blind tests. The resulting articles have chronicled the successive development and growth of the CSP community and gives a reasonably fair account of successes and failures that would otherwise probably not have been published.

Chapter 4 describes the prediction of crystal structures for two of the target systems in the sixth blind test, held in 2015. Both 0 K lattice energies and room-temperature free energies were used in these predictions.

1.3.3 Crystal structure prediction methods

1.3.3.1 Five steps to predict crystal structures

The blind tests allow us to summarize the most common and successful methods for CSP. All methods consist of essentially the same five steps. First, three-dimensional conformations of the individual molecule(s) are generated. These are then used in the

generation of a large number of trial structures in the second step, although in some algorithms the generation of molecular conformation and crystal structure are done simultaneously [136]. The trial structures are geometry-optimised, usually by minimising the structures' energy in a third step. Many trial crystals will energy-minimise to the same structure, but possibly in different space group representations. In the fourth step, clusters of (nearly) identical structures are identified by calculating some kind of similarity score, and duplicate structures removed. Finally, in the fifth and last step, the predicted structures are ranked by a final scoring function that, at least in principle, estimates the probability of the structure being observed in experiments.

Each step is briefly reviewed below. Several methods have been used for each step and only the most common or successful are mentioned here. For a more comprehensive treatment, see the blind test papers [107, 130, 134, 135, 25, 5] and related works by Day, Price, Neumann and others [137, 138, 139, 140, 141, 142, 108].

1.3.3.2 Molecular conformers

The first part of a crystal structure prediction is to generate 3-dimensional molecular conformations of the molecule (or molecules for multicomponent systems) that are generally specified by a name, a 2D drawing, a SMILES string [143] or similar descriptors. To just generate *a* conformation can easily be achieved with common software like Open Babel or Tinker [144, 145]. However, for flexible molecules it is important to sample conformational space intelligently [146].

The packing forces in a crystal structure can distort the molecule away from a stable conformer [16, 43]. This makes it crucially important to consider, and predict, which conformations are the most likely to occur in the solid state. In rare cases, the conformer in the crystal structure is far from an energy-minimum of the isolated molecule [147]. Failing to predict which conformer occurs in the solid state is not an uncommon reason for failed predictions [5]. The PhD-thesis by Thompson describes strategies for sampling the molecular conformational space in detail [148].

Methods to sample conformational space include grid-based searches where the geometry, perhaps represented as a Z-matrix, is perturbed by small changes to the angles and dihedrals.

For large flexible molecules, the number of conformations can be very large. Conformations can then be chosen based on the the occurrence of molecular fragment geometries in databases [149, 150].

Monte Carlo simulations can be used to sample a large number of conformations. Energy-minimisation of these leads to a number of conformers that are normally used in CSP. The computational cost of MC (or MD) however require simple force fields,

which may not be accurate enough so that some shallow energy minima conformers can be missed. Enhanced sampling of conformational space in MC can be obtained by performing the simulation by perturbing normal modes rather than Z-matrix coordinates [151, 152]. See Chapter 4 for an application of this method.

In many cases, molecular symmetries can greatly affect how conformers and crystal space groups need to be sampled. Chiral molecules necessarily crystallise in one of the Sohnke space groups, of which there are only 65. In many cases, one molecular conformer is achiral, while others are not.

Pyramidalisation of amino groups can cause particular problems [43]. Can $-NH_2$-groups be treated as planar, saving considerable efforts, or is it necessary to generate crystals of both pyramidal geometries? In experimental crystal structures, amino groups are often found to be planar, but this is probably often an erroneous artefact. The protons are in reality disordered and tunnel between the two pyramidal energy-minima, but this cannot be resolved by X-ray crystallography. This issue affected the CSP performed in Chapter 6, where the amino group was planar in a force field model, pyramidal when treated with DFT and planar in the experimental structure.

In acids, salts and zwitterions, the protonation state of the molecule must be carefully considered. Is the position of acidic protons in the crystal known or should it be predicted by considering every possibility? In these cases, a final geometry-optimisation and scoring based on DFT may be preferable [153, 5].

1.3.3.3 Structure generation

The second step in crystal structure prediction is to construct crystal structures of the molecular conformations from the previous step. This can be done a several ways, only a couple of the most common methods are mentioned here.

A naive method is to simply randomly orient a variable number of molecules in space and then add random lattice vectors, followed by energy-minimisation. That is, no space group symmetry is assumed *a priori*, but $P1$-structures are generated and allowed to relax and only then is the resulting symmetry determined. This has been done [13], but has proven inefficient since the $P1$ search space is so vast. Virtually all methods used today sample a predetermined set of space groups.

One commonly used method is Monte Carlo simulated annealing. It was one of the first methods employed for CSP and has since been used in several successful predictions [127, 154, 106, 155]. Crystal structures of the molecule(s) are subjected to a Metropolis Monte Carlo simulation. Initially, the temperature is set much higher than the boiling point, thereby all Monte Carlo moves are accepted, ensuring ergodicity of the sampling. The temperature is then slowly decreased and the system condenses to an energy-minimum.

The lattice energy is calculated with some force field, like CHARMm or COMPASS [154, 127]. A method of this kind in used in Chapter 7.

Another approach, advocated by Desiraju *et al.*, is to start with a single molecule, and then successively add molecules in energetically favourable, or otherwise commonly observed, positions, thus building up plausible packing motifs [156, 157].

More recently, methods based on a random sampling of the crystalline degrees of freedom have become more popular [25, 5]. The molecules are packed into crystal structures, where the position and orientation of the molecule(s) in the asymmetric unit is determined by a random number generator. Similarly, the lattice parameters are generated randomly. Often low-discrepancy quasi-random numbers are used, since these allow a faster and more uniform sampling of the search space [158, 159, 160]. A pseudo-random structure-generator is used in Chapters 4 and 6.

Both the quasi-random search method and Monte Carlo can also be used to sample intramolecular degrees of freedom.

1.3.3.4 Energy minimisation

The randomly generated structures can be relatively far from an energy minimum. The structures must therefore be geometry-optimised and this is often done my minimising the energy. Successive geometry optimisations can be performed by using increasingly more accurate force fields [2], by allowing intramolecular flexibility [161], or by accounting for polarization *etc.*

Because of the large number of geometry-optimisations that must be performed, relatively simple force field methods are the most common, as they are computationally affordable. Multipole-based force fields are sometimes used in the later stages of a CSP. Recently, the use of tailor-made non-transferable force fields, parametrised to perform optimally for a specific molecule has attracted attention [162].

An alternative is to use a genetic algorithm [163, 164]. The crystal structure is represented by a genome in the form of a binary string. Optimisation is then performed by introducing sexual reproduction of the genome, adding mutations *etc.* The fitness of a structure is determined by a scoring function, which usually include estimates of the lattice energy and artificial potentials based on commonly observed close contact geometries [132].

1.3.3.5 Clustering

Many of the generated crystals will optimise to the same structure. The duplicates need to be identified and removed. For this we need some measure of crystal structure

similarity. In this thesis I use a method based on extracting and overlaying clusters of molecules from the crystal structures and calculating their similarity by a root mean square distance between atoms [165].

Other methods for calculating crystal structure similarity are discussed in Chapter 2.

1.3.3.6 Final ranking

The set of final unique predicted structures are then often re-ranked using a more accurate, and more costly, scoring function than could be used during the optimisation. After clustering, the number of structures are much fewer, typically a few hundred, allowing computationally far more expensive methods to be used to rank the structures.

Sometimes a point charge model is used for modelling electrostatic interactions during a CSP and a more accurate multipole-based force field is used to calculate the lattice energy of the final set of predicted crystal structures. As the blind tests of crystal structure prediction have clearly demonstrated, highly accurate methods for calculating the crystal lattice energy are necessary for successful predictions [25, 5]. Hence, CSP is strongly tied to the study of intermolecular forces and the development of highly accurate force fields [166, 167, 168, 2].

Recently, two methods for the final ranking have attracted the most interest and have performed well in objective blind tests [135, 25]. The first is to add a vibrational contribution to the lattice energy, so as to estimate free energy differences between the predicted crystals. Because of the computational cost of free energy calculations, force fields are still commonly used, and this is the method that will be explored in this thesis.

The other method, which was quite common in the last blind test [5], is to use periodic electronic structure calculations in an attempt to calculate more accurate lattice energies. The PBE density functional with some dispersion correction is the most commonly used. Using periodic DFT for the final scoring became popular when Neumann *et al.* demonstrated great successes with such methods in the fourth and fifth blind tests [135, 108, 25, 142].

1.4 Experimental techniques for the study of crystals

1.4.1 Single crystal X-ray diffraction

X-ray diffraction remains the most important method for determining the structure of crystals. The methodology is nowadays largely automated and proceeds essentially as follows.

An individual crystal of good quality and at least 0.1 mm on the side is mounted on a goniometer. The goniometer rotates the crystal in a beam of monochromatic X-rays. The crystal diffracts a small fraction of the X-rays and these are detected. Historically photographic paper was used, but now large CCD arrays are used. The X-ray tube, goniometer and CCD-image sensor are integrated into a single computer-controlled crystal diffractometer instrument. Data analysis is performed by largely automated computer software on ordinary desktop computers.

To reduce thermal vibrations in the crystal and to reduce damage to the crystal by the X-rays, the crystal may be cooled in a flow of cold nitrogen gas from a reservoir of liquid nitrogen.

The recorded images of diffraction spots are processed by Fourier transformation to yield a 3D electron charge density. Determining the phases of the transform was historically a major problem. For crystals of small organic molecules it is normally not a problem any more, since the development of fast computers allow computationally expensive so-called 'direct methods' to be used [9]. The theory of X-ray diffraction will not be repeated here, excellent textbooks have been written on the subject [9].

There is one final note the reader should remember. It is the electrons that diffract X-rays, the nuclei are not detected, but their locations are inferred from local maxima in the charge density. However, hydrogen atoms are actually not located at the measured charge density maxima but about 0.1 Å further out. *Ab initio* calculations and neutron diffraction yields the true hydrogen positions. This has consequences for the parametrisation and accuracy of empirical force fields, as will be discussed in Chapter 3.

1.4.2 Powder X-ray diffraction

Diffraction experiments can also be carried out on micro-crystalline powders. The random orientation of the crystallites causes the X-rays to be diffracted in so-called Laue cones, resulting in concentric circles on the detector. The angles of the diffraction cones and the relative intensity of the diffracted radiation is measured, resulting in a 2-dimensional 'diffraction pattern', the intensity is plotted as a function of the cone angle, conventionally denoted 2θ.

Each cone corresponds to a (Miller) plane in the crystal lattice, such that each peak in the diffraction pattern can be associated with an ordered triple of lattice indices. Assigning Miller indices to diffraction spots or cones is called indexing and allows the determination of the unit cell dimensions and space group symmetry. This is almost always the first step in the determination of a crystal structure by diffraction experiments.

Powder patterns can serve as unique fingerprints of crystal structures. In Chapter 6, the structures of new polymorphs of chloridazon are determined by directly comparing experimental powder patterns with simulated patterns of a set of predicted crystal structures.

1.4.3 Differential scanning calorimetry

In differential scanning calorimetry (DSC), a sample (powder or single crystal) and an inert reference (often indium, gallium, caffeine and benzoic acid) are heated by individual electric heating elements. The instrument heats the sample such that the temperature of the sample and reference increases at a specific (usually constant) rate. If there is a phase transition in the sample, this will result in the adsorption or release of heat, requiring the instrument to change the heat flow by adjusting the effect of the sample's heating element.

The measured quantity is hence the heat flow as a function of either time or temperature. This offers a greater precision than the older differential thermal analysis (DTA) method in which the instrument instead measured the temperature and heated the sample by a fixed effect or heat flow.

Integrating the peaks in a DSC curve directly gives the enthalpy change ΔH of the transition. It is often useful to run cycles of heating and cooling to see if any observed phase transition is reversible.

Comprehensive reviews of the use of thermal analysis techniques in the characterization of organic crystal structures have been published by Giron and Threlfall [169, 170].

1.4.4 Hot stage microscopy

Optical microscopes are useful for the study of crystals. With a microscope it is easy to see the shape of the macroscopic crystal, the so-called *habit*. The habit depends on the relative growth rates, *i.e.* how easily a molecule attached to, of the different (Miller) surfaces of the structure.

It is also often possible to see if there are more than one polymorph present in a sample. If the microscope is also equipped with a heating element so that the temperature of the sample can be controlled, it becomes a very powerful tool. A hot stage microscope allows the investigator to directly see phase transitions in minute samples and to precisely determine the temperature of such transitions. Polarized light also greatly helps as the birefringence and refraction of crystals changes abruptly in most phase transitions.

Hot stage microscopy was largely developed at the Innsbruck school, founded by the Koflers [20]. The Kofler hot stage, invented in 1954 [171] but based on earlier work

[172], has been extensively used in Innsbruck by the Koflers, Kuhnert-Brandstätter, Burger and Griesser [171, 173, 174, 175].

1.5 Outline of this thesis

Now that the reader has been introduced to the terminology, background and available computational and experimental methods, I can describe the main purpose of this thesis: To demonstrate that clathrate crystal structures, and their properties, can be predicted computationally.

The prediction methodology and the theory behind it will be covered in more detail in Chapter 2. The existing crystal structure prediction methods that have already been introduced will be used. To these are added lattice dynamics free energy calculations and the focus is on performing accurate, yet affordable, calculations on hundreds of crystal structures. Special emphasis is laid on the convergence properties and accuracy of such free energy calculations.

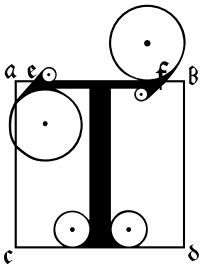
The accuracy of the methods will be thoroughly tested in Chapter 3. I will then apply the algorithms described to the study of polymorphs in Chapters 4, 5 and 6. Successful blind predictions of several compounds are described throughout the thesis. Predictions of clathrate structures are described in Chapter 7. Calculations of temperature-dependent properties of hypothetical clathrates are described in Chapter 8.

The thesis concludes in Chapter 9 by summarising the current state of the art in computational crystal prediction and reiterates the most important lessons learned. A number of appendices follow, containing details of some software programs that were developed in the course of this thesis.

In the end, I hope to have convinced the reader that the methods described here are suitable, accurate enough, economical and useful for the scientific study of polymorphs and clathrates.

Chapter 2

Theory and Methods



This chapter describes in greater detail the methods that have been used in the preparation of this thesis. Theoretical justifications for the practical models are discussed, as well as important shortcomings where simplifications cause appreciable errors in the calculations. Accurate methods for calculating the cohesive energy holding the molecules together in the solid state are introduced, focusing on dispersion-corrected density functional approximations and anisotropic force fields.

The rigid-body lattice dynamics method for calculating vibrational contributions to the free energy is described. Approximate methods that account for phonon dispersion and improves the convergence of lattice dynamical calculations form a significant part of this thesis and are discussed here in some detail.

Methods for comparing crystal structure similarity and for calculating the volume and shape of crystalline cavities are also covered.

The chapter ends with an brief introduction to solid-state nuclear magnetic resonance spectroscopy and the computational modelling thereof. The focus is limited to ^{129}Xe solid state NMR that is applied in Chapter 7 for the study of clathrates.

2.1 Intermolecular interactions in the solid state

2.1.1 Theory of intermolecular interactions

Fundamentally, all interactions between molecules are electronic in nature, they are caused by the positions and movements of the electrons and nuclei. However, it is useful to separate the total interaction into a few different parts according to the physical effects they are mediated by and the distances at which they are most pronounced.

The total cohesive energy holding a crystal structure together, here called the lattice energy E_{latt}, can be written as a sum of pairwise interactions, three-body interactions, four-body interactions *et cetera* between molecules $k, l, m \ldots$ in the system.

$$E_{\text{latt}} = \sum_{k<l} E_{kl} + \sum_{k<l<m} E_{klm} + \sum_{k\ldots} E_{k\ldots} + \ldots . \tag{2.1}$$

This expression is truncated after pairwise or three-body interactions in practical calculations. The higher-order terms can to some extent be absorbed into an *effective* pairwise-additive model potential.

Note however, that already the pairwise term contains interactions between *all pairs* of molecules in the structure. Recently, we have seen models where interactions only between nearest neighbour molecules are considered [176, 177, 178]. There is not much physical justification for such methods since intermolecular interactions in the solid state, especially electrostatic interactions, are long-range and extend far beyond the distance between nearest neighbour molecules. Such methods can be misleading [179] and may lead to even qualitatively incorrect results and conclusions. Analogously, atom-atom close contacts and supramorepulsion haslecular synthons [180] are often interpreted as strong intermolecular interactions and even the *cause* of crystal stability [181, 182]. However, Dunitz and Gavezzotti have convincingly argued against this view [179, 183].

A similar problem arises in attempts to understand intermolecular interactions in the solid state from local features in the electron density. So-called Hirschfeld surfaces [184, 185] or methods identifying stationary points in the charge density [186, 187] cannot be used to draw conclusions about the properties of crystals or the aetiology of polymorphism. See also Gavezzotti [188] for a critical discussion of common crystallographic misconceptions.

This thesis aims to avoid the pitfalls mentioned above by the use of methods for calculating the full lattice energy by considering interactions between all molecules. Intermolecular interactions can be grouped depending on their distance-dependence. The long-range potentials depend on the intermolecular distance r as r^{-n}, where n is an integer. They are listed in Table 2.1.

Table 2.1: Long-range intermolecular interactions. Table adapted from [85].

Interaction	Additive	Comment
Electrostatic	yes	Anisotropic
Induction	no	Many-body effect
Dispersion	approx.	
Magnetic	yes	Very small

The electrostatic interaction is due to the attraction and repulsion between charges, dipoles and higher multipole moments. It is often responsible for a large part of the

total lattice energy of organic crystals. Hydrogen bonds are mainly electrostatic in nature. In salts or crystals of molecules with large dipole moments, such as water or zwitterionic amino acids, the electrostatic interactions are very dominant.

The induction, or polarization, is the mutual perturbation of molecular charge densities due to the electromagnetic fields surrounding them. It is a quantum-mechanical many-body effect and because of that challenging to model in force field methods. It is however very well accounted for in DFT calculations.

The dispersion is due to electron correlation. According to London's theory [90], dispersionattraction arises because of the mutual attraction between instantaneous and induced dipoles, quadrupoles *etc* [85]. The most important term, the dipole-induced dipole interaction decays with the distance raised to the sixth power. Higher multipole moments give rise to r^{-8}, r^{-10}-terms *etc*.

The short range forces arise due to overlap of the electron orbitals. Because they are fermions, the electrons can not occupy the same state, leading to a strong repulsive force when the atoms come so close to each other that their electron clouds overlap. This is the Pauli repulsion. A small attractive force due to electron exchange is usually incorporated in this term as well [85].

Short range interactions decay approximately exponentially with the distance, $V \sim \exp(-Br)$, a direct consequence of the exponential decay of the charge density with respect to distance from the nucleus. The most important short range forces are summarised in Table 2.2.

There can also be a partial delocalisation, or transfer, of electrons between molecules, as in molecular salts, or in strong hydrogen bonds. This charge-transfer is non-additive and gives rise to an attractive force. It is difficult to model with classical force fields, but most force fields have more than one 'type' of each element. This allows some effects of the immediate surroundings of an atom to be captured, such as whether an atom is part of a hydrogen bond or not.

Table 2.2: Short-range intermolecular interactions. Adapted from [85].

Interaction	Additive	Comment
Pauli repulsion	yes	Large at very short distances
Exchange-induction	approx.	Part of exchange-repulsion
Charge transfer	no	Many-body effect

Since the strength of the interactions decay with distance, it is practical to introduce a cutoff distance and only consider short-range interactions between molecules at distances smaller than the cutoff, typically 15 or 20 Å. This does not work well for electrostatic interactions between charges and dipoles, however, see below.

2.1.2 Electrostatic interactions

Electrostatic interactions between atoms and molecules are most often modelled as point-like partial charges at the positions of the nuclei. The electrostatic potential energy $V_e(\mathbf{r})$ of a system of n point charges q_i, $i = 1, \ldots n$ at positions $\mathbf{r}_i$, $i = 1, \ldots n$ is

$$V_e(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \sum_{i=1}^{n} \frac{q_i}{|\mathbf{r} - \mathbf{r}_i|} \tag{2.2}$$

Many common force fields, for instance AMBER, OPLS and COMPASS, use this approach [89, 189, 88]. The partial charges can be fitted parametrically to experimental data, or calculated to reproduce the electrostatic potential of a molecular charge density [190].

An important drawback of atomic partial charge models is that they do not adequately model the directionality of electrostatic interactions. Strongly directional interactions, such as the dipole-charge interaction in hydrogen bonds, or the interactions between overlapping p-orbitals in $\pi - \pi$-stacking are often not reproduced accurately. Including dipoles and higher multipoles provides a significantly better model and is known to improve lattice energies of crystal structures [191, 137].

The theory of multipole moment representations of intermolecular interactions has been described in great detail by Stone [85]. Only a brief summary is included here.

The electrostatic potential between the charge densities $\rho(\mathbf{r})$ of two molecules A and B is

$$V_e = \int \frac{\rho_A(\mathbf{r}_1)\rho_B(\mathbf{r}_2)}{\mid \mathbf{r}_1 - \mathbf{r}_2 \mid} \quad \mathrm{d}^3\mathbf{r}_1 \mathrm{d}^3\mathbf{r}_2 \tag{2.3}$$

Evaluating this integral with the complete charge densities would be far too slow for practical simulations. It is possible to discretize the charge density into a number of density 'pixels' (actually voxels), an approach used in Gavezzotti's PIXEL program [192].

An alternative is to consider the molecule's multipole expansion. The multipoles Q_{lk} are obtained from the charge density by

$$Q_{lk} = \int \rho(\mathbf{r}) R_{lk}(\mathbf{r}) \quad \mathrm{d}^3\mathbf{r} \tag{2.4}$$

Where R_{lk} are the regular solid harmonic functions. The index l denote the rank of the multipole and k enumerates the $2l + 1$ non-zero moments.

There is however a serious problem in using molecular multipoles. The expression for the electrostatic potential at an arbitrary point R converges only if R is located outside of the sphere encircling the molecule. Hence, the electrostatic interactions will be correctly calculated only for molecules that are relatively far apart. This is clearly not the case in molecular crystals.

One solution to this convergence problem is to distribute the multipoles over several sites in the molecule, each site representing a smaller part of the molecular charge distribution. An obvious and convenient choice is to place the distributed multipoles at the nuclear sites.

In this thesis, extensive use will be made of distributed multipoles, calculated from electron charge densities obtained from DFT calculations. The distributed multipole analysis by Stone will be employed [193], which uses properties of Gaussian basis sets for the calculations [194].

Molecular wavefunctions are often expressed as a sum of Gaussian basis functions located at atomic sites $\mathbf{a}_s$ of the form

$$\psi(\mathbf{r}) = \sum_s c_s R_{lm}(\mathbf{r} - \mathbf{a}_s) \exp\left(-\zeta(\mathbf{r} - \mathbf{a}_s)^2\right) \tag{2.5}$$

Here c_s are coefficients, R_{lm} are regular harmonic functions, and ζ a parameter for the radial decay of the Gaussian.

Since the charge density is $\rho(\mathbf{r}) = \langle \mathbf{r}|\psi^*\psi|\mathbf{r}\rangle$ we consider pairwise products of basis functions. The product of two Gaussian basis functions centred at coordinates $\mathbf{a}$ and $\mathbf{b}$ is another Gaussian centred at a point p between $\mathbf{a}$ and $\mathbf{b}$. The pre-exponential regular harmonic $R_{lm}(\mathbf{r} - \mathbf{r}')$ can always be moved to the same point, where it gives a linear combination of spherical harmonics up to rank l. Hence, using a Gaussian basis set is particularly suitable for deriving distributed multipoles [194]. The multipoles at the p-sites thus give an exact representation of the molecular multipoles in a finite series of distributed multipoles.

Since a large number of basis functions are used for the wavefunction, the number of multipole sites becomes very large. The multipole sites p that do not coincide with the position of a nucleus are therefore moved to the nearest nuclear site, resulting in series of spherical harmonics with Clebsch-Gordan coefficients up to rank $l + l'$ at the nucleus [195, 196]. Moving a multipole like this results in higher moments in the spherical harmonics, so that the multipoles must be truncated, leading to a small error. In this thesis, multipoles up to rank $l = 4$, hexadecapoles, are normally used on all nuclear sites.

The convergence of the electrostatic potential with respect to increasing multipole moments has been studied. The distributed multipole analysis used here gives a relatively fast convergence to the reference potential of the full charge density [85]. Alternative methods to derive multipoles, such as the iterated stockholder algorithm, may have faster convergence still or yield zeroth moments that correspond better to classical partial charges [197].

The computer program GDMA [198, 199], by Stone, is used in this thesis to perform the distributed multipole analysis from molecular charge densities calculated with Gaussian. Other programs for calculating distributed multipoles include MULFIT and OPEP [200, 201].

It is important to remember that the electrostatic potential decays as r^{-1} and is therefore very long-range. This makes the use of a cutoff-distance hard to justify. The practical implementation of such a cutoff also leads to difficulties as the sum over all pairwise electrostatic interactions does not converge smoothly with respect to the cutoff radius. A very common solution is to employ an Ewald summation, where the discrete point charges are replaced by Gaussian functions and the sum is performed in periodic boundary conditions [202]. This also allows using the convolution theorem to perform the calculation in the lattice's Fourier space, the reciprocal lattice. Multipole moments can also be calculated with Ewald summation [203].

One possible problem regarding Ewald summation for molecular crystals is that if the unit cell has a dipole moment, then the energy of the dipole-dipole interactions depends on the shape of the macroscopic crystal, the habit [204]. The effect is almost always neglected, and this negligence has been both criticised [205] and defended [206]. The most important effect that limits the error is that polar crystals are typically produced by crystallisation from water solutions with high dielectric constants. In such conditions, the crystal's dipole moment will be screened by the dielectric medium. Also, assuming that crystal growth occurs in a way that minimises the free energy, polar crystals will tend to have needle-like habits, which also minimises the effect [206]. Hence, for crystal structure prediction, this error is probably not large.

In Chapter 3 we will see how force fields with multipoles and point charges perform in calculations of crystal structure properties.

2.1.3 Force fields

2.1.3.1 Origin of force field methods

The potential energy of a chemical system can be modelled by a set of simple analytical functions of atomic coordinates. Such methods are known as force field methods and were introduced by Hill [86]. Important contributions were also made by Lifson and Warshel [207], Williams and Kitaigorodskii [208, 122]. Below I go through the main features of force fields for molecular mechanics calculations.

2.1.3.2 Bonded interactions

Classical force fields consist of intramolecular (bonded) and intermolecular (non-bonded) terms, which are all strictly additive.

$$E_{\text{bonded}} = E_{\text{bonds}} + E_{\text{angles}} + E_{\text{dihedrals}} \tag{2.6}$$

The intramolecular terms model covalent bonds and angles and dihedrals between covalent bonds. Covalent bonds between two atoms of type ι and κ are most often modelled with a harmonic Hookean potential.

$$E_{\text{bond}} = \sum_{\text{bonds}} k_{\iota\kappa}(r_{\iota\kappa} - r_{ik})^2 \tag{2.7}$$

Here $k_{\iota\kappa}$ and $r_{\iota\kappa}$ are two atom type-dependent parameters that describe the strength of the potential and the equilibrium distance between the atoms. The potential energy hence varies quadratically with respect to the distance from the equilibrium distance $r_{\iota\kappa}$. An alternative that more accurately describes the anharmonicity in covalent bonds and allows bond-breaking is the Morse potential, but it will not be used in this thesis.

Bond angles are similarly modelled as harmonic, with an equilibrium angle and a quadratic potential with respect to deviations from the equilibrium.

Dihedral angles, describing torsional strain, cannot be based on a Taylor expansion as the previous expressions, since the potential must be 360°-periodic. A Fourier expansion type functional form is therefore used.

$$E_{\text{dihedrals}} = \sum_{\text{dihedrals}} k_{\iota\kappa} \left(1 + \cos\left(n\tau + \delta\right)\right) \tag{2.8}$$

Additional terms for out-of-plane bending are used for rings in molecules.

2.1.3.3 Non-bonded interactions

Non-covalent intermolecular interactions are modelled as a sum of electrostatic interactions E_{elec} and van der Waals interactions E_{vdW}. Electrostatic interactions are most often modelled as Coulombic interactions between electric partial charges assigned to the atomic nuclei, but in this thesis extensive use of multipole representations will be used instead. These have been shown to be more reliable and accurate for molecular crystals [191, 168, 2].

Expressions for interaction energies, forces and torques between multipoles have been tabulated elsewhere and are not reproduced here [195, 85, 209].

Non-electrostatic intermolecular interactions, primarily short-range repulsion and long range attractive dispersion are commonly described by a Lennard-Jones potential function:

$$E = 4\epsilon_{\iota\kappa}\left(\frac{\sigma_{\iota\kappa}^6}{r_{ik}^6} - \frac{\sigma_{\iota\kappa}^{12}}{r_{ik}^{12}}\right) \tag{2.9}$$

or a Buckingham potential:

$$E = A_{\iota\kappa}\exp\left(-B_{\iota\kappa}r_{ik}\right) - C_{\iota\kappa}r_{ik}^6 \tag{2.10}$$

The exponential shape of the short-range repulsion comes from the exchange-repulsion between overlapping electron densities, which to a good approximation has an exponential distance-dependence [210]. This 'Born-Mayer' term thus has a better physical justification. Even more sophisticated expressions for the short range repulsion have been derived [211]. The Lennard-Jones potential has historically been more common because it is much faster to compute.

The r^{-6}-dependence of the attractive van der Waals term comes from the theory of London dispersion [90]. A closer look at that theory reveals that the r^{-6}-term is only the first in a series expression, it is the distance-dependence of dipole-dipole interactions. The next two terms, with r^{-8} and r^{-10}-dependence, come from dipole-quadrupole and quadrupole-quadrupole interactions respectively that may also be significant. Such terms are included in some of the most accurate dispersion correction schemes for density functional theory, see below.

A problem with the dispersion models we have covered so far is that London's derivation of the functional form is only valid at relatively large distances. The interatomic potential should have a Born-Mayer shape at very small distances, but as r goes to 0, the Buckingham potential approaches $-\infty$. This can be solved by adding a damping function to the dispersion model. A suitable damping function which does not introduce new parameters was proposed by Tang & Toennies in 1984 [212] and has the form:

$$f(n) = 1 - \exp(-Br)\sum_{k=0}^{n}\frac{(Br)^k}{r}, \tag{2.11}$$

where n denotes the distance dependence of the dispersion, normally 6, 8 and 10. The B is the same exponent as in the repulsion term.

An important limitation we have imposed in the formulation of these force fields is that we assume that the interactions are pairwise additive. This is not true, since the interactions arise because of quantum-mechanical effects that are fundamentally many-body effects. The most important non-additive contribution is the polarization. The leading term is a dipole-induced dipole interaction, which has a r^{-6} dependence. This is fortunate, because it means that in empirical force fields, some of the polarization will be captured by the dispersion model.

Other important non-pairwise additive terms are the three-body dispersion terms. Such interactions have been shown to sometimes be important for highly accurate modelling of organic crystals [100]. Expressions for the dispersion between three interacting bodies were derived by Axilrod, Teller and Muto in 1943 [213]. These terms are sometimes included in sophisticated models for dispersion, such as the many-body dispersion correction for density functional approximations developed by Reilly and Tkatchenko [214].

One final caveat in the calculation of intermolecular interactions is the halogen bond. It has been observed that in crystal structures, the distance between a halogen atom and its nearest neighbour atoms tends to be shorter when the angle of the close contact is close to 180° with respect to the halogen atom's covalent bond. Hence, halogen atoms appear the be flattened and to have an anisotropic repulsion. The repulsive wall tends to have a smaller radius in the direction of the bond [215, 216]. The halogens appear to be flattened and have a spheroid shape. To account for this the exponential term in the Buckingham potential may be replaced by a function of the contact angle θ [166, 167].

$$r(\theta) = ab\sqrt{(a^2\cos^2\theta + b^2\sin^2\theta)^{-1}} \tag{2.12}$$

where a is the atomic major radius and the radius perpendicular ($\theta = 90°$) to the C–X bond axis. The minor axis b is parallel ($\theta = 0°$) to the C–X bond.

The reason is that the electron density around halogen atoms is anisotropic. The covalent σ-bond causes an electron deficiency at the opposite side of the halogen atom, a so-called 'σ-hole', which enhances the strength of interactions with electron-rich atoms at short range. Much has been written about halogen bonds and their importance, or lack thereof [217, 218, 219].

The anisotropy is small for fluorine but significant for the heavier halogens where the difference between the polar and equatorial radii is about 1 Å [216].

Also oxygen, sulfur and selenium exhibits anisotropic van der Waals radii [220, 216] because of their unpaired electrons, this effect should however be captured adequately by the electrostatic multipoles.

2.1.4 The Williams force fields

In this thesis, we will make extensive use of empirical force fields introduced by Williams. The first force field parameters were published already in 1966 [208]. Successive reparametrisations and improvements have been published in several articles since [221, 222, 223, 92, 224, 93, 167, 225].

These force fields have been used extensively for the modelling of organic molecular crystals and have over time become successively more accurate. The force fields contain

parameters for modelling the non-electrostatic interactions. Electrostatic interactions must be added separately either as point charges or as distributed multipoles. These force fields have withstood the test of time and have been used in many successful computational studies of crystal structures.

Williams parametrized the W84 and W99 force fields to experimental crystal structures and sublimation energies. Fitting force field parameters to experimental data of crystals introduces several important errors. These must be carefully considered both during the parametrization and in the subsequent use of force fields. Some of the sources of errors are described below.

X-ray diffraction experiments systematically fail to determine the positions of hydrogen atoms. The covalent bond of the hydrogen appears to be about 0.1 Å too short when compared to neutron diffraction data or quantum chemical calculations. This means that a force field parametrized from X-ray structures will not be accurate for quantum-chemically geometry-optimized molecules. One way to counteract this error is perform the force field calculations on molecular conformations where the hydrogen positions have been foreshortened.

A second error is introduced because of the X-ray diffraction experiments are performed at a non-zero temperature. The crystal structures expand because of thermal motions, and this effect is then erroneously absorbed by the force field parameters. When fitting a force field, structures determined at low temperatures (< 100 K) are preferred, but the force field error due to thermal expansion is probably significant.

A third error is that the sublimation enthalpy ΔH_{sub} is often used instead of the lattice energy when fitting force fields. The quantities differ because of the energy of thermal vibrations.

$$\Delta H_{\mathrm{sub}} \simeq -E_{\mathrm{latt}} - 2RT \tag{2.13}$$

Here R is the general gas constant.

All these errors can in principle be alleviated. If the experimental structures were to be optimised with an accurate *ab initio* calculation before the fitting, the force field would correspond to structures at 0 K. However, since a large number of crystal structures are needed, the computational cost would be quite considerable.

For the purpose of CSP, the Williams force fields have proven very useful when combined with electrostatic multipoles. Then these force fields often reproduce experimental crystal structures with errors in lattice parameters of a few percent or less. In this thesis, an extensive benchmarking study of the Williams force fields will be presented, see Chapter 3.

An interesting possibility is to replace the parameters in the force field with functions of temperature [226]. In particular, the pre-exponential term in the Buckingham potential

can be made to depend on temperature such that the apparent van der Waals radii [227] of the atoms increase at higher temperatures [221]. This way thermal expansion can be accounted for. This thesis will use another method however, based on minimising the free energy with respect to unit cell dimensions.

Parameters for chlorine were not determined by Williams, but have been parametrized later using *ab initio* calculations [167]. For halogen atoms and sulfur, a modified form of the repulsion term can be used to account for the anisotropy. The repulsion is made dependent on the contact angle between atoms. This is used throughout the thesis for fluorine and chlorine.

Recently, the Williams99 force field [92, 93] was carefully re-parametrized to improve the accuracy of hydrogen bond strength and directionality. Atom types that participate in hydrogen bond formation were re-parametrized using a selected set of high-quality low-temperature crystal structures. In addition, the cross terms were explicitly parametrized instead of relying on mixing rules. Electrostatic interactions were treated with multipoles calculated from B3LYP/6-311G** charge densities, and the sublimation enthalpies were adjusted with $2RT$ to account for thermal contributions. This re-parametrisation also considered multipoles calculated in a polarising environment, deriving separate force field parameters to be used with polarised and non-polarised multipoles respectively. Significant improvements in the directionality of hydrogen bond interactions could be seen in these new revised force fields [225].

2.2 Deducing the chemical contents of a crystal structure

The common file formats for storing crystallographic data, such as the Shelx and IUCr's cif formats [228, 229], do not contain any information about the atomic connectivity, functional groups, the molecules and their stoichiometry. We need to determine this from interatomic distances only.

Since crystals are periodic, a very commonly occurring issue in computational crystallography is that interatomic distances in periodic boundary conditions need to be calculated often and quickly. Since the fractional coordinates describing the crystal structure are periodic and may be outside of the [0, 1) range, it may not be enough to consider atoms in the unit cell and its 26 neighbouring cells.

Instead of calculating distances from one atom to another in every periodic image, there is a much more efficient algorithm. Periodic boundary conditions in (possibly) non-orthorhombic unit cells are applied by calculating the minimum-image distance in fractional coordinate space. Since it is of general interest, and the original description of the method is difficult to find [230], the algorithm for calculating distances in non-orthorhombic periodic boundary conditions is reproduced here.

A crystal structure, or any parallelepipedic periodic boundary conditions, can be described by three lattice vectors **a**, **b**, **c** and a set of (atomic) fractional coordinates. If only the lattice parameters are given, the corresponding lattice vectors can be obtained from standard text-book equations.

Let $\mathcal{S}$ denote the fractional coordinate space. The transformation of fractional coordinates $\mathbf{s}_i$ to Cartesian ("real") space $\mathcal{R}$ is

$$\mathbf{r}_i = \mathbf{H}\mathbf{s}_i \tag{2.14}$$

where the transformation matrix $\mathbf{H}$ is formed by the three lattice vectors as columns.

The key to calculating distances in non-orthorhombic periodic unit cells is to apply the periodic boundary conditions in fractional space and subsequently transform the minimum image distance vector to real space. The minimum distance between points $\mathbf{s}_i$ and $\mathbf{s}_j$ in fractional space is

$$\mathbf{s}_{ij} = (\mathbf{s}_i - \mathbf{s}_j) - \mathrm{fint}(2(\mathbf{s}_i - \mathbf{s}_j) - \mathbf{1}), \tag{2.15}$$

where fint() is the (element-wise) Fortran INT() function, defined as

$$\mathrm{fint}(x) = \mathrm{sgn}(x)\lfloor |x| \rfloor = \begin{cases} \lfloor x \rfloor & x > 0 \\ -\lfloor x \rfloor + 1 & x < 0 \end{cases} \tag{2.16}$$

The vector $\mathbf{s}_{ij}$ is then transformed to Cartesian space, using $\mathbf{H}$, and its norm is calculated.

The chemical contents of a unit cell can then be deduced by graph theoretical methods. A connectivity matrix is built up by comparing the distance between every pair of atoms in the unit cell with a set of cutoff values. The connectivity matrix is then used to describe a graph. The subgraphs of this (possibly disconnected) graph are the molecules in the unit cell. The subgraphs can be conveniently extracted as individual graph objects using the SciPy and NetworkX libraries [231, 232]. This is mathematically equivalent to identifying the connected components of a disconnected graph, a problem for which efficient algorithms exist [233].

By subsequently solving the graph isomorphism problem [234, 235] for every pair of molecular graphs, we can deduce the formula unit and get a complete description of the unit cell contents. This is particularly useful for multicomponent crystals and structures with non-integer Z'.

This provides an efficient method for deducing the chemical contents of a crystal structure given only the asymmetric unit, the lattice parameters and the symmetry operations, which are specified in all computer files for crystal structures.

These algorithms have been implemented in Python and have been used extensively in performing the calculations described in this thesis. In particular, these methods are useful for automating large scale calculations on crystal structures of different compounds, deducing the stoichiometry of solvates and clathrates and in automatically choosing which degrees of freedom to optimise in CrystalOptimizer [236].

2.3 Automatic selection of molecular flexibility

An algorithm for automatically choosing flexible degrees of freedom, *i.e.* bond lengths, bond angles and dihedrals, was devised and implemented in Python. The algorithm essentially proceeds as follows.

Default hard-coded input files for CrystalOptimizer are written to disk. CrystalOptimizer is then called, without specifying any degrees of freedom, prompting the program to generate a Z-matrix for each molecule in the asymmetric unit and write these to disk together with a list of atomic connectivities. The algorithm used for constructing the Z-matrix is described in Appendix A of the PhD-thesis by Kazantsev [237].

The output file from this first dry run we can call 'FirstRun.out'. The FirstRun.out file is read and the connectivity data is used to construct connectivity matrices. Using the SciPy library a graph object is initialized from the connectivity sparse matrix.

By breaking bonds, *i.e.* setting elements in the connectivity matrix to 0, and identifying the presence of distinct subgraphs [231, 232], covalent bonds that are not part of a ring can be identified.

Rings can be transversed by successively moving to the nearest neighbour atoms and testing if its bonds are part of a ring or not. In the algorithm implemented here, condensed ring systems are not identified or dealt with, and this does not seem to be necessary for effectively choosing a suitable set of degrees of freedom.

The following rules have been found to adequately capture the flexibility of molecules. The rules were implemented as the AutoDof program, see Appendix A.

- Covalent bond lengths are optimised without considering packing forces;
- All angles and dihedrals containing a polar hydrogen atom (–OH, –NH, –SH[1]) are optimised under the influence of packing forces;
- All exocyclic bonds are considered rotatable and are optimised under the influence of crystal packing forces;

[1]There are no thiols in the structure set, as very few polymorphic thiol compounds are known. Also, the force field used in this work does not contain parameters for thiol hydrogen atoms.

- Dihedrals and angles in 3- and 4-membered rings are optimised without considering packing forces;
- Dihedrals and angles in 5- and 6-membered rings consisting of 3-coordinated carbon atoms and nitrogen in any combination are unaffected by packing forces, except dihedrals and angles that contain a polar hydrogen atom;
- Dihedrals and angles in 5-membered rings containing sulfur or oxygen[2] atoms bonded to two 3-coordinated carbon atoms are optimised without considering packing forces;
- Any remaining dihedrals are optimised with respect to packing forces;
- Linear angles in –C≡N and –C≡C– groups are constrained to 179.99°.

For completely rigid molecules we allow the first angle to be optimized with respect to packing forces, since CrystalOptimizer needs at least one flexible degree of freedom. We use CrystalOptimizer also on formally rigid molecules in order to optimise them to the minimum DFT energy conformer. In particular this optimises the hydrogen positions in a consistent way.

2.4 Density Functional Theory

2.4.1 Fundamentals of density functional theory

Of the different quantum-chemical methods, density functional theory has become the most popular for modelling molecules, because it is accurate enough and affordable for most applications. Here a very brief introduction is given. More detailed descriptions are available in the cited literature.

Let ρ denote the electron density of a system. The energy of the system is then a functional of the density:

$$E[\rho] = V_{\mathrm{ne}}[\rho] + T[\rho] + V_{\mathrm{ee}}[\rho], \tag{2.17}$$

Here the density is itself a function of the spatial coordinates, and represented by a superposition of more fundamental basis functions.

The electrostatic interaction energy between the electron density and the nuclei is denoted V_{ne}. By $T[\rho]$ we denote the kinetic energy, a scalar functional of the density, and V_{ee} is the electron-electron interaction energy.

[2]Specifically furan, thiophene, 1,3-oxazole, 1,3-thiazole, 1,3,4-oxadiazole and 1,3,4-thiadiazole rings.

The nuclear-electronic interaction is trivial to calculate, but for the kinetic energy and the electron-electron energy exact functionals are not known. They depend on many-body interactions between the electrons. A practical approximate solution was provided by Kohn and Sham in 1965 [238]. Instead of the real system at hand, imagine a fictitious system of *non-interacting* electrons in a set of orbitals ϕ_i, having the same electron density as the real system.

The density is then, naturally,

$$\rho(\mathbf{r}) = \sum_i |\phi_i(\mathbf{r})|^2 \tag{2.18}$$

and the kinetic energy is

$$T_s[\rho] = -\frac{1}{2}\sum_i \langle\phi_i|\nabla^2|\phi_i\rangle\,, \tag{2.19}$$

where we use atomic units and $e = m_\mathrm{e} = \hbar = 1$.

The classical Coulomb repulsion between the electrons is easy to calculate and is denoted $J[\rho]$. We can now rewrite the energy functional as

$$E[\rho] = V_\mathrm{ne}[\rho] + T_s[\rho] + J[\rho] + E_\mathrm{xc}[\rho], \tag{2.20}$$

where E_xc is the *exchange-correlation* functional, which is still unknown. However, we know that the system obeys the Schrödinger-like eigenvalue equation $\hat{\mathcal{K}}\phi_i = \varepsilon_i\phi_i$. Hence, we can write a Kohn-Sham (Hamiltonian) operator $\hat{\mathcal{K}}$ as

$$\hat{\mathcal{K}} = -\frac{1}{2}\nabla^2 + v(\mathbf{r}) + v_\mathrm{xc}(\mathbf{r}) + \int \frac{\rho(\mathbf{r})}{\mid \mathbf{r} - \mathbf{r}' \mid} \quad \mathrm{d}\mathbf{r}' \tag{2.21}$$

Here the first term on the right hand side is the kinetic energy operator, $v(\mathbf{r})$ is the external potential acting on the system, *i.e.* the electrostatic potential of the nuclei. The last term is the Coulomb self interaction. The exchange-correlation potential is, by definition, the functional derivative

$$v_\mathrm{xc}(\mathrm{r}) = \frac{\delta E_\mathrm{xc}[\rho(\mathbf{r})]}{\delta\rho(\mathbf{r})} \tag{2.22}$$

Since the electron density is now expressed in a determinantal wavefunction, they can be solved iteratively until self-consistency with the Hartree-Fock algorithm.

The wavefunction is expressed as a sum of basis functions. In this thesis, the linear combination of atomic orbital ansatz is used for DFT calculations on molecules and the Pople 6-311G** set of Gaussian basis functions is used for first-row elements [239]. For Cl and S, the (negative ion) McLean-Chandler basis set was used [240].

For DFT calculations on periodic systems, such as crystals, a basis set of plane waves is often used, see below.

2.4.2 Exchange-correlation functionals

The exact *exchange-correlation* functional is unknown and we must choose an appropriate function in its place. It is generally assumed that the exchange and the correlation are also additive, $E_{\mathrm{xc}} = E_{\mathrm{x}} + E_{\mathrm{c}}$, and that these functionals can be determined separately.

Early attempts were functions of the local scalar charge density only. Such Local Density Approximations (LDA) are however not accurate enough. Instead, the exchange and correlation functionals are approximated with functions of both the local charge density and its local gradient. DFT methods based on this *generalized gradient approximation* (GGA) are commonly used for periodic calculations on the organic solid state, the PBE-functional by Perdew, Becke and Ernzerhof [241] being particularly popular.

Another, more costly, alternative is to calculate the electron exchange at the Hartree-Fock level of theory, using the same orbitals as in the DFT-calculation [242]. A fraction of E_{xc} is then replaced with the exchange calculated with Hartree-Fock. In this thesis, we will make extensive use of one such hybrid functional, B3LYP [243, 244, 245].

In hybrid functionals, a GGA-functional is combined with a fixed fraction of Hartree-Fock exchange E_{x}^{HF}. In 1993, Becke published the B3LYP hybrid functional [244]. The B3LYP exchange-correlation energy E_{xc}^{B3LYP} is a combination of Hartree-Fock exchange, Becke 1988 (B88) exchange E_{x}^{B88} [246], the Lee-Yang-Parr (LYP) correlation E_{c}^{LYP} [243]. In addition, some Slater exchange E_{x}^{S} and Vosko-Wilk-Nusair correlation E_{c}^{VWN} is included.

$$E_{\mathrm{xc}}^{B3LYP} = a_0 E_{\mathrm{x}}^{HF} + (1 - a_0) E_{\mathrm{x}}^{S} + a_{\mathrm{x}} E_{\mathrm{x}}^{B88} + (1 - a_{\mathrm{c}}) E_{\mathrm{c}}^{VWN} + a_{\mathrm{c}} E_{c}^{LYP} \tag{2.23}$$

where the three fitted parameters a_0, a_{x}, and a_{c} are 0.20, 0.72, and 0.81, respectively.

B3LYP has been the most popular DFT functional for a long time because it gives accurate results for a wide variety of chemical properties. It should however be noted that hybrid functionals become impractical and computationally very costly when plane waves are used as basis set, making their use in periodic systems less common.

In this thesis, the GGA-type PBE functional is used on periodic systems. The hybrid B3LYP functional is used on individual molecules, while BHandHLYP [246, 244] is used on small clusters of molecules. BhandHLYP has $a_0 = 0.5$ and so contains a larger fraction of Hartree-Fock exchange than B3LYP. Alternative methods, used by other research groups, are considered and compared in Chapter 3.

Exchange-correlation functionals based on the local density and gradient fail to account for long-range electron correlation which causes attractive dispersion interactions between atoms. Hence, density functional theory in the GGA approximation, including hybrid functionals, underestimate the dispersion.

The missing dispersion causes quite significant errors in calculations on condensed phases and cannot be ignored. Several methods have been proposed to correct for this. The earliest were empirically parametrized force fields to be added to the density functional energy [97, 98]. Recently, sophisticated algorithms for calculating the missing dispersion have been published [99, 100]. Some of these are described below.

2.4.3 Dispersion corrections

For individual small molecules, the density functionals described here are accurate enough, provided a large enough basis set is used. Problems arise for large molecules and condensed phases where long range dispersion is important. For crystal structures this error is so large that density functional approximations are not useful unless additional measures are taken add the missing dispersion.

A simple way to correct for the missing dispersion in DFT is to add a force field type dispersion term to the energy. An atom-atom pairwise potential of the $C_6 r^{-6}$ type is added to the density functional [97], yielding a 'DFT-D' functional. This method has become popular since Grimme published suitable parameters to be used for GGA density functionals in 2006 [98]. Grimme has since improved the dispersion model and the newer methods are often denoted 'D2' and 'D3'.

The D3 dispersion model is not empirical, but the dispersion parameters are calculated on the fly from atomic polarizabilities, using a partitioning of the electron density between atoms [247]. A short range damping of the dispersion has also been added [99].

Tkatchenko and Sheffler have also suggested a dispersion correction, 'TS', that has become popular [248]. Similar to the D3 correction, it contains a damped $C_6 r^{-6}$ term that is calculated on the fly from atomic polarizabilities. The TS correction is known for reproducing experimental crystal structures quite accurately [249, 115].

Marom *et al.* also considered many-body dispersion and developed the 'MBD' model for dispersion correction [100]. The derivation of the full many-body interaction is beyond the scope of this thesis [250, 251].

In 2012, Otero-de-la-Roza & Johnson suggested a dispersion correction for solid state DFT calculations based on exchange-hole dipoles [252]. An exchange hole is the reduction in probability of finding an electron of the same spin at a position r under the assumption of the presence of a reference electron at a position r'. The exchange hole

and the electron form a dipole, which is assumed to point in the direction of the nearest nucleus. From the exchange hole dipoles, (expectation values of) atomic multipoles can be calculated. Together with a Hirschfeld partitioning of the electron density and atomic polarizabilities, C_6, C_8 and C_{10} dispersion coefficients can be calculated. The XDM method in principle also allows the calculation of Axilrod-Teller-Muto three-body terms, but because of difficulties in formulating a damping functions this is not done in generally available implementations [253].

The accuracy of DFT functionals with the D2, D3, TS, MBD and XDM dispersion corrections have all been investigated, using a set of 23 experimental crystal structures and sublimation enthalpies as benchmark. We will review these benchmark studies in Chapter 3.

2.5 Elastic properties

2.5.1 Theory of elasticity in anisotropic materials

The theory of elasticity in crystalline materials has been described elsewhere [254]. Here, a simple introduction is given in order for the reader to appreciate the contents of this thesis.

Imagine a solid continuum and some external forces acting on it. The forces will lead to deformations of the material, while the internal forces between atoms and molecules in the material will resist such deformations. Here we will assume that we are dealing with such small forces and deformations that we can consider the material to be perfectly elastic and that Hook's law of linearity between stress and strain applies.

At any point P within the continuum, imagine an infinitesimally small cube. The stress at point P depends on the force acting on P and the direction of that force relative to some reference coordinate system. The force vector is projected onto the sides of the cube, resulting in components perpendicular to each side. The stress components can be written σ_{ij} where i and j are sides of the cube, i the side on which the force is acting and j the side that the component is directed towards. Hence, when i and j are unequal, we are dealing with a shearing component of the force. Naturally, $\sigma_{ij} = \sigma_{ji}$ (unless the crystal is accelerating!) and the stress at a point can be completely described by

$$\sigma_{ij} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix} \tag{2.24}$$

The reader should now be in a position to accept that the strain can be described by a similar tensor ε_{kl}. Hook's law may then be written

$$\sigma_{ij} = C_{ijkl}\varepsilon_{kl} \tag{2.25}$$

where C_{ijkl} is the stiffness tensor, containing all elastic constants c_{ij}. The stiffness tensor has $3 \times 3 \times 3 \times 3 = 81$ elements, but a maximum of 36 are independent because of symmetry. Using Voigt notation, these 36 elements can be written as a 6×6 matrix. The inverse of the elastic stiffness tensor is the compliance tensor, denoted S_{ijkl}.

We are interested in elastic properties of crystals, because of the relationships between elastic stiffness, the propagation of acoustic waves and phonon dispersion; that is the change in phonon frequency with respect to wavevector, see below. We will utilize these relationships in order to improve the convergence of lattice dynamical calculations.

The relationship between the elastic stiffness tensor and the propagation of sound in an anisotropic medium is described by the elastodynamic equation and the Christoffel equation [255]. The elastodynamic wave-equation is a system of three partial differential equations.

$$\rho\frac{\partial^2 u_i}{\partial t^2} - C_{ijkl}\frac{\partial^2 u_k}{\partial x_j \partial x_l} = 0, \quad k \in \{1,2,3\} \tag{2.26}$$

Here, ρ is the density of the medium, u_i is the displacement vector, t is time and x_i are the Cartesian coordinates and summation over repeated indices is implied. The solutions are in the form of plane waves of the form

$$u_k = U_k \exp\Big(i\omega\Big(\frac{n_j x_j}{v} - t\Big)\Big), \quad k \in \{1,2,3\} \tag{2.27}$$

Substituting this into the elastodynamic equations leads to a system of three equations which can be written as a matrix equation called the Christoffel equation [256].

$$\begin{pmatrix} G_{11} - \rho v^2 & G_{12} & G_{13} \\ G_{21} & G_{22} - \rho v^2 & G_{23} \\ G_{31} & G_{32} & G_{33} - \rho v^2 \end{pmatrix} \begin{pmatrix} U_1 \\ U_2 \\ U_3 \end{pmatrix} = \mathbf{0} \tag{2.28}$$

Here ρ is the density and v is the phase velocity of sound. The elements of the Christoffel equation are $G_{ij} = C_{ijkl}n_j n_i$ where $n_j n_i$ is a direction cosine of unit length describing the direction of wave propagation. The Cristoffel equation is solved as an eigenvalue equation. The eigenvalues are guaranteed to be positive.

$$\left|\mathbf{\Gamma} - \rho v^2 \mathrm{I}\right| = 0 \tag{2.29}$$

Averages of elastic constants are used to estimate the elastic properties of polycrystalline aggregates, and are important for the tabletability of pharmaceuticals *etc* [27].

The Voigt average bulk modulus is

$$K_V = (c_{11} + c_{22} + c_{33}) + 2\,(c_{12} + c_{23} + c_{31})\,/9 \tag{2.30}$$

and the Reuss average is

$$K_R = (s_{11} + s_{22} + s_{33}) + 2\,(s_{12} + s_{23} + s_{31})^{-1} \tag{2.31}$$

Similarly, the Voigt and Reuss shear moduli are:

$$G_V = (c_{11} + c_{22} + c_{33}) - (c_{12} + c_{23} + c_{31}) + 3\,(c_{44} + c_{55} + c_{66})\,/15 \tag{2.32}$$

$$G_R = (4\,(s_{11} + s_{22} + s_{33}) - 4\,(s_{12} + s_{23} + s_{31}) + 3\,(s_{44} + s_{55} + s_{66})\,/15)^{-1} \tag{2.33}$$

The Voigt- and Reuss averages are upper and lower estimates of the properties of a powder [257]. A more accurate estimate is the Hill average [258], which is the arithmetic average of the Reuss and Voigt estimates, respectively.

2.5.2 Previous work on elastic properties of crystals

Derivations of expressions for elastic constants in the rigid-molecule atom-atom force field formalism was described by Walmsley and used to calculate elastic constants of solid carbon dioxide [259]. The method uses second-derivatives of the potential (lattice) energy with respect to deformations of the unit cell to calculate elements of the elastic tensor.

This method was applied to a crystal structure of hexakis(phenylthio)benzene, a clathrate host structure. A Buckingham-type force field was used, without explicit electrostatic interactions. The velocities of the acoustic waves were used to derive the full set of elastic constants and the bulk modulus [260].

The same method was later implemented in the DMAREL program, later DMACRYS and used to calculate elastic properties of six molecular crystals [261]. Comparisons to experimental reference data showed that the calculated stiffness of the crystal structure depends primarily on close contacts and that therefore, a force field with an accurate model for the repulsion and dispersion interactions is important. It was also shown that the calculated stiffness was overestimated, because of the neglect of thermal vibrations and thermal expansion.

A difficulty in estimating the accuracy of elastic property calculations is that there is a lack of high-accuracy experimental data. The most common experimental method is Brillouin scattering, that relies on the scattering of optic photons by phonons [262].

Calculations of elastic tensors can also be done by periodic DFT [263]. A large set of elastic properties of inorganic crystals have been computed with GGA DFT [257]. Several DFT methods for calculating elastic properties have recently been benchmarked [264].

In Chapter 3 we will calculate the speed of sound and use it to account for phonon dispersion using a model proposed by Debye in 1912 [265], thereby improving the convergence of free energy calculations. Reuss-Voigt-Hill bulk and shear moduli and their temperature-dependence are calculated for 864 crystal structures in Chapter 5, forming the first database of elastic properties for organic crystals.

2.6 Lattice Dynamics

2.6.1 Background to lattice dynamics

In real crystals, the atoms constantly vibrate. Even at 0 K, the atoms vibrate, leading to a zero-point vibrational energy. With increasing temperature, the vibrational modes are increasingly populated, according to standard statistical mechanics.

In this thesis, we will use a special formulation of lattice dynamics, in which the whole molecules are treated as vibrating rigid bodies. The main reason for this rigid-body approach is that it is well-suited for the calculation and study of low-frequency lattice vibrations. Since covalent intramolecular bonds are much stronger than the weak intermolecular interactions, the intramolecular vibrations are much higher in frequency. In order to estimate entropy- and free energy differences between polymorphs, the low-frequency lattice modes are the most important. Another reason is that the rigid-body approach allows us to perform the lattice dynamics calculations with a multipole-based force field, which should greatly enhance the accuracy [266, 2].

The theory of lattice dynamics was comprehensively described by Born & Huang [111] in 1954. Their book is still used as an important reference work. The mathematical notation is a bit cumbersome, but has become somewhat of a standard.

The rigid-body lattice dynamics method used in this thesis was introduced by Califano, Walmsley and co-workers in 1975–1981 [267, 268, 269, 270]. The method was applied to calculate the phonon spectrum of an ethanediamine crystal structure [271], which will be revisited in Chapter 5.

Walmsley also described elastic tensor calculations of molecular crystals in the same formalism [259], and we will make use of that too. Warshel described calculations of lattice energies and thermodynamic properties of crystals of alkanes using lattice dynamics in 1970 [207], in a paper that was ahead of its time.

Contributions in this field was also made by Bonadeo *et al.* [112], who derived thermodynamic functions, expressions for the free energy and entropy, from phonon densities of state calculated with atom-pair potentials. Modified expressions that also consider phonon dispersion were derived by Day [114] and will be revisited in this thesis.

More recent, and far more accessible, sources to the theory of lattice dynamics are the review by Fultz [109] and textbooks by Dove [110] and Simon [272].

Lattice dynamical calculations have previously been used to obtain phonon spectra for several polymorphs of carbamazepine. These calculations showed that different polymorphs have clearly distinct spectra, which can also be observed experimentally through terahertz spectroscopy [273, 274]. This implies that polymorphs react differently to temperature, that they may have different entropy and that their relative stability may change with temperature. This is the main motivation for performing lattice dynamics on polymorphic crystals.

2.6.2 Harmonic approximation lattice dynamics

We will now cover the essence of the lattice dynamics method. The idea is to model vibrations in a crystal as a set of harmonic oscillators. In our case, vibrating and rotating rigid molecules. We follow the notation and terminology of Born & Huang and van Eijck [111, 13].

Consider a crystal consisting of a large number $l = 1, 2, 3 \ldots$ of unit cells. Each cell contains a number of molecules k, each with mass m_k. Assuming the crystal is an energy-minimum, the equilibrium position of (the centre of mass of) molecule k in unit cell l is denoted $\mathbf{x}\binom{l}{k}$, where $\mathbf{x} = (x, y, z)$.

The equilibrium energy is the lattice energy E_{latt} and E the instantaneous energy of the perturbed lattice. We start with a Taylor expansion of the lattice energy around the equilibrium geometry. The potential energy is smooth and changes harmonically for small displacements $\mathbf{u}_\alpha\binom{l}{k}$. Here $\alpha \in \{1 \ldots 6\}$ is the vibrational degree of freedom and runs over three translational and three rotational coordinates for each molecule, unless the molecule is linear or a single atom, for which the number of degrees of freedom are 5 and 3, respectively.

The Taylor expansion of the energy with respect to displacements in the degrees of freedom can be written

$$E = E_{\text{latt}} + \sum_{l,k,\alpha} \frac{\partial E}{\partial \mathbf{u}_\alpha\binom{l}{k}} + \\ + \frac{1}{2} \sum_{\substack{l,k,\alpha,\\ l',k',\alpha'}} \frac{\partial^2 E}{\partial \mathbf{u}_\alpha\binom{l}{k} \partial \mathbf{u}_{\alpha'}\binom{l'}{k'}} + \ldots \tag{2.34}$$

The first-derivative term is exactly zero because there are no forces or torques on the molecules at their equilibrium positions. The Taylor expansion is truncated after the second-derivative term. This is the harmonic approximation, we assume that the vibrations are harmonic and that the higher order terms are small. At low temperatures this is a good approximation, but it fails at high temperatures. We will address this problem below. Born & Huang introduced a (slightly) simpler notation for the force constants Φ.

$$\Phi_{\alpha\alpha'}\begin{pmatrix} l-l' \\ kk' \end{pmatrix} = \left(\frac{\partial^2 E}{\partial \mathbf{u}_\alpha \binom{l}{k} \mathbf{u}_{\alpha'} \binom{l'}{k'}} \right)_0 \tag{2.35}$$

The force constants are obtained by differentiating the force field energy. Thus, the force field is now required not only to have the correct value, but also the correct curvature. Because of this, free energy calculations depend strongly on the accuracy of the force field. This makes temperature-dependent properties more difficult to calculate. It also makes experimentally observed property-temperature dependencies ideal for benchmarking the accuracy of computational methods.

For simple force fields, the second-derivatives can be calculated analytically, but tensor-expressions for the derivatives of the higher multipole interactions require additional work to implement [275]. The calculations are therefore performed numerically using finite displacements [114].

The equations of motion have wave-like solutions.

$$\mathbf{u}\begin{pmatrix} l \\ k \end{pmatrix} = \mathbf{u}\begin{pmatrix} 1 \\ k \end{pmatrix} \exp\left(2\pi i \mathbf{k} \cdot \mathbf{x}\begin{pmatrix} l \\ k \end{pmatrix} - i\omega t \right) \tag{2.36}$$

The dynamical matrix is built up $\mathbf{D}\binom{k}{kk'}$ from its elements, given by

$$\mathbf{D}_{\alpha\alpha'}\begin{pmatrix} k \\ kk' \end{pmatrix} = \frac{1}{\sqrt{m_k m_{k'}}} \sum_l \Phi_{\alpha\alpha'}\begin{pmatrix} l-l' \\ kk' \end{pmatrix} \exp\left(-2\pi i \mathbf{k} \cdot \left(\mathbf{x}\begin{pmatrix} l \\ k \end{pmatrix} - \mathbf{x}\begin{pmatrix} 1 \\ k \end{pmatrix} \right) \right) \tag{2.37}$$

For every vector $\mathbf{k}$, the phonon angular frequency $\omega_{k,\alpha} = 2\pi v$ can be obtained from the eigenvalue equation

$$\left| \mathbf{D}_{\alpha\alpha'}\begin{pmatrix} k \\ kk' \end{pmatrix} - \omega_{k,\alpha}^2 \delta_{kk'} \delta_{\alpha\alpha'} \right| = 0 \tag{2.38}$$

Here δ signify Kronecker deltas.

Hence, there are $n = N_\mathrm{d} Z$ phonons for a crystal with Z molecules per unit cell and N_d degrees of freedom per molecule at every $\mathbf{k}$-point. Three of the eigenvalues will always be zero, corresponding to translations of the whole crystal.

Because of periodic boundary conditions in a primitive Bravais lattice, the $\mathbf{k}$ that are

independent are restricted to one unit cell of the reciprocal lattice, defined by the reciprocal lattice vectors $\mathbf{a}^*, \mathbf{b}^*, \mathbf{c}^*$

$$\mathbf{a}^* = 2\pi \frac{\mathbf{b} \times \mathbf{c}}{\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})} \tag{2.39}$$

$$\mathbf{b}^* = 2\pi \frac{\mathbf{a} \times \mathbf{c}}{\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})} \tag{2.40}$$

$$\mathbf{c}^* = 2\pi \frac{\mathbf{a} \times \mathbf{b}}{\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})} \tag{2.41}$$

The unit cell in real space transforms to a parallelepiped in reciprocal space with $\mathbf{a}^*, \mathbf{b}^*, \mathbf{c}^*$ as lattice vectors. Only $\mathbf{k}$-points in one such reciprocal cell need to be considered. It is conventional to consider the Wigner-Seitz-cell of the reciprocal lattice, which is centred at the origin Γ and is called the (first) Brillouin zone [10].

Point group symmetries in the unit cell can be used so that only one symmetry-independent slice of the Brillouin zone, the *irreducible zone* needs sampling. This is not done in this thesis, but is a quite common method to speed up quantum-chemical calculations in periodic systems [276, 113]. Next, I describe different methods for sampling the Brillouin zone that are used in this thesis.

2.6.3 Brillouin zone sampling

Phonon frequencies vary with wave vector, the dispersion of a phonon $\omega_{i,\mathbf{k}}$ is

$$\tilde{\omega}_{i,\mathbf{k}} = \left(\frac{\partial \omega_i}{\partial \mathbf{k}} \right)_{\mathbf{k}} \tag{2.42}$$

and thus itself a periodic 3-vector in reciprocal space.

The dispersion of phonons is non-trivial. Generally, crystals exhibit a phonon band structure with two separate bands on which phonons live. Phonons that cross the gamma point are known as acoustic phonons, the others as optic phonons, see Fig. 2.1. The phonon frequency γ for several vibrational modes are shown as a function of the vector $\mathbf{k}$. The full 3-dimensional band structure is hard to visualise, and this figure shows a simplified dimensional reduction, where $\mathbf{k}$ has been projected to $|\mathbf{k}|$.

This means that in order to correctly calculate the vibrational energy and entropy we must in principle calculate phonon frequencies at every point in one reciprocal unit cell and evaluate the physical properties as integrals over the phonon spectrum [113]. This is obviously not feasible in practice, and approximations must be made.

Brillouin zone sampling is however not always done and phonons are sometimes calculated only at the gamma point [277, 278]. This can lead to uncertainty about the accuracy and reliability of the obtained results. In the benchmark paper by Otero-de-la-Roza & Johnson, lattice dynamics was used to calculate and subtract thermal

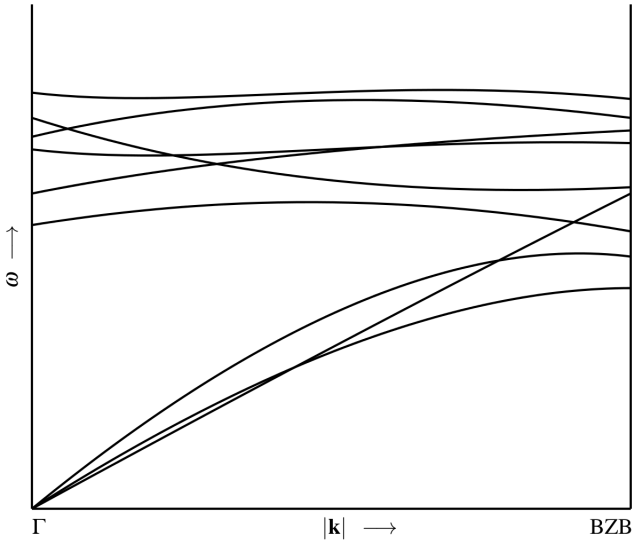


Figure 2.1: Schematic 2-dimensional illustration of a typical phonon band structure in a crystal. Three phonon modes exhibit frequencies ω going to 0 as $|\mathbf{k}| \rightarrow \mathbf{0}$ at the gamma point. Optical modes have non-zero frequencies at the gamma point.

contributions to experimental sublimation enthalpies in other to obtain temperature-free lattice energies for benchmarking purposes [115]. These were soon corrected by more reliable calculations [214].

In one recent example, the phonon **k**-point sampling was mistaken for the Monkhorst-Pack grid of the electronic plane-wave basis set! The similarity in terminology can lead to confusion and incorrect calculations and conclusions. A convergence test where the free energy was converged with respect to the electronic basis may have caused an incorrect conclusion that phonon dispersion is not important for lattice dynamics calculations [278]. Calculations were however also performed on a $1 \times 1 \times 2$ supercell, leading to diverging results and contradicting the conclusion, see the supplementary information to reference [278] for details.

A common 'excuse' for neglecting Brillouin zone sampling is that the applied quantum-chemical electronic structure method is so expensive, in terms of computational effort, that it is only possible to calculate phonons at Γ. As we shall see, the errors caused by neglecting **k**-point sampling in calculating the vibrational energy may be larger than the accuracy in the static lattice energy gained by using such electronic structure methods!

Methods for Brillouin zone sampling are an important part of this thesis and below I describe two old methods suggested by Debye and Einstein as well as a new method based on non-parametric statistics.

A common method for sampling **k**-points is to construct a supercell of the crystal's unit cell. The lattice parameters of the supercell corresponds to a smaller Brillouin

zone. A $k \times l \times m$ supercell samples $N_\mathrm{k} = klm$ **k**-points in a lattice spanning reciprocal space. This method has been used for force field methods [135] and relatively expensive fragment-based QM/MM methods [279, 280].

A Python program was written as a part of this thesis to automatically determine an optimal supercell expansion. The program prepares input files for DMACRYS to perform lattice dynamics calculation with the supercell method. The program is described in Appendix B.

Lattice dynamics calculations on supercells work best with unit cells that do not have very acute or obtuse angles. Suitable 'reduced cells' can be obtained with the Krivý-Gruber algorithm [281, 282], or the ADDSYMM routine in PLATON [283, 284]. Experimental crystal structures rarely require any modification.

Approximate methods for interpolating and extrapolating phonon dispersion were derived by Debye and Einstein [265, 285], these methods are reviewed below. A hybrid method that combines the two have also been used [112, 114].

In 1969, Karo & Hardy used a very simple method to improve the convergence of the phonon density of states for a lattice dynamics calculation on a sodium fluoride crystal. The density of states was first represented by a histogram, but then each bin was replaced by a Gaussian function in order to smooth the spectrum somewhat. This improved the convergence. This method was later elaborated upon by Gilat in several works [286, 287, 288], leading to a method where the local dispersion was calculated from finite differences in phonon frequency between adjacent **k**-points in a lattice. This method, called 'linear approximation' by Gilat, yields a piecewise linear dispersion through each Wigner-Seitz cell in the grid.

2.6.3.1 Debye model

It is the low frequency modes that makes that largest contributions to the entropy and vibrational energy. Because of this, we would like to somehow account for the phonon dispersion of the three acoustic modes.

In 1912 Peter Debye suggested an approximative method for doing so [265]. Debye wanted to calculate the specific heat capacity of solids, which was experimentally known to have a T^3-dependence in the low-temperature limit. In estimating the specific heat capacity of a solid, he assumed that the frequency of acoustic phonons is proportional to the wavelength, *i.e.* that the speed of sound is constant and the phonon dispersion linear. He then calculated the speed of sound and the maximum possible frequency of an acoustic wave, the Debye frequency, from elastic constants. For a rigorous derivation of the Debye model see references [111] and [289].

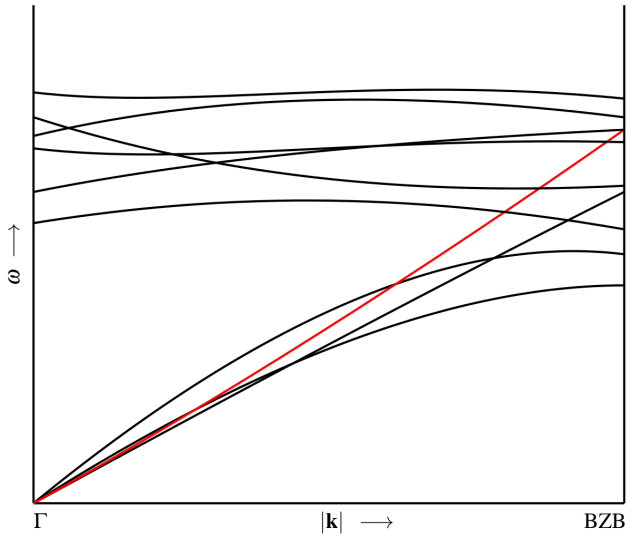


Figure 2.2: Schematic illustration of the Debye approximation. The dispersion of the acoustic modes is modelled as a straight line from 0 at Γ to the Brillouin zone boundary where the frequency is the Debye frequency.

The Debye model leads to the correct temperature dependence of the heat capacity in both the low and high temperature limits, but is inaccurate at intermediate temperatures.

The Debye frequency can be obtained from the (effective) velocity of sound through the crystal, since the group velocity corresponds to $\partial\omega/\partial\mathbf{k}$. In the Debye approximation, where dispersion is assumed to be linear near the Brillouin zone centre, the group and phase velocities are equal.

As can be seen in Figure 2.2, extrapolating all the way to the Brillouin zone boundary can lead to an over-estimation of the acoustic phonon frequency for large $\mathbf{k}$. Thus, a scaling of the Debye frequency with a factor of $2/\pi$ has been suggested [114].

The speed of sound is calculated by solving the Christoffel equation, see above. Since organic crystals typically are highly anisotropic, we calculate the velocity in 13 different directions and average over these to obtain an effective velocity of sound. These directions are vectors pointing from the gamma point to the Brillouin zone boundary towards the centre of 13 neighbouring Brillouin zones. Any unit cell has 26 neighbouring unit cells and correspondingly the first Brillouin zone has 26 adjacent zones. Since the velocity of a phonon with wave vectors $\mathbf{k}$ and $-\mathbf{k}$, only half of them need to be sampled. The vectors are used to calculate 13 different Christoffel matrices which are then diagonalised and the velocities averaged over. The 13 vectors used in this thesis are listed in Table 2.3.

We then calculate direction cosine vectors $\hat{\mathbf{k}}_m$ of these vectors with respect to the reciprocal axis system. When using supercells to sample additional $\mathbf{k}$-points, the reciprocal

Table 2.3: Direction vectors used in calculating the average speed of sound through anisotropic crystals.

$(1/2, 0, 0)$,	$(1/2, 0, -1/2)$,
$(0, 1/2, 0)$,	$(0, 1/2, -1/2)$,
$(0, 0, 1/2)$,	$(1/2, 1/2, 1/2)$,
$(1/2, 1/2, 0)$,	$(1/2, 1/2, -1/2)$,
$(1/2, 0, 1/2)$,	$(1/2, -1/2, 1/2)$,
$(0, 1/2, 1/2)$,	$(-1/2, 1/2, 1/2)$
$(1/2, -1/2, 0)$,	

axes are those of a supercell with the expansion coefficients being those of the longest supercell in each lattice direction.

The 13 direction cosine vectors point from Γ to the surface of an ellipsoid, the anisotropy of which is due to the anisotropy in lattice dimensions and speed of sound. If supercells were used in the calculation, the first three vectors will coincide with the nearest explicitly sampled **k**-points.

2.6.3.2 Einstein model

Einstein introduced a simple model for Brillouin zone integration [285]. He assumed that optical phonons have negligible dispersion, and that the frequency is independent of **k**, see Figure 2.3.

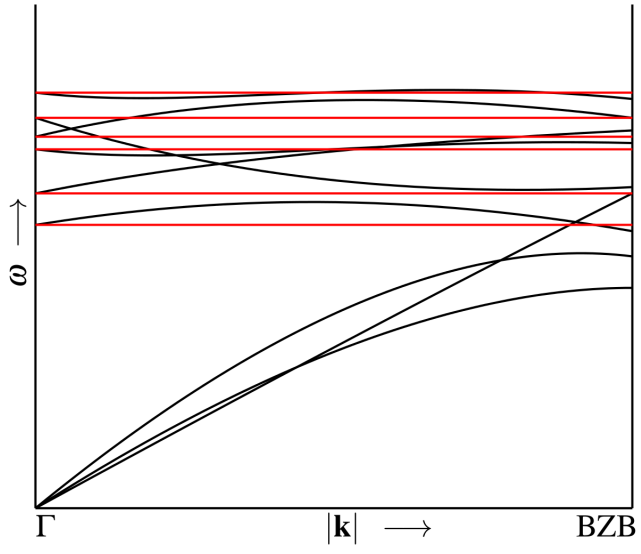


Figure 2.3: The Einstein dispersion approximation. The dispersion of optic modes are assumed to be zero and the phonon frequencies constant throughout the Brillouin zone.

The Einstein model allows an analytic integration over the Brillouin zone, facilitating calculations by hand. The model leads to correct results in the high-temperature limit.

2.6.3.3 Hybrid model

It is possible to apply the Einstein and Debye models simultaneously. The Debye method is used to account for dispersion of acoustic modes, and the Einstein model is applied to the optic phonons. This leads to improved expressions for thermodynamic properties that converge faster [112, 114]. In fact, reasonably accurate free energies can be obtained from $\mathbf{k} = \mathbf{0}$ phonons only, see Figure 2.9. This method was used in the last blind test by the Price group [5].

In this thesis, the KDE method and sampling several **k**-points is proposed instead of the Einstein model, and we will not consider the latter any further.

2.6.3.4 Supercell method

A simple way to sample **k**-points is to perform the lattice dynamics calculation on a supercell of the crystal unit cell. This method is however computationally impractical when supercells are expanded in all three lattice dimensions simultaneously. Such 3-dimensional supercells makes the computational cost increase very quickly, while long wavelength phonons still cannot be calculated. Even force field calculations are limited to supercells with 2 or 3 unit cells per side. This is not enough to sample long wavelength phonons. Another problem is that when performing calculations on different crystal structures, such as a set of polymorphs, the free energy may converge for some structure but not others, making the error in the relative free energy large.

With rectangular supercells, the convergence is stepwise. As an example, consider a calculation of one $2 \times 3 \times 3$ supercell, sampling 18 **k**-points. Making a larger supercell, $3 \times 3 \times 3$, would sample 27 **k**-points. This would lead to quite different results for the calculated thermodynamics properties.

For a set of polymorphs with possibly large differences in unit cell sizes and dimensions, the naive supercell method is not feasible. Here we will instead use a modified method that makes the convergence smoother and allows a more fine-grained choice of the number of sampled **k**-points.

We will use a method where the unit cell is expanded into linear supercells expanded along **a**, **b** or **c** only. The expansion coefficient n along each lattice vector λ is chosen such that the length of $\lambda^*/\mathbf{n}$ is smaller than some target parameter.

Realistic choices of supercell coefficients for linear supercells is 3–10, depending on the unit cell dimensions. Clearly, this allows the sampling of much longer phonon modes. The lattice dynamics calculations on the three supercells can now also be run in parallel, allowing a far more efficient use of modern processor architectures.

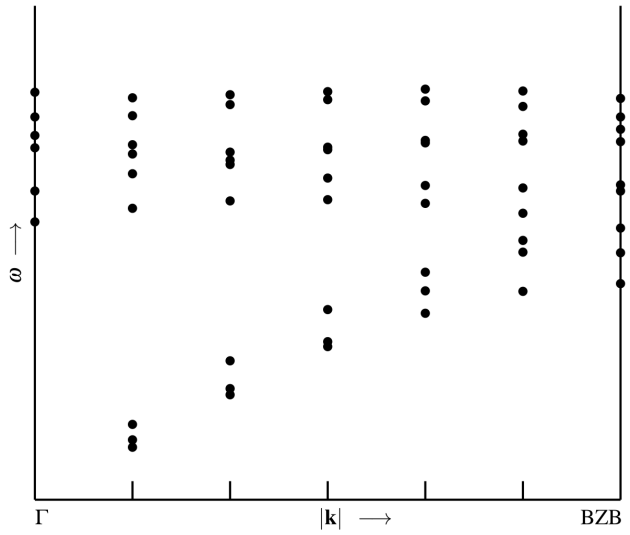


Figure 2.4: Schematic illustration of how a supercell sampling of the Brillouin zone leads to phonon dispersion. Phonon frequencies are calculated on a regularly spaced grid.

The supercell method can however be improved by splitting the three supercells into several smaller supercells. As long as these supercells all have expansion coefficients that are all mutually relatively prime, *i.e.* they have no common divider, they will all sample unique **k**-points (with the exception of $\mathbf{k} = 0$), making the calculations maximally efficient.

One algorithm for selecting split supercells is the following. If $s < 6$ the lattice dynamic calculations is performed on the supercell as is. For $s \geqslant 6$ the supercell is split into n smaller supercells (1x1xk, 1x1xℓ, 1x1xm ...) such that $k, \ell, m...$ are all mutually co-prime and $k+\ell+m-(n-1) \geqslant s$. This ensures that at least the same number of **unique** **k**-points are sampled. The phonons for $\mathbf{k} = \mathbf{0}$ will be calculated in each supercell, and we only include these phonons from one of the split supercells. The long linear supercells are split into 2, 3 or 4 co-prime supercells according to the scheme in Table 2.4. Note that this is by no means the only possible choice, and we make no claim that this is the best or computationally most efficient splitting. Splitting the supercells means that the sampled **k**-points will no longer be equidistantly placed along the reciprocal axes, but this has a very small effect on the results.

The split supercells allows more than 20 **k**-points to be sampled in each direction, but this will only happen for unit cells with one very small lattice parameter. In practical calculations 15–25 **k**-points are quite affordable and anything above a dozen **k**-points leads in practice to well-converged results.

There is often more than one way of choosing the co-prime supercells. In order to make the convergence smooth, the splitting scheme shown in Table 2.4 is proposed. This has

been found to perform well in the calculations described in this thesis.

Table 2.4: Linear supercells were split into 2, 3 or 4 smaller supercells with mutually co-prime expansion coefficients in this way.

2	→	2	15	→	5, 6, 7
3	→	3	16	→	3, 4, 5, 7
4	→	4	17	→	5, 7, 8
5	→	5	18	→	5, 7, 8
6	→	3, 4	19	→	3, 5, 7, 8
7	→	3, 5	20	→	3, 5, 7, 8
8	→	4, 5	21	→	4, 5, 7, 9
9	→	3, 4, 5	22	→	4, 5, 7, 9
10	→	3, 4, 5	23	→	5, 7, 8, 9
11	→	3, 4, 7	24	→	5, 7, 8, 9
12	→	3, 4, 7	25	→	5, 7, 8, 9
13	→	3, 5, 7	26	→	5, 7, 8, 9
14	→	4, 5, 7			

2.7 Kernel density estimation

2.7.1 Fundamentals of kernel densities

Kernel density estimation is a common non-parametric statistical method for estimating an unknown probability density from a sample. The method was introduced by Parzen [290] and Rosenblatt [291] and has been very actively discussed in the statistics community, as documented in several reviews and books [292, 293, 294].

Here we will only consider the univariate case. Given a random sample of n independent observations $X_i \in \mathbb{R}$ drawn from an unknown continuous probability distribution, the kernel density estimate of the population distribution is:

$$KDE(x,h) = \frac{1}{nh}\sum_{i}^{n} K\Big(\frac{x - X_i}{h}\Big) \tag{2.43}$$

Here $K(\cdot)$ is a kernel function, which generally has the following properties:

- Positive: $K(x) \geq 0, \quad x \in \mathbb{R}$
- Real: $x \in \mathbb{R} \rightarrow K(x) \in \mathbb{R}$
- Normal: $\int K(x)\, dx = 1$
- Symmetric: $K(x) = K(-x)$

The bandwidth h is the width of the kernel function and determines the level of smoothing applied to the sample.

The choice of kernel usually has a limited effect on the accuracy of the estimate. The bandwidth is more important, but also much more difficult to choose. The subject of optimal kernel and bandwidth selection has received considerable attention in the statistics literature [295, 296, 297, 298].

2.7.2 Choice of kernel

Histograms are commonly used to represent phonon densities of states. Note that a histogram can be written as a kernel density with a top hat function as kernel. A kernel density is a natural generalization of the histogram method.

The Gaussian kernel is probably the most commonly used, since it is well-known and computationally accessible. A simple and computationally convenient way to smooth the phonon density of states is to replace each histogram bin with a Gaussian function, and this has been used in real lattice dynamics applications [299, 287]. A true kernel density however places one kernel function at every individual observation.

I have used Gaussian kernel densities to show that it does in fact improve the convergence properties of quasi-harmonic lattice dynamics calculations [2]. However, the Gaussian kernel is hard to justify scientifically. We would like the shape of the kernel function to be justified by the physics of the system.

As described above, it is necessary to sample several **k**-points. The most common approach is to sample a small number of **k**-points in a regular grid in the Brillouin zone, and we will assume that a method of that kind is used.

Around each sampled **k**-point, we would like a Debye-like extrapolation of phonon frequencies. Following Taylor, the phonon dispersion can be assumed to be linear within a small neighbourhood around each **k**-point. Thus, we imagine a *Debye sphere* centred at each sampled **k**-point, and that each sphere have a constant phonon frequency gradient through it.

To derive the kernel needed to model this Debye-like dispersion, we begin by considering the humble unit circle.

$$x^2 + y^2 = 1, \tag{2.44}$$

or equivalently;

$$y = \pm\sqrt{1 - x^2}, \quad -1 \leq x \leq 1 \tag{2.45}$$

Rotating y around the x-axis produces a unit sphere in 3 dimensions. The area of a cross-section through the unit sphere through x and parallel to the yz-plane is πy^2

$$A(x) = \pi(1 - x^2), \quad -1 \leq x \leq 1 \tag{2.46}$$

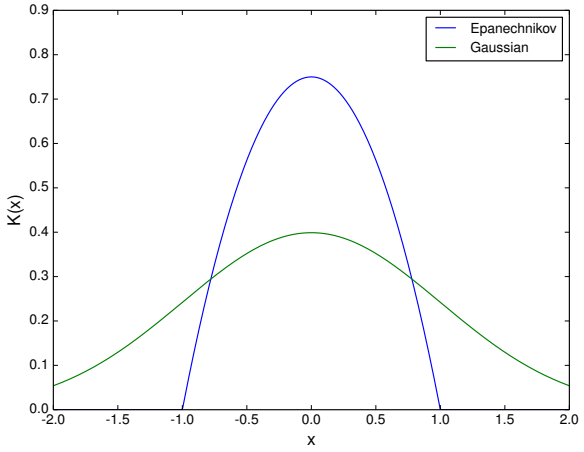


Figure 2.5: The Epanechnikov and Gaussian kernels plotted with unit bandwidth.

Equation 2.46 will form the basis of our kernel, it is the distribution that results from a linear gradient through a sphere, regardless of the direction of the gradient. The spheres may be rotated, and this will not affect the shape of the kernel. We say the kernel is rotatable.

Equation 2.46 does not obey the requirements for a kernel stated above. The function is not positive for all x, and its integral is not normal. We solve this by introducing the indicator function $\mathrm{I}(x)$ defined by

$$\mathrm{I}(x) = \begin{cases} 0, & \text{if } |x| \geq 1 \\ 1, & \text{if } |x| \leq 1, \end{cases}$$

By normalizing the integral we obtain the proper kernel function

$$K(x) = \frac{3}{4}(1 - x^2)\mathrm{I}(x) \tag{2.47}$$

This kernel is known as the Epanechnikov kernel [295]. In figure 2.5 we can see the shape of the Epanechnikov kernel in comparison to the Gaussian. The Gaussian kernel is rather wide but has the advantage of being everywhere continuous and smooth. The Epanechnikov kernel is continuous but its second derivative is not. It may therefore be more difficult to integrate numerically.

The Epanechnikov kernel was found by Epanechnikov when he in 1969 was searching for a kernel that minimises the mean integrated squared error (MISE) between the kernel density of a random sample and the population probability distribution [295].

Phonon frequencies calculated by lattice dynamics are not independent and not random, but van Eijck sampled **k**-points randomly [300], and in that case, the Epanechnikov

kernel density is the optimal solution and will result in the fastest possible convergence to the true phonon density of states [295, 301].

We place the kernel function centred at each sample value X_i and scale the kernel width with the bandwidth, h. To keep the integral unchanged, we must also divide by the bandwidth. The full expression for the kernel density thus becomes:

$$KDE(x) = \frac{3}{4nh} \sum_{i}^{n} \left(1 - \left(\frac{X_i - x}{h}\right)^2\right) \mathrm{I}\left(\frac{X_i - x}{h}\right) \tag{2.48}$$

Where n is the number of samples X_i and h is the bandwidth, or equivalently, the largest difference in phonon frequency between the centre of the Debye sphere and its surface. Kernel densities are normalised to unity, so phonon densities must be renormalised to $N_\mathrm{d}Z$.

2.7.3 Choice of kernel bandwidth

2.7.3.1 Methods for bandwidth selection

We now need a way to determine the optimal kernel bandwidth. This is a difficult problem that has provoked intense research and debate in the statistical community [302, 301, 303, 296, 298].

A couple of methods for choosing kernel bandwidth are investigated in this thesis. We begin by attempting an automatic optimisation of the bandwidth by least-squares cross-validation. Then a method is proposed that uses real phonon frequencies calculated at the Brillouin zone centre and three boundary points (X, Y, Z for cubic and orthorhombic cells) and average over the phonon dispersion between these points.

2.7.3.2 Least-squares cross-validation

One way to optimise the bandwidth using the phonon frequencies themselves is cross-validation. An implementation of such a method was performed and is described here.

We begin by noting that there are reasons to make the bandwidth dependent on sample size. In general, the bandwidth should obey the following relations:

$$\lim_{n\to\infty} h(n) = 0 \tag{2.49}$$

$$\lim_{n\to\infty} nh(n) = \infty \tag{2.50}$$

Making the bandwidth proportional to the fifth root of the sample size has been shown to be optimal for large samples and asymptotically lead to the exact density [301]. We

will therefore set the bandwidth to $h = c\sqrt[5]{n}$, and try to optimize the bandwidth scale factor c. To optimize c, we minimise the mean integrated squared error MISE(h) by leave-one-out least-squares cross-validation [302, 303, 292].

The method essentially means that for every phonon frequency ω_i, we calculate the KDE using all phonons except ω_i. These KDEs are then compared with the 'true' density from all phonons. The integrated squared errors are calculated and averaged over. This error is then minimised numerically with respect to the bandwidth scaling factor.

Real phonon frequencies were obtained from rigid-molecule lattice dynamics calculations. Four crystal structures obtained from the Cambridge Structural Database [304] were used. The four structures are crystals of 2,2'-dipyridylamine (DPYRAM02), acridine (ACRDIN04), theophylline (BAPLOT02) and paracetamol (HXACAN12). The number of sampled **k**-points were 17, 21, 18, and 21, respectively. The number of non-zero phonon frequencies is $n = 6ZN_\mathrm{k} - 3$, or 813, 501, 861 and 501 phonon frequencies, respectively, for the four structures.

Using these crystal structure and their phonon frequencies, optimal kernel bandwidths for the kernel density estimate for the complete phonon density of states were then calculated with cross-validation.

The cross-validation requires a large number of KDEs to be calculated, which is quite expensive. However, by noting that the KDE is a convolution of the data points with the kernel, it is possible to derive a convenient approximate loss function $M_1(h)$ for rapid calculations [301, 292].

$$M_1(h) = \frac{1}{hn^2}\sum_i^n\sum_j^n\left(K^{(2)}\left(\frac{\omega_i-\omega_j}{h}\right)\right) \tag{2.51}$$

$$-\frac{2}{hn^2}\sum_i^n\sum_j^n\left(K\left(\frac{\omega_i-\omega_j}{h}\right)\right) \tag{2.52}$$

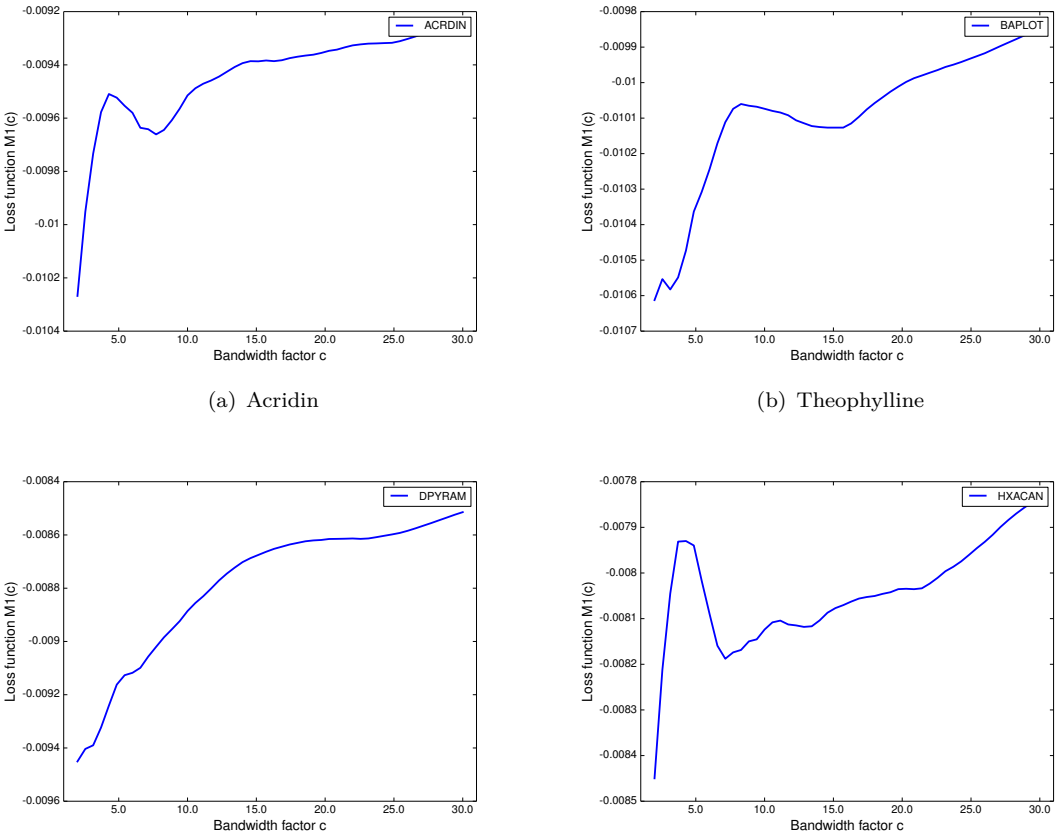
$$+\frac{2K(0)}{hn} \tag{2.53}$$

Here $K^{(2)}(x)$ denotes the convolution of the Epanechnikov kernel with itself, which is:

$$K^{(2)}(x) = \int_{-\infty}^{\infty} K(t)K(x-t)dt \tag{2.54}$$

$$= \begin{cases} -3/160\cdot(x-2)^3(x^2+6x+4), & 2>x>0 \\ 3/160\cdot(x+2)^3(x^2-6x+4), & 0>x>-2 \\ 0, \quad \text{otherwise} \end{cases} \tag{2.55}$$

The minimisation of the M_1 loss function is well-known to be problematic. The function often has several minima and can diverge towards $-\infty$ when the bandwidth approaches



(a) Acridin

(b) Theophylline

(c) dipyridylamine

(d) Paracetamol

Figure 2.6: The M_1 loss as a function of the bandwidth factor c for four crystal structures. Note that the loss function diverges for small bandwidths and has a singularity at $h = 0$.

zero. Numerical minimisation by Newton-Raphson methods is risky and should not be applied without careful monitoring.

In Figure 2.6 the M_1 loss functions for the four crystal structures are plotted as a function of the bandwidth factor. The loss function unfortunately diverges for small bandwidths and has a singularity at $h = 0$. In addition, there are often several and very shallow minima. This is why the function is so difficult to minimize.

Computer simulations of the loss function for randomly generated subsets of the complete phonon spectra were also performed. These simulations showed that the loss function is very sensitive to small changes in the data set (results not shown). Because of the large sensitivity of the loss function to random perturbations in the underlying data, and because of the practical difficulties in the automatic optimisation of the bandwidth, the author cannot recommend that such methods be used for the purpose of lattice dynamics calculations.

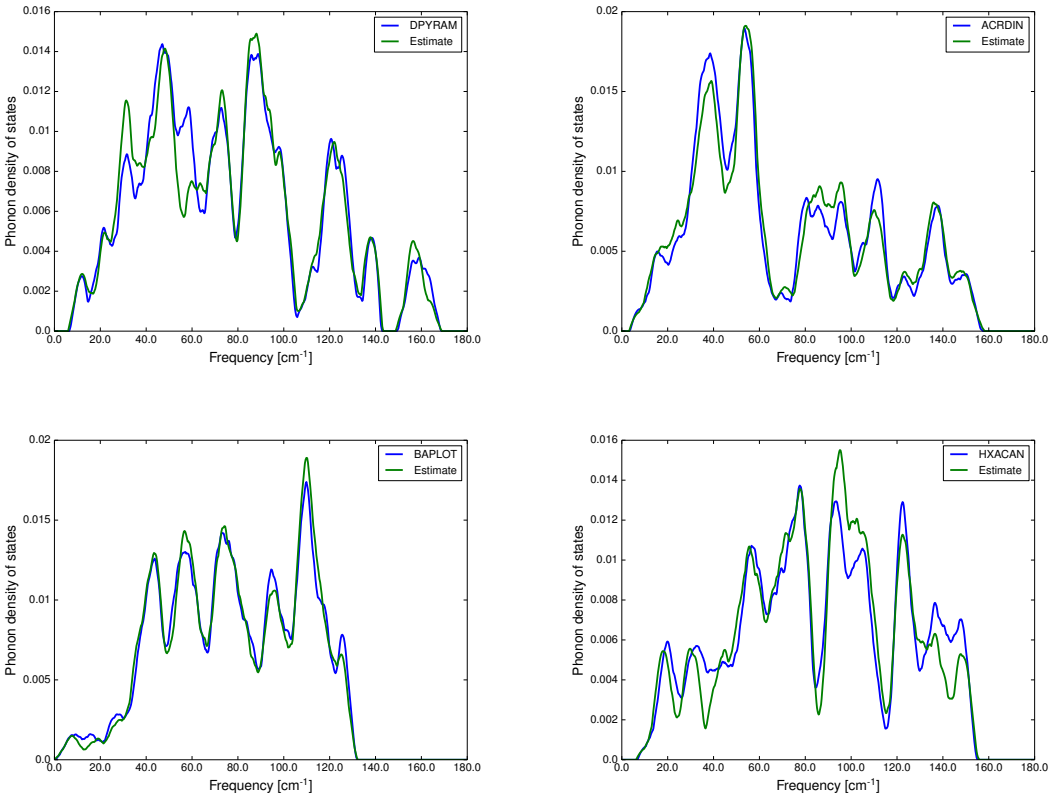


Figure 2.7: Phonon densities of states (blue curves) of four crystal structures calculated as Epanechnikov kernel densities of all phonons using a bandwidth factor $c = 15$, and kernel estimates (green curves) using only half of the data points.

2.7.3.3 Rule-of-thumb bandwidth

It has been suggested [292] that the optimal c should be in the interval

$$\frac{1}{4}\sigma < c_{opt} < \frac{3}{4}\sigma \tag{2.56}$$

where σ is the sample standard deviation. This is however a pretty wide interval and of little guidance for our calculations. The simulation results indicate that the loss function usually has a shallow minimum between $5 < c < 20$, but as seen above, this is not always so.

However, a rule-of-thumb bandwidth factor of around 15 seems quite reasonable. This gives good results in the four cases described above. In fact, using this bandwidth, the complete phonon density of states can be reproduced reasonably well with as little as 50 % of the calculated phonon frequencies. In Fig 2.7 the phonon densities of states are shown calculated from all phonons together with estimates calculated from a random subset of phonons. The random sample contains only half as many data points as the full densities. Despite using so few phonons, the full densities are estimated remarkably

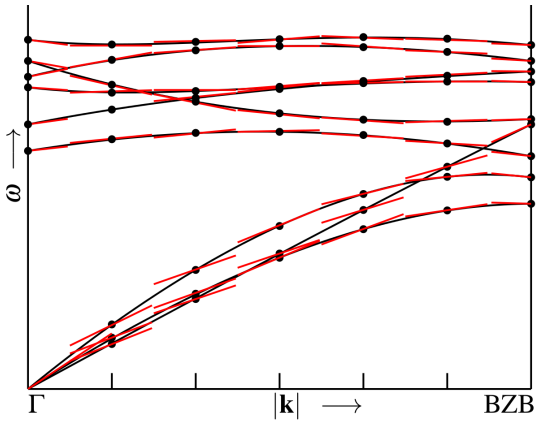


Figure 2.8: Illustration of how a combined use of the supercell, Debye and Epanechnikov KDE together models the phonon dispersion as piecewise linear. The agreement between the true (black) dispersion and the model (red) is now excellent.

well when using a bandwidth factor of 15. This convincingly demonstrates that kernel densities help converge phonon spectra.

In Figure 2.8 I illustrate how the KDE method complements the supercell sampling of the Brillouin zone. Each phonon at every **k**-point is extrapolated in a piecewise linear manner.

2.7.3.4 Mean dispersion of optical phonons

Since the least-squares cross-validation method is often numerically cumbersome, it may instead be preferable to calculate a mean dispersion of optical phonons and use this to determine a suitable bandwidth.

It is computationally trivial to calculate optical phonon frequencies at the Γ-point and at the the Brillouin zone boundary points X, Y, Z. Calculating phonons for $1\times1\times2$, $1\times2\times1$ and $2\times1\times1$ supercells yields phonons at the boundary points and we can calculate a mean dispersion in each principal direction $\lambda \in \{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$ from finite differences between the boundary and Γ.

$$\tilde{\omega}_\lambda = \frac{1}{n}\sum_i^n \frac{|\omega_{i,\lambda} - \omega_{i,\Gamma}|}{|\lambda^*|} \quad [\mathrm{cm}^{-1}/\mathrm{\AA}^{-1}] \tag{2.57}$$

This means that we assume the dispersion is linear, but not constant between the Brillouin zone centre and boundary.

Since this does not take the direction of the dispersion gradient into account it will underestimate the total rate of dispersion. In other words, we assume the phonon gradient is aligned along the principal direction. To 'rotate the spheres' we use the

Euclidean distance.

$$\tilde{\omega} = \frac{1}{\sqrt{3}} \left(\sum_{\lambda} \tilde{\omega}_{\lambda}^2 \right)^{1/2} \tag{2.58}$$

This gives a total mean rate of dispersion throughout the Brillouin zone.

A suitable kernel bandwidth can then be determined from the mean dispersion and the average distance L between sampled **k**-points.

$$h = \tilde{\omega} L/2 \quad (\mathrm{cm}^{-1}) \tag{2.59}$$

This directly gives the bandwidth, but is based on a rather crude approximation of a single average dispersion for each phonon and throughout the Brillouin zone. We assume that the dispersion can be described by a single average for all vibrational modes.

A natural extension is to replace averaging over the three principal directions with constructing the total phonon density as a sum of three kernel densities, one for each principal direction and using different bandwidths for each. Even better would be to use one bandwidth for every phonon mode. This should better capture the fact that some phonons, the acoustic ones, have a larger dispersion than the optic ones.

2.7.4 The bounded interval problem

The kernel density is defined on the interval $(-\infty, +\infty)$, but phonon frequencies cannot be negative, and the density should vanish at zero frequency:

$$\lim_{x \to 0} \mathrm{KDE(x)} = 0 \tag{2.60}$$

We need the density to be bounded on the interval $[0, +\infty)$. This can be accomplished, for instance, by transformation of the phonon frequencies by a logarithm to yield a bounded KDE, KDE_b.

$$KDE_b(x) = \frac{1}{x} KDE(\log x) \tag{2.61}$$

However, several of the methods for creating bounded kernel densities impose the condition that the *derivative* of the kernel density goes to zero [292], rather than the density itself. A method that respects the necessary boundary condition in Eq. 2.60 is the *negative reflection* [292]. The negative reflection method results in this estimate:

$$KDE_b(x) = \frac{1}{nh} \sum_{i=1}^{n} \left(K\left(\frac{x - X_i}{h} \right) - K\left(\frac{x + X_i}{h} \right) \right) \tag{2.62}$$

This may be useful if the Gaussian or other heavy-tailed kernels are used. For the Epanechnikov kernel, it should not be necessary, since phonon frequencies rarely fall below the bandwidth. Note that Eq. 2.62 is not guaranteed to integrate to unity and may need to be renormalized. Bounded kernel densities have not been used in this thesis.

I have now derived a method for kernel density estimation of phonon densities of states that is independent of empirical parameters. This results in an enhanced Brillouin zone integration that is justified by physics. Under the condition that phonon frequencies can be considered a random sample of the density of states, the Epanechnikov kernel density is an unbiased and maximally efficient estimate of the phonon density of states.

The KDE method with the Epanechnikov kernel is based on the assumption of a piecewise linear phonon dispersion throughout the Brillouin zone. This is a natural extension of Debye's method to arbitrary **k**-points. It is an obvious improvement over Einstein's assumption of constant dispersion.

It is possible, but challenging, to optimise the kernel bandwidth with cross-validation methods. However, it is found that the optimal bandwidth is typically close to $15n^{-1/5}$ for crystals of rigid molecules. This can be used as a rule of thumb.

An alternative method to determine a suitable bandwidth is to calculate a mean dispersion rate of optical phonons from a finite difference between phonons at the Brillouin zone centre and boundaries only. This gives a non-empirical method.

By using a kernel density estimate, it is possible to obtain a good approximation of a very well-converged phonon density using a much smaller number of explicitly calculated phonons. In addition, this method makes the convergence of the density and calculated thermodynamic properties smooth. This means that considerable savings in terms of time and computational cost can be achieved with these methods. This should be particularly useful for quantum-chemical methods.

2.8 Thermodynamic expressions

2.8.1 From phonons to energies

Thermodynamic properties can be calculated from the phonon densities of states. Here we present expressions for calculating the vibrational energy, entropy and heat capacity from phonon frequencies when the different methods for accounting for phonon dispersion are used. These are based on derivations by Born & Huang, Bonadeo and Day [111, 112, 114]. Expressions for the KDE approximation have been published as supplementary information to reference [2].

2.8.2 Conventional expressions

For rigid molecules with 6 degrees of freedom, there are $n = 6ZN_\mathrm{k}$ phonon frequencies ω_i. Here Z is the number of molecules per unit cell. The Boltzmann constant is denoted k_B and $\hbar$ is the reduced Planck's constant. The number of sampled **k**-points is denoted N_k.

The following derivation is based on standard textbook statistical mechanics, later we will modify these to include the phonon dispersion from the Debye and KDE methods.

Phonons are bosons and the partition function of a macroscopic crystal at low temperatures, such as those below the melting point of organic crystals is [114, 111]

$$Z_{\text{vib}} = \prod_{i,\mathbf{k}} \frac{\exp\left(\frac{-\hbar\omega_{i,\mathbf{k}}}{2k_BT}\right)}{1-\exp\left(\frac{-\hbar\omega_{i,\mathbf{k}}}{k_BT}\right)} \tag{2.63}$$

The product is over all vibrational modes i and all sampled points in reciprocal space $\mathbf{k}$.

The vibrational energy contribution to the free energy F_{vib} is then

$$F_{\text{vib}}(T) = -\frac{k_BT}{N_{\text{k}}} \ln Z_{\text{vib}} \tag{2.64}$$

which can be calculated directly as

$$F_{\text{vib}}(T) = \frac{1}{2N_{\text{k}}} \sum_{i,\mathbf{k}} \hbar\omega_{i,\mathbf{k}} + \frac{k_BT}{N_{\text{k}}} \sum_{i,\mathbf{k}} \ln\left(1-\exp\left(\frac{-\hbar\omega_{i,\mathbf{k}}}{k_BT}\right)\right). \tag{2.65}$$

The first term at the right hand side is the vibrational zero point energy ZPE, which is completely temperature-independent in the harmonic approximation.

An alternative expression for the vibrational energy that is often seen in the literature is

$$F_{\text{vib}} = \frac{k_BT}{N_{\text{k}}} \sum_{i,\mathbf{k}} \ln\left(2\sinh\left(\frac{\hbar\omega_{i,\mathbf{k}}}{2k_BT}\right)\right) \tag{2.66}$$

We will not entertain this form. The Helmholtz lattice free energy A is obtained as

$$A(T) = E_{\text{latt}} + F_{\text{vib}}(T) \tag{2.67}$$

The Gibbs free energy is recovered later, in the quasi-harmonic approximation, where thermal expansion and a PV-term is added, see below.

The lattice-vibrational entropy (there are additional contributions to the entropy from disorder, *etc*, which are not considered here) is the (negative) derivative of the free energy with respect to temperature, which is

$$\begin{aligned} S(T) = & -\frac{k_BT}{N_{\text{k}}} \sum_{i,\mathbf{k}} \ln\left(1-\exp\left(\frac{-\hbar\omega_{i,\mathbf{k}}}{k_BT}\right)\right) \\ & + \frac{1}{N_{\text{k}}T} \sum_{i,\mathbf{k}} \hbar\omega_{i,\mathbf{k}} \left(\frac{\exp\left(\frac{-\hbar\omega_{i,\mathbf{k}}}{k_BT}\right)}{1-\exp\left(\frac{-\hbar\omega_{i,\mathbf{k}}}{k_BT}\right)}\right) \end{aligned} \tag{2.68}$$

The heat capacity is calculated as:

$$C_{\mathrm{v}}(T) = \frac{k_B}{N_{\mathrm{k}}} \sum_{i,\mathbf{k}} \left(\frac{\hbar\omega_{i,\mathbf{k}}}{k_B T}\right)^2 \frac{\exp\left(\frac{\hbar\omega_{i,\mathbf{k}}}{k_B T}\right)}{\left(\exp\left(\frac{\hbar\omega_{i,\mathbf{k}}}{k_B T}\right) - 1\right)^2} \tag{2.69}$$

2.8.3 Expressions for continuous phonon densities

If the entire Brillouin zone were to be sampled, we would obtain an essentially continuous phonon density of states $g(\omega)$, such that $g(\omega)d\omega$ is the number of phonon modes in the frequency range from ω to $\omega + d\omega$. The density is normalised such that

$$\int_0^\infty g(\omega)\, d\omega = N_{\mathrm{d}} Z \tag{2.70}$$

The sums in the expressions above become integrals over $g(\omega)$. This is also the case for when kernel densities are calculated. The discrete phonon frequencies are replaced by a continuous kernel density function that is used to estimate the phonons density of states $g(\omega) \mathrel{\hat{=}} \mathrm{KDE}(\omega_i)$.

The expression for the vibrational energy is then

$$\begin{aligned} F_{\mathrm{vib}} = &\frac{1}{2}\int_0^\infty \hbar\omega g(\omega)\, d\omega \\ &+ k_B T \int_0^\infty g(\omega) \ln\left(1 - \exp\left(\frac{-\hbar\omega}{k_B T}\right)\right)\, d\omega \end{aligned} \tag{2.71}$$

and the entropy becomes

$$\begin{aligned} S(T) = &-k_B \int_0^\infty g(\omega) \ln\left(1 - \exp\left(\frac{-\hbar\omega}{k_B T}\right)\right)\, d\omega \\ &+ \frac{1}{T}\int_0^\infty \hbar\omega g(\omega) \left(\frac{\exp\left(\frac{-\hbar\omega}{k_B T}\right)}{1 - \exp\left(\frac{-\hbar\omega}{k_B T}\right)}\right)\, d\omega \end{aligned} \tag{2.72}$$

The expression for the heat capacity with continuous phonon densities is

$$C_{\mathrm{v}}(T) = k_B \int_0^\infty \left(\frac{\hbar\omega}{k_B T}\right)^2 \frac{\exp\left(\frac{\hbar\omega}{k_B T}\right) g(\omega)}{\left(\exp\left(\frac{\hbar\omega}{k_B T}\right) - 1\right)^2}\, d\omega \tag{2.73}$$

2.8.4 Expressions for Debye contributions

The Debye model is used to calculate the contributions from acoustic phonons near the Brillouin zone centre. Using the speed of sound, an average phonon frequency on an ellipsoid around the Brillouin zone centre, the Debye frequency ω_{D} is calculated. The

Debye frequency is then used in following expressions to obtain the Debye contributions to the zero-point energy, the thermal contribution to the free energy, the entropy and heat capacity.

When the Debye method is used, contributions from the three acoustic modes that cross the Γ-point are accounted for with these expressions. Hence, these three modes should not be included in the expressions above, which are now considered to account for the optic phonons only. The simplest way to avoid the double-counting is to include a factor of $n/(n+3)$ in the right hand side of the above equations.

The expressions below contain the (third) Debye function $\mathrm{D}(x)$, defined by

$$\mathrm{D}(x) = \frac{3}{x^3}\int_0^x \frac{t^3}{\exp(t)-1}\,dt \tag{2.74}$$

Integrals over the Debye function generally have no known analytical solutions, but must be calculated numerically.

The Debye contribution to the zero point energy is:

$$ZPE_{\mathrm{D}} = \frac{9\hbar\omega_{\mathrm{D}}}{8N_{\mathrm{k}}} \tag{2.75}$$

The Debye contribution to the thermal energy is:

$$\begin{aligned} F_{\mathrm{D}}(T) = {} & \frac{3k_BT}{N_{\mathrm{k}}}\ln\left(1-\exp\left(\frac{-\hbar\omega_{\mathrm{D}}}{k_BT}\right)\right) \\ & - \frac{k_BT}{N_{\mathrm{k}}}\mathrm{D}\left(\frac{\hbar\omega_{\mathrm{D}}}{k_BT}\right) \end{aligned} \tag{2.76}$$

The Debye contribution to the entropy is $S_{\mathrm{D}}(T) = -\partial F_{\mathrm{D}}(T)/\partial T$, which becomes:

$$\begin{aligned} S_{\mathrm{D}}(T) = {} & -\frac{3k_B}{N_{\mathrm{k}}}\ln\left(1-\exp\left(\frac{-\hbar\omega_{\mathrm{D}}}{k_BT}\right)\right) \\ & + \frac{3}{N_{\mathrm{k}}T}\hbar\omega_{\mathrm{D}}\left(\frac{\exp\left(\frac{-\hbar\omega_{\mathrm{D}}}{k_BT}\right)}{1-\exp\left(\frac{-\hbar\omega_{\mathrm{D}}}{k_BT}\right)}\right) \\ & + \frac{k_B}{N_{\mathrm{k}}}\mathrm{D}\left(\frac{\hbar\omega_{\mathrm{D}}}{k_BT}\right) \\ & - \frac{\hbar\omega_{\mathrm{D}}}{N_{\mathrm{k}}T}\mathrm{D}'\left(\frac{\hbar\omega_{\mathrm{D}}}{k_BT}\right) \end{aligned} \tag{2.77}$$

Here $\mathrm{D}'(x)$ is the derivative of the Debye function, which is

$$\mathrm{D}'(x) = \frac{3}{\exp(x)-1} - \frac{1}{x^4}\int_0^x \frac{t^3}{\exp(t)-1}\,dt \tag{2.78}$$

The Debye contribution to the heat capacity is

$$C_{\mathrm{D}}(T) = \frac{12k_B}{N_{\mathrm{k}}}\mathrm{D}\left(\frac{\hbar\omega_{\mathrm{D}}}{k_BT}\right) - \frac{9\hbar\omega_{\mathrm{D}}}{N_{\mathrm{k}}T}\left(\exp\left(\frac{\hbar\omega_D}{k_BT}\right) - 1\right)^{-1} \tag{2.79}$$

These contributions from the acoustic modes should be added to the above standard expressions with either discrete phonons or continuous phonon densities, after they have been scaled by $n/(n+3)$, where n is the number of phonons.

For atomic calculations where the atoms have three translational degrees of freedom, the mean squared isotropic displacement u^2 of an atom from its equilibrium position is [305]

$$u^2 = \frac{3\hbar}{2M\pi}\int_0^\infty \left(\frac{1}{\exp\left(\frac{\hbar\omega}{k_BT}\right) - 1} + \frac{1}{2}\right)\frac{g(\omega)}{\omega}\,d\omega \tag{2.80}$$

Here M is the mass of the atom. This expression becomes more complicated for rigid-molecule calculations where the rotation of the molecule around its centre of mass must be considered, leading to anisotropic displacements and a quite involved algebra [306, 307]. Because of this, calculating displacement parameters has not been attempted in this thesis.

2.8.5 Convergence of lattice dynamics calculations

The convergence of the vibrational energy with respect to the Brillouin zone sampling was examined for four crystal structures. The X23 benchmark crystal structures [214] (see Chapter 3) of benzene, acetic acid, cyanamide and urea were used to test the convergence of free energy calculations when different **k**-points were sampled and when the Debye method and Gaussian kernel density estimation methods were used. A kernel bandwidth based on the phonon standard deviation was used, $h = \mathrm{std}(\omega_i)/20$.

The results are displayed in Figure 2.9. The standard textbook expression with discrete phonon frequencies is labelled 'Neat'. The others are with only the Gaussian KDE method, only the Debye method, and finally with both of these methods used together.

Note that the leftmost data points in each diagram in Figure 2.9 are for a single **k**-point only, the Γ-point. These clearly give erratic and unreliable results quite far from the converged results in the right hand part of the diagrams. Hence, sampling several **k**-points is necessary.

The enhanced Brillouin zone sampling methods, both the Debye and KDE methods and especially using both simultaneously, result in dramatically improved convergence. In fact, it seems quite possible to use $\mathbf{k} = \mathbf{0}$ lattice dynamics if the KDE and Debye methods are both used. Based on the results here, this method gives nearly the same results as the much costlier calculations that sample many **k**-points.

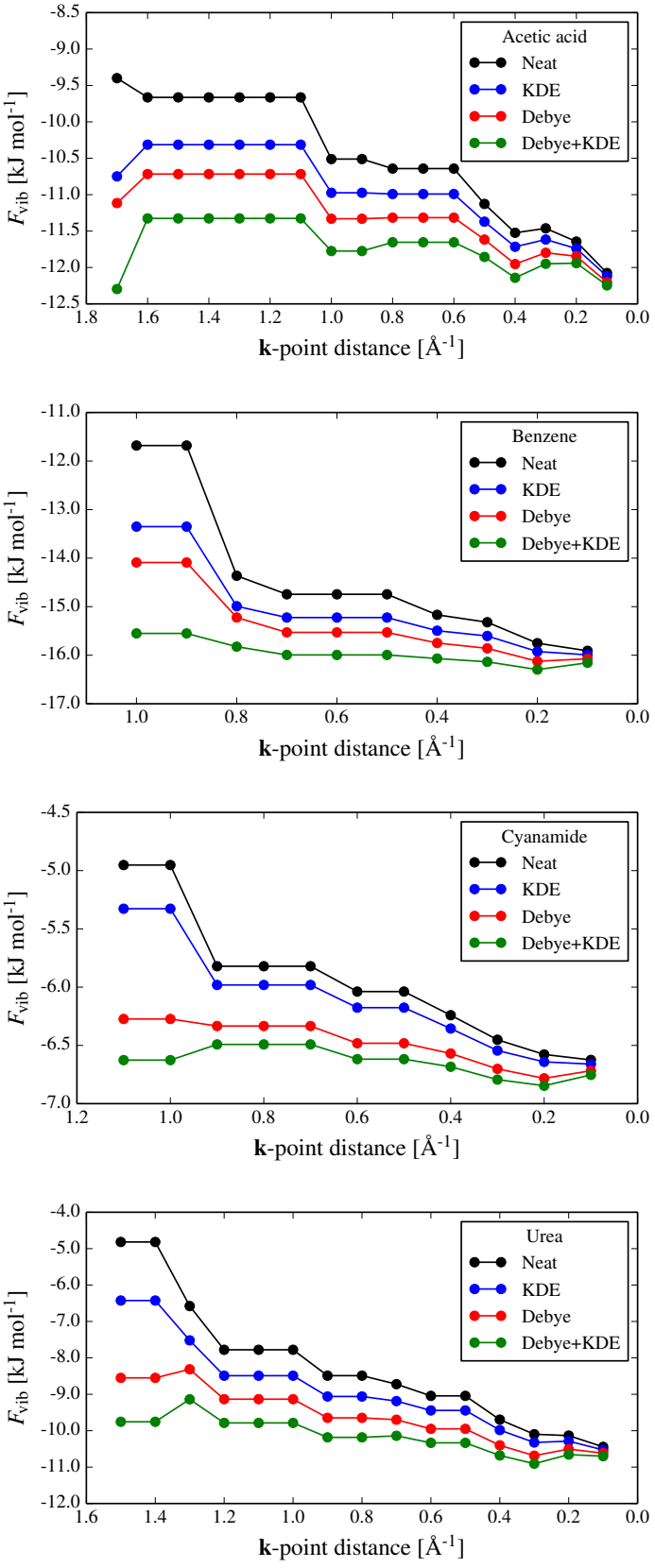


Figure 2.9: Convergence of the vibrational energy with different Brillouin zone sampling methods.

2.8.6 Quasi-harmonic approximation lattice dynamics

Harmonic approximation lattice dynamics is based on an assumption that the potential energy changes quadratically with respect to displacements of atoms, or molecules in our case. This is clearly not accurate since the interatomic potential has a steep repulsive wall and a much softer r^{-6}-dependence on either side of the equilibrium distance, the potential is clearly anharmonic.

The neglect of anharmonicity in lattice dynamics limits the accuracy of free energy calculations in the high-temperature limit. One can also imagine crystal structures where the potential energy is strongly anharmonic, such as the potential well experienced by a guest molecule inside a clathrate cavity.

An important consequence of the harmonic approximation is that the crystal geometry does not change with temperature, *i.e.* there is no thermal expansion. Furthermore, the calculated phonon frequencies and elastic properties are constant with respect to temperature, all in contradiction to physical reality. The free energy will be that for a constant volume, the Helmholtz free energy, not the desired Gibbs free energy.

Several ideas have been proposed for modelling anharmonic lattice dynamics in primarily ionic crystals [308, 309, 310, 311]. These models are quite intricate, and since they apply primarily to inorganic materials, beyond the scope of this thesis.

A simpler method for including the effects of anharmonicity is the quasi-harmonic approximation (QHA) [110, 312, 272]. Here, the phonon modes are still modelled as harmonic, but the crystal lattice dimensions are adjusted to account for thermal expansion. If experimental data on the lattice expansion is available, the unit cell can be restricted to the observed dimensions.

Constant volume calculations however do not allow the calculation of elastic constants and often experimental data on the lattice parameters are not available. In this thesis we are interested in the *a priori* prediction of both lattice parameters and elastic properties, why this method is not adequate.

An alternative is to perform harmonic lattice dynamics at several unit cell volumes. The free energy minimum can then be located for a flexible unit cell [300]. The method is cumbersome, as a large number of calculations must be performed for different cell geometries. The results can be used to fit an equation-of-state model for the crystal structure, relating unit cell volume, pressure and bulk modulus by, for instance, a Murnaghan equation [313, 312]:

$$P(V) = \frac{K}{K_0'}\left(\left(\frac{V_0}{V}\right)^{K_0'} - 1\right) \tag{2.81}$$

where K is the bulk modulus and K' its first derivative

$$K' = \left(\frac{\partial K}{\partial P}\right)_T \tag{2.82}$$

These methods are cumbersome because of the anisotropic thermal expansion of molecular crystals. One cannot assume that the lattice angles are constant, leading to not only an expansion, but also a contortion of the unit cell. It would in principle be necessary to fit separate equations for states for the individual lattice parameters.

A simpler method, that allows anisotropic contortion of unit cells and the calculation of elastic tensors is to use a negative pressure to expand the unit cell at elevated temperatures. The *thermal pressure* P_{th} is [314]:

$$P_{\text{th}} = -\left(\frac{\partial F_{\text{vib}}}{\partial V}\right)_T \tag{2.83}$$

We can obtain a thermal pressure from only two or three harmonic lattice dynamics calculations at different volumes and approximate the derivatives with finite differences [115]. Geometry-optimising the crystal structure at $-P_{\text{th}}$ results in a realistic anisotropic expansion.

This quasi-harmonic approximation method facilitates the calculation of the minimum in free energy, the Gibbs free energy $G(T, P)$, at arbitrary constant ambient pressures P_0 through only three harmonic calculations. Two in order to obtain the thermal pressure, and then a final one after geometry-optimisation at $P = P_0 - P_{\text{th}}$.

The accuracy of the thermal pressure method depends on whether the vibrational energy varies linearly with volume. This is shown to be the case for volumetric expansions up to *ca* 15% for three crystal structures in Figure 2.10. Here the vibrational energy was calculated with linear supercells to a **k**-point distance of 0.35 Å. The clear linearity means that it is safe to use only two unit cell volumes in calculating the thermal pressure.

Recently, quasi-harmonic approximation lattice dynamics was used to computationally reproduce the thermal expansion and thermodynamic properties of carbon dioxide at a very high level of theory [279]. The lattice energy was calculated as a sum of the energy of individual molecules, pairwise short-range interactions and long range pairwise and many body interactions. The first two terms were calculated at MP2 or CCSD(T) level of theory while the long range interactions were modelled with the AMOEBA force field.

The quasi-harmonic approximation was implemented in a way different to the ones described above [280]. Phonon frequencies vary with unit cell volume and this is modelled with Grüneisen parameters γ and calculated for every phonon ω_i

$$\gamma_i = \frac{\partial \ln \omega_i}{\partial \ln V} \tag{2.84}$$

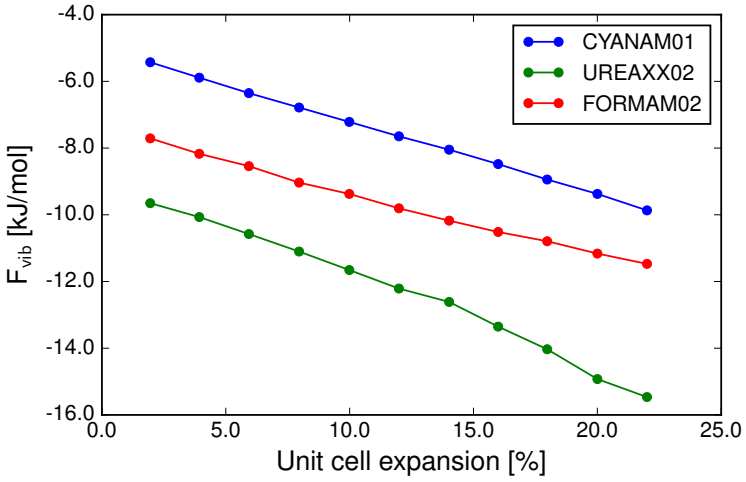


Figure 2.10: Linearity of the lattice-vibrational energy with respect to unit cell volume for three crystal structures.

The phonon frequency at an arbitrary volume is then

$$\omega_{i,T} = \omega_{i,0} \left(\frac{V_T}{V_0} \right)^{\gamma_i} \tag{2.85}$$

Using these methods the thermal expansion could be calculated for crystals of carbon dioxide, water ice Ih, acetic acid and imidazole. It was found that thermal expansion contributes 1–2 kJ/mol to the free energy near room temperature and that this may be sufficiently large to affect free-energy rankings among energetically close polymorphs.

In Chapter 5 we will present quasi-harmonic calculations on a very large set of organic crystal structures to obtain statistical distributions on how large these effects are.

2.9 Free energy curves

Here I will review the method of free energy curves, introduced at the Innsbruck school. The possibility of diverging free energy curves and the accuracy of empirical rules of thumb introduced by Burger, Ramberger and the Koflers are discussed.

At any given pressure, a pair of polymorphs is either enantiotropic or monotropic. It does not make any sense to talk about *one* enantiotropic form, it is enantiotropic relative to another form. Each polymorph may also be monotropic relative to one form, while simultaneously enantiotropic to another. These relationships may change with pressure, since polymorphs have different density and may be re-ranked by the pressure-volume contribution to the Gibbs free energy.

In monotropic pairs, one polymorph has a larger (less negative) free energy than the other at all temperatures below the melting point, see Figure 2.11. This means that the higher

energy polymorph is always thermodynamically unstable and will eventually transform to a more stable form. Consequently, monotropic phase transitions are irreversible and exothermic.

In enantiotropic pairs, both polymorphs are thermodynamically stable relative to each other, but in different temperature intervals. At one particular temperature, their free energy curves cross and both polymorphs have exactly the same free energy. This can be seen experimentally as reversible solid-to-solid phase transitions [315], iso-energetic polymorphism [316] and concomitant crystallisation [317]. Enantiotropic transitions are *in principle* reversible, but rarely in practice. The transition is endothermic when going from the low-temperature form to the high-temperature polymorph. There are often significant deviations from ideal thermodynamic behaviour because of lattice defects, impurities and kinetic energy barriers [318].

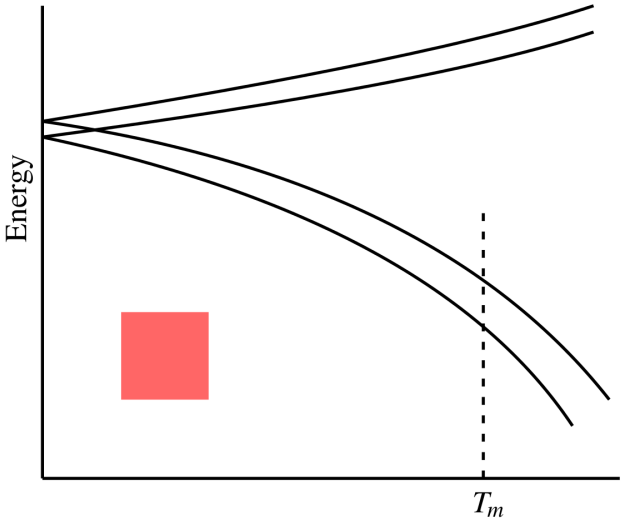


Figure 2.11: Schematic free energy curves for a polymorph pair. The free energy is diverging as the temperature increases. The unstable form has a lower entropy than the stable form and the polymorphs are monotropic. Note the color-coding.

Free energy curves and a few general empirical rules, the Burger-Ramburger rules, were introduced by the Innsbruck school and used to enhance the understanding of experimental results on the stability and phase transitions of polymorphs. The fundamentals of thermodynamics of polymorphic crystals were discussed in detail by Burger and Ramburger [319]. A compilation of experimental results was also assembled and contains exampled of where the methods work, and where they fail [320]. Free energy curves have since been used to interpret and understand polymorphs. One aspect however, seems to have been forgotten and will be addressed here. I ask whether free energy curves can diverge with temperature, see Figure 2.11. This possibility has rarely been addressed in the literature. Bernstein discusses free energy curves in general, but does

not comment explicitly on whether they can diverge, though he does not exclude the possibility [20].

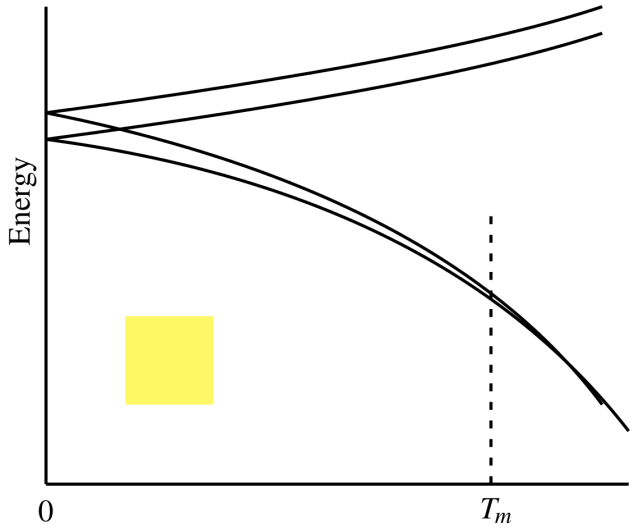


Figure 2.12: Schematic free energy curves for a polymorph pair with converging curves that cross above the melting point. These will experimentally be found to be monotropic. Note the color-coding.

Figures 2.11 and 2.12 introduces a color-coding of polymoprh pairs. Pairs with diverging free energy cures are 'red', while converging pairs are yellow or green (not shown). The green polymorphs are enantiotropic, *i.e.* their free energy curves converge such that they cross below the melting point. Yellow pairs have converging curves, but they converge too slowly to cross below the melting point, see Figure 2.12. These color-codes are used in Figure 5.29 in Chapter 5.

The Burger-Ramberger rules [319] are often used to deduce thermodynamic relationships between polymorphs, and are held in high regard in the organic solid state community. There are however some misconceptions about them. For example, one paper discussing free energy curves claim that free energy curves always converge [321], citing Burger and Ramberger as support for the claim.

I want to, in this thesis, argue that a careful reading of Burger and Ramberger will show that such conclusions cannot be drawn from the arguments and evidence presented by Burger & Ramberger. In fact, there are two different points I want to make. The first is that Burger and Ramberger *did* consider diverging free energy curves and *did not* rule them out as impossible. The second point I want to make is that even if the Burger-Ramburger rules had ruled out diverging curves, the rules themselves are based on approximate derivations and rely on simplifying assumptions. Hence, the rules must be considered as empirical guide lines, not as infallible laws of nature. In fact, several exceptions to the rules are documented by Burger and Ramberger themselves [320].

On page 260 [319], Burger & Ramburger point out that near a first-order transition point, ΔH and ΔS have the same sign and that the free energy curves are necessarily converging. Note however, that this paragraph concerns only enantiotropic pairs, which by definition have converging curves.

On page 262, a paragraph begins by making the assumption that the zero-point energy is negligible and that $\Delta H(0) = \Delta E_{latt}$. Real zero-point energies are calculated for 1061 polymorphs in Chapter 5 of this thesis. The zero point energy is typically 2.5 kJ/mol, and differences in zero point energy may be as large as 0.5 kJ/mol, see Figure 5.10 in Chapter 5. This is large enough to invalidate Burger & Ramburger's assumption, the lattice energy difference and the enthalpy difference at 0 K may even have different signs!

On page 262, a reference is made to the harmonic approximation lattice dynamics method, clearly stating that this is a 'first approximation'. Diverging Gibbs free energy curves are found in this thesis by the quasi-harmonic approximation, see Figure 5.29.

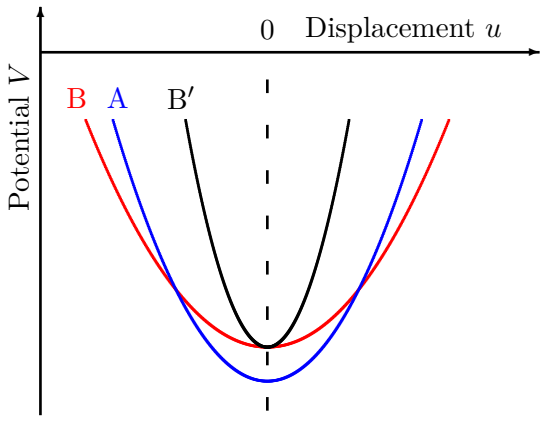


Figure 2.13: Schematic potential energy wells in thee polymorphs A, B, B′. The frequency of vibrations and the resulting entropy depends on the shape of the potential energy. The potential well in B′ has a larger force constant ($\partial^2 V/\partial u^2$) than A, despite being higher in energy at the equilibrium position, resulting in diverging free energies for polymorphs A and B′. Adapted from reference [319].

In their figure 2, reproduced here as Figure 2.13, Burger & Ramburger describe the (harmonic approximation) potential energy well experienced by a molecule in three polymorphs and the relationship between the shape of the potential energy and the vibrational frequency. The frequency of vibrations and the resulting entropy depends on the shape of the potential energy. The potential well in B′ has a larger force constant ($\partial^2 V/\partial u^2$) than in A, despite being higher in energy at the equilibrium position. This is exactly what would be needed to have diverging free energy curves for forms A and B′. Hence, Burger & Ramberger did consider this possibility, noting that it is 'exceptional, but they did not rule it out as impossible. Indeed, there is no reason to believe that the

force constant of a lattice vibrational mode must be related to the equilibrium potential energy.

On page 263 [319], Burger & Ramberger 'assume for simplicity' that the phonon frequency of every vibrational mode in polymorph A is higher than the frequency of the corresponding mode in B, as in Figure 2.13, and that ΔS is monotonically increasing with temperature, leading to strictly converging curves. This is then immediately modified, by saying that for real crystals the harmonic approximation is not strictly valid, and that the relationship $\nu_A > \nu_B$ is not always true. Indeed, such a relationship between phonon frequencies cannot be postulated. Being aware of this, they go on to say that "*the relation holds only on the average; moreover, the frequency depends on the width as well as the depth of the potential well, so a smaller potential depth may not necessarily imply a lower frequency*". This is case for B′ in Figure 2.13, and leads to diverging free energy curves.

Another article, by Yu [322], derives rules for deducing thermodynamic stabilities from experimental heats of fusion and melting temperatures, avoiding some of the assumptions and simplifications made by Burger & Ramberger. Yu considered the possibility of diverging curves, noting that "*monotropy is not necessarily due to the divergence of G_A and G_B as T increases, but may result from their converging not before T_{mp}.*" A method for calculating the transition temperature T_t is then derived, and "*if $T_t < 0$, it formally implies the case in which as T decreases, [the negative] ΔG increases, but at a rate insufficient to reach zero at a positive T. [...] This case is considered undecided.*" Hence, Yu's method does not apply for such cases, but he did consider diverging curves possible. Yu lists a number of experimentally known polymorphic phase transitions. Several of them fall in the category 'undecided' and have transition points formally below 0 K. Consequently, they are inferred, *from experimental data*, to have diverging free energy curves.

Polymorphs of ethyl β-aminocrotonate (I/II) [323], might have diverging curves [322]. Burger & Ramburger knew about these polymorphs and give them as an example that violates the density rule [320]. Yu also mentions polymorphs of ethyl biscoumacetate [324, 325], a commercial pharmaceutical as possibly diverging. Polymorphs of glutaronitrile [326, 327] is also given as an example of polymorphs with transition temperature below 0 K.

Gu & Grant [328] referred to Yu regarding diverging free energy curves: "*The rare possibility that the transition temperature lies below absolute zero has been considered by Yu and is discussed in the Appendix.*" In the Appendix, Gu & Grant give a few thermodynamic relations that must be true for diverging polymorphs. They certainly consider diverging free energy curves possible, and give cyclopenthiazide I/II [329] an example of possibly diverging polymorphs.

I have now discussed the possibility of diverging free energy curves in some detail. The four or five literature sources mentioned here may well constitute a comprehensive review. That's how little is known about diverging free energy polymorphs. We will come back to computational results in Chapter 5 that will drastically change the picture.

2.10 Crystalline voids and guest enclathration

2.10.1 Calculations of cavity volumes

Crystal structures that contain empty voids are of special interest in this thesis. Cavities in crystal structures and their volumes can be determined computationally.

The most common method is to computationally construct a lattice or grid throughout one unit cell and for each grid point test if it falls within the van der Waals-radius of any atom in the structure. The fraction of such lattice points is the packing coefficient of the structure, it is typically 0.7 for ordinary close-packed structures. Truly porous structures tend to have packing coefficients between 0.5 and 0.6. This grid-based method will be used in this thesis.

Other methods for void calculations exist. The Zeo++ program and software library is a competent tool which allows void calculations by stochastic and ray-tracing methods. In addition, it can determine the dimensionality of interconnected void structures and other properties of cavities in crystal lattices [330, 331, 332, 333]. Since such calculations can be computationally costly, implementations using graphical processing units have been made [334, 335].

The packing coefficient neglects the size of the guest molecule itself and the fact that the space in the cusps near the intersection of atoms is not available to a guest of finite size. Void calculations therefore add a radius to the probe particle at the lattice points, most commonly 1.2 Å.

The calculation is highly sensitive to the particle size. Using a large probe will underestimate the volume and will often result in an erroneous topology of the void space. Small molecules may diffuse through narrow channels between larger cavities. We have chosen a probe size of 1.2 Å radius, corresponding to the kinetic diameter of a H_2 molecule, this is also the default setting in Platon. The choice of probe radius is difficult, since too small values may cause an overestimation of the pore volume available to more realistic guest molecules, or underestimations of the volume since the algorithm neglects dynamic porosity caused by thermal fluctuations and expansion of the pore in response to the presence of a guest. Methods for void calculations and problems associated with them have been discussed by McKeown, Barbour, Gavezzotti, Haranczyk and Spek [69, 336, 337, 330, 283].

A coarse grid will also result in inaccurate volumes. A convergence test was carried out on a set of clathrates in the Cambridge Structural Database. The convergence of void volume calculations with respect to grid spacing is shown in Fig 2.14. The relative errors of the calculated void volume for different grid spacings were compared to the result for a fine grid spacing of 0.1 Å.

$$\text{Error} = 100 \cdot \left(\frac{V_{\text{grid}} - V_{0.1}}{V_{0.1}} \right) \tag{2.86}$$

The computational cost prevents the finest grid spacings to be used routinely. The convergence test indicated that the void volume calculation is slowly converging for grid spacings smaller than 0.4 Å. Grid spacings larger than 0.3 Å produces unacceptably large errors. A grid spacing of 0.2 Å is adequate, as it consistently gives volumes with small errors of approximately 5% relative to the 0.1 grid spacing. By extrapolating the curves in the convergence diagram, the true errors relative to a perfect grid spacing of zero can be estimated. The volumes calculated with a 0.2 Å grid spacing are systematically underestimated by 5–10%.

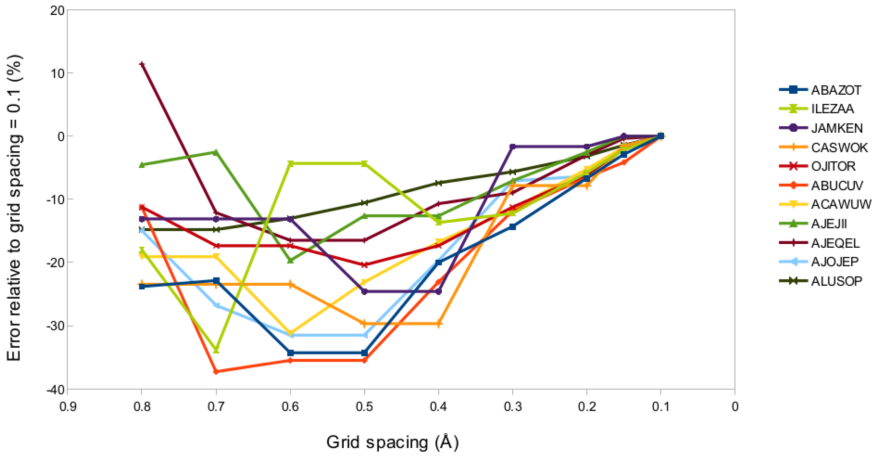


Figure 2.14: Convergence of void volume calculations.

Void calculations such as these can also be used to calculate the volumes of individual cavities in a unit cell, which may be important. There can for instance be two different cavities in a unit cell, only of which is large enough to host a guest molecule. The total packing coefficient is then not a good predictor of inclusion behaviour. Hence, it may be necessary to calculate the volume of each distinct pore in a structure. The position and overall shape of the cavities can be determined from the centre-of-mass or centroid and the eigenvectors of the grid points forming each cavity.

The lattice- or free energy cannot be used as the scoring function in crystal structure predictions using the guest-insertion method. We need to know which of all the predicted structures are likely inclusion compound host structures. One way is to rely on the

empirical rule that the guest molecule in such structures typically occupy 55% of the volume [338, 339].

Cruz-Cabeza pioneered this methodology of searching for solvates and inclusion compounds. She suggested that a combination of the relative lattice energy and solvent-accessible volume be used in selecting likely inclusion frameworks from crystal structure predictions. She noted that the methodology can be for the anticipation of inclusion behaviour in organic molecules prior to experiments [75]. Calculations such as these have been applied to predicted crystal structures of urea, Dianin's compound, theobromine, carbamazepine and hydroquinone [75, 340, 46]. It was found that predicted sets of polymorphs often contain porous crystal structures. Since these are not close-packed, they are always unstable and typically found several kJ/mol above the global lattice energy minimum.

Also Braun has successfully predicted solvates and inclusion behaviour of 4-aminoquinaldine and of the alkaloids brucine and strychnine with this method [341, 342].

2.10.2 Guest-to-host volume ratios in clathrates

In order to develop a scoring function for inclusion compounds in predicted structure landscapes, and to confirm the validity of the 55% rule [338] using the void calculation method adopted here, the CSD was searched for organic clathrate structures. Disordered structures, metallo-organic structures, and hydrates were excluded. In addition, the hydroquinone SO_2 and hydroquinone C_{60}-fullerene clathrates were used. The set of selected structures was chosen only to be large enough to estimate the spread in guest-to-host ratio and to represent a variation in host guest chemistry.

The guest molecules were removed from the structure and a void calculation was performed using a 1.2 Å probe radius and 0.2 Å grid spacing. Guest molecule volumes other than xenon were calculated with the Molinspiration property engine v2013.09, by Molinspiration Cheminformatics 2014. It provides volumes calculated from summing atomic group volumes fitted to vacuum AM1 semiempirical conformations of a large training set. For the volume of xenon, the van der Waals volume 42.2 $Å^3$ was used. Alternatively, the volume of molecules can be calculated by Connolly's method [343].

Guest to host volume ratios are normally distributed around 60% with a standard deviation of 8 percentage points.

2.10.3 Computational guest enclathration

Molecular volumes were calculated as the sum of atomic fragment contributions using the Molinspiration property calculator available online [344] and a volume of 42.2 $Å^3$ was

Table 2.5: Structures and void properties used in determining selection rules for inclusion hosts. In summary: 58.9 ± 8.0 ($1\ \sigma$). TCE = 1,1,2,2-tetrachloroethane

CSD code	# of guests	Guest	Guest Vol [Å^3]	Void Vol [Å^3]	Ratio [%]
ABAZOT	2	3-pentanone	98.344	327	60.14
ABUCUV	2	acetonitrile	46.055	158	58.28
ACAWUW	2	ethanol	54.016	164	65.86
AJEJII	3	acetonitrile	46.055	193	71.58
AJEQEL	4	chloroform	70.07	553	50.68
AJOJEP	1	chloroform	70.07	119	58.88
ALUSOP	2	chloroform	70.07	275	50.96
ANIRUL	1	butan-1-ol	87.619	171	51.23
ASOKOI	1	TCE	100.434	154	65.21
AVESOJ	4	diethylamine	91.763	694	52.88
BABYIN	2	acetonitrile	46.055	171	53.86
BAFSOQ	2	urea	54.196	154	70.38
BAZJES	1	acetone	64.74	122	53.06
BEGZAP	2	propionic acid	72.999	285	51.22
BERYED	4	methane	28.644	193	59.36
BICYCLO	2	diethylether	88.345	271	65.18
BIFQAI01	2	benzonitrile	100.903	408	49.46
BIJVOf	3	acetid acid	46.055	217	63.66
BOSVEK01	1	sulphur dioxide	41.314	61	67.72
BUSPAG	1	methylisocyanide	51.287	79	64.92
CAKGUQ	1	acetonitrile	46.055	85	54.18
CAVMES	1	diethylether	88.345	271	65.18
GAGNEJ	2	DMSO	71.433	216	66.14
GIRBOY	3	xenon	42.2	226	56.01
HQ-C_{60}	2	C_{60}	539.5	1945	55.46
HQUACN	1	acetonitrile	46.055	77	59.81
HQ β	1	sulphur dioxide	41.314	68	60.75
ILEZAA	1	xenon	42.2	84	50.23
JAMKEN	1	xenon	42.2	60	70.33
LOZNUJ	1	xenon	42.2	55	76.72
OJITOR	1	xenon	42.2	92	45.86

used for xenon. Crystalline cavity volumes were calculated by placing probe particles at positions in a regular grid with 0.2 Å spacing in the unit cell and testing if the particle was within any atom's van der Waals radius [345, 336], as implemented in the program PLATON [283]. Convergence of the calculated void with respect to grid spacing is discussed below. PLATON allows not only the total void volume per unit cell to be determined, but also the volume of each individual pore, and that is the volume used in this work.

A practical algorithm for computationally inserting guest molecules into crystal pores is to first do a calculation of the voids, determining the volume and the centroid coordinates for each cavity in the unit cell. Since guest insertion frequently breaks the space group

symmetry of the host structure, is is necessary to perform these calculations on a $P1$ representation of the crystal structure.

The center of mass and the volume of the guest molecule are also determined. The volume of a molecule can be calculated in several ways. The naive method of calculating the volume of overlapping atomic spheres is possible, but cumbersome [346, 347]. The volume of a molecule is estimated in many QM-codes as the volume bounded by an electron isodensity surface. A relatively simple method is to use a Monte Carlo calculation: enclose the molecule in a cuboid and randomly insert a test point. The volume times the fraction of test points that fall within the van der Waals radius of any atom is the van der Waals volume of the molecule.

In this thesis we will use a molecular fragment contribution method. The volumes of the fragments have been fitted to reproduce volumes of electron isodensity surfaces calculated with the AM1 semiempirical QM method. The fragment model is implemented in the molinspiration software, accessible online.

The centre of mass of the molecule is translated to the centroid point of a cavity and randomly rotated until there are no distances between guest and host atoms smaller than 0.85 times their combined van der Waals radii. The guest insertion may be repeated until there are not more cavities large enough to host a guest. Once guests have been inserted in this way, the entire crystal structure should be energy-minimised to allow the structure to relax to an energy minimum.

For solvates and especially hydrates, the guest insertion may have to be repeated several times to find all possible geometries and adsorption sites. For clathrates such as the ones studied in this thesis however, this method works well.

2.11 Structure comparisons

When the trial structures are energy minimised, many will converge to the same potential energy minimum. The crystal structures may have different unit cell representations and symmetries but have converged to essentially the same crystal structure. We need to identify and remove the duplicate structures are keep exactly one structure from each group of similar structures.

Several different algorithms exist for calculating crystal structure similarity. One of the most commonly used methods is to calculate the root mean square distance (RMSD) between atoms in a group of molecules. The COMPACK program [165] rotates a cluster of molecules taken from one structure onto a reference structure. Once the two clusters of 15 to 20 molecules have been overlayed the RMSD is calculated. This RMSD-value is a convenient estimate of structure similarity. For instance, it is practical to demonstrate

a small RMSD-value between one experimental and one predicted structure to show that the structures match.

There are also other methods. For instance, one could compare the atomic radial distribution functions or calculated powder X-ray diffraction patterns [348, 349, 350]. One method that also considers space group symmetry is to solve for the affine mapping of coordinates from one structure or another [351], a method implemented in the program KPLOT [352].

There are however problems with most of these methods. The COMPACK algorithm depends on which molecules are selected to be overlay and gives quite different results based on which molecules are chosen. Polymorphs can contains virtually identical layers that may be stacked on top of each other in different arrangements. If a large fraction of the molecules chosen for comparison belong to the same layer, the RMSD-value will be very small despite the structures being quite different. Such polytypic polymorphs [353] may not be identified correctly. Problems can also arise when comparing structures with differing Z'-values. Many of the practical problems of structure comparisons and clustering have been discussed by van Eijck [140], in a paper that is well worth reading.

2.12 Solid-state nuclear magnetic resonance spectroscopy

Nuclear magnetic resonance spectroscopy utilizes the magnetic properties of fermionic nuclei to extract information about the local environment surrounding the nuclei. A half-integer spin immersed in an external magnetic field will have two energy states corresponding to the nuclear spin being either parallel or anti-parallel to the external field. Other nuclei in the vicinity will locally perturb the field and shield the nucleus, resulting in a shift in the energy-difference between the two states. Transitions between the states can be observed as peaks in the radio-frequency power spectrum after the sample has been excited by an RF pulse. The spectral peak positions and their shapes consequently contain information about the chemical shielding and the local structure surrounding the fermionic nuclei [354].

Here we will limit the discussion to spin-1/2 nuclei, such as ^{1}H, ^{13}C, and ^{129}Xe in particular. The reader should however be aware that spin-3/2 nuclei such as ^{35}Cl and ^{37}Cl are also of interest for the study of crystal structures [355, 356].

The reason for why we are interested in NMR is because crystal structure prediction is potentially very useful as a method for determining the crystal structure for which experimental NMR data is available. The measured chemical shifts can be compared to computed chemical shielding tensors of predicted crystal structures, and because the NMR shielding is so sensitive to the crystalline structure, it may be possible to single out which of the predicted structures that correspond to the experimental material.

A couple of such attempts have been made, using ^{1}H, ^{13}C and ^{35}Cl solid state NMR experiments in combination with CSP [357, 358].

In Chapter 7, a method for the crystal structure determination of clathrates will be presented, that relies on CSP in combination with ^{129}Xe NMR. For this purpose, we need to introduce the fundamentals of solid-state NMR and the *ab initio* calculation of nuclear magnetic shielding. A comprehensive treatment of NMR theory and computation is not possible to include here, but the reader should consult the literature cited below.

The chemical shielding of a nucleus in a molecule is anisotropic because of the generally anisotropic chemical environment around the atom. However, because of random orientations and thermal motions of molecules in a liquid, only an averaged chemical shift can be observed in NMR experiments of liquid samples.

The theory of NMR has been reviewed elsewhere [354]. In summary, it suffices to note that the energy of a nuclear spin in a local magnetic field $\mathbf{B}$ is given by the NMR Hamiltonian H

$$H = -\gamma \mathbf{I} \cdot \mathbf{B} \tag{2.87}$$

or equivalently, after separating the external applied field $\mathbf{B}^{\text{ext}}$ and the local field at the nuclear position due to shielding.

$$H = -\gamma \mathbf{I}(1-\sigma)\mathbf{B}^{\text{ext}} \tag{2.88}$$

Here γ is the gyromagnetic ratio, $\hbar\mathbf{I}$ the spin angular momentum and σ the chemical shielding.

The chemical shielding is a 3×3 tensor. An external homogeneous magnetic field $\mathbf{B}^{\text{ext}}$ will induce an electric current J^{in} in a sample in the field. The induced current in turn produces a magnetic field $\mathbf{B}^{\text{in}}$ at the nuclear position r. We assume the response in linear and write:

$$\mathbf{B}^{\text{in}}(r) = -\sigma(r)\mathbf{B}^{\text{ext}}, \tag{2.89}$$

where $\sigma(r)$ is the local shielding tensor. NMR experiments of liquids give access only to the isotropic chemical shielding $\sigma_{iso} = \text{Tr}(\sigma)/3$. The experimentally observed chemical shift δ is referenced to an absolute chemical shift standard σ_{ref}, which for xenon is the chemical shift of the pure gas extrapolated to zero pressure

$$\delta = \frac{\sigma_{\text{ref}} - \sigma}{\mathbf{I} - \sigma_{\text{ref}}} \simeq \sigma_{\text{ref}} - \sigma \tag{2.90}$$

Solid state NMR of powder samples can provide the diagonal elements $\delta_{xx}, \delta_{yy}, \delta_{zz}$. Only NMR experiments of single crystals can also reveal the three remaining independent elements, related to the orientation of the tensor.

Experimentally, the tensor properties are revealed as an asymmetric line broadening of the chemical shift spectral peak. This makes it convenient to consider the tensor asymmetry η and the chemical shift asymmetry $\Delta\delta$.

$$\eta = \frac{\delta_{yy} - \delta_{xx}}{\delta_{zz} - \delta_{\text{iso}}} \tag{2.91}$$

$$\Delta\delta = \delta_{zz} - \frac{\delta_{xx} + \delta_{yy}}{2} \tag{2.92}$$

Chemical shielding tensors can also be calculated. In 1996 Mauri *et al.* published a solution to the problem of calculating chemical shifts in the solid state with an all-electron Hamiltonian. Their paper contains a theory for the *ab initio* calculation of chemical shifts in condensed systems using periodic boundary conditions [359]. In practical implementations however, it is used with pseudopotentials for the core electrons, a development made possible by work of Blöchl and van de Valle [360, 361]. A popular method for calculating accurate NMR tensors in periodic boundary conditions was presented by Pickard & Mauri in 2001 [362]. It is based on Blöchl's projector augmented wave method (PAW) [363] and the theory of chemical shift calculations by Mauri et al. [359]. They called their method gauge including projector augmented wave method (GIPAW) and it facilitates the calculation of chemical shielding calculations in plane wave GGA DFT with pseudopotentials. Their article include an explicit derivation [362]. The method is included in density functional computer programs such as CASTEP and QUANTUM ESPRESSO [364, 365, 361]. The calculation proceeds by calculating the induced current J^{in} and the induced magnetic field $\mathbf{B}^{\text{in}}$, but the shielding tensor can also be written as the more intuitive quantity [354, 366]

$$\sigma_{ij} = \frac{\partial^2 E}{\partial \mathbf{m}_i \partial \mathbf{B}_j^{\text{ext}}} \tag{2.93}$$

where E is the (free) energy of the electronic structure and $\mathbf{m}$ is the magnetic moment of the nucleus. A complete coverage of the GIPAW method is too involved to be included in this thesis. The reader is encouraged to consult review articles on the topic [367, 368, 369, 370, 361].

Xenon is a particularly interesting element for NMR and for the study of porous materials. The use of xenon NMR for the study of zeolites was introduced by Ito & Fraissard in 1988 [371]. Ripmeester *et al.* used Xe NMR for the study of hydrate clathrates and proposed methods for using Xe NMR for the study of clathrates, inclusion complexes *etc.* The ^{129}Xe isotope has spin 1/2, while ^{131}Xe has spin 3/2, which can be useful as this allows measurements of the local electric field gradient.

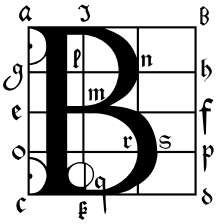
There is a general and relatively simple relationship between the ^{129}Xe isotropic chemical shift and the pore size distribution in porous silica materials [372]. The chemical shift anisotropy depends on the shape and symmetry of the adsorption site [373].

The use of xenon NMR for the study of porous materials, in particular silica gels and porous glasses, has been investigated by Jokisaari's group in Oulu, Finland, and published in several articles [374, 375, 376, 84, 83]. Their method exploits the high sensitivity of the ^{129}Xe isotope chemical shift to its local environment. They used an inert liquid to slow the diffusion of xenon so that the NMR signal is characteristic of the properties of a single pore, and the signals of all atoms in the sample represent distributions of properties.

The hydroquinone xenon clathrate has also been investigated by the Finnish group. The pore size and shape of clathrates can be determined by xenon NMR and it was found that the xenon atom perturbs the characteristic hydrogen bonds in the clathrate cage [82].

Chapter 3

Benchmarking of computational methods



enchmarking is crucial for ensuring the accuracy of computational methods. In this chapter several different force fields are benchmarked against experimental reference data in order to determine how accurate they are for the calculation of lattice energies and structures of molecular crystals. I compare the performance of the force fields to several of the most popular DFT-D methods. In addition, I test the implementation of quasi-harmonic lattice dynamics calculations described in Chapter 2 by comparing computationally thermally expanded crystals to experimental structures recorded at non-zero temperatures. I acknowledge Orla Sheehan Pundyke, whom I had the pleasure of supervising, for performing a preliminary study. It was her results that prompted me to undertake the benchmarking reported here. This chapter is based on work previously published as reference [2].

3.1 Background

Interactions between molecules in the solid state are of great importance for the development and prediction of crystal forms. Understanding and modelling intermolecular interactions help facilitate supramolecular chemistry, drug formulation and crystal engineering [377]. Computational methods have been developed with this aim, and perhaps with the ultimate goal of *ab initio* crystal structure prediction (CSP) [138].

Computational modelling of molecular crystals has a long history. Atom-atom force field methods were employed as soon as computers became accessible [87, 378, 208, 207, 379]. More recently a trend favouring density functional theory (DFT) calculations has

emerged. This is a direct consequence of the development of computationally affordable dispersion-corrected DFT methods. In the solid state, the dispersion cannot be neglected. The trend is perhaps most notable in the series of blind tests of crystal structure prediction organized by the CCDC [135, 25, 5]. Accurate lattice- and free energy calculations with DFT-D are currently the most successful prediction methods [162, 108, 5].

As I will show in Chapter 5, the lattice energy difference between polymorph pairs is usually smaller than 2 kJ/mol [1, 19, 3], so that relatively small contributions, such as many-body dispersion [380, 100, 250], polarization [381, 382], lattice vibrations [300, 1, 3] and pressure [24, 383] affect the relative stability of polymorphs.

Virtually all current methods for crystal structure prediction work under the assumption that it is the thermodynamically most stable form(s) that crystallise. This implies that the final scoring function of the predicted structures should be the Gibbs free energy. Calculating accurate Gibbs free energies is challenging and requires intricate computational modelling. A further obstacle is that since the predicted crystal energy landscapes typically contain hundreds of plausible polymorphs [384, 22], the computational cost of evaluating each crystal structure cannot be too large.

For these reasons, it is desirable to have computational methods that are computationally affordable while also accurate enough. Benchmarking studies are used to compare the performance of different methods [385, 386, 249, 387, 388, 389, 214, 115, 390, 2, 279, 391]. However, it is difficult to rigorously benchmark computational models for molecular crystals, because there is a rather limited amount of high-accuracy experimental reference data available. Experimental sublimation enthalpies have been estimated to have errors of ± 4.9 kJ/mol [392]. This implies that the accuracy of calculated lattice energies cannot be assessed to better than this margin of error. Consequently, it is common to compare computational methods to each other and rank them by the level of theory.

In this chapter I will address the challenge of calculating Gibbs free energies in a computationally very affordable way. I will benchmark the accuracy of the force fields used throughout the thesis in reproducing experimental lattice energies, crystal structures, thermal expansion coefficients and elastic constants. This should provide a comprehensive assessment of the overall accuracy of the methods in this thesis.

The benchmark study by Otero-de-la-Roza & Johnson [115] compared several popular plane wave DFT-D methods and their own exchange-hole dipole moment dispersion correction method (XDM) using 21 organic molecular crystal structures (C21) which accurate structures and sublimation enthalpies are known. The vibrational contribution to the sublimation enthalpy was calculated and subtracted to obtain 0 K benchmark lattice energies. Quasi-harmonic lattice dynamics was used to also compare calculated crystal structures to finite temperature experimental structures. However, only $\mathbf{k} = \mathbf{0}$

phonons were calculated and at a geometry that was not an energy minimum [115, 214], making the lattice dynamics of that study unreliable.

The C21 benchmark set was corrected with recalculated vibrational energies and extended with two additional crystal structures by Reilly and Tkatchenko to form the X23 benchmark [214]. Several DFT-D methods have been assessed using the X23 set. The DFT methods include the popular GGA functional PBE, the hybrid functionals PBE0, HSE06, B3LYP and M06-2X, non-local van der Waals functionals and the meta-GGA functional TPSS. Dispersion corrections such as the Tkatchenko-Scheffler (TS), Grimme's D2, D3 and a many-body dispersion (MBD) model by Reilly & Tkatchenko [214, 250, 393, 394, 395]. In addition, the empirically corrected PBEh-3c functional and density functional tight binding (DFTB) have been benchmarked with the X23 set [395, 396]. These studies together make the X23 benchmark by far the largest and best for assessing the methods used here.

Since the interactions between organic molecules are relatively weak, temperature dependent vibrations have a significant influence on crystal structures and their properties. The quasi-harmonic approximation [397] is often used to model crystal structures at non-zero temperatures. However, force fields that are parametrised to reproduce experimental crystal structures absorb some of the effects of lattice dynamics. The thermal expansion due to zero point and thermal vibrations will make the force field overestimate the volume of 0 K structures. Likewise, a force field parametrised to sublimation enthalpies will underestimate absolute lattice energies if this is not corrected for.

The force fields used in this thesis are based on those parametrised by Donald Williams [91, 223, 92, 93]. Later re-parametrisations have been performed to enhance the compatibility with distributed multipoles [225]. These force fields were parametrised in a way that would have absorbed some thermal effects, although the later works tried to mitigate this error. Despite this, it is plausible that modelling temperature explicitly can result in a double-counting of the thermal expansion and zero-point energy. Here we will compare computationally thermally expanded crystal structures to finite-temperature experimental data. This should allow us to estimate how large the error due to this double counting is.

In principle, the methods used here could also be benchmarked against experimentally determined elastic properties. That has been done previously [261]. It was found that when accurate potential models are used, the elastic constants are generally overestimated. For instant, the root mean square relative error in the diagonal elements of the stiffness matrix of a durene crystal was 38.3% and 50.8% respectively, for the W99_ESP and W99_DMA force fields. In general, overestimation by about 40% was seen for the FIT and W84 force fields for six molecular crystals.

It is likely that the neglect of thermal motion in molecules may account for much of the error in the calculated elastic constants [398, 261]. Because of the scarcity of suitable

experimental reference data and the temperate dependence of elastic constants, benchmarking against elastic properties has not been attempted here. We will revisit the temperature-dependence of elastic properties in Chapter 5.

3.2 Methods

3.2.1 Benchmarked force fields

In this chapter the accuracy of the following force fields is assessed by reproducing the X23 benchmark lattice energies:

- **FIT**: The FIT force field, [223] which is used with electrostatics described by distributed, atomic multipoles.
- **W99_ESP**: The W99 force field [93, 92], with atomic partial charges fitted to the molecular electrostatic potential.
- **W99_DMA**: The W99 force field with distributed atomic multipoles.
- **W99rev6311**: A revised version of the W99 force field [225], which was parametrised to improve hydrogen bonding when using atomic multipolar electrostatics.
- **W99rev6311P3**: A revised W99 force field [225], parametrised to be used with multipoles from a charge density calculated within a polarisable continuum model with dielectric constant $\varepsilon_r = 3.0$.
- **W99rev6311P5**: The same as W99rev6311P3, but using $\varepsilon_r = 5.0$ during the PCM calculation for the electrostatic model.

The FIT and W99rev6311P5 are also benchmarked with respect to their ability to reproduce crystal structure geometries, thermal expansion and elastic constants. This extended benchmark makes use of quasi-harmonic approximation lattice dynamics calculations. This means that the results depend both on the force field parameters and the specific implementation of the quasi-harmonic lattice dynamics presented in Chapter 2. For instance, these results also depend on the **k**-point sampling and the thermal pressure method for thermal expansion. Hence, the results presented here provide a comprehensive test of the entire computational methodology used in this thesis.

The force fields benchmarked here are hybrid energy models, consisting of a DFT model for the molecular geometry and energy, and an intermolecular atom-atom force field. All force fields have an *exp-6* (Buckingham) potential for non-electrostatic intermolecular interactions. Intermolecular electrostatics is modelled with atomic point charges or distributed atomic multipoles generated from a quantum chemical calculation.

In their original forms, FIT, W99 and W99rev provide repulsion-dispersion parameters for carbon, hydrogen, oxygen and nitrogen only. The original W99 force field [92, 93] was parametrised using an atomic partial charge electrostatic model, but in the later Chapters of this thesis, an atomic multipole model of electrostatics will be used. Therefore, the W99 force field was benchmarked with both multipoles (DMA) and atomic point charges.

Recently, a revision of the W99 parameters was published [225]. The pre-exponential Born-Mayer parameter of atom types involved in strong hydrogen bonds were re-parametrised to perform optimally when the force field is used together with multipoles. This lead to significant improvements in the geometry of hydrogen bonds. Different versions of the force field were parametrised for use with multipoles derived either from a ground state charge density of the isolated molecule (W99rev6311), or from a polarised charge density obtained by embedding the molecular calculation in a polarising environment (W99rev6311P).

In order to account for some polarization of the molecules in the crystal, it may be preferable to perform the molecular DFT calculations in a polarisable continuum model [399, 400, 401] (PCM). This should, to some extent, model an average polarisation of the molecular electron density [382]. Polarised multipoles can then be derived from the polarised charge density using the distributed multipole analysis. This method would be a mean field approximation of molecular polarisation in crystals, ignoring local polarisation of functional groups around hydrogen bonds. If a force field is parametrised specifically to be used together with such PCM-derived multipoles, the localized polarisation should partly be absorbed in the repulsion-dispersion force field parameters. This was attempted in parametrising the W99rev6311P force field. A relative permittivity of 3.0 was used in the polarisable continuum model. Here I present results for $\varepsilon_r = 3.0$ and 5.0.

The FIT force field [223] is also included. It is still used [5], but not much is known about its accuracy. The FIT force field is a revision of the W84 force field [91], and was re-parametrised to perform optimally with atomic multipole electrostatics.

The force fields investigated here were parametrised with foreshortened hydrogen positions, except the FIT force field, where foreshortening of hydrogen atoms was omitted in the re-parametrisation. Foreshortening moves the interaction centre of hydrogen atoms off the nucleus by 0.1 Å along the covalent bond towards the heavy atom to which the hydrogen is bonded. The distributed multipoles and point charges should be fitted at the foreshortened positions. The original reason for hydrogen foreshortening was that Williams used experimental structures from X-ray diffraction in the parametrisation. X-rays diffract of the electron density and this has a maximum at the foreshortened position, not the nuclear position. There is a good reason for keeping the foreshortening in

the force fields, as the centre for the repulsion and dispersion interactions is the charge density maxima.

The X23 set consists of 23 crystals of small organic molecules and temperature-free benchmark lattice energies. The crystal structures were collected from the Cambridge Structural Database (CSD) [304], previous benchmark articles [115, 214] and from primary sources [402, 403, 404]. Reilly & Tkatchenko added hexamine and succinic acid structures, but neglected to specify exactly which experimental structure they used. Using COMPACK [165] it was determined that the benchmark hexamine structure corresponds to the experimental structure HXMTAM09 [405] and the succinic acid structure is the monoclinic β polymorph, best represented by SUCACB02 [406].

3.2.2 Lattice energy calculations

The lattice energy consists of the intermolecular cohesive energy and an intramolecular relaxation energy, the difference in energy between the in-crystal geometry and the lowest energy (gas phase) geometry.

$$E_{\text{latt}} = E_{inter}^{\text{atom}-\text{atom}} + E_{intra}^{\text{DFT}} \tag{3.1}$$

Care was taken to use the correct lowest energy conformer of oxalic acid, succinic acid and cytosine. Molecular geometries and energies were calculated at the B3LYP / 6-311G(d,p) level of theory using Gaussian 09[407] revision D.01. The molecules are very small so a dispersion correction is not necessary. When polarised multipoles are used (for the W99rev6311P3 and W99rev6311P5 force fields), the relaxation energy also contains the energy due the electronic relaxation from the polarised to unpolarised gas phase electron density.

The intermolecular cohesive energy between any two molecules N and M was calculated as a sum over atom-atom pair interactions. Non-electrostatic interactions were modelled with a Buckingham type model potential. Electrostatic interactions were modelled using either atomic partial charges or distributed multipoles. Partial charges were fitted to the molecular electrostatic potential using the program Mulfit 2.1 [200]. Distributed multipoles [408] were derived from the molecular charge density using Gdma 2.2.06 [198].

The crystal structures (except carbon dioxide) were first lattice energy-minimised using the program CrystalOptimizer 2.4.2 [236, 161], using the W99rev6311 force field. CrystalOptimizer minimises the sum of intra- and intermolecular energies with respect to crystal packing degrees of freedom and a set of selected intramolecular degrees of freedom (dihedrals and bond angles). The intramolecular degrees of freedom were chosen automatically using the algorithm described in Appendix A. The optimised structures

were then re-optimised with the different intermolecular force fields, using DMACRYS 2.0.4, keeping the molecular geometries rigid.

For the W99rev6311P3 and W99rev6311P5 force fields, the multipoles were calculated from a charge density obtained in a polarisable continuum model (PCM) with isotropic relative permittivity of 3.0 and 5.0, respectively.

Carbon dioxide was treated specially. A molecular geometry-optimisation at the B3LYP / 6-311G** level of theory was performed in GAUSSIAN 09 and partial charges were calculated with the CHelpG routine [409]. Multipoles and a molecular local axis system were constructed by manually adding non-interacting mass-less dummy atoms equatorially around the carbon atom. This is necessary as DMACRYS requires an intramolecular axis system for the multipoles.

Charge-charge, charge-dipole and dipole-dipole interactions were calculated using Ewald summation, while repulsion-dispersion interactions and all higher multipole-multipole interactions were calculated between whole molecules to a center-of-mass cutoff distance of 20 Å.

3.2.3 Quasi-harmonic lattice dynamics calculations

Thermal expansion can significantly change the crystal structure at elevated temperatures and for high-accuracy calculations of crystal properties, this should not be neglected. Lattice vibrational free energy calculations in the quasi-harmonic approximation was implemented, so that the zero-point energy, vibrational frequencies and thermal expansion can be calculated. This allows an explicit modelling of thermal effects and facilitates the use of free energy as the scoring function in crystal structure prediction.

All lattice dynamics calculations were performed in the rigid-molecule approximation, as implemented in DMACRYS [261, 266], with the FIT and W99rev6311P5 force fields.

The Brillouin zone was sampled by forming several linear supercells, each elongated along one lattice vector. The supercell expansion in each direction was chosen such that the distance between the **k**-points was less than some target distance in reciprocal space. A target distance of 0.12 $Å^{-1}$ was used in the benchmarking described here. This results in between 17 and 34 unique **k**-points being sampled for the crystals in the benchmark, and quite well-converged calculations, see Figure 2.9. In addition to the supercell method, the Debye method and a Gaussian KDE method were also used to converge the free energy calculations.

As described in Chapter 2, the Debye approximation interpolates the phonon dispersion between the Brillouin zone centre and the nearest explicitly sampled **k**-point. A Debye frequency ω_D is calculated from the elastic stiffness tensor C_{ikjl} obtained from DMACRYS

[261, 114]. From the elastic tensor Christoffel matrices $\mathbf{\Gamma}$ are calculated for 13 direction cosine vectors $\hat{\mathbf{k}}_m$.

$$\mathbf{\Gamma}(\hat{\mathbf{k}}_m) = \sum_{k,l} C_{ikjl}\hat{\mathbf{k}}_m \tag{3.2}$$

From the eigenvalues of the Christoffel matrices, we obtain the phase velocity of sound ν in direction $\hat{\mathbf{k}}_m$.

$$|\mathbf{\Gamma}(\hat{\mathbf{k}}_m) - \rho\nu^2(\hat{\mathbf{k}}_m)\mathbf{I}| = 0 \tag{3.3}$$

The crystal's density is denoted ρ and the identity matrix $\mathbf{I}$. Average over the three eigenvalues produces a mean velocity of sound $\overline{\nu}(\hat{\mathbf{k}}_m)$. Assuming a linear phonon dispersion reduced by $2/\pi$ around Γ, the Debye frequency $\omega_D(\hat{\mathbf{k}}_m)$ is calculated from

$$\omega_D(\hat{\mathbf{k}}_m) = \frac{2\overline{\nu}(\hat{\mathbf{k}}_m)|\hat{\mathbf{k}}_m|}{\pi} \tag{3.4}$$

Averaging over the 13 directions yields the mean Debye frequency ω_D, which is an approximation of the average phonon frequency on an ellipsoid around the Γ-point, extending to the nearest sampled $\mathbf{k}$-points. Encouraged by the work of Karo, Hardy and Gilat [299, 288], we replaced each discrete phonon frequency with a narrow Gaussian distribution. The phonon density of states was approximated with the kernel density KDE(ω).

$$g(\omega) = \frac{6Z}{nh\sqrt{2\pi}} \sum_{i}^{n} \exp\Big(\frac{-(\omega_i - \omega)^2}{2h^2}\Big) \tag{3.5}$$

The free energy at constant volume was calculated as

$$A(T) = E_{\text{latt}} + F_{\text{vib}}(T) \tag{3.6}$$

The vibrational contribution to the free energy $F_{\text{vib}}(T)$ for one unit cell with rigid molecules, each with 6 degrees of freedom can then be calculated from the Debye frequency and the density of states $g(\omega)$ as:

$$\begin{aligned} F_{\text{vib}}(T) \; &= 3\Big(\frac{n}{n+3}\Big)\int_0^\infty \hbar\omega g(\omega)\, d\omega \\ &+ 6\Big(\frac{n}{n+3}\Big)k_BT\int_0^\infty \ln\Big(1-\exp\Big(\frac{-\hbar\omega}{k_BT}\Big)\Big)g(\omega)d\omega \\ &+ \frac{9\hbar\omega_D}{8N_{\text{k}}} \\ &+ \frac{3k_BT}{N_{\text{k}}}\ln\Big(1-\exp\Big(\frac{-\hbar\omega_D}{k_BT}\Big)\Big) \\ &- \frac{k_BT}{N_{\text{k}}}\text{D}\Big(\frac{\hbar\omega_D}{k_BT}\Big) \end{aligned} \tag{3.7}$$

where $\text{D}(x)$ is the Debye function

$$\text{D}(x) = \frac{3}{x^3}\int_0^x \frac{t^3}{\exp(t)-1}\, dt \tag{3.8}$$

Carbon dioxide has only 5 intermolecular degrees of freedom, and was treated separately.

Inspired by Otero-de-la-Roza [115], a thermal pressure method was chosen for the quasi-harmonic implementation. A thermal pressure [410] was calculated as a finite difference in $F_{\mathrm{vib}}(T)$ between unit cells with different volumes. An isotropic scaling of the unit cell volume for the thermal pressure calculation I believe can lead to errors, since molecular crystals can have strongly anisotropic thermal expansion.

Instead, it was decided to expand the unit cells by geometry-optimising the crystal structure at a pressure different from the ambient pressure P_0. I have chosen -300 MPa, taking a negative pressure as this should result in a crystal structure that is as similar to a thermally expanded structure as possible. The negative pressure causes a volume expansion between 2.4 and 7.7% in the benchmark crystal structures. The vibrational energy is highly linear for such small volume changes, facilitating a finite difference from only two volumes.

It was found that because of the stepwise convergence of lattice dynamics calculations with respect to the **k**-point sampling, it is imperative that the same **k**-points are sampled in both unit cell volumes. The thermal pressure is then [115, 410]

$$P_{\mathrm{th}}(T) = -\frac{\Delta F_{\mathrm{vib}}(T)}{\Delta V} \tag{3.9}$$

Again geometry-optimising to minimise $E_{\mathrm{latt}} + PV$ at $P = P_0 - P_{\mathrm{th}}$ results in a crystal structure close to the Gibbs free energy minimum at ambient pressure. The Gibbs free energy can then be calculated as

$$G(P_0, T) = E_{\mathrm{latt}}(P) + F_{\mathrm{vib}}(P, T) + P_0 V(P) \tag{3.10}$$

3.3 Benchmark results

3.3.1 Lattice energies

In Figure 3.1 a comparison of the performance of the different force fields in reproducing the X23 benchmark lattice energies is displayed. Table 3.1 shows the mean absolute deviation (MAD), mean absolute relative deviation (MA%D), the systematic mean deviation (MD) and the random error in the form of one standard deviation (SD). The figures and tables here are reproduced from Ref. [1].

All tested force fields systematically underbind the crystals. This is expected since the force fields were parametrised to thermally expanded experimental structures and sublimation enthalpies, as explained in Chapter 2. Expressed as a percentage mean

Energy model	MA%D	MAD	MD	SD
FIT	10.27	9.22	−7.95	8.63
W99rev6311P5	15.72	14.19	−13.97	9.58
W99rev6311P3	16.79	15.20	−14.99	10.18
W99_DMA	17.47	15.70	−15.52	11.27
W99rev6311	18.28	16.38	−16.21	10.70
W99_ESP	25.27	22.01	−20.75	13.92

Table 3.1: Error in lattice energy
Mean absolute relative deviation (MA%D), mean absolute deviation (MAD, kJ/mol), systematic error (MD, kJ/mol) and standard deviation (SD, kJ/mol) in benchmark lattice energies for the different force fields. A negative MD represents underbinding.

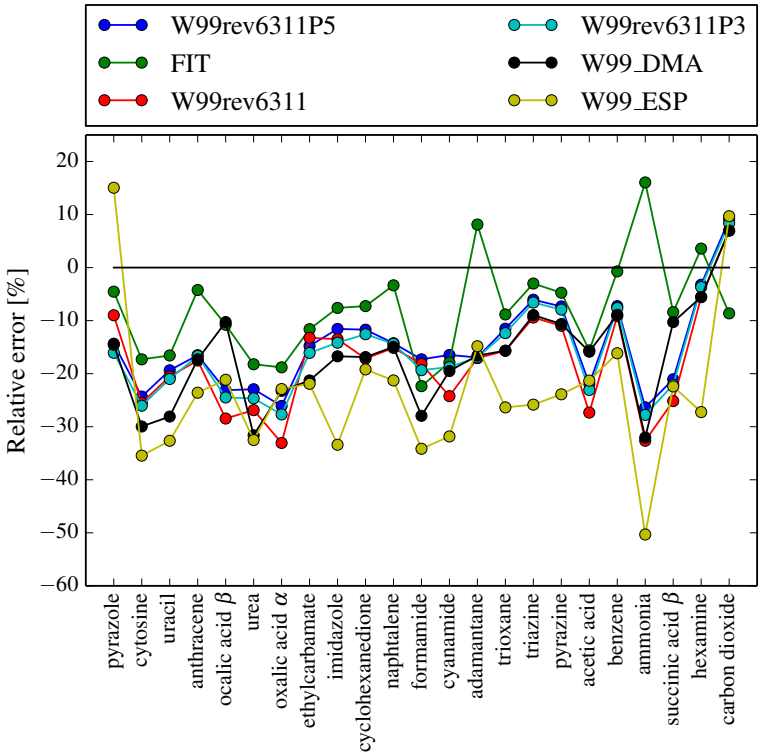


Figure 3.1: Relative errors in relative lattice energies calculated with different force fields.

signed error, we have between −7.9% and and −23.3% too small lattice energies for the various force fields.

The multipole electrostatics of W99_DMA significantly improves on the point charge model used in the W99_ESP force field. The errors are reduced by about a third, confirming the great advantage of using a multipole model for electrostatic intermolecular interactions.

The recent re-parametrisation of W99 [225] to W99rev6311 did not significantly improve the accuracy of lattice energies. An improvement was seen in how well the geometries of the structures were reproduced, especially in the directionality of hydrogen bonds. The

revised force fields utilizing polarised multipoles perform better than the non-polarised version since the polarisation generally increases the strength of directional electrostatic interactions, principally hydrogen bonds, and this reduces the systematic underbinding.

The best models in terms of lattice energy are the revised force fields with polarised multipoles (W99rev6311P5) and the FIT force field. The latter is the most accurate of the tested force fields, and it performs well both with respect to systematic and random errors.

The calculated lattice energies depends partly on the level of theory used for the intramolecular energies. However, for most crystal structures studied in this thesis, the in-crystal and vacuum molecular geometries differ only slightly. Because of this, errors in the intramolecular energies are a small contribution to overall errors.

In Table 3.2 and Figure 3.2 I compare the two best force fields with several DFT-D methods previously benchmarked against the X23 or C21 sets. For published C21 results, the errors were recalculated relative to the X23 benchmark lattice energies. Benchmark results for hexamine and succinic acid in the PBE-XDM and PBE-D2 have not been published, so such calculations were performed separately by Graeme Day using QuantumEspresso [365].

Energy model	MA%D	MAD	MD	SD
TPSS-D3 [393]	5.21	3.84	0.71	5.13
PBE-D3 [393]	5.75	4.47	1.63	5.65
HSE06-D3 [393]	6.29	4.80	2.67	4.11
PBE0-D3 [393]	6.43	4.67	2.20	5.85
PBE-XDM [115]	7.68	6.51	−2.54	8.67
PBEh-3c* [395]	7.8	5.4	0.4	7.0
PBE-MBD [214]	8.05	5.92	4.73	5.14
PBE-D2 [115]	9.96	7.83	6.24	8.15
FIT	10.27	9.22	−7.95	8.63
DFTB-D3 [396]	11.92	9.86	−0.32	12.22
W99rev6311P5	15.72	14.19	−13.97	9.58
PBE-TS [214]	17.22	13.40	13.13	8.62
B3LYP [395]	34.7	27.2	−24.6	28.0

Table 3.2: Mean absolute relative deviation (MA%D), mean absolute deviation (MAD, kJ/mol), mean deviation (MD, kJ/mol) and standard deviation (SD, kJ/mol) in benchmark lattice energies for FIT, W99rev6311P5 and several dispersion corrected DFT methods. *Anthracene and naphthalene not included.

The FIT and W99rev6311P5 force fields reproduce lattice energies better than PBE-TS and B3LYP without dispersion correction. The random error, which is the most important for CSP purposes, is of the same magnitude for most of the benchmarked methods.

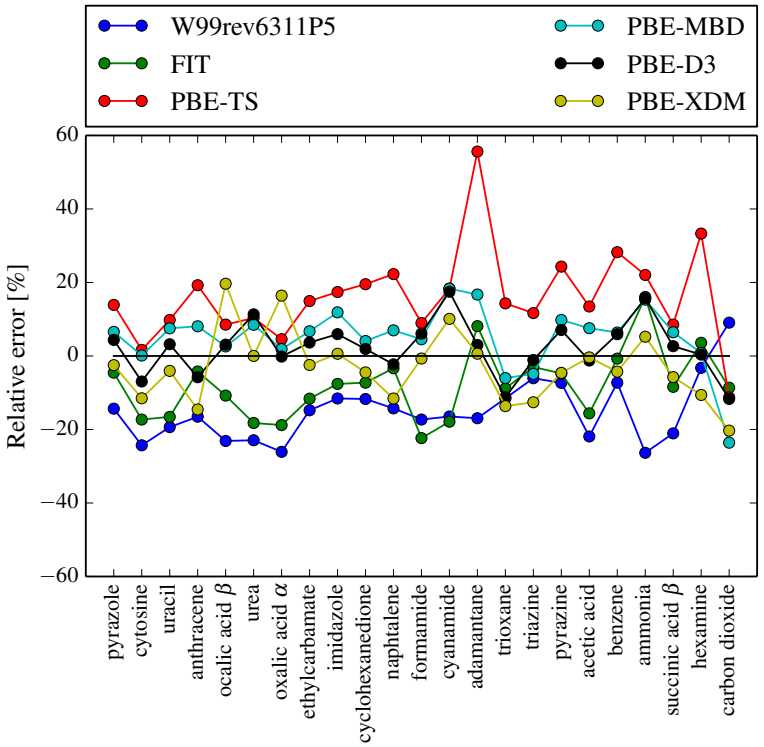


Figure 3.2: Relative errors in relative lattice energies calculated with the FIT and W99r6311P5 force fields and the PBE density functional with different dispersion corrections.

Most solid state calculations on molecular crystals are performed using GGA functionals, such as PBE. This is due to the large cost of evaluating exact exchange in a plane-wave basis sets. Consequently, hybrid functionals are rarely an option for periodic calculations.

Only when used together with the most accurate dispersion correction methods: exchange-hole dipole (XDM), many-body dispersion (MBD) and Grimme's D3 method, is the PBE functional significantly better than the force fields we have tested here. We note that the force field-like dispersion correction used in these DFT methods has a more complex form than what is used in the current force fields: XDM uses a sum of C_6, C_8 and C_{10} dispersion terms, while MBD includes non-pairwise dispersion contributions. It seems reasonable to expect that the future development of force fields including such terms could reduce the gap between the current force fields and the best DFT-D methods.

The hybrid functionals HSE06 and PBE0 and the meta-GGA functional TPSS together with D3 or many-body dispersion methods are also significantly better than the force fields [393]. However, even the best of the DFT-D methods described here have errors that are only about 2 or 3 times smaller than the errors of the best force fields.

3.3.2 Lattice parameters and geometries

In Table 3.3 we compare the performance of the force fields in reproducing experimental lattice parameters, without any thermal adjustments. Symmetry-independent lattice vector lengths for 22 of the crystal structures are used, excluding urea. With the force field models, the experimental tetragonal $P\bar{4}2_1m$ urea structure (UREAXX02) is found to be a vibrationally averaged structure corresponding to a saddle point on the potential energy surface [261]. Breaking the symmetry results in a stable orthorhombic $P2_12_12$ structure. This prevents a direct comparison of lattice parameters and elastic constants for this structure.

Energy model	MAD [Å]	MA%D [%]
PBE-D2 [115]	0.10	1.22
PBE-TS* [115]	0.10	1.58
B86b-XDM* [115]	0.12	1.76
TPSS-D3 [393]	0.13	1.73
FIT	0.17	2.48
PBE-XDM [115]	0.19	2.64
W99rev6311	0.20	2.80
W99_DMA	0.21	2.94
W99rev6311P3	0.22	3.02
W99rev6311P5	0.22	3.02
W99_ESP	0.41	5.67

Table 3.3: Mean absolute deviation (MAD) and mean absolute relative deviation (MA%D) in lattice vector length for the different force fields and a selection of DFT-D methods. The urea structure was not included in the force field calculations. *Hexamine and succinic acid not included.

The FIT force field performs very well in reproducing the experimental lattice parameters. It is comparable in accuracy to PBE-XDM. The B86b-XDM functional, PBE-TS and PBE-D2 are better than the force field methods in reproducing lattice parameters. The W99rev6311P5 force field also reproduces the crystal structures well, with the exception of benzene and naphthalene. The W99-based force fields do not perform well for crystals with aromatic ring systems in T-stacked configuration [390], see Table S3. However, systems in π-stacked configurations are well reproduced [387]. The point charge force field W99_ESP performs significantly worse than the other methods and the errors in lattice parameters are about twice as large as the multipole-based force fields.

The errors are not systematic; the mean error in lattice vector lengths are merely -0.03 Å and $+0.04$ Å for the FIT and W99rev6311P5 force fields respectively.

PBE-TS reproduces crystal geometries very well [249, 115], but not lattice energies [214]. In Chapter 7 PBE-TS is used to optimize clathrate crystal structures for NMR calculations, where structural accuracy is of utmost importance.

Table 3.4: Calculated and experimentally observed volumetric thermal expansion coefficients in units of 10^6 K^{-1}. The coefficients are calculated as a linear thermal expansion between 0 to 350 K for predictions and over the indicated temperature range for experimental data.

Structure	Calc. α	Exp. α	Temp. range [K]
MNPHOL mc	211.9	214.6 ± 14.9	95–350
ETDIAM Iα	270.5	193.0 ± 38.7	130–274
FURHUV β	187.3	154.6 ± 24.6	100–218
GLURAC β	164.6	206.6 ± 21.5	120–373
HXACAN I	154.4	156.9 ± 15.3	20–330
ADIPAC I	173.8	198.4 ± 11.3	100–293
DPYRAM mc	178.9	164.8	150–293
PYRZIN β	155.3	181.1 ± 12.3	90–293

3.3.3 Thermal expansion

For the FIT force field, the deviations in X23 lattice vector lengths for the thermally expanded structures are MAD = 0.188 Å and MA%D = 2.82%. The W99rev6311P5 force field yields MAD = 0.264 Å and MA%D = 3.89%. After thermal expansion, experimental crystal structure geometries are reproduced with an average $RMSD_{20}$ of 0.234 Å and 0.314 Å for the two force fields respectively.

The thermal expansion rates can also be compared to published experimental data. Crystal structures of the following compounds were used: 1,2-ethanediamine (ETDIAM) [411], *m*-nitrophenol (MNPHOL) [412, 413, 414], paracetamol (HXACAN) [415, 416], glutaric acid (GLURAC) [417, 418, 419], octa-sulfur (FURHUV) [420, 421, 422], adipic acid (ADIPAC) [423, 424], 2,2′-dipyridylamine (DPYRAM) [425] and pyrazine-2-carboxamide (PYRZIN) [426]. Extensive experimental reference data is available for these crystal structures.

In Table 3.4, the calculated volumetric thermal lattice expansion of these structures are compared to experimental variable temperature X-ray diffraction data. Calculated thermal expansion were taken as a finite difference between 0 and 350 K (except ETDIAM, for which 250 K was used). The experimental thermal expansions were calculated from a linear regression of experimental data over the indicated temperature intervals. The mean absolute relative error in calculated thermal expansion coefficients is estimated to be 15%. The calculations do not systematically over- or underestimate the thermal expansion, the mean signed error is merely 2.5%. These results have previously been published in reference [3].

3.4 Concluding remarks

This chapter demonstrates the overall reasonable accuracy of the computational methods employed in this thesis. Lattice energies, lattice parameters and crystal structures have been used to benchmark several force fields against the X23 set and electronic structure methods. In addition, thermal expansion has been benchmarked, demonstrating the soundness of the implemented quasi-harmonic lattice dynamical method.

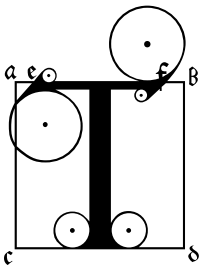
Since the force fields were parametrised to reproduce experimental crystal structures, the explicit modelling of thermal expansion here constitutes a double counting of some of the thermal expansion, although low temperature (< 100 K) crystal structures were deliberately chosen in re-parametrising W99 to minimise this effect. This should cause a systematic positive error in the lattice vector lengths. However, the mean errors are only +0.14 Å and +0.05 Å for the W99rev6311P5 and the FIT force fields respectively. Again the FIT force field performs very well and does not have a large systematic overestimation of unit cell sizes.

It seems reasonable to expect that future development of force fields including higher order dispersion terms, short-range damping of the dispersion and many-body interactions could reduce the gap between the current force fields and DFT-D methods, offering highly accurate calculations are small computational costs. Such methods would be highly attractive for CSP applications beyond what is currently feasible, such as for large molecules or multicomponent crystals.

These results confirm the soundness of the computationally very efficient thermal pressure method and that the error caused by the double-counting of implicit and explicit thermal expansion is negligible.

Chapter 4

Free energy as scoring function in the sixth blind test



The sixth blind test of crystal structure prediction was announced in 2014. Research groups were openly invited to participate by performing one or several blind predictions of crystal structures of five selected compounds [427]. No less than 25 groups participated, demonstrating the growing interest in CSP methods. The blind tests have become the most important, unbiased and fair assessment of the methods used for crystal structure prediction. The results of the sixth test were published in 2016 [5].

4.1 Background

The Day group participated and submitted predictions for all target molecules [5]. The calculations were performed in close collaboration between group members. The conformational search, structure generation and rigid-body geometry optimisations of molecules XXII and XXV were primarily performed by Angeles Pulido and David McMahon. The author performed lattice dynamical free energy calculations on the predicted polymorphs. These predictions are discussed in this Chapter, with a special focus on the experiences of using the lattice-vibrational free energy as scoring function in the blind test predictions.

In the end, the methods and results of other groups that also used estimates of the free energy as scoring functions in the sixth blind test are also discussed. The chapter ends with some remarks about lessons to be learned from this communal exercise.

4.2 Structure prediction and lattice dynamics

4.2.1 Molecular geometries and conformer searches

4.2.1.1 Conformations of XXII

Molecular conformers of target molecule XXII (tricyano-1,4-dithiino[c]-isothiazole) and target XXV (a 1:1 co-crystal of 3,5-dinitrobenzoic acid and 2,8-dimethyl6H,12H-5,11-methanodibenzo[b,f][1,5]diazocine) were generated with a normal mode conformational search method [151, 152], as implemented in MacroModel [428]. Due to the large number of energy evaluations and minimisations required for an exhaustive search, the initial conformer search was performed using a force field method. The OPLS2005 force field [429, 89] was used to calculate the energy of molecular conformations and the normal modes of conformers. The OPLS2005 force field is a variant of the OPLS-AA force field [430]. Duplicate molecular geometries were removed if the all-atom RMS deviation of atomic positions was smaller than 0.02 Å.

All unique conformers within 50 kJ/mol of the force field global minimum were then geometry optimised with DFT-D in Gaussian 09 [407] using the B3LYP [244, 243] density functional. The McLean-Chandler basis set [240] was used for sulfur and 6-311G** for the carbon atoms. A dispersion energy correction was included using Grimme's D3 method and short-range damping [247, 99]. The DFT-D optimised conformers were clustered to remove duplicates, using all-atom RMS deviations of atomic positions of 0.4 Å as selection criteria.

For target molecule XXII, two symmetry-related, buckled molecular conformers were found in the force field conformational search. A planar molecular conformation was also identified as a saddle point between the two minima. After DFT-D geometry optimisation, the saddle point was found to have an energy of 5.7 kJ/mol relative to the lowest-energy minimum.

4.2.1.2 Conformations of XXV

The molecules in target XXV are essentially rigid, so no conformational search was performed. Instead suitable conformers were chosen manually. Three conformers of the 3,5-dinitrobenzoic acid with different orientations of the carboxyl group were used together with a single conformer of 2,8-dimethyl6H,12H-5,11-methanodibenzo[b,f][1,5]diazocine. B3LYP-D3 / 6-311G**-optimal geometries were used.

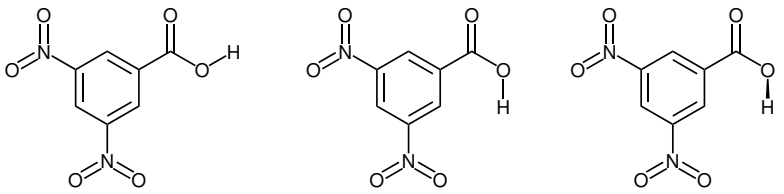


Figure 4.1: 3,5-dinitrobenzoic acid conformations A, B and C.

4.2.2 Crystal Structure generation

The Global Lattice Energy Explorer program was used to generate trial crystal structures [160]. The program essentially implements the following algorithm. For each structure, the degrees of freedom to be sampled are determined by mapping a quasi-random vector, drawn from a Sobol sequence [159], onto the relevant parameters. The positions of the (centre of mass) of the molecules in the asymmetric unit are simply random linear translations. Orientations are sampled according to Shoemake's method [431], ensuring a uniform sampling of Euler angles on the sphere.

Unit cell angles that are not fixed by the space group are also randomly sampled. A random number is mapped through the arccosine function to angles in the interval [35°, 145°]. To sample unit cell lengths, a nominal 'volume' is assigned to each molecule by projecting the atomic positions onto the eigenvectors of the moments-of-inertia matrix. A cuboid volume is determined from the largest and smallest atomic coordinates along each axis. The volume of the sum of the cuboids of each molecule in the asymmetric unit is multiplied by the number of symmetry elements of the space group to yield a target volume for the unit cell.

In order to generate crystal structures in which the molecules do not overlap, a convex hull of each molecule is constructed, and the unit cell dimensions are expanded to relieve molecular clashes up to a maximum volume of 2.5 times the target volume. Overlaps of the molecular convex hulls were checked using the separating axis theorem, which also provides the direction for expansion of the unit cell. See reference [160] for more details of how the structure generator program works.

As is seen below, flexible molecules with several conformers and multicomponent crystals with unknown stoichiometry require very extensive sampling of the available phase space, as the size of the search space increases geometrically with respect to the number of such degrees of freedom.

4.2.2.1 Structure generation of XXII

For molecule XXII, the buckled conformer is chiral. The 94 most common space groups were sampled, including both space groups from enantiomorphic space groups. The 25

most common space groups [304, 432] had 5000 valid structures generated in each, with 2000 structures in the remaining 69. The planar conformer is not chiral, so only one of each chiral space group pair needs sampling, resulting in 87 space groups.

4.2.2.2 Structure generation of XXV

The three conformers of the acid component of target XXV were used in combination with the single conformer of the base compound. Only 1:1 stoichiometry co-crystals were generated, as this stoichiometry was given [427].

For the lowest energy gas-phase conformer of the acid component (labelled A), 10000 structures were generated for the co-crystal in the 25 most common space groups. An additional 10000 structures were generated in space groups 2, 7, 9, 14, 15, 19, 29, 33, 43, 60, 61, as these space groups tended to yield particularly low-energy crystal structures.

For the higher-energy conformers of the acid (B and C), 10000 structures were generated for the co-crystal in the 25 most common space groups.

Conformer C of the acid and the base are both chiral. This necessitated generating structures in the subset of those 25 space groups which are enantiomorphic (19, 4, 5, 1, 18, 76, 96, 92, 78, 144, 145) using both possible enantiomers of C. 10000 structures for each pair of enantiomers were generated in each enantiomorphic space group.

4.2.3 Crystal structure geometry optimisation

Rigid-body lattice-energy minimisations were performed with DMACRYS. Distributed multipoles up to rank 4 on all atoms were derived from the DFT charge density. The W99rev_6311 force field was used [433, 225, 2].

4.2.3.1 Flexible molecule optimisation of XXII

After rigid body lattice energy minimisation, flexible molecule lattice energy minimisation was performed with CRYSTALOPTIMIZER. For molecule XXII, calculations were performed on a subset of crystal structures within an energy window of 15 kJ/mol above the minimum energy structure of each conformer. During flexible molecule lattice energy minimisation, all bond lengths and angles (except linear angles in the cyano-groups) were allowed to relax to their DFT equilibrium values. The cyano-groups were constrained to an angle of 179.99°.

After flexible-molecule geometry-optimisation, duplicate crystal structures were removed and a total of 753 unique XXII polymorphs were obtained.

4.2.3.2 Flexible molecule optimisation of XXV

For target XXV, structures containing conformer A of 3,5-dinitrobenzoic acid, the lowest 20 kJ/mol were optimised with CrystalOptimizer. For structures containing conformers B and C of the 3,5-dinitrobenzoic acid only the lowest 10 kJ/mol of structures were optimised further.

The base is essentially rigid, but has two methyl groups that were allowed to rotate in response to packing forces. In the acid molecule, the nitro- and carboxyl groups were allowed to rotate, and the angle and dihedral of the carboxy-hydroxyl group were also allowed to relax.

The final list of predicted structures contains 239 structures, 208 of which are below 20 kJ/mol lattice energy and 225 are below 20 kJ/mol free energy. This means that the free energy landscape is 'compressed' compared to the corresponding lattice-energy landscape, a result that has been seen elsewhere [434].

4.2.4 Free energy calculations

The Helmholtz free energy was calculated at 300 K for all predicted structures of molecules XXII and XXV. The harmonic approximation lattice dynamics method implemented in Dmacrys was used [168, 114].

The co-prime split linear supercell method described in Chapter 2 was used with a target **k**-point distance of 0.1 $Å^{-1}$ resulting in an average of 21 unique **k**-points being sampled. The minimum number of **k**-points sampled in any one structure was 12, maximum 33.

Crystal structures with highly anisotropic dimensions were converted to reduced Niggli cells with Platon's ADDSYMM routine [284, 281, 435, 282]. This enhances the sampling of the Brillouin zone and reduces the computational cost of lattice dynamics calculations.

Atom-centred distributed multipoles [85, 191, 137] were calculated from the DFT charge density, calculated in a polarisable continuum model [401, 399]. The W99rev_6311P3 force field was used for all lattice dynamics calculations [2].

Many of the predicted crystal structures were initially found to be located at saddle points on the potential energy surface, yielding imaginary phonon frequencies. Such structures were symmetry-reduced to $P1$ [113], *i.e.* all point group symmetry was removed. If the $P1$ structure was also unstable, indicating that the structure is unstable with respect to translational symmetry breaking [436, 437], the unit cell was doubled and calculations were performed on the $1 \times 1 \times 2$, $1 \times 2 \times 1$ and $2 \times 1 \times 1$ supercells. The space group in which the structure was first generated can therefore differ from the space group of the final optimised structure.

The fact that unstable structures corresponding to saddle points on the potential energy surface are so commonly generated in crystal structure prediction is one of several possible reasons why CSP yields many more crystal structures than are observed experimentally [22]. One possible reason is that approximate quasi-Newton-Raphson algorithms are typically used in geometry-optimisations [438, 439, 168]. The identification and exclusion of unstable structures based on imaginary phonon frequencies has been advocated earlier [384, 140].

4.2.5 Final ranking of structures

Two lists of 100 structures, ranked by lattice- and free energy respectively were submitted to the blind test. The reason for submitting these two lists were to see if the free energy list would have a better predictive ability.

As this blind test category was stated to be $Z' = 1$ [5], the resulting $Z' > 1$ structures that result from symmetry breaking (see above) are not expected to match the observed structure, whereas they do constitute true structures on the lattice energy surface.

4.3 Results and discussion

4.3.1 Prediction results for target XXII

Once all participating groups had submitted their predictions, the experimental data were revealed by the test organisers. The predictions were assessed by whether they contained the experimentally observed crystal structure, how well the geometry of the structure had been predicted and the predicted stability [5].

The predicted crystal energy landscapes of target XXII are shown in Figures 4.2 and 4.3.

The experimentally known crystal structure of XXII (tricyano-1,4-dithiino[c]-isothiazole) was revealed after the predictions had been submitted. The experimental structure is believed to be the thermodynamically stable form at ambient conditions. In our predicted landscapes, the experimental structure was found ranked third, merely 0.65 kJ/mol above the lowest-energy structure by lattice energy. The correct structure was found as the global minimum by Helmholtz free energy. Hence, in this case, including the vibrational energy in the scoring function improved the predictive ability. The predicted and experimentally observed crystal structures differed by $\mathrm{RMSD}_{20} = 0.267$ Å.

The Kendall's τ_b rank correlation [440] between the lists ranked by lattice energy and free energy is 0.83 for the structures below 20 kJ/mol lattice energy. The correlation between the two lists is shown in Figure 4.4.

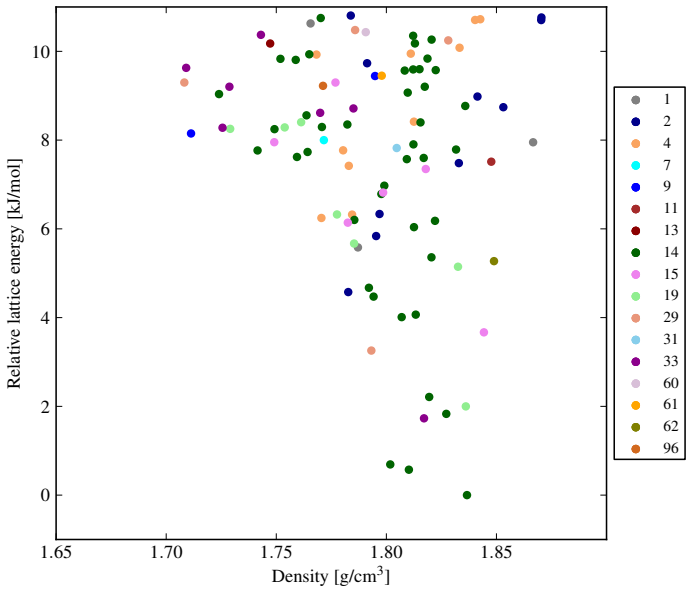


Figure 4.2: Lattice energy landscape of target XXII. Only the lowest 100 structures are included. Data labels denote space group number.

4.3.2 Prediction results for target XXV

The original experimental data for target XXV were collected at room temperature. The structure was remeasured after the blind test at 100 K, which revealed that there is substantial disorder or proton transfer from the carboxylic acid group to the base's amine group [5].

Such charge transfer is not modelled with the methods used here. In the prediction calculations, the molecules were assumed to be neutral, as this was the information originally stated in the announcement of the blind test. It is however an important lesson. The aim of CSP is the *ab initio* prediction of crystal structures, without any prior experimental knowledge. This was hardly possible for this structure. Groups that used quantum-chemical methods that are able to reproduce the proton transfer performed better in this case.

Energy landscapes for target XXV are presented in Figures 4.5 and 4.6. We were not able to predict the structure of target XXV. Subsequent analysis has revealed that the structure generator did indeed generate a trial structure that geometry-optimises towards the experimental structure, but that trial structure was prematurely removed as a candidate during clustering. Clustering of large numbers of structures is known to be a hard problem [140].

If the experimental structure had been retained, it would have been ranked as the 8th structure by lattice energy, 5.5 kJ/mol above global lattice energy minimum. In Figure

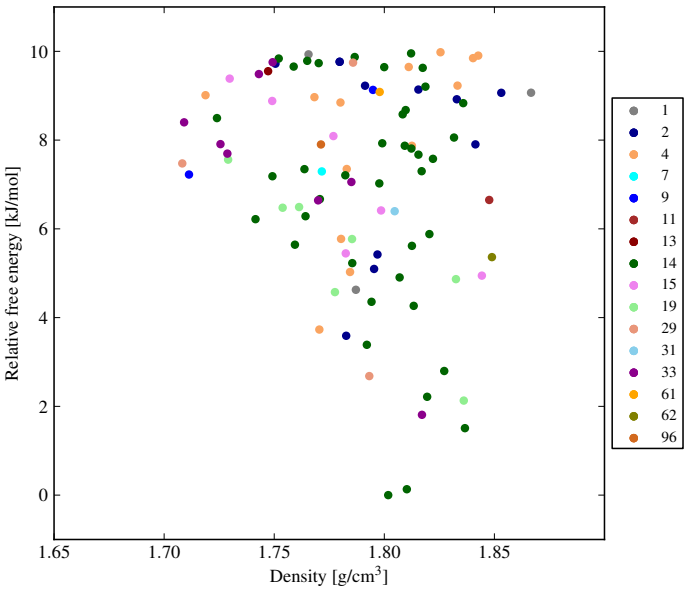


Figure 4.3: Free energy landscape of target XXII at 300 K. Only the 100 lowest ranked structures included. Labels denote space group number.

4.6, the experimental structure would have been located in the red circle, ranked no 15 and 6.1 kJ/mol above the predicted free energy minimum. That is, in this case including the vibrational energy did not improve the predicted stability ranking. Errors of about 5 kJ/mol in the relative free energy between polymorphs are to be expected.

The lattice vibrational energy causes some re-ranking among the predicted structures. The structures below 20 kJ/mol in lattice energy were ranked by lattice energy and free energy. The Kendall's τ_b rank correlation [440] between these lists is 0.8, indicating that one fifth of all polymorph pairs are ranked discordantly when ranked by free energy instead of lattice energy.

4.4 Other submissions using free energies

A few other groups submitted predictions using lattice-vibrational free energies as scoring function.

The Pickard group submitted predictions for target XXII. They used plane-wave basis PBE with the TS and MBD dispersion corrections. Zero point and vibrational energies were calculated at $\mathbf{k} = \mathbf{0}$ for a handful of crystal structures. It seems $\mathbf{k}$-point sampling for phonon modes was not considered. Anharmonic contributions were also calculated for one structure. It was found that the vibrational contributions did not substantially alter the ranking, but that many-body dispersion may have a significant effect.

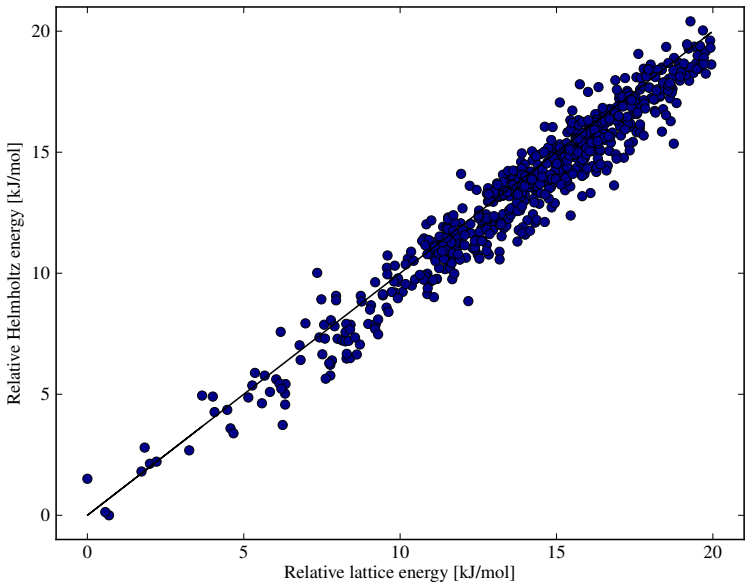


Figure 4.4: The free energy and the lattice energy are strongly correlated for target XXII, as expected.

The Price group used methods similar to the ones described here. Electrostatic multipoles were derived from a hybrid-functional PBE0/6-31G** charge density in a polarizable continuum model. The FIT force field was used, despite it not being parametrised to be used with polarized multipoles. Rigid-body $\mathbf{k} = \mathbf{0}$ lattice dynamics calculations were performed with DMACRYS. Phonon dispersion of acoustic modes was modelled with the Debye approximation.

The Price group confirms the re-ranking of the experimental XXII structure by vibrational energy, the ranking of the correct structure was improved when this was accounted for. The same kind of free energy calculations were also performed for targets XXIII, XXV and XXVI. This group also observed Born instabilities, with about 10% of the predicted structures exhibiting imaginary phonons. The correct structure of form XXV was successfully predicted, although here the vibrational energy was inconsequential to this result [5].

The Tkatchenko group performed lattice dynamics in DFT-D. 19 structures of target XXIII and 20 structures of XXII were subjected to harmonic approximation $\mathbf{k} = \mathbf{0}$ phonon calculations at the PBE-TS level of theory. The vibrational energy was added to PBE-MBD static energies. No attempt was made to account for phonon dispersion. Their results confirm the re-ranking of the stable polymorph of XXII. The thermodynamically stable form was correctly ranked with the free energy, and ranked third by static energy. The calculations were quite costly, the 19 structures of target XXIII required 1.9 million CPU-hours for harmonic lattice dynamics calculations [5].

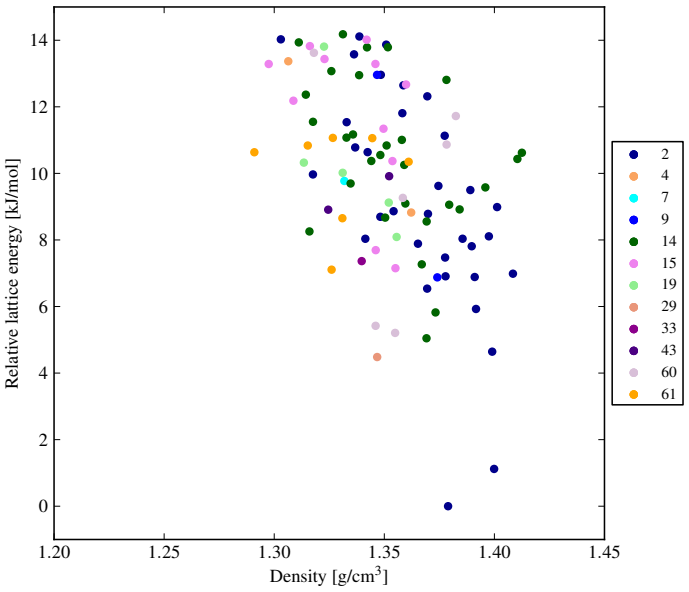


Figure 4.5: Lattice energy landscape of target XXV. Only the lowest 100 structures are shown. The labels are space group numbers.

4.5 Concluding remarks

The lattice vibrational energy can sometimes cause important differences in the ranking of polymorphs [300, 1]. Because of this, it is likely that the free energy is a better scoring function in crystal structure predictions that the static lattice energy.

In the sixth blind test of crystal structure prediction, we performed harmonic approximation lattice dynamics calculations on targets XXII and XXV. It was confirmed that crystal landscapes are substantially re-ranked when the lattice-vibrational energy contribution is accounted for. The Kendall rank correlation between the lists ordered by lattice- and free energies are 0.85 and 0.8 respectively. That is, the lattice energy accounts for 80–85% of the free energy variance at 300 K.

Several groups used free energy as the final scoring function in the blind test. Remarkably, all groups that calculated free energies found that the experimental polymorph of XXII was stabilized by entropy in such a way that it was the free energy minimum, but not the lattice energy minimum. This is probably the first example where we can definitively conclude that the vibrational energy causes a re-ranking of the experimentally known structure. This is strong evidence in support of using free energy as scoring function in blind structure predictions.

A couple of groups performed lattice dynamics calculations at DFT-D level of theory. These suffer from several problems. The computational cost was so large that only small subsets of the predicted landscapes could be ranked by free energies. The Tkatchenko

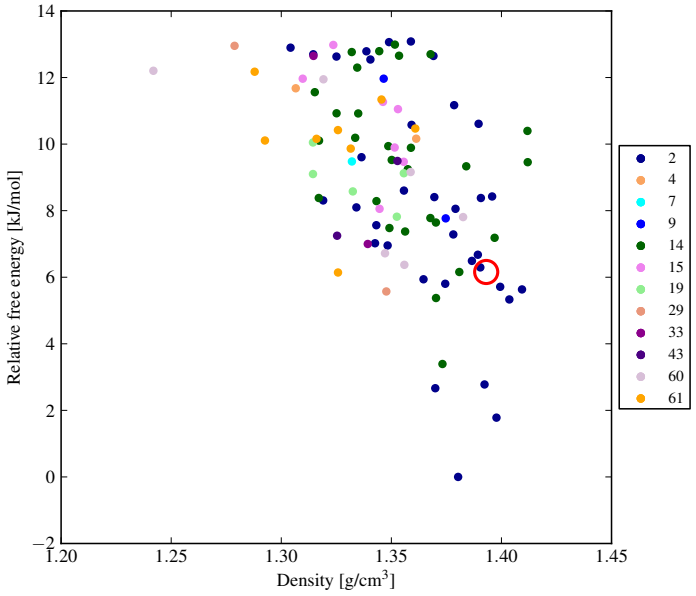


Figure 4.6: Free energy landscape of target XXV. Only the 100 lowest energy structures are shown. Labels denote space group number. The position of the experimental structure, which was not found, is encircled in red.

group used 1.9 million CPU-hours to re-rank 19 crystal structures. None of the groups that used DFT-D for lattice dynamics attempted to account for phonon dispersion.

The experimentally observed thermodynamically stable crystal structure of tricyano-1,4-dithiino[c]-isothiazole was successfully predicted.

The structure of target XXV could not be predicted however. This is likely due to three difficulties. The structure is a co-crystal, making the search space far larger. The experimental structure displayed significant proton transfer, which was not considered in the computational models. A flaw in how crystal structures were compared may have caused trial structures to be erroneously rejected.

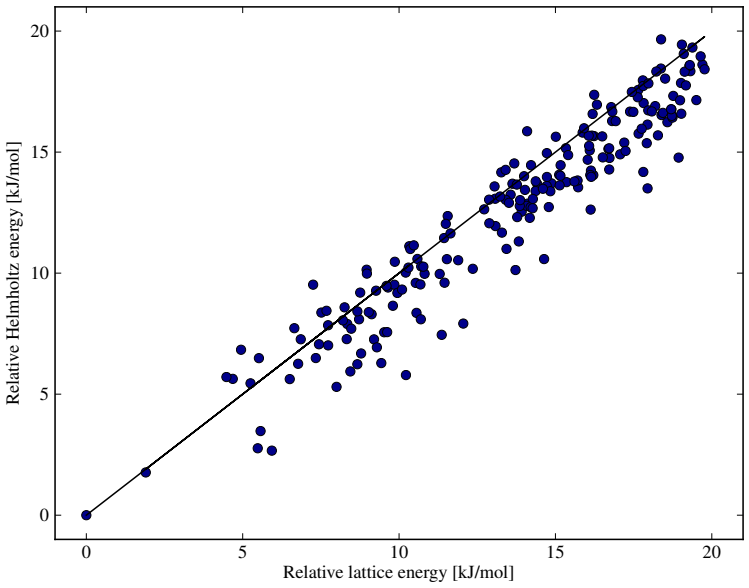
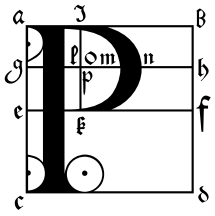


Figure 4.7: Rank correlation between relative lattice- and free energies of target XXV. Only structures below 20 kJ/mol shown. The diagonal line has unit slope.

Chapter 5

Differences between polymorphs



olymorphs and their properties are of great importance for the development of several classes of materials. Pharmaceuticals, organic semiconductors, chocolate, explosives and pigments are all greatly affected by polymorphism [20]. Unexpected polymorphism can have far-reaching consequences and affect economic, medical [28, 29], and intellectual rights [39, 40] aspects of the product. Because of this, we are interested in the properties of organic molecular crystals in general and in the differences between polymorphs in particular.

5.1 Background

We want to know how much polymorphs can differ in (free) energy and in physical properties, and how these differences change with temperature. This is of crucial importance for pharmaceuticals. The relative difference in free energy directly relates to the difference in solubility and bioavailability between polymorphs. This was the reason for the costly failure of ritonavir and rotigotine [29, 38, 35, 28]. A thermodynamically more stable form unexpectedly crystallized and this new form was significantly less soluble, requiring a reformulation of the drug.

In computational crystal structure prediction, a large number of plausible crystal forms are generated and ranked by a scoring function, increasingly often the free energy [441, 141]. The relative free energy determines which polymorph is thermodynamically favourable under the given thermodynamic conditions. However, crystallisation does not always lead to the most stable structure, but kinetic factors sometimes cause an alternative crystal packing to crystallize before, or faster than, the stable form. How large an energy penalty can be overcome by such kinetic factors? To answer this, and in order to interpret predicted landscapes and select the most likely crystal forms, it would be helpful to know by how much polymorphs generally differ in free energy.

This chapter is primarily about the overall trends in properties of polymorphic crystals. I determine how common it is that polymorphs are re-ranked by temperature and how common enantiotropic polymorphism is. Highly resolved distributions of the properties, and property differences, of polymorphic crystals are determined. In addition, I study a number of compounds in greater detail, calculating full free energy curves, the phonon densities of state and the thermal expansion. The chapter is based on work that has previously been published as references [1] and [2]. I present a number of examples of polymorph systems with possibly diverging free energy curves that were identified during the calculations. In the end I draw some conclusions about possible implications for the study of polymorphism.

Polymorph energy differences are generally assumed to be less than 10 kJ/mol [20, 23], based on experience from experimental determinations of polymorph relative energies, for example from sublimation enthalpies [392], melting data [322], differential scanning calorimetry [442], and relative solubilities [443]. However, experimental collection of a large quantity of high quality polymorph energy differences is hardly feasible. Computational modelling provides an alternative approach to obtaining polymorph energy differences.

In computational studies of molecular crystals, crystal structure prediction in particular, the free energy is often approximated with the static lattice energy, calculated either with force field or electronic structure methods. While relatively accurate lattice energies can be achieved with electronic structure methods, this ignores the energetic contributions related to the vibration of atoms and molecules about their equilibrium positions. Omitting the vibrational contribution to the free energy is convenient, since it greatly simplifies calculations, and tempting from the computational cost point of view. Neglecting thermal contributions has previously been justified by a claim that the difference in vibrational energy between polymorphs is so small that it never, or rarely, causes a re-ranking of the relative stability [23].

That the vibrational energy does not re-rank the order of stability between polymorphs was one of the conclusions of Gavezzotti's and Filippini's seminal work from 1995 on polymorph energy differences [23]. They calculated harmonic approximation free energy differences in 204 polymorph pairs using a simple force field. Their conclusion however, stands in contradiction to the many experimentally known temperature-dependent polymorphic transitions. There are also later computational studies that have demonstrated the importance of vibrational zero-point energy- and entropy-differences between crystal structures, principally by van Eijck [300]. Very little is known about how common enantiotropism is. What fraction of polymorph pairs are monotropic or have free energy curves that cross?

Mechanical properties of crystals are also important. For compressing a micro-crystalline powder into tablets, the elastic properties are crucial. Polymorphs of pharmaceutical

compounds sometimes differ significantly, making one polymorph preferable because of its better tabletability [27]. Because of this, I also look at how the bulk and shear moduli differ between polymorphs and how they change with temperature. Again, the focus is on the general trends.

Very few studies have been made of the anharmonic vibrational contribution to polymorph free energy ranking. The study of polymorphs of glycol and glycerol by van Eijck [300] is unique. In examining their influence on the ranking of predicted structures, van Eijck found that the harmonic approximation vibrational zero point- and thermal vibrational energy can have an important effect, but that the energetic impact of thermal expansion generally does not affect rankings. Because of the large computational cost and the small size of the anharmonic effect it was concluded that it may be preferable to neglect thermal expansion in CSP.

More recently, Heit and Beran have applied high level of theory fragment-based quantum chemical methods to investigate the importance of thermal expansion in modelling the thermodynamic properties of four crystal structures of small molecules [280]. The errors in free energies that they associated with ignoring thermal expansion were small, due to a cancellation of effects between the increasing enthalpy and entropy, which have opposite effects on the total free energy. Because of the huge computational cost, only a handful of relatively simple systems were accessible with this method.

The number of molecular crystal structures for which the importance of thermal expansion has been assessed is consequently very small. A larger sample of polymorph pairs is needed to estimate its importance in modelling the energetics and properties of organic molecular crystals. The large sample used here should provide good estimates of the general trends for polymorphic crystals, even if the computational methods cannot reliably resolve lattice- and free energy differences in the individual pairs.

The influence of thermal expansion on polymorph relative stabilities will depend on how different the volumetric thermal expansion coefficients are between polymorphs, and this in turn depends on differences in elastic properties of the materials. The magnitude of typical differences in elastic properties between polymorphs are therefore also of interest.

In this Chapter, I present results of large scale application of rigid-body lattice-vibrational free energy calculations at 0 K, room temperature and at the respective melting point temperature. A very large number of organic molecular polymorphic crystals have been used. Doing so, I hope to be able to provide answers to many of the questions posed above. The study reported here is likely the largest ever on polymorph pairwise differences.

5.2 Methods

The main objective of this study is to use a large selection of known polymorph pairs and for each pair calculate the difference in lattice-vibrational free energy and other properties at 0 K, at room temperature and at the melting point temperature T_{m}. The results are used to determine how often temperature causes a re-ranking of the thermodynamic stability, thus obtaining an estimate for how commonly polymorph pairs are enantiotropic. In addition, polymorph pairs with diverging free energies may be found.

Below I describe the set of polymorph pairs and the calculations that were performed on them. I describe how the structures were first energy-minimised on a realistic potential energy surface and how rigid-body harmonic (HA) and quasi-harmonic approximation (QHA) lattice dynamics calculations were carried out. I also explain how melting point temperatures were obtained for all the crystal structures.

The main focus is on general trends in the large set of polymorph pairs, but I also performed more detailed studies of a number of structures and polymorph families. The calculated properties of individual systems are compared to experimental literature data, which allows us to estimate the overall accuracy of the results.

Both harmonic approximation and quasi-harmonic approximation lattice dynamics were used. The harmonic approximation was used to study crystal properties at absolute zero and room temperature, and how the free energy changes in this interval. At higher temperatures, anharmonic effects become significant. In order to study property differences over the whole temperature interval from 0 K to the melting point, quasi-harmonic calculations were employed at the melting point temperature of each structure.

5.2.1 Polymorph selection

Crystal structures were obtained from the Cambridge Structural Database (CSD), version 5.35 (Nov. 2013). The compounds and crystal structures are referred to by their database (CSD) reference codes [304]. Reference code families of polymorphs in van de Streek's 'best hydrogens list' [444, 445] containing only H/D, C, N, O, F, S, and Cl were selected, the elements for which accurate Williams-type force fields exist, see Chapter 3. Multicomponent crystals and structures of radicals, *i.e.* molecules that have unpaired electrons, were removed.

Since the lattice dynamical treatment used here only includes rigid molecule motions, differences in intramolecular vibrations between polymorphs are assumed to be negligible. This approximation may lead to errors when there are significant molecular geometry differences between polymorphs, such as in conformational polymorphism. Therefore, a root-mean-square deviation (RMSD) of atomic coordinates for molecules in different

polymorphs was used to exclude structures with large intramolecular differences. The RMSD was calculated with the program TORMAT [446], ignoring the positions of hydrogen atoms. Cruz-Cabeza & Bernstein [16] showed that polymorphs tend to have similar molecular conformations, and that an RMSD of atomic coordinates of up to 0.225–0.3 Å usually corresponds to slight adjustments of the same molecular conformer, *i.e.* a large majority of polymorphs are packing polymorphs. Essentially the same distribution of intramolecular RMSD-values was observed in this study, see Figure 5.1 and Figure S12 in the supplementary information of reference [16]. An RMSD limit of 0.25 Å was therefore used to exclude structure pairs where molecular geometry differences exceeded this value.

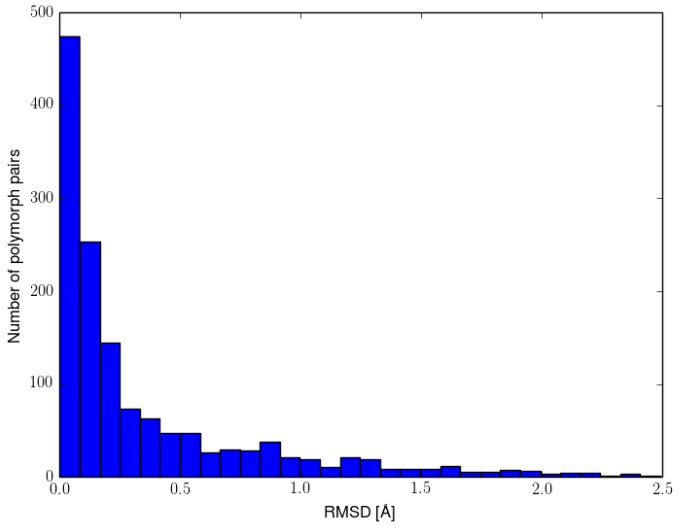


Figure 5.1: Distribution of root mean square deviations in molecular geometry between polymorphs. Most polymorphs have essentially the same molecular conformations.

The symmetries of structures with non-integer Z′ were modified to include whole molecules in the asymmetric unit, in order to facilitate density functional- and multipole calculations. Structures with more than 65 atoms in the asymmetric unit were excluded to limit the computational cost of the flexible molecule energy minimisation. Missing hydrogen atoms were added based on conventional geometric criteria whenever this could be done unambiguously. Deuterium atoms were substituted by hydrogen in deuterated structures determined by neutron-diffraction. Structures with disorder or incorrectly placed hydrogen atoms were excluded. A number of structures had to be excluded because of limitations in the energy-minimisation method applied here, as described below.

The final structure set consisted of 1061 crystal structures in 508 polymorphic clusters forming 466 polymorph pairs, 39 triplets and 3 quadruplets, yielding 601 pairwise comparisons of polymorphs. The structures were analysed for the presence of hydrogen

bonds with the program PLATON [283]. Hydrogen bonds were found in 654 crystal structures, forming 310 clusters, and 507 structures in 198 clusters did not contain hydrogen bonds.

For quasi-harmonic calculations, a somewhat smaller subset of these structures were used. The final structure set used for quasi-harmonic calculations consisted of 864 crystal structures of 418 different compounds that successfully completed all calculations. There are 391 polymorph pairs, 26 triplets and 1 quadruplet, resulting in a total of 475 pairwise comparisons.

The set of polymorph pairs used here is very well suited for computational studies of polymorphism. Additional information about these systems obtained during this study is available. The data consists of SMILES strings [143] and the chemical formula of the compounds, calculated thermal pressures, experimental and predicted melting points, elastic moduli, volume thermal expansion coefficients, unit cell volumes, Z and Z'-values. This data has been published as the *Nyman Polymorph Library* (NPL2016) in reference [3], see Appendix D for details.

5.2.2 Energy model

The computational framework of the rigid-molecule hybrid force field with distributed multipoles, as described in Chapter 2, was used. The lattice dynamical methods have been described in Chapters 2 and 3.

The lattice energy consists of the intermolecular cohesive energy and a quantum chemical evaluation of the intramolecular energy, calculated with a hybrid density functional.

The W99rev6311 and W99rev6311P force fields were used [225, 2], supplemented with parameters for sulfur, fluorine and chlorine [433, 222, 167]. Halogen atoms were modelled with an anisotropic repulsion [167]. For flexible-molecule geometry-optimization and harmonic approximation lattice dynamics the W99rev6311 force field was used. For the quasi-harmonic parts the more accurate W99rev6311P5 force field with polarized multipoles was used.

The intramolecular electronic structure and energy was calculated with GAUSSIAN 09 at the B3LYP/6-311G** level of theory. For the quasi-harmonic calculations, a Becke-Johnson damped version of Grimme's D3 dispersion correction was also included [407, 99].

5.2.3 Flexible molecule optimisation

Crystal geometry optimisations were performed with CrystalOptimizer [236, 237]. CrystalOptimizer minimises the lattice energy while allowing a set of selected intramolecular degrees of freedom (bond lengths, angles and dihedrals) and unit cell dimensions to change in response to intermolecular forces. All other intramolecular degrees of freedom are optimised to their DFT(-D) equilibrium values. The optimisation is performed numerically by iterating between inter- and intramolecular calculations. Intermolecular packing forces were calculated with Dmacrys, using the W99rev6311 force field and distributed multipoles, as described above.

Two slightly different versions of the AutoDof algorithm, described in Appendix A, were used to select suitable flexible degrees of freedom. For harmonic calculations, three-, four- and five-membered rings were treated as more flexible. The algorithm was later improved to recognize small rigid rings. The latter algorithm was used to optimise structures prior to quasi-harmonic lattice dynamics calculations.

When a hybrid density functional, such as B3LYP, is used for the structure optimization, crystals with more than 55–65 atoms per asymmetric unit, or crystals of large flat molecules or macrocycles are often prohibitively expensive to optimize with CrystalOptimizer due to convergence difficulties in calculating the molecular Hessian matrix. A few problematic structures had to be excluded because of this. The 6-311G** Gaussian basis set were used for all calculations. For large molecules, the GGA functional PBE with the D3 dispersion correction may be more suitable than hybrid functionals.

5.2.4 Lattice dynamics

All crystal structures obtained from CrystalOptimizer were re-optimized with the force field used for lattice dynamics calculations. Lattice dynamics requires the structure to be an energy minimum. Rigid-body lattice dynamics were performed with Dmacrys using the W99rev6311 (harmonic) and W99rev6311P5 (quasi-harmonic) force fields [2, 225].

For all crystal structures, the lattice-vibrational free energy at 0 K and 300 K were calculated with harmonic approximation lattice dynamics. Calculations were also applied at the respective melting point of each structure with both the harmonic and quasi-harmonic methods. The co-prime split linear supercell method was used for Brillouin zone sampling. Target **k**-point distances of 0.12 (harmonic) and 0.10 Å (quasi-harmonic) were used. For quasi-harmonic calculations, the Debye and Gaussian KDE methods were also used to improve the convergence of the free energy. To model thermal expansion, the thermal pressure method was used. These methods have been described in detail in Chapter 2.

Elastic tensors were calculated simultaneously from the second derivative of the potential energy with respect to unit cell changes, as implemented in DMACRYS, using algorithms described elsewhere [259, 261].

5.2.5 Melting point predictions

Melting points for many of the polymorphs in the set were available in the Cambridge Structural Database [304]. The melting temperature differences between polymorphs are smaller than the accuracy needed for the purposes of this study. Therefore, I assume that polymorphs have the same melting point, and only one temperature per polymorph family was determined. The experimentally determined melting temperature was used in the quasi-harmonic calculations whenever possible. This was the case for 225 (44%) of the 508 polymorph families. It is possible that additional experimental melting points are known and included in other databases, principally by NIST. The NIST databases however require searching by compound name, which was not attempted.

For all other polymorph families, the melting points were predicted. Two different methods were used for the predictions. The first method is based on molecular fragment contributions, and is implemented in the MTBTNT program in the EPASUITE [447] software, developed by the US Environmental Protection Agency. The program internally uses two algorithms for estimating the melting temperature and returns a weighted average of these. The first algorithm is based on Joback's method [448]. The second algorithm estimates the melting point as $T_\mathrm{m} = 0.5839 \cdot T_\mathrm{b}$ as suggested by Gold & Ogle [449], where T_b is the boiling point, which is also estimated from fragment contributions using a method by Stein & Brown [450]. The MTBTNT program takes SMILES strings [143] as input. These were obtained from the crystal structure's atomic coordinates (xyz-files) *via* the OPEN BABEL software [144].

Molecular fragment based methods for predicting melting points are known to be inaccurate. The standard deviation of the error in estimated melting points is 63.9 K, according to the MTBTNT manual [447]. Therefore, I attempted to improve the accuracy of predicted temperatures by using a second method relying on the correlation between lattice enthalpies and melting points.

Since we have access to an accurate energy model for crystals, trying to predict melting points based on the lattice energy is attractive. A straight proportionality is sometimes fitted between these quantities [451], but this does not result in an adequate fit, the residual errors are non-normal and heteroscedastic. Only the intermolecular part of the lattice energy was used for this purpose. Intramolecular contributions arising from the small geometric distortions of the molecules in the crystals (see above) are not included here for simplicity.

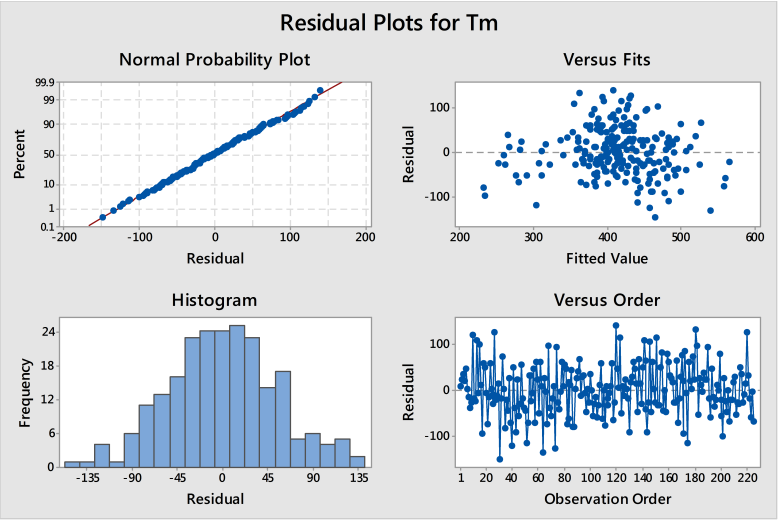


Figure 5.2: Residual analysis for the regression model. The residual errors not accounted for by the regression model are normally distributed (left subfigures) and have a uniform, homoscedastic, variance (right subfigures). The regression model thus adequately models the trend in the underlying data.

To obtain an adequate regression model, it is necessary to perform a variance stabilizing transformation. Assuming the data is Poisson-distributed, the variance can be made homoscedastic with the Anscombe transform [452]:

$$x \mapsto 2\sqrt{x + 3/8} \tag{5.1}$$

In this case the 3/8 term is small and simply taking the square root of the lattice energies is adequate and gives virtually identical results. A regression analysis in Minitab 17 between 225 experimental melting points and the crystals' W99rev6311 intermolecular energies E_{inter} yielded the following model for the melting points:

$$T_{\mathrm{m}} = 39.0\ \mathrm{K} \cdot \sqrt{-E_{inter} \cdot \mathrm{mol/kJ}} \tag{5.2}$$

The residual standard deviation for this model is 54.5 K. The residuals are normally distributed and homoscedastic, see Figure 5.2. It is remarkable that such a simple regression model yields more accurate melting point predictions than the molecular fragment method. Details of the regression analysis are included as output from the Minitab program, see Figure 5.3 and Table 5.1.

We have used the arithmetic mean of the two methods as our predicted melting temperatures. The mean absolute error of this average was estimated to be 39 K, or less than 10%. Although still not quantitatively accurate, it is good enough for the purposes of this study. Both predicted and experimental melting temperatures are included in the *Nyman Polymorph Library*, see Appendix D.

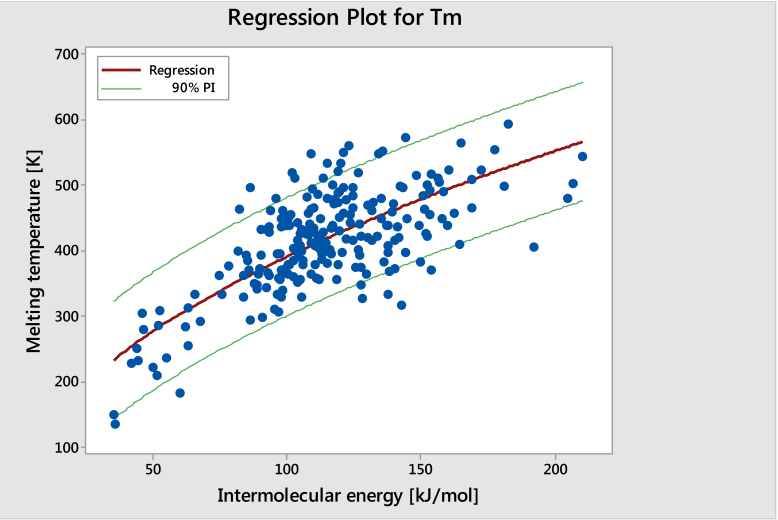


Figure 5.3: Non-linear fit between experimental melting points and intermolecular energies.

5.2.6 Detailed study of selected systems

In-depth studies of several selected systems were performed. These were chosen primarily as families with several experimentally well characterized polymorphs.

Plenty of experimental data have been reported for polymorphs of the following compounds: acridine (ACRDIN) [453, 454], 1,2-ethanediamine (ETDIAM) [411, 271], *m*-nitrophenol (MNPHOL) [412, 413, 414], acetaminophen/paracetamol (HXACAN) [415, 416], glutaric acid (GLURAC) [417, 418, 419], octa-sulfur (FURHUV) [420, 421, 422, 455], adipic acid (ADIPAC) [423, 424], 2,2′-dipyridylamine (DPYRAM) [425], theophylline (BAPLOT) [456, 48, 53, 457, 458, 459] and pyrazine-2-carboxamide (PYRZIN) [426].

Polymorphs of these compounds were studied computationally in greater detail. Free energy curves were calculated, and in addition thermal expansion rates (see Chapter 3) and phonon densities of states. The lattice energy was calculated relative to the energy of the gas phase molecular conformer. Continuous free energy curves were obtained by Hermite piecewise polynomial interpolation [460], using both the free energy and entropy at several temperatures to fit continuous free energy curves. These results were then compared to published experimental data. This allowed me to identify weaknesses and strengths in the methods used, which are discussed below.

5.3 Results and Discussion

The results for the general trends and property-distributions for the entire set of polymorphs are presented first. Results for particular polymorph systems are presented later, together with comparisons to experimental reference data.

```
Starting Values for Parameters
Parameter  Value
Theta1        37
Theta2       0.5*
* Locked.

Equation
Tm = 38.9938 * 'Lattice energy' ^ 0.5

Parameter Estimates
Parameter  Estimate  SE Estimate
Theta1      38.9938     0.339152
Theta2       0.5000             *

Tm = Theta1 * 'Lattice energy' ^ Theta2

Lack of Fit
Source          DF      SS       MS     F      P
Error          224  665027  2968.87
  Lack of Fit  222  660987  2977.42  1.47  0.492
  Pure Error     2    4040  2020.00

Summary
Iterations        2
Final SSE    665027
DFE             224
MSE         2968.87
S           54.4873
```

Table 5.1: Table of regression analysis results produced by MINITAB.

5.3.1 Contributions to crystal stability

We first inspect the distribution of lattice energy differences between polymorphs, shown in Figure 5.4. These are the total differences in the sum of intermolecular and intramolecular energies, calculated for the static, lattice energy minimised structures at 0 K. The results confirm the validity of the commonly held belief that polymorph lattice energy differences are commonly less than 10 kJ/mol [20, 23], only 1.5% of polymorph pairs included in this study exceed 10 kJ/mol in relative lattice energy. Indeed, most lattice energy differences are much smaller: more than half (52.7%) of all polymorph pairs are separated by less than 2 kJ/mol and 95% by less than 7.2 kJ/mol. This can be used as a general rule-of-thumb for organic crystals.

Lattice energy differences are generally dominated by differences in intermolecular interactions, with 68.5% of polymorph pairs differing in intramolecular energy by less than 1 kJ/mol, see Figure 5.5. This is largely due to our selection of polymorphs with small intramolecular geometry differences. However, most polymorphs are packing

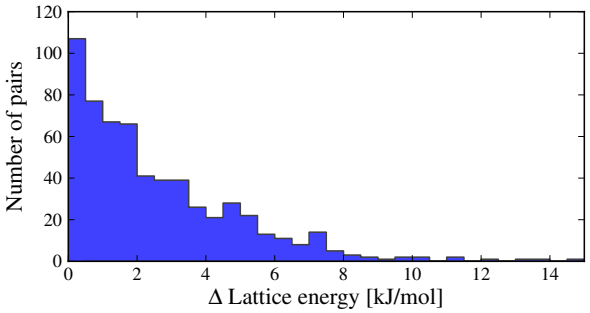


Figure 5.4: Distribution of the absolute value of calculated lattice energy differences between 601 polymorph pairs. Lattice energy differences rarely exceed 8 kJ/mol.

polymorphs and this result is likely to be quite general. The range of intramolecular energies is in general agreement with another study of conformational energies in molecular crystals [147]. Only in rare cases can intramolecular energy differences reach 15–20 kJ/mol, where a high-energy conformation is found in one polymorph. The largest intramolecular energy differences are associated with different hydrogen atom positions that lead to a switch between inter- and intramolecular hydrogen bonding. It is difficult to computationally predict this type of conformational polymorphism [461]. In all cases of large intramolecular energy differences, the intramolecular penalty is compensated by stronger intermolecular interactions in the polymorph containing the higher energy molecular conformation.

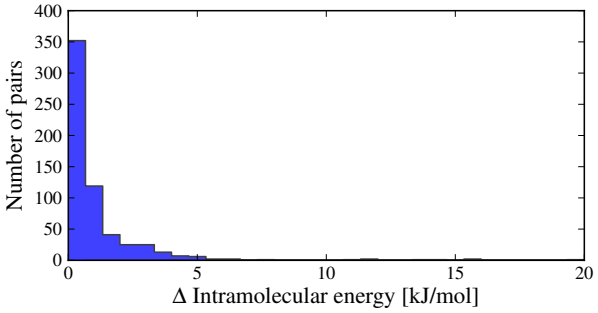


Figure 5.5: Distribution of intramolecular energy differences in 601 pairs of packing polymorphs.

Differences in the vibrational contribution to the free energy at 300 K, calculated in the harmonic approximation, are generally smaller in magnitude than lattice energy differences, see Figure 5.6. In fact, $|\Delta F_{\mathrm{vib}}|$ was found to be smaller than 1 kJ/mol in more than 70% of polymorph pairs and greater than 2 kJ/mol in fewer than 6% of cases. Due to the small magnitude of ΔF_{vib} between polymorphs, and recalling that $\Delta A = \Delta E_{\mathrm{latt}} + \Delta F_{\mathrm{vib}}$, the overall distribution of free energy differences, see Figure 5.7, closely resembles the distribution of lattice energy differences: 56.6% of pairs are separated by less than 2 kJ/mol in calculated free energy, 95% are below 6.4 kJ/mol

and the free energy difference of only 0.5% of polymorph pairs exceed 10 kJ/mol.

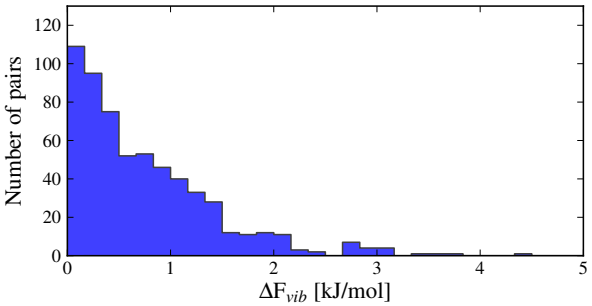


Figure 5.6: Distribution of the absolute value of calculated differences in lattice vibrational energies between 601 pairs of polymorphs at 300 K calculated with harmonic approximation lattice dynamics.

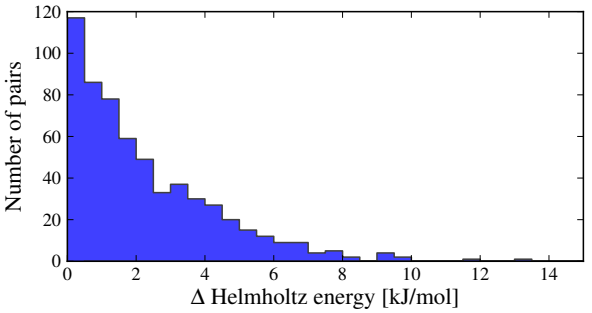


Figure 5.7: Distributions of the absolute value of calculated differences in Helmholtz free energies between polymorphs at 300 K.

There are cases where F_{vib} reinforces the static lattice energy difference as well as cases where ΔF_{vib} and ΔE_{latt} have opposite sign. The latter case, where the vibrational energy contribution counteracts the lattice energy difference, leading to converging free energy curves, is more common: dynamical energy contributions generally reduce the energy difference between polymorphs. This is the case in 69% of the pairs studied here. In 9% of cases does the vibrational energy have the opposite sign of, and is larger than, the lattice energy difference. That is, in the harmonic approximation, 9% of polymorphs are re-ranked in stability below 300 K, see the green field in Figure 5.8.

Most crystal structures in this study have melting points between 350 and 470 K, see Figure 5.9. Since these temperatures are substantially higher than room temperature, we expect to see a larger vibrational contribution to the free energy and more re-ranking of polymorph stabilities compared to the results above and other studies using room temperature free energies [23]. This is also the reason why quasi-harmonic calculations were performed. In order to estimate how often polymorphs are re-ranked at higher temperatures, it is conceivable that the thermal expansion cannot be neglected. Results of quasi-harmonic calculations are presented below.

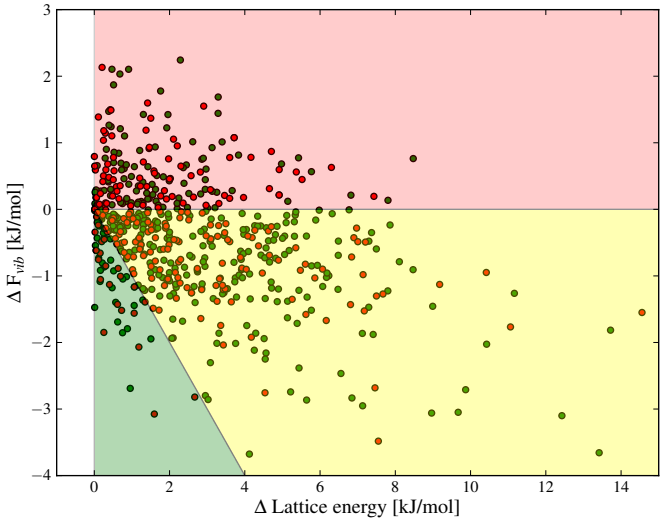


Figure 5.8: Re-ranking of polymorph pairs at 300 K calculated in the harmonic approximation. Green and red data points indicate the presence and absence of hydrogen bonds in the crystals, respectively. Background shade indicates whether the pair has diverging (red) or converging (green & yellow) free energy curves.

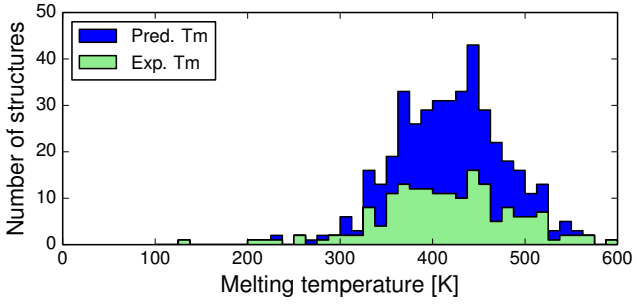


Figure 5.9: The distribution of melting points for the 418 polymorph families used for quasi-harmonic calculations. 44% of the polymorphs had experimentally determined melting points (green), while the rest had their melting points predicted (blue).

5.3.2 Contributions to vibrational energy differences

The total vibrational contribution to the free energy difference between polymorphs is a sum of the zero point energy, the thermal contribution to the internal energy and an entropic contribution:

$$\Delta F_{\mathrm{vib}}(T) = \Delta ZPE + \int_0^T \Delta C_{\mathrm{v}}(T)\, dT - T\Delta S \tag{5.3}$$

At high temperatures, anharmonic contributions also arise as a larger entropy due to the softening of the vibrational frequencies and a less favourable lattice energy due to

thermal expansion of the crystal. These effects are accounted for in the quasi-harmonic approximation. The relative importance of these different terms are examined next.

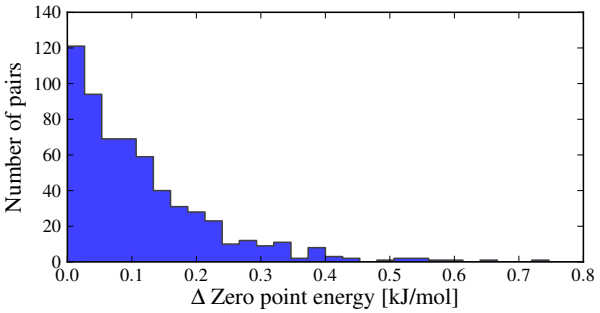


Figure 5.10: Distribution of the differences in vibrational zero point energy between polymorphs.

The zero point energy is a small contribution to the relative stability of polymorphs, see Figure 5.10. The zero point energy difference ΔZPE is less than 0.33 kJ/mol in 95% of polymorph pairs and the largest calculated difference is just over 0.7 kJ/mol. However, as has been demonstrated with *ab initio* calculations, even the zero point vibrations cause a significant thermal expansion [279, 280], which is often not considered when modelling crystals with electronic structure methods.

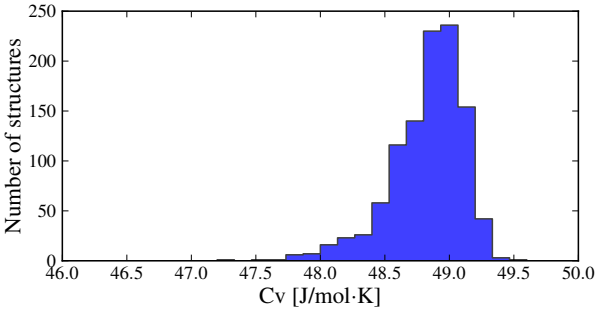


Figure 5.11: Distribution of heat capacities calculated in the harmonic approximation at 300 K for all 1061 crystal structures. Only intermolecular vibrational mode contributions were included.

Molar heat capacities do not vary greatly from the expected equipartition value. Since our lattice dynamical treatment excludes intramolecular vibrations, each molecule is modelled as having exactly six vibrational degrees of freedom (3 translational and 3 rotational) and the calculated $C_v \approx 6R$, see Figure 5.11. This is an important drawback of the rigid-body approach, and makes the calculated absolute heat capacities unreliable.

Consequently, specific heat capacity differences between polymorphs are small, see Figure 5.12. In 95% of polymorph pairs, C_v differs by less than 0.46 J/mol·K. Note that here the differences between packing polymorphs are shown and these should still be highly

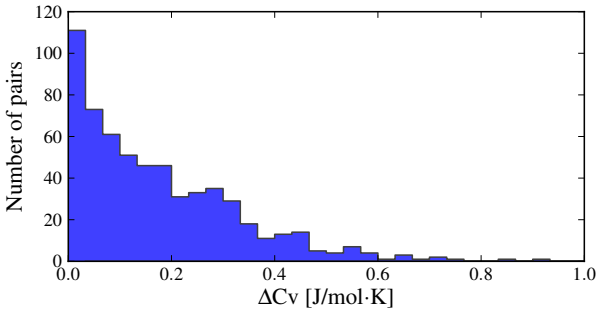


Figure 5.12: Distributions of the differences in vibrational heat capacity between packing polymorphs at 300 K calculated in the harmonic rigid-body approximation.

reliable, despite the limitations of the rigid-body model. Conformational polymorphs can be expected to have larger ΔC_v-values than those in Figure 5.12.

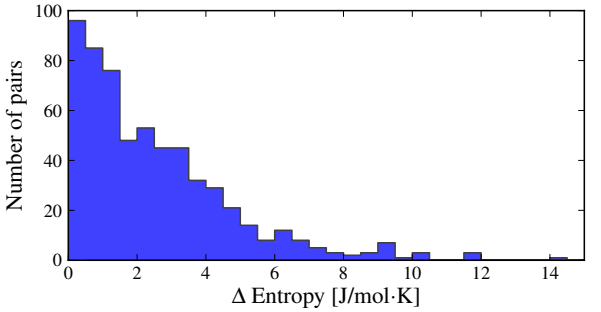


Figure 5.13: Distributions of the differences in vibrational entropy between polymorphs calculated at 300 K.

Entropy differences at room temperature are about an order of magnitude larger than the thermal contribution to the internal energy, see Figure 5.13, and entropy is by far the most important vibrational contribution to polymorph free energy differences at room temperature and above. As with all the observed distributions, entropy differences have a tendency to be small: 51% of calculated ΔS are below 2 J/mol·K, corresponding to only 0.6 kJ/mol at 300 K. However, ΔS is greater than 5.4 J/mol·K in 10% of polymorph pairs and exceeds 6.8 J/mol·K in 5% of pairs, corresponding to a 2 kJ/mol contribution to the room temperature free energy difference. These largest entropy differences are important when compared to the static lattice energy differences, which are less than 2 kJ/mol in over half of the the polymorph pairs. Vibrational effects are most important in the cases where large entropy differences are coupled with small lattice energy differences, possibly causing an enantiotropic phase transition.

An important limitation of the method for calculating entropy is that non-vibrational contributions to the entropy have not been included. Disordered hydrogen atoms give a significant contribution to the absolute entropy [462, 463, 464, 278], but again, the contribution of disorder to the entropy-difference between polymorphs should be small.

Only in cases where one crystal form is disordered and another is not should this effect be significant.

Thermal expansion makes the vibrational frequencies lower, which increases the entropy. On the other hand, the structure moves away from the lattice energy minimum, so that the enthalpy and entropic effects largely cancel out [280]. Indeed, the free energy contributions due to anharmonic effects are small, see Figure 5.14. These contributions were calculated as the difference between the harmonic and quasi-harmonic free energies at the respective melting point of each polymorph. Hence, the contributions shown in Figure 5.14 should be seen as an upper limit. At room temperature, the anharmonic contribution should normally be quite small.

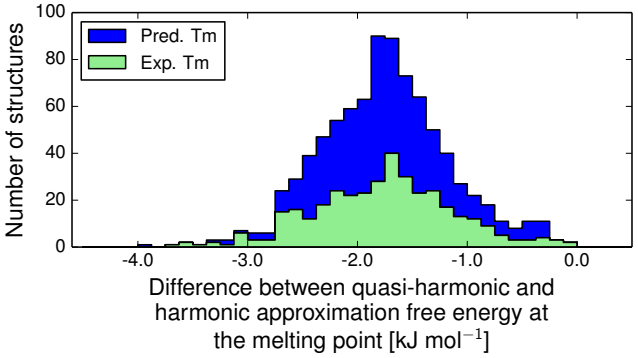


Figure 5.14: The distribution of contributions to Gibbs free energy due to thermal expansion at the respective melting points (T_m). The values are differences between quasi-harmonic (thermally expanded) and harmonic free energies, $A^\mathrm{QHA}(T_\mathrm{m}) - A^\mathrm{HA}(T_\mathrm{m})$. The distribution is split into green and blue fields for results based on experimental and predicted melting points.

The contribution from thermal expansion to the pairwise difference in free energy between polymorphs is shown in Figure 5.15. For ranking the relative stability of polymorphs, as is done in crystal structure prediction, the thermal expansion can safely be neglected. That is, the results obtained here confirm the conclusion drawn by van Eijck [300].

5.3.3 Physical origins of differences in properties

The thermodynamic properties considered above are functions of the phonon density of states. Differences between polymorphs in entropy, heat capacity, zero point energy, and to some extent the thermal expansion, arise from variations in the distribution of lattice vibrational frequencies. Calculated phonon frequencies generally fall in the range between 4 and 220 cm^{-1}. The distribution of frequencies in this range can vary substantially between polymorphs. As an example, Figure 5.16 shows the calculated vibrational frequency distributions of theophylline forms I and II. Form II is favoured

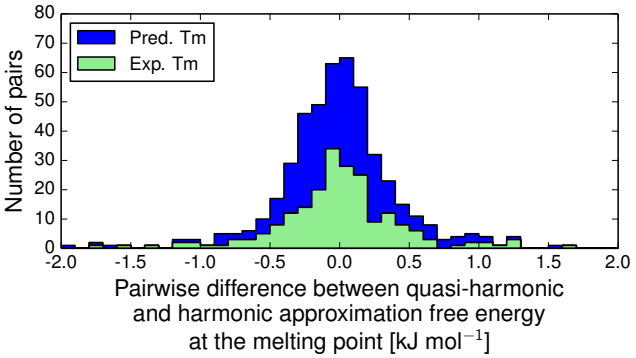


Figure 5.15: The distribution of pairwise differences in thermal expansion-contributions to the relative thermodynamic stability of polymorphs at their respective melting temperatures, $\Delta(A^{\mathrm{QHA}}(T_{\mathrm{m}}) - A^{\mathrm{HA}}(T_{\mathrm{m}}))$. The distribution is split into green and blue fields depending on whether the melting points were experimental or predicted.

by lattice energy and is known to be more stable than I at low temperatures. However, form I has a frequency distribution that is shifted towards lower frequencies. This leads to a relatively large difference in entropy, $\Delta S = 6.5$ J/mol·K, favouring form I and an enantiotropic phase transition between these structures [457]. A long-term goal of the free energy methods developed in this thesis is to be able to predict such phase transitions.

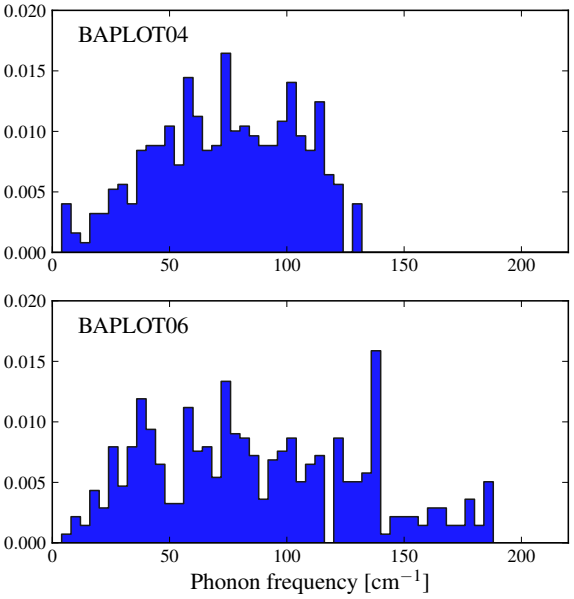


Figure 5.16: Phonon densities of states for theophylline polymorphs I and II (CSD: BAPLOT04 and BAPLOT06). The target **k**-point distance was 0.08 Å^{-1}. The difference in frequency distribution causes a relatively large difference in entropy.

It is generally not possible to immediately see what causes these differences in vibrational frequencies. The theophylline polymorphs I and II differ in which atoms are involved in hydrogen bonding. Such differences in strong intermolecular interactions may lead to different vibrational spectra and, sometimes, significant differences in thermodynamic properties. However, trying to deduce such causal relationships from nearest neighbour interactions only is an oversimplification.

As was elaborated upon in Chapter 2, it is commonly assumed that there is a general correlation between the lattice energy and density [319, 321]. In the results obtained here however, such correlations are very weak indeed, and cannot be used to make predictions about the thermochemical properties of crystals. The results indicate that there is an extremely weak, but due to the large number of data points statistically highly significant, correlation between lattice energy- and density differences between polymorphs, see Figure 5.18. The denser polymorph has a slight tendency to have a lower (*i.e.* more stable) lattice energy.

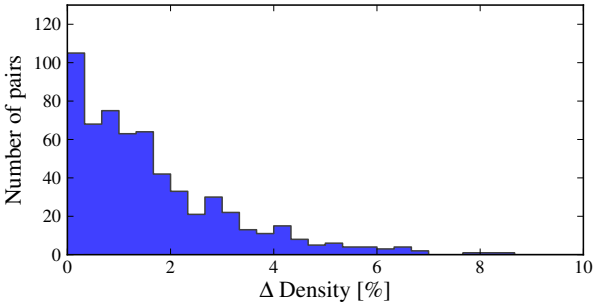


Figure 5.17: Distribution of differences in density between polymorphs.

Crystal structures with low density are often assumed to vibrate with lower frequencies, due to molecules having more empty space to vibrate in, and hence might be expected to have higher entropy. There is such a trend in this study, see Figure 5.19. Again, the correlation is extremely weak but statistically significant. The lower density polymorph has a small tendency to have the larger vibrational entropy.

Both correlations represent trends across a large set of polymorph pairs and are not predictive for individual cases. The correlations are so small that no conclusions can be drawn about the relative energy or entropy of polymorphs based on their densities. This has potentially important consequences for the interpretation and validity of the Burger & Ramburger rules [319, 321], see below.

Strong directional interactions, hydrogen bonds in particular, might interfere with the close-packing of crystals and cause differences in density and entropy between polymorphs. I therefore examined if there are any differences in the distributions of ΔF_{vib} and ΔS between polymorphs for molecules with and without hydrogen bonds (green and

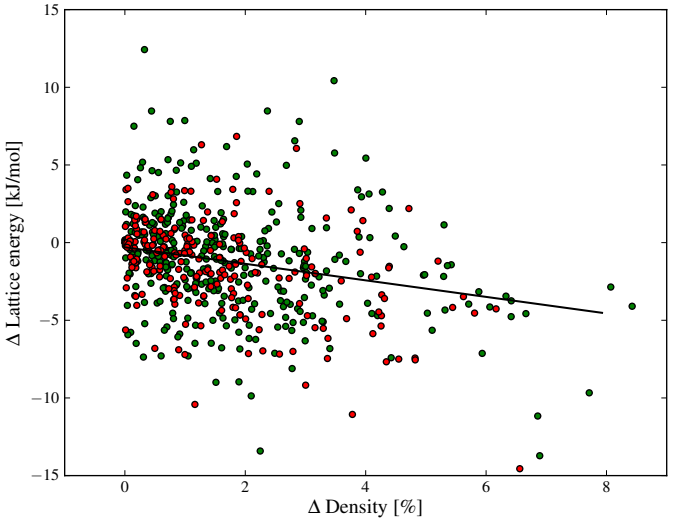


Figure 5.18: Differences in lattice energy *vs* density differences between polymorphs. Polymorphs in each pair are ordered such that the density difference is positive. Green and red data points denote structures with and without hydrogen bonding. The straight line shows the least-squares regression.

red data points, respectively, in Figure 5.8). No evidence for a difference in distributions of thermodynamic property-differences between polymorphs based on the presence or absence of hydrogen bonding was found in this study.

The apparently random scatter in Figures 5.8, 5.18 and 5.19 suggests that other factors, such as non-specific or non-nearest-neighbour interactions contribute to the bulk trend. Nevertheless, entropy can perhaps be expected to contribute a stabilising effect for porous crystal structures, such as the clathrates and inclusion compounds described in Chapter 7 and 8 of this thesis.

5.3.4 Thermal expansion and elastic properties

The volumetric thermal expansion coefficient of each crystal structure was calculated as a finite difference between the 0 K and the thermally expanded structure. Figure 5.20 displays the distribution of expansion coefficients for the whole set of structures. The thermal expansion coefficient is typically between 80–240 $\times$ 10^{-6} K^{-1}, corresponding to 0.8 and 2.4% per 100 K, which can be used as a rule of thumb for organic molecular crystals in general.

Unusually large thermal expansions were seen for a few weakly bound crystals of organic solvents, such as acetonitrile, chloroethane, cyclobutanol and 2,5-dimethylpyrazine.

The pairwise differences in volumetric thermal expansion coefficient between polymorphs are shown in Figure 5.21. Differences in thermal expansion coefficients are less than 15.8

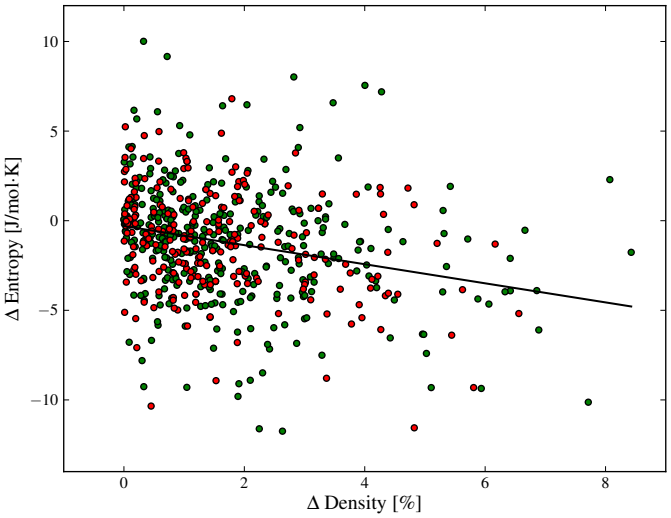


Figure 5.19: Differences in lattice-vibrational entropy *vs* density differences between polymorphs. Polymorphs in each pair are ordered such that the density difference is positive. Green and red data points denote structures with and without hydrogen bonding, respectively. The straight line shows the correlation model.

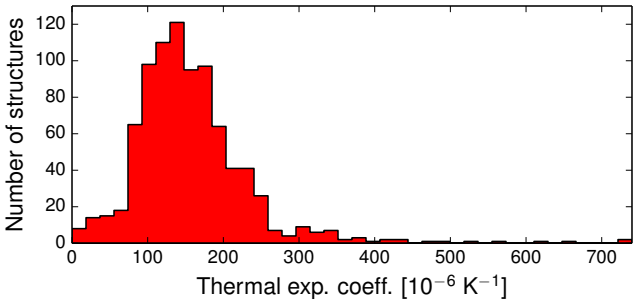


Figure 5.20: Distribution of thermal expansion coefficients.

$\times 10^{-6}$ K^{-1} for 50% of pairs, and rarely differ by more than 50×10^{-6} K^{-1}. The thermal expansion tends to be similar between polymorphs, only rarely is it relatively large.

The thermal expansion of a crystal is related to its elastic properties and differences in bulk and shear moduli are often of interest, for example in the manufacturing of drug tablets. The Hill averaged bulk and shear moduli calculated for all structures at 0 K are displayed in Figures 5.22 and 5.23. The calculated bulk and shear moduli for crystals of small organic molecules are usually 8–15 GPa and 2–8 GPa respectively at low temperature.

Bulk modulus differences exceed 2 GPa in only 10% of polymorph pairs and are less than 0.5 GPa in 41% of pairwise comparisons, see Figure 5.24. These small differences in the resistance to volume changes partly explain why polymorphs also tend to have similar

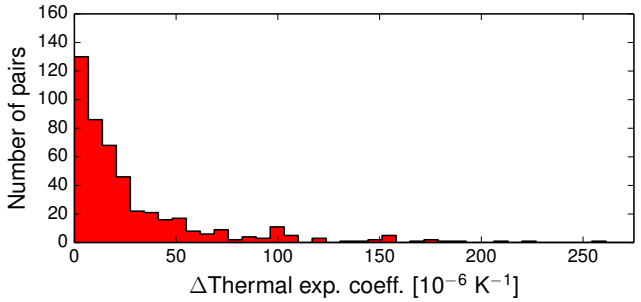


Figure 5.21: Distribution of pairwise differences in thermal expansion coefficient for 475 polymorph pairs.

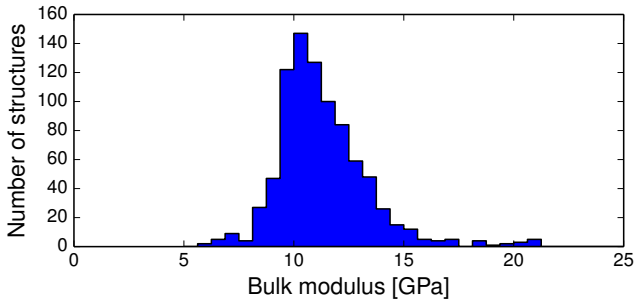


Figure 5.22: Distribution of Hill-averaged calculated bulk moduli for 864 crystal structures at their respective lattice energy minimum.

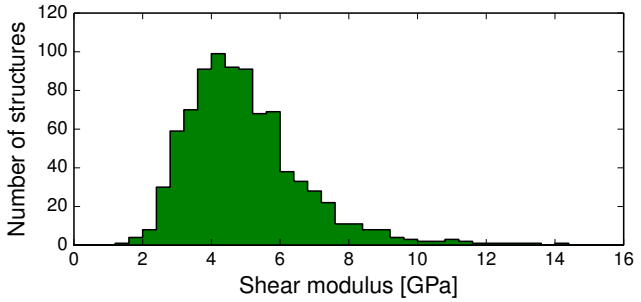


Figure 5.23: Distribution of calculated Hill-averaged shear moduli for 864 crystal structures at the lattice energy minimum.

thermal expansion coefficients. We find that the shear moduli also rarely differ by more than 2 GPa between polymorphs of the same compound, although this corresponds to a larger relative difference because of the typically smaller values of shear moduli. Since low shear moduli can relate the facile plastic deformation, the relatively large differences in shear moduli, see Figure 5.25, demonstrate why polymorph differences have important consequences for the tabletability of pharmaceuticals [27].

Pairwise absolute differences in elastic moduli between polymorphs at 0 K are shown in Figures 5.24 and 5.25.

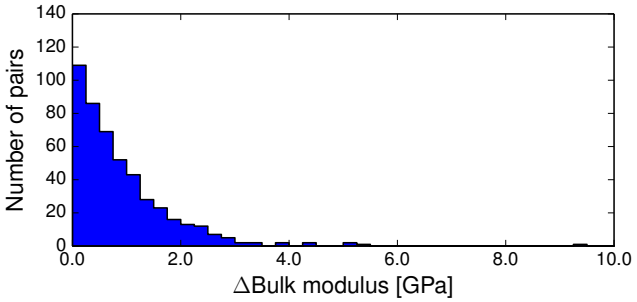


Figure 5.24: Distribution of differences in calculated 0 K bulk moduli for 475 polymorph pairs.

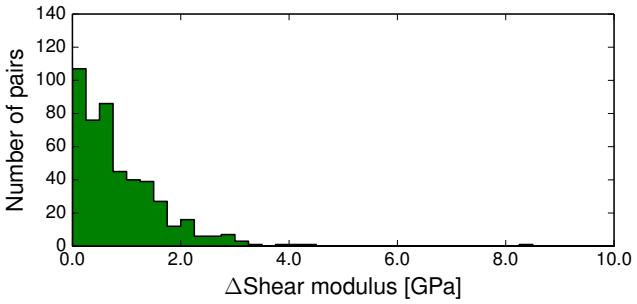


Figure 5.25: Distribution of differences in calculated 0 K shear moduli for 475 polymorph pairs.

Thermal expansion leads to a sizeable softening of the crystal structure. The bulk and shear moduli decrease rapidly with temperature so that near the melting point, the bulk modulus has reduced to on average 50% of the 0 K value, see Figure 5.26. The corresponding data for the shear moduli are shown in Figure 5.27.

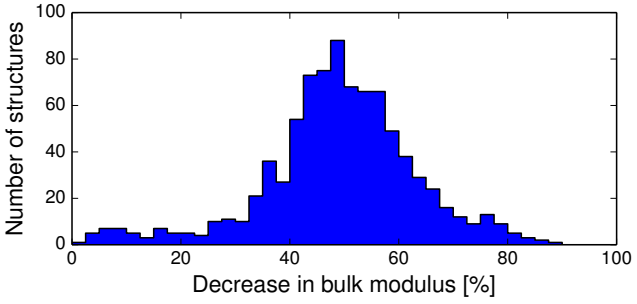


Figure 5.26: The bulk moduli at the melting point as a percentage of the 0 K value.

5.3.5 Re-ranking of polymorph stability with temperature

Re-ranking of polymorph, corresponding to crossing free energy curves, are of fundamental importance. This relates to the polymorphs being enantiotropic, and the frequency

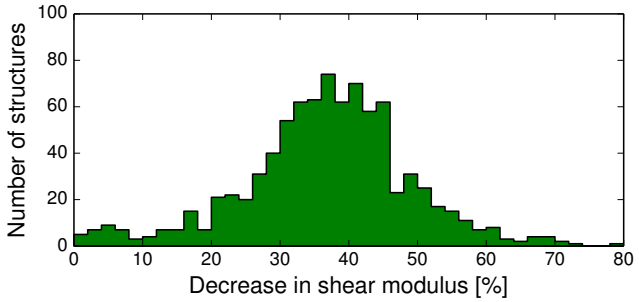


Figure 5.27: The shear moduli at the melting point as a percentage of the 0 K value.

with which such events occur relates to the predictive ability of crystal structure prediction methods that use temperature-free scoring functions.

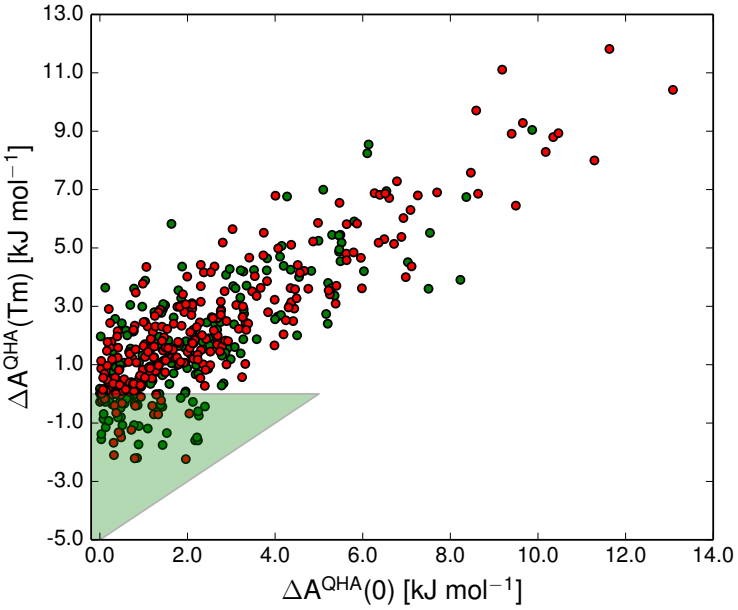


Figure 5.28: The correlation between the free energy difference in each polymorph pair at 0 K (including vibrational zero-point energy) and the free energy difference at the melting point temperature (T_{m}) in the quasi-harmonic approximation. The shaded green triangular region marks the 21% of polymorph pairs that were re-ranked by vibrational energy. Green and red data points denote structures with experimentally measured and predicted melting points, respectively.

In Figure 5.28 we show the correlation between polymorph pair-wise energy differences at 0 K and at the melting point calculated in the quasi-harmonic approximation. Even at the melting temperature, there is a strong correlation between the temperature-dependent free energy and the 0 K energy, just as has been shown previously at room temperature [1].

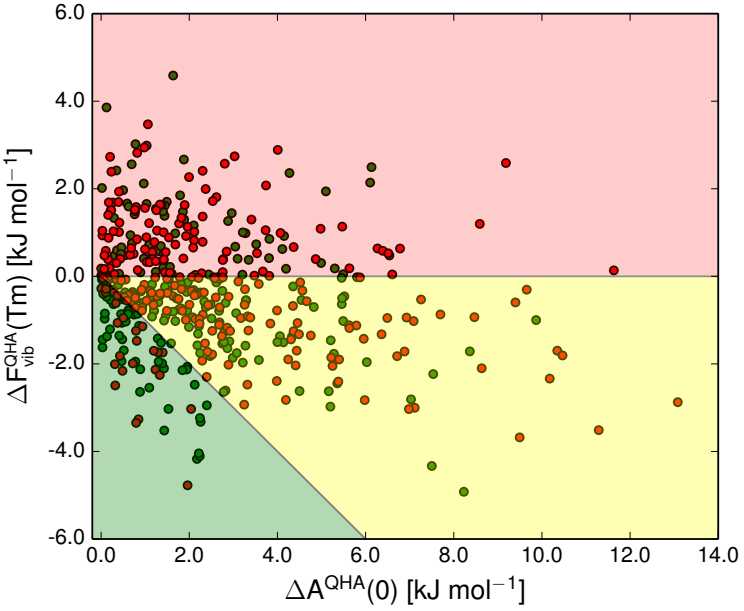


Figure 5.29: The quasi-harmonic vibrational energy differences at the melting point (T_m) relative to the 0 K relative stability between pairs of polymorphs. The 0 K energies include vibrational zero-point energy. The background colours indicate pairs that have diverging (red) and converging (yellow and green) free energy curves, and which pairs are re-ranked (green). Green and red data points denote polymorphs with measured and predicted melting points, respectively.

Figure 5.29 shows a scatter plot of the lattice-vibrational energy, including the effects of thermal expansion, at the melting point temperature of each crystal with respect to the relative stability at 0 K.

The shaded green regions in Figures 5.28 and 5.29 shows where a re-ranking of the relative thermodynamic stability occurs for the polymorphs between 0 K and the melting point. Here the computational method predicts an enantiotropic relationship between polymorphs. In the quasi-harmonic approximation 21% of the pairs fall within this region, compared to 17% in the harmonic approximation. This is substantially more than the 9% that were found to be re-ranked below 300 K, which is expected since the melting points are substantially higher than 300 K for the vast majority of the crystals studied.

5.3.6 Diverging free energy curves

As was argued in Chapter 2, free energy curves of polymorphs can either diverge with increasing temperature or converge. Diverging free energy curves, corresponding to points in the red field of Figure 5.29, have previously been described as extremely rare, and only a few polymorph pairs with possibly diverging curves have been studied experimentally [319, 322, 328]. The results obtained during this study on the other hand, perhaps

surprisingly, predict that diverging free energy curves should be common, occurring in 38% of cases.

One possibility that might explain this discrepancy is that the result is caused by incorrectly ranked lattice energies for some of the polymorphs. That is, the relative lattice energies are not accurately reproduced with our energy model, which might lead to an overestimation of diverging polymorph pairs. For the same reason, the fraction of enantiotropic polymorph pairs (21%) may be underestimated.

In order to investigate the predicted diverging free energy curves, I extracted a number of pairs that were predicted to have diverging free energy curves. These were extracted from the results of the quasi-harmonic calculations. In Chapter 3 these methods were benchmarked and shown to be reasonably accurate. There are however still significant errors in the lattice energy, often of several kJ/mol. This means that we cannot be certain about any particular polymorph pair. However, there are several examples of diverging free energy curves where the energy and entropy differences between the polymorphs are so large that it is unlikely they are due only to errors in the energy model.

Below I have include details of some polymorph pairs that were predicted to be diverging. Pairs whose lattice energies differ by at least 1.0 kJ/mol and have free energy curves diverging by at least an additional 2.0 kJ/mol at their melting point were chosen. A 1 kJ/mol difference in lattice energy is too small to be calculated reliably by almost any available computational methods, so even these results are not conclusive, but might help focus future investigations.

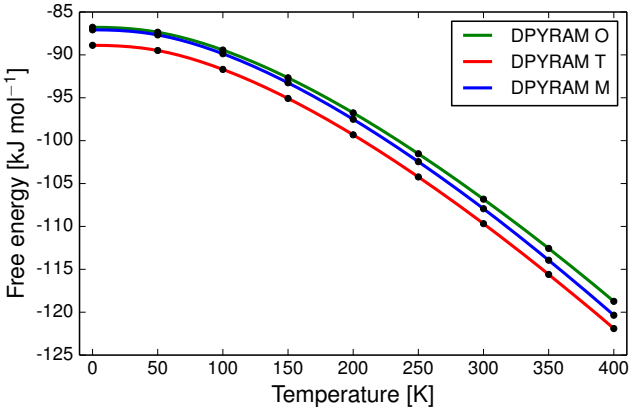


Figure 5.30: Diverging quasi-harmonic free energy curves of 2,2'-dipyridylamine polymorphs

In Figures 5.30 and 5.31 we show calculated quasi-harmonic free energy curves for three polymorphs of 2,2'-dipyridylamine (DPYRAM/01/02). Complete Gibbs free energy curves were calculated up to the melting point at 357 K and plotted as Hermite polynomials fitted to both the free energy and the entropy at several temperatures. The

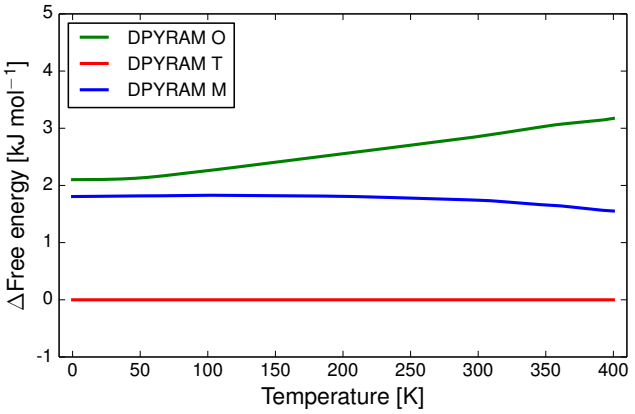


Figure 5.31: Predicted diverging free energy difference curves for polymorphs of 2,2'-dipyridylamine calculated in the quasi-harmonic approximation.

orthorhombic and triclinic polymorphs diverge significantly and the lattice energy difference is about 2 kJ/mol. These can be compared to qualitative curves based on experimental data [425]. The orthorhombic polymorph should be unstable at high temperatures, so the fact that this is the least entropically stabilized form is correct. However, it should be the most stable form below 263 K, indicating that our lattice energy ranking of the polymorphs is incorrect. Thus, the relative entropies seem to be accurate, but the static energy differences between the three polymorphs are incorrect.

We find several examples of polymorphs that diverge. It is unlikely that all of these are due to errors in the energy model. These polymorph pairs are described below in order to give a number of possible examples of crystal forms of the B′-kind, as described by Burger & Ramburger [319]. Unfortunately, experimental results to confirm or reject the predictions below are not currently available.

Two polymorphs of phenylisocyanurate (1,3,5-triphenylperhydro-1,3,5-triazine-2,4,6-trione) are known [465, 466]. The monoclinic form PHISCY is here predicted to be the more stable. The orthorhombic form PHISCY01 should be monotropically related to the stable form. At 0 K, the free energy difference is 1.0 kJ/mol, increasing to 3.9 kJ/mol at the melting point. The more stable form is predicted to have a substantially larger entropy, 12 J/mol·K, at the melting point 454 K.

Cyclo-(L-Glu-L-Glu), is a commercially available dipeptide. It has two known polymorphs deposited in the CSD [467]. According to the predictions made here, polymorph 2 (BUVKEJ01) is the more stable at all temperatures. At 0 K the free energy difference is 2.8 kJ/mol, increasing to 5.2 kJ/mol at the predicted melting point of 530 K. The free energy curves are diverging and the stable form has a higher entropy by about 5 J/mol·K at the melting point.

Two different polymorphs are known of 4-phenyl-1H-imidazole, a monoclinic form (NENJOF) and a trigonal form (NENJOF03) [468, 469, 470]. Computationally, these polymorphs are predicted to have an energy difference of 4.3 kJ/mol at 0 K, with the monoclinic form being the more stable. At the melting point, 403 K, the free energy curves have diverged and the free energy difference increased to 6.8 kJ/mol. The stable form has the higher entropy, the difference being 10.3 J/mol·K at the melting point. This is one of the largest entropy differences seen in this study.

There are two known polymorphs of 5-((4-methylphenyl)diazenyl)salicylaldehyde, one monoclinic (QANTUV) [471] and one orthorhombic (QANTUV02) [472]. According to the free energy calculations performed here, the most stable form should be the orthorhombic form (QANTUV02). At 0 K it is more stable by 1.9 kJ/mol, increasing to 4.4 kJ/mol at the melting point (685 K). The stable form hence has a higher entropy, ΔS is 8.1 J/mol·K at the melting point.

The compound eldoral, with systematic name 5-ethyl-5-(piperidin-1-yl)-pyrimidine-2,4,6-(1H,3H,5H)-trione, has at least three polymorphs. Two of them were recently studied by Gelbrich *et al.* in Innsbruck [473]. The same polymorphs were studied by Brandstätter-Kuhnert & Aepkers, also in Innsbruck in 1962 [474], and polymorph I*a* (denoted I here) was found to have a higher melting point than form I*b* (denoted II). This shows that form I*a* is more thermodynamically stable than I*b* at the melting point. The I*a* polymorph (XAQXOE) was predicted to be the more stable form at all temperatures. At 0 K it is 3.9 kJ/mol more stable than form I*b*. The energy difference increases with temperature such that the free energy difference becomes 6.8 kJ/mol at the predicted melting point of 474 K. The difference in entropy at this temperature is predicted to be 7.3 J/mol·K.

I have here attempted to confirm the existence of diverging free energy curves. Unfortunately, this was not possible because of a lack of experimental data on the systems mentioned above. Future experimental work on these compounds is highly desirable and can hopefully shed some light on the thermodynamics of these crystal structures. Also further computational studies using even more accurate energy models would be useful.

5.3.7 Studies of other selected systems

I now present some selected results of more in-depth studies of polymorphic systems. Besides the calculations performed for the benchmarking in Chapter 3, phonon densities of states and free energy curves were calculated for several polymorphs of the following compounds: acridine (ACRDIN), 1,2-ethanediamine (ETDIAM), *m*-nitrophenol (MNPHOL), paracetamol (HXACAN), glutaric acid (GLURAC), cyclo-octa-sulfur (FURHUV), adipic acid (ADIPAC), 2,2′-dipyridylamine (DPYRAM), theophylline (BAPLOT) and pyrazine-2-carboxamide (PYRZIN).

Phonon densities of states were plotted for several crystal structures. In Figure 5.32 and 5.33, the phonon spectrum, and its temperature-dependence, is shown as Gaussian kernel densities of the phonon frequencies calculated in the quasi-harmonic approximation at each temperature. The quasi-harmonic method results in a very realistic softening of the vibrational modes. Note that using a single Grüneisen parameter to model the frequency shift is probably inadequate, since the temperature-dependence varies for different frequencies. Additional phonon spectra were published in reference [3] and are not reproduced here.

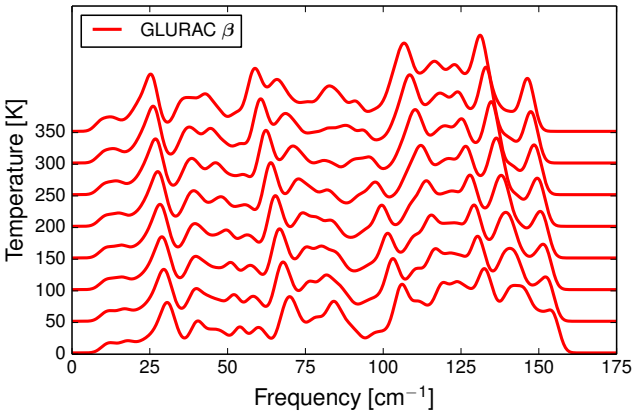


Figure 5.32: Phonon densities of state at different temperatures for glutaric acid β (GLURAC03).

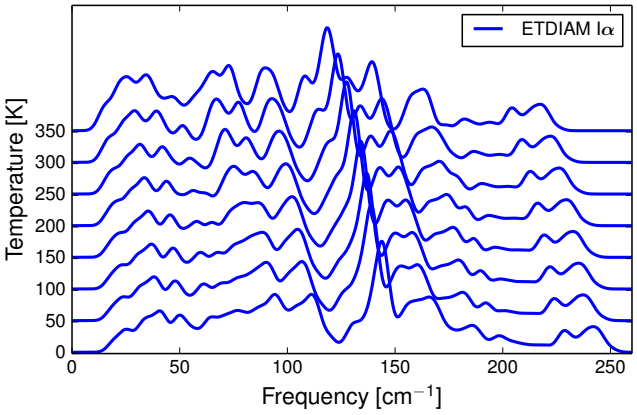


Figure 5.33: Phonon densities of state at different temperatures for 1,2-ethanediamine polymorph Iα.

The calculated free energy curves for adipic acid are shown in Figure 5.34. These can be compared to experimental data. Adipic acid (1,6-hexanedioic acid) has three experimentally known polymorphs [423, 424]. There is a known enantiotropic phase transition between forms I and II between 130 and 136 K. Form I is thermodynamically stable at room temperature, while form II is stable below 130 K. Computationally, form II became unstable at high temperatures, as the elastic tensor became indefinite (Born

instability) [436, 437]. No transitions have been described in the triclinic phase III and it appears to be monotropic relative to the other forms.

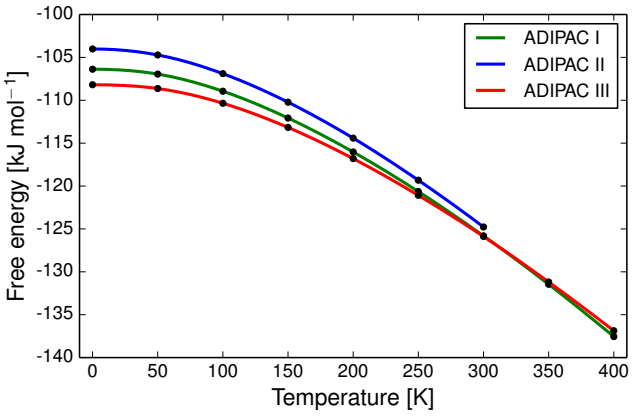


Figure 5.34: Quasi-harmonic free energy curves for polymorphs of adipic acid.

Form III of adipic acid is predicted to be more stable than the others below 310 K. This possibility is not strictly ruled out by experimental data. The calculated free energy curves are incorrect primarily because of the 0 K lattice energy ordering of the three polymorphs, while the relative entropies are plausible.

Figure 5.35 show the predicted relative stabilities of three polymorphs of theophylline (BAPLOT). Form I is experimentally known to be a high-temperature polymorph, enantiotropically related to form II. Form II was believed to be the thermodynamically most stable form at ambient conditions until form IV was discovered. There is a (monotropic) transition from IV to II [52]. The structure of form IV becomes unstable around 450 K in the quasi-harmonic approximation. Theophylline is an interesting system because of the plenitude of available experimental and computational data. It has for instance been the subject of CSP calculations, leading to the discovery of at least two new unstable polymorphs (named VI and VII) [50, 51, 475].

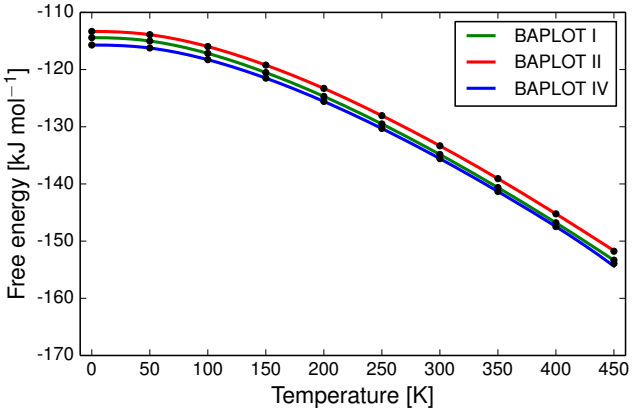


Figure 5.35: Quasi-harmonic free energy curves for polymorphs of theophylline. Polymorph IV changes structure slightly above 450 K, becoming less stable.

The available experimental data is not unambiguous, but suggests theophylline form IV is thermodynamically stable between 295 and 483 K. Form II is then the most stable between 513 and 536, and form I is favoured up to the melting point. The calculated free energy curves correctly identify from IV as the most stable at low temperatures, and that form I has a higher entropy than the other structures. Form I is however incorrectly found to be monotropically related to form II. Note the very small energy differences between these polymorphs, which are smaller than the expected errors in the energy model. Again, it seems the entropies are trustworthy, but the stability ranking fails because of incorrect ranking by the static lattice energy.

5.4 Concluding remarks

The most important observation is that polymorphs have a strong tendency to have very similar properties. Due to compensation between enthalpies and entropies, free energy differences between polymorphs are typically smaller than lattice energy differences, making it very challenging to correctly reproduce temperature-dependent polymorph stability rankings. The methods for calculating accurate vibrational energies and entropies are quite well developed, but the errors in calculated static 0 K lattice energy are still an important limitation. Even more accurate force fields are still needed for organic molecular crystals.

The results of this study should be valuable for discussion and analysis of polymorphism in general. An understanding of the expected energy differences between polymorphs forms a foundation for examinations of the influence of specific structural features and particular interactions on polymorph relative stabilities.

The full set of crystal structures used in this Chapter and the calculated properties have been published as the *Nyman Polymorph Library* (NPL2016) in the supplementary information to reference [3]. Further details of the contents are available, see Appendix D. This library of polymorph pairs should facilitate further studies into general properties of polymorphs.

One area where an understanding of polymorph energies is crucial is the *ab initio* prediction of crystal structures, which is usually performed by ranking computer-generated crystal structures by their calculated lattice energies [441, 25, 141, 5]. The fact that lattice energy differences tend to be so small demonstrates the challenge involved in correctly ranking the energies of predicted structures. The differences in calculated energies between observed and unobserved predicted crystal structures are typically as small as the polymorph energy differences seen here. Furthermore, given that observed polymorphs can differ by up to 10 kJ/mol in lattice energy, all predicted crystal structures within this energy range from the most stable structure can be seen as potentially

observable polymorphs. Such an energy range frequently includes large numbers of putative crystal structures [384, 22]. In fact, our distribution of polymorph lattice energy differences closely matches the relative lattice energies of observed crystal structures in crystal structure prediction studies of small organic molecules [476]. Some reasons why predicted polymorphs outnumber observed polymorphs have been discussed by Price [22].

A further observation relates to the use of lattice energy versus free energy in predicting relative stabilities of predicted polymorphs. Free energies should be used to assess the true thermodynamic stabilities of structures, but are often approximated by calculated lattice energies, due in large part to the added complexity and computational expense of free energy calculations. However, the small magnitude of lattice energy differences that we find between known polymorphs highlights the fact that it takes very little vibrational energy to cause a re-ranking of polymorph stability. While a small number of studies have shown promising results from the inclusion of dynamical effects in crystal structure prediction, either through a lattice dynamics [300, 477, 135, 478] or molecular dynamics [479, 480, 461] approach, lattice energy-based predictions are still the norm.

Lattice energy minimisation and rigid-molecule harmonic lattice dynamics have been applied to understand thermodynamic property differences for a large set of experimentally determined, non-disordered, packing polymorphs of organic molecules. The main results are those summarised in Figure 5.4–5.7 and 5.29, showing the distribution of static and vibrational energy differences between polymorphs. Given the large size of the structure set (1061 crystal structures in 508 polymorph families), and quality of the energy model, see Chapter 3 these should faithfully reflect true energy differences and be of value in discussing and understanding polymorphism.

Unsurprisingly, the lattice energy differences between polymorphs are typically very small and are less than 7.2 kJ/mol in 95% of polymorph pairs. Entropies dominate the vibrational contribution to relative free energies and, while these contributions to relative free energies are typically small ($|\Delta F_{\mathrm{vib}}| < 2$ kJ/mol in most cases), they can be large enough to significantly affect the calculated relative stability of polymorphs.

ΔF_{vib} and ΔE_{latt} are of opposite sign in 69% of the polymorph pairs, so that polymorph free energies usually converge with increasing temperature and will eventually cross. By $T = 300$ K, we find that vibrational contributions swap the stability order of 9% of polymorph pairs. At least 21% of polymorph pairs are enantiotropic. A sizeable fraction of polymorph pairs should also have diverging free energy curves. Predicted diverging free energy curves could however not be confirmed by comparisons to experimental data.

Correlations of energy and entropy differences with density are weak, and there is no evidence that polymorphs of hydrogen bonding and non-hydrogen bonding molecules show different trends in thermodynamic differences. Therefore, it is difficult to predict relative stability or thermodynamic behaviour based on these coarse descriptions of

structure. The empirical rules suggested by Burger & Ramburger may have a weaker predictive ability than previously thought.

Based on the evidence presented here, computational studies of polymorph stability, including crystal structure prediction, should not be restricted to static lattice energy calculations. The influence of lattice vibrations is important in judging the true stability difference between polymorphs and a necessary consideration for the anticipation of temperature-driven phase transitions. Since the errors introduced by restricting lattice dynamics to a single **k**-point are large, the energy methods that are developed for use in the context of crystal structure prediction should be sufficiently efficient to allow adequate sampling of phonons without an unacceptably large computational expense. Currently, only atom-atom potentials meet these criteria and we believe that the continued development of accurate atom-atom models for such studies is necessary.

Calculated vibrational energies and entropies are realistic and sometimes in good agreement with experimental data. The complete free energy curves are however often not in agreement with experimental data. This is primarily because of errors in the 0 K relative lattice energies; if polymorphs are incorrectly ranked by lattice energy the free energy relationship will also be wrong.

Quasi-harmonic lattice dynamics calculations have been performed on a large set of polymorphs of organic molecules, as a general assessment of the temperature-dependence of their properties, and to gauge the magnitude of typical property differences between polymorphs. The force field based lattice dynamical methods described here offer a good compromise between accuracy and computational cost, facilitating well-converged lattice-vibrational calculations on hundreds of crystal structures.

Temperature-dependent predictions for individual polymorph systems are still unreliable. However, because of the large set of structures studied here, the overall trends should be resolved well. Accordingly, it is concluded that at least one fifth of molecular polymorph pairs are enantiotropic. To my knowledge, this is the first prediction of how common enantiotropism is in general.

Bulk and shear moduli have been calculated for the entire set of crystal structures. These mechanical properties have a very pronounced temperature-dependence and, while polymorphs tend to have very similar elastic moduli, the differences can be important– polymorph differences of up to about 2 GPa were seen in the low temperature bulk and shear moduli.

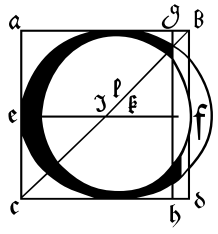
Realistic thermal expansions and the temperature-dependence of vibrational spectra were reproduced, often in good agreement with experimental data.

In crystal structure prediction, harmonic approximation free energies are becoming more commonplace [5]. Since thermal contributions can be significant, this should improve the predictive ability, and also provide additional insight regarding the properties of the

predicted structures. Even when accounting for thermal expansion, lattice vibrations never change the relative stability between polymorphs by more than 5 kJ/mol, at any temperature, see Figure 5.29. While vibrational energy contributions can, therefore, be important and should not generally be left out of polymorph energy studies, the impact of thermal expansion on the vibrational energy contributions is small, especially when considering free energy differences between polymorphs. Thus, thermal expansion can safely be ignored for routine CSP-purposes, but as we saw in Chapter 3 it should be included when modelling crystal structures at high temperatures, or for benchmarking purposes [2].

Chapter 6

Prediction of polymorphs of chloridazon



hloridazon is an old herbicide. In this chapter I describe the crystal structure prediction of chloridazon and the subsequent experimental confirmation of several of the predicted structures. The focus is on the combined use of CSP and experimental work for the structure-determination of organic solid forms.

The work described here was performed as a collaboration between the author and a group at Innsbruck University. The author performed the CSP-calculations. The experimental work was performed by Ana Ivanović in Ulrich J. Griesser's laboratory in Innsbruck.

6.1 The compound chloridazon

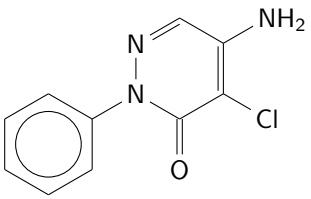


Figure 6.1: Chloridazon.

Chloridazon (also pyrazon, 5-amino-4-chloro-2-phenyl-pyridazin-3(2H)-one), see Figure 6.1 is a herbicide that inhibits photosynthesis in plants. The compound has been widely applied to beets as a pest control since 1964 [481].

Chloridazon research has focused on the herbicidal effect and the environmental degradation of the compound [482]. Recent experimental and computational results suggest that chloridazon intercalates with double-stranded DNA [483], potentially making it more toxic than previously thought.

One crystal structure of chloridazon was determined by Jacobson *et al.* in 1978 [484]. That structure is deposited in the Cambridge Structural Database as PYRZON. Apparently no other crystal structures of chloridazon have been observed, despite the compound being researched, manufactured and used on an industrial scale for several decades.

A computational prediction of polymorphs of chloridazon was performed without knowledge of experimental results (except the previously mentioned crystal structure). The prediction was followed by an experimental screening for solid forms. The results of the combined computational and experimental polymorph screening are described here. In addition, selected results of differential scanning calorimetry (DSC) and powder X-ray diffractometry (PXRD) are presented.

6.2 Crystal Structure prediction of chloridazon

6.2.1 Molecular conformations

A conformational search was performed with a Monte Carlo simulation. The molecular geometry was perturbed in normal mode coordinates to enhance the conformational sampling. The conformers were geometry-optimised by minimising the energy calculated with the OPLS-AA 2005 force field [89]. Generated conformers were clustered using an RMSD > 0.2 Å criteria.

Two conformers were found in the Monte Carlo search. Both conformers were geometry-optimised at the B3LYP-D3 / 6-311G** level of theory in Gaussian 09 [407]. During DFT geometry-optimisation, the two conformers converged to the same (global) minimum. The global minimum gas phase conformer has a dihedral angle of 37.407° around the bond between the two rings of the molecule.

In the force field model, the amino group was planar, but at the B3LYP level of theory, the lowest-energy conformation has a pyramidal amino group. Ideally, crystal structure prediction should be performed such that all amino group geometries are considered in order to sample all of crystal phase-space [43]. This was not done here, however flexible-molecule geometry-optimisations will to some extent alleviate this by allowing the amino groups to flex in response to crystal packing forces, see below.

A determination of the conformational energy with respect to the central dihedral angle was performed at the B3LYP-D3 / 6-311G** level of theory to study the stiffness of the central dihedral. The dihedral angle is soft; to twist the molecule a full turn requires less than 8 kJ/mol, see Figure 6.2. Therefore, all conformations should be considered possible. Most polymorphs have molecular geometries that deviate little from an energy-minimum, but intermolecular packing forces can occasionally stabilise conformers up to

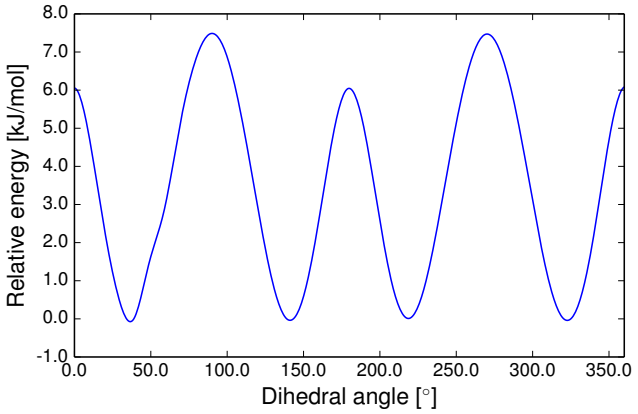


Figure 6.2: The relative energy of the chloridazon conformation with respect to the value of the central C_C_N_N dihedral angle.

about 20 kJ/mol from the minimum energy conformer [147]. I therefore manually made two additional conformations with dihedral angles distorted by ±20° from the minimum energy conformer, see Figure 6.3.

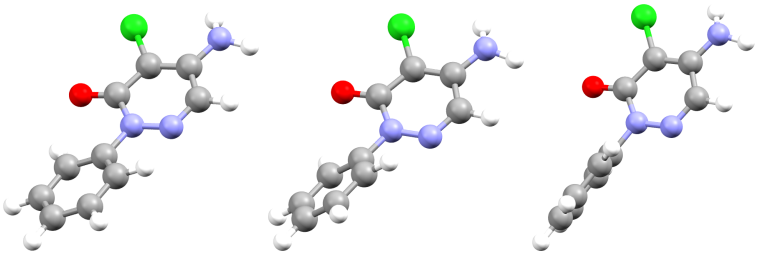


Figure 6.3: Chloridazon conformations used in the CSP. The central dihedral angle is 17.4°, 37.4° and 57.4°, respectively.

Most space groups are achiral because they contain inversions. Ignoring the pyramidalisation of the amino group, only the central dihedral angle confers chirality to the molecule. The dihedral angle of chloridazon is 180°-periodic, because the phenyl group in symmetric. These symmetry considerations greatly reduce the number of conformations that must be sampled explicitly. Hence, the three conformers were deemed enough to sample the intramolecular flexibility of chloridazon.

6.2.2 Trial structure generation

Trial crystal structures were generated using an in-house Python program, Glee, written by Peter J. Bygrave, Joshua Campbell and David H. Case. It has been described elsewhere [160]. The program samples crystalline degrees of freedom by random numbers drawn from a quasi-random low-discrepancy Sobol sequence [159]. This ensures an unbiased and uniform sampling of crystal packing arrangements. The molecule is positioned in the asymmetric unit by random translations. The molecular rotation is

sampled by Shoemake's method to give a uniform sampling of positions on a sphere [431].

Since the chosen molecular conformations are similar, unit cell dimensions were sampled independently for each conformer, starting from different random number seeds. Unit cell angles were sampled between 45° and 135°. In order to sample unit cell lengths, the program assigns a volume to each molecule by projecting each atomic position onto the eigenvectors of the moments-of-inertia matrix, and determining a cuboid from the maxima and minima along each axis. The volume of the cuboid is multiplied by the number of symmetry operations to yield a target volume for the unit cell. In order to generate structures where molecules do not overlap, the convex hull of each molecule is constructed and the unit cell is expanded until the hulls no longer overlap. If the expansion results in a unit cell volume greater than 2.1 times the target volume, the structure is deemed to be too far from any realistic energy-minimum and rejected.

Trial structures with a single molecule in the asymmetric unit ($Z' = 1$) were generated. Ten thousand crystal structures were generated for each of the three conformations in the six most common space groups (2, 14, 15, 29, 33, 61). Less common space groups (1, 4, 5, 7, 9, 13, 18, 19, 43, 56, 60, 76, 78, 88, 92, 96, 144, 145, 148) were sampled by generating 5000 structures of each conformation. In total 465,000 trial structures were generated.

Structures with more than one molecule in the asymmetric unit were not sampled explicitly, but some will still have been obtained from $Z' = 1$ structures that broke symmetry during geometry-optimisation in DMACRYS [168], see below.

6.2.3 Geometry optimisation

The trial structures were geometry-optimised by minimising the lattice energy in several stages. The hybrid force field energy model and the general methodology for these calculations were described in Chapter 2.

The first geometry-optimisation used a rigid-molecule lattice energy minimisation. The W99rev6311 force field and distributed multipoles were used for intermolecular interactions [225], see Chapters 2 and 3. Calculations were performed with the program DMACRYS [168]. The minimiser uses a quasi-Newton-Raphson algorithm (BFGS) [438, 439, 485], which means that the second derivative matrix (the Hessian) will normally be approximate during the minimisation. Convergence to a true energy-minimum was enforced by recalculating the Hessian matrix when the optimisation converged to a stationary point. The Hessian matrix gives information about the local curvature of the potential energy surface. If the stationary point was a saddle point, additional optimisation steps were taken along the lowest-energy eigenvector, allowing symmetry breaking when necessary [168, 486]. This allows the minimisation to proceed past saddle

points on the potential energy surface. In other words, space group symmetry was not enforced.

As an experiment, symmetry breaking during geometry optimisation was explicitly allowed in this CSP. It was hypothesised that this may be an efficient way to obtain $Z' > 1$ structures, since they tend only to be slightly symmetry-broken [12]. The purpose of the symmetry breaking was twofold. Firstly, it ensures that the predicted structures correspond to true energy-minima with positive-definite Hessian matrices. Secondly, the symmetry-broken structures generated in this way should correspond to realistic pseudo-symmetric $Z' > 1$ structures [12], which are otherwise difficult to obtain in CSP calculations, because of the much larger space of packing possibilities for such structures.

In order to see how efficient this method is for generating $Z' = 2$ structures, and to estimate how common it is that a crystal structure optimises past a saddle point, a sample of 5000 crystals in each of the following space groups were investigated: $P\bar{1}$, $P2_1/c$, $P2_12_12_1$, $Pbca$, $R\bar{3}$ and $I4_1/a$. The geometry-optimisation can occur through two different mechanisms in DMACRYS. The first mechanism, here referred to as 'SEIG1', means that the optimisation proceeds past a saddle point by taking additional optimisation steps along a totally symmetric representation [10], and the space group symmetry and the size of the asymmetric unit are preserved. For the second method, 'RMEIG', the representation is not totally symmetric, requiring a change of space group and a doubling of the asymmetric unit contents. In such cases, NEIGHCRYS was used to remove the symmetry operation corresponding to the negative Hessian eigenvalue and DMACRYS was restarted.

Spacegroup	#Symm	SEIG1	RMEIG	Tot
$P\bar{1}$	2	1747	0	35%
$P2_12_12_1$	4	1344	119	27%
$P2_1/c$	4	1229	568	25%
$R\bar{3}$	6	424	44	8%
$Pbca$	8	1489	544	30%
$I4_1/a$	8	2141	205	43%

Table 6.1: The number of saddle point passages by two different methods, SEIG1 and RMEIG, during geometry optimisation of 5,000 structures of 6 space groups. #Symm denotes the number of symmetry operations in the space group.

The results of this exercise are summarised in Table 6.1. It is clear from these results that it is common for a crystal structure to first converge to a saddle point and that for the structures examined here, 30% in total could be successfully optimised to a more stable structure with the SEIG and RMEIG methods. Most likely, this is because the crystal structures coming out of the trial structure generator GLEE are often far from an energy minimum. By verifying that the Hessian is really positive definite, about 98% of the 465,000 trial structures could be successfully optimised to true energy-minima, a number significantly higher that normally achievable. About 5% broke symmetry

and yielded $Z' > 1$ structures. This suggests that for CSP purposes it is useful to calculate the second-derivative matrix and to make sure that it does not have negative eigenvalues. Hence, to calculate phonon frequencies serves two purposes, to make sure that the predicted structures are stable, and to calculate free energies.

6.2.4 First clustering

Duplicate crystal structures were removed by comparing crystal similarity calculated as a root mean square deviation in atomic coordinates for a cluster of 20 overlayed molecules using the program COMPACK. Hydrogen atom positions were not included in the calculation. Because of the large number of structures, clustering was at this stage only performed within each conformer and within each space group.

6.2.5 Flexible-molecule optimisation

A subset of 1770 structures below 15 kJ/mol from the lowest-energy structure were chosen for flexible molecule optimisation. The sum of the intramolecular B3LYP-D3 and the intermolecular force field energy was minimised with respect to lattice parameters and intramolecular Z-matrix parameters. The dihedral angle between the molecule's rings and dihedral angles and bond angles of the polar hydrogen atoms of the amino group were free to adjust in response to packing forces. The calculations were performed with CRYSTALOPTIMIZER. The W99rev6311 force field was used for intermolecular interactions.

6.2.6 Final clustering

After flexible-molecule optimisation, a final clustering was performed across all conformations and all space groups. The crystal packing similarity tool in MERCURY was used to calculate the RMS distance between atomic coordinates in a cluster of 15 overlayed molecules.

6.2.7 Final scoring

The experimentally known polymorph, here referred to as form A, and deposited in the CSD as PYRZON [484], was found in the predicted crystal landscape. It was ranked second by the W99rev6311 lattice energy, 0.55 kJ/mol above the lowest-energy structure. Re-optimising with the more accurate W99rev6311P3 force field, see Chapter 3, the structure that matches PYRZON becomes the global energy minimum, see landscape in Figure 6.4. The geometry of the experimental structure was reproduced with an

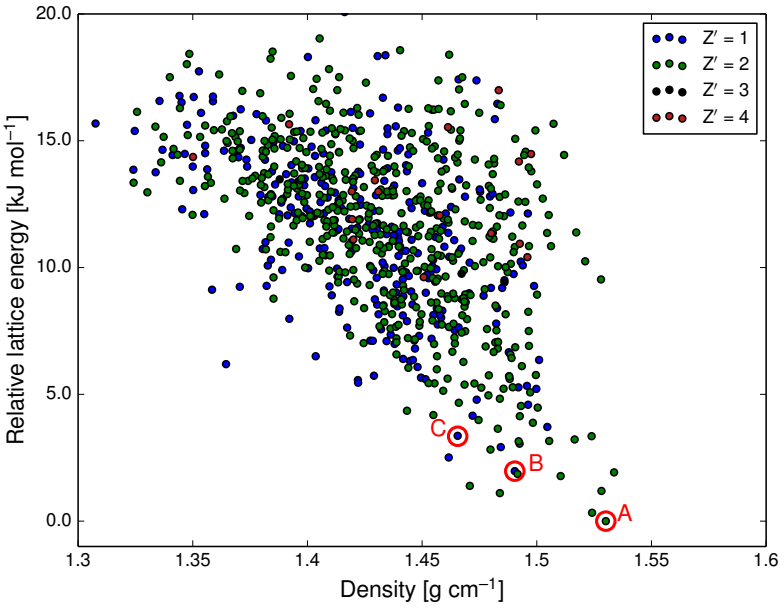


Figure 6.4: Crystal lattice energy landscape of chloridazon. Polymorphs A, B and C encircled in red.

$RMSD_{20}$ of 0.335 Å. An overlay of molecules extracted from the experimental and the predicted structures are shown in Figure 6.5.

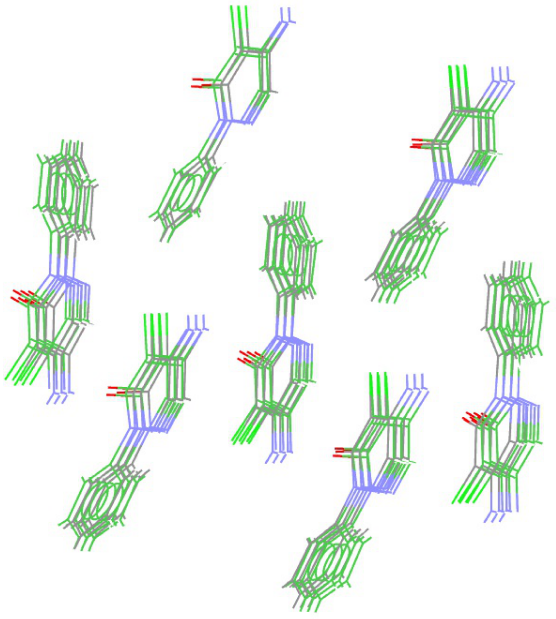


Figure 6.5: Best matching overlay between the 1978 crystal structure and the predicted structure of polymorph A.

The 100 lowest lattice energy structures were subjected to harmonic approximation lattice dynamics calculations. Linear co-prime split supercells were used to sample **k**-points. The W99rev6311P3 force field was used. The vibrational free energy was

calculated at 300 K. The results are displayed in Figure 6.6. The vibrational contribution to the ranking had a small effect. The experimentally known polymorph PYRZON was ranked second by the Helmholtz free energy, *i.e.* in this case, including the vibrational energy did not improve on the predicted stability ranking. This again indicates that the force fields used here are not able to resolve energy differences on the order of 1 kJ/mol.

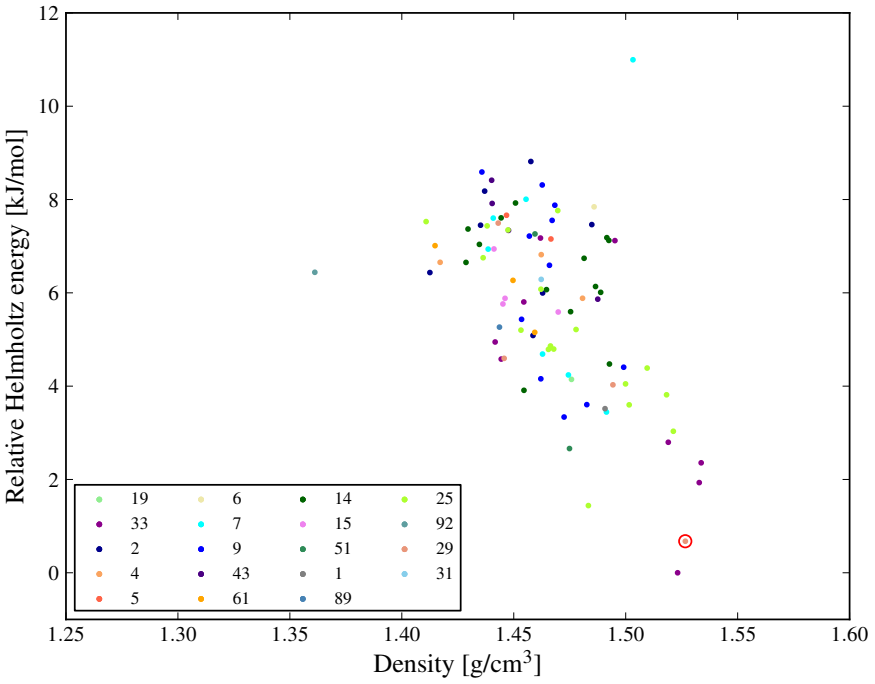


Figure 6.6: Crystal free energy landscape of the 100 most stable structures at 300 K. The experimentally known polymorph A (CSD PYRZON) is ranked second (encircled). Data point colour denote space group number.

The most common space groups among the lowest-energy structures are number 33, 29, 9, 14, and 2, see Figure 6.6. These are also very common space groups for molecular crystals in general. Many of the predicted low-energy structures have two molecules in the asymmetric unit, more so than one would expect for organic crystal structures in general [487]. An explicit search of $Z' = 2$ structures can therefore be expected to result in additional low-energy structures.

6.3 Experimental screening of polymorphs of chloridazon

The experimental work consisted of a comprehensive screening for polymorphs by crystallization experiments with a large number of organic solvents and subsequent analyses of the obtained crystal forms by PXRD, DSC and TGA, see Chapter 2.

The experimental work was performed by Ana Ivanović under the supervision of Ulrich J. Griesser at the University of Innsbruck. Experimental results were kindly provided

by Ana Ivanović, and some are reproduced here in order to compare them to some of the computational predictions.

6.3.1 Observed crystal forms

Chloridazon, which was not previously known to be polymorphic, exhibited no less than seven distinct solid forms when crystallised from organic solvents. The crystal structures were named A–G, with A being the previously known structure [484]. Forms B, C, D and E are new polymorphs of pure chloridazon. Form F is a highly unstable 1,4-dioxane solvate and form G is a benzene solvate, apparently iso-structural with the 1,4-dioxane solvate, see Figure 6.7.

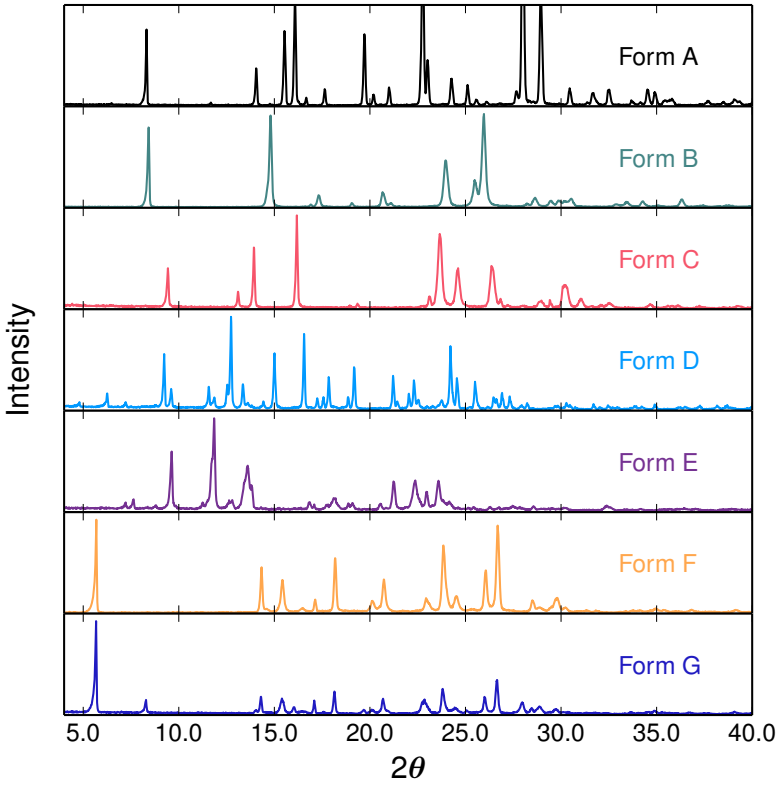


Figure 6.7: Normalised powder X-ray diffraction patterns of all seven chloridazon solid forms. Note the apparent iso-structurality of the solvate forms F and G.

Forms D and E were too unstable for further experimental characterisation as they spontaneously and rapidly transformed to form C. They could not be obtained again after the first crystallisation experiment, *i.e.* they are examples of disappearing polymorphs, where trace amounts of relatively stable crystal forms are believed to work as sites of nucleation and growth, preventing less stable forms to crystallize [32, 31].

6.3.2 Powder X-ray diffraction analysis

Powder X-ray diffraction patterns were used to identify the polymorphs. Diffraction of X-rays by micro-crystalline powders were measured using Cu K_α radiation with a wavelength of 1.5419 Å. Computationally simulated diffraction patterns obtained from the predicted structures allowed the confirmation of the structure of form A and structure determination of the novel polymorphs B and C.

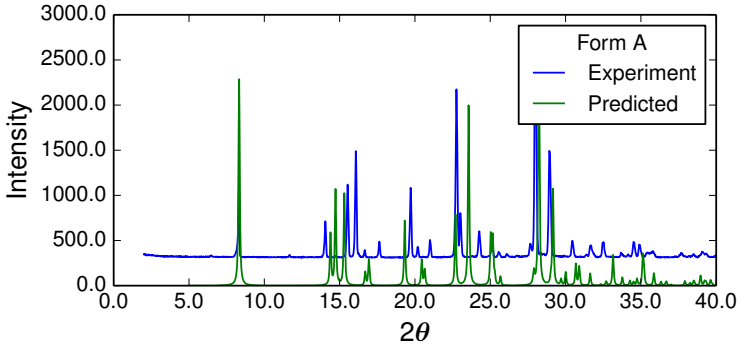


Figure 6.8: Powder X-ray diffraction pattern of chloridazon form A and the best matching predicted structure. The patterns have been normalised to have equal area under the curve. The experimental pattern has been shifted upwards to aid comparison by eye.

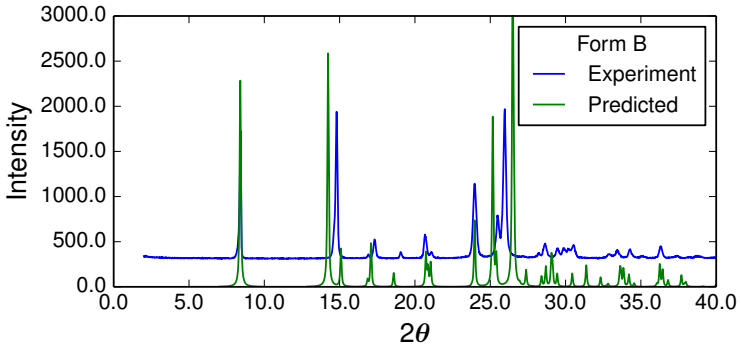


Figure 6.9: Powder X-ray diffraction pattern of chloridazon form B and the best matching predicted structure. The patterns have been normalised to have equal area under the curve. The experimental pattern has been shifted upwards to aid comparison by eye.

The experimental patterns could be directly compared to the simulated patterns by eye, after they had been normalised to have equal area under the curve, see Figures 6.8, 6.9 and 6.10. This way the crystal structures of forms B and C could be determined. Note that it is the positions of the diffraction peaks that are the most important for crystal structure comparisons, not the relative intensities. Automated computational comparisons of powder patters were not attempted here, but would help in cases where

the number of structures is large [349, 350]. Indexing of the powder pattern peaks to Miller indices was performed by Doris Braun to conclusively confirm the matches.

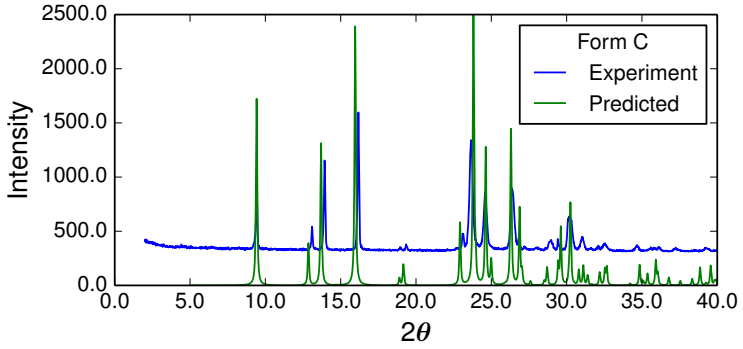


Figure 6.10: Powder X-ray diffraction pattern of chloridazon form C and the best matching predicted structure. The patterns have been normalised to have equal area under the curve. The experimental pattern has been shifted upwards to aid comparison by eye.

For form A, an overlay picture in Figure 6.5 shows the structural agreement between the originally published structure [484] and the structure of polymorph A obtained in the prediction. The $RMSD_{20}$ is 0.335 Å.

The agreement between the predicted structure of form B and the structure solved from the experimental powder pattern is shown in Figure 6.11. The $RMSD_{20}$ is 0.367 Å. The experimental structure suggests a planar conformation of the amino group, as is commonly seen experimentally due to the temporal averaging of the positions of the hydrogen atoms that tunnel between the two pyramidal conformations.

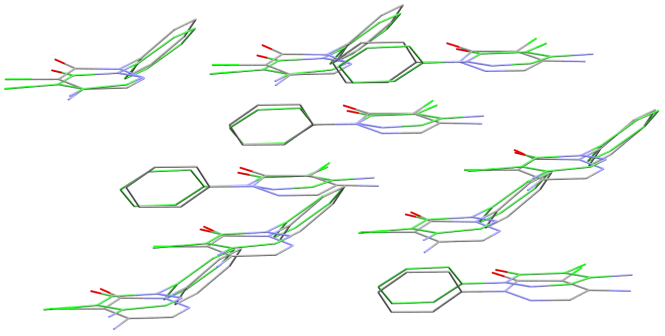


Figure 6.11: Overlay picture showing the structural similarity between the predicted and experimental structures of chloridazon form B.

The overlay between the predicted structure of polymorph C and the solved experimental structure is displayed in Figure 6.12. The $RMSD_{20}$ is 0.190 Å, which is an unusually good match, see Chapter 3.

Attempts were also made to identify plausible empty iso-structural framework structures

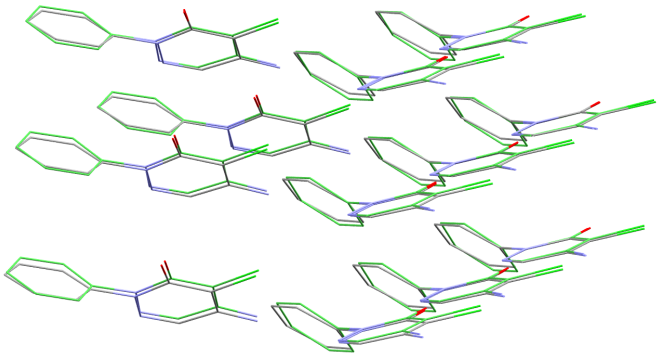


Figure 6.12: Overlay picture showing the structural similarity between the predicted and experimental structures of chloridazon form C.

of the two solvates, forms F and G. The solvates were very unstable and could not be obtained in desolvated form, indicating that the empty framework is not an energy minimum. Hence, the efforts to find the apo-host in the CSP set were futile. This method of predicting multicomponent structures is likely better suited for weakly interacting guest molecules in true clathrates, rather than solvates, see Chapter 7.

6.3.3 Differential scanning calorimetry

Differential scanning calorimetry was performed on polymorphs A, B and C of chloridazon. The results are shown in Figures 6.13 and 6.14. The scans revealed an endothermic solid-to-solid phase transition from form A to Form B at about 160–170°C, suggestive of an enantiotropic relationship between forms A and B.

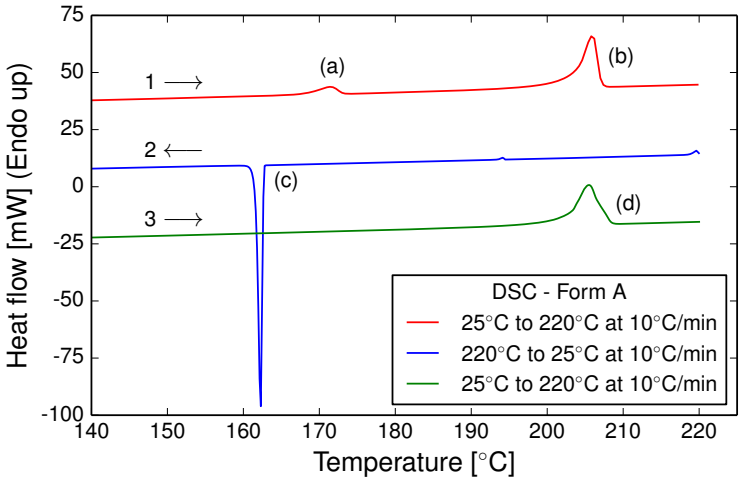


Figure 6.13: Differential scanning calorimetry starting with polymorph A, with the endothermic transition (a) to form B. Point (c) indicates crystallisation of form B from the melt. Points (b) and (d) show the melting of form B. To aid the eye, the curves 1 and 3 have been moved vertically by ±25 mW, respectively.

Upon heating, Form C transitions exothermically to form A, see Figure 6.14 (a). This indicates a monotropic relationship between form A and C. The thermodynamic relationships were confirmed by additional experiments, which are not reported here.

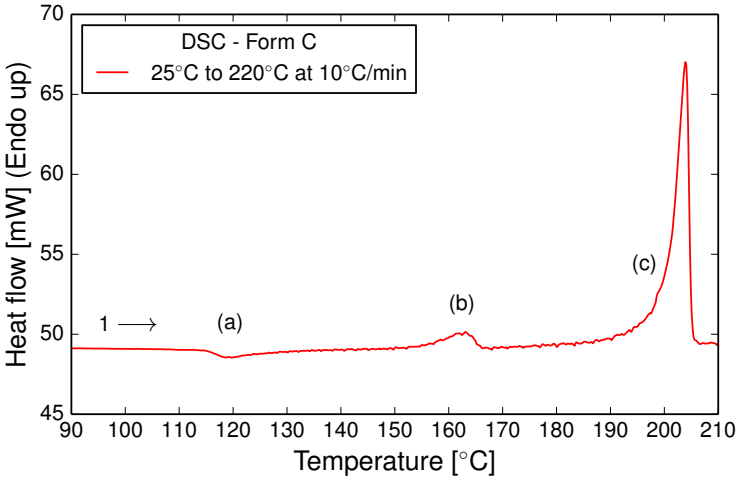


Figure 6.14: Differential scanning calorimetry starting with polymorph C, with a transition at (a) to form A. The event at (b) is the transition of form A to B, and (c) is the melting of form B.

The results indicated that polymorphs A and B are enantiontropic, and that form B is thermodynamically stable at temperatures above *ca* 170°C. Form C appears to be monotropically related, and metastable with respect to both form A and B. Form A was confirmed to be thermodynamically stable at ambient conditions.

The enthalpy-difference between form C and A is 1.1 kJ/mol at 167°C. The lattice energy difference should be practically the same, since heat capacities do not differ much between polymorphs [1]. Form B has 3.5 kJ/mol higher enthalpy than form A at 113°C. These results are in agreement with typical energy-differences between polymorphs, see Chapter 5.

The computational energy landscape correctly predicted form A to be the more stable form, but incorrectly ranked the order of stability between forms B and C. The experimental results indicate that these are separated by 2.4 kJ/mol of enthalpy. The W99rev_6311P5 force field yields instead a (lattice) energy difference of −1.4 kJ/mol. Hence the energy model used in the calculations cannot reliably resolve these small energy differences.

6.4 Concluding remarks

This chapter demonstrates the combined use of CSP and experimental PXRD for the structure determination of novel crystal forms. The set of predicted structures can

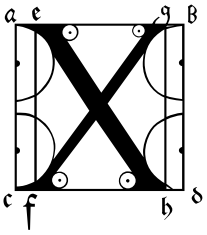
be searched for powder patterns that match experimental data and used as starting point for Miller-indexing or Rietveld refinement, allowing conclusive determination of the structure of polymorphs even from poor or ambiguous experimental results.

The relative stability of polymorphs B and C was incorrectly predicted, as revealed by differential scanning calorimetry. This is not unexpected, the energy models currently available are often not accurate enough to correctly rank polymorphs with energy differences smaller than 3 kJ/mol [391], as was the case here. Despite this, the crystal structure prediction was still a very useful complement to the experimental polymorph screening and diffraction experiments.

The latent polymorphism of chloridazon is remarkable. Despite being used on an industrial scale for decades, no polymorphs had been observed previously. This is probably because no active attempts to search for alternative crystal forms have been made until now. This is likely the case for many, perhaps most, organic compounds. Crystal structure prediction has taught us that there could be a large number of polymorphs for every compound.

Chapter 7

Prediction of fluorophenol xenon clathrates



enon nuclear magnetic resonance spectroscopy and crystal structure prediction are combined in this Chapter for the determination of the crystal structures of two fluorophenol xenon clathrates. The author performed CSP calculations while experimental and computational work on xenon NMR were performed by a group at Oulu University, Finland. The NMR experiments were performed by Marcin Selent and calculations of NMR shielding tensors were performed by Marcin Selent and Perttu Lantto. The results of this collaboration have been published as reference [4].

7.1 Background

Clathrates are unusual crystalline materials that form by the complexation between porous crystalline host materials and gases [488, 489, 490, 491]. The term clathrate originates from the work of Powell [488]. The term is specific to structures with discrete cage-like adsorption sites, unlike the more general term inclusion compound [6].

Phenol and several of its derivatives are known to form clathrates. Clathrates of, for example, phenol, *p*-cresol [73], *p*-fluorophenol [74] and hydroquinone (quinol, 1,4-dihydroxybenzene) have been described. Hydroquinone and its clathrates have been very extensively studied, both experimentally [80, 492, 493] and computationally [494, 76].

Only clathrates of water (hydrate clathrates) are more studied than hydroquinone. There is a great interest in the properties of hydrate clathrates because of the technical problems and economic costs associated with the formation of methane hydrate clathrate around natural gas- and petroleum wells [495].

In 1972 and 1973 Kazankin *et al.* reported the formation of xenon clathrates by several monosubstituted phenol compounds, including *o*- and *m*-fluorophenol. The results of their studies, including phase diagrams showing the pressure-temperature intervals of clathrate stability were published in Russian literature [496, 497]. Xenon enclathration by hydroquinone [498, 490, 82] is well-known but the fluorophenol complexes have not been described in English literature.

Combining ^{1}H and ^{35}Cl solid state nuclear magnetic resonance (ssNMR) and computational methods has previously been shown to be a practical method for distinguishing between polymorphs. This is because of both the dependence of the NMR chemical shielding to the molecular environment in a crystal structure, and the increasing reliability of computational methods for predicting chemical shielding parameters [357, 499, 500, 356, 501].

The method used in this Chapter involves first predicting which crystal structures are available to the host molecule by means of crystal structure prediction, and then performing chemical shielding calculations by density functional theory. The calculated shieldings are then compared to measured chemical shifts of the material in question. Previous applications of such methods to organic crystals have demonstrated that the isotropic ^{1}H chemical shifts are usually sufficient to unambiguously identify which of the predicted structures corresponds to the material under investigation [500].

The NMR properties of xenon [502, 503, 82, 371] included in porous structures gives a potentially very sensitive probe of clathrate structures. Xenon NMR has been shown to be useful as a method for porosimetry of materials, allowing the determination of pore size distributions [376, 374, 84].

Solid-state powder ^{129}Xe NMR is used here to determine the crystal structure of *o*- and *m*--fluorophenol xenon clathrates. The chemical shielding of ^{129}Xe is known to be highly sensitive to the local environment, and the variability of the shielding stems exclusively from environmental effects. The sensitivity comes primarily from the large and easily polarizable electron cloud. Although not used here, hyperpolarization by optical pumping can increase the ^{129}Xe polarizability by up to five orders of magnitude, making it an extremely sensitive probe of pore geometry [504, 505].

Computational modelling of ^{129}Xe NMR in clathrates began in 2004 [506, 507] with studies of clathrate hydrate structures. The calculations were performed on molecular clusters extracted from the crystal structure, not on periodic structures. Since the chemical shielding is a local effect, this approximation is often acceptable, and facilitates the use of hybrid density functionals, which are computationally cumbersome in periodic calculations. Later studies have shown that the relativistic speeds of electrons in xenon and thermal vibrational motion of atoms can significantly affect the results of NMR shielding calculations [508, 509, 510, 511]. Such effects are considered in the work

presented here, and confirmed to be so large that they must be considered if quantitative accuracy in the shielding tensor is desired.

Crystal structure prediction has previously been focused primarily on predicting one or a few low energy polymorphs [107, 130]. The most stable crystal structure of a given molecule is almost always close-packed [512], leaving no room for the inclusion of guest molecules. Dianin's compound is one notable exception to this rule [513].

Computational predictions of porous molecular crystals [514, 515] or solvates [516, 340, 341] are still rare. The difficulties lie in both the need for a very extensive sampling of possible crystal packings to find the high-energy porous framework structures, and in the need to sample multicomponent systems with variable or unknown stoichiometry [478]. The much larger space of packing possibilities for multicomponent crystals requires an efficient sampling in the generation of trial structures. Which space groups to explore in the prediction of multicomponent structures can also be an issue [432]. Common arguments about which space groups to sample based on Bayesian arguments may not be valid for clathrates.

The empty host molecule frameworks of several observed inclusion structures have however been shown to exist as local minima on the lattice energy surface [75, 44, 341]. This suggests a very efficient approach for the prediction of stoichiometric clathrate structures with weakly interacting guests: searching for stable, empty frameworks using CSP methods and subsequently inserting the guest into the pores.

This chapter describes comprehensive CSP calculations of *o*- and *m*-fluorophenol with the specific aim to predict their xenon clathrate structures. The success of the method is demonstrated by calculating ^{129}Xe NMR parameters that are then compared to experimentally measured spectra. The comparison of calculated and measured NMR parameters allowed the determination of the crystal structures of these two clathrates.

I begin by describing the crystal structure prediction. Then I describe how ^{129}Xe NMR parameters were calculated for the predicted structures. First a preliminary screening of predicted clathrates was performed by calculating NMR parameters with cluster models in order to rule out structure candidates. Then NMR shielding tensors were calculated in periodic boundary conditions with the GGA-density functional PBE, dispersion-corrected with the method by Tkatchenko and Sheffler, to yield a smaller set of plausible candidates. Finally, calculations at a high level of theory that included relativistic effects and the motion of the xenon atom were performed to pin-point the correct clathrate structure of each compound.

The experimental measurements and computational modelling of NMR tensors were performed by Marcin Selent, Perttu Lantto and others at the University of Oulu, Finland.

7.2 Crystal structure prediction of *o*- and *m*-fluorophenol

7.2.1 Molecular conformations

Both molecules are essentially rigid, except for the orientation of the hydroxyl group. The two compounds *o*-fluorophenol and *m*-fluorophenol both have a single dihedral angle (H-O-C-C) and the dihedral angle is likely to have a large impact on the possible crystal structures and their lattice energy. The dihedral angles of both molecules were examined by performing single-point energy calculations in steps of 5°, using GAUSSIAN 09 and the B3LYP hybrid density functional and Møller-Plesset second order perturbation theory [407, 246, 243, 517]. The Gaussian 6-311G** basis set was used.

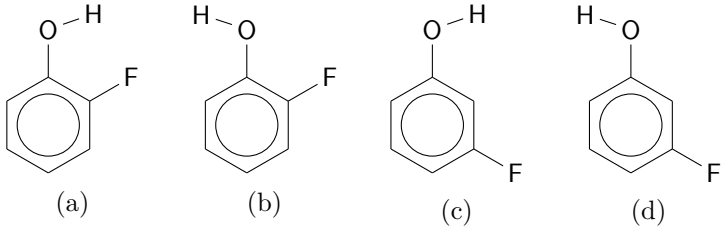


Figure 7.1: Conformers of the two fluorophenol isomers. a) *cis*- and b) *trans*-*o*-fluorophenol c) *cis*- and d) *trans*-*m*-fluorophenol.

The *cis*/0° and *trans*/180° conformers, where the –OH hydrogen is pointing towards the fluorine atom and away from it, respectively, have the lowest energy. In *o*-fluorophenol the difference in energy between the *cis* and *trans* conformers is 13 kJ/mol and in *m*-fluorophenol, it is only 0.1 kJ/mol. Hence, both conformers of *m*-fluorophenol can be assumed to be equally likely to occur in crystal structures. Both conformers of each isomer were used in the CSP.

7.2.2 Trial structure generation

A large number of hypothetical crystal structures were generated of the rigid B3LYP-optimised molecular geometries using Monte Carlo simulated annealing [106] with the POLYMORPH PREDICTOR module in Accelrys' MATERIALS STUDIO software. Searches were performed using space group symmetry in the most commonly observed space groups of known crystal structures in the Cambridge Structure Database (CSD) [304]. There are reasons to believe that clathrate structures occur selectively in certain space groups [21, 432], but searches were performed in a wide selection of space groups. Searches with one molecule in the asymmetric unit ($Z' = 1$) were performed in the 25 most commonly observed space groups ($P2_1/c$, $P2_12_12_1$, $P\bar{1}$, $P2_1$, $Pbca$, $C2/c$, $Pna2_1$, Cc, $C2$, $Pca2_1$, $P1$, $Pbcn$, $P4_12_12$, $P2_12_12$, Pc, $P4_1$, $Fdd2$, $Pccn$, $P2/c$, $P6_1$, $I4_1/a$, $P3_2$, $R3$, $R\bar{3}$, $C222_1$) and were repeated for both planar conformers of both molecules.

For structures with $Z' = 2$, searches were performed in $P1$, $P\bar{1}$, $P2_1$, $P2_1/c$ and $P2_12_12_1$ with the three possible combinations of conformers for each molecule.

Hypothetical crystal structures were generated by simulated annealing in the POLYMORPH module in MATERIALS STUDIO version 6.1.200. Several Monte Carlo simulated annealing calculations were performed starting with different random number seeds. About 10 runs were performed for each space group with $Z' = 1$ and 20 runs per space group for $Z' = 2$. The convergence of the search was monitored by looking at how the lowest-energy structures were found several times. The generated structures were optimised with the COMPASS force field, using partial charged fitted to the electrostatic potential with the CHelpG-method [88, 409, 190]. A cutoff radius of 18.5 Å was used for van der Waals interactions, while electrostatic interactions were calculated with Ewald summation.

7.2.3 Lattice energy minimisation

Structures with a COMPASS lattice energy within a 15 kJ/mol window from the lowest energy structure of each conformer were further refined with an anisotropic intermolecular potential model. The W99rev6311 force field with distributed multipoles was used throughout the CSP, as described in previous Chapters of this thesis. Fluorine atoms were treated as having an anisotropic repulsion, as described previously [216, 167].

Duplicate crystal structures were removed using the COMPACK program [518]. Interatomic RMS distances between clusters of 15 molecules from each crystal structure were used as a score of structure similarity.

The resulting unique structures were re-optimised using the CRYSTALOPTIMIZER [519] program to treat flexibility of the hydroxyl group. CRYSTALOPTIMIZER minimises the sum of intra- and intermolecular energies by calculating the strain due to crystal packing forces on the hydroxyl dihedral angle and allows this to relax. The minimisation is performed by iteratively computing the intermolecular forces on the flexible dihedral angle in DMACRYS, and optimising the intramolecular geometry in GAUSSIAN under the influence of the packing forces.

7.2.4 Xenon insertion

Plausible clathrate hosts were identified by searching for porous crystal structures. The theory behind void calculations was covered in Chapter 2. Void volumes and their centroid coordinates were calculated by placing a 1.2 Å probe at every point in a regular 0.2 Å grid spanning the unit cell. The PLATON software was used for void calculations [284]. This grid spacing results in well-converged cavity volumes, see Figure 2.14.

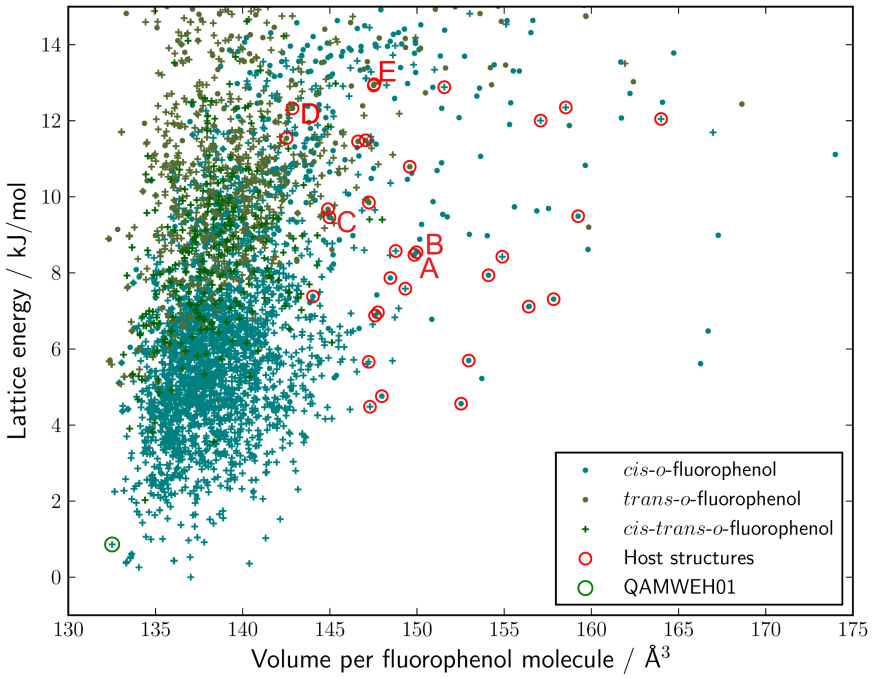


Figure 7.2: Crystal energy landscape of *o*-fluorophenol. The 33 porous clathrate host structures are encircled in red. The previously known high-pressure polymorph QAMWEH01 is encircled in green. The 5 candidates for which periodic modelling was performed are marked A–E, see the text for details.

Guest xenon atoms were inserted as described in Chapter 2. The host structures were converted to $P1$, *i.e.* all point group symmetry was relaxed. A single xenon atom was placed at the centroid coordinated of every cavity with a suitable volume [338, 339]. As described in Chapter 2, clathrates have a tendency to have cavity-to-guest volume ratios R_g around 59% with a standards deviation of 8 percentage points. Structures with cavities yielding host-to-guest volume ratios R_g of 59% ± 24 percentage-points, *i.e.* within 3 standard deviations. This corresponded to cavity volumes of 48–121 Å^3, which should be compared to the van der Waals volume of xenon, which is 42.2 Å^3 [220].

The obtained clathrate structures were again geometry-optimised to relax the position of the xenon atom and the host framework geometry in response to guest interactions. Although a Buckingham potential for xenon has been parametrised by Williams [221], the author performed a new parametrisation by minimising the error in reproducing several known van der Waals complex geometries and intermolecular interaction energies. Van der Waals dimers of Xe$\cdots$Xe [520], Xe$\cdots$H$_2$O [521], Xe$\cdots$CH$_4$ [522], Xe$\cdots$N$_2$ [523], Xe$\cdots$O$_2$ [524] and Xe$\cdots$C$_6$H$_6$ [525] were used. The parametrisation was performed with the Forcite module in Materals Studio.

The obtained xenon Buckingham parameters that were used in this work are shown in Table 7.1. Force field terms for interactions between xenon atoms and other atom types

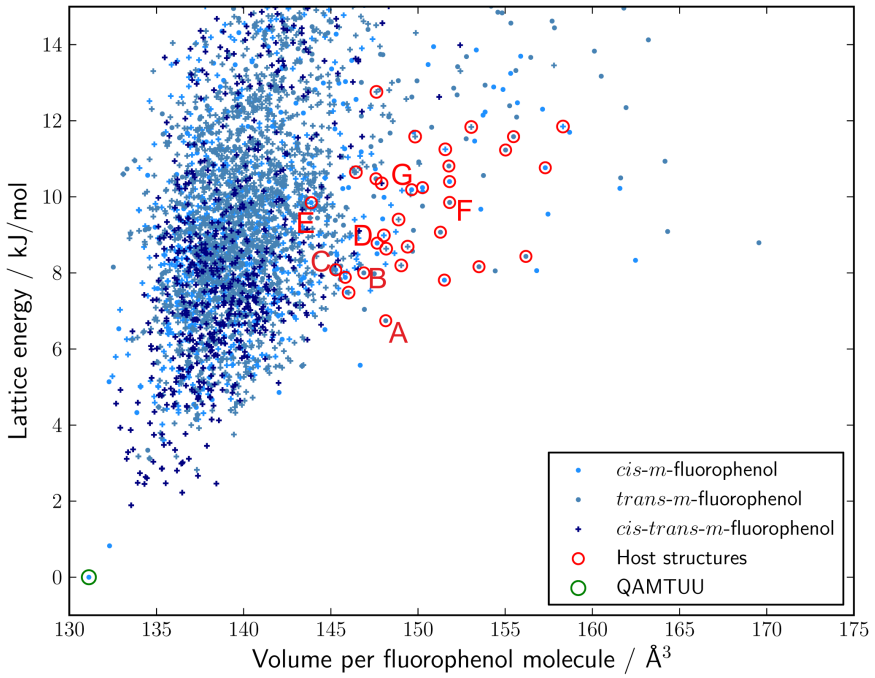


Figure 7.3: Crystal energy landscape of *m*-fluorophenol. The previously known stable form QAMTUU is encircled in green. The 32 predicted clathrate host structures are encircled in red, and the 7 candidates for which periodic modelling was performed are marked A–G, see the text for details.

Table 7.1: Xenon Buckingham potential parameters

Coefficient	Fitted value	
A_{Xe}	8216.849	eV
B_{Xe}	0.348606	Å^{-1}
C_{Xe}	339.32	eV·Å^{-6}

κ were obtained as:

$$A_{\mathrm{Xe},\kappa} = \sqrt{A_{\mathrm{Xe}} A_{\kappa}} \tag{7.1}$$

$$B_{\mathrm{Xe},\kappa} = \frac{A_{\mathrm{Xe}} + A_{\kappa}}{2} \tag{7.2}$$

$$C_{\mathrm{Xe},\kappa} = \sqrt{C_{\mathrm{Xe}} C_{\kappa}} \tag{7.3}$$

7.3 Xenon-129 shielding calculations

Chemical shielding tensors were calculated in cluster models with a hybrid functional and in periodic structures with the GIPAW method and GGA DFT-D [362, 359, 361]. An introduction to the theory behind such calculations was given in Chapter 2 and is not repeated here. The 3×3 shielding tensor σ, and the shielding of a lone xenon atom σ_{Xe} as reference, were used to calculate the chemical shift tensor $\delta = \sigma_{\mathrm{Xe}} - \sigma$. From this the isotropic shift δ_{iso}, the chemical shift anisotropy $\Delta\delta$ and the tensor asymmetry

η was calculated.

$$\delta_{\text{iso}} = \frac{\delta_{xx} + \delta_{yy} + \delta_{zz}}{3} \tag{7.4}$$

$$\Delta\delta = \delta_{zz} - \frac{\delta_{xx} + \delta_{yy}}{2} \tag{7.5}$$

$$\eta = \frac{\delta_{yy} - \delta_{xx}}{\delta_{zz} - \delta_{\text{iso}}} \tag{7.6}$$

The three ^{129}Xe NMR parameters of the 65 predicted clathrate structures were calculated and compared with the experimental data in several stages. First an initial screening was performed using molecular clusters extracted from the crystal structures. This allowed several candidates to be rejected as clearly incompatible with the experimental data. Specifically, $Z' = 2$ structures had substantial tensor asymmetries that were at odds with the experimental results.

Secondly, periodic boundary condition GGA DFT-D calculations were performed at the PBE-TS level of theory to further single out a few likely candidates. PBE-TS is particularly suitable for this, since it yields highly accurate crystal structures and the chemical shifts are highly sensitive to the local geometry [249, 115]. The best candidates were subjected to high level of theory hybrid functional calculations and with corrections for xenon dynamics and electron relativity. The different stages are described below.

The NMR calculations were performed by Marcin Selent and Pertuu Lantto at Oulu University, Finland. The results are reproduced here to demonstrate the predictive ability of the crystal structure prediction.

7.3.1 Initial screening of structures

Xenon NMR parameters of 65 plausible clathrate structures were first calculated using molecular clusters consisting of a single xenon atoms and its nearest-neighbour molecules, see examples in Figure 7.4. Non-relativistic shielding tensor calculations were performed using TURBOMOLE [526] with the BHandHLYP [246, 244] hybrid density functional and the MHA basis set [$22s17p14d2f/15s13p11d2f$] for the xenon atom and def2-SVP basis sets for all other atoms. The BHandHLYP functional contains 50% Hartree-Fock exchange. It is known to perform well for shielding tensor calculations [510]. However, it has been shown to systematically underestimate xenon chemical shifts and chemical shift anisotropies [527, 528, 529].

Generalized gradient functionals such as PBE and BLYP [241, 243, 246] on the other hand, are known to overestimate chemical shifts in different environments [510, 511, 530]. This overestimation in GGA functionals is counteracted in hybrid functionals in a way that depends on the fraction of Hartree-Fock exchange. The reader should remember that B3LYP contains 20% Hartree-Fock exchange and BHandHLYP 50%.

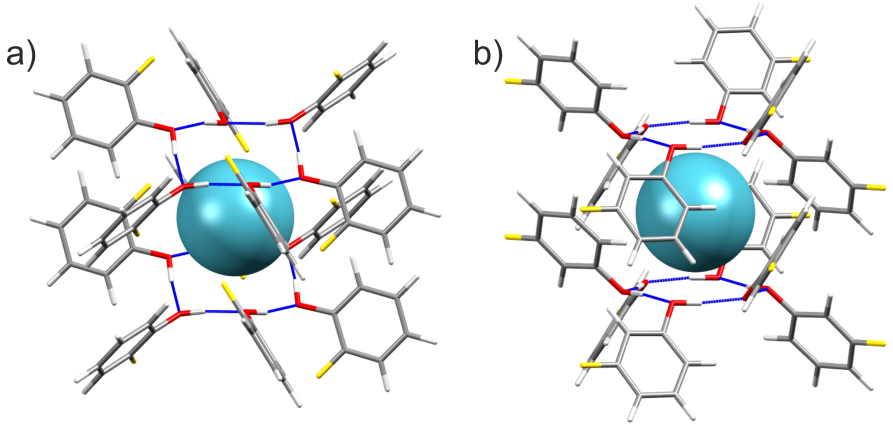


Figure 7.4: Typical cluster models of clathrates. A single xenon atom and its nearest neighbour fluorophenol molecules were used for shielding tensor calculations. Cluster a) corresponds to structure *o*F_D and b) is from *m*F_A. Note the characteristic hydrogen bond graphs surrounding the xenon atoms (blue lines).

In Table 7.2 the calculated ^{129}Xe NMR chemical shifts δ_{iso}, shift anisotropies $\Delta\delta$ and tensor asymmetry factors η are shown for the 33 predicted *o*-fluorophenol clathrate structures. These results were compared to the experimental results, which were: $\delta_{\text{iso}} = 256.0$ ppm, $\Delta\delta = -47.5$ ppm, and $\eta = 0.0$.

The corresponding screening results for the 32 *m*-fluorophenol clathrate structure candidates are shown in Table 7.3. The experimental results for the *m*-fluorophenol clathrate were $\delta_{\text{iso}} = 228.5$ ppm, $\Delta\delta = 183.3$ ppm, and $\eta = 0.0$.

Of the 65 predicted clathrate structures 5 *o*-fluorophenol and 7 *m*-fluorophenol crystal structures with predicted NMR parameters close to the experimentally observed values were selected for additional calculations. Other structures, primarily $Z' = 2$ structures, could be rejected as they had substantial tensor asymmetries $\eta > 0.3$, or chemical shift anisotropies which are incompatible with the experimental observations. The isotropic chemical shift was less discriminatory.

7.3.2 Periodic DFT-D modelling of likely candidates

The 12 remaining candidates, all of space group $R\bar{3}$, were labelled *o*F_A–E and *m*F_A–G according to Table 7.4. These candidates are also specially marked in Figures 7.2 and 7.3.

The chosen structures were geometry-optimised with periodic DFT-D in a plane wave basis, as implemented in the program CASTEP [364, 531]. The GGA-functional PBE was used with the Tkatchenko-Scheffler (TS) dispersion correction [248]. PBE-TS is known to yield accurate crystal geometries, which is important here, since the NMR parameters are so sensitive to the structure.

Table 7.2: Xenon chemical shifts δ_{iso}, chemical shift anisotropies $\Delta\delta$ and tensor asymmetries η calculated with cluster models for 33 predicted *o*-fluorophenol clathrate structures, in units of ppm.

Name	δ_{iso}	$\Delta\delta$	η
oF0_R-3_46_Xe	108.7	−12.9	0.0
oF0_R-3_24_Xe	178.2	−43.6	0.0
oF0_R-3_20_Xe	147.0	15.2	0.0
oF180_R-3_51_Xe	181.4	−10.0	0.0
oF180_R-3_44_Xe	241.4	27.0	0.0
oF0_R-3_17_Xe	116.2	−20.6	0.0
oF0_R-3_38_Xe	180.4	167.2	0.0
oF0_R-3_6_Xe	240.4	67.4	0.0
oF0_R-3_7_Xe	246.1	200.0	0.0
oF180_R-3_11_Xe	291.4	39.7	0.0
oF180_R-3_19_Xe	304.5	25.3	0.0
oF180_R-3_1_Xe	286.2	196.8	0.0
oF180_R-3_2_Xe	131.7	31.8	0.0
oF180_R-3_3_Xe	286.8	31.2	0.0
oF180_R-3_9_Xe	226.6	188.1	0.0
oF0_R-3_10_Xe	137.6	177.3	0.0
oF0_11591_Xe	133.3	47.6	0.0
oF0_R-3_14_Xe	125.9	124.6	0.0
oF180_R-3_15_Xe	161.6	143.9	0.0
oF0_1258_Xe	430.3	−148.8	0.1
oF0_R-3_39_Xe	72.8	−0.4	0.4
oF0_13826_Xe	265.1	69.2	0.6
oF0_5586_Xe	264.4	69.4	0.6
oF00_P-1_14655_Xe	223.8	−140.9	0.4
oF00_P-1_14867_Xe	83.6	−37.4	0.4
oF00_P-1_1752_Xe	77.8	−88.4	0.4
oF00_P-1_1773_Xe	126.8	−36.2	0.6
oF00_P-1_1990_Xe	149.8	−144.8	0.7
oF00_P-1_25199_Xe	62.0	35.8	0.5
oF00_P-1_2991_Xe	87.7	−28.4	0.8
oF00_P-1_7665_Xe	98.6	57.7	0.9
oF00_P21-C_3118_Xe	205.0	−122.0	0.7
oF00_P21-C_5438_Xe	195.9	134.5	0.1
Experimental	256.0	−47.5	0.0

Table 7.3: Calculated xenon chemical shift parameters for 32 predicted *m*-fluorophenol clathrate structures. Calculations were performed on cluster models using the BHandHLYP functional. In units of ppm.

Name	δ_{iso}	$\Delta\delta$	η
mF180_R-3_8_Xe	159.7	157.9	0.0
mF180_R-3_1_Xe	303.3	170.8	0.0
mF0_R-3_11_Xe	134.4	113.1	0.0
mF0_R-3_1_Xe	240.0	125.8	0.0
mF0_R-3_7_Xe	118.7	117.5	0.0
mF180_R-3_5_Xe	313.2	148.1	0.0
mF0_R-3_5_Xe	332.6	168.2	0.0
mF0_R-3_6_Xe	202.1	73.0	0.0
mF0_R-3_9_Xe	146.0	87.8	0.0
mF180_119952_Xe	109.2	49.2	0.0
mF180_R-3_11_Xe	158.5	42.9	0.0
mF180_R-3_22_Xe	127.0	52.6	0.0
mF0_R-3_10_Xe	123.7	35.7	0.0
mF180_120304_Xe	233.4	95.2	0.2
mF0_18479_Xe	158.6	81.3	0.2
mF0_R-3_4_Xe	57.6	−0.3	0.2
mF180_119925_Xe	132.7	86.1	0.2
mF180_120263_Xe	130.1	−76.9	0.1
mF180_R-3_18_Xe	157.1	152.4	0.1
mF00_P-1_14373_Xe	161.6	−89.4	0.7
mF00_P-1_256_Xe	104.6	−63.6	0.3
mF00_P21_4252_Xe	210.4	−100.0	0.6
mF0180_P-1_4959_Xe	136.5	−98.2	0.3
mF180180_P-1_11297_Xe	137.3	71.9	0.7
mF180180_P-1_12542_Xe	287.7	175.3	0.4
mF180180_P-1_20220_Xe	175.9	126.2	0.6
mF180180_P-1_22201_Xe	255.3	122.9	0.4
mF180180_P21_3636_Xe	200.1	72.7	0.8
mF180180_P21-C_1194_Xe	181.1	75.3	0.7
mF180180_P21-C_8331_Xe	164.0	86.7	0.5
mF180180_P21-C_9677_Xe	141.4	87.6	0.8
mF180180_P21-C_9808_Xe	214.5	77.4	0.9
Experimental	228.5	183.3	0.0

Table 7.4: Abbreviated names of likely clathrate structures.

ID	Name
*o*F_A	oF0_R-3_46_Xe
*o*F_B	oF0_R-3_24_Xe
*o*F_C	oF0_R-3_20_Xe
*o*F_D	oF180_R-3_51_Xe
*o*F_E	oF180_R-3_44_Xe
*m*F_A	mF180_R-3_8_Xe
*m*F_B	mF180_R-3_1_Xe
*m*F_C	mF0_R-3_11_Xe
*m*F_D	mF0_R-3_1_Xe
*m*F_E	mF0_R-3_7_Xe
*m*F_F	mF180_R-3_5_Xe
*m*F_G	mF0_R-3_5_Xe

Following geometry-optimization, NMR shielding tensors were calculated with the PBE functional using the gauge-including projector augmented wave method [532, 533]. The periodic boundary calculations include the effects of long range interactions in the crystal structure and also the scalar relativistic effects on the xenon shielding tensor *via* the zeroth-order regular approximation [534, 535]. The results are shown in Tables 7.5 and 7.6.

Table 7.5: Xenon isotropic chemical shifts δ_{iso} and chemical shift anisotropies $\Delta\delta$ from periodic calculations of five likely *o*-fluorophenol clathrate structures at their PBE-TS optimal geometries. The asymmetry factor η is zero for all structures.

Structure	δ_{iso} (ppm)	$\Delta\delta$ (ppm)
*o*F_A	229.8	−38.9
*o*F_B	273.2	−86.7
*o*F_C	236.4	−15.4
*o*F_D	285.0	−46.5
*o*F_E	402.7	2.0
Experimental	256.0	−47.5

The reader is reminded that the periodic GGA-functional results are upper limits on ^{129}Xe chemical shifts and shift anisotropies, since GGA-functionals overestimate the magnitudes of both parameters [510, 527, 536, 529, 530].

7.3.3 Detailed NMR-modelling of the most probable structures

After periodic NMR calculations, the most probable clathrate structures were selected for extensive modelling of NMR shieldings using the Amsterdam Density Functional program (ADF) [537, 538]. In addition to the non-relativistic and scalar relativistic quantum chemistry, ADF provides the 2-component spin-orbit relativistic effect

Table 7.6: Calculated ^{129}Xe chemical shifts δ and chemical shift anisotropies $\Delta\delta$ from periodic GIPAW/PBE calculations for seven likely *m*-fluorophenol clathrate structures at their PBE-TS optimized geometries. The asymmetry parameter η is zero for all structures.

Structure	δ_{iso} (ppm)	$\Delta\delta$ (ppm)
*m*F_A	270.6	226.7
*m*F_B	295.2	216.0
*m*F_C	229.2	224.8
*m*F_D	271.4	163.4
*m*F_E	272.9	192.4
*m*F_F	314.1	203.7
*m*F_G	337.2	242.8
Experimental	228.5	183.3

[534, 539]. The details of these calculations are beyond the scope of this thesis, but are described in full in reference [4].

Calculations were performed on cluster models with hybrid functionals that contain different amounts of Hartree-Fock exchange. Relativistic effects and the effect of the thermal motion of the xenon atom were accounted for. In addition, the effect of the crystal lattice was included by scaling the cluster results by a factor obtained from the difference between cluster and periodic calculations at the PBE level of theory.

Finally, the effect due to thermal motion of the xenon atoms in the cages was accounted for. Shielding tensors were calculated for several geometries with different xenon atom positions, and the results averaged over a Boltzmann partitioning at 300 K.

Comparisons of calculated NMR spectral peak shapes to experimentally observed data are shown in Figures 7.5 and 7.6. The relationship between the powder spectral peak shape and the NMR tensor elements are described elsewhere [540]. For the single best candidate of each compound, results at several different levels of theory are shown. The highest level of theory, peak 4, in Figures 7.5 and 7.6 include the effects of scalar and spin-orbit relativistic effects, the effect of the periodic lattice, and the effect due to the thermal motion of the xenon atom at 300 K. The results at this level of theory are tabulated in Table 7.7.

The results of the highest level of theory calculations agree well with the experimental results. The best agreements for each compound are for structures *o*F_D and *m*F_A. For *o*-fluorophenol, the alternative candidates from the PBE-TS level of theory have quite different periodic BHandHLYP ^{129}Xe chemical shifts or chemical shift anisotropies as compared to the experimental ones, leaving only one possible matching structure.

For *m*-fluorophenol the closest alternative candidates overestimate the chemical shift (*m*F_F) or the shift anisotropy (*m*F_B) at the highest level of theory, when relativistic and thermal effects are also included, see Table 7.7.

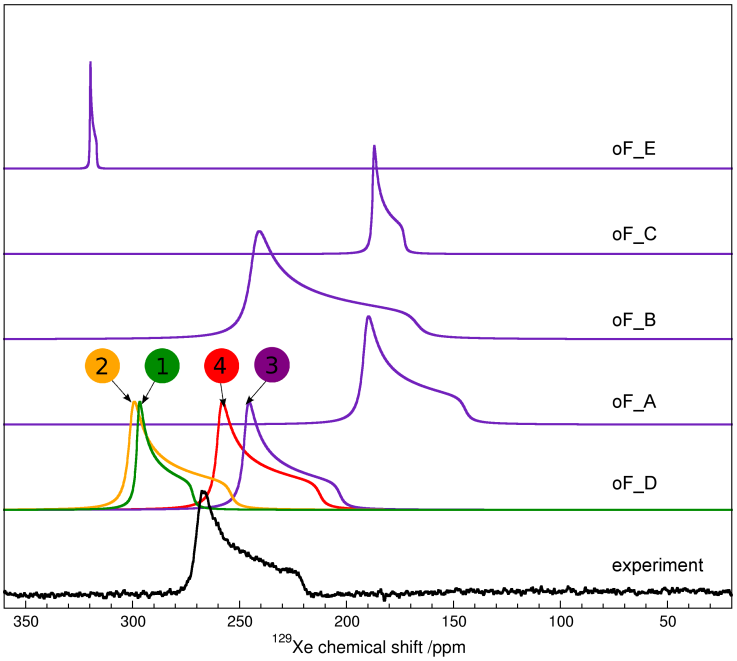


Figure 7.5: Comparison of calculated ^{129}Xe NMR spectral peak shapes to experimental NMR spectroscopy data of *o*-fluorophenol xenon clathrate. The peaks 1–4 were calculated at different levels of theory; (1) PBE at force field geometry, (2) PBE at PBE-TS geometry, (3) BHandHLYP at PBE-TS geometry, (4) BHandHLYP at PBE-TS including effects due to xenon dynamics at $T = 300$ K.

Table 7.7: Calculated NMR parameters for the best matches at the highest level of theory.

Structure	δ (ppm)	$\Delta\delta$ (ppm)	η
*o*F_B	237.8	−85.4	0
*o*F_D	250.5	−42.6	0
Exp. *o*F	256.0	−47.5	0
*m*F_A	235.9	188.9	0
*m*F_B	224.7	195.6	0
*m*F_F	247.0	186.7	0
Exp. *m*F	228.5	183.3	0

Thus, we conclude that the crystal structures for *o*- and *m*-fluorophenol clathrates are *o*F_D and *m*F_A. The most accurate level of modelling almost quantitatively reproduces the experimental NMR parameters. Since the NMR tensor is so sensitive to the crystal structure geometry, it is highly unlikely that another crystal structure could have the same NMR parameters. This is also the reason for why such a high level of theory is necessary in order to reproduce the experimental results computationally.

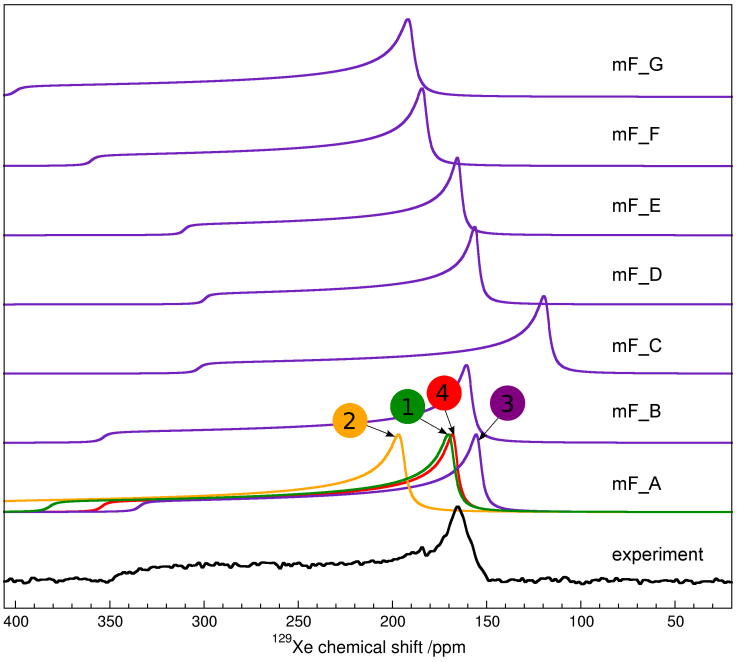


Figure 7.6: Comparison of calculated ^{129}Xe NMR spectral peak shapes to experimental NMR spectroscopy data of *m*-fluorophenol xenon clathrate. The peaks 1–4 were calculated at different levels of theory; (1) PBE at force field geometry, (2) PBE at PBE-TS geometry, (3) BHandHLYP at PBE-TS geometry, (4) BHandHLYP at PBE-TS including effects due to xenon dynamics at $T = 300$ K.

7.3.4 Description of the fluorophenol xenon clathrate structures

The structures are depicted in Figures 7.7, 7.8 and 7.4. The two clathrate structures determined here both belong to space group $R\bar{3}$. This seems to be a general characteristic for several phenolic clathrates. Hydroquinone, *p*-cresol, *p*-fluorophenol, phenol, and even benzene all form essentially iso-structural clathrates [82, 73, 74, 541].

The space group symmetry has a very prominent 3-fold screw axis along the **c** lattice vector. Space groups with 3-, 4- and 6-fold screw axes were particularly prone to yield clathrate host structures in the CSP described above. The screw axis facilitates strong hydrogen bonding motifs of phenol-derivatives, with $R\bar{3}$ in particular forming very strong $R_6^6(12)$-motifs [542]. Hence, for the prediction of clathrates and inclusion compounds, space groups with such screw axes should be prioritized.

Both structures have three adsorption sites per unit cell, with volumes of 61.4 and 88.5 Å^3 each for *o*- and *m*-fluorophenol, respectively. These host volumes yields R_g-values of 68.7% and 47.7% for xenon. These are quite compatible with the previously observed 55%-rule [338, 339].

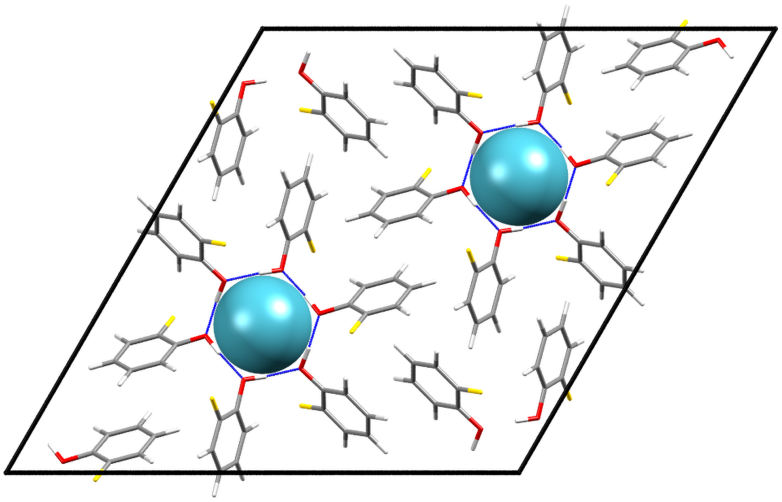


Figure 7.7: The crystal packing diagram of the predicted *o*-fluorophenol xenon clathrate *o*F_D. Note the distinctive $\overline{3}$ point group symmetry and the R_6^6-hydrogen bond motifs.

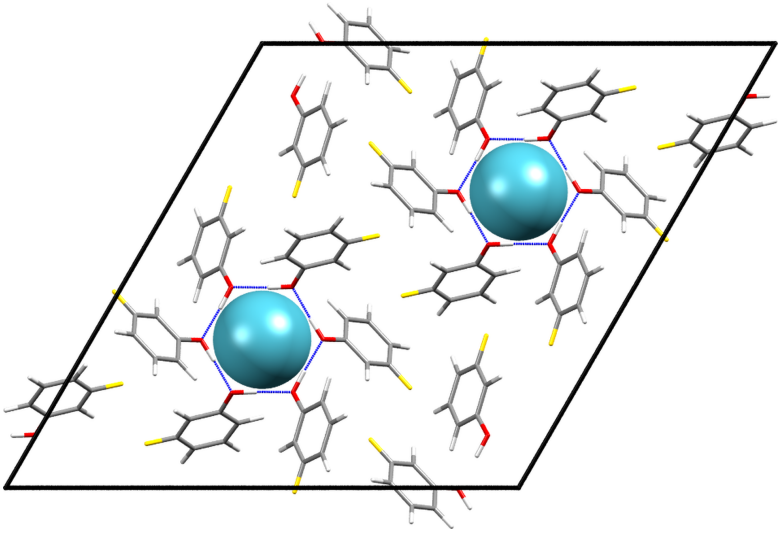


Figure 7.8: The crystal packing diagram of the predicted *m*-fluorophenol xenon clathrate *m*F_A. Note the distinctive $\overline{3}$ point group symmetry and the R_6^6-hydrogen bond motifs.

The sign of the chemical shift anisotropy relates to the shape of the cavities. An oblate cavity leads to a negative $\Delta\delta$, while prolate cavities gives $\Delta\delta$ larger than zero.

7.4 Concluding remarks

A combined application of crystal structure prediction and ^{129}Xe NMR has been used to determine the crystal structures of two *o*- and *m*-fluorophenol xenon clathrates. The exceptional sensitivity of the ^{129}Xe chemical shift tensor to its local environment allows

direct comparisons between observed and calculated NMR chemical shift parameters and was used to directly reject many hypothetical clathrate structures.

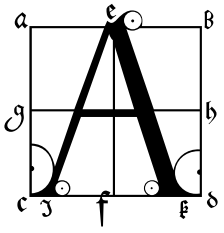
The crystal structure prediction proved difficult because of the very large number of predicted structures. The search for porous high-energy structures, and the consideration of $Z' = 2$ candidates made the predictions demanding. When a CSP calculation is not targeted at the most stable polymorph, great care must be exercised in the sampling of packing possibilities. An unexpectedly large number of quite realistic clathrate candidates were obtained. The 12 plausible candidates that could not be ruled out from screening calculations were all $Z' = 1$, $R\overline{3}$ structures.

Alternative experimental techniques, such as magic-angle spinning NMR or X-ray powder diffraction would not have been enough to single out the correct crystal structure. X-ray powder diffraction was attempted, but could not yield powder patterns of sufficient quality for structure determination, or even for comparisons to simulated patterns of CSP-candidates (data not shown here, see reference [4]). Magic-angle spinning NMR would yield only the isotropic chemical shift, which would not have been enough to discriminate between the predicted candidates. Static solid-state NMR was in this case necessary.

The method presented here is potentially a powerful approach for structure determination of porous materials when powder diffraction patterns cannot be obtained, or are insufficient for structure determination. The success of the method is related to the high sensitivity of the xenon NMR chemical shift and chemical shift anisotropy to minute details of the cavity geometry, so that even structurally similar clathrates can have vastly different chemical shift tensors. This large sensitivity also necessitates sophisticated NMR calculations in order to obtain quantitatively accurate NMR parameters that can be used for comparison to experiments.

Chapter 8

Prediction of inclusion behaviour



dsorption free energies of gaseous guests in porous crystals conclude my investigation of the prediction of structures and properties of polymorphs and clathrates. In this Chapter, I apply the Epanechnikov kernel density method to lattice dynamical calculations on hypothetical clathrates generated by crystal structure prediction. The influence of temperature and the number of degrees of freedom of the guest molecule on adsorption behaviour is studied. The purpose is to show some examples of what can be done with the various computational methods explored in this thesis.

8.1 Background

In the previous Chapters, I have described and applied crystal structure prediction and lattice vibrational calculations for the study of the stability and properties of experimentally known, and hypothetical, crystal structures. In this Chapter I will apply lattice dynamics to a number of various clathrate structures that have been generated by CSP methods. This Chapter also demonstrates the use of the Epanechnikov kernel density estimation and a method for automatically choosing the kernel bandwidth, as described in Chapter 2.

The statistical mechanics of clathrates was elucidated by van der Waals and Platteeuw [543]. Their articles are still cited as *the* theory of clathrate formation. Belosludov and Schneider have since made additional contributions to the theoretical understanding of clathrate thermodynamics [544, 545].

Hydroquinone (quinol, 1,4-dihydroxybenzene) is perhaps the best known clathrate-forming molecule, and have been extensively studied both experimentally and computationally [76, 546, 547, 492, 79]. In addition, phenol and several *para*-substituted phenol

compounds are known to readily form clathrates with a multitude of gases at low temperatures, and because of their selective affinities for different guests, they can be used for gas separation [74]. Hydroquinone is particularly promiscuous, forming clathrates with a large number of guests. Separation of ethane/methane and CO_2/N_2 by hydroquinone has been demonstrated [493, 77]. Phenol with substitutions at the *ortho*- and *meta*-positions are less prone to forming clathrates, but as I described in Chapter 7, *o*- and *m*-fluorophenol forms clathrates with xenon [4, 496, 497].

The hydroquinone dihydrogen clathrate has an interesting history. The constituent compounds and the procedure for synthesising the clathrate have been, at least in principle, experimentally accessible to any chemist for some 150 years. Despite this, it was found experimentally only after a computational study predicted that such clathrates can form and that they may have multiple H_2 occupancy [79], obviously of great interest for gas storage applications. The computational predictions have recently been confirmed experimentally. The clathrate exhibits fast and reversible absorption of hydrogen at room temperature and has double and triple occupancy of H_2 at elevated pressures [80, 81]. This demonstrates how *a priori* computational predictions can greatly aid the discovery process.

Several researchers have noted that the size of the guest is important. The binding affinity of a guest in a clathrate is related to the guest volume, or the guest-to-pore volume ratio [337, 339, 338, 548]. Others have suggested that the adsorption free energy depends on the guest's molar refractivity A [549, 492], *i.e.* the polarizability α of the guest molecule. It should be noted that that the molar refractivity and the molar volume are strongly correlated, since

$$A = \frac{4\pi}{3} N_\mathrm{A} \alpha \simeq \frac{M}{\rho} \frac{n^2 - 1}{3}, \tag{8.1}$$

where n is the refractive index, ρ the density and M the molar mass of the guest [550]. Hence, for structures with only non-directional interactions between the host and a gaseous guest, we would expect the guest's and cavity's volumes to be directly related to the binding affinity. This suggests that the simple method of docking guest molecules into porous host structures obtained by CSP is a sound method for predicting such clathrates.

Computational studies of clathrates have mainly focused on the calculation of guest adsorption energies by force field methods. The adsorption energy is often calculated as the static (0 K) energy [494], or sometimes as adsorption free energies through molecular dynamics (MD) [551] or Monte Carlo simulations (MC) [547]. Calculations with *ab initio* quantum-chemical methods have also been performed [546], but such studies are still quite rare.

Lattice dynamics is an alternative to the above methods, and as we have seen in the previous Chapters, well-converged estimates for the free energy can be obtained with highly accurate force fields, and at a moderate computational cost [2]. Lattice dynamics studies of clathrates are rare, but Dang [552] and Belosludov [76] performed lattice dynamics on hydroquinone clathrates with several guest molecules. Computationally obtained vibrational frequencies were compared to experimental data. There is also a lattice dynamics study of hexakis(phenylthio)benzene that also includes the calculation of elastic properties [260].

Lattice dynamics does however rest on an assumption that there is only one adsorption site per cavity and that the guest is not disordered. This has been shown not to be true for some clathrates [553, 554, 555, 545]. For instance, acetonitrile is disordered over three 120° rotationally related energy-degenerate positions in the β-hydroquinone clathrate [556]. Furthermore, both the harmonic and quasi-harmonic approximation assume that the lattice vibrations are harmonic, which is generally an excellent approximation for close-packed molecular crystals, see Chapter 3 and reference [279], but probably less accurate for clathrates [557]. Guest molecules have been described to have a rattling motion where the molecule move between different adsorption sites, rather than a harmonic vibration.

Solvates, hydrates and clathrates have occasionally been the subject of crystal structure prediction (CSP) studies [75, 341, 342, 340, 46]. Predictions of multi-component crystals are rarely attempted because of the difficulty in sampling the vast space of all possible relative orientations and stoichiometric ratios between the different components, and predictions of multicomponent crystal structures are often not successful [5].

The goal of CSP was initially about finding the most stable form and possibly some additional polymorphs. These are generally located less than about 6 kJ/mol above the global energy minimum. In Chapter 7, I showed that CSP can be used to also predict highly unstable crystal forms. The porous clathrand structure of *o*-fluorophenol was found more than 12 kJ/mol above the most stable crystal form. My interest in these high-energy crystal forms is part of a general trend in CSP research. There has recently been a change in the focus and purpose of CSP studies [5]. Now, a much larger part of the predicted crystal energy landscape is considered and structures with specific properties may be targeted.

In this chapter I will present results of lattice dynamics calculations on a relatively large number of clathrate structures. A study of the general behaviour of clathrates with respect to temperature and the number of degrees of freedom of the guest molecule is presented. As host structures, I use *o*- and *m*-fluorophenol structures that were generated in the previous Chapter. Xenon, carbon dioxide and methane are used as guests. The three guests have 3, 5 and 6 vibrational degrees of freedom respectively, allowing me to study how this affects the free energy of adsorption. In particular, the

effect of temperature and cavity-to-guest volume ratio on the thermodynamic driving force for guest enclathration is studied.

I then present results for additional computationally predicted clathrate structures. Predicted structures of resorcinol (1,3-dihydroxybenzene) [558] and triptycenetrisbenzimidazolone [559, 560] (TTBI, Figure 8.1), obtained from CSP calculations are used. Resorcinol is an isomer of hydroquinone and was hypothesised to also be able to form clathrates. The experimentally known form of TTBI is a channel-type inclusion compound that adsorbs solvent molecules [559]. It was hypothesised that it might be able to form additional inclusion compounds.

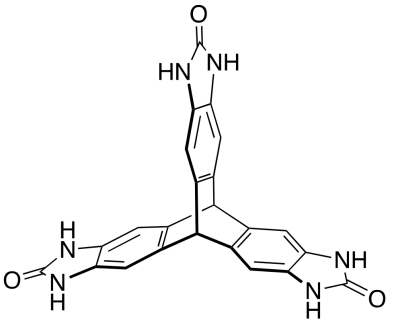


Figure 8.1: Triptycenetrisbenzimidazolone (TTBI).

The main purpose of this Chapter is to demonstrate the combined use of crystal structure prediction and lattice dynamics methods with enhanced Brillouin zone sampling for the prediction of clathrate structures and their properties.

8.2 Methods

8.2.1 Studied systems

The prediction of hypothetical crystal structures of *o*- and *m*-fluorophenol was described in the previous Chapter. From their crystal energy landscapes, plausible clathrate host structures were selected. The selection was based on the pore size and pore dimensionality as calculated with PLATON [284]. The porosity calculations were performed by placing a probe at every point in a 0.15 Å lattice throughout the unit cell and testing whether the probe is within the van der Waals radius plus a 1.2 Å probe radius of any atom [561, 336]. Only structures with discrete cavities and where all the cavities in the unit cell are of the same size and suitable for the adsorption of a single guest molecule per cavity were chosen. Channel-type inclusion compounds were excluded.

Using sets of predicted structures facilitates studies of the general behaviour of clathrates with respect to cavity size and the type of guest, while effects due to the chemistry should be minimal. That is, using a large number of polymorphic structures should allow the determination of adsorption behaviour with respect to only the geometry and size of the cavities. This is not possible if using only experimentally determined structures.

Calculations were also performed on a set of 50 hypothetical host structures of TTBI, obtained from reference [560]. Six hypothetical structures of resorcinol were obtained

from reference [558]. The structures of TTBI and resorcinol were generated with CSP, using methods quite similar to those described in Chapter 4.

Xenon atoms, methane and carbon dioxide were used as guests for all structures. The CO_2 and CH_4 geometries were optimised with B3LYP-D3 / 6-311G** and the three guest species were inserted into the host structures using a simple docking mechanism, as described in Chapter 2. The centre of mass of the guest molecule is placed at the centroid coordinates of the cavity, as determined with PLATON. For guests other than xenon, the molecule was then randomly rotated until there were no clashes between atoms, using 85% of the combined van der Waals radii as criterion for a clash. Guest insertion is done with the crystal structures in space group $P1$, so that no point group symmetry constraints are imposed on the resulting clathrate structure. The docking was repeated until there were no more empty cavities in the unit cell large enough to contain another guest molecule.

The guest insertion method in this study means that I only consider the case of fully occupied, and stoichiometric, clathrates. Partially occupied clathrates can be modelled computationally either with Molecular Dynamics or Monte Carlo simulations, or by scaling the host-guest force field interaction terms. Such methods potentially allow a global optimisation of the Gibbs free energy also with respect to the guest occupancy. Such simulations are however challenging to perform with an accurate multipole-based force field, and beyond the scope of this thesis.

The carbon dioxide molecules were supplemented with four non-interacting and massless dummy atoms located equidistantly around the carbon atom, since the program DMACRYS cannot handle linear molecules. A new atom type 'V_01' was implemented to represent dummy atoms, which were also used to automatically construct an intramolecular coordinate system for the distributed multipoles. This allows energy-minimisation and lattice dynamics with multipoles also for structures containing linear molecules, although special care is needed to deal with the additional fictitious phonons resulting from a linear molecule rotating around its axis.

8.2.2 Lattice dynamics

Geometry-optimisations and rigid-body harmonic approximation lattice dynamics calculations were performed with DMACRYS version 2.0.4 [168, 486]. The W99rev6311P5 force field was used, see Chapter 3. Distributed multipoles were derived from the B3LYP-D3 / 6-311G** charge density, calculated in a polarizable continuum model with relative permittivity $\varepsilon_r = 5.0$. For xenon atoms, the Buckingham potential described in the previous Chapter was used.

The static intermolecular energy of the crystal structures calculated with the force field is denoted E_{latt}. In this Chapter, intramolecular energies were not considered. Calculations

were performed in the rigid-body approximation, meaning that the molecules' geometries are assumed not to change significantly when guest molecules are inserted into the porous host structures, the clathrands.

In order to sample the Brillouin zone, three supercells were used, each expanded in **a**, **b** or **c** only. Thus, **k**-points were sampled equidistantly along three axes in the Brillouin zone. This is computationally more costly than the co-prime split supercell method used in Chapters 3–5, but it results in a more uniform sampling of the Brillouin zone. The supercell expansion coefficient in each direction was chosen such that the distance between adjacent **k**-points was strictly smaller than a target distance. This distance was 0.15 Å^{-1} for resorcinol and fluorophenol structures, and 0.1 Å^{-1} for TTBI structures. The smaller **k**-point distance ensured an adequate sampling also for TTBI's relatively large unit cells.

The main difficulty in lattice dynamics lies in the slow convergence of the vibrational energy and entropy with respect to Brillouin zone sampling [2]. To further enhance the convergence, Debye's method was used to model phonon dispersion of the three acoustic modes near the Brillouin zone centre Γ, as described in Chapter 2. The linear dispersion relation was determined from the elastic tensor and the Christoffel equation, and used to calculate a mean phonon frequency ω_D on a Debye sphere around Γ. The sphere extends out to the nearest explicitly sampled **k**-points.

Dispersion of optic phonons around other **k**-points was modelled with a kernel density estimate [292, 290]. Gaussian kernel density estimates were used in the previous Chapters and have also been used by other researchers [299, 288, 562]. However, invoking Debye's assumption of a linear phonon dispersion within a small sphere around each **k**-point [265], it can be shown that the Epanechnikov kernel is a better choice [295]. A derivation of this result is given in Chapter 2. Hence, in this work, the phonon dispersion was modelled by substituting each discrete phonon frequency with an Epanechnikov kernel function.

$$K(x) = \frac{3}{4}\left(1 - x^2\right)\mathrm{I}(x) \tag{8.2}$$

where $\mathrm{I}(x)$ denotes the indicator function

$$\mathrm{I}(x) = \begin{cases} 0, & \text{if } |x| \geq 1 \\ 1, & \text{if } |x| < 1 \end{cases} \tag{8.3}$$

The phonon density of states $g(\omega)$ can then be estimated from the phonon frequencies ω_i as:

$$\frac{g(\omega)}{6Z + N_\mathrm{d}N_\mathrm{g}} = \frac{3}{4nh}\sum_i^n \left(1 - \left(\frac{\omega - \omega_i}{h}\right)^2\right)\mathrm{I}\left(\frac{\omega - \omega_i}{h}\right) \tag{8.4}$$

Where Z is the number of host molecules in the unit cell, n the number of phonons and h is the kernel bandwidth. N_g and N_d are the number of guest molecules and the number of degrees of freedom of the guest, respectively.

Choosing a suitable kernel bandwidth h can be difficult. Much work has been invested on the problem in the statistics community [296, 293, 301, 297, 298]. However, to my knowledge, no consideration has previously been given to the specific problem at hand; the (automatic) choice of a kernel bandwidth for the estimation of phonon dispersion. I explored the possibility of optimizing the bandwidth through least-squares cross-validation [302, 292, 303] but this proved impractical, see Chapter 2.

The bandwidth was instead determined from the average phonon gradient of optic modes over the Brillouin zone in each crystal structure. Phonon frequencies were calculated at Γ and at the Brillouin zone boundary points X, Y, Z in the three principal directions **a**, **b** and **c**. This is easily done, as it requires only three small supercells, each with expansion coefficient 2. The mean dispersion in each of the three directions was calculated as a mean absolute difference in phonon frequencies between the Brillouin zone centre Γ and the boundary point. For each non-zero phonon mode i at Γ, the phonon mode j at the zone boundary in direction $\lambda \in \{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$ that have eigenvectors $\mathbf{e}_{j,\lambda}$ that are maximally collinear with $\mathbf{e}_i$ is identified, using

$$\frac{1}{Z}\sum_{i=1}^{Z}(\mathbf{e}_i \cdot \mathbf{e}_{j,\lambda})^2 \in [0,1] \tag{8.5}$$

as a score of collinearity between phonon modes i and j.

The absolute differences in phonon frequencies $\Delta\omega_{i,\lambda} = |\omega_i - \omega_{j,\lambda}|$ are averaged to give one $\overline{\Delta\omega}$ in each direction. This is a simplified version of the linear approximation method by Gilat [562, 288]. The major improvement is the use of a rotatable kernel, which greatly simplifies the method.

The dispersion in each Debye sphere was modelled as a single kernel bandwidth h and taken to be

$$h = \frac{1}{2\sqrt{3}}\sqrt{\sum_{\lambda}\left(\frac{\overline{\Delta\omega}(\lambda)}{N_\mathrm{k}(\lambda)}\right)^2} \tag{8.6}$$

Where $N_\mathrm{k}(\lambda)$ denotes the supercell expansion coefficient in direction λ.

The Epanechnikov kernel was originally derived as the solution to a completely different problem [295]. Given a random and independent sample drawn from an arbitrary probability density function, it is possible to estimate the probability density function by a kernel density estimate [290]. Using the integrated square error between the estimate and the true probability density as cost function, which kernel makes the estimate converge most rapidly with respect to the size of the sample? Under mild assumptions, the answer is the Epanechnikov kernel. Hence, if the calculated phonon frequencies can

be considered a random sample of the true frequency distribution, or if **k**-points are sampled randomly in the Brillouin zone [300], then the Epanechnikov kernel density is an optimal estimate of the phonon density of states.

The total number of non-zero phonons ω_i is $n = 6ZN_\mathrm{k} + N_\mathrm{d}N_\mathrm{g}N_\mathrm{k} - 3$, where Z is the number of host molecules per unit cell, N_k is the number of sampled **k**-points, N_d, the number of degrees of freedom for the guest molecule, and N_g the number of guest molecules per unit cell.

The vibrational energy per unit cell was calculated as:

$$\begin{aligned} F_\mathrm{vib}(T) = & \frac{1}{2}\left(\frac{n}{n+3}\right)\int_0^\infty \hbar\omega g(\omega)\, d\omega \\ & + k_B T\left(\frac{n}{n+3}\right)\int_0^\infty g(\omega)\ln\left(1-\exp\left(\frac{-\hbar\omega}{k_B T}\right)\right)\, d\omega \\ & + \frac{9\hbar\omega_D}{8N_\mathrm{k}} \\ & + \frac{3k_B T}{N_\mathrm{k}}\ln\left(1-\exp\left(\frac{-\hbar\omega_D}{k_B T}\right)\right) \\ & - \frac{k_B T}{N_\mathrm{k}}\mathrm{D}\left(\frac{\hbar\omega_D}{k_B T}\right) \end{aligned} \tag{8.7}$$

where $\mathrm{D}(x)$ is the Debye function

$$\mathrm{D}(x) = \frac{3}{x^3}\int_0^x \frac{t^3}{\exp(t)-1}\, dt. \tag{8.8}$$

The guest adsorption energies were calculated as the difference in lattice- and vibrational energy between the guest-filled clathrate and the empty host framework.

$$\Delta E_\mathrm{latt}^\mathrm{bind} = \frac{1}{N_\mathrm{g}}\left(E_\mathrm{latt}^\mathrm{clathrate} - E_\mathrm{latt}^\mathrm{host}\right) \quad [\mathrm{kJ/mol\ guest}] \tag{8.9}$$

The vibrational contribution to guest adsorption is

$$\Delta F_\mathrm{vib}^\mathrm{bind} = \frac{1}{N_\mathrm{g}}\left(F_\mathrm{vib}^\mathrm{clathrate} - F_\mathrm{vib}^\mathrm{host}\right) \quad [\mathrm{kJ/mol\ guest}] \tag{8.10}$$

8.3 Results

8.3.1 Changes in phonon spectra upon guest enclathration

I begin by showing how the lattice vibrational spectra of two representative structures are affected by the presence of different guest molecules. The purpose is to determine if the vibrations of the clathrate host and guest are essentially additive, or if the intermolecular interactions are such that the phonon spectrum of the clathrand changes

upon enclathration of a guest. In particular, does the enclathration of guests lead to a stiffening of the structures, with increased elastic moduli and higher vibrational frequencies?

The changes in the phonon DOS of a typical $R\bar{3}$ *o*-fluorophenol clathrate structure upon adsorption of carbon dioxide, methane and xenon are displayed in Figure 8.2. A packing diagram of this particular clathrate structure is shown in Figure 8.3.

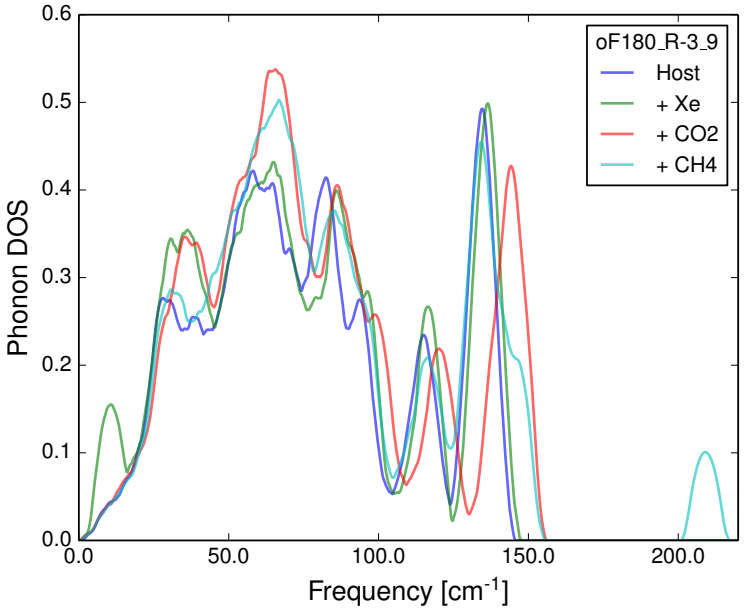


Figure 8.2: Phonon densities of state for a *o*-fluorophenol clathrand structure (oF180_R-3_9) with adsorbed Xe, CO_2 and CH_4 guests. The vibrational modes of the guests and the host are partially additive.

The first observation to make from Figure 8.2 is how well-converged and highly resolved the phonon densities of states are. The Epanechnikov kernel density estimate results in high-quality spectra.

Xenon has three vibrational degrees of freedom, but no rotational. The lowest-frequency mode, near a mere 10 cm^{-1}, most likely correspond to vibrations parallel to the crystallographic **c**-vector, where there is very little resistance to movements. This mode should contribute substantially to the entropy. The two modes perpendicular to the **c**-axis are nearly degenerate at 35–40 cm^{-1} because of the high 6-fold symmetry of the cavity. The xenon atoms in this clathrate, and all other $R\bar{3}$ clathrates, are located at crystallographic special positions and since the xenon atoms are spherically symmetric, the $R\bar{3}$ space group symmetry is preserved. In general however, guest enclathration often breaks the point group symmetry of the clathrand, see below.

In the analogous carbon dioxide clathrate, the guest molecules in the unit cell are not necessarily oriented along the **c**-axis, which breaks the symmetry. Disorder over multiple such adsorption sites will in reality contribute additional entropy which is not captured

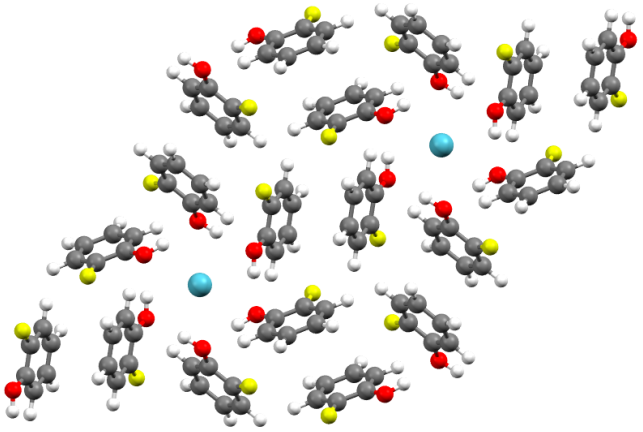


Figure 8.3: Packing diagram showing the crystal structure of the hypothetical *o*-fluorophenol xenon clathrate named *o*F180_R-3_9. The crystallographic **c**-axis is perpendicular to the paper plane.

by harmonic approximation lattice dynamics [555]. This effect has also been seen in hydroquinone clathrates with methanol and acetonitrile, with phase transitions occurring between ordered and disordered clathrates [553, 554, 556].

The methane molecules appear to have a high-frequency mode near 210 cm^{-1}, most likely due to rotations of the guest molecule. The vibrational modes of the guests are only partially additive, or perhaps it is better to say that some modes are additive while others are not. The presence of the guest molecules perturb some of the lattice modes of the host. This means that the same guest molecule can give different thermal contributions to the free energy of different clathrate structures, even when they are polymorphic.

8.3.2 Changes in bulk modulus upon guest enclathration

Table 8.1: Reuss-Voigt-Hill averaged bulk moduli of *m*-fluorophenol clathrates in units of GPa.

Structure	Host	+Xe	+CH_4
mF00_P-1_256	7.09	8.14	-
mF00_P21_4252	8.56	-	9.10
mF0_18479	8.59	-	9.02
mF0_R-3_11	8.33	9.24	9.17
mF180180_P-1_11297	8.12	8.08	-
mF180180_P-1_12542	8.64	-	8.63
mF180180_P-1_20220	7.88	-	8.27
mF180180_P-1_22201	8.47	8.49	8.53
mF180180_P21-C_8331	8.65	9.33	-
mF180180_P21_3636	8.84	9.03	9.15
mF180_120304	8.53	8.21	-
mF180_R-3_22	7.84	8.93	8.73

The presence of a guest generally leads to a stiffening of a clathrand, as seen in the increase in the Reuss-Voigt-Hill averaged bulk moduli of fluorophenol clathrates in Table 8.1. The increase in bulk modulus is however sometimes small and in some cases the clathrate is not harder than the empty host framework.

The changes in the phonon spectra of the host lattice and elastic moduli are relatively small, which means that the thermodynamic contributions of the guest depends mostly on the vibrational energy and static interaction of the guest itself in a mostly additive manner, rather than through anharmonic interactions between the guest and the framework, although such effects do exist.

8.3.3 Volume change upon guest enclathration

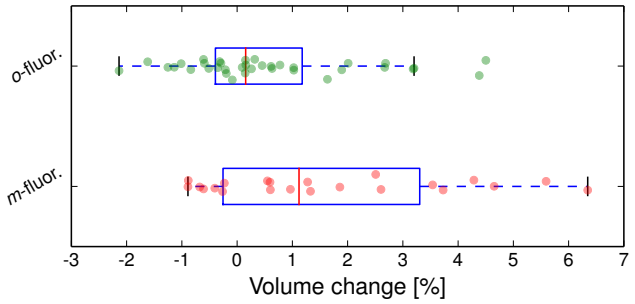


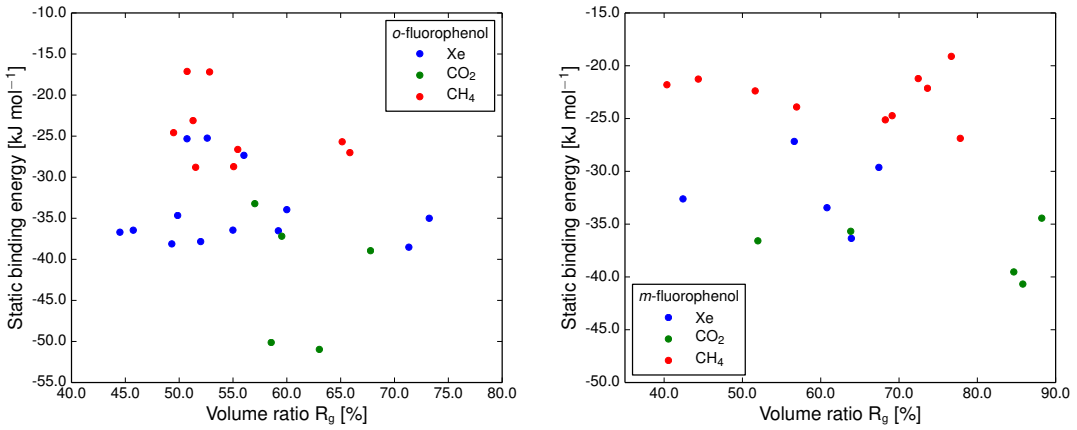
Figure 8.4: The changes in unit cell volume upon guest enclathration for *o*- and *m*-fluorophenol clathrates.

The percent difference in unit cell volumes between the empty host structures and their respective Xe-, CO_2- and CH_4-clathrates are shown in Figure 8.4. Data are shown for 22 different *m*-fluorophenol clathrates and 36 *o*-fluorophenol clathrates. The results obtained here agree well with the empirical observation that in general clathrate structures do not change much in volume upon guest enclathration [490].

8.3.4 Adsorption free energies

The temperature-dependence of adsorption free energies of the three different guests in the clathrate hosts are shown in Figure 8.6. The static temperature-independent adsorption energies are shown in Figure 8.5.

Perhaps surprisingly, there is no trend between the adsorption free energy and volume ratio. This seems to contradict the close-packing principle and the empirical rule that cavity-to-guest volume ratios near 55% are thermodynamically favoured [339, 338]. Naively, one would expect that the temperature-independent part of binding energy should increase (absolutely) as the cavity becomes tighter. Also, there is no evidence that a larger cavity results in lower-frequency vibrations and a larger entropy. This is



(a) (b)

Figure 8.5: The static lattice binding free energy of three guest molecules in several fluorophenol clathrates. There is no correlation between the binding energy and the host-to-guest volume ratio R_g.

consistent with the finding in Chapter 5 that the correlation between density-differences and entropy-differences between polymorphs is very small, see Figure 5.19.

Hence, we see again that the empirical rules of Kitaigorodsky and Burger & Ramberger [512, 319] may not be as solid as previously thought. A likely explanation is that the variance in guest-to-host volume ratio R_g is too small. Clathrates tend to have R_g-values between 45 and 85 and this interval is too narrow for the expected correlations to be of importance.

There is however a trend in that methane clathrates appear to be more affected by temperature than xenon clathrates. Comparing Figures 8.6(a) to (e) and (b) to (f), it can be seen that at 100 K, xenon generally binds stronger than methane, but at 300 K the opposite is true. Xenon, having only three degrees of freedom is less affected by temperature than carbon dioxide and methane, with 5 and 6 degrees of freedom respectively. This has the effect that at higher temperatures, the methane and carbon dioxide clathrates are thermodynamically favoured over their iso-structural rare gas clathrates. Hence, the number of degrees of freedom for the guest molecule is an important factor to the free energy of enclathration and the relative stability of clathrates with different guests.

8.3.5 Predicted resorcinol clathrates

Six porous crystal structures with cavities suitable for the adsorption of xenon, carbon dioxide and methane were found in the set of structures from reference [558]. Five of these structures could be geometry-optimised to true energy-minima. This indicates that resorcinol may be able to form clathrates, similarly to many other phenol derivatives.

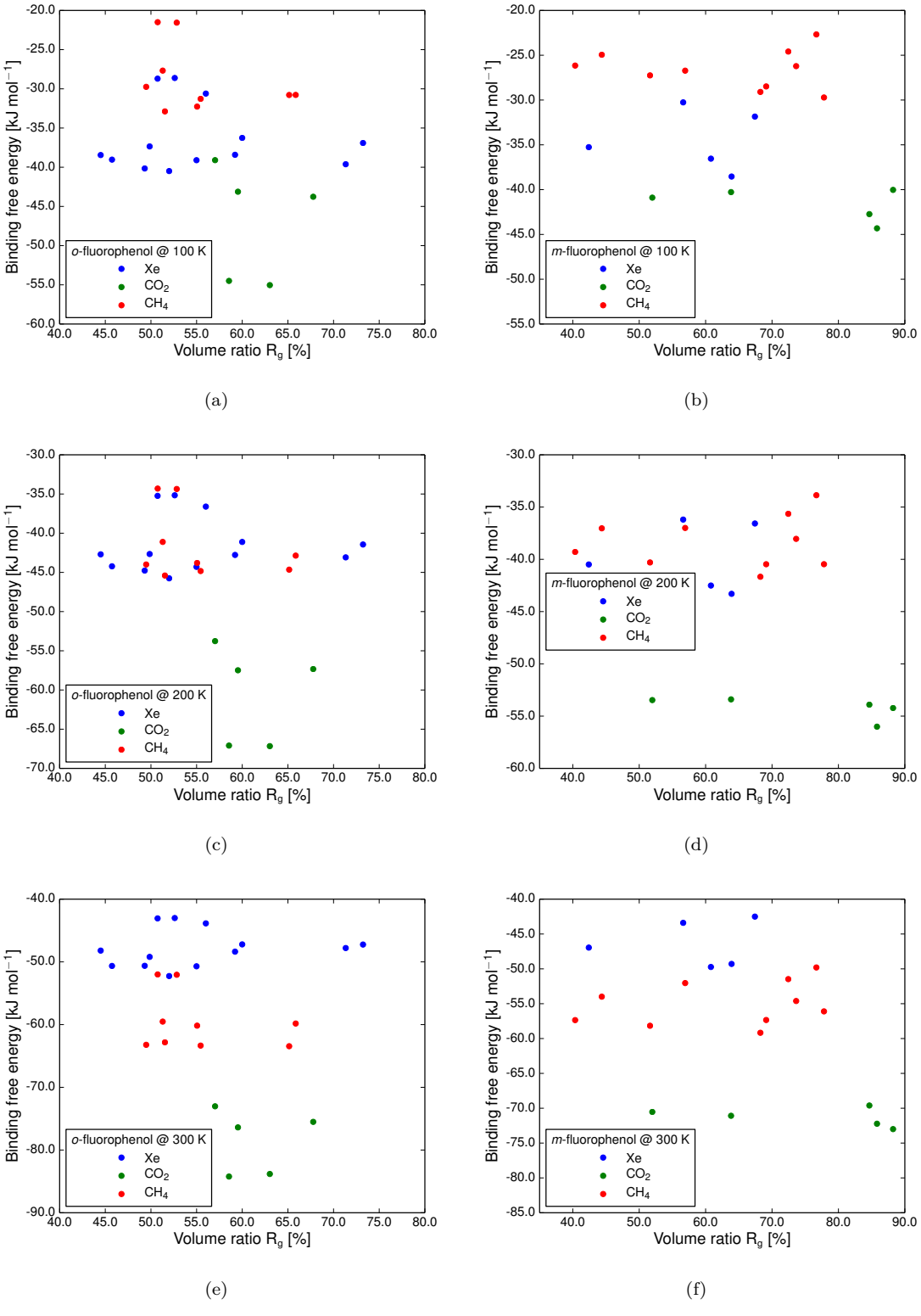
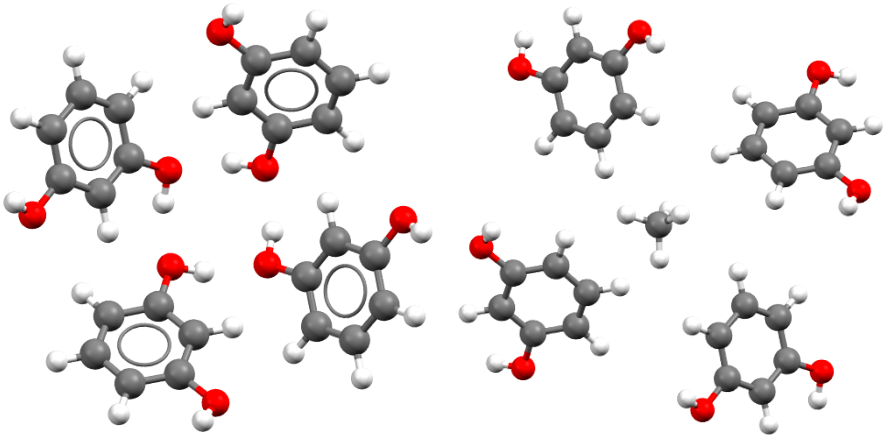


Figure 8.6: The binding free energy of three guest molecules in several fluorophenol clathrates at 100, 200 and 300 K. Because of the different number of vibrational degrees of freedom, CH_4 has a more temperature-dependent adsorption behaviour, and binds more strongly at higher temperatures, than Xe.

One of the predicted structures is shown in Figure 8.7. The empty framework of resorcinol structure 78_3643 has $P4_3$-symmetry and a prominent 4-fold screw axis infinite hydrogen bonded chain motif [563, 564]. Upon guest enclathration, the guest is adsorbed between the 4-fold chains, breaking the point group and lattice symmetries to $P1$.



(a) The resorcinol host structure 78_3643 has $P4_3$ symmetry.

(b) Resorcinol methane clathrate 78_3643_CH4 in space group $P1$.

Figure 8.7: The porous resorcinol structure 78_3643 obtained by CSP is predicted to be able to absorb xenon, methane and carbon dioxide. The guests adsorb between columns held together by chains of hydrogen bonds, breaking the $P4_3$ symmetry.

Predicted static and vibrational adsorption energies of the five host structures and their respective clathrates are shown in Tables 8.2 and 8.3. It is interesting to note that the vibrational contribution to the binding energies is large. For carbon dioxide and methane, the entropic contribution is larger than the static binding energy. It is however difficult to assess the accuracy of these results, beyond the general benchmarking of the force field performed in Chapter 3.

Table 8.2: Predicted lattice- and vibrational energies at 300 K of five predicted porous structures of resorcinol. Energies E_{latt} and F_{vib} in units of kJ/mol resorcinol. The kernel bandwidth is denoted h and has units of cm^{-1}, N_{k} is the number of sampled **k**-points, and S.g. is the space group number.

Host	E_{latt}	F_{vib}	h	N_{k}	S.g.
2_71899	−77.885	−17.467	7.79	18	2
14_41283	−75.593	−17.480	8.49	16	14
76_557	−72.650	−19.018	10.47	16	76
76_3511	−72.544	−18.918	10.92	16	76
78_3643	−75.290	−18.277	10.25	16	78

Further experimental work with resorcinol may reveal if these predicted clathrate phases can exist. An experimental procedure for preparing clathrates by dissolving the host compound in a solvent and bubbling the guest gas through the saturated solution may

be successful in obtaining new clathrates [565].

Table 8.3: Predicted adsorption energies of several predicted clathrates of resorcinol. Binding energies are given in kJ per mol of the guest. The kernel bandwidth is denoted h and has units of cm^{-1}, N_k is the number of sampled **k**-points. The space group number is denoted S.g., and H:G denotes the stoichiometric ratio between the host molecules H and the guest G.

Clathrate	E_{latt}^{bind}	F_{vib}^{bind}	h	N_k	S.g.	H:G
2_71899_Xe	−20.946	−10.138	6.80	17	2	4:1
14_41283_Xe	−10.448	−18.522	9.56	14	14	4:1
76_557_Xe	−27.090	−18.654	11.33	16	1	4:1
76_3511_Xe	−27.259	−18.180	12.04	16	1	4:1
78_3643_Xe	−24.154	−13.408	10.22	17	1	4:1
76_557_CO2	−28.198	−38.036	9.83	16	1	4:1
78_3643_CO2	−27.355	−33.983	7.74	16	1	4:1
76_557_CH4	−19.647	−39.000	8.31	17	1	4:1
76_3511_CH4	−19.907	−38.497	9.55	17	1	4:1
78_3643_CH4	−21.998	−27.192	9.88	16	1	4:1

8.3.6 Predicted TTBI clathrates

TTBI is known to form inclusion compounds with solvents, where the solvent molecules are adsorbed in relatively large channels. Because of this, and the unusual symmetric shape of the molecule, it was hypothesised that additional porous crystal structures may exist.

Several clathrate structures of TTBI were found as distinct lattice energy minima. Their predicted properties are displayed in Table 8.5. Figure 8.8 shows a typical TTBI methane clathrate. The iso-structural CO_2-clathrate was also stable, but the Xe-clathrate was not.

TTBI was recently subjected to CSP (providing the host structures used here) and an experimental search for new crystal structures as part of a large scale application of CSP for the discovery of new porous molecular materials. At the time of writing, the results of these efforts are not yet publicly available but a manuscript of Pulido *et al.* has been accepted for publication in *Nature Chemistry* [560].

Table 8.4: Predicted lattice- and vibrational energies at 300 K of several predicted porous structures of TTBI. Energies E_{latt} and F_{vib} in units of kJ/mol TTBI. The kernel bandwidth is denoted h and has units of cm^{-1}, N_{k} is the number of sampled **k**-points, and S.g. is the space group number.

Host	E_{latt}	F_{vib}	h	N_{k}	S.g.
9_2151	−188.024	−21.987	4.95	14	13
14_1647	−186.759	−22.445	5.21	21	18
14_3767	−188.113	−21.470	3.98	20	13
14_4224	−164.323	−22.405	8.32	15	14
14_507	−181.410	−21.587	4.02	20	7
15_280	−166.258	−23.987	7.51	11	15
15_8777	−161.991	−25.318	5.39	11	15
15_9011	−171.256	−23.350	6.40	11	15
29_2806	−186.767	−22.791	4.87	14	18
33_4095	−186.761	−22.789	5.24	14	18
41_7385	−149.295	−23.674	7.33	11	41
52_11062	−156.258	−24.885	6.64	11	52
56_3303	−157.924	−23.701	5.75	12	33
60_3700	−167.150	−22.951	5.50	11	14
61_2696	−156.810	−24.065	6.70	12	61
154_3352	−148.320	−24.326	7.01	13	154

Table 8.5: Predicted static and vibrational adsorption energies of several TTBI clathrates. Energies E_{latt} and F_{vib} in units of kJ/mol TTBI. The kernel bandwidth is denoted h and has units of cm^{-1}, N_{k} is the number of sampled **k**-points, and S.g. is the space group number.

Clathrate	$E_{\text{latt}}^{\text{bind}}$	$F_{\text{vib}}^{\text{bind}}$	h	N_{k}	S.g.	H:G
14_1647_Xe	−37.010	−16.753	3.25	21	18	1:1
15_9011_Xe	−10.183	−7.761	4.72	12	5	2:1
29_2806_Xe	−37.011	−16.745	4.11	14	18	1:1
52_11062_Xe	−39.583	−13.120	5.00	11	1	8:5
60_3700_Xe	−24.487	−14.671	3.18	11	1	8:9
14_3767_CO2	−37.629	−45.950	3.49	18	7	1:1
14_507_CO2	−44.572	−46.207	3.90	20	7	1:1
29_2806_CO2	−40.791	−44.598	3.64	14	18	1:1
33_4095_CO2	−40.840	−43.658	2.93	14	1	1:1
56_3303_CO2	−43.710	−25.189	2.39	12	1	4:3
9_2151_CH4	−22.012	−40.629	4.09	14	13	1:1
14_1647_CH4	−28.454	−36.233	4.20	21	18	1:1
14_3767_CH4	−21.784	−40.233	3.67	17	13	1:1
14_4224_CH4	−23.908	−17.403	3.35	15	1	4:9
14_507_CH4	−28.681	−40.476	3.90	20	13	1:1
15_280_CH4	−29.833	−36.238	9.31	12	15	2:1
15_8777_CH4	−24.768	−18.470	3.98	11	2	2:3
15_9011_CH4	−29.300	−34.630	3.51	11	1	1:1
29_2806_CH4	−28.436	−36.397	5.09	14	18	1:1
33_4095_CH4	−28.455	−36.377	5.11	14	18	1:1
41_7385_CH4	−23.373	−24.091	3.28	11	1	4:5
52_11062_CH4	−26.353	−16.096	2.39	11	1	8:11
61_2696_CH4	−21.515	−14.010	3.41	11	1	8:13
154_3352_CH4	−23.252	−19.306	4.84	13	1	6:5

8.4 Concluding remarks

This Chapter has demonstrated how predicted single-component porous crystal structures are potential clathrate host structures. A large number of plausible clathrates were relatively easily obtained by inserting guest molecules and performing geometry-optimisations. This is a viable method for predicting clathrates in general, and possibly also solvates. Especially for weakly interacting guests, it is likely that the apo-structure is a local lattice energy minimum that can be found by CSP-algorithms that are already in common use.

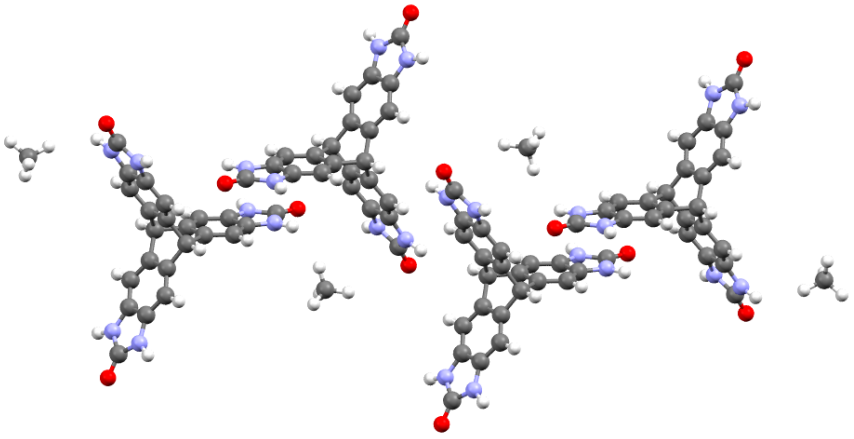


Figure 8.8: Packing diagram of the predicted TTBI methane clathrate structure 14_507_CH4.

Harmonic approximation lattice dynamics has been used to further confirm that the predicted structures are at least meta-stable and thermodynamically plausible. The Epanechnikov kernel density method introduced in Chapter 2 was used to improve the convergence of calculated phonon spectra and lattice-vibrational energies. Highly resolved phonon densities of states and adsorption free energies were calculated for a large number of clathrate host-guest combinations.

Suitable bandwidths were relatively easily obtained from an average dispersion between the Γ-point and the Brillouin zone boundary. The method is similar to, but easier to implement computationally, than the linear approximation method introduced by Gilat & Raubenheimer [562, 566]. The obtained bandwidths are remarkably consistent and fall in a limited range between 2.4 and 9.3 cm^{-1} for TTBI structures and between 6.8 and 12.0 cm^{-1} for resorcinol structures. That TTBI would tend to have smaller bandwidths was expected, since a denser **k**-point sampling was used. This suggests that Epanechnikov kernel bandwidth between 3 and 10 cm^{-1} can probably be used as a rule of thumb. The Gaussian kernel can probably be used with smaller bandwidth than these, since the kernel shape itself is wider.

The lattice vibrations of the guest molecule are partially additive, the lattice modes of the framework are only slightly shifted by the presence of a guest. There is not a large overall shift of the frequency spectrum towards higher frequencies that would make enclathration thermodynamically unfavourable. Instead, only some modes are perturbed, while others are not.

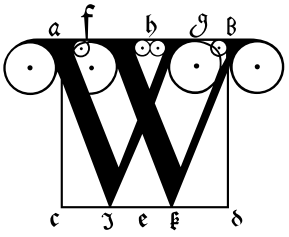
The results indicate that stable clathrates tend to have host-to-guest volume ratios of 45–85%, and that within this range, the effect of the cavity size on the static and vibrational contributions to the free energy of adsorption are small. There is no narrow and deep free energy optimum that makes one particular volume ratio much more favourable, but all volume ratios within reasonable bounds are thermodynamically plausible. This indicates that the close-packing principle is at least somewhat forgiving. Instead, it is the number of rotational degrees of freedom of the guest molecule that has a large effect, such that at elevated temperatures, methane is entropically stabilised substantially more than xenon.

It was confirmed that clathrates generally do not expand or contract much upon guest enclathration, and it was found that elastic moduli tend to increase slightly, but cases where the clathrate is softer than the apo-host exist.

Crystal structure prediction has over the last 20 years taught us that organic compounds have a large latent polymorphism. Every molecule, it seems, has a large number of crystal structures into which it potentially could crystallise. This study have given us a first peek into the latent reservoir of available clathrates of organic molecules.

Chapter 9

Summary and conclusion



e have seen how crystal structures and properties of polymorphs and clathrates can be predicted by computer calculations. In this thesis, large-scale computational crystal structure prediction and lattice dynamics calculations have been applied for the *a priori* prediction of crystal structures and their properties. Structures of both polymorphs and clathrates have been predicted and subsequently experimentally confirmed by collaborating research groups.

Temperature-dependent properties of many structures have been calculated and, when possible, compared to experimental results. This has revealed that, despite the relatively good performance of the energy model, the predictions generally do not reach quantitative accuracy. Predicted lattice parameters have errors of a few percent, while absolute lattice energies have relatively large errors of 16%. The error in the (free) energy-difference between polymorphs is about 5 kJ/mol. An error of this size would explain most, if not all, erroneous predictions seen in this work. Despite these errors, reliable predictions are possible for trends among large sets of structures, where the errors can be treated as random and normally distributes such that statistical methods are applicable. It is however unfortunate to have to rely on such fortuitous cancellations or errors.

Efforts should be invested in developing a new generation of highly accurate force fields. The Benchmark presented in Chapter 3 shows that force fields can be competitive in accuracy with density functional theory. Future force fields that explicitly include terms for short-distance damping of the dispersion and higher-order dispersion terms could probably be just as accurate as DFT-D at a dramatically smaller computational cost. Such force fields would probably greatly improve the predictive ability of the algorithms presented in this thesis.

Chapter 4 presented my contributions to the latest blind test of crystal structure prediction [5]. This largest test ever of the current state of the art in CSP research showed

that two scoring functions have emerged as the most reliable; DFT-D lattice energies and harmonic approximation free energies calculated either with DFT-D or Williams type force fields. Lattice dynamics with DFT-D is however so computationally costly that normally only $\mathbf{k} = \mathbf{0}$ phonons are accessible. This causes an error in the calculated vibrational energy that is of the same size as the accuracy gained by using quantum-chemical methods instead of force fields. Several research groups were able to predict the crystal structures of the smaller, relatively rigid, molecules. Large flexible molecules and multicomponent systems are still quite challenging and predictions of such systems are far less reliable.

Highly resolved distributions for property-differences between polymorphs were calculated in Chapter 5. This study derived general properties and trends among organic packing polymorphs, which should be relevant to many future studies of polymorphism. Several of the trends observed in the computations in this thesis has made me more sceptical of some of the empirical guidelines and rules suggested by earlier works. The conclusion of Gavezzotti [23] that the thermal vibrational energy never causes a re-ranking of polymorph stability was shown to be incorrect. Instead, it was concluded that at room temperature about 9% of polymorphs are re-ranked, and at high temperatures up to 21% are re-ranked. This is probably in agreement with experimental results, though no estimate of the commonality of enantiotropism has been found in the literature. It was also shown that the rules of Burger & Ramberger [319] should probably be seen as the most common behaviour of polymorphs, not as generally applicable laws. In particular, one cannot assume that the most densely packed polymorph is always the most stable, or having the least entropy [512].

The problem of Brillouin zone sampling in lattice dynamics has been dealt with extensively in this thesis. Computationally efficient methods have been developed that greatly improve the convergence and applicability of lattice-vibrational free energy calculations. The kernel density approximation was shown to have a scientifically sound justification, and was used together with the Debye approximation, leading to a realistic modelling of both acoustic and optic phonon modes. These methods were then used to predict adsorption free energies of clathrates in Chapter 8.

The methods for phonon dispersion modelling proposed here should also be useful in quantum-chemical lattice-vibrational calculations where the improved convergence might allow routine application of well-converged lattice dynamics calculations at the DFT-D or *ab initio* fragment-based levels of theory. I strongly recommend that the KDE and Debye methods are applied to periodic DFT-D calculations.

Several methods and algorithms used and developed in this thesis have been implemented in three computer programs that are made freely available. These are described in Appendices A–C. The programs greatly simplify the calculations described in this thesis and should facilitate further studies of the organic solid state.

Crystal structure prediction has evolved over time. From a search for *the* crystal structure in the 1990s, the scientific community is now focusing on the prediction of the entire landscape of polymorphs, hydrates, solvates and other crystal forms. There is now convincing evidence for the notion that most molecules have a large reservoir of latent polymorphs waiting to be discovered. Because of the large number of plausible polymorphs, and the small energy-differences between them, it is unlikely that any one scoring function can be used to rank the predicted crystal forms. Instead, the properties of all low-energy forms should be computed and the most desirable forms be targeted in rationally designed crystallisation experiments.

Put in other words, CSP will probably never become completely reliable in the sense that one scoring function consistently ranks the most stable form as the global optimum. It is important to realize that this is not necessary for CSP to be practically useful. The methods used in this thesis are clearly not reliable in that regard, yet every CSP calculation described here yielded *a priori* knowledge and information that aided subsequent experimental work, and provided a deeper insight and understanding of measurements.

The combined use of crystal structure prediction and experimental work, of almost any kind, has proven to be scientifically highly productive. In particular, CSP allows structure determination from otherwise inconclusive experimental data. The interface between CSP and experimental crystal form screening should be fertile soil for future doctorates.

Porous crystal structures are rare, but as we have seen in this thesis, perfectly ordinary and simple molecules may have a large number of plausible clathrand structures. The relatively small number of currently known clathrates may be due to an experimental bias. Only a small number of molecules have been thoroughly tested for their ability to form clathrates. CSP studies have revealed latent polymorphism in many compounds. It is my opinion that there also exist a large number of latent solvates and clathrate structures of ordinary organic compounds. These latent reservoirs of crystal structures are most likely accessible at pressure-, temperature- and time scales other than those typically encountered in Earthly laboratories.

Instead, it may be the case that exoplanets form the natural habitats of these crystal structures. A growing body of evidence is pointing in this direction. For instance, crystal structure prediction of inorganic phases at high pressures by the Oganov group have revealed a large number of materials that may be stable in exoplanetary environments [164, 567, 568]. A recently discovered benzene ethane clathrate has been suggested to form in the hydrocarbon lakes on Titan [541]. That clathrate structure and its properties could probably have been predicted by the methods presented in this thesis, but in this case the experimentalists beat me to it.

Appendices

Appendix A

The AutoDof program

A Python script, AutoDof, was written to automatically select the most relevant flexible degrees of freedom from the Z-matrix generated by CrystalOptimizer. The algorithm used for constructing the Z-matrix is described in Appendix A of the PhD-thesis by Kazantsev [237].

The program chooses a suitable set of flexible bond angles and dihedrals for CrystalOptimizer 2.4.x [236, 237]. The AutoDof script is compatible with both Python 2.7 and 3.5. It requires only the SciPy sparse matrices library.

The script reads a file in the current directory that must be named 'FirstRun.out'. This file is the output file that CrystalOptimizer produces when no degrees of freedom are specified. This file contains a Z-matrix and details of the atomic connectivity and Williams force field types.

All angles and most dihedrals can be determined to be flexible or rigid only from the atom types of the constituent atoms. Dihedrals in rings may require some more analysis. To determine if a dihedral belongs to a ring, a connectivity matrix is constructed of the molecule. The bond in the middle of the dihedral is cut by setting the corresponding elements in the connectivity matrix to zero. It is then used to initialize a SciPy **csgraph** object via a SciPy sparse matrix. The csgraph class contains a **shortest_path()** function which can be used to obtain the atoms in the ring. The atom types of the atoms in the ring can then be used to deduce if the dihedral should be flexible or not.

The results, the selected degrees of freedom, are written to the file **autodof.out**. A second call to CrystalOptimizer, using the automatically chosen degrees of freedom can then be made:

CrystalOptimizer structure.res $(< autodof.out) < CrystOpt.input > SecondRun.out

The AutoDof script is intended to implement the selection rules below, first developed and used to geometry-optimise large sets of diverse crystal structures [2, 3]. The selected degrees of freedom are suitable for most molecules, and gives economical calculations for large molecules.

Each bond angle is determined to be either flexible, rigid or constrained. These are the only possibilities. Dihedral angles can be flexible, rigid or duplicate. A duplicate dihedral is a dihedral angle around a bond that has already been found to be flexible by the program. Normally, only one dihedral per rotatable bond is set as flexible, the exception is dihedrals with polar hydrogen atoms, which are always flexible.

Only angles and dihedrals in the Z-matrix in the FirstRun.out file are considered. CrystalOptimizer's internal algorithm for generating the Z-matrix is quite suitable. There is no need to construct alternative Z-matrices. All degrees of freedom that are flexible according to the rules will be considered by AutoDof. This is an important advantage of these specific rules.

The selection rules are:

- Covalent bond lengths are always optimised without considering packing forces.
- All angles and dihedrals containing a polar hydrogen atom (–OH, –NH, –SH) are optimised under the influence of packing forces.
- Exocyclic bonds are always considered rotatable and are optimised under the influence of crystal packing forces.
- Dihedrals and angles in 3- and 4-membered rings are optimised without considering packing forces.
- Dihedrals and angles in 5- and 6-membered rings consisting only of 3-coordinated carbon atoms and nitrogen in any combination are unaffected by packing forces, except dihedrals and angles that contain a polar hydrogen atom.
- Dihedrals and angles in 5-membered rings containing sulfur and/or oxygen atoms each bonded to two 3-coordinated carbon atoms are optimised without considering packing forces.
- If several dihedral angles in the Z-matrix have same two atoms in the middle, only the first dihedral in the Z-matrix is chosen as flexible and unless the additional dihedral also include a polar hydrogen atom, in which case it is flexible.
- Any remaining dihedrals are optimised with respect to packing forces.
- Linear angles around triple bonds are constrained to 179.99 degrees.

For completely rigid molecules we allow the first angle to be optimized with respect to packing forces, since CrystalOptimizer needs at least one flexible degree of freedom.

The AutoDOF program and its source code has been made freely available to the scientific community under a permissive MIT license.

Appendix B

The AutoLD program

AutoLD is a Python 2.7/3.5 program for automatically making input files to DMACRYS [168] for lattice dynamics calculations on supercells. The script creates dmain files containing supercells, the size of which can be either input as an argument, or selected automatically according to methods described in ref. [1] and [2]. Files generated with AutoLD have been tested and works with DMACRYS versions 2.0.4 and 2.2.2.

The AutoLD program takes the following arguments, all of which are optional.

-i A string specifying the name of the input file. A DMACRYS (dmain) input file. This should be generated by other means, such as with NEIGHCRYS. If no name is specified and there is exactly one dmain file in the working directory, that file will be used.

-k The target **k**-point distance in units of $Å^{-1}$. The default is 0.12.

-l Select 'yes' or 'no', to generate linear supercells or not. This is 'yes' by default

-s Choose 'yes' or 'no', to split linear supercells into shorter cells with mutually co-prime expansion coefficients. This defaults to 'yes'. This has no effect if rectangular cells are made.

-c Cells, a list of supercell dimensions to be generated. Should be a text string of the form [1,1,2][1,2,1].

The default method for choosing expansion coefficients is to specify a target distance between **k**-points in reciprocal space, typically around 0.2 ($Å^{-1}$). Three supercells, expanded along lattice vectors **a**, **b** or **c** respectively are then generated. A lattice vector **a** is expanded until $\mathbf{a}^*$ is strictly shorter than the target distance.

Lattice dynamics calculations work best with unit cells that do not have very acute or obtuse angles. Suitable 'reduced cells' can be obtained with the Krivý-Gruber algorithm

[282], or the ADDSYMM routine in PLATON. Experimental crystal structures rarely require any modification.

If the three linear supercells are very long, it can be cost-effective to split them into several smaller supercells, each sampling different **k**-points. A co-prime splitting method [2] is implemented and is used by default. This allows a larger number of **k**-points to be sampled with distances down to about 0.05 $Å^{-1}$. Up to 26 **k**-points can be sampled in each direction.

It is also possible to generate a single rectangular supercell, which samples **k**-points in a regular grid in reciprocal space. Again, the optimal expansion coefficients are selected in order to make the supercell as close to a cube as possible. A target **k**-point distance of about 0.45 is recommended for rectangular supercells.

The dmain files may contain Williams (C_W3___), FIT (C_F1___) or custom atom types (C_C1___ or C_01___). These atom types work for H, C, N, O, F, S, Cl, Xe and 'virtual' atoms with atom type V_01, which facilitate calculations with linear molecules.

A manual and some example dmain files are provided for testing purposes and for demonstrating the capabilities of the program.

The AutoLD program and its source code has been made freely available to the scientific community under a permissive MIT license.

Appendix C

The AutoFree program

AutoFree is a Python 2.7/3.5 program for calculating free energies and properties of crystals from DMACRYS output files.

The program takes the following arguments:

-i Input file, a zip archive containing dmain and dmaout files.

-t Temperature in Kelvin. Properties are calculated at this temperature. Default is 300 K.

-b Kernel bandwidth in units of cm^{-1}. This value is used for the Epanechnikov kernel. Half this bandwidth is use for the Gaussian kernel. Default is 6.0.

The program reads lattice vectors, phonon eigenvectors and eigenvalues, elastic tensors and Z-values from one or several dmaout files where lattice dynamics calculations have been performed. DMACRYS versions between 2.0.4 and 2.2.2 are recommended. Files without LD-calculations are ignored.

The speed of sound and the Debye frequency is calculated from the elastic tensor. The Debye frequency is used to improve the convergence of the thermodynamic properties.

A kernel density estimate can also be used to improve the convergence of the phonon density of states. Gaussian and Epanechnikov kernel estimates are calculated. The kernel bandwidth should be set with some care. The phonon frequencies and kernel densities are written to disk which facilitate plotting the phonon spectrum.

It is fairly common that LD-calculations fail because the crystal structure is not a true energy minimum and the Born stability criterion is not met. It may be necessary to re-optimise the unit cell structure without the NOPR directive but with the SYMM and SEIG 1 commands to push the structure off a potential energy saddle point.

The AutoFree program and its source code has been made freely available to the scientific community under a permissive MIT license.

Appendix D

The Nyman Polymorph Library

As supplementary data to reference [3], a large data set was published. The purpose was to facilitate future computational studies of polymorph differences.

The optimised crystal structures at both 0 K and at the melting point of each structure is provided in the form of two crystallographic information files (.cif). Please see the files **NPL2016_0K.cif** and **NPL2016_Tm.cif** for these structures. We also provide a list of all Cambridge Structural Database (CSD) reference codes as a txt/gcd. The gcd file can be opened in CSD software (Mercury or Conquest) to view all of the structures.

Additional data is provided in the comma-separated value file (.csv) **NPL2016_data.csv**. This is a text file that can be opened in any text editor or spread sheet program. Each row contains the data for one crystal structure. The data fields are:

- The CSD reference code
- Z, the number of formula units (molecules) per unit cell
- Z′, the number of formula units per asymmetric unit
- The chemical formula of the molecule.
- The calculated unit cell volume of the 0 K optimised structure, in $Å^3$
- Experimentally determined melting point temperature in K.
- Predicted melting point temperature in K.
- The thermal pressure at the melting point temperature, in GPa.
- The SMILES string of the molecule.
- The calculated bulk modulus at 0 K in GPa, from the Hill average of the elastic tensor.

- The calculated shear modulus at 0 K in GPa, from the Hill average of the elastic tensor.
- The calculated volumetric thermal expansion coefficient in 10^{-6} K^{-1}.

This set of crystal structures is highly suitable for further studies of polymorphism by other computational methods. The structures have been carefully chosen as the best known representation of each polymorph. Experimental artefacts such as missing or disordered hydrogen atoms have been excluded. The symmetries have been modified so that all structures have integer Z'. Thermal pressures and other data are included so that quasi-harmonic calculations are not necessary to make in possible much more expensive computational methods. The chemical formula and SMILES strings are provided so that the structures can easily be categorised into subsets depending on the size or chemistry of the constituent molecules.

References

[1] J. Nyman and G. M. Day. Static and lattice vibrational energy differences between polymorphs. *CrystEngComm*, 17:5154–5165, 2015.

[2] J. Nyman, O. Sheehan Pundyke, and G. M. Day. Accurate force fields and methods for modelling organic molecular crystals at finite temperatures. *Physical Chemistry Chemical Physics*, 18:15828–15837, 2016.

[3] J. Nyman and G. M. Day. Modelling temperature-dependent properties of polymorphic organic molecular crystals. *Physical Chemistry Chemical Physics*, 18:31132–31143, 2016.

[4] M. Selent, J. Nyman, J. Roukala, M. Ilczyszyn, R. Oilunkaniemi, P. J. Bygrave, R. Laitinen, J. Jokisaari, G. M. Day, and P. Lantto. Clathrate structure determination by combining crystal structure prediction with computational and experimental 129Xe NMR spectroscopy. *Chemistry – A European Journal*, 23(22):5258–5269, 2017.

[5] Anthony M. Reilly and Richard I. Cooper *et al.* Report on the sixth blind test of organic crystal structure prediction methods. *Acta Crystallographica Section B*, 72(4):439–459, Aug 2016.

[6] A. D. McNaught, A. Wilkinson, M. Nic, J. Jirat, B. Kosata, and A. Jenkins. Compendium of chemical terminology (The Gold Book), 1997.

[7] International Tables for Crystallography Volume A: Space-group symmetry, 2006.

[8] U. Müller. *Symmetry relationships between crystal structures.* Oxford University Press, 1 edition, 2013.

[9] P. Luger. *Modern X-ray Analysis on Single Crystals: A Practical Guide.* Walter de Gruyter, 2014.

[10] B. Grenier and R. Ballou. Crystallography: Symmetry groups and group representations. In *EPJ Web of Conferences*, volume 22, page 00006. EDP Sciences, 2012.

[11] P. Niggli. *Geometrische Kristallographie des Diskontinuums*. Gebrüder Borntraeger, 1919.

[12] A. Gavezzotti. Structure and energy in organic crystals with two molecules in the asymmetric unit: causality or chance? *CrystEngComm*, 10(4):389–398, 2008.

[13] B. P. van Eijck and J. Kroon. Structure predictions allowing more than one molecule in the asymmetric unit. *Acta Crystallographica Section B*, 56(3):535–542, 2000.

[14] G. S. Papaefstathiou, A. J. Kipp, and L. R. MacGillivray. Exploiting modularity in template-controlled synthesis: a new linear template to direct reactivity within discrete hydrogen-bonded molecular assemblies in the solid state. *Chem. Commun.*, pages 2462–2463, 2001.

[15] J. Haleblian and W. McCrone. Pharmaceutical applications of polymorphism. *Journal of Pharmaceutical Sciences*, 58(8):911–929, 1969.

[16] A. J. Cruz-Cabeza and J. Bernstein. Conformational polymorphism. *Chemical reviews*, 114(4):2170–2191, 2013.

[17] K. R. Seddon. Pseudopolymorph: a polemic. *Crystal Growth & Design*, 4(6):1087–1087, 2004.

[18] A Nangia. Pseudopolymorph: retain this widely accepted term. *Crystal Growth & Design*, 6(1):2–4, 2006.

[19] A. J. Cruz-Cabeza, S. M. Reutzel-Edens, and J. Bernstein. Facts and fictions about polymorphism. *Chemical Society Reviews*, 44(23):8619–8635, 2015.

[20] J. Bernstein. *Polymorphism in Molecular Crystals*. Oxford University Press, 2002.

[21] A. I. Kitaigorodsky. *Molecular Crystals and Molecules*. Physical Chemistry. Academic Press, Inc, 1973.

[22] S. L. Price. Why don't we find more polymorphs? *Acta Crystallographica Section B: Structural Science, Crystal Engineering and Materials*, 69(4):313–328, 2013.

[23] A. Gavezzotti and G. Filippini. Polymorphic forms of organic crystals at room conditions: Thermodynamic and structural implications. *Journal of the American Chemical Society*, 117:12299–12305, 1995.

[24] E. Boldyreva. High-pressure polymorphs of molecular solids: when are they formed, and when are they not? some examples of the role of kinetic control. *Crystal Growth & Design*, 7(9):1662–1668, 2007.

[25] D. A. Bardwell, C. S. Adjiman, Y. A. Arnautova, and E. *et al.* Bartashevich. Towards crystal structure prediction of complex organic compounds – a report on the fifth blind test. *Acta Crystallographica Section B*, 67(6):535–551, Dec 2011.

[26] N. Blagden, M. De Matas, P. T. Gavan, and P. York. Crystal engineering of active pharmaceutical ingredients to improve solubility and dissolution rates. *Advanced drug delivery reviews*, 59(7):617–630, 2007.

[27] E. Joiris, P. Di Martino, C. Berneron, A.-M. Guyot-Hermann, and J.-C. Guyot. Compression behaviour of othorhombic paracetamol. *Pharmaceutical Research*, 15(7):1122–1130, 1998.

[28] S. R. Chemburkar, J. Bauer, K. Deming, H. Spiwek, K. Patel, J. Morris, R. Henry, and S. Spanton *et al.* Dealing with the impact of ritonavir polymorphs on the late stages of bulk drug process development. *Organic Process Research & Development*, 4(5):413–417, 2000.

[29] J. Bauer, S. Spanton, R. Henry, J. Quick, W. Dziki, W. Porter, and J. Morris. Ritonavir: an extraordinary example of conformational polymorphism. *Pharmaceutical research*, 18(6):859–866, 2001.

[30] S. L. Morissette, S. Soukasene, D. Levinson, M. J. Cima, and Ö. Almarsson. Elucidation of crystal form diversity of the HIV protease inhibitor ritonavir by high-throughput crystallization. *Proceedings of the National Academy of Sciences*, 100(5):2180–2184, 2003.

[31] D.-K. Bučar, R. W. Lancaster, and J. Bernstein. Disappearing polymorphs revisited. *Angewandte Chemie International Edition*, 54(24):6972–6993, 2015.

[32] J. D. Dunitz and J. Bernstein. Disappearing polymorphs. *Accounts of Chemical Research*, 28(4):193–200, 1995.

[33] American Thyroid Association, The Endocrine Society, and American Association of Clinical Endocrinologists. U.S. Food and Drug Administration's decision regarding bioequivalence of levothyroxine sodium. *Thyroid*, 14(7):486, August 2004.

[34] S. Mondal and G. Mugesh. Structure elucidation and characterization of different thyroxine polymorphs. *Angewandte Chemie*, 127(37):10983–10987, 2015.

[35] K. R. Chaudhuri. Crystallisation within transdermal rotigotine patch: is there cause for concern? *Expert opinion on drug delivery*, 5(11):1169–1171, 2008.

[36] H.-M. Wolff, L. Quere, and J. Riedner. Polymorphic form of rotigotine and process for production, 2012. US Patent 8,232,414.

[37] S. Krivonos and A. Weisman. Crystalline rotigotine base and preparation process therefor, January 1 2013. US Patent 8,344,165.

[38] I. B. Rietveld and R. Céolin. Rotigotine: unexpected polymorphism with predictable overall monotropic behavior. *Journal of pharmaceutical sciences*, 104(12):4117–4122, 2015.

[39] W. Cabri, P. Ghetti, G. Pozzi, and M. Alpegiani. Polymorphisms and patent, market, and legal battles: Cefdinir case study. *Organic Process Research & Development*, 11(1):64–72, 2007.

[40] J. Bernstein. Structural chemistry, fuzzy logic, and the law. *Israel Journal of Chemistry*, 2016.

[41] A. L. Grzesiak, M. Lang, K. Kim, and A. J. Matzger. Comparison of the four anhydrous polymorphs of carbamazepine and the crystal structure of form I. *Journal of Pharmaceutical Sciences*, 92(11):2260–2271, 2003.

[42] J.-B. Arlin, L. S. Price, S. L. Price, and A. J. Florence. A strategy for producing predicted polymorphs: catemeric carbamazepine form V. *Chemical Communications*, 47(25):7074–7076, 2011.

[43] A. J. Cruz Cabeza, G. M. Day, W. D. S. Motherwell, and W. Jones. Amide pyramidalization in carbamazepine: A flexibility problem in crystal structure prediction? *Crystal Growth & Design*, 6(8):1858–1866, 2006.

[44] A. J. Cruz Cabeza, G. M. Day, W. D. S. Motherwell, and W. Jones. Prediction and observation of isostructurality induced by solvent incorporation in multicomponent crystals. *Journal of the American Chemical Society*, 128(45):14466–14467, 2006.

[45] M. M. J. Lowes, M. R. Caira, A. P. Lotter, V. Watt, and J. G. Der. Physicochemical properties and x-ray structural studies of the trigonal polymorph of carbamazepine. *Journal of pharmaceutical sciences*, 76(9):744–752, 1987.

[46] A. J. Cruz Cabeza, G. M. Day, W. D. S. Motherwell, and W. Jones. Solvent inclusion in form II carbamazepine. *Chemical Communications*, 16:1600–1602, 2007.

[47] F. P. A. Fabbiani, L. T. Byrne, J. J. McKinnon, and M. A. Spackman. Solvent inclusion in the structural voids of form II carbamazepine: single-crystal X-ray diffraction, NMR spectroscopy and Hirshfeld surface analysis. *CrystEngComm*, 9(9):728–731, 2007.

[48] L. Seton, D. Khamar, I. J. Bradshaw, and G. A. Hutcheon. Solid state forms of theophylline: presenting a new anhydrous polymorph. *Crystal Growth & Design*, 10(9):3879–3886, 2010.

[49] C. Roy, A. Vega-González, and P. Subra-Paternault. Theophylline formulation by supercritical antisolvents. *International Journal of Pharmaceutics*, 343(1–2):79–89, 2007.

[50] M. D. Eddleston, K. E. Hejczyk, E. G. Bithell, G. M. Day, and W. Jones. Determination of the crystal structure of a new polymorph of theophylline. *Chemistry – A European Journal*, 19(24):7883–7888, 2013.

[51] M. D. Eddleston, K. E. Hejczyk, A. M. C. Cassidy, G. M. Day, H. P. G. Thompson, and W. Jones. Understanding the growth of triangular crystals of theophylline: Morphological evidence for a new theophylline polymorph. Manuscript in preparation, 2014.

[52] D. Khamar, I. J. Bradshaw, G. A. Hutcheon, and L. Seton. Solid state transformations mediated by a kinetically stable form. *Crystal Growth & Design*, 12(1):109–118, 2011.

[53] D. Khamar, R. G. Pritchard, I. J. Bradshaw, G. A. Hutcheon, and L. Seton. Polymorphs of anhydrous theophylline: stable form IV consists of dimer pairs and metastable form I consists of hydrogen-bonded chains. *Acta Crystallographica Section C*, 67(12):o496–o499, Dec 2011.

[54] P. Rudel, S. Odiot, J. C. Mutin, and M. Peyrard. Structure cristalline et caractère détonique des cristaux moléculaires nitrés. *Journal de chimie physique*, 87(7-8):1307–1344, 1990.

[55] R. M. Vrcelj, J. N. Sherwood, A. R. Kennedy, H. G. Gallagher, and T. Gelbrich. Polymorphism in 2-4-6 trinitrotoluene. *Crystal Growth & Design*, 3(6):1027–1032, 2003.

[56] J. P. Lewis, T. D. Sewell, R. B. Evans, and G. A. Voth. Electronic structure calculation of the structures and energies of the three pure polymorphic forms of crystalline HMX. *The Journal of Physical Chemistry B*, 104(5):1009–1013, 2000.

[57] D. C. Sorescu, B. M. Rice, and D. L. Thompson. Isothermal-isobaric molecular dynamics simulations of 1,3,5,7-tetranitro-1,3,5,7-tetraazacyclooctane (HMX) crystals. *The Journal of Physical Chemistry B*, 102(35):6692–6695, 1998.

[58] D. C. Sorescu, B. M. Rice, and D. L. Thompson. Theoretical studies of the hydrostatic compression of RDX, HMX, HNIW, and PETN crystals. *The Journal of Physical Chemistry B*, 103(32):6783–6790, 1999.

[59] C. Loisel, G. Keller, G. Lecq, C. Bourgaux, and M. Ollivon. Phase transitions and polymorphism of cocoa butter. *Journal of the American Oil Chemists' Society*, 75(4):425–439, 1998.

[60] J. Bricknell and R. W. Hartel. Relation of fat bloom in chocolate to polymorphic transition of cocoa butter. *Journal of the American Oil Chemists' Society*, 75(11):1609–1615, 1998.

[61] S. Kishimoto, N. Nagashima, M. Naruse, and K. Toyokura. The 'bundlelike' crystals in aspartame crystallization. In *Process Technology Proceedings*, volume 6, pages 511–14, 1989.

[62] A. Whitaker. X-ray powder diffraction of synthetic organic colorants. In *Analytical chemistry of synthetic colorants*, pages 1–48. Springer, 1995.

[63] L. Yu. Polymorphism in molecular solids: an extraordinary system of red, orange, and yellow crystals. *Accounts of chemical research*, 43(9):1257–1266, 2010.

[64] S. Chen, I. A. Guzei, and L. Yu. New polymorphs of ROY and new record for coexisting polymorphs of solved structures. *Journal of the American Chemical Society*, 127(27):9881–9885, 2005.

[65] S. S. Labana and L. L. Labana. Quinacridones. *Chemical Reviews*, 67(1):1–18, 1967.

[66] E. F. Paulus, F. J. J. Leusen, and M. U. Schmidt. Crystal structures of quinacridones. *CrystEngComm*, 9(2):131–143, 2007.

[67] A. I. Kitaigorodskii. The principle of close packing and the condition of thermodynamic stability of organic crystals. *Acta Crystallographica*, 18(4):585–590, 1965.

[68] T. C. Hales. An overview of the Kepler conjecture. *arXiv preprint math/9811071*, 1998.

[69] N. B. McKeown. Nanoporous molecular crystals. *Journal of Materials Chemistry*, 20(47):10588–10597, 2010.

[70] J. Tian, P. K. Thallapally, and B. P. McGrail. Porous organic molecular materials. *CrystEngComm*, 14(6):1909–1919, 2012.

[71] M. J. Bojdys, M. E. Briggs, J. T. A. Jones, D. J Adams, S. Y. Chong, M. Schmidtmann, and A. I. Cooper. Supramolecular engineering of intrinsic and extrinsic porosity in covalent organic cages. *Journal of the American Chemical Society*, 133(41):16566–16571, 2011.

[72] G. O. Lloyd, M. W. Bredenkamp, and L. J. Barbour. Enclathration of morpholinium cations by Dianin's compound: salt formation by partial host-to-guest proton transfer. *Chemical Communications*, pages 4053–4055, 2005.

[73] S.-P. Kang and H. Lee. Equilibrium dissociation pressures for p-cresol + methane and p-cresol + nitrogen clathrates at temperatures above the normal melting temperature of p-cresol. *J. Chem. Eng. Data*, 42(3):467–469, 1997.

[74] R. M. Barrer and V. H. Shanson. Clathration by para-substituted phenols. *Journal of the Chemical Society, Faraday Transactions 1: Physical Chemistry in Condensed Phases*, 72:2348–2354, 1976.

[75] A. J. Cruz-Cabeza, G. M. Day, and W. Jones. Predicting inclusion behaviour and framework structures in organic crystals. *Chemistry – A European Journal*, 15(47):13033–13040, 2009.

[76] R. V. Belosludov, V. R. Belosludov, and V. E. Zubkus. The lattice dynamics of β-hydroquinone clathrate. *The Journal of chemical physics*, 103(8):2773–2778, 1995.

[77] J.-W. Lee, P. Dotel, J. Park, and J.-H. Yoon. Separation of CO2 from flue gases using hydroquinone clathrate compounds. *Korean Journal of Chemical Engineering*, 32(12):2507–2511, 2015.

[78] J.-P. Torré, R. Coupan, M. Chabod, E. Pere, S. Labat, A. Khoukh, R. Brown, J.-M. Sotiropoulos, and H. Gornitzka. CO2–hydroquinone clathrate: Synthesis, purification, characterization and crystal structure. *Crystal Growth & Design*, 16(9):5330–5338, 2016.

[79] J. L Daschbach, T.-M. Chang, L. R. Corrales, L. X. Dang, and P. McGrail. Molecular mechanisms of hydrogen-loaded β-hydroquinone clathrate. *The Journal of Physical Chemistry B*, 110(35):17291–17295, 2006.

[80] K. W. Han, Y.-J. Lee, J. S. Jang, T.-I. Jeon, J. Park, T. Kawamura, Y. Yamamoto, T. Sugahara, T. Vogt, and J.-W. *et al.* Lee. Fast and reversible hydrogen storage in channel cages of hydroquinone clathrate. *Chemical Physics Letters*, 546:120–124, 2012.

[81] V. F. Rozsa and T. A. Strobel. Triple guest occupancy and negative compressibility in hydrogen-loaded β-hydroquinone clathrate. *The journal of physical chemistry letters*, 5(11):1880–1884, 2014.

[82] M. Ilczyszyn, M. Selent, and M. M. Ilczyszyn. Participation of xenon guest in hydrogen bond network of β-hydroquinone crystal. *The Journal of Physical Chemistry A*, 116(12):3206–3214, 2012.

[83] V.-V. Telkki, J. Lounila, and J. Jokisaari. Determination of pore sizes and volumes of porous materials by 129Xe NMR of xenon gas dissolved in a medium. *The Journal of Physical Chemistry B*, 109(51):24343–24351, 2005. PMID: 16375434.

[84] V.-V. Telkki, J. Lounila, and J. Jokisaari. Xenon porometry at room temperature. *The Journal of Chemical Physics*, 124(3):034711, 2006.

[85] A. Stone. *The Theory of Intermolecular Forces*. Oxford University Press, 2 edition, 2013.

[86] T. L. Hill. Steric effects. I. Van der Waals potential energy curves. *The Journal of Chemical Physics*, 16(4):399–404, 1948.

[87] A. J. Pertsin and A. I. Kitaigorodskii. The atom–atom potential method: Applications to organic molecular solids, 1987.

[88] H. Sun. Compass: an ab initio force-field optimized for condensed-phase applications overview with details on alkane and benzene compounds. *The Journal of Physical Chemistry B*, 102(38):7338–7364, 1998.

[89] G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, and W. L. Jorgensen. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *The Journal of Physical Chemistry B*, 105(28):6474–6487, 2001.

[90] F. London. The general theory of molecular forces. *Trans. Faraday Soc.*, 33:8b–26, 1937.

[91] D. E. Williams and S. R. Cox. Nonbonded potentials for azahydrocarbons: the importance of the Coulombic interaction. *Acta Crystallographica Section B*, 40(4):404–417, Aug 1984.

[92] D. E. Williams. Improved intermolecular force field for molecules containing H, C, N, and O atoms, with application to nucleoside and peptide crystals. *Journal of Computational Chemistry*, 22(11):1154–1166, 2001.

[93] D. E. Williams. Improved intermolecular force field for crystalline oxohydrocarbons including O-H$\cdots$O hydrogen bonding. *Journal of Computational Chemistry*, 22(1):1–20, 2001.

[94] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Physical review*, 136(3B):B864, 1964.

[95] A. D. Becke. Perspective: Fifty years of density-functional theory in chemical physics. *The Journal of chemical physics*, 140(18):18A301, 2014.

[96] W. Koch and M. C. Holthausen. *A chemist's guide to density functional theory*. John Wiley & Sons, 2015.

[97] S. Grimme. Accurate description of van der Waals complexes by density functional theory including empirical corrections. *Journal of computational chemistry*, 25(12):1463–1473, 2004.

[98] S. Grimme. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *Journal of Computational Chemistry*, 27(15):1787–1799, 2006.

[99] S. Grimme, S. Ehrlich, and L. Goerigk. Effect of the damping function in dispersion corrected density functional theory. *Journal of Computational Chemistry*, 32(7):1456–1465, 2011.

[100] N. Marom, R. A. DiStasio, V. Atalla, S. Levchenko, A. M. Reilly, J. R. Chelikowsky, L. Leiserowitz, and A. Tkatchenko. Many-body dispersion interactions in molecular crystal polymorphism. *Angewandte Chemie International Edition*, 52(26):6629–6632, 2013.

[101] A. E. Gray, G. M. Day, M. Leslie, and S. L. Price. Dynamics in crystals of rigid organic molecules: contrasting the phonon frequencies calculated by molecular dynamics with harmonic lattice dynamics for imidazole and 5-azauracil. *Molecular Physics*, 102(9-10):1067–1083, 2004.

[102] P. Chantrenne and J.-L. Barrat. Finite size effects in determination of thermal conductivities: comparing molecular dynamics results with simple models. *Journal of Heat Transfer*, 126(4):577–585, 2004.

[103] J. A. van den Ende and H. M. Cuppen. Dynamics of the α and β polymorphs of dl-norleucine at different temperatures: Sliding to a partial phase transition. *Crystal Growth & Design*, 14(7):3343–3351, 2014.

[104] J. A. van den Ende, M. M. H. Smets, D. T. de Jong, S. J. T. Brugman, B. Ensing, P. T. Tinnemans, H. Meekes, and H. M. Cuppen. Do solid-to-solid polymorphic transitions in DL-norleucine proceed through nucleation? *Faraday discussions*, 179:421–436, 2015.

[105] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092, 1953.

[106] H. R. Karfunkel, F. J. J. Leusen, and R. J. Gdanitz. The ab initio prediction of yet unknown molecular crystal structures by solving the crystal packing problem. *Journal of Computer-Aided Materials Design*, 1(2):177–185, 1994.

[107] J. P. M. Lommerse, W. D. S. Motherwell, H. L. Ammon, J. D. Dunitz, A. Gavezzotti, D. W. M. Hofmann, F. J. J. Leusen, W. T. M. Mooij, S. L. Price, B. Schweizer, M. U. Schmidt, and B. P. van Eijck *et al.* A test of crystal structure prediction of small organic molecules. *Acta Crystallographica Section B*, 56(4):697–714, Aug 2000.

[108] M. A. Neumann, F. J. J. Leusen, and J. Kendrick. A major advance in crystal structure prediction. *Angewandte Chemie International Edition*, 47(13):2427–2430, 2008.

[109] B. Fultz. Vibrational thermodynamics of materials. *Progress in Materials Science*, 55(4):247–352, 2010.

[110] M. T. Dove. *Introduction to lattice dynamics*, volume 4. Cambridge university press, 1993.

[111] M. Born and K. Huang. *Dynamical theory of crystal lattices*, volume 188. Clarendon Press Oxford, 1954.

[112] H. Bonadeo, E. D'alessio, E. Halac, and E. Burgos. Lattice dynamics, thermodynamic functions, and phase transitions of p-dichloro- and 1,2,4,5-tetrachlorobenzene. *The Journal of Chemical Physics*, 68(10):4714–4721, 1978.

[113] G. Filippini, C. M. Gramaccioli, M. Simonetta, and G. B. Suffritti. Lattice-dynamical applications to crystallographic problems: consideration of the Brillouin zone sampling. *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography*, 32(2):259–264, 1976.

[114] G. M. Day. *Lattice Dynamical Studies of Molecular Crystals with Application to Polymorphism and Structure Prediction*. PhD thesis, University College London, 2003.

[115] A. Otero-de-la-Roza and E. R. Johnson. A benchmark for non-covalent interactions in solids. *The Journal of Chemical Physics*, 137(5):054103, 2012.

[116] J. Maddox. Crystals from first principles. *Nature*, 335(6187):201, Sep 1988.

[117] G. M. Day and W. D. S. Motherwell. An experiment in crystal structure prediction by popular vote. *Crystal growth & design*, 6(9):1985–1990, 2006.

[118] D. Hilbert. Mathematical problems. *Bulletin of the American Mathematical Society*, 8(10):437–479, 1902.

[119] D. E. Williams. Crystal packing of molecules. *Science*, 147(3658):605–606, 1965.

[120] D. E. Williams. A method of calculating molecular crystal structures. *Acta Crystallographica Section A*, 25(3):464–470, May 1969.

[121] D. E. Williams. Molecular packing analysis. *Acta Crystallographica Section A*, 28(6):629–635, Nov 1972.

[122] A. I. Kitaigorodsky and K. V. Mirskaya. Prediction of the structure of an organic crystal. *Materials Research Bulletin*, 7(11):1271 – 1280, 1972.

[123] A. V. Dzyabchenko. Theoretical structures of crystalline benzene: The search for a global minimum of the lattice energy in four space groups. *Journal of Structural Chemistry*, 25(3):416–420, 1984.

[124] A. Gavezzotti and G. Filippini. Crystal packing and lattice energies of polythienyls: calculations and predictions. *Synthetic Metals*, 40(2):257 – 266, 1991.

[125] A. Gavezzotti. The prediction of the crystal packing of organic molecules: geometry and intermolecular potentials. In *Electron crystallography of organic molecules*, pages 77–83. Springer, 1991.

[126] A. Gavezzotti. Are crystal structures predictable? *Accounts of chemical research*, 27(10):309–314, 1994.

[127] R. J. Gdanitz. Prediction of molecular crystal structures by Monte Carlo simulated annealing without reference to diffraction data. *Chemical physics letters*, 190(3):391–396, 1992.

[128] B. P. van Eijck, W. T. M. Mooij, and J. Kroon. Attempted prediction of the crystal structures of six monosaccharides. *Acta Crystallographica Section B: Structural Science*, 51(1):99–103, 1995.

[129] D. W. M. Hofmann and T. Lengauer. A Discrete Algorithm for Crystal Structure Prediction of Organic Molecules. *Acta Crystallographica Section A*, 53(2):225–235, Mar 1997.

[130] W. D. S. Motherwell, H. L. Ammon, J. D. Dunitz, A. Dzyabchenko, P. Erk, A. Gavezzotti, D. W. M. Hofmann, F. J. J. Leusen, J. P. M. Lommerse, and W. T. M. Mooij *et al.* Crystal structure prediction of small organic molecules: a second blind test. *Acta Crystallographica Section B: Structural Science*, 58(4):647–661, 2002.

[131] J. Apostolakis, D. W. M. Hofmann, and T. Lengauer. Derivation of a scoring function for crystal structure prediction. *Acta Crystallographica Section A*, 57(4):442–450, Jul 2001.

[132] D. W. M. Hofmann and J. Apostolakis. Crystal structure prediction by data mining. *Journal of Molecular Structure*, 647(1):17–39, 2003.

[133] D. W. M. Hofmann and J. Apostolakis. Crystal structure prediction by data mining. *Journal of Molecular Structure*, 647(1):17–39, 2003.

[134] G. M. Day, W. D. S. Motherwell, H. L. Ammon, S. X. M. Boerrigter, R. G. Della Valle, E. Venuti, A. Dzyabchenko, J. D. Dunitz, B. Schweizer, and B. P. van Eijck *et al.* A third blind test of crystal structure prediction. *Acta Crystallographica Section B: Structural Science*, 61(5):511–527, 2005.

[135] G. M. Day, T. G. Cooper, A. J. Cruz-Cabeza, K. E. Hejczyk, H. L. Ammon, S. X. M. Boerrigter, J. S. Tan, and R. G. Della Valle *et al.* Significant progress in predicting the crystal structures of small organic molecules – a report on the fourth blind test. *Acta Crystallographica Section B*, 65(2):107–125, Apr 2009.

[136] M. Habgood, I. J. Sugden, A. V. Kazantsev, C. S. Adjiman, and C. C. Pantelides. Efficient handling of molecular flexibility in ab initio generation of crystal structures. *Journal of chemical theory and computation*, 11(4):1957–1969, 2015.

[137] W. T. M. Mooij and F. J. J. Leusen. Multipoles versus charges in the 1999 crystal structure prediction test. *Physical Chemistry Chemical Physics*, 3(22):5063–5066, 2001.

[138] G. M. Day. Current approaches to predicting molecular organic crystal structures. *Crystallography Reviews*, 17(1):3–52, 2011.

[139] S. L. Price. The computational prediction of pharmaceutical crystal structures and polymorphism. *Advanced drug delivery reviews*, 56(3):301–319, 2004.

[140] B. P. van Eijck. Comparing hypothetical structures generated in the third Cambridge blind test of crystal structure prediction. *Acta Crystallographica Section B*, 61(5):528–535, Oct 2005.

[141] S. L. Price. Predicting crystal structures of organic compounds. *Chemical Society Reviews*, 43(7):2098–2111, 2014.

[142] A. Asmadi, M. A. Neumann, J. Kendrick, P. Girard, M.-A. Perrin, and F. J. J. Leusen. Revisiting the blind tests in crystal structure prediction: accurate energy ranking of molecular crystals. *The Journal of Physical Chemistry B*, 113(51):16303–16313, 2009.

[143] D. Weininger. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *Journal of chemical information and computer sciences*, 28(1):31–36, 1988.

[144] N. M. O'Boyle, M. Banck, A. James, C. Morley, T. van der Meersch, and G. R. Hutchison. Open Babel: An open chemical toolbox. *Journal of Cheminformatics*, 3(1):1–14, 2011.

[145] J. W. Ponder. TINKER - software tools for molecular design. http://dasher.wustl.edu/ffe.

[146] C. Ouvrard and S. L. Price. Toward crystal structure prediction for conformationally flexible molecules: The headaches illustrated by aspirin. *Crystal Growth & Design*, 4(6):1119–1127, 2004.

[147] H. P. G. Thompson and G. M. Day. Which conformations make stable crystal structures? Mapping crystalline molecular geometries to the conformational energy landscape. *Chemical Science*, 5:3173–3182, 2014.

[148] H. P. G. Thompsson. *Extending Crystal Structure Prediction Methods towards Flexible Molecules*. PhD thesis, University of Cambridge, 2014.

[149] T. G. Cooper, W. Jones, W. D. S. Motherwell, and G. M. Day. Database guided conformation selection in crystal structure prediction of alanine. *CrystEngComm*, 9:595–602, 2007.

[150] A. Gavezzotti. Normalized one-and two-dimensional distribution functions for interatomic distance and angle data from crystallographic databases. *Journal of Applied Crystallography*, 43(3):429–433, 2010.

[151] I. Kolossváry and W. C. Guida. Low mode search. An efficient, automated computational method for conformational analysis: Application to cyclic and acyclic alkanes and cyclic peptides. *Journal of the American Chemical Society*, 118(21):5011–5019, 1996.

[152] I. Kolossváry and W. C. Guida. Low-mode conformational search elucidated: Application to C39H80 and flexible docking of 9-deazaguanine inhibitors into PNP. *Journal of Computational Chemistry*, 20(15):1671–1684, 1999.

[153] J. A. Chisholm, S. Motherwell, P. R. Tulip, S. Parsons, and S. J. Clark. An ab initio study of observed and hypothetical polymorphs of glycine. *Crystal Growth & Design*, 5(4):1437–1442, 2005.

[154] S. Kirkpatrick and M. P. Vecchi *et al.* Optimization by simulated annealing. *science*, 220(4598):671–680, 1983.

[155] M. A. Neumann, C. Tedesco, S. Destri, D. R. Ferro, and W. Porzio. Bridging the gap – structure determination of the red polymorph of tetrahexylsexithiophene by Monte Carlo simulated annealing, first-principles DFT calculations and Rietveld refinement. *Journal of applied crystallography*, 35(3):296–303, 2002.

[156] J. A. R. P. Sarma and G. R. Desiraju. The supramolecular synthon approach to crystal structure prediction. *Crystal growth & design*, 2(2):93–100, 2002.

[157] A. Dey, M. T. Kirchner, V. R. Vangala, G. R. Desiraju, R. Mondal, and J. A. K. Howard. Crystal structure prediction of aminols: advantages of a supramolecular synthon approach with experimental structures. *Journal of the American Chemical Society*, 127(30):10545–10559, 2005.

[158] B. Hayes. Quasirandom ramblings. *American Scientist Magazine*, 99:282–287, 2011.

[159] I. M. Sobol. On the distribution of points in a cube and the approximate evaluation of integrals. *Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki*, 7(4):784–802, 1967.

[160] D. H. Case, J. E. Campbell, P. J. Bygrave, and G. M. Day. Convergence properties of crystal structure prediction by quasi-random sampling. *Journal of Chemical Theory and Computation*, 12(2):910–924, 2016. PMID: 26716361.

[161] A. V. Kazantsev, P. G. Karamertzanis, C. S. Adjiman, and C. C. Pantelides. Efficient handling of molecular flexibility in lattice energy minimization of organic crystals. *Journal of Chemical Theory and Computation*, 7(6):1998–2016, 2011.

[162] M. A. Neumann. Tailor-made force fields for crystal-structure prediction. *The Journal of Physical Chemistry B*, 112(32):9810–9829, 2008.

[163] N. L. Abraham and M. I. J. Probert. A periodic genetic algorithm with real-space representation for crystal structure and polymorph prediction. *Physical Review B*, 73(22):224104, 2006.

[164] A. R. Oganov and C. W. Glass. Crystal structure prediction using ab initio evolutionary techniques: Principles and applications. *The Journal of chemical physics*, 124(24):244704, 2006.

[165] J. A. Chisholm and S. Motherwell. COMPACK: a program for identifying crystal structure similarity using distances. *Journal of applied crystallography*, 38(1):228–231, 2005.

[166] A. J. Stone and S. L. Price. Some new ideas in the theory of intermolecular forces: anisotropic atom-atom potentials. *The Journal of Physical Chemistry*, 92(12):3325–3335, 1988.

[167] G. M. Day and S. L. Price. A nonempirical anisotropic atom–atom model potential for chlorobenzene crystals. *Journal of the American Chemical Society*, 125(52):16434–16443, 2003. PMID: 14692787.

[168] S. L. Price, M. Leslie, G. W. A. Welch, M. Habgood, L. S. Price, P. G. Karamertzanis, and G. M. Day. Modelling organic crystal structures using distributed multipole and polarizability-based model intermolecular potentials. *Physical Chemistry Chemical Physics*, 12:8478–8490, 2010.

[169] D. Giron. Thermal analysis and calorimetric methods in the characterisation of polymorphs and solvates. *Thermochimica acta*, 248:1–59, 1995.

[170] T. L. Threlfall. Analysis of organic polymorphs. A review. *Analyst*, 120(10):2435–2460, 1995.

[171] L. Kofler, A. Kofler, and M. Brandstätter. *Thermo-mikro-methoden zur kennzeichnung orgahischer stoffe und stoffgemische*. Verlag Chemie, 1954.

[172] A Kofler. Thermal analysis with a hot-stage microscope. *Z. Phys. Chem. Abteilung A*, 187:201–10, 1941.

[173] M. Kuhnert-Brandstätter, E. Junger, and A. Kofler. Thermo-microscopic and spectrophotometric determination of steroid hormones. *Microchemical Journal*, 9(2):105–133, 1965.

[174] M. Kuhnert-Brandstätter. Thermomicroscopy of organic compounds. *Wilson and Wilson's Comprehensive Analytical Chemistry*, 16:329, 1982.

[175] U. J. Griesser, D. Weigand, J. M. Rollinger, M. Haddow, and E. Gstrein. The crystal polymorphs of metazachlor. *Journal of thermal analysis and calorimetry*, 77(2):511–522, 2004.

[176] M. J. Turner, S. P. Thomas, M. W. Shi, D. Jayatilaka, and M. A. Spackman. Energy frameworks: insights into interaction anisotropy and the mechanical properties of molecular crystals. *Chemical Communications*, 51(18):3735–3738, 2015.

[177] D. Dey, S. P. Thomas, M. A. Spackman, and D. Chopra. 'quasi-isostructural polymorphism' in molecular crystals: inputs from interaction hierarchy and energy frameworks. *Chemical Communications*, 52(10):2141–2144, 2016.

[178] M. W. Shi, S. P. Thomas, G. A. Koutsantonis, and M. A. Spackman. Supramolecular recognition and energy frameworks in host–guest complexes of 18-crown-6 and sulfonamides. *Crystal Growth & Design*, 15(12):5892–5900, 2015.

[179] J. D. Dunitz and A. Gavezzotti. How molecules stick together in organic crystals: weak intermolecular interactions. *Chemical Society Reviews*, 38(9):2622–2633, 2009.

[180] G. R. Desiraju. Supramolecular synthons in crystal engineering—a new organic synthesis. *Angewandte Chemie International Edition in English*, 34(21):2311–2327, 1995.

[181] T. S. Thakur, R. Dubey, and G. R. Desiraju. Intermolecular atom–atom bonds in crystals – a chemical perspective. *IUCrJ*, 2(2):159–160, Mar 2015.

[182] C. Jelsch and Y. Bibila Mayaya Bisseyou. Atom interaction propensities of oxygenated chemical functions in crystal packings. *IUCrJ*, 4(2), Mar 2017.

[183] J. D. Dunitz. Intermolecular atom–atom bonds in crystals? *IUCrJ*, 2(2):157–158, Mar 2015.

[184] M. A. Spackman and D. Jayatilaka. Hirshfeld surface analysis. *CrystEngComm*, 11(1):19–32, 2009.

[185] J. J. McKinnon, D. Jayatilaka, and M. A. Spackman. Towards quantitative analysis of intermolecular interactions with hirshfeld surfaces. *Chemical Communications*, (37):3814–3816, 2007.

[186] E. Espinosa, E. Molins, and C. Lecomte. Hydrogen bond strengths revealed by topological analyses of experimentally observed electron densities. *Chemical Physics Letters*, 285(3–4):170 – 173, 1998.

[187] C. Garau, A. Frontera, D. Quiñonero, P. Ballester, A. Costa, and P. M. Deyá. A topological analysis of the electron density in anion–π interactions. *ChemPhysChem*, 4(12):1344–1348, 2003.

[188] A. Gavezzotti. The "sceptical chymist": intermolecular doubts and paradoxes. *CrystEngComm*, 15:4027–4035, 2013.

[189] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case. Development and testing of a general AMBER force field. *Journal of computational chemistry*, 25(9):1157–1174, 2004.

[190] L. E. Chirlian and M. M. Francl. Atomic charges derived from electrostatic potentials: A detailed study. *Journal of Computational Chemistry*, 8(6):894–905, 1987.

[191] G. M. Day, W. D. S Motherwell, and W. Jones. Beyond the isotropic atom model in crystal structure prediction of rigid molecules: Atomic multipoles versus point charges. *Crystal Growth & Design*, 5(3):1023–1033, 2005.

[192] A. Gavezzotti. Calculation of lattice energies of organic crystals: the PIXEL integration method in comparison with more traditional methods. *Zeitschrift für Kristallographie – Crystalline Materials*, 220(5/6):499–510, 2005.

[193] A. J. Stone. Distributed multipole analysis, or how to describe a molecular charge distribution. *Chemical Physics Letters*, 83(2):233–239, 1981.

[194] A. T. Amos and R. J. Crispin. Intermolecular forces between large molecules. *Molecular Physics*, 31(1):159–176, 1976.

[195] C. G. Gray. On the theory of multipole interactions. *Canadian Journal of Physics*, 46(2):135–139, 1968.

[196] M. J. Caola. Solid harmonics and their addition theorems. *J. Phys. A: Math. Gen*, 11(2), 1978.

[197] A. J. Misquitta, A. J. Stone, and F. Fazeli. Distributed multipoles from a robust basis-space implementation of the iterated stockholder atoms procedure. *Journal of Chemical Theory and Computation*, 10(12):5405–5418, 2014.

[198] A. J. Stone. Distributed multipole analysis of Gaussian wavefunctions. GMDA version 2.2.02.

[199] A. J. Stone. Distributed multipole analysis: Stability for large basis sets. *Journal of Chemical Theory and Computation*, 1(6):1128–1132, 2005.

[200] G. G. Ferenczy, C. A. Reynolds, P. J. Winn, and A. J. Stone. MULFIT: a program for calculating electrostatic potential-fitted charges, 1998.

[201] J. G. Ángyán, C. Chipot, F. Dehez, C. Hättig, G. Jansen, and C. Millot. OPEP: A tool for the optimal partitioning of electric properties. *Journal of Computational Chemistry*, 24(8):997–1008, 2003.

[202] P. P. Ewald. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Annalen der Physik*, 369(3):253–287, 1921.

[203] P. G. Cummins, D. A. Dunmur, R. W. Munn, and R. J. Newham. Applications of the ewald method. I. calculation of multipole lattice sums. *Acta Crystallographica Section A*, 32(5):847–853, Sep 1976.

[204] B. P. van Eijck and J. Kroon. Coulomb energy of polar crystals. *The Journal of Physical Chemistry B*, 101(6):1096–1100, 1997.

[205] B. P. van Eijck and J. Kroon. Comment on "crystal structure prediction by global optimization as a tool for evaluating potentials: role of the dipole moment correction term in successful predictions". *The Journal of Physical Chemistry B*, 104(33):8089–8089, 2000.

[206] W. J. Wedemeyer, Y. A. Arnautova, J. Pillardy, R. J. Wawak, C. Czaplewski, and H. A. Scheraga. Reply to "comment on 'crystal structure prediction by global optimization as a tool for evaluating potentials: role of the dipole moment correction term in successful predictions'" by b. p. van eijck and j. kroon. *The Journal of Physical Chemistry B*, 104(33):8090–8092, 2000.

[207] A. Warshel and S. Lifson. Consistent force field calculations. II. Crystal structures, sublimation energies, molecular and lattice vibrations, molecular conformations, and enthalpies of alkanes. *The Journal of Chemical Physics*, 53(2):582–594, 1970.

[208] D. E. Williams. Nonbonded potential parameters derived from crystalline aromatic hydrocarbons. *The Journal of Chemical Physics*, 45(10):3770–3778, 1966.

[209] D. M. Elking. Torque and atomic forces for cartesian tensor atomic multipoles with an application to crystal unit cell optimization. *Journal of Computational Chemistry*, 37(22):2067–2080, 2016.

[210] M. Born and Joseph E. Mayer. Zur Gittertheorie der Ionenkristalle. *Zeitschrift für Physik*, 75(1):1–18, 1932.

[211] M. J. van Vleet, A. J. Misquitta, A. J. Stone, and J. R. Schmidt. Beyond Born-Mayer: Improved models for short-range repulsion in ab initio force fields. *arXiv preprint arXiv:1606.00734*, 2016.

[212] K. T. Tang and J. P. Toennies. An improved simple model for the van der Waals potential based on universal damping functions for the dispersion coefficients. *The Journal of chemical physics*, 80(8):3726–3741, 1984.

[213] B. M. Axilrod and E. Teller. Interaction of the van der Waals type between three atoms. *The Journal of Chemical Physics*, 11(6):299–300, 1943.

[214] A. M. Reilly and A. Tkatchenko. 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*, 139(2):024705, 2013.

[215] N. Ramasubbu, R. Parthasarathy, and P. Murray-Rust. Angular preferences of intermolecular forces around halogen centers: preferred directions of approach of electrophiles and nucleophiles around carbon-halogen bond. *Journal of the American Chemical Society*, 108(15):4308–4314, 1986.

[216] S. C. Nyburg and C. H. Faerman. A revision of van der Waals atomic radii for molecular crystals: N, O, F, S, Cl, Se, Br and I bonded to carbon. *Acta Crystallographica Section B*, 41(4):274–279, Aug 1985.

[217] A. C. Legon. The halogen bond: an interim perspective. *Physical Chemistry Chemical Physics*, 12(28):7736–7747, 2010.

[218] H. L. Nguyen, P. N. Horton, M. B. Hursthouse, A. C. Legon, and D. W. Bruce. Halogen bonding: a new interaction for liquid crystal formation. *Journal of the American Chemical Society*, 126(1):16–17, 2004.

[219] J. P. M. Lommerse, A. J. Stone, R. Taylor, and F. H. Allen. The nature and geometry of intermolecular interactions between halogens and oxygen or nitrogen. *Journal of the American Chemical Society*, 118(13):3108–3116, 1996.

[220] A. Bondi. van der Waals volumes and radii. *The Journal of Physical Chemistry*, 68(3):441–451, 1964.

[221] D. E. Williams. Direct calculations of crystalline thermal expansion and molecular reorientation from nonbonded interatomic potential anharmonicity and thermal amplitudes. *Acta Crystallographica Section A*, 28(1):84–88, Jan 1972.

[222] D. E. Williams and D. J. Houpt. Fluorine nonbonded potential parameters derived from crystalline perfluorocarbons. *Acta Crystallographica Section B: Structural Science*, 42(3):286–295, 1986.

[223] D. S. Coombes, S. L. Price, D. J. Willock, and M. Leslie. Role of electrostatic interactions in determining the crystal structures of polar organic molecules. A distributed multipole study. *The Journal of Physical Chemistry*, 100(18):7352–7360, 1996.

[224] A. Abraha and D. E. Williams. Spherical and aspherical intermolecular force fields for sulfur allotropes. *Inorganic Chemistry*, 38(19):4224–4228, 1999.

[225] E. O. Pyzer-Knapp, H. P. G. Thompson, and G. M. Day. An optimized intermolecular force field for hydrogen bonded organic molecular crystals using atomic multipole electrostatics. *Acta Crystallographica Section B*, 72:477–487, 2016.

[226] A. K. Al-Matar and H. Binous. Vapor–liquid phase equilibrium diagram for uranium hexafluoride (UF6) using simplified temperature dependent intermolecular potential parameters (TDIP). *Journal of Radioanalytical and Nuclear Chemistry*, pages 1–16, 2016.

[227] A. Bondi. Thermal properties of molecular crystals. I. Heat Capacity and Thermal Expansion. *Journal of Applied Physics*, 37(13):4643–4647, 2004.

[228] G. M. Sheldrick. A short history of SHELX. *Acta Crystallographica Section A: Foundations of Crystallography*, 64(1):112–122, 2008.

[229] S. R. Hall, F. H. Allen, and I. D. Brown. The crystallographic information file (CIF): a new standard archive file for crystallography. *Acta Crystallographica Section A*, 47(6):655–685, Nov 1991.

[230] W. Smith. The minimum image convention in non-cubic MD cells, 1989. (This is an unpublished document that was sent to a CCP5 mailing list.).

[231] D. J. Pearce. An improved algorithm for finding the strongly connected components of a directed graph. Technical report, Technical report, Victoria University, Wellington, NZ, 2005.

[232] A. A. Hagberg, D. A. Schult, and P. J. Swart. Exploring network structure, dynamics, and function using NetworkX. In *Proceedings of the 7th Python in Science Conference (SciPy2008)*, pages 11–15, Pasadena, CA USA, August 2008.

[233] J. Hopcroft and R. Tarjan. Algorithm 447: Efficient algorithms for graph manipulation. *Commun. ACM*, 16(6):372–378, June 1973.

[234] L. P. Cordella, P. Foggia, C. Sansone, and M. Vento. An improved algorithm for matching large graphs. In *3rd IAPR-TC15 workshop on graph-based representations in pattern recognition*, pages 149–159, 2001.

[235] L. P. Cordella, P. Foggia, C. Sansone, and M. Vento. A (sub) graph isomorphism algorithm for matching large graphs. *IEEE transactions on pattern analysis and machine intelligence*, 26(10):1367–1372, 2004.

[236] A. V. Kazantsev, P. G. Karamertzanis, C. C. Pantelides, and C. S. Adjiman. CrystalOptimizer: An efficient algorithm for lattice energy minimization of organic crystals using isolated-molecule quantum mechanical calculations. *Process Systems Engineering: Volume 6: Molecular Systems Engineering*, 6, 2010.

[237] A. Kazantsev. *Molecular Flexibility in Crystal Structure Prediction*. PhD thesis, Imperial College London, 2011.

[238] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Physical review*, 140(4A):A1133, 1965.

[239] R. B. J. S. Krishnan, J. S. Binkley, R. Seeger, and J. A. Pople. Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions. *The Journal of Chemical Physics*, 72(1):650–654, 1980.

[240] A. D. McLean and G. S. Chandler. Contracted gaussian basis sets for molecular calculations. I. second row atoms, Z = 11–18. *The Journal of Chemical Physics*, 72(10):5639–5648, 1980.

[241] J. P. Perdew, K. Burke, and M. Ernzerhof. Generalized gradient approximation made simple. *Physical review letters*, 77(18):3865, 1996.

[242] J. P. Perdew, M. Ernzerhof, and K. Burke. Rationale for mixing exact exchange with density functional approximations. *The Journal of Chemical Physics*, 105(22):9982–9985, 1996.

[243] C. Lee, W. Yang, and R. G. Parr. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical review B*, 37(2):785, 1988.

[244] A. D. Becke. Density-functional thermochemistry. III. The role of exact exchange. *The Journal of chemical physics*, 98(7):5648–5652, 1993.

[245] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and Michael J. Frisch. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *The Journal of Physical Chemistry*, 98(45):11623–11627, 1994.

[246] A. D. Becke. Density-functional exchange-energy approximation with correct asymptotic behavior. *Physical review A*, 38(6):3098, 1988.

[247] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg. 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*, 132(15), 2010.

[248] A. Tkatchenko and M. Scheffler. Accurate molecular van der Waals interactions from ground-state electron density and free-atom reference data. *Physical review letters*, 102(7):073005, 2009.

[249] J. Binns, M. R. Healy, S. Parsons, and C. A. Morrison. Assessing the performance of density functional theory in optimizing molecular crystal structure parameters. *Acta Crystallographica Section B*, 70(2):259–267, Apr 2014.

[250] A. M. Reilly and A. Tkatchenko. van der Waals dispersion interactions in molecular materials: beyond pairwise additivity. *Chemical Science*, 6(6):3289–3301, 2015.

[251] A. Tkatchenko, R. A. DiStasio, R. Car, and M. Scheffler. Accurate and efficient method for many-body van der Waals interactions. *Phys. Rev. Lett.*, 108:236402, Jun 2012.

[252] A. Otero-de-la-Roza and E. R. Johnson. Van der Waals interactions in solids using the exchange-hole dipole moment model. *The Journal of Chemical Physics*, 136(17), 2012.

[253] A. Otero-de-la-Roza and E. R. Johnson. Many-body dispersion interactions from the exchange-hole dipole moment model. *The Journal of Chemical Physics*, 138(5), 2013.

[254] L. D. Landau and E. M. Lifshitz. Theory of elasticity, vol. 7. *Course of Theoretical Physics*, 3:109, 1986.

[255] M. J. P. Musgrave. The propagation of elastic waves in crystals and other anisotropic media. *Reports on Progress in Physics*, 22(1):74, 1959.

[256] M. J. P. Musgrave. Crystal acoustics, 1970.

[257] M. De Jong, W. Chen, T. Angsten, A. Jain, R. Notestine, A. Gamst, M. Sluiter, C. K. Ande, S. Van Der Zwaag, and J. J. *et al.* Plata. Charting the complete elastic properties of inorganic crystalline compounds. *Scientific data*, 2, 2015.

[258] R. Hill. The elastic behaviour of a crystalline aggregate. *Proceedings of the Physical Society. Section A*, 65(5):349, 1952.

[259] S. H. Walmsley. Lattice vibrations and elastic constants of molecular crystals in the pair potential approximation. *The Journal of Chemical Physics*, 48(4):1438–1444, 1968.

[260] D. Michalski, D. R. Swanson, and C. J. Eckhardt. Elasticity, bulk modulus, and mode Grüneisen parameters of the HPTB molecular crystal: computational investigation of a clathrate precursor. *The Journal of Physical Chemistry*, 100(22):9506–9511, 1996.

[261] G. M. Day, S. L. Price, and M. Leslie. Elastic constant calculations for molecular organic crystals. *Crystal Growth & Design*, 1(1):13–27, 2001.

[262] S. Speziale, H. Marquardt, and T. S. Duffy. Brillouin scattering and its application in geosciences. *Reviews in Mineralogy and Geochemistry*, 78(1):543–603, 2014.

[263] J. Zhao, J. M. Winey, Y. M. Gupta, and W. Perger. Elastic properties of molecular crystals using density functional calculations. *AIP Conference Proceedings*, 706(1):429–434, 2004.

[264] M. Råsander and M. A. Moram. On the accuracy of commonly used density functional approximations in determining the elastic constants of insulators and semiconductors. *The Journal of Chemical Physics*, 143(14), 2015.

[265] P. Debye. Zur theorie der spezifischen Wärmen. *Annalen der Physik*, 344(14):789–839, 1912.

[266] G. M. Day, S. L. Price, and M. Leslie. Atomistic calculations of phonon frequencies and thermodynamic quantities for crystals of rigid organic molecules. *The Journal of Physical Chemistry B*, 107(39):10919–10933, 2003.

[267] S. Califano. *Lattice dynamics and intermolecular forces*, volume 55. Academic Press, 1975.

[268] S. H. Walmsley. Basic theory of the lattice dynamics of molecular crystals, 1975.

[269] N. Neto, R. Righini, S. Califano, and S. H. Walmsley. Lattice dynamics of molecular crystals using atom—atom and multipole—multipole potentials. *Chemical Physics*, 29(1):167–179, 1978.

[270] S. Califano, V. Schettino, and N. Neto. *Lattice dynamics of molecular crystals*, volume 26. Springer-Verlag Berlin, 1981.

[271] R. Righini and S. Califano. The vibrational optical spectrum of ethylenediamine crystal. calculation of the one-phonon spectrum of the high-temperature crystal form. *Chemical Physics*, 17(1):45–57, 1976.

[272] S. H. Simon. *The Oxford Solid State Basics*. Oxford University Press, 1 edition, 2013.

[273] G. M. Day, J. A. Zeitler, W. Jones, T. Rades, and P. F. Taday. Understanding the influence of polymorphism on phonon spectra: Lattice dynamics calculations and terahertz spectroscopy of carbamazepine. *The Journal of Physical Chemistry B*, 110(1):447–456, 2006. PMID: 16471555.

[274] D. Tomerini. *A Computational analysis of the vibrational absorption of molecular solids in the terahertz range*. PhD thesis, University of Cambridge, 2012.

[275] C. Hättig. On the calculation of derivatives for Coulomb-type interaction energies and general anisotropic pair potentials. *Chemical physics letters*, 268(5):521–530, 1997.

[276] H. J. Monkhorst and J. D. Pack. Special points for Brillouin-zone integrations. *Physical review B*, 13(12):5188, 1976.

[277] S. R. Whittleton, A. Otero-de-la Roza, and E. R. Johnson. The exchange-hole dipole dispersion model for accurate energy ranking in molecular crystal structure prediction. *Journal of Chemical Theory and Computation*, 2016.

[278] M. Rossi, P. Gasparotto, and M. Ceriotti. Anharmonic and quantum fluctuations in molecular crystals: A first-principles study of the stability of paracetamol. *Physical Review Letters*, 117(11):115702, 2016.

[279] Y. N. Heit, K. D. Nanda, and G. J. O. Beran. Predicting finite-temperature properties of crystalline carbon dioxide from first principles with quantitative accuracy. *Chemical Science*, 7(1):246–255, 2016.

[280] Y. N. Heit and G. J. O. Beran. How important is thermal expansion for predicting molecular crystal structures and thermochemistry at finite temperatures? *Acta Crystallographica Section B*, 72(4):514–529, Aug 2016.

[281] I. Křivý and B. Gruber. A unified algorithm for determining the reduced (Niggli) cell. *Acta Crystallographica Section A*, 32(2):297–298, Mar 1976.

[282] R. W. Grosse-Kunstleve, N. K. Sauter, and P. D. Adams. Numerically stable algorithms for the computation of reduced unit cells. *Acta Crystallographica Section A*, 60(1):1–6, Jan 2004.

[283] A. L. Spek. Single-crystal structure validation with the program *PLATON*. *Journal of Applied Crystallography*, 36(1):7–13, Feb 2003.

[284] A. L. Spek. Platon, an integrated tool for the analysis of the results of a single crystal structure determination. *Acta Crystallographica Section A: Foundations of Crystallography*, 46:c34–c34, 1990.

[285] A. Einstein. Die Plancksche Theorie der Strahlung und die Theorie der spezifischen Wärme. *Annalen der Physik*, 327(1):180–190, 1906.

[286] G. Gilat and G. Dolling. A new sampling method for calculating the frequency distribution function of solids. *Physics Letters*, 8(5):304–306, 1964.

[287] G. Gilat. Analysis of methods for calculating spectral properties in solids. *Journal of Computational Physics*, 10(3):432 – 465, 1972.

[288] G. Gilat and B. J. Alder, editors. *Methods in Computational Physics*, volume Vibrational Properties of Solids of *Advances in research and applications*, chapter Methods of Brillouin zone integration. Academic Press, 1976.

[289] M. Shubin and T. Sunada. Geometric theory of lattice vibrations and specific heat. *arXiv preprint math-ph/0512088*, 2005.

[290] E. Parzen. On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33(3):1065–1076, 09 1962.

[291] M. Rosenblatt. Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, 27(3):832–837, 1956.

[292] B. W. Silverman. *Density estimation for statistics and data analysis*, volume 26. CRC press, 1986.

[293] M. P. Wand and M. C. Jones. *Kernel smoothing*. Crc Press, 1994.

[294] D. E. Seaman and R. A Powell. An evaluation of the accuracy of kernel density estimators for home range analysis. *Ecology*, 77(7):2075–2085, 1996.

[295] V. A. Epanechnikov. Non-parametric estimation of a multivariate probability density. *Theory of Probability & Its Applications*, 14(1):153–158, 1969.

[296] P. Hall, S. J. Sheather, M. C. Jones, and J. S. Marron. On optimal data-based bandwidth selection in kernel density estimation. *Biometrika*, 78(2):263–269, 1991.

[297] S. J. Sheather and M. C. Jones. A reliable data-based bandwidth selection method for kernel density estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 683–690, 1991.

[298] M. C. Jones, J. S. Marron, and S. J. Sheather. A brief survey of bandwidth selection for density estimation. *Journal of the American Statistical Association*, 91(433):401–407, 1996.

[299] A. M. Karo and J. R. Hardy. Lattice dynamics of NaF. *Physical Review*, 181(3):1272, 1969.

[300] B. P. van Eijck. Ab initio crystal structure predictions for flexible hydrogen-bonded molecules. Part III. Effect of lattice vibrations. *Journal of Computational Chemistry*, 22(8):816–826, 2001.

[301] C. J. Stone. An asymptotically optimal window selection rule for kernel density estimates. *The Annals of Statistics*, pages 1285–1297, 1984.

[302] A. W. Bowman. An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, 71(2):353–360, 1984.

[303] P. Hall and J. S. Marron. Extent to which least-squares cross-validation minimises integrated square error in nonparametric density estimation. *Probability Theory and Related Fields*, 74(4):567–581, 1987.

[304] F. H. Allen. The cambridge structural database: a quarter of a million crystal structures and rising. *Acta Crystallographica Section B: Structural Science*, 58(3):380–388, 2002.

[305] S. Grimvall and G. Grimvall. Einsteinmodellen – en introduktion till fasta tillståndets fysik]. In *Kosmos 1971*, pages 29–36. Almqvist & Wiksell, 1971.

[306] V. Schomaker and K. N. Trueblood. On the rigid-body motion of molecules in crystals. *Acta Crystallographica Section B*, 24(1):63–76, Jan 1968.

[307] B. T. M. Willis and G. S. Pawley. The temperature factor of an atom in a rigid vibrating molecule. I. Isotropic thermal motion. *Acta Crystallographica Section A*, 26(2):254–259, Mar 1970.

[308] N. R. Werthamer. Self-consistent phonon formulation of anharmonic lattice dynamics. *Phys. Rev. B*, 1:572–581, Jan 1970.

[309] J. E. Turney, E. S. Landry, A. J. H. McGaughey, and C. H. Amon. Predicting phonon properties and thermal conductivity from anharmonic lattice dynamics calculations and molecular dynamics simulations. *Phys. Rev. B*, 79:064301, Feb 2009.

[310] R. A. Cowley. The lattice dynamics of an anharmonic crystal. *Advances in Physics*, 12(48):421–480, 1963.

[311] R. A. Cowley. Anharmonic crystals. *Reports on Progress in Physics*, 31(1):123, 1968.

[312] A. Otero-de-la Roza, D. Abbasi-Pérez, and V. Luaña. GIBBS2: a new version of the quasiharmonic model code. II. Models for solid-state thermodynamics, features and implementation. *Computer Physics Communications*, 182(10):2232–2248, 2011.

[313] F. D. Murnaghan. The compressibility of media under extreme pressures. *Proceedings of the National Academy of Sciences*, 30(9):244–247, 1944.

[314] O. L. Anderson. The volume dependence of thermal pressure in solids. *Journal of Physics and Chemistry of solids*, 58(2):335–343, 1997.

[315] F. H. Herbstein. On the mechanism of some first-order enantiotropic solid-state phase transitions: from Simon through Ubbelohde to Mnyukh. *Acta Crystallographica Section B*, 62(3):341–383, Jun 2006.

[316] E. V. Boldyreva, S. G. Arkhipov, T. N. Drebushchak, V. A. Drebushchak, E. A. Losev, A. A. Matvienko, V. S. Minkov, D. A. Rychkov, Y. V. Seryotkin, and J. *et al.* Stare. Isoenergetic polymorphism: The puzzle of tolazamide as a case study. *Chemistry–A European Journal*, 21(43):15395–15404, 2015.

[317] J. Bernstein, R. J. Davey, and J.-O. Henck. Concomitant polymorphs. *Angewandte Chemie International Edition*, 38(23):3440–3461, 1999.

[318] K. Kawakami. Reversibility of enantiotropically related polymorphic transformations from a practical viewpoint: thermal analysis of kinetically reversible/irreversible polymorphic transformations. *Journal of pharmaceutical sciences*, 96(5):982–989, 2007.

[319] A. Burger and R. Ramberger. On the polymorphism of pharmaceuticals and other molecular crystals. I. *Microchimica Acta*, 72(3-4):259–271, 1979.

[320] A. Burger and R. Ramberger. On the polymorphism of pharmaceuticals and other molecular crystals. II. *Microchimica Acta*, 72(3-4):273–316, 1979.

[321] A. Grunenberg, J.-O. Henck, and H. W. Siesler. Theoretical derivation and practical application of energy/temperature diagrams as an instrument in preformulation

studies of polymorphic drug substances. *International Journal of Pharmaceutics*, 129(1-2):147 – 158, 1996.

[322] L. Yu. Inferring thermodynamic stability relationship of polymorphs from melting data. *Journal of Pharmaceutical Sciences*, 84(8):966–974, 1995.

[323] S. A. Glickman and A. C. Cope. Structure of β-amino derivatives of α,β-unsaturated lactones and esters. *Journal of the American Chemical Society*, 67(6):1017–1020, 1945.

[324] M Kuhnert-Brandstätter and A Martinek. Über den Einfluß der Polymorphie auf die Löslichkeit von Arzneimitteln. *Microchimica Acta*, 53(5-6):909–919, 1965.

[325] R. Cameroni, G. Gamberini, V. Ferioli, and M. T. Bernabei. Polymorphism of ethylbiscumacetate. *Il Farmaco; edizione pratica*, 32(3):125–138, 1977.

[326] M. Kubota and G. O. Spessard. Glutaronitrile. The metastable modification. *The Journal of Physical Chemistry*, 70(3):941–942, 1966.

[327] I. Matsubara. Infrared spectrum and molecular configuration of glutaronitrile. *The Journal of Chemical Physics*, 35(1):373–374, 1961.

[328] C.-H. Gu, V. Young, and D. J. W. Grant. Polymorph screening: Influence of solvents on the rate of solvent-mediated polymorphic transformation. *Journal of Pharmaceutical Sciences*, 90(11):1878–1890, 2001.

[329] J. J. Gerber, J. G. van der Watt, and A. P. Lötter. Physical characterisation of solid forms of cyclopenthiazide. *International Journal of Pharmaceutics*, 73(2):137 – 145, 1991.

[330] M. Haranczyk and J. A. Sethian. Automatic structure analysis in high-throughput characterization of porous materials. *Journal of Chemical Theory and Computation*, 6(11):3472–3480, 2010.

[331] A. J. Jones, C. Ostrouchov, M. Haranczyk, and E. Iglesia. From rays to structures: Representation and selection of void structures in zeolites using stochastic methods. *Microporous and Mesoporous Materials*, 181:208–216, 2013.

[332] M. Haranczyk and J. A. Sethian. Navigating molecular worms inside chemical labyrinths. *Proceedings of the National Academy of Sciences*, 106(51):21472–21477, 2009.

[333] K. Theisen, B. Smit, and M. Haranczyk. Chemical hieroglyphs: Abstract depiction of complex void space topology of nanoporous materials. *Journal of Chemical Information and Modeling*, 50(4):461–469, 2010.

[334] T. F. Willems, C. H Rycroft, M. Kazi, J. C. Meza, and M. Haranczyk. Algorithms and tools for high-throughput geometry-based analysis of crystalline porous materials. *Microporous and Mesoporous Materials*, 149(1):134–141, 2012.

[335] R. L. Martin, D. D. Donofrio, J. A. Sethian, and M. Haranczyk *et al.* Accelerating analysis of void space in porous materials on multicore and GPU platforms. *International Journal of High Performance Computing Applications*, 26(4):347–357, 2012.

[336] L. J. Barbour. Crystal porosity and the burden of proof. *Chemical communications*, 11:1163–1168, 2006.

[337] A. Gavezzotti. The calculation of molecular volumes and the use of volume analysis in the investigation of structured media and of solid-state organic reactivity. *Journal of the American Chemical Society*, 105:5220–5225, 1983.

[338] S. Mecozzi and J. Rebek Jr. The 55% solution: A formula for molecular recognition in the liquid state. *Chemistry – A European Journal*, 4(6):1016–1022, 1998.

[339] K. Nakano, K. Sada, Y. Kurozumi, and M. Miyata. Importance of packing coefficients of host cavities in the isomerization of open host frameworks: Guest-size-dependent isomerization in cholic acid inclusion crystals with monosubstituted benzenes. *Chemistry – A European Journal*, 7(1):209–220, 2001.

[340] A. J. Cruz-Cabeza, S. Karki, L. Fabian, T. Friscic, G. M. Day, and W. Jones. Predicting stoichiometry and structure of solvates. *Chem. Commun.*, 46:2224–2226, 2010.

[341] D. E. Braun, T. Gelbrich, V. Kahlenberg, and U. J. Griesser. Solid state forms of 4-aminoquinaldine – from void structures with and without solvent inclusion to close packing. *CrystEngComm*, 17(12):2504–2516, 2015.

[342] D. E. Braun and U. J. Griesser. Why do hydrates (solvates) form in small neutral organic molecules? exploring the crystal form landscapes of the alkaloids brucine and strychnine. *Crystal Growth & Design*, 2016.

[343] M. L. Connolly. Computation of molecular volume. *Journal of the American Chemical society*, 107(5):1118–1124, 1985.

[344] Molinspiration Cheminformatics 2014. http://www.molinspiration.com/cgi-bin/properties.

[345] H. Küppers, F. Liebau, and A. L. Spek. The determination of pore volumes, pore shapes and diffusion paths in microporous crystals. *Journal of Applied Crystallography*, 39(3):338–346, Jun 2006.

[346] R. Ott, J. Bijma, C. Hemleben, and M. Signes. A computer method for estimating volumes and surface areas of complex structures consisting of overlapping spheres. *Mathematical and Computer Modelling*, 16(12):83 – 98, 1992.

[347] J. Buša, J. Džurina, E. Hayryan, S. Hayryan, C.-K. Hu, J. Plavka, I. Pokorný, J. Skřivánek, and M.-C. Wu. ARVO: A Fortran package for computing the solvent accessible surface area and the excluded volume of overlapping spheres via analytic equations. *Computer Physics Communications*, 165(1):59 – 96, 2005.

[348] E. L. Willighagen, R. Wehrens, P. Verwer, R. de Gelder, and L. M. C. Buydens. Method for the computational comparison of crystal structures. *Acta Crystallographica Section B*, 61(1):29–36, Feb 2005.

[349] D. W. M. Hofmann and L. Kuleshova. New similarity index for crystal structure determination from X-ray powder diagrams. *Journal of Applied Crystallography*, 38(6):861–866, Dec 2005.

[350] R. de Gelder, R. Wehrens, and J. A. Hageman. A generalized expression for the similarity of spectra: application to powder diffraction pattern classification. *Journal of Computational Chemistry*, 22(3):273–289, 2001.

[351] B. P. van Eijck and J. Kroon. Fast clustering of equivalent structures in crystal structure prediction. *Journal of Computational Chemistry*, 18(8):1036–1042, 1997.

[352] R. Hundt, J. C. Schön, and M. Jansen. CMPZ–an algorithm for the efficient comparison of periodic structures. *Journal of applied crystallography*, 39(1):6–16, 2006.

[353] B. B. Zvyagin. Polytypism of crystal structures. *Computers & Mathematics with Applications*, 16(5):569 – 591, 1988.

[354] C. Bonhomme, C. Gervais, F. Babonneau, C. Coelho, F. Pourpoint, T. Azaïs, S. E. Ashbrook, J. M. Griffin, J. R. Yates, F. Mauri, and C. J. Pickard. First-principles calculation of nmr parameters using the gauge including projector augmented wave method: A chemist's point of view. *Chemical Reviews*, 112(11):5733–5779, 2012. PMID: 23113537.

[355] S. C. Pérez, L. Cerioni, A. E. Wolfenson, S. Faudone, and S. L. Cuffini. Utilization of pure nuclear quadrupole resonance spectroscopy for the study of pharmaceutical crystal forms. *International Journal of Pharmaceutics*, 298(1):143 – 152, 2005.

[356] H. Hamaed, J. M. Pawlowski, B. F. T. Cooper, R. Fu, S. H. Eichhorn, and R. W. Schurko. Application of solid-state 35Cl NMR to the structural characterization of hydrochloride pharmaceuticals and their polymorphs. *Journal of the American Chemical Society*, 130(33):11056–11065, 2008.

[357] E. Salager, G. M. Day, R. S. Stein, C. J. Pickard, B. Elena, and L. Emsley. Powder crystallography by combined crystal structure prediction and high-resolution 1H solid-state NMR spectroscopy. *Journal of the American Chemical Society*, 132(8):2564–2566, 2010.

[358] J. Brus, J. Czernek, L. Kobera, M. Urbanova, S. Abbrent, and M. Husak. Predicting the crystal structure of decitabine by powder NMR crystallography: Influence of long-range molecular packing symmetry on NMR parameters. *Crystal Growth & Design*, 2016.

[359] F. Mauri, B. G. Pfrommer, and S. G. Louie. Ab initio theory of NMR chemical shifts in solids and liquids. *Physical review letters*, 77(26):5300, 1996.

[360] C. G. van de Walle and P. E. Blöchl. First-principles calculations of hyperfine parameters. *Phys. Rev. B*, 47:4244–4255, Feb 1993.

[361] T. Charpentier. The PAW/GIPAW approach for computing NMR parameters: A new dimension added to NMR study of solids. *Solid State Nuclear Magnetic Resonance*, 40(1):1 – 20, 2011.

[362] C. J. Pickard and F. Mauri. All-electron magnetic response with pseudopotentials: NMR chemical shifts. *Physical Review B*, 63(24):245101, 2001.

[363] P. E. Blöchl. Projector augmented-wave method. *Physical Review B*, 50(24):17953, 1994.

[364] S. J. Clark, M. D. Segall, C. J. Pickard, P. J. Hasnip, M. J. Probert, K. Refson, and M.C. Payne. First principles methods using CASTEP. *Z. Kristall.*, 220:567–570, 2005.

[365] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, and D. Ceresoli *et al.* QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *Journal of Physics: Condensed Matter*, 21(39):395502, 2009.

[366] W. van den Heuvel and A. Soncini. NMR chemical shift as analytical derivative of the Helmholtz free energy. *The Journal of chemical physics*, 138(5):054113, 2013.

[367] M. Bühl, M. Kaupp, O. L. Malkina, and V. G. Malkin. The DFT route to NMR chemical shifts. *Journal of Computational Chemistry*, 20(1):91–105, 1999.

[368] L. B. Casabianca and A. C. de Dios. Ab initio calculations of NMR chemical shifts. *The Journal of Chemical Physics*, 128(5):052201, 2008.

[369] J. Vaara. Theory and computation of nuclear magnetic resonance parameters. *Physical Chemistry Chemical Physics*, 9:5399–5418, 2007.

[370] J. R. Yates. Prediction of NMR J-coupling in solids with the planewave pseudopotential approach. *Magnetic Resonance in Chemistry*, 48(S1):S23–S31, 2010.

[371] J. Fraissard and T. Ito. 129Xe nmr study of adsorbed xenon: A new method for studying zeolites and metal-zeolites. *Zeolites*, 8(5):350 – 361, 1988.

[372] V. V. Terskikh, I. L. Moudrakovski, S. R. Breeze, S. Lang, C. I. Ratcliffe, J. A. Ripmeester, and A. Sayari. A general correlation for the 129Xe NMR chemical shift-pore size relationship in porous silica-based materials. *Langmuir*, 18(15):5653–5656, 2002.

[373] J. A. Ripmeester, C. I. Ratcliffe, and J. S. Tse. The nuclear magnetic resonance of 129 Xe trapped in clathrates and some other solids. *Journal of the Chemical Society, Faraday Transactions 1: Physical Chemistry in Condensed Phases*, 84(11):3731–3745, 1988.

[374] V.-V. Telkki, J. Lounila, and J. Jokisaari. Influence of diffusion on pore size distributions determined by xenon porometry. *Physical Chemistry Chemical Physics*, 8:2072–2076, 2006.

[375] V.-V. Telkki, J. Lounila, and J. Jokisaari. Behavior of acetonitrile confined to mesoporous silica gels as studied by 129Xe NMR: A novel method for determining the pore sizes. *The Journal of Physical Chemistry B*, 109(2):757–763, 2005. PMID: 16866438.

[376] V.-V. Telkki, J. Lounila, and J. Jokisaari. Xenon porometry: a novel method for the derivation of pore size distributions. *Magnetic Resonance Imaging*, 25(4):457 – 460, 2007.

[377] G. R. Desiraju. Crystal engineering: a holistic view. *Angewandte Chemie International Edition*, 46(44):8342–8356, 2007.

[378] D. E. Williams. Crystal packing of molecules. *Science*, 147(3658):605–606, 1965.

[379] A. Gavezzotti. *Molecular Aggregation*. Oxford University Press, 2007.

[380] T. Risthaus and S. Grimme. Benchmarking of london dispersion-accounting density functional theory methods on very large molecular complexes. *Journal of Chemical Theory and Computation*, 9(3):1580–1591, 2013.

[381] G. W. A. Welch, P. G. Karamertzanis, A. J. Misquitta, A. J. Stone, and S. L. Price. Is the induction energy important for modeling organic crystals? *Journal of Chemical Theory and Computation*, 4(3):522–532, 2008.

[382] T. G. Cooper, K. E. Hejczyk, W. Jones, and G. M. Day. Molecular polarization effects on the relative energies of the real and putative crystal structures of valine. *Journal of Chemical Theory and Computation*, 4(10):1795–1805, 2008.

[383] R. D. L. Johnstone, M. Ieva, A. R. Lennie, H. McNab, E. Pidcock, J. E. Warren, and S. Parsons. Pressure as a tool in crystal engineering: inducing a phase transition in a high-Z′ structure. *CrystEngComm*, 12:2520–2523, 2010.

[384] G. M. Day, J. Chisholm, N. Shan, W. D. S. Motherwell, and W. Jones. An assessment of lattice energy minimization for the prediction of molecular organic crystal structures. *Crystal Growth & Design*, 4(6):1327–1340, 2004.

[385] D. J. Carter and A. L. Rohl. Benchmarking calculated lattice parameters and energies of molecular crystals using van der Waals density functionals. *Journal of Chemical Theory and Computation*, 10(8):3423–3437, 2014.

[386] G. J. O. Beran and K. Nanda. Predicting organic crystal lattice energies with chemical accuracy. *The Journal of Physical Chemistry Letters*, 1(24):3480–3487, 2010.

[387] T. S. Totton, A. J. Misquitta, and M. Kraft. A first principles development of a general anisotropic potential for polycyclic aromatic hydrocarbons. *Journal of Chemical Theory and Computation*, 6(3):683–695, 2010.

[388] J. G. Brandenburg and S. Grimme. Dispersion corrected hartree–fock and density functional theory for organic crystal structure prediction. In *Prediction and Calculation of Crystal Structures*, pages 1–23. Springer, 2014.

[389] C. E. S. Bernardes and A. Joseph. Evaluation of the OPLS-AA force field for the study of structural and energetic aspects of molecular organic crystals. *The Journal of Physical Chemistry A*, 119(12):3023–3034, 2015. PMID: 25733134.

[390] P. Grančič, R. Bylsma, H. Meekes, and H. M. Cuppen. Evaluation of all-atom force fields for anthracene crystal growth. *Crystal Growth & Design*, 15(4):1625–1633, 2015.

[391] J. G. Brandenburg and S. Grimme. Organic crystal polymorphism: a benchmark for dispersion-corrected mean-field electronic structure methods. *Acta Crystallographica Section B*, 72(4):502–513, Aug 2016.

[392] J. S. Chickos and W. E. Acree. Enthalpies of sublimation of organic and organometallic compounds. *Journal of Physical and Chemical Reference Data*, 31(2):537–698, 2002.

[393] J. Moellmann and S. Grimme. DFT-D3 study of some molecular crystals. *The Journal of Physical Chemistry C*, 118(14):7615–7621, 2014.

[394] R. Sure, J. G. Brandenburg, and S. Grimme. Small atomic orbital basis set first-principles quantum chemical methods for large molecular and periodic systems: A critical analysis of error sources. *ChemistryOpen*, 5(2):94–109, 2016.

[395] S. Grimme, J. G. Brandenburg, C. Bannwarth, and A. Hansen. Consistent structures and interactions by density functional theory with small atomic orbital basis sets. *The Journal of Chemical Physics*, 143(5), 2015.

[396] J. G. Brandenburg and S. Grimme. Accurate modeling of organic molecular crystals by dispersion-corrected density functional tight binding (DFTB). *The Journal of Physical Chemistry Letters*, 5(11):1785–1789, 2014.

[397] M. T. Dove. *Structure and Dynamics: an atomic view of materials*. Oxford University Press, 2003.

[398] M. Sanquer and C. Ecolivet. *Elastic Constants in Molecular Crystals: Experiments and Intermolecular Potentials*. Elsevier: Amsterdam, 1987.

[399] M. Cossi, G. Scalmani, N. Rega, and V. Barone. New developments in the polarizable continuum model for quantum mechanical and classical calculations on molecules in solution. *The Journal of Chemical Physics*, 117(1):43–54, 2002.

[400] J. Tomasi, B. Mennucci, and R. Cammi. Quantum mechanical continuum solvation models. *Chemical Reviews*, 105(8):2999–3094, 2005.

[401] B. Mennucci, J. Tomasi, R. Cammi, J. R. Cheeseman, M. J. Frisch, F. J. Devlin, S. Gabriel, and P. J. Stephens. Polarizable continuum model (PCM) calculations of solvent effects on optical rotations of chiral molecules. *The Journal of Physical Chemistry A*, 106(25):6102–6113, 2002.

[402] R. Boese, D. Bläser, R. Latz, and A. Bäumen. Acetic acid at 40 K. *Acta Crystallographica Section C*, 55(2), Feb 1999.

[403] R. Boese, N. Niederprüm, D. Bläser, A. Maulitz, M. Y. Antipin, and P. R. Mallinson. Single-crystal structure and electron density distribution of ammonia at 160 K on the basis of X-ray diffraction data. *The Journal of Physical Chemistry B*, 101(30):5794–5799, 1997.

[404] A. Simon and K. Peters. Single-crystal refinement of the structure of carbon dioxide. *Acta Crystallographica Section B: Structural Crystallography and Crystal Chemistry*, 36(11):2750–2751, 1980.

[405] S. P. Kampermann, J. R. Ruble, and B. M. Craven. The charge-density distribution in hexamethylenetetramine at 120 K. *Acta Crystallographica Section B*, 50(6):737–741, Dec 1994.

[406] J.-L. Leviel, G. Auvert, and J.-M. Savariault. Hydrogen bond studies. A neutron diffraction study of the structures of succinic acid at 300 and 77 K. *Acta Crystallographica Section B*, 37(12):2185–2189, Dec 1981.

[407] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, and V. Barone *et al.* Gaussian 09 Revision A.02, 2009. Gaussian Inc. Wallingford CT 2009.

[408] A. J. Stone and M. Alderton. Distributed multipole analysis methods and applications. *Molecular Physics*, 100(1):221–233, 2002.

[409] C. M. Breneman and K. B. Wiberg. Determining atom-centered monopoles from molecular electrostatic potentials. the need for high sampling density in formamide conformational analysis. *Journal of Computational Chemistry*, 11(3):361–373, 1990.

[410] O. L. Anderson. *Equations of state of solids for geophysics and ceramic science*. Oxford University Press, 1995.

[411] A. Budzianowski, A. Olejniczak, and A. Katrusiak. Competing hydrogen-bonding patterns and phase transitions of 1,2-diaminoethane at varied temperature and pressure. *Acta Crystallographica Section B*, 62(6):1078–1089, Dec 2006.

[412] G. Wójcik, J. Holband, J. J. Szymczak, S. Roszak, and J. Leszczynski. Interactions in polymorphic crystals of *m*-nitrophenol as studied by variable-temperature X-ray diffraction and quantum chemical calculations. *Crystal Growth & Design*, 6(1):274–282, 2006.

[413] G. Wójcik, J. Giermańska, Y. Marqueton, and C. Ecolivet. Polarized infrared and raman spectra of low-frequency vibrations of two phases of crystalline *m*-nitrophenol. *Journal of Raman spectroscopy*, 22(7):375–381, 1991.

[414] G. Wójcik, B. Jakubowski, M. M. Szostak, K. Hołderna-Natkaniec, J. Mayer, and I. Natkaniec. Neutron diffraction and direct dilatometric studies of two polymorphs of meta-nitrophenol crystal. *physica status solidi (a)*, 134(1):139–150, 1992.

[415] C. C. Wilson. Variable temperature study of the crystal structure of paracetamol (p-hydroxyacetanilide), by single crystal neutron diffraction. *Zeitschrift für Kristallographie – Crystalline Materials*, 215(11):693–701, 2000.

[416] B. A. Kolesov, M. A. Mikhailenko, and E. V. Boldyreva. Dynamics of the intermolecular hydrogen bonds in the polymorphs of paracetamol in relation to crystal packing and conformational transitions: a variable-temperature polarized Raman spectroscopy study. *Physical Chemistry Chemical Physics*, 13(31):14243–14253, 2011.

[417] P. Espeau, P. Négrier, and Y. Corvis. Crystallographic and pressure-temperature state diagram approach for the phase behavior and polymorphism study of glutaric acid. *Crystal Growth & Design*, 13(2):723–730, 2013.

[418] S. Bhattacharya, V. G. Saraswatula, and B. K. Saha. Thermal expansion in alkane diacids - another property showing alternation in an odd-even series. *Crystal Growth & Design*, 13(8):3651–3656, 2013.

[419] J. Grip and E. J. Samuelsen. A raman study of crystalline glutaric acid. *Physica Scripta*, 29(6):556, 1984.

[420] L. M. Goldsmith and C. E. Strouse. Molecular dynamics in the solid state. The order-disorder transition of monoclinic sulfur. *Journal of the American Chemical Society*, 99(23):7580–7589, 1977.

[421] W. I. F. David, R. M. Ibberson, S. F. J. Cox, and P. T. Wood. Order–disorder transition in monoclinic sulfur: a precise structural study by high-resolution neutron powder diffraction. *Acta Crystallographica Section B*, 62(6):953–959, 2006.

[422] B. Eckert and R. Steudel. *Elemental Sulfur und Sulfur-Rich Compounds II*, chapter Molecular Spectra of Sulfur Molecules and Solid Sulfur Allotropes, pages 31–98. Springer Berlin Heidelberg, Berlin, Heidelberg, 2003.

[423] R. Srinivasa Gopalan, P. Kumaradhas, and G. U. Kulkarni. Structural phase transition in adipic acid. *Journal of Solid State Chemistry*, 148(1):129–134, 1999.

[424] H.-K. Fun and S. Chantrapromma. A triclinic polymorph of hexanedioic acid. *Acta Crystallographica Section E*, 65(3):o624, Mar 2009.

[425] H. Schödel, C. Näther, H. Bock, and F. Butenschön. Trimorphism of 2,2'-dipyridylamine: structures, phase transitions and thermodynamic stabilities. *Acta Crystallographica Section B*, 52(5):842–853, Oct 1996.

[426] S. Cherukuvada, R. Thakuria, and A. Nangia. Pyrazinamide polymorphs: Relative stability and vibrational spectroscopy. *Crystal Growth & Design*, 10(9):3931–3941, 2010.

[427] C. R. Groom and A. M. Reilly. Sixth blind test of organic crystal-structure prediction methods. *Acta Crystallographica Section B*, 70(4):776–777, Aug 2014.

[428] F. Mohamadi, N. G. J. Richards, W. C. Guida, R. Liskamp, M. Lipton, C. Caufield, G. Chang, T. Hendrickson, and W. C. Still. Macromodel – an integrated software system for modeling organic and bioorganic molecules using molecular mechanics. *Journal of Computational Chemistry*, 11(4):440–467, 1990.

[429] Schrodinger LLC, New York, NY, *MacroModel, V9.9.013*, 2014.

[430] W. L. Jorgensen, Maxwell D. S., and J. Tirado-Rives. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.*, 118:11225–11236, 1996.

[431] K. Shoemake. Uniform random rotations. In *Graphics Gems III*, pages 124–132. Academic Press Professional, Inc., 1992.

[432] A. J. Cruz Cabeza, E. Pidcock, G. M. Day, W. D. S. Motherwell, and W. Jones. Space group selection for crystal structure prediction of solvates. *CrystEngComm*, 9:556–560, 2007.

[433] N. L. Abraham and M. I. J. Probert. A periodic genetic algorithm with real-space representation for crystal structure and polymorph prediction. *Physical Review B*, 73(22):224104, 2006.

[434] M. Vasileiadis. *Calculation of the free energy of crystalline solids*. PhD thesis, Imperial College London, 2013.

[435] B. Gruber. The relationship between reduced cells in a general Bravais lattice. *Acta Crystallographica Section A*, 29(4):433–440, Jul 1973.

[436] M. Born. Thermodynamics of crystals and melting. *The Journal of Chemical Physics*, 7(8):591–603, 1939.

[437] M. Born. On the stability of crystal lattices. I. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 36, pages 160–172. Cambridge University Press, 1940.

[438] C. G. Broyden. Quasi-Newton methods and their application to function minimisation. *Mathematics of Computation*, pages 368–381, 1967.

[439] J. D. Head and M. C. Zerner. A Broyden-Fletcher-Goldfarb-Shanno optimization procedure for molecular geometries. *Chemical physics letters*, 122(3):264–270, 1985.

[440] M. G. Kendall. A new measure of rank correlation. *Biometrika*, 30(1-2):81–93, 1938.

[441] G. M. Day. Current approaches to predicting molecular organic crystal structures. *Crystallography Reviews*, 17(1):3–52, 2011.

[442] L. Yu, G. A. Stephenson, C. A. Mitchell, C. A. Bunnell, S. V. Snorek, J. J. Bowyer, T. B. Borchardt, J. G. Stowell, and S. R. Byrn. Thermochemistry and conformational polymorphism of a hexamorphic crystal system. *Journal of the American Chemical Society*, 122(4):585–591, 2000.

[443] C.-H. Gu and D. J. W. Grant. Estimating the relative stability of polymorphs and hydrates from heats of solution and solubility data. *Journal of pharmaceutical sciences*, 90(9):1277–1287, 2001.

[444] J. van de Streek and S. Motherwell. Searching the Cambridge Structural Database for polymorphs. *Acta Crystallographica Section B*, 61(5):504–510, Oct 2005.

[445] J. van de Streek. Searching the Cambridge Structural Database for the 'best' representative of each unique polymorph. *Acta Crystallographica Section B*, 62(4):567–579, Aug 2006.

[446] Z. F. Weng, W. D. S. Motherwell, and J. M. Cole. Tormat: a program for the automated structural alignment of molecular conformations. *Journal of Applied Crystallography*, 41(5):955–957, 2008.

[447] US EPA. 2015. Estimation programs interface suite for microsoft windows, v 4.11.

[448] K. G. Joback and R. C. Reid. Estimation of pure-component properties from group-contributions. *Chemical Engineering Communications*, 57(1-6):233–243, 1987.

[449] P. I. Gold and G. J. Ogle. Estimating thermophysical properties of liquids. 4. boiling freezing and triple-point temperatures. *Chemical Engineering*, 76(1):119, 1969.

[450] S. E. Stein and R. L. Brown. Estimation of normal boiling points from group contributions. *Journal of Chemical Information and Computer Sciences*, 34(3):581–587, 1994.

[451] M. Salahinejad, T. C. Le, and D. A. Winkler. Capturing the crystal: Prediction of enthalpy of sublimation, crystal lattice energy, and melting points of organic compounds. *Journal of Chemical Information and Modeling*, 53(1):223–229, 2013. PMID: 23215043.

[452] F. J. Anscombe. The transformation of Poisson, binomial and negative-binomial data. *Biometrika*, pages 246–254, 1948.

[453] X. Mei, , and C. Wolf. Formation of new polymorphs of acridine using dicarboxylic acids as crystallization templates in solution. *Crystal Growth & Design*, 4(6):1099–1103, 2004.

[454] D. Braga, F. Grepioni, L. Maini, P. P. Mazzeo, and K. Rubini. Solvent-free preparation of co-crystals of phenazine and acridine with vanillin. *Thermochimica Acta*, 507-508:1–8, 2010.

[455] G. Gautier and M. Debeau. Spectres de diffusion raman du soufre β monoclinique. *Spectrochimica Acta Part A: Molecular Spectroscopy*, 32(5):1007–1010, 1976.

[456] P. C. Upadhya, K. L. Nguyen, Y. C. Shen, J. Obradovic, K. Fukushige, R. Griffiths, L. F. Gladden, A. G. Davies, and E. H. Linfield. Characterization of crystalline phase-transformations in theophylline by time-domain terahertz spectroscopy. *Spectroscopy letters*, 39(3):215–224, 2006.

[457] B. Legendre and S. L. Randzio. Transitiometric analysis of solid II/solid I transition in anhydrous theophylline. *International Journal of Pharmaceutics*, 343(1-2):41–47, 2007.

[458] M. Karjalainen, S. Airaksinen, J. Rantanen, J. Aaltonen, and J. Yliruusi. Characterization of polymorphic solid-state changes using variable temperature x-ray powder diffraction. *Journal of Pharmaceutical and Biomedical Analysis*, 39(1-2):27–32, 2005.

[459] E. D. L. Smith, R. B. Hammond, M. J. Jones, K. J. Roberts, J. B. O. Mitchell, S. L. Price, R. K. Harris, D. C. Apperley, J. C. Cherryman, and R. Docherty. The determination of the crystal structure of anhydrous theophylline by X-ray powder diffraction with a systematic search algorithm, lattice energy calculations, and 13C and 15N solid-state NMR: A question of polymorphism in a given unit cell. *The Journal of Physical Chemistry B*, 105(24):5818–5826, 2001.

[460] F. T. Krogh. Efficient algorithms for polynomial interpolation and numerical differentiation. *Mathematics of Computation*, 24(109):185–190, 1970.

[461] P. G. Karamertzanis, P. Raiteri, M. Parrinello, M. Leslie, and S. L. Price. The thermal stability of lattice-energy minima of 5-fluorouracil: metadynamics as an aid to polymorph prediction. *The Journal of Physical Chemistry B*, 112(14):4298–4308, 2008. PMID: 18341322.

[462] B. P. van Eijck. Crystal structure predictions for disordered halobenzenes. *Physical Chemistry Chemical Physics*, 4:4789–4794, 2002.

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

[464] A. J. Cruz-Cabeza, G. M. Day, and W. Jones. Structure prediction, disorder and dynamics in a DMSO solvate of carbamazepine. *Physical Chemistry Chemical Physics*, 13:12808–12816, 2011.

[465] A. Usanmaz. The crystal structure of 1,3,5-triphenyl-1,3,5-perhydrotriazine-2,4,6-trione(phenylisocyanurate). *Acta Crystallographica Section B*, 35(5):1117–1119, May 1979.

[466] M. B. Mariyatra, K. Panchanatheswaran, J. N. Low, and C. Glidewell. Orthorhombic and monoclinic polymorphs of 1,3,5-triphenylperhydro-1,3,5-triazine-2,4,6-trione at 120 K: chains and sheets formed by C—H$\cdots\pi$(arene) hydrogen bonds. *Acta Crystallographica Section C*, 60(9):o682–o685, Sep 2004.

[467] A. P. Mendham, R. A. Palmer, B. S. Potter, T. J. Dines, M. J. Snowden, R. Withnall, and B. Z. Chowdhry. Vibrational spectroscopy and crystal structure analysis

of two polymorphs of the di-amino acid peptide cyclo (L-Glu-L-Glu). *Journal of Raman Spectroscopy*, 41(3):288–302, 2010.

[468] R. J. Staples and L. H. Sonderegger. Crystal structure of 4-phenylimidazole, C9H8N2. *Zeitschrift für Kristallographie-New Crystal Structures*, 216(1-4):323–324, 2001.

[469] R. M. Claramunt, M. D. Santa Maria, L. Infantes, F. H. Cano, and J. Elguero. The annular tautomerism of 4(5)-phenylimidazole. *J. Chem. Soc., Perkin Trans. 2*, pages 564–568, 2002.

[470] D. J. Nielsen, C. Pettinari, B. W. Skelton, and A. H. White. 4-Phenyl-1*H*-imidazole (a low-temperature redetermination), 1-benzyl-1*H*-imidazole and 1-mesityl-1*H*-imidazole. *Acta Crystallographica Section C*, 60(8):o542–o544, Aug 2004.

[471] M. Bakir, G. R. Harewood, A. Holder, I. Hassan, T. P. Dasgupta, P. Maragh, and M. Singh-Wilmot. 5-[(4-Methylphenyl)diazenyl]salicylaldehyde. *Acta Crystallographica Section E*, 61(6):o1611–o1613, Jun 2005.

[472] T. S. Basu Baul, S. Kundu, H. D. Arman, and E. R. T. Tiekink. An orthorhombic polymorph of 5-[(4-methylphenyl)diazenyl]salicylaldehyde. *Acta Crystallographica Section E*, 65(12):o3061, Dec 2009.

[473] T. Gelbrich, D. Rossi, and U. J. Griesser. Two polymorphs and the diethylammonium salt of the barbiturate eldoral. *Acta Crystallographica Section C: Crystal Structure Communications*, 68(2):o65–o70, 2012.

[474] M. Brandstätter-Kuhnert and M. Aepkers. Molekülverbindungen, mischkristallbildung und neue polymorphiefälle bei barbituraten. *Microchimica Acta*, 50(6):1055–1074, 1962.

[475] M. D. Eddleston, K. E. Hejczyk, E. G. Bithell, G. M. Day, and W. Jones. Polymorph identification and crystal structure determination by a combined crystal structure prediction and transmission electron microscopy approach. *Chemistry - A European Journal*, 19(24):7874–7882, 2013.

[476] G. M. Day, W. D. S. Motherwell, and W. Jones. Beyond the isotropic atom model in crystal structure prediction of rigid molecules: atomic multipoles versus point charges. *Crystal Growth & Design*, 5(3):1023–1033, 2005.

[477] A. T. Anghel, G. M. Day, and S. L. Price. A study of the known and hypothetical crystal structures of pyridine: why are there four molecules in the asymmetric unit cell? *CrystEngComm*, 4:348–355, 2002.

[478] A. J. Cruz-Cabeza, G. M. Day, and W. Jones. Towards prediction of stoichiometry in crystalline multicomponent complexes. *Chemistry - A European Journal*, 14(29):8830–8836, 2008.

[479] A. Gavezzotti. A molecular dynamics test of the different stability of crystal polymorphs under thermal strain. *Journal of the American Chemical Society*, 122(43):10724–10725, 2000.

[480] P. Raiteri, R. Martoňák, and M. Parrinello. Exploring polymorphism: The case of benzene. *Angewandte Chemie International Edition*, 44(24):3769–3773, 2005.

[481] C. B. Sayre, S. A. Anderson, and R. F. Becker *et al.* Effect of post-emergence application of pyrazon on yield and weed control in red beets, 1964. In *Proceedings of the Northeastern Weed Control Conference, 1965*, pages 17–19, 1965.

[482] G. Buttiglieri, M. Peschka, T. Frömel, J. Müller, F. Malpei, P. Seel, and T. P. Knepper. Environmental occurrence and degradation of the herbicide n-chloridazon. *Water research*, 43(11):2865–2873, 2009.

[483] F. Ahmadi, N. Jamali, S. Jahangard-Yekta, B. Jafari, S. Nouri, F. Najafi, and M. Rahimi-Nasrabadi. The experimental and theoretical QM/MM study of interaction of chloridazon herbicide with ds-DNA. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 79(5):1004 – 1012, 2011.

[484] R. A. Jacobson, M. S. Virant, and R. G. Baughman. Crystal and molecular structure of herbicides. 1. Pyrazon. *Journal of Agricultural and Food Chemistry*, 26(2):511–512, 1978.

[485] J. Dennis Jr. and J. Moré. Quasi-Newton methods, motivation and theory. *SIAM Review*, 19(1):46–89, 1977.

[486] M. Leslie. *Advanced Users DMACRYS manual. Manual release July 2013, referring to Program release DMACRYS 2.0.8*. University College London, 2013. http://www.chem.ucl.ac.uk/cposs/dmacrys/manuals/dmacrys.pdf.

[487] T. Steiner. Frequency of Z' values in organic and organometallic crystal structures. *Acta Crystallographica Section B*, 56(4):673–676, Aug 2000.

[488] H. M. Powell. The structure of molecular compounds. Part IV. Clathrate compounds. *Journal of the Chemical Society (Resumed)*, pages 61–73, 1948.

[489] D. D. MacNicol, J. J. McKendrick, and D. R. Wilson. Clathrates and molecular inclusion phenomena. *Chemical Society Reviews*, 7(1):65–87, 1978.

[490] L. Mandelcorn. Clathrates. *Chemical Reviews*, 59(5):827–839, 1959.

[491] Yu A. Dyadin, I. S. Terekhova, T. V. Rodionova, and D. V. Soldatov. Half-century history of clathrate chemistry. *Journal of Structural Chemistry*, 40(5):645–653, 1999.

[492] H. F. Clausen, M. R. V. Jørgensen, S. Cenedese, M. S. Schmøkel, M. Christensen, Y.-S. Chen, G. Koutsantonis, J. Overgaard, M. A. Spackman, and B. B. Iversen. Host perturbation in a β-hydroquinone clathrate studied by combined x-ray/neutron charge-density analysis: Implications for molecular inclusion in supramolecular entities. *Chemistry–A European Journal*, 20(26):8089–8098, 2014.

[493] J.-W. Lee, S.-P. Kang, and J.-H. Yoon. Competing occupation of guest molecules in hydroquinone clathrates formed from binary C2H4 and CH4 gas mixtures. *The Journal of Physical Chemistry C*, 118(14):7705–7709, 2014.

[494] V. E. Zubkus, I. L. Shamovsky, and E. E. Tornau. Computer-simulation studies of β-quinol clathrate with various gases. Molecular interactions and crystal structure. *The Journal of chemical physics*, 97(11):8617–8627, 1992.

[495] E. Dendy S. Jr. and C. Koh. *Clathrate hydrates of natural gases*. CRC press, 2007.

[496] Yu. N. Kazankin, A. A. Palladiev, and A. M. Trofimov. Clathrate compound of meta-fluorophenol with xenon. *Zhurnal Obshchei Khimii*, 41(11):2268–2370, 1972. in Russian.

[497] Yu. N. Kazankin, A. A. Palladiev, and A. M. Trofimov. Clathrate compound of ortho-fluorophenol with xenon. *Zhurnal Obshchei Khimii*, 43(12):2670–2671, 1973. in Russian.

[498] T. Birchall, C. S. Frampton, G. J. Schrobilgen, and J. Valsdóttir. β-Hydroquinone xenon clathrate. *Acta Crystallogr. C*, 45(6):944–946, Jun 1989.

[499] M. Baias, J.-N. Dumez, P. H. Svensson, S. Schantz, G. M. Day, and L. Emsley. De novo determination of the crystal structure of a large drug molecule by crystal structure prediction-based powder NMR crystallography. *Journal of the American Chemical Society*, 135(46):17501–17507, 2013.

[500] M. Baias, C. M. Widdifield, J.-N. Dumez, H. P. G. Thompson, T. G. Cooper, E. Salager, S. Bassil, R. S. Stein, A. Lesage, and G. M. *et al.* Day. Powder crystallography of pharmaceutical materials by combined crystal structure prediction and solid-state 1H NMR spectroscopy. *Physical Chemistry Chemical Physics*, 15(21):8069–8080, 2013.

[501] M. Hildebrand, H. Hamaed, A. M. Namespetra, J. M. Donohue, R. Fu, I. Hung, Z. Gan, and R. W. Schurko. 35Cl solid-state NMR of HCl salts of active pharmaceutical ingredients: structural prediction, spectral fingerprinting and polymorph recognition. *CrystEngComm*, 16(31):7334–7356, 2014.

[502] W. Proctor and F. Yu. On the nuclear magnetic moments of several stable isotopes. *Phys. Rev.*, 81:20–30, Jan 1951.

[503] T. Ito and J. Fraissard. In *Proceedings of the 5th International Conference on Zeolites*, page 510. Heyden, London, 1981.

[504] T. G. Walker and W. Happer. Spin-exchange optical pumping of noble-gas nuclei. *Reviews of Modern Physics*, 69(2):629, April 1997.

[505] C. D. Keenan, M. M. Herling, R. Siegel, N. Petzold, C. R. Bowers, E. A. Rössler, J. Breu, and J. Senker. Porosity of pillared clays studied by hyperpolarized 129Xe NMR spectroscopy and Xe adsorption isotherms. *Langmuir*, 29(2):643–652, 2013.

[506] D. Stueber and C. J. Jameson. The chemical shifts of Xe in the cages of clathrate hydrate structures I and II. *J. Chem. Phys.*, 120(3):1560–1571, 2004.

[507] C. J. Jameson and D. Stueber. The nuclear magnetic resonance line shapes of Xe in the cages of clathrate hydrates. *J. Chem. Phys.*, 120(21):10200–10214, 2004.

[508] J. Autschbach and E. Zurek. Relativistic density-functional computations of the chemical shift of 129Xe in Xe@C60. *The Journal of Physical Chemistry A*, 107(24):4967–4972, 2003.

[509] J. Autschbach. Relativistic effects on NMR parameters. *High Resolution NMR Spectroscopy—Understanding Molecules and their Electronic Structures*, 3:69, 2013.

[510] M. Straka, P. Lantto, and J. Vaara. Toward calculations of the 129Xe chemical shift in Xe@C60 at experimental conditions: relativity, correlation, and dynamics. *J. Phys. Chem. A*, 112(12):2658–2668, 2008. cited By (since 1996)30.

[511] S. Standara, P. Kulhánek, R. Marek, and M. Straka. 129Xe NMR chemical shift in Xe@C60 calculated at experimental conditions: Essential role of the relativity, dynamics, and explicit solvent. *Journal of Computational Chemistry*, 34(22):1890–1898, 2013. cited By (since 1996)3.

[512] A. Kitaigorodsky. *Molecular crystals and molecules*, volume 29. Elsevier, 2012.

[513] J. L. Flippen, Jerome Karle, and I. L. Karle. Crystal structure of a versatile organic clathrate. 4-p-hydroxyphenyl-2,2,4-trimethylchroman (dianin's compound). *Journal of the American Chemical Society*, 92(12):3749–3755, 1970.

[514] J. T. A. Jones, T. Hasell, X. Wu, J. Bacsa, K. E. Jelfs, M. Schmidtmann, S. Y. Chong, D. J. Adams, A. Trewin, F. Schiffman, F. Cora, B. Slater, A. Steiner, G. M. Day, and A. I. Cooper. Modular and predictable assembly of porous organic molecular crystals. *Nature*, 474:367–371, 2011.

[515] E. O. Pyzer-Knapp, H. P. G. Thompson, F. Schiffmann, K. E. Jelfs, S. Y. Chong, M. A. Little, A. I. Cooper, and G. M. Day. Predicted crystal energy landscapes of porous organic cages. *Chem. Sci.*, 5:2235–2245, 2014.

[516] A. J. Cruz-Cabeza, G. M. Day, and W. Jones. Towards prediction of stoichiometry in crystalline multicomponent complexes. *Chem. Eur. J.*, 14(29):8830–8836, 2008.

[517] M. Head-Gordon, J. A. Pople, and M. J. Frisch. MP2 energy evaluation by direct methods. *Chemical Physics Letters*, 153(6):503 – 506, 1988.

[518] J. A. Chisholm and S. Motherwell. *COMPACK*: a program for identifying crystal structure similarity using distances. *Journal of Applied Crystallography*, 38(1):228–231, Feb 2005.

[519] A. V. Kazantsev, P. G. Karamertzanis, C. S. Adjiman, C. C. Pantelides, S. L. Price, P. T. A. Galek, G. M. Day, and A. J. Cruz-Cabeza. Successful prediction of a model pharmaceutical in the fifth blind test of crystal structure prediction. *International journal of pharmaceutics*, 418(2):168–178, 2011.

[520] P. Slavıček, R. Kalus, P. Paška, I. Odvárková, P. Hobza, and A. Malijevskỳ. State-of-the-art correlated ab initio potential energy curves for heavy rare gas dimers: Ar_2, Kr_2, and Xe_2. *The Journal of chemical physics*, 119(4):2102–2119, 2003.

[521] Q. Wen and W. Jäger. Rotational spectroscopic and ab initio studies of the Xe-H_2O van der Waals dimer. *The Journal of Physical Chemistry A*, 110(24):7560–7567, 2006.

[522] W. A Alexander and D. Troya. Theoretical study of the Ar-, Kr-, and Xe-CH_4,-CF_4 intermolecular potential-energy surfaces. *The Journal of Physical Chemistry A*, 110(37):10834–10843, 2006.

[523] Q. Wen and W. Jäger. Microwave spectra of the Xe–N_2 van der Waals complex: A comparison of experiment and theory. *The Journal of chemical physics*, 122(21):214310, 2005.

[524] V. Aquilanti, D. Ascenzi, D. Cappelletti, M. de Castro, and F. Pirani. Scattering of aligned molecules. The potential energy surfaces for the Kr-O_2 and Xe-O_2 systems. *The Journal of chemical physics*, 109(10):3898–3910, 1998.

[525] P. Hobza, O. Bludskỳ, H. L. Selzle, and E. W. Schlag. Ab initio second- and fourth-order Møller-Plesset study on structure, stabilization energy, and stretching vibration of benzene$\cdots$X (X= He, Ne, Ar, Kr, Xe) van der Waals molecules. *The Journal of chemical physics*, 97(1):335–340, 1992.

[526] TURBOMOLE V6.6 2014, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from
`http://www.turbomole.com`.

[527] P. Lantto and J. Vaara. (129)Xe chemical shift by the perturbational relativistic method: xenon fluorides. *J. Chem. Phys.*, 127(8):084312, 2007.

[528] M. Straka, P. Lantto, M. Rasanen, and J. Vaara. Theoretical predictions of nuclear magnetic resonance parameters in a novel organo-xenon species: chemical shifts and nuclear quadrupole couplings in HXeCCH. *J. Chem. Phys.*, 127(23):234314, 2007.

[529] P. Lantto, S. Standara, S. Riedel, J. Vaara, and M. Straka. Exploring new 129Xe chemical shift ranges in HXeY compounds: hydrogen more relativistic than xenon. *Phys. Chem. Chem. Phys.*, 14:10944–10952, 2012.

[530] J. Roukala, J. Zhu, C. Giri, K. Rissanen, P. Lantto, and V.-V. Telkki. Encapsulation of xenon by a self-assembled Fe4L6 metallosupramolecular cage. *J. Am. Chem. Soc.*, 137(7):2464–2467, 2015. PMID: 25671394.

[531] E. R. McNellis, J. Meyer, and K. Reuter. Azobenzene at coinage metal surfaces: Role of dispersive van der waals interactions. *Phys. Rev. B*, 80:205414, Nov 2009.

[532] C. J. Pickard and F. Mauri. All-electron magnetic response with pseudopotentials: NMR chemical shifts. *Phys. Rev. B*, 63:245101, 2001.

[533] J. R. Yates, C. J. Pickard, and F. Mauri. Calculation of NMR chemical shifts for extended systems using ultrasoft pseudopotentials. *Phys. Rev. B*, 76:024401, 2007.

[534] E. van Lenthe, E.-J. Baerends, and J. G. Snijders. Relativistic regular two-component Hamiltonians. *The Journal of chemical physics*, 99(6):4597–4610, 1993.

[535] S. K. Wolff, T. Ziegler, E. van Lenthe, and E. J. Baerends. Density functional calculations of nuclear magnetic shieldings using the zeroth-order regular approximation (ZORA) for relativistic effects: Zora nuclear magnetic resonance. *J. Chem. Phys.*, 110(16):7689–7698, 1999.

[536] M. Straka, P. Lantto, M. Rasanen, and J. Vaara. Theoretical predictions of nuclear magnetic resonance parameters in a novel organo-xenon species: chemical shifts and nuclear quadrupole couplings in HXeCCH. *J. Chem. Phys.*, 127(23):234314, 2007.

[537] G. te Velde, F. M. Bickelhaupt, E. J. Baerends, C. Fonseca Guerra, S. J. A. van Gisbergen, J. G. Snijders, and T. Ziegler. Chemistry with ADF. *Journal of Computational Chemistry*, 22(9):931–967, 2001.

[538] ADF2016, SCM, Theoretical Chemistry, Vrije Universiteit, Amsterdam, The Netherlands, http://www.scm.com.

[539] E. van Lenthe, E.-J. Baerends, and J. G. Snijders. Relativistic total energy using regular approximations. *The Journal of chemical physics*, 101(11):9783–9792, 1994.

[540] H. Saitô, I. Ando, and A. Ramamoorthy. Chemical shift tensor–the heart of NMR: insights into biological aspects of proteins. *Progress in nuclear magnetic resonance spectroscopy*, 57(2):181, 2010.

[541] Helen E. Maynard-Casely, Robert Hodyss, Morgan L. Cable, Tuan Hoang Vu, and Martin Rahm. A co-crystal between benzene and ethane: a potential evaporite material for Saturn's moon Titan. *IUCrJ*, 3(3):192–199, May 2016.

[542] M. C. Etter, J. C. MacDonald, and J. Bernstein. Graph-set analysis of hydrogen-bond patterns in organic crystals. *Acta Crystallographica Section B*, 46(2):256–262, Apr 1990.

[543] J. H. van der Waals. The statistical mechanics of clathrate compounds. *Transactions of the Faraday Society*, 52:184–193, 1956.

[544] V. R. Belosludov, Yu A. Dyadin, G. N. Chekhova, B. A. Kolesov, and S. I. Fadeev. Hydroquinone clathrates and the theory of clathrate formation. *Journal of inclusion phenomena*, 3(3):243–260, 1985.

[545] V. E. Schneider, E. E. Tornau, A. A. Vlasova, and A. A. Gurskas. Clathrate formation and phase transitions in β-hydroquinone. *Journal of inclusion phenomena*, 3(3):235–242, 1985.

[546] K. Hermansson. Host–guest interactions in an organic crystal: β-hydroquinone clathrate with Ne and HF guests. *The Journal of Chemical Physics*, 112(2):835–840, 2000.

[547] P. Dansas and P. Sixou. Liberté de reorientation et mise en ordre cooperative dans les clathrates d'hydroquinone. *Molecular Physics*, 31(5):1297–1318, 1976.

[548] L. von Szentpály, I. L. Shamovsky, R. Ghosh, V. E. Zubkus, and E. E. Tornau. Computer simulation studies on β- quinol clathrates with various gases. Interdependence of host and guest molecules. *The Journal of Chemical Physics*, 101(1):683–692, 1994.

[549] V. V. Gorbatchuk, A. G. Tsifarkin, I. S. Antipin, B. N. Solomonov, and A. I. Konovalov. Influence of the guest molecular size on the thermodynamic parameters of host–guest complexes between solid tert-butylcalix[4]arene and vapours of organic compounds. *Mendeleev communications*, 9(1):11–13, 1999.

[550] M. Born and E. Wolf. *Principles of optics: electromagnetic theory of propagation, interference and diffraction of light*. CUP Archive, 2000.

[551] A. Nemkevich, M. A. Spackman, and B. Corry. Simulations of guest transport in clathrates of Dianin's compound and hydroquinone. *Chemistry–A European Journal*, 19(8):2676–2684, 2013.

[552] L. X. Dang and B M. Pettitt. A theoretical study of the inclusion complexes of β-quinol. *The Journal of chemical physics*, 89(2):968–974, 1988.

[553] T. Matsuo, H. Suga, and S. Seki. Phase transitions in the quinol clathrate compounds. I. The quinol hydrogen cyanide clathrate compound. *Journal of the Physical Society of Japan*, 30(3):785–793, 1971.

[554] T. Matsuo. Phase transitions in the quinol clathrate compounds. II. The quinol methanol clathrate compound. *Journal of the Physical Society of Japan*, 30(3):794–805, 1971.

[555] S. Hirokawa, T. Imasaka, and T. Matsuo. Quantum effects on the orientational ordering of H_2S and D_2S molecules enclosed in β-quinol clathrate. *Journal of the Physical Society of Japan*, 63(2):593–601, 1994.

[556] A. Nemkevich, M. A. Spackman, and B. Corry. Mechanism of concerted hydrogen bond reorientation in clathrates of Dianin's compound and hydroquinone. *Journal of the American Chemical Society*, 133(46):18880–18888, 2011. PMID: 21988055.

[557] J. S. Tse, D. D. Klug, J. Y. Zhao, W. Sturhahn, E. E. Alp, J. Baumert, C. Gutt, M. R. Johnson, and W. Press. Anharmonic motions of Kr in the clathrate hydrate. *Nature materials*, 4(12):917–921, 2005.

[558] Q. Zhu, A. G Shtukenberg, D. J. Carter, T.-Q. Yu, J. Yang, M. Chen, P. Raiteri, A. R. Oganov, B. Pokroy, and I. Polishchuk *et al.* Resorcinol crystallization from the melt: A new ambient phase and new "riddles". *Journal of the American Chemical Society*, 138(14):4881–4889, 2016.

[559] M. Mastalerz and I. M. Oppel. Rational construction of an extrinsic porous molecular crystal with an extraordinary high specific surface area. *Angewandte Chemie International Edition*, 51(21):5252–5255, 2012.

[560] A. Pulido, L. Chen, T. Kaczorowski, D. Holden, M. A. Little, and S. Y. Chong *et al.* Additional computational data (related to "functional materials discovery using energy–structure–function maps" manuscript). *Nature Chemistry*, 2017.

[561] P. van der Sluis and A. L. Spek. BYPASS: an effective method for the refinement of crystal structures containing disordered solvent regions. *Acta Crystallographica Section A*, 46(3):194–201, Mar 1990.

[562] G. Gilat and L. J. Raubenheimer. Accurate numerical method for calculating frequency-distribution functions in solids. *Physical Review*, 144(2):390, 1966.

[563] M. C. Etter. Encoding and decoding hydrogen-bond patterns of organic compounds. *Accounts of Chemical Research*, 23(4):120–126, 1990.

[564] M. C. Etter. Hydrogen bonds as design elements in organic chemistry. *The Journal of Physical Chemistry*, 95(12):4601–4610, 1991.

[565] J. E. Mock, J. E. Myers, and E. A. Trabant. Crystallization of the rare-gas clathrates. *Industrial & Engineering Chemistry*, 53(12):1007–1010, 1961.

[566] L. J. Raubenheimer and G. Gilat. Accurate numerical method of calculating frequency distribution functions in solids. II. Extension to hcp crystals. *Physical Review*, 157(3):586, 1967.

[567] A. R. Oganov, S. Ono, Y. Ma, C. W. Glass, and A. Garcia. Novel high-pressure structures of $MgCO_3$, $CaCO_3$ and CO_2 and their role in earth's lower mantle. *Earth and Planetary Science Letters*, 273(1):38–47, 2008.

[568] G. Gao, A. R. Oganov, P. Li, Z. Li, H. Wang, T. Cui, Y. Ma, A. Bergara, A. O. Lyakhov, and T. Iitaka *at al.* High-pressure crystal structures and superconductivity of Stannane (SnH4). *Proceedings of the National Academy of Sciences*, 107(4):1317–1320, 2010.